



**UNIVERSIDAD
DE ANTIOQUIA**

**Caracterización de los espacios e instancias de participación ciudadana en Antioquia:
Aportes desde el análisis exploratorio de datos y machine learning**

Cristian David Alvarez Urrea

Trabajo de grado presentado para optar al título de Politólogo

Asesor

Juan Esteban Lopera Morales, Magíster (MSc) en Ciencia Política

Universidad de Antioquia
Facultad de Derecho y Ciencias Políticas
Ciencia Política
Medellín, Antioquia, Colombia
2025



UNIVERSIDAD DE ANTIOQUIA

Cita

(Alvarez Urrea, 2025)

Referencia

Estilo APA 7 (2020)

Alvarez Urrea, C. D. (2025). *Caracterización de los espacios e instancias de participación ciudadana en Antioquia: Aportes desde el análisis de datos y machine learning* [Informe de práctica]. Universidad de Antioquia, Medellín, Colombia.



Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Tabla de contenido

Resumen	10
Abstract	11
Introducción	12
SECCIÓN 1. PRESENTACIÓN DEL ESCENARIO DE PRÁCTICA, ACTIVIDADES REALIZADAS Y CONTENIDOS TEÓRICO-METODOLÓGICOS DE LA SISTEMATIZACIÓN.....	13
1.1. Secretaría de Participación y Cultura Ciudadana	13
1.1.1. Observatorio de Participación Ciudadana de Antioquia – Política Pública de Participación Ciudadana y Democrática	14
1.1.2. Caracterización de Espacios e Instancias de Participación Ciudadanas Municipales en el Departamento	15
1.2. Pertinencia de la sistematización.....	16
1.3. Pregunta de sistematización	17
1.4. Objeto de la Sistematización	18
1.5. Objetivos de la Sistematización	18
1.6. Pasos Metodológicos en la Sistematización de la Experiencia de Practica	19
1.7. Referentes Conceptuales Empleados.....	20
SECCIÓN 2. DESARROLLO DE LA SISTEMATIZACIÓN.....	27
2.1. Descripción de procesos – Caracterización de Espacios e Instancias de Participación Ciudadana.....	27
2.1.1. El diseño del instrumento de recolección de información	27
2.1.2. El diseño de las herramientas para el análisis	31
2.1.3. El diseño y publicación de los resultados	33
2.2. Ruta de trabajo diseñada e implementada para el análisis. Cualificación de la gestión de información dentro del observatorio	36
2.2.1. Preparación de los Datos.....	36

Construcción de Data Set:	36
Construcción de la MetaData de Espacios e Instancias de Participación Ciudadana:	38
2.2.2. Arquitectura de Datos	39
Estandarización de la información:	39
Construcción de bases de datos:	40
2.2.3. Análisis exploratorio de datos (EDA)	42
Aplicación de EDA a la Caracterización de Espacios e Instancias de Participación con Python:	43
2.2.4. Aplicación del EDA para Determinar las Necesidades Formativas en el Departamento:	47
Necesidades formativas por subregión:	51
2.2.5. Influencia Presencia de Espacios e Instancias de Participación Ciudadana Concretos en las Respuestas de Necesidades Formativas de los Municipios del Departamento – Tablas de Contingencia.	54
2.2.6. Espacios e Instancias de Participación Ciudadana, Búsqueda de Patrones y Relaciones en los Municipios del Departamento	63
2.2.7. Espacios e Instancias de Participación – Temáticas Formativas - Tablas de Contingencia	65
2.3. Aplicación de Árboles de Decisión de Clasificación para Detección de Patrones – Machine Learning	67
2.4. Modelos Robustos para Detección de Patrones - Random forest	72
2.5. Machine learning para Detectar Tipologías de Municipios - Clusters kmodes.	74
2.6. Análisis y Patrones Detectados de la Aplicación de EDA y Machine Learning.....	79
2.6.1. Análisis Necesidades Formativas Departamento y Subregiones	79
2.6.2. Análisis. Influencia de la Presencia de Espacios e Instancias de Participación Ciudadana en las Respuestas de Necesidades Formativas de los Municipios del Departamento	81
2.6.3. Análisis. Espacios e Instancias de Participación Ciudadana, Búsqueda de Patrones y Relaciones en los Municipios del Departamento:	83

2.6.4.	Arbol de Decision y Random Forest	84
2.6.5.	Análisis. Clasificación de Municipios por Tipologías – Clusters kmodos.....	86
2.7.	Reflexión Critica	89
2.7.1.	Analítica de Datos y Análisis Exploratorio de Datos	89
2.8.	Necesidades Formativas y Encuentros de Capacitación – Cualificación de Inferencias	90
2.9.	Modelos de Machine Learning, Pertinencia Espacios/Instancias – Temáticas Formativas	91
2.10.	Recomendaciones	93
2.10.1.	Proceso de Caracterización de Espacios e Instancias de Participación:	93
2.10.2.	Integración de Herramientas y Métodos de Analítica de Datos y Machine Learning	95
	Referencias	96
	Anexos.....	100

Lista de tablas

Tabla 1 Cobertura de la encuesta de caracterización de espacios e instancias de participación del departamento de Antioquia	45
Tabla 2: Numeración necesidades formativas.....	47
Tabla 3: Lista códigos de preguntas por espacios e instancias de participación presentes o ausentes por municipio.....	55
Tabla 4: Instancias o espacios priorizados para analizar	57
Tabla 5: Clusters – tipologías de municipios y respuesta a temáticas formativas	78
Tabla 6: Temáticas formativas óptimas para cada subregión – pregunta PC_40_1	80
Tabla 7: Clusters de municipios por KModes	87

Lista de figuras

Figura 1: Función para importar la librería de pandas	43
Figura 2: Función para importar archivos a Google Colab desde la computadora.....	44
Figura 3: Función para el alojamiento de los datos en un Dataframe utilizando pandas.....	44
Figura 4: Función para la creación de la tabla de municipios por subregión	44
Figura 5: Resultado de la función tabla de municipios por subregión Google Colab.	45
Figura 6: Mapa de Municipios encuestados por la caracterización de espacios e instancias de participación	46
Figura 7: Función porcentajes de respuestas necesidades formativas priorización #1.	48
Figura 8: Resultado de la función porcentajes de respuestas necesidades formativas priorización #1.	48
Figura 9: Función Creación de las etiquetas de datos necesidades formativas departamento	49
Figura 10: Función Creación del grafico circular distribución necesidades formativas #1 Departamento - Porcentajes	49
Figura 11: Gráfico Distribución necesidades formativas #1 – Departamento – Porcentajes	50
Figura 12: Gráfico Distribución necesidades formativas #2 Departamento – Porcentajes	50
Figura 13: Gráfico Distribución necesidades formativas #3 Departamento – Porcentajes	51
Figura 14: Función división de los porcentajes de los municipios por subregión - #1.....	52
Figura 15; Función creación de la tabla de porcentajes de los municipios por subregión - #1 ...	52
Figura 16 Función creación del gráfico Distribución necesidades #1 Subregiones – porcentajes	52
Figura 17: Gráfico distribución necesidades #1 Subregiones - Porcentajes	53
Figura 18: Gráfico distribución necesidades #2 Subregiones - Porcentajes	53
Figura 19: Gráfico distribución de necesidades #3 Subregiones – Porcentajes.....	54
Figura 20: Función preparación y comprobación de los Dataframes cargados PC_40_1	58

Figura 21: Función revisión y consolidación de las columnas entre P_1 a P_38	58
Figura 22: Función designación columna P_1 a Variable IEN_1.....	59
Figura 23: Función creación de la tabla de contingencia Filas: 0/1 de la instancia; columnas: categorías de necesidad	59
Figura 24: Función conversión del conteo en porcentaje por fila (porcentaje dentro de P_1=0 y P_1=1)	59
Figura 25: Función visualización de tablas de contingencia IEN_Conceo_PC_40_1:	60
Figura 26: Función creación de gráfico de barras apiladas para las tablas de contingencia P_1 – PC_40_1:.....	60
Figura 27: Gráfico distribución de PC_40_1 según (P_1) en %.....	60
Figura 28: Gráfico distribución de PC_40_2 según (P_1) en %.....	61
Figura 29: Gráfico distribución de PC_40_1 según (P_15) en %.....	61
Figura 30: Gráfico distribución de PC_40_1 según (P_33) en %.....	62
Figura 31: Gráfico distribución de PC_40_1 según (P_34) en %.....	62
Figura 32: Gráfico distribución de PC_40_1 según (P_36) en %.....	63
Figura 33: Función creación de los Dataframes de presencia y ausencia de espacios e instancias de participación	63
Figura 34: Función creación de la tabla de ausencia y presencia de espacios e instancias de participación	64
Figura 35: Gráfico porcentaje de presencia y ausencia de espacios e instancias de participación (P_1 a P_38).....	64
Figura 36: Función creación de la variable Inst_Cols	65
Figura 37: Función matriz Presencia porcentual de instancias y espacios (P_1 a P_38) según necesidad formativa (PC_40_1)”	66
Figura 38: Función creación del mapa de calor “Presencia porcentual de instancias y espacios (P_1 a P_38) según necesidad formativa (PC_40_1)”	66
Figura 39: Gráfico mapa de calor presencia porcentual de instancias y espacios (P_1 a P_38) según necesidades formativas (PC_40_1).....	67
Figura 40: Función importación librerías sklearn para árboles de decisión	69

Figura 41: Función definición de variables x, y para árbol de decisión	69
Figura 42: Función división de datos para entrenamiento y test – árbol de decisión	70
Figura 43: Función creación árbol de decisión para espacios e instancias de participación ciudadana – temáticas formativas	70
Figura 44: Función precisión del modelo – árbol de decisión espacios e instancias de participación – temáticas formativas	70
Figura 45: Gráfico árbol de decisión ausencia y presencia de espacios e instancias de participación ciudadana y temáticas formativas.....	71
Figura 46: Función importación de las librerías para creación de modelos Random Forest	72
Figura 47: Función creación del modelo Random Forest para espacios e instancias de participación – temáticas formativas.....	73
Figura 48: Función #26 Precisión Random Forest conjunto train y test.....	73
Figura 49: Función instalación e importación del paquete Kmodes	75
Figura 50: Función creación del modelo Kmodes para los municipios del departamento	75
Figura 51: Función creación de la columna Clusters_Kmodes en el dataframe.....	76
Figura 52: Función proporción de municipios por clusters Kmodes	76
Figura 53: Función centroides – modas clusters Kmode	76
Figura 54: Función municipios del departamento por clusters de kmode	77
Figura 55: Función respuesta de los clusters para tematicas formativas #1	78

Resumen

El presente informe de sistematización de experiencias tiene como fin evidenciar el trabajo realizado durante la práctica académica de excelencia realizada en el Observatorio de Participación Ciudadana de la Secretaría de Participación y Cultura Ciudadana de la Gobernación de Antioquia, específicamente en el proceso de Caracterización de Espacios e Instancias de Participación Ciudadana realizada durante el segundo semestre del año 2024. El objetivo de la sistematización giró en torno a la necesidad de cualificar inferencias para la toma de decisiones informadas en el ámbito público, para ello se presenta un flujo de trabajo para la cualificación de los procesos internos, partiendo desde la preparación de los datos hasta el análisis exploratorio de datos y la aplicación de modelos de machine learning para predicciones y clasificaciones, todo ello con el fin de cualificar las inferencias resultantes de los ejercicios emprendidos por el Observatorio. Además, se hace presente una reflexión crítica sobre los hallazgos, y los métodos utilizados y sugeridos y se culmina con recomendaciones para futuras iniciativas emprendidas por el Observatorio de Participación.

Palabras clave: cualificación de inferencias, participación ciudadana, gobernación de Antioquia, análisis exploratorio de datos, machine learning, observatorio de participación ciudadana

Abstract

This systematization report aims to present the work carried out during the academic excellence internship conducted at the Citizen Participation Observatory of the Secretariat of Participation and Citizen Culture of the Government of Antioquia, specifically within the process of Characterization of Spaces and Instances of Citizen Participation developed during the second semester of 2024. The objective of the systematization focused on the need to enhance the quality of inferences for informed decision-making in the public sphere. To this end, a workflow is proposed for the improvement of internal processes, ranging from data preparation to exploratory data analysis and the application of machine learning models for prediction and classification. These actions seek to strengthen the quality of the inferences derived from the exercises undertaken by the Observatory. Additionally, a critical reflection on the findings and on the methods used and proposed is presented, concluding with recommendations for future initiatives developed by the Citizen Participation Observatory.

Keywords: inference qualification, citizen participation, Government of Antioquia, exploratory data analysis, machine learning, Citizen Participation Observatory.

Introducción

El presente informe de sistematización de experiencias tiene como fin evidenciar el trabajo realizado durante la práctica académica de excelencia realizada en el Observatorio de Participación Ciudadana de la Secretaría de Participación y Cultura Ciudadana de la Gobernación de Antioquia, específicamente en el proceso de Caracterización de Espacios e Instancias de Participación Ciudadana realizada durante el segundo semestre del año 2024.

El texto está dividido en dos secciones, tal cual lo pide el reglamento de trabajos de grado del Pregrado en Ciencia Política para el caso de sistematización de experiencias de práctica, la primera hace una presentación del escenario de práctica; además se plantea la pregunta, el objeto y los objetivos de la sistematización asociados a la necesidad de cualificar inferencias para la toma de decisión informada en el ámbito público; así como las referencias teórico- metodológicas que guían la investigación.

La segunda sección desarrolla la propuesta de sistematización, allí se hace una descripción de los procesos internos de gestión de la información dentro del Observatorio para presentar el lugar en donde se ubica la sistematización; así como se presenta el flujo de trabajo para la cualificación de los procesos internos, partiendo desde la preparación de los datos por medio de estandarización y creación de data sets hasta el análisis exploratorio de datos con tablas de contingencia y gráficos, y la aplicación de modelos de machine learning como árboles de decisión, random forest para predicciones y clusterinng para clasificar, todo ello con el fin de cualificar las inferencias resultantes de los ejercicios emprendidos por el Observatorio. Aunado esto, igualmente se hace una reflexión crítica sobre los hallazgos, y los métodos utilizados y sugeridos, además se culminará con recomendaciones para futuras iniciativas. Siendo así, se dará paso a la descripción del lugar de la práctica académica, la Secretaría de Participación y Cultura Ciudadana.

SECCIÓN 1. PRESENTACIÓN DEL ESCENARIO DE PRÁCTICA, ACTIVIDADES REALIZADAS Y CONTENIDOS TEÓRICO-METODOLÓGICOS DE LA SISTEMATIZACIÓN

1.1. Secretaria de Participación y Cultura Ciudadana

La Participación Ciudadana desempeña un papel fundamental en la vida en sociedad, ella gesta nuevas formas de accionar en el territorio acorde a problemáticas, necesidades e intereses en común. Por ello, es indispensable tener en cuenta las diferentes opiniones, creencias y saberes de los sujetos que conocen la realidad y le apuestan a mejorar la calidad de vida de las personas a través de sus intervenciones, dichas intervenciones deben ser acompañadas y guiadas por entes o herramientas orientadas a fortalecer la apropiación y uso de la información y el conocimiento relacionado a los territorios con diversos insumos para la toma de decisiones informadas y el planteamiento creativo de alternativas de actuación sobre los problemas asociados a la participación ciudadana.

Así pues, los entes estatales con el fin de cubrir este campo, han diseñado diferentes estrategias o lugares específicos para tal fin, en el caso del departamento de Antioquia, este con la evolución de la gestión pública en el territorio instauró la Secretaría de Participación y Cultura Ciudadana, siendo esta la Secretaría encargada entre otras cosas de organizar y fortalecer el desarrollo social del departamento en coordinación con los municipios, promoviendo tanto la convivencia pacífica como la gestión participativa del desarrollo. Además, dicha Secretaría coordina y concreta la adopción y ejecución de planes, programas y proyectos entre el sector público y privado, además, motiva a la participación ciudadana en el ciclo de la gestión pública intentando involucrar a la ciudadanía en la toma de decisiones y acciones públicas durante las etapas de diagnóstico, formulación, implementación, evaluación y seguimiento (Gobernación de Antioquia, sf). Siendo así, una de las apuestas más grandes de la Secretaría de Participación Ciudadana y la Ordenanza 025 de 2023 es el Observatorio de Participación Ciudadana, el cual se expondrá en el siguiente apartado.

1.1.1. Observatorio de Participación Ciudadana de Antioquia – Política Pública de Participación Ciudadana y Democrática

En los últimos años, la Secretaría de Participación y Cultura Ciudadana de la Gobernación de Antioquia por medio de la Dirección de Participación Ciudadana, Cultura y Gestión social ha implementado la estrategia denominada como “El Observatorio de Participación Ciudadana”, el cual nace de la necesidad por medir el cumplimiento de los instrumentos de participación en los municipios de Antioquia y con el fin de informar al ciudadano oportunamente y con claridad sobre el derecho a la participación ciudadana, contenidos, formas y procedimientos para su ejercicio. Siendo así, respecto a su vigencia, esta estrategia se enmarcó en el Plan de Acción del año 2024, el cual cuenta además con 32 actividades, las cuales junto con el Observatorio de Participación (OPC) apuntan al cumplimiento de los proyectos indicativos de las líneas estratégicas de la Política Pública de Participación Ciudadana y Democrática vigente desde el año 2023.

La Política Pública de Participación Ciudadana y Democrática o también llamada Ordenanza 025 de 2023 tiene como objetivo entre otros fortalecer las capacidades de la ciudadanía en general entendida como el conjunto de individuos, poblaciones, grupos, organizaciones e instituciones, de manera tal que facilite y promueva el ejercicio del derecho a la participación ciudadana y la generación de espacios de diálogo público, permitiendo su incidencia en la toma de decisiones. Dicha Política Pública, se encuentra vinculada con la implementación de una ley de carácter nacional (Ley 1757 de 2015 de Participación Democrática) por ello se entiende de carácter estatal; y debe asumirse e implementarse como política pública de largo plazo, lo cual implica que su ejecución sea de manera continua, sostenible y coherente con sus objetivos, principios y enfoques independiente de la administración departamental que la ejecute, por ello, se espera que el Observatorio de Participación Ciudadana (OPC) diseñado en un primer momento para la promoción y/o apoyo de la sistematización y divulgación de la información asociada a la participación ciudadana deba tener esta misma continuidad y cierta autonomía respecto a las administraciones y gobiernos por los cuales funcione o exista.

Respecto a la relación entre la Política Pública de Participación Ciudadana y Democrática y el Observatorio de Participación Ciudadana, este se posiciona como proyecto indicativo de la línea estratégica 2 “Información y comunicación para la participación ciudadana en las instancias y espacios de Desarrollo social”, siendo este una herramienta de información diseñada para permitir

analizar el estado de la participación ciudadana, identificar las causas de la desigual participación, comprender los patrones de participación en los territorios, destacar diferencias en la implementación del Sistema de Participación Ciudadana y reconocer las prioridades de la población Antioqueña. Siendo así, para el año 2024, dicho observatorio implementó un nuevo ejercicio de caracterización, el cual fue llamado “Caracterización de Espacios e Instancias de Participación Ciudadanas Municipales en el Departamento”, el cual se expondrá en el siguiente apartado.

1.1.2. Caracterización de Espacios e Instancias de Participación Ciudadanas Municipales en el Departamento

Como se menciona anteriormente, una de las más grandes e importantes iniciativas diseñadas, desarrolladas e implementada por el Observatorio de Participación en el año 2024 es la llamada “Identificación de Espacios e Instancias de Participación Ciudadanas Municipales en el Departamento”, la cual fue diseñada

“Con el ánimo de reconocer aquellas expresiones ciudadanas de participación presentes en cada uno de los municipios antioqueños y contar así con un panorama de su estado actual pensando en futuras actividades de acompañamiento enfocadas al fortalecimiento de estos espacios e instancias de participación, en asocio constante con las Administraciones Municipales”. (Observatorio de Participación Ciudadana de Antioquia, 2024.)

Relacionado además con el Plan de Desarrollo “2024-2027 Por Antioquia Firme” específicamente en su línea estratégica 1. Seguridad desde la democracia y la justicia; componente 1.1 Participación ciudadana; programa 1.1.1 Fortalecimiento de los organismos de acción comunal y la participación ciudadana en la gestión pública, el cual le apunta a “Promover la participación ciudadana en la gestión pública, legitimando procesos políticos, la democracia, la autonomía de las organizaciones y la comunidad que presentan, de manera que sea posible impulsar el desarrollo local y la democracia con participación e incidencia territorial”, propósito que se materializa con instrumentos y herramientas como el Observatorio de Participación en Antioquia.

El ejercicio de identificación o caracterización además es previsto como una herramienta de información para el uso de los Municipios, sus habitantes y los tomadores de decisiones, en la cual estos últimos pueden encontrar datos de primera mano de sus espacios e instancias de

participación y de posibles ofertas formativas desde la administración departamental, por ello, se entiende que la iniciativa sigue el principio de información pública, el cual busca en últimas brindar a la ciudadanía la capacidad de consultar sobre la información pública disponible de sus territorios. En consonancia con lo anterior, el ejercicio procura también apoyar la reactivación o conformación de aquellos espacios o instancias, que, por diversas razones, aún no estén constituidos o se encuentren en estado de inactividad, por ello, la iniciativa se entiende como una expresión de control social, el cual según función pública tiene por objeto el seguimiento y evaluación de lo público, así como la prestación de los servicios públicos, y los espacios institucionalizados de explicación y justificación de las decisiones, acciones y resultados en el ciclo de la gestión pública, es de destacar que al momento de la realización del proceso no se contaba con una metodología para la realización de caracterizaciones o procesos similares dentro del Observatorio de Participación Ciudadana, por lo cual se siguieron diferentes pasos propuestos por los funcionarios públicos encargados del Observatorio de Participación conforme iba avanzando el proceso. Siendo así, en el siguiente apartado se procederá a exponer las actividades realizadas durante la práctica de excelencia y durante el proceso de caracterización de espacios e instancias de participación ciudadana.

1.2. Pertinencia de la sistematización

La falta de metodologías y pasos claros para el ejercicio de caracterización de espacios e instancias de participación, lleva a pensar sobre que métodos deberían ser implementados en la misma, por ello, se propone el análisis exploratorio de datos (AED), el cual se presenta como un método comúnmente utilizado por la ciencia de datos para analizar y resumir conjuntos de datos permitiendo conocer sus características principales permitiendo comprender las estructuras e identificar relaciones entre variables. Dicho método es considerado como una fase diagnóstica y estratégica que permite elegir caminos más pertinentes para los análisis de datos, permitiendo con resúmenes numéricos y herramientas visuales proporcionar información esencial para orientar los pasos del proceso de análisis de datos, este paso puede ser una contribución esencial para guiar y proporcionar una base metodológica a los procesos desarrollados por el Observatorio de Participación con ayuda de técnicas de aprendizaje estadístico más adecuadas.

Dicha mejora aplicada dentro de la herramienta de política pública llamada Observatorio de Participación Ciudadana, específicamente dentro de la caracterización de espacios e instancias de participación ciudadana sería esencial para la cualificación del proceso y la mejora de las inferencias resultantes del mismo, entendiendo que los entes estatales o el sector público dispone con información limitada del contexto social, por ello es necesario lograr que los ejercicios emprendidos por los mismos y que terminan siendo la base para sus decisiones cuenten con una cualificación tal que logre que las inferencias tomadas por los mismos y entendidas como la información no explícita del ejercicio oriente a los “tomadores de decisión” a tomar decisiones lo más informadas, claras y pertinentes posibles para impactar de forma positiva el territorio.

Respecto a la cualificación de inferencias, varios autores han trabajado este concepto desde diferentes áreas del conocimiento, en este caso es esencial mencionar a Gamboa Rocabado 2010, quien en su artículo “Metodología para el análisis político: un enfoque flexible a partir de problemas, mecanismos e inferencias causales” en su apartado “La búsqueda de inferencias causales y el diseño de la investigación en ciencia política” afirma que es fundamental comprender que el diseño eficaz de la investigación debe representar en gran medida la posibilidad de llegar a la obtención de un modelo de inferencias causales para sistematizar y comprender fenómenos estudiados, el énfasis y fortaleza de los enfoques causales radica en la identificación de los mecanismos causales que van a explicitar la relación entre una causa y su correspondiente efectos, sin embargo dichos mecanismos están a su vez influenciados por voluntades, creencias, poderes y restricciones que caracterizan a los individuos y cuyas acciones influyen los fenómenos sociales, dichos mecanismos además están directamente unidos a la significativa conducta intencional que los individuos le dan a sus acciones, por ello en el ámbito público al igual que en el investigativo la necesidad de evidencia clara es esencial para la construcción de inferencias acertadas y cercanas a la realidad de los territorios, este debe ser uno de los fines mayores de los ejercicios emprendidos por las herramientas de política pública.

1.3. Pregunta de sistematización

Una vez mencionado lo anterior y retomando la necesidad de mejora en el proceso de caracterización de espacios e instancias de participación ciudadana, es importante señalar que, como herramienta de política pública, esta debe contar con capacidad de responder de forma

adecuada y con la información de mayor calidad posible a los requerimientos de la ciudadanía y ser así la base de información para las inferencias de los “tomadores de decisiones” en el ámbito público.

Siendo así, esta sistematización de experiencia de practica se pregunta por ¿De qué forma el análisis exploratorio de datos puede contribuir a cualificar las inferencias para la toma de decisiones producto de la caracterización de espacios e instancias de participación ciudadana del Observatorio de Participación Ciudadana de la Gobernación de Antioquia?

1.4. Objeto de la Sistematización

La sistematización de experiencias se podría definir siguiendo a Oscar Jara como la interpretación crítica de una o varias experiencias que a partir de su ordenamiento y reconstrucción, descubre o explica la lógica del proceso vivido, además de los factores que han intervenido en dicho proceso y como se han relacionado entre si (Jara Holliday, 2011). Siendo así, la fundación educación para el desarrollo en su guía para la sistematización de experiencias innovadoras precisa que la interpretación crítica es el resultado de todo un esfuerzo por comprender el sentido de las experiencias, tomando distancia de ellas y señala que dicha interpretación sólo es posible si previamente se ha ordenado y reconstruido el proceso vivido en esas experiencias.

Siguiendo a Jara y teniendo en cuenta que el objeto a sistematizar deben ser la o las experiencias concretas indicando el lugar donde se ha llevado a cabo, así como el periodo de tiempo que se va a escoger para esta sistematización, en este caso es el proceso llevado a cabo por la Secretaría de Participación y Cultura Ciudadana por medio del Observatorio de Participación Ciudadana en el segundo semestre del año 2024 denominado como “Caracterización de espacios e instancias de participación ciudadana municipales”.

1.5. Objetivos de la Sistematización

Retomando a Jara, el objetivo de la sistematización de experiencias es “reforzar la base empírica para la toma de decisiones por medio de la revisión del proceso llevado a cabo por la Secretaría de Participación y Cultura Ciudadana de la Gobernación de Antioquia en la caracterización de espacios e instancias de participación ciudadana municipales del Observatorio

de Participación Ciudadana”. Todo ello con el fin además de determinar como el análisis exploratorio de datos puede contribuir a construir flujos de trabajo para generar dicha base empírica por medio de una cualificación de inferencias producidas por iniciativas públicas.

Además, se analizarán y describirán las herramientas de análisis exploratorio de datos aplicables al proceso de caracterización de espacios e instancias de participación ciudadana del observatorio de participación (OPC) de la Gobernación de Antioquia, se explorará sobre las del uso de herramientas de análisis exploratorio de datos en los programas públicos como el OPC desde una perspectiva politológica. Además, se explorará el potencial de las herramientas de análisis exploratorio de datos para la cualificación de las inferencias derivadas de la caracterización de espacios e instancias de participación ciudadana, y su incidencia en la toma de decisiones en el ámbito público, logrando distinguir los desafíos y limitaciones de la implementación de nuevas herramientas de análisis exploratorio de datos en el sector público.

1.6. Pasos Metodológicos en la Sistematización de la Experiencia de Practica

Respecto a la metodología de la sistematización de experiencias, se seguirá la planteada por Oscar Jara Holliday en su artículo llamado “Orientaciones teórico – prácticas para la sistematización de experiencias”, dicho autor plantea diferentes pasos a seguir para plasmar una sistematización de experiencia adecuada, entre ellos destacan: El punto de partida, el cual consiste en la experiencia vivida, las preguntas iniciales, la recuperación del proceso vivido, las reflexiones de fondo (la explicación de por qué pasó lo que pasó en la experiencia), y finalmente los puntos de llegada.

Siendo así, Jara caracteriza el punto de partida como la experiencia vivida, es decir, afirma que se trata de partir de la experiencia propia, de lo que se hace y de lo que se piensa, siendo indispensable haber participado de manera activa en la experiencia que se va a sistematizar, para ello plantea diferentes preguntas iniciales, las cuales pueden servir de guía, algunas de ellas son: ¿Para qué queremos hacer esta sistematización?, ¿Qué experiencia (s) queremos sistematizar?, ¿Qué aspectos de la (s) experiencia (s) nos interesan más?, y ¿Qué fuentes de información se va a utilizar?.

Una vez concluido este paso, Jara da paso a la recuperación del proceso vivido, el cual se trata de una reconstrucción histórica ordenada de lo que fue aconteciendo durante la experiencia

vivida, esto permite una visión global de los principales sucesos que ocurrieron durante el periodo comprendido por la práctica, siendo así, es importante mencionar además que en este punto es importante identificar las etapas que siguió el proceso de la experiencia para poder narrarlo de manera óptima, dentro de esta reconstrucción no se debe emitir conclusiones o interpretaciones adelantadas, es decir, la idea es ser lo más descriptivos posibles.

En este punto se continua con la reflexión de fondo, la cual se trata de iniciar la fase interpretativa sobre todo lo que se ha descrito y reconstruido previamente en la experiencia, es decir, analizar el comportamiento de cada componente por separado y luego establecer relaciones entre estos para que surjan puntos críticos e interrogantes, dichos interrogantes pasaran por una interpretación critica, la cual permite entender la lógica que ha tenido la experiencia y confrontar estos hallazgos con otras experiencias y establecer relaciones. Es importante destacar que este punto además busca entender los factores claves o fundamentales; es decir, entender, explicitar o descubrir la lógica que ha tenido la experiencia.

Para terminar, respecto a los puntos de llegada Jara propone formular conclusiones y recomendaciones y elaborar productos de comunicación. Respecto a lo primero explica que se trata de arribar a las principales afirmaciones que surgen como resultado del proceso de sistematización, estas pueden ser tanto formulaciones teóricas como prácticas y son el punto de partida para nuevos aprendizajes, por lo que explica que pueden ser inquietudes abiertas debido a que no se trata de conclusiones definitivas.

1.7. Referentes Conceptuales Empleados

Respecto a los referentes conceptuales empleados, gran variedad de conceptos han sido trabajados desde la disciplina de la ciencia política en torno a la acción de los entes públicos y el estado, sin embargo en el contexto actual de transformación institucional, resulta de gran importancia integrar enfoques tales como los de las Políticas Públicas Basadas en Evidencias (PPBE), la analítica de datos y la cualificación de inferencias como parte de una agenda orientada a mejorar la transparencia, eficiencia y legitimidad de diferentes acciones públicas. Es así como por medio de las políticas sustentadas en información confiable y rigurosa, donde la analítica de datos se convierte en una herramienta clave para transformar grandes volúmenes de información en conocimiento útil para los tomadores de decisión. No obstante, para que el conocimiento

generado logre incidir de forma directa y robustezca las bases para la toma de decisiones es indispensable cualificar las inferencias derivadas de los análisis, asegurando que las relaciones identificadas entre variables estén fundamentadas en criterios válidos, replicables y sustentables.

Siendo así, entendiendo que el Observatorio de Participación es producto y herramienta de la política pública de participación ciudadana o ordenanza 025 de 2024, es necesario explicar el enfoque de las políticas públicas basadas en evidencias (PPBE), sobre ellas se ha trabajado desde diferentes enfoques, y se ha resaltado su importancia, ejemplo de ello es Jenkins (2023) quien argumenta que las políticas públicas bien fundamentadas deben desarrollarse a partir de una amplia gama de evidencias de alta calidad, siendo así, autores como Sutcliffe y Court (2006) proponen diferentes definiciones, una de ellas es concebirlas como un conjunto de métodos, cuyo objetivo es entre otros aportar información al proceso de elaboración de políticas, enmarcando su fin en que las personas a quienes corresponde tomar decisiones dentro del ciclo lo hagan a partir de la evidencia disponible recolectada a partir de la investigación rigurosa.

Además de estos autores, es de destacar Brancho (2010), quien nos da una definición de Políticas Públicas Basadas en Evidencia relacionándolas al uso de manera intencional de evidencia producto de la investigación, que emplea procedimientos rigurosos y sistemáticos para la recolección de los datos y los transforma en conocimiento formal útil, asistiendo a la toma de decisiones bien informadas en políticas, programas y proyectos, vinculada a esta definición Chaverri y Arguadas dentro de su trabajo “Políticas Públicas Basadas en Evidencia: Una revisión del concepto y sus características” argumentan que la importancia de formular PPBE ha ido creciendo dentro de los gobiernos nacionales, agencias internacionales y organizaciones independientes en los últimos años. Siendo así, estos autores hacen un esfuerzo en resaltar la importancia del rol y la contribución que los investigadores pueden cumplir en todas las etapas de las políticas públicas, desde su concepción hasta su evaluación, incluyendo la retroalimentación entre la valoración y la toma de decisiones sobre el futuro de las mismas, además, estos autores concluyen afirmando que las PPBE se presentan como una opción para mejorar la calidad de las decisiones políticas, permitiendo imprimir mayor claridad, precisión y confianza en la toma de decisiones con las que se busca resolver los diversos problemas sociales.

Una vez ampliado el panorama respecto a las PPBE, es de rescatar la definición propuesta por Flores – Crespo (2013), la cual se acercan más a iniciativas como el OPC, estos argumentan que las PBE son un enfoque que busca transformar la evidencia científica en conocimiento

utilizable, para promover un mejor desempeño de los gobiernos en la conducción de asuntos públicos, mediante políticas y programas más acertados.

Siendo así, la intención de ejercicios como la caracterización de espacios e instancias de participación dentro de un sector como el público entendido como gestor del bienestar ciudadano y actor que toma decisiones informadas y basadas en evidencia, genera la obligatoriedad y exigencia de integrar y actuar aplicando principios de las PPBE, esto entiendo además que la Secretaría de Participación Ciudadana siguiendo la aplicación de la política pública de participación ciudadana (Ordenanza 025 de 2023) utilice herramientas eficientes como el Observatorio de Participación para la toma de decisiones informadas y como base para sus intervenciones en los diferentes municipios del departamento, por ello dentro del ejercicio de caracterización de espacios e instancias de participación ciudadana, se incluyeron preguntas enfocadas en la mejora de la eficiencia de las intervenciones municipales, como ejemplo de esto encontramos dentro de la encuesta aplicada por la caracterización la pregunta #39 enfocada en detectar e inferir que prioridades de capacitación necesitaban los municipios del departamento, para lo cual se les puso a elegir entre diferentes opciones, entre ellas: Marco conceptual y legal de la participación y mecanismos de participación ciudadana, atención a la ciudadanía desde el enfoque de participación, control social y gestión pública, socialización de la política pública de participación ciudadana y democrática de Antioquia, veedurías ciudadanas – ley 850 de 2003 y rendición de cuentas. Una vez se tuvo la información completa de los municipios junto a la Escuela Superior de Administración Pública (ESAP), se impartieron encuentros formativos en cada subregión del departamento priorizando las respuestas de los municipios a la encuesta, es decir, se realizaron encuentros de Control Social y Gestión Pública pese a que esta opción no era la más eficiente.

Por ejemplos como el anterior, se evidencia la necesidad de la mejora y cualificación de los métodos empleados por los ejercicios emprendidos por las herramientas de la Secretaría de Participación Ciudadana como el Observatorio de Participación Ciudadana, entendiendo que al momento de la realización de la práctica no existían protocolos definitivos ni una ruta metodológica a seguir clara para la Caracterización de Espacios e Instancias de Participación Ciudadana, para ello es esencial y potencialmente necesario el uso de herramientas de analítica de datos con el fin de cualificar el diseño, la captura y el análisis de la información producida por los ejercicios emprendidos por el Observatorio de Participación, sin mencionar que la introducción de dichos

métodos alimentan una base procedimental y aportan a un diseño metodológico para esta y futuras iniciativas.

Siendo así y en búsqueda de herramientas para la cualificación de los ejercicios, encontramos el análisis exploratorio de datos (EDA), el cual se entiende como el método comúnmente utilizado para analizar y resumir conjuntos de datos y sus principales características para ayudar a determinar la mejor manera de manipular las fuentes de datos para obtener las respuestas más adecuadas a los requerimientos de los analistas, siendo así, entre los autores que han trabajado este concepto, es de destacar a Eugene Horber y Dominique Ladiray, quienes plantearon que el razonamiento exploratorio de datos es un esquema de análisis que en el ámbito investigativo logra enriquecer las posibilidades del investigador para hallar nuevas respuestas a los problemas que se planteen, esta afirmación se puede extrapolar al sector público, logrando que con la implementación de dichos métodos dentro de las iniciativas públicas se logre conseguir información de gran calidad que contribuyan a generar inferencias adecuadas para responder de forma acertada a las necesidades del accionar público.

Sobre la intervención científica y cualificada en los procesos públicos se ha escrito mucho, como ejemplo de ello encontramos el artículo de Eloisa del Pino Matute presidenta del Consejo Superior de Investigaciones Científicas llamado “Es necesario que existan intermediarios entre los científicos y los políticos” (Del Pino Matute, 2024) en donde expone que es relevante entender que el tipo de información que necesitan las personas que toman decisiones por lo general tiene poco que ver con el tipo de producto que habitualmente elaboran los científicos, sin embargo, los tomadores de decisiones requieren evidencias directas sobre los fenómenos sobre los cuales van a interferir, es decir necesitan cualificar sus inferencias. Pese a los numerosos análisis que explican los puntos de fricción entre políticos y científicos, del Pino explica que los productos científicos deben ser traducidos con el fin de que realmente sean útiles para los cargos públicos que toman decisiones, todo ello llevara a una cualificación de sus inferencias y por ende a una mejor intervención pública.

Las inferencias han adquirido tanta relevancia que actualmente se consideran el núcleo de la comprensión e interpretación de la realidad, y por tanto, uno de los pilares de la cognición humana (León, Solari, Olmos, Escudero, 2011.) por ello diferentes autores han hecho esfuerzos por explicar las implicaciones de dicho termino en la investigación social, como ejemplo de ello encontramos el artículo de Cortez y Solorio “Rastreo de procesos e inferencia causal en los métodos

cualitativos de la ciencia política” , quienes se enfocan en el rastreo de procesos o process traicing como la herramienta metodológica predilecta para los estudios politológicos, sin embargo, respecto a las inferencias traen el ejemplo de Gary King, Rober Keohane y Sidney Verba, quienes hacen un esfuerzo por exponer que tanto la investigación cualitativa como la cuantitativa siguen y contienen lógicas inferenciales subyacentes, es decir, los cualitativistas como los cuantitativistas comparten una misma lógica inferencial causal y solo se diferencian en cuestiones de estilos, esto ha llevado a una búsqueda por el refinamiento de las herramientas y formas entorno a los métodos cualitativos de investigación como herramienta válida para la inferencia causal en las ciencias políticas.

Siendo así, esta búsqueda por el refinamiento de las herramientas conduce necesariamente a la cualificación de las inferencias por medio de procedimientos que logren producir evidencias acertadas y cercanas a la realidad, uno de los autores que escribe sobre este problema es David Collier, Henry E. Brady y Jason Seawright en el capítulo “Source of leverage in Causal Inference: Toward an Alterative View of Methodology, donde presentan observaciones sobre el equilibrio de prioridades metodológicas ante la tecnificación continua del método y la teoría en las ramas de las ciencias sociales, siendo así, afirman que una inferencia causal efectiva requiere portar la mayor cantidad de evidencia posible, por lo cual buscan mostrar como las tradiciones cualitativas y estadísticas pueden ayudar a establecer una base metodológica más sólida para el progreso en las ciencias sociales.

Pese a los compromisos que enfrentan los investigadores a la luz del conocimiento metodológico y el peligro de la degeneración de la tecnificación en obscurantismo intelectual¹, Collier, Brady y Seawright afirman que a la hora de perseguir una inferencia descriptiva y causal no existe un conjunto de herramientas universales, si no que la existencia de los propios compromisos requiere un reconocimiento de que se necesitan diversas herramientas analíticas en la investigación social, concluyendo que la mayor promesa para el progreso en las ciencias sociales radica en una visión ecléctica de la metodología, la cual reconozca las contribuciones potenciales de diversas herramientas para cumplir con los diferentes estándares compartidos.

Los entes estatales, al igual que las políticas públicas, programas y proyectos se nutren directamente de los ejercicios investigativos y muchas veces aplican sus técnicas y metodologías

¹ Estos autores hacen referencia al obscurantismo intelectual como el peligro de que una solución o una herramienta más simple en efecto logre colocar a los académicos en un contacto más cercano con los casos que están estudiando y a veces permita que los analistas descubran patrones causales no anticipados

en búsqueda de conocimiento válido y acertado para las intervenciones públicas, en el caso del Observatorio de Participación, este al ser una herramienta de información deberá contar con los métodos y procedimientos más acertados para cumplir con sus fines, que en últimas es guiar la intervención pública y servir de repositorio de información para la ciudadanía en general, por ello es necesario que sus iniciativas se nutran de herramientas como las propuestas por la analítica de datos y generen inferencias cualificadas y reales que logren mejorar la calidad de la toma de decisiones.

Sin embargo, al ser un ente público, es esencial conocer sus implicaciones políticas, por ello es necesario mencionar el artículo “The politics of descriptive inferences: contested concepts in conflict data” de M.P. Broache y Agnes Yu, quienes a partir de la investigación descriptiva de los datos respecto a la toma de decisiones en política, afirman que este proceso implica la interpretación de datos para producir inferencias, las cuales terminan siendo altamente políticas en el sentido de que las estructuras de poder formal y/o informal informan o restringen necesariamente estas decisiones, es decir, las inferencias realizadas en la investigación descriptiva pueden reflejar y reforzar, o alternativamente, desafiar narrativas dominantes y estructuras de poder, siendo así, los datos proporcionados por la investigación descriptiva contribuyen firmemente a la toma de decisiones en el ámbito político, además, los autores señalan las responsabilidades de los investigadores tanto en la recopilación como en el uso de datos descriptivos debido a que la recopilación de datos además implica necesariamente hacer inferencias sobre la aplicación de conceptos frecuentemente debatidos, por lo cual afirman que es imperativo que los investigadores sean lo más transparentes posible sobre la razón detrás de sus decisiones, tanto que permita a los usuarios de datos descriptivos evaluar las inferencias en estos conjuntos de datos, apuntando a consideraciones éticas más amplias relacionadas con los roles duales de los investigadores como usuarios de datos y productores de conocimiento.

Una vez expuestos estos argumentos, es esencial realizar una revisión del proceso vivido y aportar desde la analítica de datos y la ciencia política capacidades y recomendaciones de mejora y cualificación del proceso de caracterización de espacios e instancias de participación, todo ello debido a que los datos producidos por la misma respecto a los diferentes municipios del departamento logren cualificar las inferencias respecto a los propios datos y estas logren mejorar las decisiones de intervención pública con una base científica y metodológica fuerte. Siendo así, como ejemplo de ello podemos retomar los encuentros de formación en control social y gestión

pública impartidos en las subregiones del departamento junto a la ESAP antes mencionados, ellos partieron de la aplicación de la encuesta realizada por la caracterización, la cual logró dar un panorama general de las necesidades expresadas por los enlaces municipales de sus propios territorios, su idea era lograr un impacto positivo y directo a las necesidades de la población en general, sin embargo utilizando los métodos de EDA se evaluará esta intervención. Bajo esta lógica, una mejora en el proceso de cualificación técnica de la caracterización de espacios e instancias de participación implementando herramientas de analítica de datos que logren cualificar el proceso, específicamente el análisis y la recolección de datos, proporcionaría una base científica y metodológica fuerte para lograr que los tomadores de decisiones que consulten la información producida por el Observatorio realicen inferencias de mayor calidad y por ende exista un mayor impacto de la herramienta de política pública y por ende de la política pública de participación ciudadana.

SECCIÓN 2. DESARROLLO DE LA SISTEMATIZACIÓN

2.1. Descripción de procesos – Caracterización de Espacios e Instancias de Participación Ciudadana

En este apartado se describirá los procedimientos asociados a la gestión de la información dentro de la Secretaría de Participación y Cultura Ciudadana de la Gobernación de Antioquia desde junio a diciembre del año 2024, haciendo énfasis en el acompañamiento al Observatorio de Participación Ciudadana de Antioquia en su iniciativa llamada “Caracterización de espacios e instancias de participación ciudadana” municipales, esta descripción además está marcada por cada uno de los momentos del proceso según los pasos seguidos por el Observatorio de Participación². Por ello, además es de destacar que el texto a continuación está estructurado a partir de la experiencia práctica en cada uno de los momentos planteados para la caracterización de espacios e instancias de participación ciudadana, siendo así estos fueron: El diseño del instrumento de recolección de información, el diseño de las herramientas para el análisis y el diseño y publicación de resultados

2.1.1. *El diseño del instrumento de recolección de información*

Para comenzar, respecto al diseño del instrumento para la recolección de la información dentro de las mediciones del Observatorio de Participación Ciudadana de la Gobernación de Antioquia, por lo general se utiliza el método de encuestas o formularios diseñados en los aplicativos de la propia Gobernación de Antioquia, en este caso, este fue implementado dentro de la etapa del diseño del instrumento para la recolección de la información de los espacios e instancias de participación municipales del departamento de Antioquia.

Siendo así, se diseñó un formulario o encuesta con preguntas estratégicas sobre la existencia de espacios e instancias de participación municipales pre establecidos en las aplicación ArcGIS Survey Connect y ArcGIS Survey 123, las cuales forman parte de la nube geoespacial de Esri, mostrándose como una solución completa basada en formularios que permite la creación y el

² Es necesario mencionar que a la fecha de realización de la práctica los métodos y las metodologías para llevar a cabo la sistematización no estaban claras, y se fueron construyendo conforme avanzaba el proceso

análisis de encuestas, con la posibilidad de captura de datos a través de la web o dispositivos móviles, en el caso de ArcGIS Survey Connect permitiendo además la especificación de formularios XLS, es decir, posibilita la edición a fondo permitiendo integrar Geoshapes, Códigos de barras, files, rangos, y parámetros de XLSForms³ (Esri, s.f.), estos parámetros se utilizaron con el fin de describir el orden y los tipos de campo de la encuesta, para ello, se debió completar los campos “type”, “name”, “label” dentro de una hoja de cálculo, siguiendo las instrucciones de Survey Connect, esta hoja de cálculo hace de base de la encuesta, por ello, fue crucial diseñar y completar con especial cautela todas sus variables.

Ahora bien, la creación de este instrumento de medición fue guiada por el componente “Recolección de la información clave para profundizar sobre el conocimiento en la participación ciudadana en el departamento de Antioquia”, dando como producto el formulario llamado “Identificación espacios e instancias de participación ciudadana existentes en Antioquia”, alojado en Survey123 de ArcGIS y en “Mi Contenido” de ArcGIS Online donde se configuró de forma tal que cualquier persona pudiese responder el formulario⁴, es decir, se admitió dentro de los parámetros de ArcGIS la recolección de datos públicos generando además una capa de entidades alojadas específica para los datos recolectados en la encuesta, si bien la creación del formulario fue planeada con más de un mes de antelación, a la hora de compartirlo hubo problemas para contactar con los enlaces municipales de la Secretaría de Participación y Cultura Ciudadana dentro de las diferentes subregiones del departamento.

El diseño del instrumento se enfocó en el objetivo de conocer si los municipios encuestados contaban con alrededor de 38 diferentes espacios e instancias de participación ciudadana, además se preguntó sobre la existencia de políticas públicas de participación ciudadana y/o comunales dentro de los 125 municipios del departamento con el fin de identificarlas y clasificarlas; también hubo un apartado específico para que cada municipio eligiera 3 opciones respecto a temáticas formativas ofrecidas por la Secretaría de Participación Ciudadana. Siendo así, los espacios o instancias caracterizados fueron:

Asambleas Municipales Constituyentes entendida como iniciativas ciudadanas creadas con el fin de mejorar o solucionar alguna situación coyuntural en el respectivo territorio que haya

³ Estos parámetros complementan las apariencias que admite un tipo de pregunta, pudiendo ser útiles para modificar además los comportamientos de las preguntas presentadas, sin embargo, no todos los parámetros son compatibles tanto con la aplicación de campo de ArcGIS Survey123 como con la aplicación web.

⁴ Enlace del formulario: <https://arcg.is/1Wj8vP>

afectado el orden democrático y social (Lattino, 2017), Cabildos del Adulto Mayor reconocido como espacio de interacción con la comunidad, el Estado y los organismos no gubernamentales, el cual analiza, adopta y propone la implementación de políticas, programas y proyectos a fin de mejorar las condiciones de vida de la población adulta mayor (Alcaldía de Jardín, 2024), Comité de Participación Social en Salud COPACO, Comités de Planeación y Presupuesto Participativo entendidos como mecanismos mediante los cuales la sociedad civil decide el rumbo de las políticas, planes, programas, proyectos o trámites de acuerdo con sus necesidades (Gobernación de Antioquia, s.f.).

Comités Interinstitucional para la Lucha Contra la Trata de Personas entendido como el organismo consultivo y ente coordinador de las acciones que desarrolla el municipio a través de la Estrategia Nacional para la Lucha contra la Trata de Personas (Ministerio, s.f.), Comités Municipal de Convivencia Escolar, Comité Municipal de Deportes, Comité Municipal de Discapacidad, Comité Municipal de Libertad Religiosa y Culto.

Comité Municipal de Política social COMPOS como instancia de coordinación interinstitucional e intersectorial para la definición, concertación seguimiento y evaluación de la política social, la toma de decisiones sustentadas en la responsabilidad de la articulación funcional del Sistema Nacional de Bienestar Familiar (Alcaldía de El Peñón, 2020), Consejo Consultivo de Diversidad Sexual y Género, Consejo Consultivo de Ordenamiento Territorial, Consejo Consultivo de Planificación de Territorios Indígenas.

Consejo Consultivo Municipal de Mujeres, Consejo Municipal de Cultura, Consejo Municipal de Desarrollo Rural comprendido como la instancia de concertación y coordinación entre las autoridades locales, las comunidades rurales y las entidades públicas en materia de políticas y programas para el desarrollo de los territorios rurales (Lattino, 2017), Consejos Municipal de Gestión del Riesgo de Desastres, Consejo Municipal de Juventud, Consejo Municipal de Participación Ciudadana, Consejo Municipal de Paz, Consejo Territorial de Planeación, Consejo Territorial de Seguridad Social en Salud, Consejo Territorial de Soberanía y Seguridad Alimentaria, Consejos Comunitarios Afrodescendientes, Juntas Municipales de Educación, Juntas Administradoras Locales, Ligas de Usuarios de Servicios de Salud, Mesas Ambientales, Mesas Diversas, Mesas Municipales del Adulto Mayor, Mesas Municipales de Derechos Humanos y Derecho Internacional Humanitario, Mesas Municipales de Mujeres, Mesas Municipales de Primera Infancia, Mesas Municipales de Víctimas, Mesas Sectoriales de Turismo, Plataformas

Municipales de Juventud, Veedurías Ciudadanas y si cuentan o no con Política Pública de Participación y/o Comunal en el municipio⁵.

Dentro de la encuesta se preguntó sobre la existencia del espacio o instancia dentro del municipio y por los datos de contacto de los voceros o representantes de estos espacios, dicha información se capturó dentro de Survey123 con el fin de ser utilizada por los funcionarios públicos para coordinar alianzas estratégicas en los territorios, además de servir para la consulta de la ciudadanía. Siendo así, respecto a las preguntas relacionadas a las necesidades formativas de los territorios, el Observatorio de Participación Ciudadana con la Secretaría de Participación y Cultura Ciudadana plantearon 7 diferentes temáticas de capacitación para que los municipios escogieran máximo 3, estas fueron: “Control social y gestión pública”, “Marco conceptual y legal de la participación, y mecanismos de participación ciudadana”, “Atención a la ciudadanía desde el enfoque de participación”, “Veedurías ciudadanas – ley 850 de 2003”, “Componente participativo en la formulación de una Política Pública de participación”, “Rendición de cuentas”, y “Socialización de la Política Pública de Participación Ciudadana y democrática de Antioquia”.

Para el proceso recolección de información se utilizó como población a todos los 125 municipios del departamento, sin embargo, se decidió por utilizar como muestra para el diligenciamiento del formulario a los promotores de desarrollo designados en cada municipio, en un primer momento ellos serían los encargados de brindar la información para el posterior análisis por parte del Observatorio de Participación Ciudadana. Siendo así, el proceso comenzó en junio y culminó en septiembre, es decir se les dio a los enlaces municipales⁶ (Promotores de desarrollo) alrededor de 4 meses para contestar la encuesta.

Debido a las pocas respuestas de los enlaces municipales se iniciaron acciones de divulgación de la encuesta, entre ellas, correos electrónicos masivos desde la cuenta de Outlook del Observatorio de Participación y llamadas desde la Secretaría de Participación y Cultura Ciudadana a los promotores de desarrollo con el fin de gestionar el diligenciamiento del instrumento de medición (encuesta). Para ello, se realizaron alrededor de 3 rondas de llamadas por cada municipio y se enviaron 3 correos distintos recordándoles su obligación, estas llamadas fueron sistematizadas dentro de una matriz de Excel, la cual contaba con los encabezados o variables

⁵ Dichos espacios e instancias de participación caracterizadas surgen de la propia identificación de la Secretaría de Participación y Cultura Ciudadana de la Gobernación de Antioquia.

⁶ Entendido como la persona encargada de brindar la información municipal al ente departamental

Municipio, Fecha de la llamada, y Respuesta al Llamado, dicha matriz fue alimentada durante todo el proceso por el Observatorio de Participación Ciudadana.

Debido a las pocas respuestas, dentro de la Secretaría de Participación y Cultura Ciudadana se decidió comenzar con la estrategia de radicar diferentes oficios a todas las Alcaldías Municipales dentro del sistema de gestión documental Mercurio 7.0⁷ (Sistema utilizado por todos los funcionarios de la Gobernación de Antioquia y los Alcaldes Municipales) para solicitar un designado para la respuesta a las mediciones futuras y existentes dentro del Observatorio de Participación Ciudadana, esta acción fue llevada a cabo desde el usuario de uno de los profesional universitario de la secretaria debido a que era necesario un tipo de permiso dentro de la aplicación y una cuenta específica, además dicha estrategia fue pensada con el fin de articular el Observatorio de Participación Ciudadana con los entes municipales y tener mayor visibilidad e incidencia dentro de los municipios y las subregiones del departamento. Una vez designado y establecido el enlace municipal por parte de las alcaldías, sus datos de contacto fueron sistematizados dentro de hojas de cálculo en excel para futuro uso por parte del Observatorio de Participación Ciudadana. Gracias a esta acción, se optimizó el proceso de recolección de la información resultante de las encuestas propuestas. Sin embargo, a la hora de finalizar el tiempo de recolección de información impuesto por la secretaria, solo 92 de los 126 municipios dieron respuesta al instrumento.

2.1.2. El diseño de las herramientas para el análisis

Referente a las herramientas para el análisis de la información resultante de la aplicación del instrumento, debido a las instrucciones y métodos utilizados anteriormente por el Observatorio de Participación Ciudadana, se diseñaron y crearon en Microsoft Excel a partir los datos brutos resultantes de la función exportar de Arcgis Survey 123, hojas de cálculo con todos los resultados de las preguntas diseñadas dentro del instrumento de recolección de información. Si bien esto no fue lo más adecuado debido a la extensión de los datos resultantes, la creación de varias matrices y tablas de análisis contribuyó al cumplimiento de los requerimientos de la Secretaría de Participación, como es el caso de la creación de bases de datos para la consulta de temáticas formativas por parte de los contratistas de la Secretaria.

⁷ Software de simplificación de procesos y gestión documental utilizado dentro de la Gobernación de Antioquia.

Si bien estas bases pudieron haber sido creadas con herramientas de analítica de datos como los entornos MySQL (Structured Query Language) , sistema de gestión de base de datos relacionadas de código abierto (Oracle, s. f.), para mejorar la eficiencia operativa debido a la relación entre los datos de las tablas resultantes de la caracterización, la creación de las bases de datos y matrices en excel contribuyó a la creación de los mapas interactivos publicados posteriormente en la página web del Observatorio de Participación, dichas matrices tenían que llevar un formato y un código específico por municipio para posteriormente realizar un join con las Feature Layers o capas de entidades alojadas de los municipios mapeados en Arcgis Pro. La base de dicho join o la variable que relacionó las dos capas fue Cod_Muni, columna esencial en las bases de datos de la Secretaria de Participación.

Dentro de las matrices para el análisis de los datos creadas, la más destacada fue la llamada “Insumo_instancias_de_participacion_PARCIAL”, debido a que allí se organizaron los datos resultantes de la encuesta por municipio y espacio e instancia de participación, es decir, por cada municipio del cual se registró respuesta se clasificó con un SI o un NO si el municipio contaba o no con el espacio o instancia de participación específico, todo ello con el fin de facilitar la información para los requerimientos de la Secretaria. Además de esta matriz, se creó la llamada “Base_Mapa_Instanceas_De_Participación_PARCIAL”, donde se organizaron los datos teniendo en cuenta el formato compatible con ArcGIS PRO, es decir, contaba con columnas o variables como Cod_muni, fuente, longitud y latitud, información de utilidad para la realización de la unión de los datos con los geoshapes existentes en las bases de datos del Observatorio de Participación, además, en el mismo archivo, en otras hojas de Microsoft Excel se crearon los gráficos dinámicos del departamento por espacio e instancia de participación por medio de tablas dinámicas, los cuales fueron agregados al informe gráfico final de la caracterización.

Siendo así, se diseñó también la matriz “Temáticas_Formativas”, donde se filtraron los datos resultantes de las respuestas de los municipios relacionadas a las preguntas referentes a las temáticas formativas propuestas por parte de la Secretaria de Participación y Cultura Ciudadana, si bien la creación de esta matriz fue muy útil para las posteriores iniciativas de la Secretaría, debido a la urgencia de su requerimiento fue creada de forma que solo se visualizará la primera opción de temática formativa escogida por cada municipio, en lugar de todas las opciones que escogieron; esta fue utilizada tanto por los funcionarios de la secretaría como por los contratistas de la misma

para encuentros formativos en cada subregión del departamento con el apoyo de la ESAP (Escuela Superior de Administración Pública).

Para terminar, la última matriz creada fue la llamada “Voceros_Municipales”, allí se filtraron los datos de contacto (Correo electrónico y teléfono) de los voceros y voceras de las instancias y espacios de participación municipales detectados, si bien esta matriz fue algo parecida a los datos expuestos en el informe “Informe de espacios e instancias de participación municipales existentes en el departamento de Antioquia”, esta fue propuesta a última hora y con las indicaciones requeridas por los contratistas de la Secretaría, para la articulación en diferentes iniciativas propuestas desde la dirección de participación ciudadana, cultura y gestión social.

2.1.3. El diseño y publicación de los resultados

Respecto a la tercera y última etapa, es decir, la presentación de los resultados de la medición, esta se llevó a cabo con el apoyo de las herramientas de ArcGIS. Es decir, se integraron y presentaron los resultados como página o botón dentro del sitio web del Observatorio de Participación Ciudadana alojado en ArcGIS HUB, esta página se nombró como “Espacios e instancias de participación”⁸, para ello, se utilizó el lenguaje HTML (Lenguaje de Marcas de Hipertexto) y las opciones de personalización del propio ArcGIS HUB. Siendo así, dicha página se constituyó con productos como: el “Informe de espacios e instancias de participación municipales existentes en el departamento de Antioquia – Septiembre”, presentado como una de las acciones propuestas en el marco del Observatorio con el fin de “reconocer aquellas expresiones ciudadanas de participación presentes en cada uno de los municipios antioqueños y contar así con un panorama de su estado actual pensando en futuras actividades de acompañamiento enfocadas al fortalecimiento de estos espacios e instancias de participación” (Observatorio de Participación, 2024.) Dicho documento, contó además con los datos de contacto⁹ de los voceros de cada espacio e instancia de participación resaltados dentro de la medición y con las preguntas presentadas en el instrumento de recolección de la información¹⁰.

⁸ Enlace: <https://opc-spcc-gobantioquia.hub.arcgis.com/pages/5339fce747bf475783c6f39e6a303b7a>

⁹ Los datos de contactos recolectados fueron: Correo electrónico y teléfono.

¹⁰ Las preguntas, se enfocaron en saber si los municipios contaban o no con los 38 espacios o instancias medidos por el Observatorio de Participación Ciudadana

Para la creación de este fueron utilizadas como instrumentos de análisis la página web de Arcgis Survey y las matrices de “Voceros_Municipales” e “Insumo_instancias_de_participacion_PARCIAL” de donde se extrajeron los datos, además se utilizaron y ajustaron mapas por cada municipio con el apoyo de diferentes insumos (Ilustraciones de mapas municipales) facilitados por la Secretaría de Planeación. Para terminar, es de destacar que la introducción del informe se diseñó en conjunto con los diferentes profesionales de la secretaria, quienes ajustaron y aportaron ideas para la construcción del mismo.

Además, se presentó el producto llamado “Informe gráfico espacios e instancias de participación municipales”, dentro del cual se expuso un panorama general del Departamento y de sus subregiones presentando gráficos dinámicos de barras con porcentajes, con el fin de comparar los datos categóricos mediante longitudes de barras proporcionales, las cuales reflejan el conteo de frecuencia en formato porcentual de las instancias y espacios de participación dentro del departamento y las subregiones ¹¹. Para la creación de este informe, se utilizó la matriz “Insumo_instancias_de_participacion ”, de la cual se extrajeron los datos sobre los cuales se crearon los gráficos dinámicos utilizados dentro del informe gráfico, estos gráficos dinámicos fueron creados gracias a una de las funcionalidades del software de Microsoft Excel y exportados al informe. Además, con apoyo del diseñador gráfico de la secretaria, se crearon diferentes mapas con colores específicos por cada subregión del departamento, resaltando los municipios que cumplieron con la encuesta, todo ello con el fin de motivar a los entes a participar en futuras iniciativas del OPC.

Para terminar, se creó, diseñó y publicó un mapa interactivo, intuitivo y de fácil acceso con la información obtenida por municipios, dicho mapa fue creado por medio de la integración y creación de capas de entidades alojadas en ArcGIS Pro¹² con las matrices de datos resultantes de la limpieza de los datos brutos del instrumento de recolección de información, es decir el producto “Base_Mapa_Instances_De_Participación_PARCIAL” con los geoshapes existentes dentro de las bases de datos del Observatorio de Participación Ciudadana. Estas capas de entidades alojadas entendidas como agrupaciones de entidades geográficas apropiadas para visualizar datos sobre los

¹¹ Es de destacar, que estos productos fueron publicados con la ayuda de las herramientas de personalización de sitios webs de ArcGIS HUB

¹² Aplicación SIG de escritorio diseñada para explorar, visualizar y analizar datos, crear mapas 2D y escenas 3D

mapas base (Esri, s, f.), fueron integradas en un mapa web en Arcgis Online para su futura publicación por parte del Observatorio de Participación Ciudadana. Dicho mapa fue publicado dentro de la página web del Observatorio de Participación Ciudadana por medio de Arcgis Dashboards, una de las opciones de Arcgis para presentar y crear cuadros de mando útiles para la presentación de datos en un formato fácil y de una lectura intuitiva para cualquier usuario, siendo así, se le aplicó un filtro de manera que seleccionando el municipio en una lista integrada en la parte izquierda, este se destacara automáticamente en el mapa interactivo, el cual fue presentado y publicado dentro del botón "Espacios e instancias de participación" del Observatorio de Participación para el libre acceso de cualquier persona que requiera o desee consultar la información allí registrada.

Finalmente, ya con toda la información pública esta se divulgó dentro de los encuentros de formación ciudadana, en este caso de control social subregionales que realizó la Secretaría de Participación y Cultura Ciudadana de la Gobernación de Antioquia, específicamente la dirección de Participación Ciudadana junto con la ESAP en el territorio, allí se les presentó a los municipios participantes el mapa interactivo con los espacios e instancias de participación presentes en sus municipios. Dichos encuentros subregionales se realizaron en los municipios de La Cebra (Oriente) , San Jerónimo (Occidente) , El Bagre (Bajo Cauca) , Venecia (Suroeste) , Puerto Berrío (Magdalena Medio) , Yarumal (Norte) y Yolombó (Nordeste) en acompañamiento con la ESAP (Escuela Superior de Administración Pública) y contaron con una media de alrededor de 50 participantes por encuentro, sin mencionar que la información contenida se socializó con los miembros de la dirección, directivos y en la semana de la participación de la Gobernación de Antioquia del año 2024. Además de estas iniciativas de divulgación, se envió un correo electrónico desde la cuenta del Observatorio de Participación y el Sistema de Participación Ciudadana de la Gobernación de Antioquia a las personas que fueron encargadas de responder el llamado del Observatorio de Participación sobre la encuesta, dicho correo contenía los informes como: “Informe gráfico espacios e instancias de participación municipales –Septiembre” e “Informe de espacios e instancias de participación municipales existentes en el departamento de Antioquia – Septiembre”, además de que se les anexo el enlace de la página web del Observatorio de Participación, específicamente de la pestaña de la caracterización con el mapa interactivo para su consulta y retroalimentación.

Una vez reconstruida la historia del proceso vivido con la caracterización de espacios e instancias de participación, siguiendo lo planteado por Jara, es necesario ordenar y clasificar la información disponible sobre los distintos componentes de la experiencia, para lo cual se planteó el uso de una matriz con los encabezados: Fuente, nombre, momento, y categoría, donde se almacenaron alrededor de 50 documentos y matrices resultantes del proceso de caracterización, como se mencionó en el anterior apartado, estos fueron extraídos del sistema de almacenamiento y gestión documental e información de la Secretaría de Participación Ciudadana.

2.2. Ruta de trabajo diseñada e implementada para el análisis. Cualificación de la gestión de información dentro del observatorio

Siguiendo la descripción anterior y la lógica planteada en los referentes conceptuales, en este apartado se describirán las primeras fases de la propuesta de cualificación de procesos asociados a la gestión de información para la toma de decisiones públicas dentro de la Secretaría de Participación, para ello, se tomara como base la caracterización de espacios e instancias de participación ciudadana y se propondrán diferentes adecuaciones metodológicas y procedimentales apoyados en las ciencias políticas y el análisis de datos, específicamente el EDA (Análisis exploratorio de datos) en todas las etapas del ejercicio descrito en el apartado anterior.

La analítica de datos como se ha venido explicando, es clave para la cualificación de inferencias en los ámbitos públicos, además es útil para la transformación de grandes volúmenes de datos en herramientas de gestión estratégicas que mejorar la toma de decisiones y optimizar los recursos, por ello a modo de diseñar una ruta metodológica y un flujo de trabajo para futuras iniciativas del Observatorio de Participación Ciudadana se tomara de ejemplo la caracterización de espacios e instancias de participación y se propondrán diferentes momentos esenciales para llevar a cabo la iniciativa con el fin de lograr un refuerzo técnico adecuado para las inferencias resultantes de los datos arrojados por la caracterización de espacios e instancias de participación ciudadana.

2.2.1. Preparación de los Datos

Construcción de Data Set:

Respecto a los resultados de la caracterización de espacio e instancias de participación, además de los informes creados, diseñados y publicados, dentro de la secretaría de participación ciudadana destacan 3 bases de datos o matrices resultantes del ejercicio, las cuales con el fin de generar recomendaciones en torno a la cualificación y tecnificación del proceso las tomaremos y las consideraremos como data sets¹³. Siendo así, la primera llamada “Departamento”, cuenta con el registro de respuestas de todos los municipios que llegaron a contestar la encuesta del Observatorio de Participación respecto a los espacios e instancias de participación existentes, esta está ordenada de forma que las respuestas de cada municipio están divididas en las columnas mientras que cada fila corresponde a cada municipio, la información se almacena de dicha forma debido a que contribuía a la creación de los gráficos y los mapas del informe publicado por el Observatorio de Participación Ciudadana.

Además de esta, por la necesidad de algunos contratistas y con el fin de guardar la información resultante de la encuesta, se diseñó otro formato para el resguardo de los datos. Este formato se llamó “Insumo_Instance_De_Participación”, en el cual se alojó la misma información de los espacios e instancias organizada de modo tal que cada fila representase una instancia o espacio de un municipio específico, con el fin de generar un intento de metadatos¹⁴ con las etiquetas de fuente, código del municipio y longitud y latitud.

Además de estos data set, encontramos otro referente a las prioridades de capacitación y formación para los municipios del departamento, esta fue diseñada de forma que cada fila representase un municipio mientras las columnas hacían referencia a quien diligenciaba la encuesta (con su cargo respectivo en la administración municipal) y las temáticas formativas encontradas en orden de prioridad para el municipio en cuestión, teniendo la posibilidad de presentar 3 como máximo. Estos data sets resultantes de la caracterización fueron muy valiosos e importantes para la toma de decisiones de la secretaria en intervenciones públicas, sin embargo, el ejercicio al carecer de una metodología y pasos claros, no existe una metadata suficiente de los espacios e instancias de participación a caracterizar para futuros ejercicios, por ello el primer paso para el diseño de una ruta metodológica, será la construcción de la Metadata.

¹³ Un data set es una colección organizada de información estructurada, típicamente en formato de tabla.

¹⁴ Se hace referencia intento de metadatos, porque no es clara la información ni suficiente las categorías para definirlos como tal.

Construcción de la MetaData de Espacios e Instancias de Participación Ciudadana:

Los Metadatos siguiendo a Bravo (2016), son entendidos como datos que describen otros datos, es decir, describe, explican y contextualizan otros datos aportando información adicional que permite interpretar correctamente la información y gestionarla de manera eficiente, su utilidad radica además en que ayudan a la recuperación de información, aportan estructura a contenidos no textuales, y organizan los recursos para su archivo e identificación de información. Este concepto ha sido trabajado por diferentes autores, entre ellos encontramos a Yarianis Cuba y Dianelis Olivera, quienes en “Los metadatos, la búsqueda y recuperación de información desde las ciencias de la información”, hacen un esfuerzo por definirlos, afirmando que estos pueden ser entendidos como conjuntos de información sobre los diferentes atributos de los recursos u objetos de información.

Siendo así, y partiendo que a la fecha de realización del ejercicio dentro de la Secretaría de Participación y Cultura Ciudadana no existe una estructura y ruta metodológica clara, es evidente la necesidad de la construcción de los metadatos referentes a los espacios e instancias de participación ciudadana, por ello como recomendación para el Observatorio de Participación Ciudadana sobre la caracterización de espacios e instancias de participación ciudadana, se partirá de la clasificación de Meta datos utilizada por Martínez Arellano y Amaya Ramírez,

Estos autores hacen un esfuerzo de distinción entre “Metadatos Administrativos”, los cuales están relacionados con el contexto administrativo de un determinado recurso de información principalmente a derechos de autor y permisos de acceso, los “Metadatos estructurales”, quienes facilitan la navegación y visualización de los recursos de información, a partir del establecimiento de las relaciones intrínsecas y extrínsecas de un recurso o conjunto de recursos y finalmente “Metadatos Descriptivos”, siendo estos los que se utilizan para identificar los principales atributos o características de los recursos de información, por ello, para este caso se utilizarán los metadatos descriptivos debido a su utilidad para describir el contenido temático de los datos facilitando su búsqueda y ayudando a organizar volúmenes de información, que en este caso serán los espacios e instancias que se buscan caracterizar con sus respectiva información.

Para la construcción de dichos metadatos se utilizó una matriz de Excel y se tomó como ejemplo los meta datos existentes dentro de la secretaría referentes a los indicadores del nivel e índice de participación ciudadana municipales, estos fueron diseñados por contratistas de la

secretaria junto al documento técnico del índice y nivel de participación municipales, por ello y con el fin de la fácil adaptación de los funcionarios encargados del Observatorio de Participación se intentó seguir la lógica de ficha y las formas existentes cambiando únicamente los datos registrados y algunos ítems descriptivos. Siendo así, dentro de los metadatos se guardaron las categorías o variables: "Nombre del espacio/instancia", "Tipo de espacio/ instancia", "ID", "Tipo de respuesta", "Definición del espacio o instancia", "Método de medición", "Periodicidad", "Entidad responsable", "Secretaria o dependencia", "Responsable del diligenciamiento", "Correo de contacto", y "Observaciones". Todo ello con el fin de crear una base sólida con la información de los datos buscados en la encuesta de Espacios e Instancias de Participación Ciudadana.

2.2.2. Arquitectura de Datos

Estandarización de la información:

Una vez creada la MetaData, con el fin de generar una estructura metodológica y procedimientos replicables para el manejo de los datos y los ejercicios del observatorio de participación ciudadana, es necesario abordar el concepto de estandarización de la información. La estandarización de la información entendida como el proceso de convertir datos de diversas fuentes a un formato, estructura, reglas comunes y uniformes, es de suma importancia en los procesos desarrollados por los entes públicos y privados, esto debido a que permite coincidir los datos de forma efectiva y facilita un formato coherente que permite realizar análisis acertados y garantiza la calidad de los datos.

Para entender este concepto hay que hacer mención a su aplicación en el ámbito empresarial, donde ha sido ampliamente trabajado por gigantes como IBM, para quien la estandarización de datos es una etapa crucial en la fase de investigación para facilitar que los datos de origen sean internamente coherentes; es decir, que cada tipo de datos tenga el mismo tipo de contenido y formato, esta etapa es esencial para el manejo y análisis de grandes cantidades de información debido a que prepara los datos para que una computadora pueda leerlos según las reglas predefinidas de la organización.

La creación de pautas, pasos, metodologías y una estandarización de datos clara además de proporcionar una base sólida para los análisis, aporta al tema o enfoque de datos abiertos, el cual

ha sido trabajado por autores como Pérez Lara y Covarrubias Moreno, quienes a propósito del tema sostienen que la analítica de datos es clave para convertir los grandes volúmenes de datos gubernamentales en herramientas de gestión estratégica que mejoren la toma de decisiones, la optimización de recursos y la anticipación a riesgos, estos autores destacan la importancia del modelo de gobierno abierto, el cual fundamentado en principios de transparencia, rendición de cuentas, participación y colaboración alude directamente al objetivo de que la cultura de la apertura se asiente en la totalidad de las instituciones públicas reduciendo así riesgos de discrecionalidad y secretismo. Al no existir un formato estándar de información para resguardar los datos resultantes del ejercicio de caracterización de espacios e instancias de participación, es necesaria la construcción de un estándar para las bases de datos resultantes.

Una vez conceptualizada la estandarización de la información, para este caso se decidió estandarizar los datos resultantes del ejercicio de caracterización de espacios e instancias de participación designándoles un código único entre P_1 y P_38, y creando códigos como P_E para los funcionarios que respondieron la encuesta, P_EN para su teléfono y P_EC para el correo electrónico de contacto. Además, se crearon las columnas PC_40_1, PC_40_2 y PC_40_3 haciendo referencia a las respuestas relacionada a la temática de formación seleccionada. Una vez hechos estos cambios, se diseñó un diccionario de datos siguiendo la estructura del Survey diseñado originalmente dentro de Arcgis Connect, además se cambiaron los valores SI/NO a valores binarios 1/0 y se le asignó un código entre 1 y 8 a cada necesidad formativa ofrecida por la Secretaría de Participación a los municipios, dicha estandarización fue necesaria para el posterior análisis de los datos cuantitativos utilizado EDA (Análisis exploratorio de datos) con ayuda de lenguajes como Python y librerías como Pandas. Siendo así, para el análisis de la información y la aplicación de la estandarización, es necesaria la creación de una base de datos sólida.

Construcción de bases de datos:

Respecto a la construcción de bases de datos, es necesario hacer una distinción y claridad sobre las formas de almacenar información, dentro del mundo de la analítica de datos existen diferentes modelos de almacenamiento de datos, entre ellos destacan las bases de datos relacionales y no relacionales. Sobre ellas han trabajado diferentes autores, entre los cuales se puede mencionar a Marrero, Olsowy, Thomas, Delia, Tesone, Fernandez Sosa y Pesado, quienes realizaron un

artículo de los procesos de diseño de las bases de datos relacionales, sobre las cuales logran definir diferentes características, entre ellas que no requieren un lenguaje de consulta estructurado (SQL) como lenguaje principal, ni una estructura fija y tabular, no soportan operaciones JOIN, no garantizan por completo las propiedades de ACID (atomicidad, consistencia, aislamiento, durabilidad), y en general su estructura se basa en escalabilidad horizontal.

Siendo así, una base de datos no relacional (NoSQL), siguiendo a Zoiner Tejada director ejecutivo y arquitecto de Microsoft, es aquella que no usa el esquema tabular de filas y columnas que se encuentra en la mayoría de los sistemas de bases de datos más tradicionales, en lugar de esto, utilizan un modelo de almacenamiento que está optimizado para los requisitos específicos del tipo de datos que se almacena, por lo cual estas bases de datos pueden trabajar con grandes volúmenes de datos no estructurados. Sin embargo, pese a su flexibilidad y escalabilidad no cuentan con un esquema fijo y tienden a tener una coherencia débil, por lo cual debido a las necesidades de almacenamiento de datos y estandarización de la información del ejercicio de espacios e instancias de participación esta opción no sería la más adecuada.

Sin embargo, respecto a las bases de datos relacionales encontramos que por lo general constan de tablas bidimensionales construidas por filas y columnas, permitiendo que los datos cuenten con independencia lógica, consistencia, confiabilidad y escalabilidad permitiendo manejar grandes volúmenes de información. Siendo así, al igual que las bases de datos no relacionales, este concepto ha sido trabajado por autores como Piñeiro Gómez, el cual parte de la definición de Piattini quien describe las bases de datos relacionales como una colección de datos integrados con redundancia controlada y con una estructura que refleje las interrelaciones y restricciones existentes en el mundo real. Agrega que los datos de estas deben contar con una definición y descripción únicas para cada tipo de dato y han de estar almacenadas junto con los mismos. Además, estas contribuyen a la integridad, confidencialidad y disponibilidad de los conjuntos de datos. Por otro lado, además de estos autores Microsoft nos ofrece una definición, la cual afirma que son tipo de bases de datos que almacena y organiza puntos de datos con relaciones definidas para un acceso rápido, es decir, utiliza las tablas para contener información sobre cada entidad y representan categorías predefinidas mediante filas y columnas.

Siendo así y entendiendo las necesidades de consulta rápida, estandarización, almacenamiento y análisis de datos por medio de EDA del Observatorio de Participación Ciudadana respecto al manejo de la información resultante del proceso de caracterización de

espacios e instancias de participación ciudadana se plantea la construcción de una base de datos relacional, la cual recoge la información de las respuestas sobre la existencia de 38 espacios o instancias de participación ciudadana (descritas en los metadatos) por cada uno de los 125 municipios del departamento de Antioquia, su nombre, su código para la creación de mapas con ArcGis, y su subregión. Además de esta información se integró el conjunto de datos referente a las temáticas formativas a elección de los municipios, para este último apartado se priorizaron las 3 primeras respuestas de cada enlace.

2.2.3. *Análisis exploratorio de datos (EDA)*

Una vez construida la Metadata, estandarizada la información y ajustadas las bases de datos, es importante traer y exponer los métodos de EDA (Análisis exploratorio de datos) para encontrar patrones útiles para la toma de decisiones de los funcionarios públicos dentro de la información resultante de la caracterización. El EDA como se explicó al inicio del trabajo será entendido como el método comúnmente utilizado para analizar y resumir conjuntos de datos y sus principales características, este concepto ha sido trabajado por diferentes autores, entre ellos Grande Atienza, quien describe el AED como una forma de analizar datos mirando números y gráficos, obteniendo un patrón a través del tratamiento de las muestras recogidas en el proceso de investigación de un campo específico.

Este autor además recoge varias características claves respecto al análisis exploratorio de datos, entre ellas, su fuerte apoyo en representaciones gráficas, su empleo preferente de los estadísticos de orden y su similitud a la estadística descriptiva tradicional sin la necesidad de una teoría matemática compleja, además autores como Solanet i Grau afirman que el análisis exploratorio de datos pone el énfasis en conocer el máximo posible de cosas de un fenómeno a través de los datos con el fin de explotar su utilidad respecto a una teoría.

Para destacar la utilidad del EDA dentro del análisis de los datos del observatorio de participación, podemos evidenciar que una de sus herramientas gráficas fue aplicada a la caracterización de espacios e instancias de participación, este es el caso del informe “Informe gráfico espacios e instancias de participación municipales”, donde se presentó un conteo de las respuestas de los municipios para cada subregión a la pregunta por la existencia de las diferentes espacios e instancias de participación, mostrando la información (conteo) por medio de gráficos de

barras con datos porcentuales. Si bien este fue un buen uso del EDA, su utilidad no se agota ahí, por ello se han de exponer diferentes herramientas útiles y aplicables a los conjuntos de datos de la caracterización de espacios e instancias de participación ciudadana.

Siendo así, las herramientas de EDA pueden ser utilizadas desde la preparación y limpieza de datos por medio de la unificación de categorías, la conversión de variables a valores binarios y la verificación de faltantes hasta la aplicación de estadística descriptiva para conocer la distribución de instancias de participación por cada subregión. La aplicación de dichas herramientas es esencial para la comparación entre municipios y subregiones por medio por ejemplo del promedio de instancias activas. Sin embargo, no es solo útil para analizar la información referente a las instancias de participación, sino también para conocer la frecuencia de los temas de capacitación solicitados por los diferentes municipios, información esencial para el accionar público de la misma secretaría de participación ciudadana.

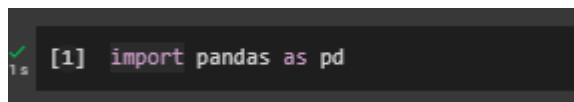
Aplicación de EDA a la Caracterización de Espacios e Instancias de Participación con Python:

Una vez explicada la utilidad del EDA para el análisis de los datos de la caracterización de espacios e instancias de participación ciudadana, para visibilizar la potencialidad de los métodos se aplicaron diferentes herramientas visuales y analíticas sobre la base de datos, todo ello en el entorno virtual de Google Colab o Colaboratory¹⁵ usando el lenguaje Python y la librería Pandas, se eligieron estos métodos debido a su facilidad de aprendizaje y su sintaxis intuitiva.

Siendo así, el primer paso para la aplicación del EDA fue el importe de la librería al entorno virtual de Google Colab:

Figura 1:

Función para importar la librería de pandas

A screenshot of a Jupyter Notebook cell. On the left, there is a green checkmark icon and the text '1 s'. To the right, the code cell contains the text '[1] import pandas as pd' in a monospaced font with syntax highlighting.

Fuente: Elaboración propia

¹⁵ Google Colab es un servicio gratuito en la nube que funciona como un entorno de Jupyter Notebook, permitiendo escribir y ejecutar código de Python directamente en un navegador web.

Seguido a esto, se importó la base de datos de espacios e instancias de participación ciudadana, la cual estaba alojada en un archivo de Excel, para ello se ejecutó la siguiente función:

Figura 2:

Función para importar archivos a Google Colab desde la computadora

```
[14] from google.colab import files  
      uploaded = files.upload()
```

Fuente: Elaboración propia

Una vez cargada la base de datos, se convirtió en un DataFrame por medio de la función de Pandas `read_excel`.

Figura 3:

Función para el alojamiento de los datos en un Dataframe utilizando pandas

```
[18] df = pd.read_excel("BD_Departamento.xlsx")
```

Fuente: Elaboración propia

Esta función hizo que los datos de la base de datos de espacios e instancias de participación se guardasen en la variable `df`, paso esencial para la futura aplicación de diferentes funciones que contribuirán al análisis de los datos resultantes de la caracterización de espacios e instancias de participación ciudadana.

Ahora bien, para la aplicación de EDA es esencial tener claridades respecto a las muestras y las bases de datos sobre las cuales se va a trabajar, por ello para sintetizar la información y la visualización general del panorama de los municipios que dieron respuesta a la encuesta se desarrolló la siguiente función:

Figura 4:

Función para la creación de la tabla de municipios por subregión

```
[18] muni_por_subregion = df.groupby("Sub")["Muni"].apply(list).reset_index()  
      muni_por_subregion.rename(columns={"Muni": "Municipios"}, inplace=True)  
      muni_por_subregion.rename(columns={"Sub": "Subregión"}, inplace=True)  
  
[19] muni_por_subregion["Total Municipios"] = muni_por_subregion["Municipios"].apply(len)  
      muni_por_subregion
```

Fuente: Elaboración propia

Por medio de la cual con ayuda de la librería pandas se organizaron los municipios por subregión, se renombraron las columnas y se agregó un conteo total de municipios en cada subregión del departamento. Esto fue realizado con el fin de conocer de que tamaño es la muestra y que tantas respuestas se logró recolectar dentro de las diferentes subregiones para su posterior análisis por medio de EDA

Figura 5:

Resultado de la función tabla de municipios por subregión Google Colab.

Index	Subregión	Municipios	Total Municipios
0	BAJO CAUCA	CACERES, CAUCASIA, ZARAGOZA	3
1	MAGDALENA MEDIO	CARACOLI, MACEO, PUERTO NARE, PUERTO TRIUNFO	4
2	NORDESTE	AMALFI, ANORI, CISNEROS, REMEDIOS, SAN ROQUE, SANTO DOMINGO, SEGOVIA, VEGACHI, YALI, YOLOMBO	10
3	NORTE	ANGOSTURA, BELMIRA, CAMPAMENTO, CAROLINA DEL PRINCIPLE, DON MATIAS, GOMEZ PLATA, ITUANGO, SAN ANDRES DE CUERQUIA, SAN PEDRO, TOLEDO, YARUMAL	11
4	OCCIDENTE	ABRIAQUI, ANZA, ARMENIA, BURITICA, CAÑAS GORDAS, DABEIBA, EBEJICO, FRONTINO, GIRALDO, LIBORINA, MARINILLA, OLAYA, PEQUE, SABANALARGA, SAN JERONIMO, SANTAFE DE ANTIOQUIA, SOPETRAN, URAMITA	18
5	ORIENTE	ABEJORRAL, CARMEN DE VIBORAL, CONCEPCION, GRANADA, GUARNE, GUATAPE, LA CEJA, LA UNION, NARIÑO, PEÑOL, RETIRO, SAN CARLOS, SAN FRANCISCO, SAN JOSE DE LA MONTAÑA, SAN LUIS, SANTUARIO, SONSON	17
6	SUROESTE	ANDES, ANGELOPOLIS, BETULIA, CARAMANTA, HISPANIA, JERICO, LA PINTADA, MONTEBELLO, SANTA BARBARA, TAMESIS, TARSO, TITIRIBI, URRAO, VALPARAISO	14
7	URABÁ	ARBOLETES, CAREPA, CHIGORODO, MURINDO, MUTATA, NECOCOLI, SAN JUAN DE URABA, TURBO, VIGIA DEL FUERTE	9
8	VALLE DE ABURRÁ	ENVIGADO, GIRARDOTA, ITAGUI, LA ESTRELLA, MEDELLIN	5

Fuente: Elaboración propia

Una vez generada la tabla podemos inferir que, de los 125 municipios existentes en el departamento de Antioquia, solo 92 dieron respuesta a la encuesta de caracterización de espacios e instancias de participación, de los cuales 3 fueron de la subregión de Bajo Cauca, 4 de Magdalena Medio, 10 de Nordeste, 11 de Norte, 18 de Occidente, 17 de Oriente, 14 del Suroeste, 9 de Urabá y 5 del Valle de Aburrá.

Tabla 1

Cobertura de la encuesta de caracterización de espacios e instancias de participación del departamento de Antioquia

Subregión	# de municipios	# de municipios encuestados	% de cobertura
Antioquia	125	92	73,6%
Bajo Cauca	6	3	50%
Magdalena Medio	6	4	66,7%
Nordeste	10	10	100%
Norte	17	11	64,7%
Occidente	19	18	94,7%
Urabá	11	9	81,8%
Suroeste	23	14	60,9%

Oriente	23	17	73,9%
Valle de Aburrá	10	5	50%

Fuente: Elaboración propia

Siendo así, con esta información podemos concluir que la cobertura que tuvo la encuesta dentro del departamento es variada, siendo que en subregiones como Nordeste u Occidente la cobertura es muy alta siendo esta del 100% y del 94,7%, prácticamente total, mientras que en el Valle de Aburrá y Bajo Cauca la cobertura es baja en términos absolutos representando solo el 50%. A partir de esta información, podemos advertir que el número de respuestas de los municipios generan implicaciones directas en el análisis de los datos subregionales, ello debido a que la baja cobertura en ciertas subregiones revela vacíos que afectan la capacidad de generalizar conclusiones, sin embargo debido a la falta de disposición de los municipios a dar respuesta al instrumento de caracterización durante la experiencia se trabajará con estos datos a modo de ejemplificar las potencialidades de la analítica de datos para la cualificación de inferencias dentro del sector público.

Figura 6:

Mapa de Municipios encuestados por la caracterización de espacios e instancias de participación



Fuente: Elaboración propia

2.2.4. Aplicación del EDA para Determinar las Necesidades Formativas en el Departamento:

Una vez importada la base de datos y teniendo claras las muestras sobre las que se va a trabajar, una de las aplicaciones del EDA será descubrir que temática formativa es la más óptima para la realización de encuentros subregionales de formación en temas participativos, todo ello debido a que la Secretaría de Participación Ciudadana realizó diferentes encuentros de capacitación en temas de Control Social y Gestión Pública en las subregiones del departamento tomando como base las respuestas generales de la encuesta entorno a las temáticas formativas. Por ello, por medio de la aplicación de EDA se concluirá cual debió haber sido la temática de los encuentros en cada subregión por medio de evidencias y un análisis claro, para ello, es necesario recordar que dentro de la encuesta de caracterización de espacios e instancias de participación se diseñaron 3 priorizaciones referentes a las necesidades formativas en orden de elección, empezando por la 1 y terminado por la 3. Siendo así, al ajustar la base de datos dichas priorizaciones fueron nombradas como PC_40_1, PC_40_2 y PC_40_3, una vez aclarado esto, se procederá a la realización de conteos con ayuda la librería PANDAS y matplotlib dentro del notebook de GoogleColab utilizando Python. En los apartados siguientes se describirá el proceso para la ejecución de los ejercicios de analítica de datos, sin embargo, su análisis se realizará en posteriormente en el apartado “**Análisis y Patrones Detectados de la Aplicación de EDA y Machine Learning**”.

Siendo así, para conocer las diferentes necesidades formativas elegidas por los municipios que contestaron la encuesta de caracterización de espacios e instancias de participación, para la aplicación de diferentes análisis univariados se procederá a aplicar la función `value_counts` de la biblioteca de pandas, la cual se encarga de contar la frecuencia de valores únicos en una columna de una serie o un dataframe, en este caso el dataframe será la base de datos resultante de la caracterización de espacios o instancias de participación ciudadana ajustada anteriormente y definido como `df`.

Teniendo en cuenta esto, es importante tener claridades sobre la base de datos, para la cual cada opción de necesidad formativa paso a representar un número entre el 1 y el 8, siguiendo el siguiente orden:

Tabla 2:
Numeración necesidades formativas

1	Control social y gestión pública
---	----------------------------------

2	Marco conceptual y legal de la participación, y mecanismos de participación ciudadana
3	Atención a la ciudadanía desde el enfoque de participación
4	Veedurías ciudadanas - ley 850 de 2003
5	Componente participativo en la formulación de una Política Pública de Participación
6	Rendición de cuentas
7	Socialización de la Política Pública de Participación Ciudadana y Democrática de Antioquia
8	Sin dato

Fuente: Elaboración propia

Este paso fue esencial debido a la presencia de variables en textos representan problemas a la hora de aplicar herramientas graficas del EDA como las barras o los histogramas y generan problemas de formato.

Para el conteo se seleccionó cada columna del dataframe Departamento (PC_40_1, PC_40_2 y PC_40_3 y se utilizó el parámetro `normalize=True`, el cual hace que el conteo se convierta en proporciones (0 a 1) en lugar de números absolutos, para luego multiplicar su resultado por 100 y obtener el porcentaje de respuesta de los municipios por cada necesidad formativa en cada una de las preguntas que hacían referencia a las mismas.

Se decidió por presentar los datos de manera porcentual dentro de una tabla de frecuencia relativa porque ello facilita la interpretación a la hora de tomar decisiones sobre la representatividad de cada opción de respuesta de la muestra de los municipios encuestados en el departamento general y en sus subregiones. Siendo así, para conocer el porcentaje de respuestas para la pregunta sobre necesidades formativas número uno, se aplicó la siguiente función:

Figura 7:

Función porcentajes de respuestas necesidades formativas priorización #1.

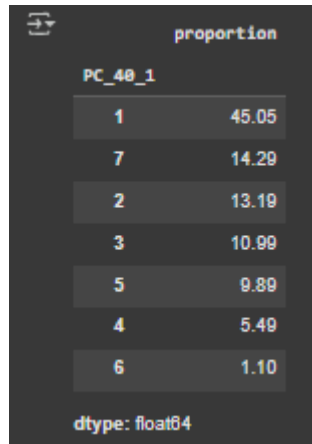
```
[ ] porcentajes_PC_40_1 = df["PC_40_1"].value_counts(normalize=True).mul(100).round(2)
    porcentajes_PC_40_1
```

Fuente: Elaboración propia

Una vez aplicada la función sobre la base de datos resultante de las respuestas de los municipios dio como resultado la siguiente tabla de frecuencia relativa:

Figura 8:

Resultado de la función porcentajes de respuestas necesidades formativas priorización #1.



PC_40_1	proportion
1	45.05
7	14.29
2	13.19
3	10.99
5	9.89
4	5.49
6	1.10

dtype: float64

Fuente: Elaboración propia

Con esta información ya se pueden realizar diferentes análisis sobre las necesidades formativas priorizadas por los diferentes municipios del departamento de Antioquia, sin embargo, para un mejor análisis de la información se optó por presentar la información de manera grafica utilizando la librería de Python Matplotlib, la cual permite crear visualizaciones de datos en dos y tres dimensiones, como histogramas, gráficos de dispersión y diagramas de barras. Siendo así, en este caso se utilizó la forma pie -chart (gráfico circular) debido a su capacidad de mostrar cómo se distribuye una variable categórica en relación el 100% del total.

Para la creación del grafico primero se establecieron las etiquetas del mismo, para lo cual se ejecutó la siguiente función con el fin de unir cada categoría con su respectivo valor:

Figura 9:

Función Creación de las etiquetas de datos necesidades formativas departamento

```
[7] labels_40_1 = [f"{cat} ({val:.2f}%)" for cat, val in zip(porcentajes_PC_40_1.index, porcentajes_PC_40_1)]
```

Fuente: Elaboración propia

Una vez creadas las etiquetas en forma de lista con las categorías y los valores, se procedió con la realización del grafico circular por medio de la librería Matplotlib y la siguiente función:

Figura 10:

Función Creación del grafico circular distribución necesidades formativas #1 Departamento - Porcentajes

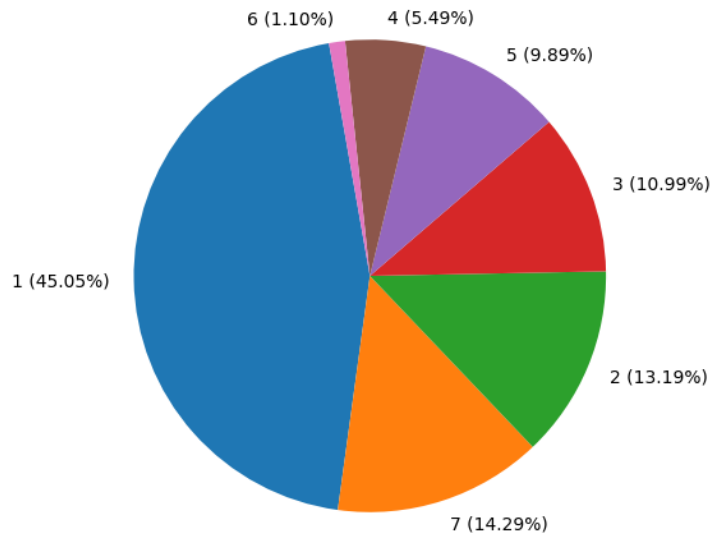
```
[16] plt.figure(figsize=(10,10
      ))
      plt.pie(
          porcentajes_PC_40_1,
          labels=labels_40_1,
          startangle=90,
      )
      plt.title("Distribución necesidades formativas #1 Departaneto - Porcentajes")
      plt.show()
```

Fuente: Elaboración propia

Dicha función dio como resultado el siguiente grafico:

Figura 11:

Gráfico Distribución necesidades formativas #1 – Departamento – Porcentajes

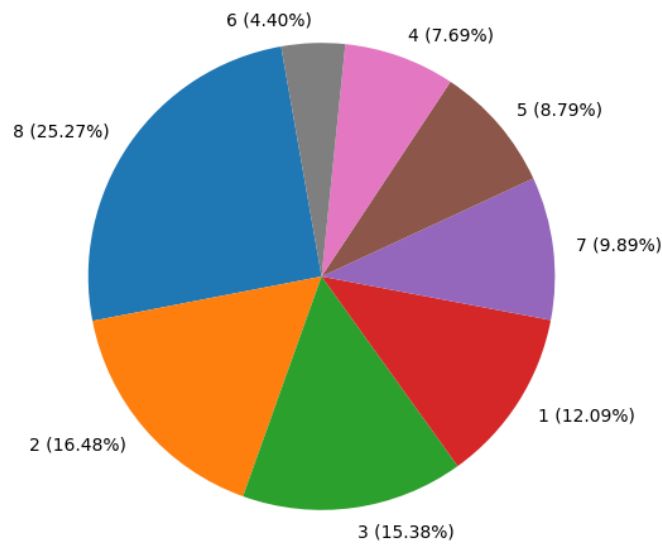


Fuente: Elaboración propia

Siendo así, para las otras dos preguntas relacionadas a las necesidades formativas (PC_40_2 y PC_40_3) se ejecutó el mismo proceso, lo cual dio como resultado el siguiente resultado grafico:

Figura 12:

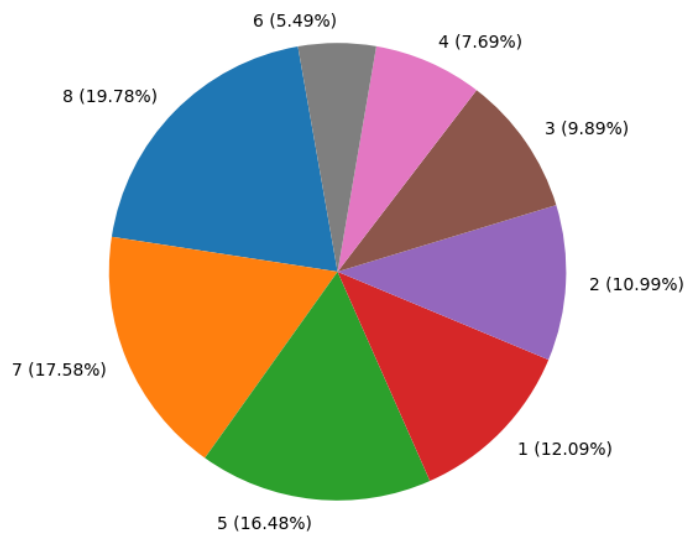
Gráfico Distribución necesidades formativas #2 Departamento – Porcentajes



Fuente: Elaboración propia

Figura 13:

Gráfico Distribución necesidades formativas #3 Departamento – Porcentajes



Fuente: Elaboración propia

Necesidades formativas por subregión:

Si bien la información resultante de los gráficos anteriores es útil y podrían arrojar conclusiones sobre las temáticas formativas del departamento, para lograr cualificar la intervención

pública por medio de encuentros formativos dentro de las diferentes subregiones es necesaria la realización del análisis de las respuestas desagregadas por subregiones. Dicho análisis se realizó con ayuda de la función GroupBy en combinación de value_counts de la librería pandas, por medio de la cual se generó un Dataframe con el porcentaje de respuesta de los municipios para cada opción en cada subregión. Para esto se diseñó la siguiente función:

Figura 14:

Función división de los porcentajes de los municipios por subregión - #1

```
[8] conteo_subregion_PC_40_1 = df.groupby("Sub")["PC_40_1"].value_counts(normalize=True).mul(100).round(2)

    conteo_subregion_PC_40_1
```

Fuente: Elaboración propia

Por medio de la cual se generó el conteo de los municipios divididos en los grupos de las subregiones, sin embargo, para exponer mejor la información fue necesaria la creación de una tabla de frecuencia relativa o porcentual, la cual se diseñó a partir de la siguiente función:

Figura 15;

Función creación de la tabla de porcentajes de los municipios por subregión - #1

```
[11] tabla_conteo_PC_40_1 = conteo_subregion_PC_40_1.reset_index(name = "Porcentaje")

    print (tabla_conteo_PC_40_1)
```

Fuente: Elaboración propia

Por medio de la anterior función fue posible convertir el conteo/porcentaje con índices en una tabla estructurada, donde el índice pasa a ser una columna normal y los valores se renombran como porcentajes, una vez generada esta tabla se procedió a diseñar el grafico con Matplotlib, con la siguiente función:

Figura 16

Función creación del gráfico Distribución necesidades #1 Subregiones – porcentajes

```
[ ] Grafico_sub_PC_40_1 = tabla_conteo_PC_40_1.pivot(index="Sub", columns="PC_40_1", values="Porcentaje").fillna(0)
Grafico_sub_PC_40_1

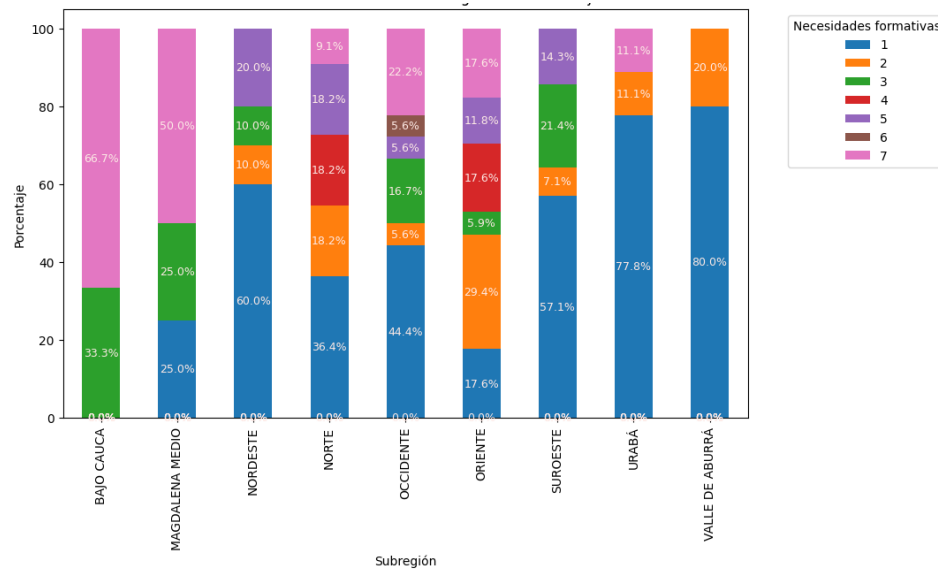
ax = Grafico_sub_PC_40_1.plot(
    kind="bar",
    stacked=True,
    figsize=(10, 6)
)

plt.title("Distribución necesidades #1 Subregiones - Porcentajes")
plt.xlabel("Subregión")
plt.ylabel("Porcentaje")
for container in ax.containers:
    ax.bar_label(container, fmt="%1fXX", label_type="center", color="#FEEBE7", fontsize= 9)
plt.legend(title="Necesidades formativas", bbox_to_anchor=(1.05, 1), loc = "upper left")
plt.show()
```

Fuente: Elaboración propia

Figura 17:

Gráfico distribución necesidades #1 Subregiones - Porcentajes

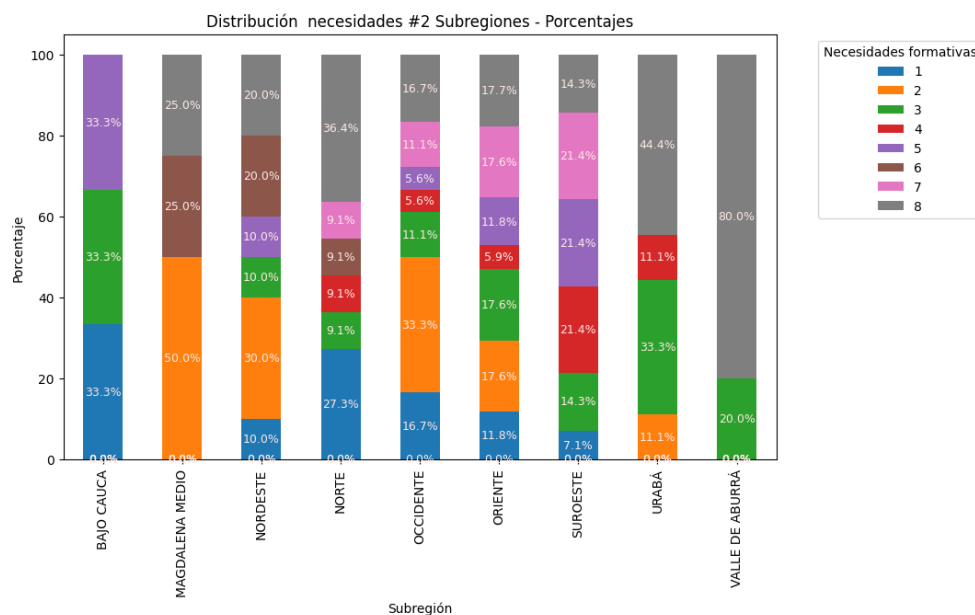


Fuente: Elaboración propia

Una vez generado el gráfico de barras apiladas para comparar proporciones entre las diferentes subregiones del departamento, se repitió el mismo procedimiento por las preguntas relacionadas a la priorización de temáticas formativas 2 y 3.

Figura 18:

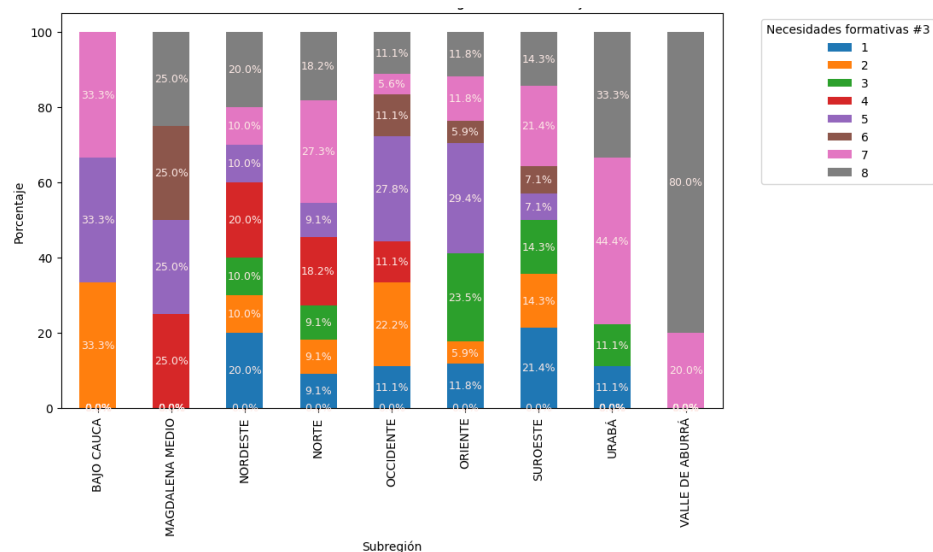
Gráfico distribución necesidades #2 Subregiones - Porcentajes



Fuente: Elaboración propia

Figura 19:

Gráfico distribución de necesidades #3 Subregiones – Porcentajes



Fuente: Elaboración propia

2.2.5. *Influencia Presencia de Espacios e Instancias de Participación Ciudadana Concretos en las Respuestas de Necesidades Formativas de los Municipios del Departamento – Tablas de Contingencia.*

Una vez conocidas las temáticas formativas de cada subregión, con el fin de conocer más información sobre los datos, como por ejemplo la relación de presencia de un espacio o instancia de participación ciudadana con la respuesta de las necesidades formativas de los municipios, es esencial definir las variables explicativas y dependientes, para este ejercicio estas serán la presencia de los espacios o instancias de participación ciudadana presentada en formato binario (1 Activo, 0 Inactivo), y las diferentes necesidades formativas en los municipios del departamento categorizadas en números entre el 1 y el 8.

Siguiendo ejemplos de la aplicación estadística como una herramienta técnica que auxilia al investigador en las ciencias políticas como el de Sánchez Ramos en “Uso metodológico de las tablas de contingencia en la Ciencia Política”, se propone la ejecución de un análisis bivariado entre variables dependientes e independientes categóricas con ayuda de tablas de contingencia, para las cuales se generará un gráfico para demostrar la relación entre la presencia o ausencia de un espacio o instancia específico y la tendencia de respuestas de los diferentes municipios encuestados a las temáticas formativas.

Siendo así, dichas tablas de contingencia son descritas por este autor como una de las técnicas más útiles para describir la covariación entre variables, siendo que el análisis bivariado se apoya en el uso de las tablas de contingencia para ir enfrentando cada variable independiente contra la variable dependiente, es decir, cada celda muestra el número de observaciones que comparten una combinación específica de categorías para ambas variables, permitiendo determinar si existe una asociación entre ellas y cuál es su intensidad. Por ello, el objetivo del ejercicio será comprobar si la proporción de municipios que reporta cierta necesidad formativa como la orientación en veedurías ciudadanas cambia con la presencia o ausencia de diferentes espacios e instancias de participación ciudadana dentro del municipio.

Para ello, como primer paso es necesario revisar el listado de espacios e instancias de participación ciudadana con el fin de elegir que espacios o instancias son prioritarios para las intervenciones de la Secretaría de Participación del departamento en los diferentes municipios.

Tabla 3:

Lista códigos de preguntas por espacios e instancias de participación presentes o ausentes por municipio.

P 1	Política Pública de Participación y/o Comunal	P_2 0	Consejo Municipal de Gestión del Riesgo de Desastres
-----	---	----------	--

P_2	Consejo Municipal de Desarrollo Rural	P_2 1	Mesa Municipal de Derechos Humanos y Derecho Internacional humanitario
P_3	Consejo Territorial de Planeación	P_2 2	Consejo Territorial de Seguridad Social en Salud
P_4	Consejo Municipal de Paz	P_2 3	Comité de Participación Social en Salud COPACO
P_5	Consejo Consultivo de Ordenamiento Territorial	P_2 4	Liga de Usuarios de Servicios de Salud
P_6	Comité Interinstitucional para la Lucha Contra la Trata de Personas	P_2 5	Consejo Territorial de Soberanía y Seguridad Alimentaria
P_7	Mesa Municipal de Mujeres	P_2 6	Mesa Sectorial de Turismo
P_8	Consejo Consultivo Municipal de Mujeres	P_2 7	Comité Municipal de Política Social COMPOS
P_9	Mesa Municipal de Adulto Mayor	P_2 8	Consejo Municipal de Juventud
P_10	Cabildo del Adulto Mayor	P_2 9	Plataforma Municipal de Juventud
P_11	Comité Municipal de Discapacidad	P_3 0	Consejo Consultivo de Diversidad Sexual y de Género
P_12	Mesa Municipal de Primera Infancia	P_3 1	Mesa Diversa
P_13	Consejo Municipal de Cultura	P_3 2	Mesa Municipal Ambiental
P_14	Comité Municipal de Deportes	P_3 3	Veedurías Ciudadanas
P_15	Consejo Municipal de Participación Ciudadana	P_3 4	Juntas Administradoras Locales
P_16	Comité Municipal de Libertad Religiosa y de Culto	P_3 5	Asamblea Municipal Constituyente
P_17	Junta Municipal de Educación	P_3 6	Comité de Planeación y Presupuesto Participativo
P_18	Comité Municipal de Convivencia Escolar	P_3 7	Consejo Consultivo de Planificación de Territorios Indígenas
P_19	Mesa Municipal de Víctimas	P_3 8	Consejos Comunitarios Afrodescendientes

Fuente: Elaboración propia

Una vez listados los espacios e instancias, con el fin de recolectar información de interés para la Secretaría de Participación Ciudadana, se realizará el análisis con 5 espacios e instancias específicos los cuales se relacionan estrechamente con la participación ciudadana y sus funciones, además de alinearse a los proyectos de la entidad.

Siendo así, entendiendo que la Secretaría de Participación Ciudadana de Antioquia es el organismo encargado de organizar y fortalecer el desarrollo social del departamento en coordinación con los municipios, dentro de sus funciones destaca el fortalecimiento de los organismos comunales para su inclusión en los procesos de desarrollo local, por ello es esencial observar la relación de la presencia de instancias como las Juntas Administradoras Locales con las necesidades formativas de los municipios, esto puede llevar a orientar la intervención pública al fortalecimiento de los organismos de participación.

Además, respecto a las Políticas Públicas de Participación, es esencial conocer qué relación tiene la elección de los municipios en torno al componente participativo en la formulación de una Política Pública de Participación en la presencia o ausencia de la misma dentro del municipio, este ejercicio además contribuirá al proceso de recolección de políticas públicas realizado por la Secretaría de Participación, ayudando a descubrir que relación hay entre la ausencia de política pública y la necesidad de formación en torno a sus componentes.

Sobre los Consejos Municipales de Participación Ciudadana, la secretaria de Participación dentro de uno de sus líneas de trabajo enfoca su esfuerzo en el fortalecimiento de la Red Institucional de Municipios con Consejos de Participación (RIAT), por ello es esencial revisar la relación de la presencia de los Consejos de Participación con las necesidades formativas para focalizar las intervenciones futuras entorno al fortalecimiento de la participación ciudadana en las zonas del departamento y determinar si es necesario orientaciones extra en torno a este espacio.

Para terminar, sobre las Veedurías Ciudadanas, es necesaria la revisión de la necesidad de orientación sobre la ley 850 de 2003 en relación a la existencia o presencia de veedurías en los municipios en cuestión, este análisis conducirá a un diagnóstico sobre la necesidad y la voluntad de los municipios en torno al fortalecimiento de dichos entes en el territorio.

Tabla 4:

Instancias o espacios priorizados para analizar

P_1	Política Pública de Participación y/o Comunal
P_15	Consejo Municipal de Participación Ciudadana
P_33	Veedurías Ciudadanas
P_34	Juntas Administradoras Locales
P_36	Comité de Planeación y Presupuesto Participativo

Fuente: Elaboración propia

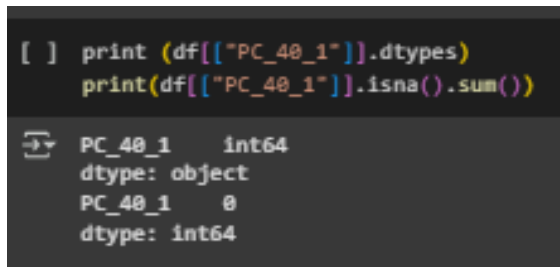
Para revisar la relación entre la presencia de espacios e instancias y las respuestas de los municipios en torno a las necesidades formativas es necesario entender la función Crosstab de Pandas, la cual fue necesaria para la creación de la tabla de frecuencias cruzada para la revisión de la relación entre dos o más variables categóricas en un conjunto de datos.

Siendo así, crosstab o más conocida como `pandas.crosstab` es la encargada de devolver las tablas de contingencia resultantes de cruzar dos o más campos de un dataframe, en este caso se realizará el cruce de las instancias o espacios de participación con las preguntas relacionadas a la temática formativa de los municipios con el fin de revisar que relación existe entre las respuestas y la presencia o ausencia de los mismos.

Para empezar, se comprobó el tipo de datos alojados en las columnas sobre las que se realizará el análisis por medio de:

Figura 20:

Función preparación y comprobación de los Dataframes cargados PC_40_1



```
[ ] print(df[["PC_40_1"]].dtypes)
    print(df[["PC_40_1"]].isna().sum())

In [ ]: PC_40_1    int64
dtype: object
PC_40_1    0
dtype: int64
```

Fuente: Elaboración propia

Por medio de dicha función se revisaron los datos y se entendió que no existen faltantes en la serie “PC_40_1” (Pregunta relacionada a la necesidad formativa #1 priorizada por los municipios que dieron respuesta a la encuesta en el departamento), Además, dicho resultado explica que los datos contenidos en la serie son de tipo Int64 (representa números enteros con signo de 64 bits) y del tipo object (Array o columna de datos que contiene elementos que son objetos de Python, como cadenas de texto, listas o tipos mixtos).

Una vez establecidos los tipos de datos contenidos en la serie PC_40_1, se procedió a la revisión de las columnas relacionadas a las preguntas por las instancias y espacios de participación ciudadana, siendo estas las columnas entre P_1 y P_38., para ello se estableció la siguiente función por la cual se creó una lista entre P_1 y P_38 para trabajar con dichos datos:

Figura 21:

Función revisión y consolidación de las columnas entre P_1 a P_38

```
[ ] Inst_cols = [f"P_{i}" for i in range(1, 39) if f"P_{i}" in df.columns]
Inst_cols
```

Fuente: Elaboración propia

Dicha función es útil para no tener que escribir manualmente los datos entre el P_1 y el P_38, sin embargo, en este caso se utilizó para la aplicación de un Dataframe. Una vez aplicada la función anterior para comenzar el análisis fue necesario designar la serie o columna P_1 como la variable IEN_1 que en este caso significará Instancia – Espacio – Necesidad 1 y hará referencia a la presencia o ausencia de política pública dentro de los diferentes municipios del departamento.

Figura 22:

Función designación columna P_1 a Variable IEN_1

```
[ ] IEN_1 = "P_1"
```

Fuente: Elaboración propia

Una vez establecida la variable a trabajar, se diseñó una función utilizando pandas.crosstab para la creación de una tabla de contingencia o tabla de frecuencias cruzadas entre las variables relacionadas a la pregunta PC_40_1 y P_1 con el fin de identificar patrones. Para ello se ejecutó la siguiente línea de código:

Figura 23:

Función creación de la tabla de contingencia Filas: 0/1 de la instancia; columnas: categorías de necesidad

```
[ ] IEN_1_Conceo_PC_40_1 = pd.crosstab(df[IEN_1], df["PC_40_1"])
```

Fuente: Elaboración propia

Una vez establecida la tabla de contingencia para el conteo de combinaciones entre presencia o ausencia de P_1 con las respuestas de PC_40_1 se estableció la necesidad de presentar la información de manera porcentual para facilitar el análisis de la información, para ello se ejecutó la siguiente función:

Figura 24:

Función conversión del conteo en porcentaje por fila (porcentaje dentro de P_1=0 y P_1=1)

```
[ ] IEN_1_Porcentaje_PC_40_1 = IEN_1_Conceo_PC_40_1.div(IEN_1_Conceo_PC_40_1.sum(axis=1), axis=0).mul(100).round(2)
```

Fuente: Elaboración propia

Una vez designadas las tablas de contingencia, para observarlas de manera grafica se ejecutaron las siguientes líneas de código:

Figura 25:

Función visualización de tablas de contingencia IEN_Cuento_PC_40_1:

```
[ ] print("Conteos por P_1", IEN_1_Cuento_PC_40_1)
    print("/nPorcentajes por P_1", IEN_1_Porcentaje_PC_40_1)

↵ Conteos por P_1 PC_40_1  1  2  3  4  5  6  7
P_1
0      29 10 10 4 9 0 9
1      12 2 0 1 0 1 4
/nPorcentajes por P_1 PC_40_1      1  2  3  4  5  6  7
P_1
0      40.85 14.08 14.08 5.63 12.68 0.0 12.68
1      60.00 10.00 0.00 5.00 0.00 5.0 20.00
```

Fuente: Elaboración propia

Visualizadas las tablas de contingencia, se propuso la creación de un gráfico de barras aplicadas para la visualización de la información de forma que facilite el análisis de los datos, para ello se ejecutó la siguiente función con ayuda de Matplotlib:

Figura 26:

Función creación de gráfico de barras apiladas para las tablas de contingencia P_1 – PC_40_1:

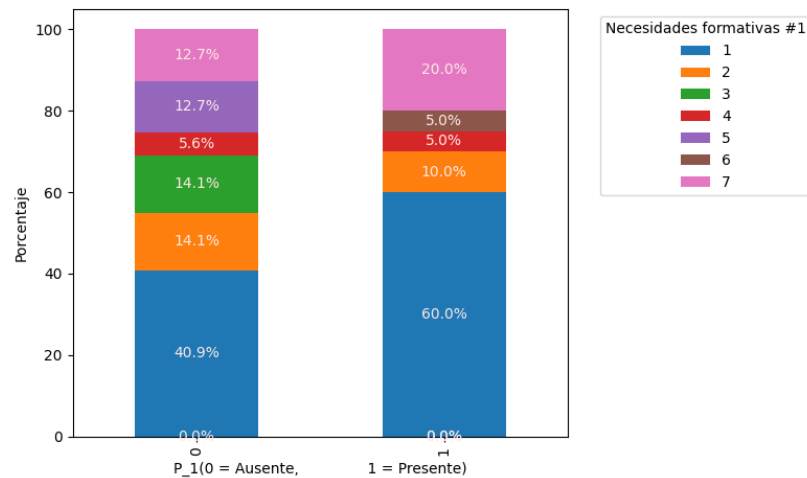
```
[ ] ax = IEN_1_Porcentaje_PC_40_1.plot(kind="bar",stacked=True, figsize=(8,5))
plt.title ("Distribución de PC_40_1 según (p_1) en %")
plt.xlabel (f"{IEN_1}(0 = Ausente, 1 = Presente)")
plt.ylabel("Porcentaje")
for container in ax.containers:
    ax.bar_label(container, fmt="%1.1f%%", label_type="center", color="#FEEBE7",fontsize= 10)
plt.legend(title="Necesidades formativas #1", bbox_to_anchor=(1.05, 1), loc = "upper left")
plt.tight_layout()
plt.show()
```

Fuente: Elaboración propia

Dicha función dio como resultado el siguiente grafico de barras apiladas:

Figura 27:

Gráfico distribución de PC_40_1 según (P_1) en %

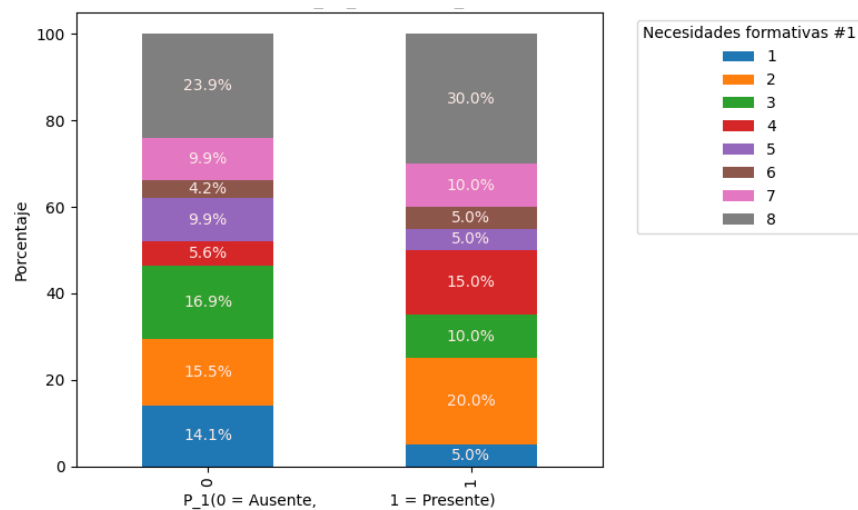


Fuente: Elaboración propia

EL proceso anterior se repitió para los datos relacionados a la priorización de formación #2, la cual dio como resultado el siguiente gráfico:

Figura 28:

Gráfico distribución de PC_40_2 según (P_1) en %

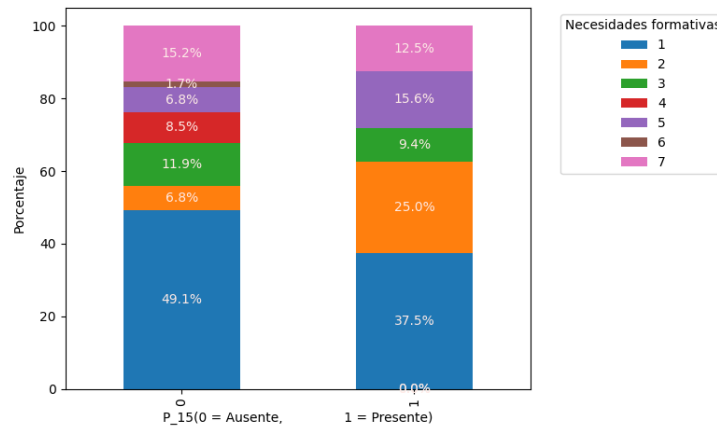


Fuente: Elaboración propia

Los anteriores pasos se repitieron para los espacios e instancias antes mencionados en la tabla 4, es decir, Consejo Municipal de Participación Ciudadana, Veedurías Ciudadanas, Juntas Administradoras Locales y Comité de Planeación y Presupuesto Participativo, lo cual arrojó como resultado los siguientes gráficos:

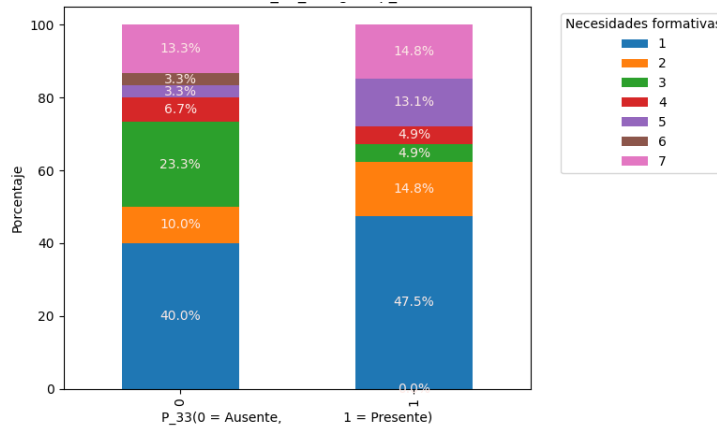
Figura 29:

Gráfico distribución de PC_40_1 según (P_15) en %



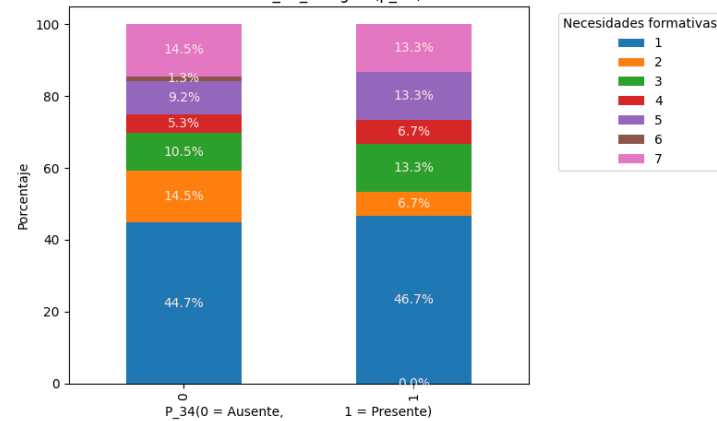
Fuente: Elaboración propia

Figura 30:
Gráfico distribución de PC_40_1 según (P_33) en %



Fuente: Elaboración propia

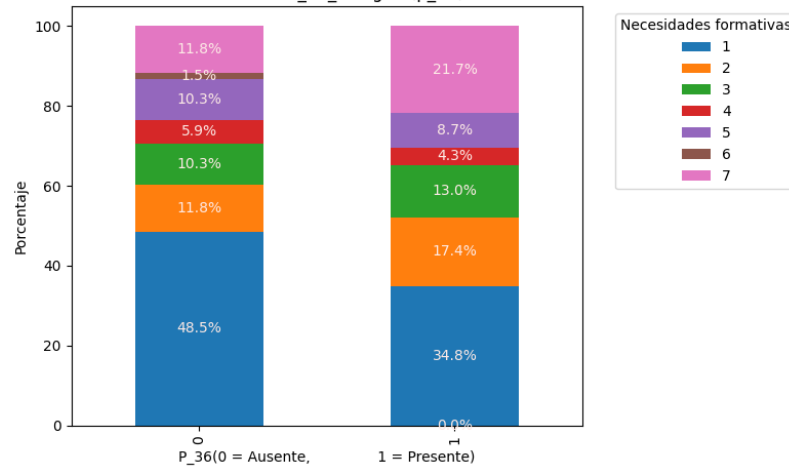
Figura 31:
Gráfico distribución de PC_40_1 según (P_34) en %



Fuente: Elaboración propia

Figura 32:

Gráfico distribución de PC_40_1 según (P_36) en %



Fuente: Elaboración propia

2.2.6. Espacios e Instancias de Participación Ciudadana, Búsqueda de Patrones y Relaciones en los Municipios del Departamento

Una vez visualizadas las relaciones entre presencia o ausencia de espacios específicos y temáticas formativas, una de las formas para poder detectar patrones en los municipios del departamento radica en conocer el porcentaje de presencia o ausencia de todos los espacios o instancias de participación, para ello se propone la ejecución de una tabla donde las columnas representen presencia o ausencia y las filas cada espacio o instancia de participación desde P_1 hasta P_38, además dicha tabla contendrá los valores en forma porcentual con el fin de realizar diferentes análisis de la información resultante.

Para la creación de esta tabla se ejecutó la siguiente función con ayuda de la variable antes definida como `Inst_cols` para la creación de los Dataframes `porcentaje_presencia` y `porcentaje_ausencia`:

Figura 33:

Función creación de los Dataframes de presencia y ausencia de espacios e instancias de participación

```
[86]
✓ 0 s
porcentaje_presencia = df[inst_cols].mean().mul(100).round(2)

porcentaje_ausencia = 100 - porcentaje_presencia
```

Fuente: Elaboración propia

Una vez definidos los Dataframes, se procedió a crear la tabla por medio de la siguiente función:

Figura 34:

Función creación de la tabla de ausencia y presencia de espacios e instancias de participación

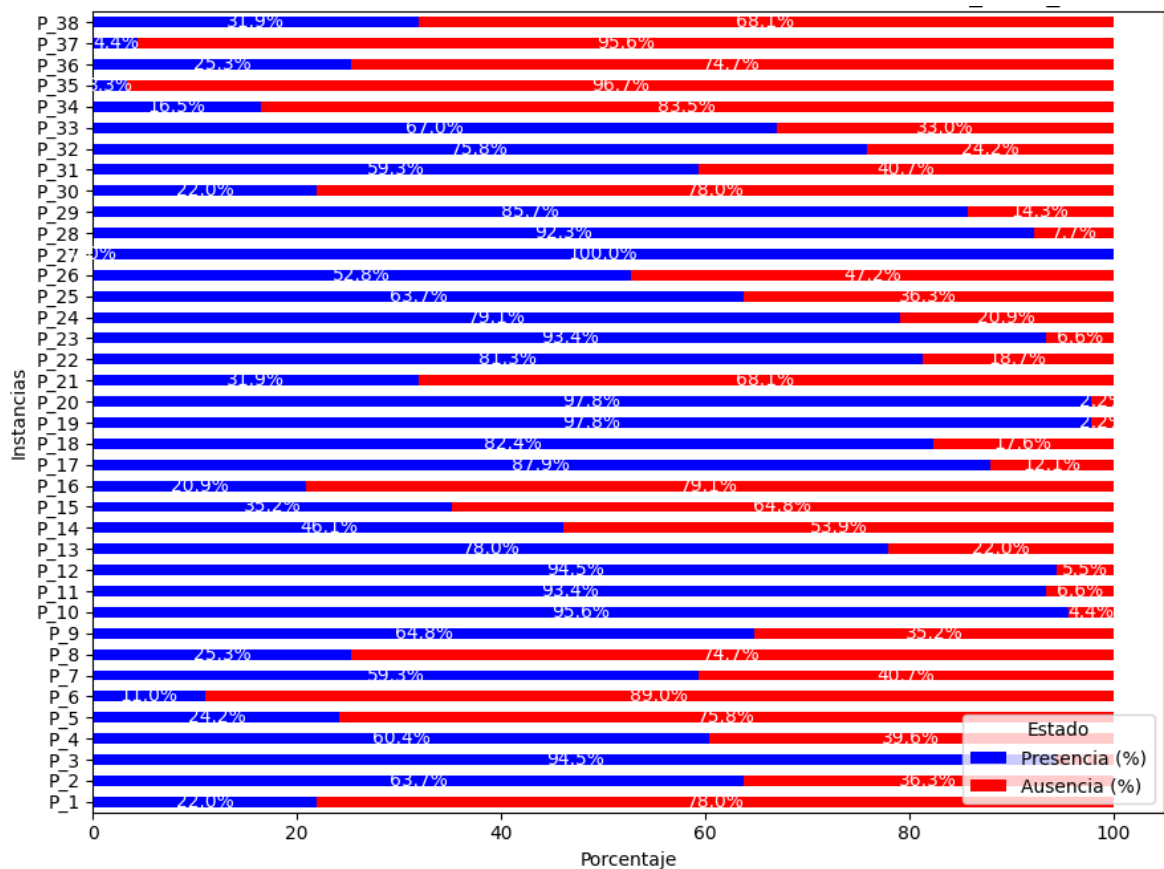
```
[87]
✓ 0s      tabla_porcentajes = pd.DataFrame({"Presencia (%)": porcentaje_presencia, "Ausencia (%)": porcentaje_ausencia })
          print (tabla_porcentajes)
```

Fuente: Elaboración propia

Una vez generada la tabla para visualizar de manera óptima la presencia o ausencia de los espacios e instancias dentro de los municipios encuestados se propuso la creación de un gráfico de barras apiladas con cada una de las instancias y espacios y el porcentaje de presencia o ausencia de las mismas dentro de los municipios 92 municipios que dieron respuesta a la encuesta:

Figura 35:

Gráfico porcentaje de presencia y ausencia de espacios e instancias de participación (P_1 a P_38)



Fuente: Elaboración propia

2.2.7. Espacios e Instancias de Participación – Temáticas Formativas - Tablas de Contingencia

Una vez establecida cual es la opción más optima de formación para cada subregión, los porcentajes de presencia y ausencia de los espacios e instancias de participación dentro de los municipios y la relación entre la presencia o ausencia de ciertos espacios e instancias específicas de participación relacionadas a las necesidades de la secretaria de participación ciudadana dentro del departamento con las temáticas formativas elegidas por los diferentes municipios, con el fin de ampliar el panorama y reconocer la mejor metodología para ofrecer formación dentro de los municipios a partir de posibles patrones, será necesario tomar como unidad de medida a los municipios en relación a las respuestas de temáticas formativas.

Para ello diseñaremos una matriz tipo tabla de patrones comparativos entre necesidades formativas y presencia de instancias y espacios de participación, donde las filas serán las categorías relacionadas a las temáticas formativas y las columnas las instancias de participación desde P_1 hasta P_38, siendo que los valores serán los porcentajes de proporción de municipios que respondieron dicha necesidad y cuentan con la presencia de una instancia específica. Esta tabla será diseñada para detectar diferencias entre las instancias presentes en los municipios que responden cierta temática específica, lo cual ayudará a crear agrupaciones naturales o detectar distinciones.

Siendo así, para la creación de la matriz se realizó una list comprehension¹⁶ en Python llamada `Inst_cols`, la cual fue diseñada para que contenga las columnas que comiencen con `P_`, cuenten con un número válido y dicho número este entre 1 y 38 que son las columnas que hacen referencia a la presencia o ausencia de espacios e instancias de participación. Es decir, la función fue diseñada para mostrar las columnas desde `P_1` hasta `P_38` del DataFrame original en una única variable.

Figura 36:

Función creación de la variable `Inst_Cols`

¹⁶ Una list comprehension hace referencia a una construcción sintáctica en Python que permite crear una nueva lista a partir de una lista existente, aplicando una operación y/o filtro a sus elementos, todo en una sola línea de código.

```
[214] inst_cols = [col for col in df.columns if col.startswith("P_") and col[2:].isdigit() and 1 <= int(col[2:]) <= 38]
```

Fuente: elaboración propia

Una vez contenidas todas las columnas referentes a espacios e instancias en la variable Inst_Cols se procedió a la generación de la matriz con los porcentajes de presencia de los espacios en los municipios del departamento en relación con las temáticas formativas elegidas, para ello se utilizó de nuevo la función groupby de pandas.

Figura 37:

Función matriz Presencia porcentual de instancias y espacios (P_1 a P_38) según necesidad formativa (PC_40_1)”

```
[215] patrones = df.groupby("PC_40_1")[inst_cols].mean().mul(100).round(2)
print(patrones)
```

Fuente: Elaboración propia

Una vez generada la matriz con la presencia porcentual de instancias y espacios según necesidad formativa para la detección de patrones se decidió aplicar un gráfico del tipo heatmap o mapa de calor para detectar la intensidad porcentual de la presencia de diferentes espacios en relación a la respuesta de cierta temática formativa de forma intuitiva. Siendo así, para la creación de dicho grafico fue necesaria la librería de Python Seaborn en conjunto con Matplotlib.

Figura 38:

Función creación del mapa de calor “Presencia porcentual de instancias y espacios (P_1 a P_38) según necesidad formativa (PC_40_1)”

```
plt.figure(figsize=(19, 8))

sns.heatmap (
    patrones, annot=True, fmt=".2f", cmap="Blues", annot_kws={"rotation": 90},
    cbar_kws={"label": "Porcentaje de municipios con la instancia activa"}
)

plt.title("Presencia porcentual de instancias y espacios (P_1 a P_38) según necesidad formativa (PC_40_1)")

plt.xlabel("Instancias y espacios P_1 a P_38")
plt.ylabel("Necesidad formativa")
plt.show()
```

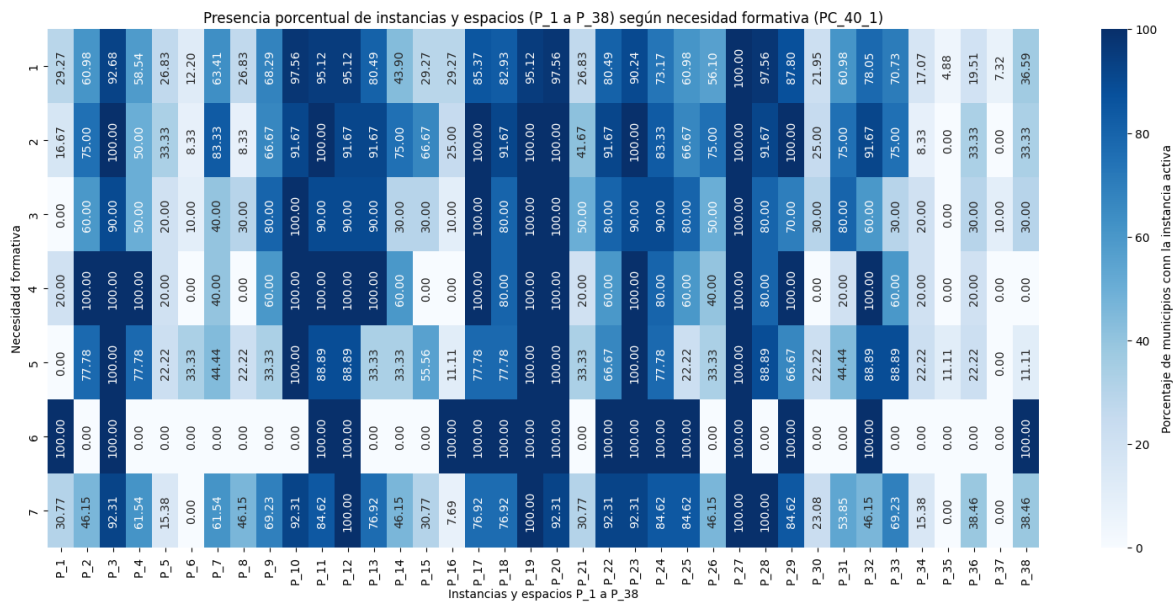
Fuente: Elaboración propia

Una vez ejecutada la función, se generó el mapa de calor para la detección de patrones dentro de los municipios del departamento en relación a las necesidades formativas. Se eligió dicho

grafico debido a que este permite ver con que intensidad hacen presencia las diferentes instancias y espacios de participación dentro de los municipios que respondieron una opción de formación específica:

Figura 39:

Gráfico mapa de calor presencia porcentual de instancias y espacios (P_1 a P_38) según necesidades formativas (PC_40_1)



Fuente: Elaboración propia

Si bien esta información arrojada por el análisis exploratorio de datos es interesante, no es suficiente para establecer si existen o no patrones relacionados a la presencia o ausencia de diferentes espacios e instancias de participación ciudadana y las temáticas formativas elegidas por los diferentes municipios. Para poder establecer esto debemos tener en cuenta que el universo de respuestas se relaciona a múltiples datos y combinaciones, con el fin de encontrar posibles cruces que arrojen patrones se propone la implementación y el uso del machine learning o aprendizaje automatizado.

2.3. Aplicación de Árboles de Decisión de Clasificación para Detección de Patrones – Machine Learning

Siguiendo a Arana, los clasificadores mediante arboles de decisión son modelos no paramétricos de aprendizaje supervisado, cuyo objetivo entre otros es poder predecir a qué clase pertenece un caso del que conocemos uno o más atributos o mediciones.

Siendo así, para autores como Rivero Suguiura los algoritmos de aprendizaje basados en arboles de decisión se consideran uno de los aventajados y más utilizados métodos de aprendizaje supervisado, explica que los datos contienen información oculta potencialmente útil que rara vez se hace explícita o se aprovecha, afirma que el uso de las técnicas de aprendizaje automático buscan entre otras cosas determinar patrones de comportamiento de los datos con el fin de que los usuarios puedan solucionar problemas al tratar analíticamente los datos, ayudando a que estos últimos cuenten con herramientas que les permitan tomar mejores decisiones.

Los árboles de decisión pueden ser concebidos también como modelos de clasificación y regresión empleados en el aprendizaje automático como parte de lo que hoy es la ciencia de datos. Siendo los problemas de clasificación aquellos en los que se intenta predecir los valores de una variable dependiente categórica a partir de ella, siendo así, los árboles de decisión siguen estructuras de árbol similares a los diagramas de flujo, donde un nodo interno representa una característica (o atributo), la rama es una regla de decisión y cada nodo u hoja simboliza el resultado, además el nodo superior se conoce como el nodo raíz. Este tipo de árbol se conoce como árbol de clasificación, que es donde cada ramificación contiene un conjunto de atributos o reglas asociadas a una etiqueta de clase específica que se halla al final de la ramificación.

Siendo así, siguiendo a Rivero, el árbol de decisión clasificatorio aprende a particionar en función del valor del atributo, este se divide de manera recursiva buscando un mínimo local de acuerdo con medidas como la entropía¹⁷ en la base o un conjunto de datos masivos de entrenamiento. Si bien estos árboles presentan ventajas como la capacidad de modelar problemas con múltiples resultados, su fiabilidad cuantificable y su facilidad de aplicarse, es importante denotar que la incorporación de datos categóricos con múltiples niveles provoca que los resultados se inclinen a favor de los atributos con mayoría de niveles, además de que los cálculos pueden volverse complejos al afrontarse con la falta de certezas y numerosos resultados relacionados.

En este caso se decidió utilizar los árboles de clasificación para lograr encontrar reglas o patrones basados en la presencia de espacios e instancias de participación que expliquen o predigan

¹⁷ Entropía hace referencia a la medida de la homogeneidad de los datos aleatorios presentes en un nodo de un árbol de clasificación. Si los datos son similares, el valor de entropía es pequeño o nulo.

la categoría de la presencia de la respuesta a necesidades formativas. Siendo así, para la aplicación del árbol de clasificación, es necesario exponer que se emplearon las siguientes variables:

Variable dependiente categórica: PC_40_1 (Temáticas formativas)

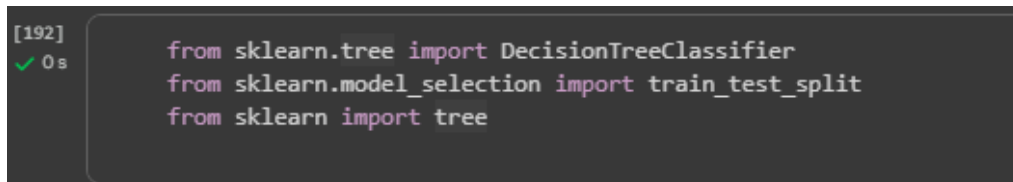
Variables independientes booleanas/binarias: P_1 a P_38 (Espacios e instancias de participación ciudadana)

Además, se utilizará la librería sklearn o scikit - learn la cual es entendida como un conjunto de rutinas escritas en Python para hacer análisis predictivo, dicha librería está basada en NumPy, SciPy y matplotlib, de forma que es fácil reaprovechar el código que use estas librerías. Dentro de la librería se pueden encontrar funciones que ayudan tanto al análisis predictivo como a los algoritmos de clusterización.

En este caso se utilizaron las funciones de DecisionTreeClassifier, train_test_split y tree, para ello se importaron por medio de las siguientes funciones:

Figura 40:

Función importación librerías sklearn para árboles de decisión



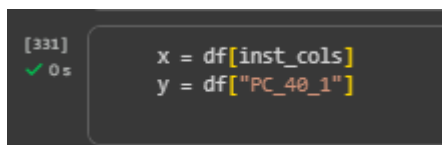
```
[192]
✓ 0s
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import tree
```

Fuente: Elaboración propia

Una vez importadas las funciones necesarias de la librería Sklearn, se procedió a definir las variables X (variable independiente) y Y (variable dependiente), para ello se ejecutó la siguiente función:

Figura 41:

Función definición de variables x, y para árbol de decisión



```
[331]
✓ 0s
x = df[inst_cols]
y = df["PC_40_1"]
```

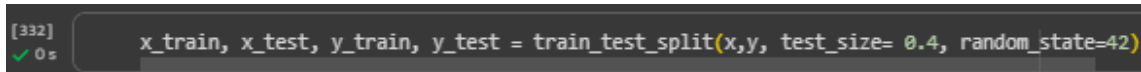
Fuente: Elaboración propia

Una vez definidas las variables, se aplicó la función train_test_split para dividir los datos en dos grupos, en el conjunto de entrenamiento (train) para entrenar el modelo y en el conjunto de prueba (test) para evaluar el modelo. Test_size hace referencia al porcentaje de datos que se dejan

para la prueba, es decir 0.3, lo que significa que el 30% de los datos se dejan para el test y 70% para el train. Ahora bien, `random_state` hace referencia a la semilla aleatoria que controla cómo se dividen los datos. Una vez aplicado el código, lo que hace es dividir la base de 92 municipios en dos partes, para poder entrenar el árbol de decisión con un grupo y luego probar que tan bien predice con municipios que nunca “vio”.

Figura 42:

Función división de datos para entrenamiento y test – árbol de decisión



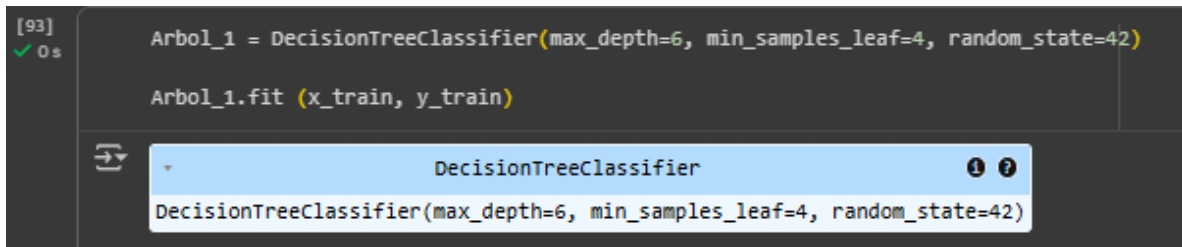
```
[332] x_train, x_test, y_train, y_test = train_test_split(x,y, test_size= 0.4, random_state=42)
```

Fuente: Elaboración propia

Una vez creados los grupos, por medio de la función `DecisionTreeClassifier`, se creó el árbol de decisión con la siguiente función:

Figura 43:

Función creación árbol de decisión para espacios e instancias de participación ciudadana – temáticas formativas



```
[93] Arbol_1 = DecisionTreeClassifier(max_depth=6, min_samples_leaf=4, random_state=42)
Arbol_1.fit (x_train, y_train)
```

DecisionTreeClassifier

DecisionTreeClassifier(max_depth=6, min_samples_leaf=4, random_state=42)

Fuente: Elaboración propia

Se decidió utilizar los parámetros `max_depth = 6`, `min_samples_leaf=4` y `random_state=42` debido a que estos fueron los parámetros que dieron una validez más alta de todas las pruebas, sin embargo, al realizar las pruebas de precisión (accuracy), solo se logró obtener el 0,611%, es decir, las predicciones del árbol estarán acertadas en un 61% de las veces. Para la realización de las pruebas de decisión se utilizaron las siguientes funciones para los datos tanto del test como del train:

Figura 44:

Función precisión del modelo – árbol de decisión espacios e instancias de participación – temáticas formativas

```
[94] ✓ 0 s
accuracy = Arbol_1.score (x_test, y_test)
print ("Preceisión del árbol_1", accuracy)

➡ Preceisión del árbol_1 0.3783783783783784

[95] ✓ 0 s
accuracy = Arbol_1.score (x_train, y_train)
print ("Preceisión del árbol_1", accuracy)

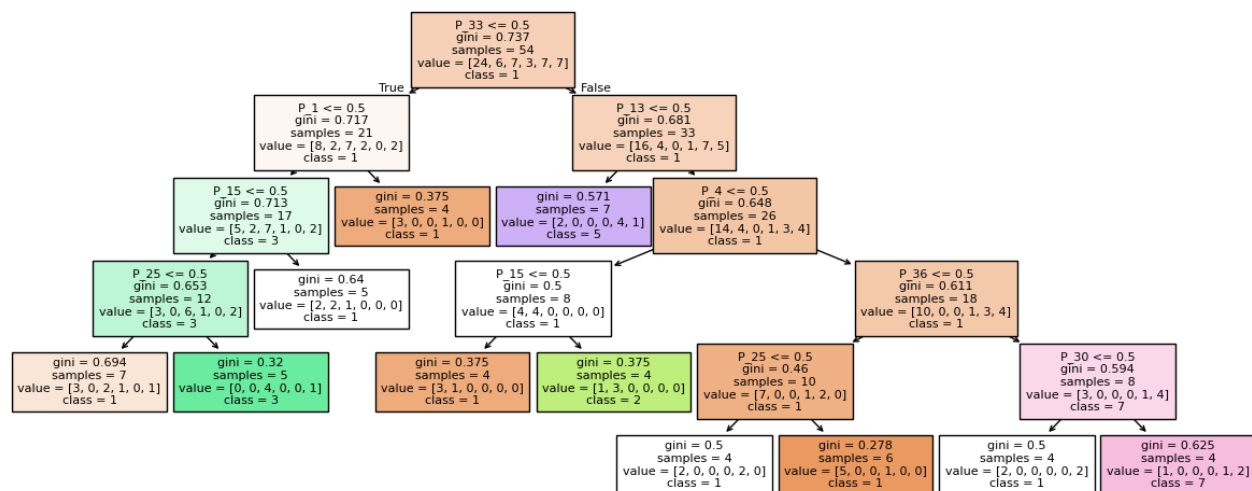
➡ Preceisión del árbol_1 0.6111111111111112
```

Fuente: Elaboración propia

Si bien la precisión o Accuracy siempre debe tender a 1, en este caso debido a la distribución y poca cantidad de datos fue de 0,61, dicho porcentaje se debe a las pocas respuestas de los municipios respecto a las temáticas #4 (5,49%) , #5(9,89%) y #6 (1,10%) dentro de la pregunta PC_40_1, por ende el modelo solo podrá predecir los municipios que tienden a responder las temáticas #1, #2, #3 y #7, contando con una capacidad de generalización del 60%. Siendo así, el resultado del árbol es el siguiente:

Figura 45:

Gráfico árbol de decisión ausencia y presencia de espacios e instancias de participación ciudadana y temáticas formativas



Fuente: Elaboración propia

Estos resultados son interesantes sin embargo debido a la poca cantidad de respuestas de los municipios en la temática formativa PC_40_1, el desbalance de clases y la baja precisión del

modelo, es necesario evaluar otros tipos de aplicación del machine learning para el conjunto de datos.

2.4. Modelos Robustos para Detección de Patrones - Random forest

El random forest es un algoritmo de aprendizaje automático de uso común que combina el resultado de múltiples árboles de decisión para llegar a un resultado único. Siendo así, siguiendo a IBM los algoritmos de bosque aleatorios tienen tres hiperparámetros principales, que deben configurarse antes del entrenamiento, estos incluyen el tamaño del nodo, la cantidad de árboles y la cantidad de características muestreadas.

Siendo así, el algoritmo de random forest se compone de una colección de árboles de decisión y cada árbol en el conjunto está compuesto por una muestra de datos extraída de un conjunto de entrenamiento con reemplazo llamado bootstrapping. Un tercio de la muestra de entrenamiento se reserva como datos de prueba, lo que se conoce como la muestra fuera de la bolsa.

Se decidió optar por probar el algoritmo de random forest debido a que el árbol de decisión simple si bien permite identificar reglas claras de clasificación, en este caso llegó a sobreajustar los datos debido a que el conjunto fue pequeño (92 municipios) y desbalanceado (Clases con 1 o 5 municipios), se esperaba que la aplicación de un Random Forest que combina múltiples árboles de decisión entrados con diferentes subconjuntos de datos y variables mejorara la estabilidad de las predicciones, ofreciendo un modelo más robusto para la clasificación de municipios en temáticas formativas pese a que se pusiera en riesgo la interpretabilidad del modelo.

Para la aplicación del algoritmo, lo primero que se hizo fue la importación de las librerías y las funciones específicas a utilizar, estas fueron:

Figura 46:

Función importación de las librerías para creación de modelos Random Forest

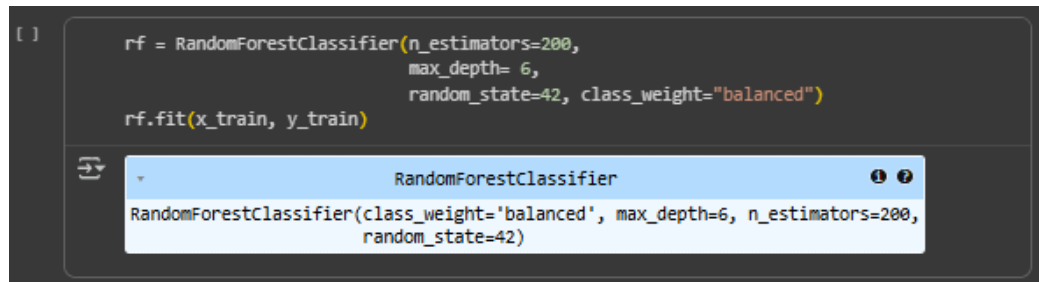
```
[ ] import pandas as pd
    from sklearn.model_selection import train_test_split, cross_val_score
    from sklearn.ensemble import RandomForestClassifier
    from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
    import matplotlib.pyplot as plt
    import seaborn as sns
```

Fuente: Elaboración propia

Una vez importadas las librerías y entendiendo que las variables `x_train` y `y_train` fueron definidas anteriormente en el árbol de decisión de clasificación, se procedió a la creación del Random Forest por medio de la función `RandomForestClassifier`

Figura 47:

Función creación del modelo Random Forest para espacios e instancias de participación – temáticas formativas



```
[ ]
rf = RandomForestClassifier(n_estimators=200,
                           max_depth= 6,
                           random_state=42, class_weight="balanced")
rf.fit(x_train, y_train)
```

Below the code cell, the Jupyter interface shows the object representation of the `RandomForestClassifier` instance:

```
RandomForestClassifier(class_weight='balanced', max_depth=6, n_estimators=200,
                       random_state=42)
```

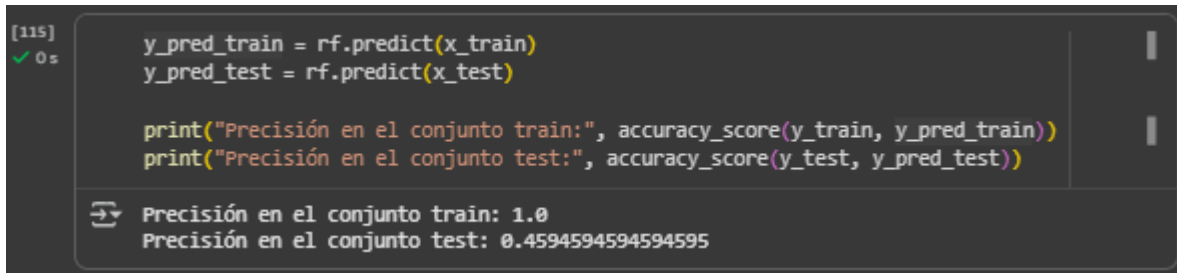
Fuente: Elaboración propia

Por medio de esta función se pretendía crear el modelo de Random Forest, el cual consiste en el entrenamiento de muchos árboles de decisión y combinación de las predicciones para mejorar la precisión y reducir el sobre ajuste, el parámetro `n_estimators` hace referencia al número de árboles en el bosque, se decidió elegir 200 debido a que proporcionan mayor estabilidad a un bajo costo computacional. La profundidad elegida fue de 6, esto debido a que fue la que mayor porcentaje de precisión dio en el anterior árbol simple, la limitación de la profundidad evita además que los árboles crezcan demasiado y memoricen los datos generando overfitting.

Además, entendiendo que las clases están desbalanceadas por la información del árbol anterior, se decidió por introducir el parámetro `class_weight="balanced"`, el cual sería el encargado de ajustar automáticamente los pesos de las clases en función de su frecuencia en los datos, ayudando a que el modelo no ignore a las clases minoritarias. Siendo así, una vez generado el random forest de 200 árboles a una profundidad máxima de 6 y con los pesos de las clases ajustados, se procedió a medir su precisión por medio de la siguiente función:

Figura 48:

Función #1 Precisión Random Forest conjunto train y test



```
[115]
✓ 0s
y_pred_train = rf.predict(x_train)
y_pred_test = rf.predict(x_test)

print("Precisión en el conjunto train:", accuracy_score(y_train, y_pred_train))
print("Precisión en el conjunto test:", accuracy_score(y_test, y_pred_test))
```

```
⇒ Precisión en el conjunto train: 1.0
Precisión en el conjunto test: 0.4594594594594595
```

Fuente: Elaboración propia

Una vez aplicada la función anterior, se puede determinar que, aunque el Random Forest es más robusto que un solo árbol de decisión, todavía memoriza en lugar de aprender patrones generales, por lo cual el modelo sufre de sobreajuste probablemente por las clases desbalanceadas y la poca cantidad de datos. Por ello, se propone el uso de aprendizaje no supervisado como el clustering para detectar tipologías de municipios y contrastarlas con las temáticas formativas.

2.5. Machine learning para Detectar Tipologías de Municipios - Clusters kmodes.

El giro a los enfoques de machine learning no supervisado trae consigo la integración de nuevos métodos como el clustering, dicho método es mencionado por Grimmer Roberts y Stewart en el artículo “Machine learning for social science: An agnostic approach”, quienes partiendo de la disponibilidad de herramientas de machine learning proponen examinar las cuestiones fundamentales en el diseño de las investigaciones.

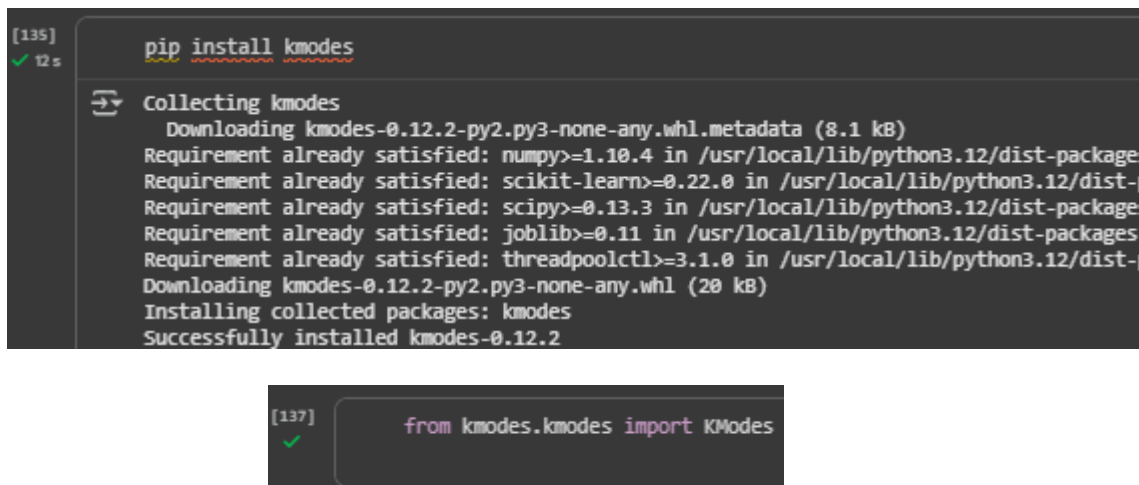
Siendo así, Grimmer, Roberts y Stewart hacen una descripción de las diferentes herramientas de machine learning y su aplicación a las ciencias sociales, entre ellas destacan la agrupación o clustering el cual divide observaciones en categorías mutuamente excluyentes o conglomerados, utilizando los principios de reducción de dimensión no supervisados, dentro de la herramienta de clustering, destaca el algoritmo k – modes, el cual es una modificación del algoritmo k – means, cambiando el uso de la media por la moda para obtener la posición de los centroides en cada iteración y empleando el número de diferencias entre dos registros como medida de la distancia en lugar de la métrica euclídea. Es decir, el algoritmo kmodes es diseñado para agrupar grandes conjuntos de datos categóricos, y tiene por objetivo obtener las kmotas que representan el conjunto.

Siendo así, al aplicar el algoritmo en la base de datos, cada municipio se asigna al cluster cuyo centroide (la moda) tiene menos distancia (hamming distance), es decir busca el cluster más parecido en términos de presencia/ausencia de las variables, luego el centroide de cada cluster no es un promedio si no la moda de las categorías (el valor más frecuente en cada variable). Así pues, con la aplicación del algoritmo se espera agrupar cada municipio según sus instancias de participación y conocer cuál es el perfil típico de cada cluster para realizar análisis y comparaciones. Siguiendo a Grimmer, Roberts y Stewart la aplicación de los métodos de clustering han sido trabajados en diferentes áreas de las ciencias políticas, destacan su uso para categorizar tipos de democracias (Ahlquist y Breunig 2012), estructuras de opinión pública (Blaydes y Linzer 2008) y para países (Wolfson 2004).

Siendo así, para aplicar clusters por K modes es necesario la instalación del paquete kmodes para la importación de las librerías necesarias por medio de las siguiente funciones:

Figura 49:

Función instalación e importación del paquete Kmodes



```
[135] pip install kmodes
✓ 12 s

Collecting kmodes
  Downloading kmodes-0.12.2-py2.py3-none-any.whl.metadata (8.1 kB)
Requirement already satisfied: numpy>=1.10.4 in /usr/local/lib/python3.12/dist-package
Requirement already satisfied: scikit-learn>=0.22.0 in /usr/local/lib/python3.12/dist-
Requirement already satisfied: scipy>=0.13.3 in /usr/local/lib/python3.12/dist-package
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.12/dist-packages
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.12/dist-
Downloading kmodes-0.12.2-py2.py3-none-any.whl (20 kB)
Installing collected packages: kmodes
Successfully installed kmodes-0.12.2

[137] from kmodes.kmodes import KModes
✓
```

Fuente: Elaboración propia

Una vez instalado el paquete kmodes e instaladas las librerías necesarias se configuro un modelo de kmodes con los parametros `n_clusters=4`, que indica el número de clusters o tipologías de municipios, se decidió 4 para evitar desbalance de clases. Se usó además el metodo de iniciación Huang, y `n_init=10`, que significa que el algoritmo va a correr 10 veces con inicializaciones diferentes y se quedará con la mejor solución, es decir, la que tenga menor costo de disimilitud.

Figura 50:

Función creación del modelo Kmodes para los municipios del departamento

```
[141] km = KModes(n_clusters=4, init="Huang", n_init=10, random_state=42)
clusters_KModes = km.fit_predict(x)
```

Fuente: Elaboración propia

Ahora bien por medio de la función `clusters_Kmodes`, lo que se hizo es utilizar la variable `x`, que contiene lo que llamamos anteriormente como `Inst_cols` para ajustar el modelo a los datos y que encuentre centroides categoricos de los clusters, así pues, `predict (x)` asigna cada fila (municipio) al cluster cuyo centroide esté más cercano.

Siendo así, se crea el array con la etiqueta de cluster para cada municipio, para lo cual se procedió a agregar una columna nueva al dataframe original con el número de cluster asignado a cada municipio por medio de la siguiente función:

Figura 51:

Función creación de la columna `Clusters_Kmodes` en el dataframe

```
[142] df['Cluster_KModes'] = clusters_KModes
```

Fuente: Elaboración propia

Una vez asignados los clusters, se hizo un recuento de porcentajes para cada uno de ellos por medio de la función `value_counts` de pandas

Figura 52:

Función proporción de municipios por clusters `Kmodes`

```
print(df["Cluster_KModes"].value_counts(normalize=True).mul(100).round(2))
```

Cluster_KModes	proportion
0	38.46
1	26.37
2	24.18
3	10.99

Name: proportion, dtype: float64

Fuente: Elaboración propia

Una vez determinada la proporción de presencia de los clusters en los municipios que respondieron la encuesta, con el fin de aportar al análisis de la información resultante se aplicó la siguiente función

Figura 53:

Función centroides – modas clusters `Kmode`

```
[145] print(km.cluster_centroids_)
✓ 0s

[[0 1 1 0 0 0 1 0 1 1 1 1 1 1 0 0 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 0 0 0
  0 0]
 [0 1 1 1 0 0 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 0 1
  0 1]
 [0 1 1 1 0 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 1 1 1 0 0 1 1 1 0 0 1 0 0 0 0
  0 0]
 [0 0 1 0 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 0 1 1 1 1 0 1 1 1 1 0 0 0 0 0 0 0
  0 0]]
```

Fuente: Elaboración propia

Dicha función contribuyó a obtener los modos categóricos o modas de cada cluster, es decir, los valores más representativos en cada variable siendo que cada fila corresponde a un cluster, cada columna a un espacio o instancia de participación y el valor entre 0 y 1 a presencia o ausencia.

Una vez obtenidos los centroides, para contribuir al análisis de la información se agrupó el dataframe por el número de clusters asignados por kmodes, es decir, se creó una serie de pandas con los municipios del departamento y sus respectivos clusters por kmodes:

Figura 54:

Función municipios del departamento por clusters de kmode

```
[158] municipios_por_clustersKModes = df.groupby("Cluster_KModes")["Muni"].apply(list)
✓ 0s
print (municipios_por_clustersKModes)
```

```

municipios_por_clustersKModes

Cluster_KModes

0      [ANGELOPOLIS, ARBOLETES, BETULIA, BURITICA, CAÑASGORDAS, CARACOLI,
      CARMEN DE VIBORAL, CAROLINA DEL PRINCIPE, CISNEROS, DABEIBA, DON
      MATIAS, GIRARDO, GRANADA, GUARNE, GUATAPE, ITUANGO, JERICO, LA
      ESTRELLA, LA UNION, MACEO, NECOCLI, PUERTO NARE, PUERTO TRIUNFO,
      SABANALARGA, SAN JERONIMO, SAN JOSE DE LA MONTAÑA, SAN PEDRO, SAN
      ROQUE, SANTAFA DE ANTIOQUIA, TAMESIS, TARSO, TOLEDO, VALPARAISO,
      VEGACHI, YALI]

1      [AMALFI, ANDES, ANORI, CAMPAMENTO, CAREPA, CAUCASIA, CHIGORODO,
      ENVIGADO, FRONTINO, ITAGUI, LA CEJA, MARINILLA, MEDELLIN, PEÑOL, RETIRO,
      SAN CARLOS, SAN JUAN DE URABA, SAN LUIS, SEGOVIA, SONSON, TURBO,
      YARUMAL, YOLOMBO, ZARAGOZA]

2      [ABEJORRAL, ANGOSTURA, BELMIRA, CACERES, CARAMANTA, CONCEPCION,
      EBEJICO, GIRALDO, GOMEZ PLATA, LA PINTADA, MONTEBELLO, MUTATA,
      NARIÑO, OLAYA, REMEDIOS, SAN ANDRES DE CUERQUIA, SAN FRANCISCO,
      SANTA BARBARA, SANTUARIO, SOPETRAN, URRAO, VIGIA DEL FUERTE]

3      [ABRIAQUI, ANZA, ARMENIA, HISPANIA, LIBORINA, MURINDO, PEQUE, SANTO
      DOMINGO, TITIRIBI, URAMITA]

dtype: object

```

Fuente: Elaboración propia

Una vez generados los clusters con el fin de observar si existe relación entre las temáticas formativas y las tipologías de municipios, se agruparon las respuestas de los municipios de cada cluster, para lo cual se utilizó la siguiente función:

Figura 55:

Función respuesta de los clusters para temáticas formativas #1

```

[138] Clusters_PC_40_1 = df.groupby("Cluster_KModes")["PC_40_1"].value_counts(normalize=True).mul(100).round(2)
✓ 0 s

```

Fuente: Elaboración propia

Lo que dio como resultado la siguiente tabla:

Tabla 5:

Clusters – tipologías de municipios y respuesta a temáticas formativas

Cluster_KModes	PC_40_1	Proporción
0	1	51,43%
	2	17,14%
	7	17,14%
	3	5,71%
	4	5,71%

	5	2,86%
1	1	45,83%
	2	16,67%
	7	16,67%
	3	12,5%
	5	8,33%
2	1	31,82%
	5	27,27%
	4	13,64%
	2	9,09%
	3	9,09%
	6	4,55%
	7	4,55%
3	1	50%
	3	30%
	7	20%

Fuente: Elaboración propia

2.6. Análisis y Patrones Detectados de la Aplicación de EDA y Machine Learning

2.6.1. Análisis Necesidades Formativas Departamento y Subregiones

Una vez aplicadas las técnicas del análisis exploratorio de datos en la base de datos de la caracterización de espacios e instancias de participación, en este apartado se realizará a modo de ejemplo el análisis de la información útil resultante de la aplicación de las técnicas.

Siendo así, sobre las tablas de frecuencia representadas por los gráficos sobre las necesidades formativas de los municipios del departamento, es evidente que para la priorización #1 de las temáticas formativas los municipios requieren capacitación en torno a control social y gestión pública (45,05%), seguida por socialización de la Política Pública de Participación Ciudadana y Democrática (14,29%) y Marco conceptual y legal de la participación, y mecanismos de participación ciudadana (13,19%). A partir de esta información se puede inferir que la Secretaría de Participación Ciudadana realizó los encuentros de formación sobre control social con la ESAP teniendo en cuenta la temática formativa con mayor presencia en el departamento, sin embargo, pese a que representan el porcentaje más alto dentro de las necesidades departamentales, estos encuentros no fueron los más óptimos.

Ahora bien, si revisamos la información arrojada por las priorizaciones 2# y #3 encontramos que el porcentaje que predomina en los gráficos es el relacionado a las respuestas sin dato, es decir, la respuesta que hace referencia a los municipios que omitieron y no dieron respuesta a la pregunta, ello siendo en la priorización #2 el 19,78%, y en la priorización #3 la alarmante cifra de 25, 27% de los municipios. A partir de esta información podemos concluir que si realizamos un análisis de estas preguntas correremos el riesgo de llegar a un sesgo de no respuesta¹⁸ que puede conllevar una distorsión de los porcentajes que conduzca a definir que la respuesta a una necesidad formativa represente una importancia mayor de lo que realmente es en el universo, comprometiendo las inferencias resultantes del ejercicio. Por ello solo se tomará en cuenta las respuestas de los municipios a la pregunta PC_40_1 o la priorización #1.

Si bien contamos con los porcentajes generales del departamento, para lograr mejorar las inferencias del ejercicio de caracterización sobre las temáticas formativas más propicias para cada territorio es necesario realizar un análisis para cada subregión del departamento, por ello se diseñaron las tablas de frecuencia con los datos desagregados por cada territorio. Siendo así, a partir de los gráficos generados de estas podemos concluir que, si bien para el departamento la temática formativa más relevante fue Control Social y Gestión Pública, estos encuentros solo serían óptimos para 6 de las 9 subregiones del departamento, siendo en estas donde el porcentaje de municipios que dieron respuesta a la encuesta eligieron mayoritariamente la opción de Control Social y Gestión Pública, las cuales son: Nordeste 60,0%, Urabá 77,8%, Valle de Aburrá 80,0%, Occidente 44,4%, Suroeste 57,1%, y Norte con 36,4%.

Para el resto de las subregiones las opciones más apropiadas como era de esperarse siguiendo la distribución del departamento fueron Socialización de la Política Pública de Participación Ciudadana y Democrática de Antioquia para Bajo Cauca 66,7% y Magdalena Medio 50,0%, y Marco Conceptual y Legal de la Participación y Mecanismos de Participación para Oriente 29,4%. Una vez aclarada esta información, las temáticas formativas más óptimas serían:

Tabla 6:

Temáticas formativas óptimas para cada subregión – pregunta PC_40_1

Subregión	Temática más óptima
Valle de Aburrá	Control Social y Gestión Pública

¹⁸ El sesgo de no respuesta ocurre cuando las personas que no participan en una encuesta o estudio son sistemáticamente diferentes a quienes si lo hacen, de maneras relevantes para los resultados del estudio

Urabá	Control Social y Gestión Pública
Nordeste	Control Social y Gestión Pública
Occidente	Control Social y Gestión Pública
Suroeste	Control Social y Gestión Pública
Norte	Control Social y Gestión Pública
Oriente	Marco conceptual y legal de la participación, y mecanismos de participación ciudadana
Bajo Cauca	Socialización de la Política Pública de Participación Ciudadana y Democrática de Antioquia
Magdalena Medio	Socialización de la Política Pública de Participación Ciudadana y Democrática de Antioquia

Fuente: Elaboración propia

2.6.2. Análisis. Influencia de la Presencia de Espacios e Instancias de Participación Ciudadana en las Respuestas de Necesidades Formativas de los Municipios del Departamento

Para poder analizar las tablas de contingencia creadas para cada instancia y espacio de participación priorizado comenzaremos con la pregunta relacionada con la Política Pública de Participación y/o Comunal, por medio de la tabla de contingencia el resultado esperado es que los municipios que no cuentan con política pública tiendan a elegir la opción formativa relacionada al componente participativo en la formulación de una política pública de participación, sin embargo al ejecutar la función y crear la tabla para la priorización #1 (PC_40_1), la respuesta más común es la relacionada al Control Social y Gestión Pública, no obstante como es de esperarse, es claro que en los municipios con ausencia de Política Pública hay un mayor porcentaje de respuestas en torno a la opción relacionada a la formulación, siendo esta opción respondida por el 12,7% de los municipios que no cuentan con la misma.

Ahora bien, sobre los Consejos Municipales de Participación Ciudadana, es importante destacar que en los municipios con presencia estos no resalta la opción relacionada al Marco conceptual y legal de la participación, y mecanismos de participación ciudadana, ello puede indicar que dentro de los municipios la presencia de este ente consultivo contribuye a la socialización y fortalecimiento tanto de mecanismos como del marco legal de la participación, sin embargo cuando no hay presencia de los Consejos de Participación, los municipios tienden a evidenciar la necesidad de un Marco conceptual y legal de la participación y mecanismos de participación ciudadana en un 25,0%.

Respecto a las Veedurías Ciudadanas, es de esperarse que en los municipios que no cuentan con esta instancia predomine la respuesta relacionada a la ley 850 de 2003, sin embargo al analizar la información resultante de la tabla de contingencia es evidente que dentro de los municipios sigue predominando la respuesta relacionada al Control Social y Gestión Pública, pese a ello en los municipios que no cuentan con Veedurías Ciudadanas la opción 4 relacionada a la ley 850 aumentó en un 1.8%, porcentaje que no llega a ser representativo. Siendo así, el único porcentaje que aumentó significativamente en los municipios que no contaron con Veedurías fue el relacionado a la atención a la ciudadanía desde el enfoque de participación, siendo de un 4,9% en los municipios con presencias de Veedurías y un 23,3% en los municipios sin Veedurías.

Sobre los Comité de Planeación y Presupuesto Participativo es de destacar que las respuestas sobre las necesidades formativas fueron más uniformes en todas sus opciones cuando la instancia hizo presencia en el territorio, sin embargo en los municipios donde no hizo presencia el espacio aumentó el porcentaje de respuestas en las temáticas de Control Social y Gestión Pública más de un 13%, lo que puede evidenciar que instancias como los comités de planeación y presupuesto participativo pueden llegar a dotar a los municipios de cierta confianza en lo público, por lo cual no es tan necesaria una capacitación en control social y pueden centrarse en temas como la socialización de la política pública de participación ciudadana y democrática que aumentó en un 10%.

Para terminar, sobre las Juntas Administradoras Locales sucede algo interesante, esto es que la distribución de respuestas es bastante similar, resaltando únicamente que en los municipios con ausencia de Juntas se precisa más capacitación en torno al Marco conceptual y legal de la participación, y mecanismos de participación ciudadana que en los que cuentan con dicho espacio,

sin embargo, con presencia o ausencia la respuesta con mayor porcentaje sigue siendo la necesidad de capacitación en torno a Control social y gestión pública.

2.6.3. Análisis. Espacios e Instancias de Participación Ciudadana, Búsqueda de Patrones y Relaciones en los Municipios del Departamento:

Para encontrar patrones relacionados a la presencia o ausencia de instancias y espacios de participación dentro de los municipios encuestados del departamento se creó la tabla “Porcentaje de Presencia y Ausencia de Instancias y Espacios (P_1 a P_38)”, a partir de la cual se pueden sacar conclusiones como:

- El espacio de participación con menor presencia en los municipios del departamento es la Asamblea Nacional Constituyente o Asambleas Constituyentes Territoriales con un porcentaje de 3.30% de presencia en el municipio, lo cual puede deberse a que fueron iniciativas surgidas en 1997 en Colombia a nivel local, municipal, regional y departamental, por lo cual puede que las formas de participar hayan mutado a otros espacios de participación ciudadana.
- El espacio de participación con mayor presencia en los municipios del departamento es la Mesa Municipal de Víctimas con una presencia del 97,80% en los municipios encuestados. Esto indica que dentro del departamento la población víctima del conflicto tiende a tener un espacio para su participación efectiva, además dicha proporción va en consonancia con el cumplimiento de las disposiciones contenidas en la ley 1448 de 2011 y sus decretos reglamentarios, conformados por Organizaciones de Víctimas (OV), Organizaciones Defensoras de Víctimas (ODV) y víctimas designadas por las autoridades tradicionales de comunidades étnicas.
- La instancia de participación con menor presencia en los municipios del departamento es el Consejo consultivo de planificación de territorios indígenas con 4.40% de presencia respecto a los 92 municipios encuestados. Esto es interesante porque pese a existir diferentes étnicas como las Embera Eyábida, Embera Chamí, en municipios como El Bagre, Remedios, Anorí, Puerto Berrio o zonas como el Magdalena Medio, Bajo Cauca o el Nordeste, hay muy poca presencia del espacio designado y clave para la planificación de proyectos y políticas para la protección y desarrollo de sus territorios.

- La instancia de participación con mayor presencia en los municipios es el Comité municipal de política social COMPOS con un porcentaje del 100%, esto puede deberse al marco legal detrás de dicha instancia, la cual es respaldado por ley 489 de 1998 artículo 6; el Decreto 1137 de 1999, la ley 136 de 1994, la ley 1098 de 2006, artículo 202 y 207.

2.6.4. *Arbol de Decision y Random Forest*

Con el fin de ampliar el análisis, presentar un panorama general y demostrar si existen patrones relacionados a la presencia de espacios e instancias específicas dentro de los municipios que eligen un tipo de temática formativa, se propuso la creación de los árboles de decisión de clasificación utilizando machine learning, para los cuales se logró detectar que para las respuestas de los municipios sobre priorización de necesidades formativas #1 (PC_40_1):

- Para comenzar, el nodo raíz (donde comienza la clasificación) es la instancia P_33 (Veedurías Ciudadanas), es decir, si P_33 es igual o menor a 0.5 (ausencia) se clasifica a la izquierda mientras que si es mayor o igual a 0.5 (presencia) se clasificará al lado derecho, dentro de este nodo recaen 54 municipios, de los cuales la mayoría pertenece a la clase 1, sin embargo, el índice Gini es muy alto 0.737, por lo cual la pureza es muy baja.

Siendo así, si seguimos hacia la izquierda (true) encontraremos el sub nodo P_1, es decir, cuando los municipios no cuentan con P_33 (Veedurías Ciudadanas) ni P_1 (Política Pública de Participación y/o Comunal) recaen 21 municipios y tienden a elegir la clase 1 (Control social y gestión pública), aquí el índice Gini 0.717 sigue siendo alto lo cual indica que hay bastante mezcla entre las clases, por lo cual el árbol seguirá clasificando nodos. Sin embargo, encontramos la primera hoja al lado derecho del nodo, la cual indica que si hay ausencia de P_33 pero presencia de P_1 los municipios tienden a elegir la clase 1, esta afirmación va acompañada de 4 ejemplos y un índice Gini de 0,375 lo cual refleja una mezcla de clases, donde la mayoría (3) son de la clase 1, pero hay un caso distinto (1 municipio en clase 4).

Continuando con el lado izquierdo, cuando no hay presencia de P_33, P_1 el próximo nodo es P_15 (Consejo Municipal de Participación Ciudadana), en el cual recaen 17 municipios que tienden a elegir la clase 3 (Atención a la ciudadanía desde el enfoque de participación), dicho nodo cuenta con un índice Gini de 0.713 el cual es muy alto, sin embargo, surge una hoja que indica que

si hay ausencia de P_33 y P_1 pero presencia de P_15 los municipios tienden a elegir la clase 1 con 5 ejemplos.

El próximo nodo del lado izquierdo sorprendentemente es el de P_25 (Consejo territorial de soberanía y seguridad alimentaria), el cual abre una bifurcación según presencia/ausencia de P_25, lo cual indica que si hay ausencia de P_33, P_1, P_15 y P_25 los municipios tienden a la clasificación 1 con 7 ejemplos y un Gini de 0.694, sin embargo, si hay presencia de P_25 tienden a la clasificación 3 con 5 ejemplos y un Gini mucho más bajo (0.32).

- Volviendo al nodo raíz, para el camino de presencia, es decir el de false encontramos que cuando hay presencia de P_33 y P_13 (Consejo municipal de cultura) encontramos 33 municipios que tienden a elegir la clase 1, sin embargo, hay una presencia de Gini alta (0.68). Ahora, cuando hay ausencia de P_13 pero presencia de P_33 el camino guía a los municipios a elegir la clase 5 (Componente participativo en la formulación de una política pública de participación).

Continuando con el lado derecho, el siguiente sub nodo es el P_4 (Consejo municipal de paz), el cual cuenta con 26 ejemplos de municipios que tienden a elegir la clase 1 y un Gini de 0.648, dicho sub nodo no tiene hojas, sin embargo, por su índice Gini se divide en otros dos nodos que dependen de la presencia o ausencia de P_4. Siendo que cuando hay ausencia de este la mayoría de los municipios tienden a elegir la clase 1 con 8 ejemplos y un Gini de 0.5 y continua el sub nodo de P_15, el cual a partir de su presencia los municipios tendieron a responder la clase 2 con un Gini de 0.375 y 4 ejemplos y con su ausencia tendieron a elegir la clase 1 con un Gini de 0.375 y 4 ejemplos.

Ahora bien, volviendo al sub nodo P_4, su presencia deriva en el sub nodo P_36, el cual cuenta con 18 ejemplos que tienden a elegir la clase 1 y un Gini alto 0.611 por lo cual este vuelve a dividirse en dos sub nodos, el P_30 del cual se derivan diferentes hojas, las cuales indican que con la presencia de P_30 los municipios tienden a elegir la clase 7, mientras que con su ausencia tienden a elegir la clase 1. Con la ausencia de P_36 se deriva el sub nodo P_25, del cual se desprenden dos hojas las cuales indican que con su presencia los municipios tienden a elegir la clase 1 con un Gini de 0.278 y 4 ejemplos y con su ausencia tienden a elegir igual la clase 1 solo que con un Gini más alto 0.5 y 4 ejemplos.

Una vez analizado el árbol de decisión, se pueden evidenciar reglas como:

Lado derecho (False):

$P_{_33} = 1 + P_{_13} = 1 + P_{_4} = 1 + P_{_36} = 1 + P_{_30} = 1 = \text{Clase 7 (Gini 0.6, Samples: 4)}$

$P_{_33} = 1 + P_{_13} = 1 + P_{_4} = 1 + P_{_36} = 1 + P_{_30} = 0 = \text{Clase 1 (Gini 0.5: Samples 4)}$

$P_{_33} = 1 + P_{_13} = 0 = \text{Clase 5 (Gini 0.571, Samples: 7)}$

$P_{_33} = 1 + P_{_13} = 1 + P_{_4} = 1 + P_{_15} = 0 = \text{Clase 1 (Gini 0.375, Samples: 4)}$

$P_{_33} = 1 + P_{_13} = 1 + P_{_4} = 0 + P_{_15} = 1 = \text{Clase 2 (Gini 0.375, Samples: 4)}$

$P_{_33} = 1 + P_{_13} = 1 + P_{_4} = 1 + P_{_36} = 1 + P_{_25} = 0 = \text{Clase 1 (Gini 0.5, Samples: 4)}$

$P_{_33} = 1 + P_{_13} = 1 + P_{_4} = 1 + P_{_36} = 1 + P_{_25} = 1 = \text{Clase 1 (Gini 0.278, Samples: 6)}$

Lado izquierdo (True):

$P_{_33} = 0 + P_{_1} = 0 + P_{_15} = 0 + P_{_25} = 0 = \text{Clase 1 (Gini 0.694, Samples 7)}$

$P_{_33} = 0 + P_{_1} = 1 = \text{Clase 1 (Gini 0.375, Samples 4)}$

$P_{_33} = 0 + P_{_1} = 0 + P_{_15} = 1 = \text{Clase 1 (Gini 0.64 Samples 5)}$

$P_{_33} = 0 + P_{_1} = 0 + P_{_15} = 0 + P_{_25} = 1 = \text{Clase 3 (Gini 0.32, Samples 5)}$

Es necesario aclarar que estas predicciones solo tienen la capacidad de generalización del 60% en train y 37% en test, porcentaje insuficiente e indicativo de que el árbol está sobreajustado, es decir, que el árbol aprendió demasiado los patrones del train, pero no generaliza suficiente al test. Siendo así, al optar por modelos más robustos como el Random Forest, evidenciamos que si bien los datos de train dan un porcentaje del 100% el modelo sigue arrojando porcentajes bajos de precisión en test, siendo estos de 45%, esto se debe a los pocos datos (92 municipios) y muchas clases (7), que a su vez están desbalanceadas (clase 1 con 41 casos, clase 4 con 5 y clase 6 con 1), esto puede ser indicativo que los modelos supervisados terminan sobreajustando, por ello se optó por dar un giro hacia el uso de enfoques no supervisados para descubrir tipologías emergentes de municipios a partir de sus características con el fin de observar y comparar los grupos con las temáticas formativas para observar concordancias y discrepancias.

2.6.5. Análisis. Clasificación de Municipios por Tipologías – Clusters kmodes

Una vez aplicados los clustering por el algoritmo de kmodes se definieron 4 tipos de municipios a partir de la presencia o ausencia de instancias o espacios de participación, siendo así los modos categóricos de cada cluster o centroides (valores más representativos en cada variable para kmodes) fueron los siguientes:

Perfiles típicos de los municipios en cada grupo:

Cluster 0 Presencia de: P_2, P_3, P_7, P_9, P_10, P_11, P_12, P_13, P_14, P_17, P_18, P_19, P_20, P_22, P_23, P_24, P_25, P_26, P_27, P_28, P_29, P_31, P_32, P_33. **Ausencia de:** P_1, P_4, P_5, P_6, P_8, P_15, P_16, P_21, P_30, P_34, P_35, P_36, P_37, P_38 - **38,46% de los municipios**

Cluster 1 Presencia de: P_2, P_3, P_4, P_7, P_9, P_10, P_11, P_12, P_13, P_14, P_15, P_17, P_18, P_19, P_20, P_21, P_22, P_23, P_24, P_25, P_26, P_27, P_28, P_29, P_31, P_32, P_33, P_36, P_38 **Ausencia de:** P_1, P_5, P_6, P_8, P_16, P_30, P_34, P_35, P_37. - **26,37% de los municipios**

Cluster 2 Presencia de: P_2, P_3, P_4, P_10, P_11, P_12, P_17, P_18, P_19, P_20, P_22, P_23, P_24, P_27, P_28, P_29, P_32, **Ausencia de:** P_1, P_5, P_6, P_7, P_8, P_9, P_13, P_14, P_15, P_16, P_21, P_25, P_26, P_30, P_31, P_33, P_34, P_35, P_36, P_37, P_38. - **24,18% de los municipios**

Cluster 3 Presencia de: P_3, P_7, P_8, P_9, P_10, P_11, P_12, P_13, P_17, P_18, P_19, P_20, P_22, P_23, P_24, P_25, P_27, P_28, P_29 **Ausencia de:** P_1, P_2, P_4, P_5, P_6, P_14, P_15, P_16, P_21, P_26, P_30, P_31, P_32, P_33, P_34, P_35, P_36, P_37, P_38. - **10,99% de los municipios.**

Siendo así, los municipios en cada grupo o cluster fueron los siguientes:

Tabla 7:

Clusters de municipios por KModes

Cluster KModes	Municipios
0	['ANGELOPOLIS', 'ARBOLETES', 'BETULIA', 'BURITICA', 'CAÑASGORDAS', 'CARACOLI', 'CARMEN DE VIBORAL', 'CAROLINA DEL PRINCIPE', 'CISNEROS', 'DABEIBA', 'DON MATIAS', 'GIRARDOTA', 'GRANADA', 'GUARNE', 'GUATAPE', 'ITUANGO', 'JERICO', 'LA ESTRELLA', 'LA UNION', 'MACEO', 'NECOCLI', 'PUERTO NARE', 'PUERTO TRIUNFO', 'SABANALARGA', 'SAN JERONIMO', 'SAN JOSE DE LA MONTAÑA', 'SAN PEDRO', 'SAN ROQUE', 'SANTAFE DE ANTIOQUIA', 'TAMESIS', 'TARSO', 'TOLEDO', 'VALPARAISO', 'VEGACHI', 'YALI']
1	['AMALFI', 'ANDES', 'ANORI', 'CAMPAMENTO', 'CAREPA', 'CAUCASIA', 'CHIGORODO', 'ENVIGADO', 'FRONTINO', 'ITAGUI', 'LA CEJA', 'MARINILLA', 'MEDELLIN', 'PEÑOL', 'RETIRO', 'SAN CARLOS', 'SAN JUAN DE URABA', 'SAN LUIS',

	'SEGOVIA', 'SONSON', 'TURBO', 'YARUMAL', 'YOLOMBO', 'ZARAGOZA']
2	['ABEJORRAL', 'ANGOSTURA', 'BELMIRA', 'CACERES', 'CARAMANTA', 'CONCEPCION', 'EBEJICO', 'GIRALDO', 'GOMEZ PLATA', 'LA PINTADA', 'MONTEBELLO', 'MUTATA', 'NARIÑO', 'OLAYA', 'REMEDIOS', 'SAN ANDRES DE CUERQUIA', 'SAN FRANCISCO', 'SANTA BARBARA', 'SANTUARIO', 'SOPETRAN', 'URRAO', 'VIGIA DEL FUERTE']
3	['ABRIAQUI', 'ANZA', 'ARMENIA', 'HISPANIA', 'LIBORINA', 'MURINDO', 'PEQUE', 'SANTO DOMINGO', 'TITIRIBI', 'URAMITA']

Fuente: Elaboración propia

Con el fin de contrastar si existe relación entre la presencia o ausencia de espacios e instancias de participación ciudadana, se distribuyeron estos clusters a partir de su respuesta a la pregunta PC_40_1, para lo cual se puede observar que no hay patrones concluyentes entre la distribución de los municipios por clusters y su elección por temática formativa (PC_40_1), por lo cual se podría asegurar que no existen patrones claros o concluyentes entre la relación de presencia de espacios o instancias de participación y la respuesta de los municipios a las temáticas formativas. Sin embargo, al realizar la tipología de municipios, se pueden destacar diferentes necesidades formativas alternas a partir de la presencia o ausencia de instancias de participación teniendo en cuenta además los alcances limitados de la Secretaria de Participación Ciudadana, estas serían: P_1 (Política Publica de Participación), P_35 (Asambleas Municipales Constituyente), P_34 (Juntas Administradoras Locales), P_30 (Consejo Consultivo de Diversidad Sexual y genero) y P_37 (Consejo consultivo de Planificación de Territorios Indígenas), las cuales repiten su ausencia en los 4 grupos de municipios. Por ello, se sugiere implementar formación o fortalecimiento en:

- Fortalecimiento en Políticas Públicas de Participación Ciudadana y Comunitaria
- Formación en participación diversa e incluyente con enfoque de genero
- Fortalecimiento de organismos comunales y de asociación con enfoque en juntas administradoras locales
- Divulgación en torno al espacio de participación ciudadana llamado asambleas municipales constituyentes
- Fortalecimiento en organismos consultivos enfocados en planificación de territorios indígenas

2.7. Reflexión Crítica

2.7.1. Analítica de Datos y Análisis Exploratorio de Datos

A modo de reflexión crítica, la falta de pasos a seguir y metodologías claras en el ejercicio de caracterización de espacios e instancias de participación ciudadana aplicada por el observatorio de participación de la Gobernación de Antioquia presenta un vacío importante y una gran oportunidad para la implementación de analítica de datos para el manejo adecuado de la información resultante de la misma en pro de obtener inferencias cualificadas que permitan una adecuada intervención de las entidades públicas, en este caso la Secretaría de Participación Ciudadana. Siendo así, dentro de este trabajo se intentó plasmar una ruta clara de tratamiento de la información, comenzando por la preparación de los datos hasta la aplicación de análisis exploratorio de datos y machine learning para detección de patrones y agrupación de tipologías de municipios, todo con el fin de mostrar un camino claro y un código que pueda ser perfeccionado con el fin de conducir a una cualificación fuerte de las inferencias resultantes del ejercicio de caracterización de espacios e instancias de participación, entendiendo además que el Observatorio de Participación como herramienta de política pública debe brindar información relevante y acertada sobre los municipios del departamento como lo plasma la Ordenanza 025 de 2023.

La integración de analítica de datos y el análisis exploratorio de datos brindó información útil y de calidad, como el alcance real de la encuesta, destacando el número de municipios encuestados, su nombre, su desagregación en cada subregión y el porcentaje de cobertura de la encuesta, lo que permitió identificar que, si bien la encuesta llegó al 73,6% de los municipios del departamento, en subregiones como Bajo Cuaca o el Valle de Aburrá solo alcanzó el 50% de los municipios. Si bien esto puede deberse a que al inicio no había un designado específico para el diligenciamiento de las encuestas en los territorios, resulta importante y necesario un ejercicio continuo de fortalecimiento a la divulgación y la importancia de la encuesta en los municipios del departamento con el fin de obtener la mayor cantidad posible de información y con ella las inferencias más acertadas y reales posibles para las intervenciones públicas.

Respecto a las necesidades formativas, el análisis exploratorio de datos contribuyó a determinar que necesidades resultaban siendo más comunes en el departamento y en cada subregión, permitiendo cualificar las inferencias para el proceso de realización de intervenciones

formativas por parte de la Gobernación de Antioquia, específicamente la Secretaría de Participación y Cultura Ciudadana y la ESAP. Además, la aplicación del análisis exploratorio de datos contribuyó por medio de las tablas de contingencia a determinar para ciertos espacios o instancias de participación ciudadana relevantes para la Secretaría de Participación y Cultura Ciudadana en su accionar, si existía relación en su presencia o ausencia respecto a las respuestas de los municipios a las temáticas formativas. Lo cual derivó en conclusiones como que dentro de los municipios que no existe política pública hay una mayor tendencia a responder la opción formativa relacionada a la formulación de componentes de la misma, o que por ejemplo los municipios que no cuentan con veedurías ciudadanas si bien en su mayoría respondieron por la necesidad formativa relacionada a la atención a la ciudadanía desde el enfoque de participación, los que no cuentan con veedurías tienden a responder mucho más la opción relacionada a la ley 850 de 2003 (Veedurías Ciudadanas). Esta información proporciona pistas esenciales para que los entes públicos intervengan en pro del fortalecimiento de los espacios e instancias de participación ciudadana.

2.8. Necesidades Formativas y Encuentros de Capacitación – Cualificación de Inferencias

Sobre la cualificación de inferencias para un adecuado accionar publico podemos concluir partiendo de que la Secretaría de Participación Ciudadana a partir de la caracterización de espacios e instancias de participación realizó diferentes intervenciones públicas como encuentros de capacitación en las subregiones del departamento, específicamente en los municipios de La Ceja (Oriente) , San Jerónimo (Occidente) , El Bagre (Bajo Cauca) , Venecia (Suroeste) , Puerto Berrío (Magdalena Medio) , Yarumal (Norte) y Yolombó (Nordeste), sobre la temática control social y gestión pública junto a la ESAP debido a que en un primer momento fue la opción que presentó el porcentaje más alto a nivel departamental, el análisis exploratorio de datos nos ayuda a esclarecer que las temáticas de estos encuentros no fueron las más óptimas en todas las subregiones del departamento.

Ello debido a que al aplicar los gráficos de distribución de necesidades para PC_40_1, se logró concluir que, si bien para las subregiones del Valle de Aburrá, Urabá, Nordeste, Occidente, Suroeste y Norte si fue adecuada la aplicación de control social y gestión pública puesto que esta opción presentó un porcentaje más alto de municipios, para subregiones como Bajo Cauca y

Magdalena Medio era necesaria capacitación en torno a socialización de la política pública de participación ciudadana y democrática de Antioquia y para el Occidente marco conceptual y legal de la participación y mecanismos de participación. Esta información obtenida a partir del análisis exploratorio de datos es un ejemplo claro de la capacidad de estos métodos para cualificar las intervenciones y las metodologías llevadas a cabo por entidades públicas como el OPC, en este caso se diseñó un paso a paso y un código para el análisis de los datos con herramientas de gran utilidad como Python, las cuales logran esclarecer información de gran utilidad para los diferentes tomadores de decisiones y la ciudadanía en general, sin embargo, es necesario un perfeccionamiento partiendo de las necesidades específicas de las entidades públicas.

2.9. Modelos de Machine Learning, Pertinencia Espacios/Instancias – Temáticas Formativas

Una vez aplicado el análisis exploratorio de datos y detectadas las temáticas formativas más óptimas para cada subregión del departamento, con el fin de expandir el análisis y buscar patrones que lleven a información de utilidad para la administración en general sobre los datos resultantes de la caracterización de espacios e instancias de participación, se aplicaron diferentes modelos de aprendizaje supervisado de machine learning. Estos fueron diseñados con el fin de generar inferencias y predicciones validas demostrando la relación entre la presencia o ausencia de los 38 espacios e instancias de participación y las respuestas de los municipios a las necesidades de temáticas formativas, intuyendo que las carencias de ciertos espacios serian el detonante directo de la elección en ciertas necesidades formativas específicas. Sin embargo, al aplicar los modelos de aprendizaje supervisado debido a la poca cantidad de datos producto de la no respuesta del 30% de los municipios del departamento, los modelos no fueron capaces de predecir de forma adecuada.

Es decir, con la aplicación del árbol de decisión de clasificación se logró obtener un máximo de precisión del 61% en los datos de entrenamiento, mientras que en los datos de test solo se alcanzó el 37%, esto es indicativo claro de sobreajuste debido a la poca cantidad de datos y a las clases desbalanceadas, sin embargo, se intentó realizar el mismo ejercicio con un modelo más robusto como el Random Forest, el cual logró aumentar la precisión de los datos de entrenamiento al 100% y los datos test a 45%, pese a ello, esto también es indicativo de sobreajuste, es decir, el modelo no lograba realizar predicciones claras debido a la poca cantidad de datos y clases desbalanceadas

probando que el problema no era el modelo, los parámetros ni el algoritmo que se utilizaba si no la cantidad de datos y la distribución de sus clases. Teniendo en cuenta esto, es importante destacar que los modelos predictivos no deben ser descartados como pasos importantes para futuras iniciativas emprendidas por entes como el Observatorio de Participación Ciudadana, la falta de datos y su distribución no permitieron que las reglas predichas por los mismos fuesen 100% acertadas, sin embargo, esto sucedió debido a la poca cantidad de respuesta de los municipios, por ello es importante ser reiterativos en fortalecer la divulgación de las encuestas con el fin de obtener muestras capaces de ser leídas de forma efectiva por los modelos predictivos para poder generar información útil a partir de este tipo de algoritmos y herramientas de análisis de datos.

Pese a que los modelos predictivos de aprendizaje supervisado no fueron los más útiles debido a la circunstancias de la base de datos, se estableció otra ruta con los datos obtenidos de la encuesta de caracterización de espacios e instancias de participación, esta fue la de generar tipologías de municipios por clustering, para ello se utilizó el modelo de KModes el cual logró agrupar los 92 municipios que dieron respuesta a la encuesta según sus espacios e instancias de participación presentes o ausentes, esto se hizo con el fin además de contrastar los grupos creados con las respuestas formativas, sin embargo, los resultados no fueron concluyentes puesto que cada grupo seguía una distribución parecida siendo que la temática formativa #1 era la más común en las respuestas de cada cluster. A partir de ello y de los patrones detectados por los modelos supervisados, se puede intuir que si bien existen diferentes tipos de municipios relacionados por sus instancias de participación estos terminan por adaptarse a las ofertas formativas de la Secretaría de Participación y no estas a las necesidades específicas de sus municipios.

Sin embargo, la aplicación de estas agrupaciones permite fortalecer el ejercicio de caracterización, logrando establecer a partir de la presencia o ausencia de espacios o instancias de participación diferentes temáticas formativas alternas, las cuales pueden ser más cercanas y acordes a las ofertadas por la Secretaría de Participación en un primer momento. Ahora bien para terminar, es necesario reevaluar la pertinencia entre la aplicación de la caracterización de espacios e instancias de participación y la elección de temáticas formativas en un mismo instrumento de recolección de la información, esto debido a que no se evidencia una relación fuerte o relevante entre la elección de temáticas formativas y la presencia o ausencia de espacios lo que puede ser indicativo además de que los enlaces municipales encargados de diligenciar el instrumento

terminan por adaptarse a las temáticas ofrecidas por la gobernación, las cuales puede que no sean las más adecuadas ni efectivas en sus municipios.

2.10. Recomendaciones

2.10.1. Proceso de Caracterización de Espacios e Instancias de Participación:

Respecto a las recomendaciones desde la experiencia como practicante dentro del Observatorio de Participación Ciudadana de Antioquia, es necesaria una revisión completa del proceso de caracterización de espacios e instancias de participación y el diseño de una metodología clara de manejo de la información que permita una efectiva cualificación de las inferencias para los tomadores de decisión, si bien en este trabajo se intenta establecer una ruta desde la preparación de los datos, el análisis exploratorio y la aplicación de modelos de machine learning, se deben evaluar a partir de las capacidades técnicas del Observatorio de Participación Ciudadana que otras fases deben integrarse en los procesos llevados a cabo por el mismo.

Siendo así, en pro del fortalecimiento de los procesos llevados a cabo por el OPC deben replantearse los softwares de almacenamiento de datos y las formas de las bases de datos, si bien dentro de este trabajo se sugirió utilizar modelos estructurados, es necesario un sistema de almacenamiento de la información interno diferente a Google Drive, ello permitirá asegurar la información en términos de privacidad, seguridad y eficiencia entendiendo que algunos de los datos proporcionados por los municipios son sensibles y de acceso privado.

Sobre la gestión de los datos, si bien en la ruta planteada por este trabajo se sugiere utilizar bases de datos estructuradas, es necesario y de vital importancia el uso de gestores de bases de datos relacionales externos a Microsoft Excel, esto debido a la necesidad de administrar, almacenar y consultar datos organizados generalmente en tablas que contienen relaciones. Siendo así, podrían recomendarse sistemas robustos como MySQL o Oracle Database, por la cantidad de datos y la magnitud de los proyectos de la Secretaria de Participación Ciudadana se sugiere el uso de Microsoft Access, este paso contribuiría de forma clave a la tecnificación y la cualificación de los procesos y las rutas de información de la Secretaria de Participación. Es de destacar que estos sistemas son compatibles y facilitan la ruta de datos plantada durante todo el documento, desde la creación de los metadatos hasta la estandarización de la información.

Respecto al análisis de los datos, se recomienda la implementación de procesos como el análisis exploratorio de datos, debido a su capacidad de mostrar información útil y de calidad en forma gráfica y porcentual como se demostró en este trabajo, además, dichos métodos deben ser acompañados por preguntas clave diseñadas por el OPC y la misma Secretaría de Participación y Cultura Ciudadana que guíen la aplicación de los métodos de analítica de datos, todo ello con el fin de que se logre cualificar cualquier intervención o inferencia que lleguen a tener los tomadores de decisión de estas entidades tan importantes del departamento.

Por otro lado, se recomienda fortalecimiento en la articulación constante y efectiva tanto del OPC como de la Secretaría de Participación con los enlaces municipales designados por las alcaldías, esto debido a que ellos son los encargados de brindar la información de primera mano de los municipios, por esto, se deben emprender acciones desde la misma Secretaría para divulgar, acompañar y brindar apoyo respecto a los procesos emprendidos por el Observatorio de Participación Ciudadana, además hacer conocer aún más el Observatorio de Participación Ciudadana y su importancia dentro de las diferentes subregiones del departamento es esencial para que los ejercicios llevados a cabo por el mismo tengan verdadera incidencia dentro de las decisiones departamentales.

Además, respecto a la publicación de los resultados es necesaria una campaña constante y fuerte de divulgación, esto debido a que el Observatorio de Participación Ciudadana como herramienta de política pública contiene la información resultante de las mediciones y caracterizaciones, la cual es de interés público y debería ser promocionada como tal de forma que la mayoría de los ciudadanos y los entes municipales se enteren de la información contenida dentro del Observatorio de Participación Ciudadana y puedan sacar inferencias lo más cualificadas posibles, en especial referentes a los espacios e instancias de participación ciudadana, de los cuales pueden y deben hacer parte.

Retomando el tema de los enlaces municipales, se recomienda sensibilizarlos y acercarlos a la Secretaría de Participación, con el fin de generar un acompañamiento efectivo a las iniciativas emprendidas por el OPC, esto debido a que es de vital importancia la recolección de los datos de la mayor cantidad de municipios posibles, con la caracterización se evidencia que los entes municipales no cuentan con cercanía a las iniciativas de este ente, por lo cual presentan problemas a la hora de responder a los instrumentos del mismo.

2.10.2. Integración de Herramientas y Métodos de Analítica de Datos y Machine Learning

Una vez visibilizadas las utilidades y potencialidades de las técnicas de analítica de datos, especialmente el análisis exploratorio de datos y el machine learning con modelos de aprendizaje supervisado para detección de patrones y no supervisado para clustering de municipios dentro de la ruta planteada por el documento, se sugiere que estos métodos sean implementados de forma efectiva en las demás iniciativas emprendidas por el Observatorio de Participación Ciudadana debido a su capacidad predictiva y de agrupación, la cual con la cantidad de datos y los parámetros adecuados puede llevar a conclusiones e inferencias acertadas y muy cercanas a la realidad, siendo que como herramienta de política pública el OPC ha de ser un elemento de consulta para los funcionarios públicos y los tomadores de decisión, la implementación de procesos analíticos de los datos y la información logran cualificar sus inferencias y hacer que la intervención pública sea lo más efectiva posible. Además, con la implementación de estas herramientas los análisis y la información serán de gran utilidad para la ciudadanía en general, permitiéndoles consultar una realidad acertada en términos de participación de sus diferentes municipios y subregiones.

Referencias

- ¿Qué es MySQL? (s. f.). Oracle Colombia. <https://www.oracle.com/co/mysql>
- Alcaldía de Jardín. (2024). Cabildo Adulto Mayor Municipal. <https://www.eljardin-antioquia.gov.co/instancias-de-participacion/cabildo-adultomayor-municipal>
- Alcaldía El Peñón. (2020). ¿Qué es el Consejo Municipal de Política Social COMPOS? <http://www.elpenon-cundinamarca.gov.co/preguntas-y-respuestas/que-es-el-consejo-municipal-de-politica-social-compos>.
- Álvarez, C. D. (2025). *BD_Departamento* [Archivo de Microsoft Excel]. Microsoft Excel. https://docs.google.com/spreadsheets/d/1PHFztf_1V4C11P_60Lw_cJ6mFjSeq0_z/edit?usp=sharing&ouid=106163912504804151262&rtpof=true&sd=true
- Álvarez, C. D. (2025). *Caracterización – Departamento* [Archivo de Microsoft Excel]. Microsoft Excel. https://docs.google.com/spreadsheets/d/1qgxHhwzEcEi_tAjlFXTXbmrTRxnrENoF/edit?gid=1361763300#gid=1361763300
- Álvarez, C. D. (2025). *MetaData_EI* [Archivo de Microsoft Excel]. Microsoft Excel. <https://docs.google.com/spreadsheets/d/1Vof-QMtJ8QFVJJcRZiFUmjVFc84SzHMC/edit?usp=sharing&ouid=106163912504804151262&rtpof=true&sd=true>
- Álvarez, C. D. (2025). *Prioridades Capacitación Municipios* [Archivo de Microsoft Excel]. Microsoft Excel. <https://docs.google.com/spreadsheets/d/110npvKjKH3F7q6Abbk9KMX16ZE7OJ7v5/edit?usp=sharing&ouid=106163912504804151262&rtpof=true&sd=true>
- Álvarez, C. D. (2025). *Trabajo de Grado Caracterización de Espacios e Instancias de Participación* [Cuaderno de Google Colab]. Google Colab., https://colab.research.google.com/drive/1ZirTi05EmkMi3uYiGDzOXtMK_1_2NZqv?usp=sharing
- Arana, C. (2021). *Modelos de Aprendizaje Automático Mediante Árboles de Decisión. CEMA Working Papers: Serie Documentos de Trabajo*, 778. Universidad del CEMA.
- Arellano, F. F. M. y Amaya, M. A. (2017). El papel de los metadatos en la Web Semántica. *Biblioteca universitaria*, 20(1).
- Asamblea Departamental de Antioquia. (2023). *Ordenanza No. 025 de 2023. Por la cual se adopta la política pública departamental de participación ciudadana y democrática, y se crea el sistema departamental de participación ciudadana*. Asamblea Departamental de Antioquia.
- Bracho, T. (2010). Políticas basadas en evidencia: la política pública como acción informada y objeto de investigación. En M. Merino, y G. Cejudo, (comps.), *Problemas, decisiones, soluciones. Enfoques de política pública* (pp. 291-319). Fondo de Cultura Económica.

- Brady, H. E., & Collier, D. (2004). *Rethinking social inquiry: Diverse Tools, Shared Standards*. Rowman & Littlefield Pub Incorporated.
- Bravo, A. A. (2016). *Tecnologías de la web semántica* [Tesis doctoral]. Universidad Complutense de Madrid, Facultad de Ciencias de la Información, España.
- Broache, M. P. (2024). Yu, A. The politics of descriptive inference: contested concepts in conflict data. *International Politics*, (2024). <https://doi.org/10.1057/s41311-024-00591-8>
- Chaverri Chaves, P., & Arguedas Ramírez, A. (2020). Políticas Públicas Basadas en Evidencia: una revisión del concepto y sus características. *Revista ABRA*, 40(60), 40-67. <https://dx.doi.org/10.15359/abra.40-60.2>
- Congreso de la República de Colombia. (2015, 6 de julio). *Ley estatutaria 1757 de 2015, por la cual se dictan disposiciones en materia de promoción y protección del derecho a la participación democrática*. Diario Oficial No. 49.565. Departamento Administrativo de la Función Pública. <https://www.funcionpublica.gov.co/eva/norma.php?i=75594>
- Corantioquia. (s. f.). *Pueblos indígenas localizados en la jurisdicción de Corantioquia: Embera Eyabida*. corantioquia.gov.co. <https://www.corantioquia.gov.co/comunidades-indigenas>
- Cortez Salinas, J., & Solorio Sandoval, I. (2022). Rastreo de procesos e inferencia causal en los métodos cualitativos de la ciencia política. *Estudios Políticos*, (55). <https://doi.org/10.22201/fcpys.24484903e.2022.55.81770>
- Cuba Rodríguez, Y., & Olivera Batista, D. (2018). Los metadatos, la búsqueda y recuperación de información desde las Ciencias de la Información. *E-Ciencias De La Información*, 8(2). <https://doi.org/10.15517/eci.v8i2.30085>
- Del Pino Matute, E. (2024, 27 agosto). *Es necesario que existan intermediarios entre los científicos y los políticos*. Almendron. <https://www.almendron.com/tribuna/es-necesario-que-existan-intermediarios-entre-los-cientificos-y-los-politicos>),
- Esri. (s. f.). *ArcGIS Survey123 | Cree encuestas y formularios inteligentes para la recopilación de datos*. <https://www.esri.co/es-co/productos/arcgis-survey123/inicio>.
- Esri. (s. f.-b). *Capas de entidades—Ayuda de ArcGIS Online | Documentación*. <https://doc.arcgis.com/es/arcgis-online/reference/feature-layers.htm>
- Flores Crespo, P. (2013). El enfoque de la política basado en la evidencia: Análisis de su utilidad para la educación de México. *Revista mexicana de investigación educativa*, 18(56), 265-290.
- Flores-Crespo, P. (2013). El enfoque de la política basada en evidencia: análisis de su utilidad para la educación de México. *Revista Mexicana de Investigación Educativa*, 18(56), 265-290. <http://www.scielo.org.mx/pdf/rmie/v18n56/v18n56a12.pdf>
- Función Pública. (s. f.). *Política de Rendición de cuentas - Función Pública*. <https://www1.funcionpublica.gov.co/politica-de-rendicion-de-cuentas>
- Gamboa Rocabado, F. (2013). Metodología para el análisis político: Un enfoque a partir de problemas, mecanismos e inferencias causales. *Temas Sociales*, (33), 245-287.

- http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S0040-29152013000100011&lng=es&tlng=es.
- Gobernación de Antioquia. (2024). *Plan de Desarrollo Departamental 2024–2027: Por Antioquia firme*. Medellín, Colombia.
- Gobernación de Antioquia. (s. f.). *Secretaría de Participación y Cultura Ciudadana*. antioquia.gov.co. <https://antioquia.gov.co/gobernacion/secretarias/secretaria-de-participacion-y-cultura-ciudadana>
- Gobernación de Antioquia. (s.f). *Participa*. <https://antioquia.gov.co/planeacion-yo-presupuesto-participativo>.
- Grande Atienza, E. (2021). *Análisis exploratorio de datos. Una variable* [Tesis de pregrado]. Universidad de Valladolid. <https://uvadoc.uva.es/handle/10324/52926>.
- Horber, E., & Ladiray, D. (1995) *Análisis Exploratorio de los datos* [Ponencia], II Seminario de Capacitación de docentes en “Producción y tratamiento de datos de Investigación en ciencias humanas”, Caracas, Venezuela.
- IBM InfoSphere Information Server. (s. f.). <https://www.ibm.com/docs/es/iis/11.5.0?topic=jobs-standardizing-data>
- Jenkins, K. (2023). Synthetic Data and Public Policy: Supporting Real-World Policymakers with Algorithmically Generated Data. *Policy Quarterly*, 19(2), 29–32. <https://ojs.victoria.ac.nz/pq/article/view/8234>
- King, G. (2014). Big data is not about the data. *PS: Political Science & Politics*, 47(1). <https://doi.org/10.1017/S1049096513001789>
- King, G., Keohane, R., & Verba, S. (2000). *El diseño de la investigación social. La inferencia en los estudios cualitativos*. Alianza Editorial.
- Latinno. (2017). *Asambleas Constituyentes Territoriales*. Berlin Social Science Center. <https://latinno.net/es/case/5119/>
- Latinno. (2017). *Consejos Municipales de Desarrollo Rural*. Berlin Social Science Center. <https://www.latinno.net/es/case/5077/#:~:text=Los%20Consejos%20Municipales%20de%20Desarrollo,desarrollo%20de%20los%20territorios%20rurales>.
- León, J., Solari, M., Olmos, R., & Escudero, I. (2011). La generación de inferencias dentro de un contexto social. Un análisis de la comprensión lectora a través de protocolos verbales y una tarea de resumen oral. *Revista de Investigación Educativa*, 29 (1), 13-42.
- Marrero, L., Olsowy, V., Tesone, F., Thomas, P. J., Corbalán, L. C., Francisco, F. S. J., & Pesado, P. M. (2023). *Un estudio de procesos de diseño de bases de datos NoSQL* [Conferencia]. XXVIII Congreso Argentino de Ciencias de la Computación (CACIC) <https://sedici.unlp.edu.ar/handle/10915/149452>
- Ministerio del Interior. (s.f). *Comité Interinstitucional para la Lucha Contra la Trata de Personas..* <https://www.mininterior.gov.co/comite-interinstitucional-para-la-lucha-contra-la-trata-de-personas/>

- Municipio de Barrancas, La Guajira. (s. f.). *¿Qué es el Consejo Municipal de Política Social?* <https://www.barrancas-laguajira.gov.co/informacion-especifica-para-grupos-de-interes-119537/que-es-el-consejo-municipal-de-politica-social>
- Observatorio de Participación Ciudadana de Antioquia [OPC]. (2024). *Espacios e instancias de participación ciudadana municipales en el departamento de Antioquia*. Observatorio de Participación Ciudadana - Secretaría de Participación y Cultura Ciudadana de Antioquia.
- Pérez Lara, J. E., Covarrubias Moreno, O. M., & Tolentino Sanjuan, A. V. (2025). Analítica de datos y gobierno abierto: hacia una gestión pública basada en evidencia. *RIESED - Revista Internacional De Estudios Sobre Sistemas Educativos*, 3(16), 762-780. <https://doi.org/10.5281/zenodo.15697620>
- Personería de Bogotá, D. C. (2023). *Elección de Mesas Locales y Mesa Distrital de Participación Efectiva de Víctimas del Conflicto Armado Interno en Bogotá, D. C.* [https://www.personeriabogota.gov.co/2023-eleccion-de-mesas-locales-y-mesa-distrital-de-participacion-efectiva-de-victimas-del-conflicto-armado-interno-en-bogota-d-c#:~:text=Son%20los%20espacios%20de%20trabajo,de%20Víctimas%20\(OV\)%2C%20Organizaciones](https://www.personeriabogota.gov.co/2023-eleccion-de-mesas-locales-y-mesa-distrital-de-participacion-efectiva-de-victimas-del-conflicto-armado-interno-en-bogota-d-c#:~:text=Son%20los%20espacios%20de%20trabajo,de%20Víctimas%20(OV)%2C%20Organizaciones)
- Piñeiro Gómez, J. M. (s. f.). *Bases de datos relacionales y modelado de datos* (1.a ed.). Ediciones Paraninfo, S.A.
- Rivero Suguiura, F. O. (2022). Árbol de Decisión en Aprendizaje Automático. *REVISTA VARIANZA*, 19(19), 39-46. <https://ojs.umsa.bo/ojs/index.php/revistavarianza/article/view/433>
- Sánchez Ramos, M. (2005). Uso metodológico de las tablas de contingencia en la ciencia política. *Espacios Públicos*, 8(16). <http://www.redalyc.org/articulo.oa?id=67681605>
- Sutcliffe, S., & Court, J. (2006) *Herramientas para diseñadores de políticas públicas en países en desarrollo*. <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3520.pdf>.
- Universitat Obertade Catalunya. (s. f.). *Espacio de recursos de ciencia de datos*. <https://datascience.recursos.uoc.edu/es/preprocesamiento-de-datos-con-sklearn/#:~:text=La%20librer%C3%ADa%20scikit-learn%2C%20tambi%C3%A9n,c%C3%B3digo%20que%20use%20estas%20librer%C3%ADas>.
- Zoiner, T. (s. f.-b). *Datos no relacionales y NoSQL - Azure Architecture Center*. Microsoft Learn. <https://learn.microsoft.com/es-es/azure/architecture/data-guide/big-data/non-relational-data>

Anexos

Anexo A

Cuaderno de Google Colab: “Trabajo de Grado Caracterización de Espacios e Instancias de Participación”

En este cuaderno de Google Colab se desarrolló todo el código por el cual se aplicó el análisis exploratorio de datos y el machine learning para la cualificación de las inferencias del Observatorio de Participación Ciudadana.

El material está disponible en el siguiente enlace:

https://colab.research.google.com/drive/1ZirTi05EmkMi3uYiGDzOXtMK_1_2NZqv?usp=sharing

[g](#)

Anexo B

Archivo de Microsoft Excel: “Caracterización – Departamento”

En este documento de Microsoft Excel se encuentra alojada la información reportada por los municipios del departamento respecto a la presencia o ausencia de espacios o instancias de participación ciudadana.

El material está disponible en el siguiente enlace:

https://docs.google.com/spreadsheets/d/1qgxHhwzEcEi_tAjIFxTXbmrTRxnrENoF/edit?gid=1361763300#gid=1361763300

Anexo C

Archivo de Microsoft Excel: “Prioridades Capacitación Municipios”

En este archivo de Microsoft Excel se encuentra alojada la información reportada por los municipios del departamento respecto las necesidades formativas ofertadas por la Secretaría de Participación en el año 2024.

El material está disponible en el siguiente enlace:

<https://docs.google.com/spreadsheets/d/110npvKjKH3F7q6Abbk9KMX16ZE7OJ7v5/edit?usp=sharing&ouid=106163912504804151262&rtpof=true&sd=true>

Anexo D

Archivo de Microsoft Excel: “MetaData_EI”

En este archivo de Microsoft Excel se encuentra la MetaData diseñada para la caracterización de espacios e instancias de participación del Observatorio de Participación Ciudadana con su respectiva información.

El material está disponible en el siguiente enlace:

<https://docs.google.com/spreadsheets/d/1Vof->

[QMtJ8QFVJJcRZiFUmjVFc84SzHMC/edit?usp=sharing&ouid=106163912504804151262&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1Vof-QMtJ8QFVJJcRZiFUmjVFc84SzHMC/edit?usp=sharing&ouid=106163912504804151262&rtpof=true&sd=true)

Anexo E

Archivo de Microsoft Excel: “BD_Departameto”

En este archivo de Microsoft Excel se encuentra la base de datos ajustada y utilizada en Google Colab para la cualificación de las inferencias del ejercicio de caracterización de espacios e instancias de participación ciudadana.

El material está disponible en el siguiente enlace:

https://docs.google.com/spreadsheets/d/1PHFztf_1V4C11P_60Lw_cJ6mFjSeq0_z/edit?usp=sharing&ouid=106163912504804151262&rtpof=true&sd=true