



Desarrollo de un modelo de aprendizaje no supervisado para identificar a los empleados con mayor riesgo de rotación en Bancolombia

Diego Alejandro Saavedra Valdivieso

Proyecto de práctica para otorgar título de Ingeniero Industrial

Modalidad de Práctica: Semestre de Industria

Asesor

Olga Cecilia Usuga Manco, PhD en Ciencias - Estadística

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería Industrial

Medellín

2024

Cita	(Saavedra Valdivieso, 2024)
Referencia	Saavedra Valdivieso, D. (2024). <i>Desarrollo de un modelo de aprendizaje no supervisado para identificar a los empleados con mayor riesgo de rotación en Bancolombia</i> [Informe de práctica].
Estilo APA 7 (2020)	Universidad de Antioquia, Medellín.



Quiero expresar mi agradecimiento al Departamento de Ingeniería Industrial por el escenario de prácticas brindado, así como a los asesores internos por su constante apoyo.

De manera especial, agradezco a Bancolombia, en particular al área de Inteligencia y Estrategia del Talento y a la Vicepresidencia de Talento y Cultura, por la oportunidad de realizar mi práctica profesional en un entorno tan enriquecedor y colaborativo. Su confianza y respaldo fueron fundamentales para el desarrollo de este proyecto.



Centro de Documentación Ingeniería (CENDOI)

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

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano/Director: Julio César Saldarriaga Molina.

Jefe departamento: Mario Alberto Gaviria Giraldo.

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.

Dedicatoria

A mi madre, por su amor incondicional, su apoyo constante y su fe inquebrantable en mí. Su fortaleza y sabiduría han sido una fuente de inspiración y motivación en cada paso de este camino.

A Dios, por guiarme y darme la fuerza necesaria para superar cada desafío. Su presencia y bendiciones han sido fundamentales para alcanzar este logro.

Agradecimientos

Quiero expresar mi más profundo agradecimiento a mi patria, por brindarme las oportunidades y el entorno necesarios para desarrollar mis habilidades y conocimientos.

A los profesores de mi carrera, quienes con su dedicación y sabiduría han sido pilares fundamentales en mi formación académica y profesional. Su guía y enseñanzas han dejado una huella imborrable en mi camino.

A mis amistades, por su apoyo incondicional, comprensión y aliento constante. Su compañía ha sido vital en este viaje.

Finalmente, a Bancolombia, por abrirme sus puertas en el escenario de prácticas. Agradezco la oportunidad de aprender y crecer profesionalmente en un entorno tan enriquecedor y colaborativo. Su confianza y apoyo han sido cruciales para el desarrollo de este proyecto.

Tabla de contenido

Resumen	9
Abstract	10
Introducción	11
1 Objetivos	12
1.1 Objetivo general	12
1.2 Objetivos específicos	12
2 Marco teórico	13
3 Metodología	15
Enfoque	15
Técnicas e Instrumentos	15
1. Recolección de Datos:	15
2. Análisis Exploratorio de Datos (EDA):	15
3. Desarrollo del modelo Predictivo:	15
4. Técnicas de Clustering:	15
5. Diseño de Estrategias de Retención:	16
6. Presentación Ejecutiva:	16
Fases	16
1. Recolección y Limpieza de Datos	16
2. Análisis Exploratorio de Datos	16
3. Desarrollo del modelo Predictivo	16
4. Técnicas de Clustering	16
5. Diseño de Estrategias de Retención	16
6. Presentación Ejecutiva	16
Cronograma de Actividades	16

Mes 1: Preparación y Recolección de Datos	17
Semana 1-2: Planificación Inicial	17
Semana 3-4: Identificación de Fuentes de Datos y Permisos	17
Mes 2: Análisis Exploratorio de Datos (EDA) y Limpieza	17
Semana 1-2: Recopilación y Limpieza de Datos	17
Semana 3-4: Análisis Exploratorio de Datos (EDA)	17
Mes 3: Desarrollo del modelo Predictivo y Técnicas de Clustering	17
Semana 1-2: Desarrollo del modelo Predictivo	17
Semana 3-4: Aplicación de Técnicas de Clustering	17
Mes 4: Diseño de Estrategias de Retención y Presentación Ejecutiva	17
Semana 1-2: Diseño de Estrategias de Retención	17
Semana 3-4: Preparación de Presentación Ejecutiva	17
Mes 5: Revisión Final y Entrega del Proyecto	18
Semana 1-2: Revisión Final y Ajustes	18
Semana 3-4: Entrega del Proyecto y Presentación Final	18
4 Resultados	18
Extracción, transformación y carga	18
Análisis de Motivos de Retiro Excluidos	22
Motivos Excluidos	22
Motivos Considerados	22
Enfoque en Investigación Profunda	22
Preparación de los datos	23
Análisis exploratorio de los datos	23
Análisis temporal de ingresos vs retiros	25
Modelo supervisado	27

Metodología para la Selección de Variables con XGBoost	27
Modelo no supervisado K-Means	29
Preparación de los datos	29
Resultados de la Función para la Selección de Clústers	30
Análisis Visual y Selección Final de Clústers	31
5 Análisis	33
Clúster 0: "Estabilidad y Experiencia" (Bajo Abandono)	33
Clúster 1: "Operativos Jóvenes y No Reconocidos" (Medio Abandono)	34
Clúster 2: "Profesionales Solteros y Con Baja Antigüedad" (Alto Abandono)	35
Clúster 3: "Operativos Solteros y No Reconocidos" (Alto Abandono)	36
Clúster 4: "Operativos con Alta Antigüedad y Moderado Reconocimiento" (Medio Abandono)	37
Estrategias específicas por clúster	38
Clúster 0: "Veteranos Estables y Moderadamente Reconocidos" (Bajo Abandono)	38
Clúster 1: "Jóvenes Operativos No Reconocidos" (Alto Abandono)	39
Clúster 2: "Profesionales Solteros con Alta Percepción de Desarrollo" (Medio Abandono)	39
Clúster 3: "Operativos Solteros y No Reconocidos" (Alto Abandono)	39
Clúster 4: "Operativos con Alta Antigüedad y Moderado Reconocimiento" (Medio Abandono)	40
Conclusiones del análisis	40
Conclusiones	42
Recomendaciones	43
Referencias	44

Lista de tablas

Tabla 1: Descripción de las columnas que fueron tomadas en cuenta	19
Tabla 2: Promedios 10 iteraciones por 10 clústers para encontrar el clúster óptimo	30

Lista de figuras

Figura 1: Boxplot abandona vs edad	24
Figura 2: Boxplot de abandona vs desempeño global.....	24
Figura 3: Boxplot de abandona vs días pendientes vacaciones.....	24
Figura 4: Boxplot de abandona vs cursos terminados acumulados.....	25
Figura 5: Ingresos vs Retiros desde 2022 hasta abril 2024	25
Figura 6: PCA con etiquetas de KMeans 6 Clusters	31
Figura 7: PCA con etiquetas de KMeans 4 Clusters	31
Figura 8: PCA con etiquetas de KMeans 5 Clusters	32

Resumen

Este proyecto de grado se enfoca en el análisis de los clústers de empleados en Bancolombia para identificar características y necesidades específicas con el fin de desarrollar estrategias que reduzcan la probabilidad de abandono y mejoren la satisfacción laboral. La metodología empleada incluye la creación de consultas para grandes volúmenes de datos, la extracción, limpieza y procesamiento de estos, y el uso de técnicas de modelado de datos y segmentación de empleados. Los objetivos principales son identificar los factores críticos que influyen en la retención de empleados y proponer intervenciones personalizadas para cada clúster.

Los resultados del análisis revelan que la falta de reconocimiento, el desarrollo profesional, la flexibilidad laboral y los beneficios personalizados son factores clave que afectan la satisfacción y la retención de los empleados. En particular, se identificaron cinco clústers: "Veteranos Estables y Moderadamente Reconocidos" (Bajo Abandono), "Jóvenes Operativos No Reconocidos" (Alto Abandono), "Profesionales Solteros con Alta Percepción de Desarrollo" (Medio Abandono), "Operativos Solteros y No Reconocidos" (Alto Abandono) y "Operativos con Alta Antigüedad y Moderado Reconocimiento" (Medio Abandono). Cada clúster requiere estrategias específicas para abordar sus necesidades particulares.

Las conclusiones destacan la importancia de implementar programas de reconocimiento, ofrecer oportunidades de desarrollo profesional, considerar la flexibilidad laboral y adaptar beneficios y compensaciones a las necesidades específicas de cada clúster. Estas estrategias personalizadas pueden mejorar significativamente la retención y satisfacción de los empleados en Bancolombia, aprovechando los puntos fuertes y abordando las áreas de mejora de cada clúster.

Palabras clave: clústers de empleados, satisfacción laboral, retención de empleados, reconocimiento, desarrollo profesional, flexibilidad laboral, beneficios personalizados, análisis de datos, Bancolombia.

Abstract

This thesis project focuses on analyzing employee clusters at Bancolombia to identify specific characteristics and needs to develop strategies that reduce the likelihood of attrition and improve job satisfaction. The methodology employed includes creating queries for large volumes of data, extracting, cleaning, and processing the data, and using data modeling and employee segmentation techniques. The main objectives are to identify critical factors influencing employee retention and propose tailored interventions for each cluster.

The analysis results reveal that lack of recognition, professional development, work flexibility, and personalized benefits are key factors affecting employee satisfaction and retention. Five clusters were identified: "Stable and Moderately Recognized Veterans" (Low Attrition), "Unrecognized Young Operatives" (High Attrition), "Single Professionals with High Development Perception" (Medium Attrition), "Unrecognized Single Operatives" (High Attrition), and "Operatives with High Tenure and Moderate Recognition" (Medium Attrition). Each cluster requires specific strategies to address its unique needs.

The conclusions emphasize the importance of implementing recognition programs, offering professional development opportunities, considering work flexibility, and tailoring benefits and compensations to the specific needs of each cluster. These personalized strategies can significantly improve employee retention and satisfaction at Bancolombia, leveraging the strengths and addressing the improvement areas of each cluster.

Keywords: employee clusters, job satisfaction, employee retention, recognition, professional development, work flexibility, personalized benefits, data analysis, bancolombia

Introducción

En el dinámico escenario laboral actual, la retención de talento se ha instituido como una tarea destacada para las organizaciones en busca de mantener su competitividad y eficacia operativa. La pérdida de colaboradores altamente capacitados no solo implica costos financieros significativos asociados al proceso de reclutamiento y formación, sino que también conlleva un impacto negativo en la productividad y la estabilidad operativa. En este contexto, surge la propuesta de implementar un modelo de inteligencia artificial focalizado en la predicción y segmentación de la rotación de empleados en Bancolombia.

Desde la Vicepresidencia de Talento y Cultura de Bancolombia, se reconoce la importancia estratégica de optimizar la retención de talento como un componente clave para el éxito continuo de la organización. Con la constante evolución del entorno laboral y la competencia por profesionales cualificados, la necesidad de implementar enfoques innovadores se hace cada vez más evidente. La propuesta se fundamenta en la premisa de que la retención efectiva no solo se trata de reconocer la rotación de empleados, sino de preverla y abordarla proactivamente.

El desafío principal que enfrenta Bancolombia es la identificación temprana de empleados con alto riesgo de rotación desde la Vicepresidencia de Talento y Cultura. La pérdida de estos profesionales no solo implica costos económicos, sino que también afecta la continuidad operativa y la capacidad para mantener un equipo sólido y experimentado. La propuesta surge como una respuesta anticipada para prevenir esta problemática, centrándose en la creación de un modelo de inteligencia artificial capaz de prever patrones de rotación y diseñar estrategias de retención personalizadas.

El objetivo central de este proyecto es implementar un modelo de inteligencia artificial que identifique proactivamente a empleados con riesgo de rotación en Bancolombia, con el fin de formular estrategias de retención personalizadas. Para alcanzar este objetivo, se plantean objetivos específicos, tales como realizar un análisis exhaustivo de datos de Bancolombia, desarrollar un modelo predictivo adaptado, aplicar técnicas de clustering para segmentar a los empleados según su riesgo de rotación y diseñar estrategias de retención personalizadas, considerando la cultura organizativa y necesidades particulares del banco.

La metodología adoptada para la ejecución de este proyecto será de enfoque mixto, combinando elementos cualitativos y cuantitativos. Se llevará a cabo una recolección y análisis de datos específicos de Bancolombia, seguido del desarrollo de un modelo predictivo y la aplicación de técnicas de clustering. La fase final incluirá el diseño de estrategias de retención personalizadas, presentadas de manera ejecutiva para su evaluación y posible implementación. Cabe destacar que, debido a consideraciones de tiempo y trámites, el proyecto no incluirá la fase de despliegue inmediato, enfocándose en proporcionar un modelo robusto y estrategias sustentadas para su futura implementación.

1 Objetivos

1.1 Objetivo general

Crear un modelo de inteligencia artificial que identifique a los empleados con mayor riesgo de rotación en Bancolombia desde la Vicepresidencia de Talento y Cultura, permitiendo la formulación de estrategias de retención personalizadas.

1.2 Objetivos específicos

- Analizar de forma exhaustiva datos demográficos, salariales, años en la empresa, respuestas a encuestas internas y movimientos de crecimiento de los empleados.
- Implementar un modelo predictivo ajustado a las particularidades de Bancolombia, utilizando variables clave para prever el riesgo de rotación de empleados.
- Segmentar a los empleados en grupos homogéneos en función de su riesgo de rotación y características relevantes utilizando técnicas de clustering adaptadas al contexto de Bancolombia. .
- Diseñar estrategias de retención personalizadas considerando la cultura organizativa y las necesidades particulares de Bancolombia, basadas en los resultados del modelo predictivo y de las técnicas de clustering.

- Comunicar de manera efectiva los resultados del modelo, las segmentaciones obtenidas y las estrategias de retención propuestas, con el objetivo de facilitar la toma de decisiones en la Vicepresidencia de Talento y Cultura en la Gerencia de Estrategia y Talento en Bancolombia.

2 Marco teórico

La retención de talento se ha destacado como un elemento crucial en la literatura especializada, ya que la lealtad y el compromiso organizacional a largo plazo son esenciales para mantener empleados altamente calificados y, por ende, la estabilidad y productividad de la organización (Allen & Meyer, 1990). Diversas investigaciones han demostrado la eficacia de los modelos predictivos en la gestión de recursos humanos. Becker, Huselid y Ulrich (2001) han encontrado que estos modelos son útiles para anticipar comportamientos, destacando aplicaciones prácticas en empresas como Google, donde se han implementado con éxito para prevenir la rotación de empleados (Bock, 2015). La experiencia de Google ilustra la relevancia y aplicabilidad de estas estrategias en entornos competitivos.

El uso de inteligencia artificial (IA) para gestionar recursos humanos es una práctica innovadora y efectiva. Empresas líderes como Google han adoptado modelos de IA para identificar patrones de comportamiento y reducir la rotación de empleados, lo que demuestra la eficacia de estos enfoques en contextos empresariales dinámicos (Bock, 2015). Una técnica valiosa para segmentar empleados es el clustering, la cual permite una comprensión profunda de los patrones de comportamiento y facilita la personalización de estrategias de retención (Kumar & Ravi, 2007). Al adaptar estas estrategias a las características específicas de cada grupo de empleados, se maximiza su efectividad y precisión.

El análisis exploratorio de datos (EDA) también es crucial en este proceso. Este método, empleado por IBM, permite una comprensión detallada de los datos específicos de la organización y revela patrones clave que influyen en la retención (Davenport, Harris, & Shapiro, 2010). La aplicación de EDA resalta la importancia de adaptar las estrategias a la situación particular de cada

empresa. Amazon ha demostrado el éxito de estrategias de retención personalizadas que se ajustan a la cultura organizativa y las necesidades específicas de la empresa (Stone & Ramsay, 2004). Esta evidencia apoya la idea de que las estrategias de retención deben estar alineadas con la identidad y valores únicos de la organización.

Además, estudios de Griffeth, Hom y Gaertner (2000) han analizado los factores que influyen en la rotación de empleados, incluyendo las condiciones laborales, oportunidades de desarrollo y el equilibrio entre trabajo y vida personal. Estos elementos son fundamentales para entender las decisiones de los empleados de permanecer o dejar una organización. La investigación de Zhang, Zhao y Qi (2018) sobre el uso de algoritmos de aprendizaje automático ha demostrado mejoras significativas en la precisión de la predicción de la rotación de empleados. Estas innovaciones tecnológicas permiten a las organizaciones tomar decisiones informadas y proactivas para retener su talento clave.

Por último, el modelo de decisión de permanencia o rotación propuesto por March y Simon (1958) sigue siendo relevante. Este modelo se basa en la satisfacción laboral y las oportunidades percibidas fuera de la organización, proporcionando un marco teórico clásico que complementa las aplicaciones modernas de IA y análisis de datos.

3 Metodología

La metodología para el desarrollo de este proyecto se fundamentará en un enfoque mixto, combinando elementos cualitativos y cuantitativos para obtener una comprensión holística de la problemática de rotación de empleados en Bancolombia.

Enfoque

Se empleará un enfoque mixto para aprovechar las fortalezas de ambas metodologías. La investigación cuantitativa permitirá el análisis de datos a gran escala, mientras que la investigación cualitativa aportará perspectivas enriquecedoras para comprender los contextos específicos de la organización.

Técnicas e Instrumentos

1. *Recolección de Datos:*

- **Cualitativa:** Entrevistas o charlas con líderes de equipo y empleados clave para obtener percepciones subjetivas sobre factores de retención.
- **Cuantitativa:** Recopilación de datos demográficos, salariales, años en la empresa, respuestas a encuestas internas y movimientos de crecimiento de empleados.

2. *Análisis Exploratorio de Datos (EDA):*

- **Cuantitativa:** Empleo de herramientas estadísticas para explorar patrones, correlaciones y variables clave en los datos cuantitativos.

3. *Desarrollo del modelo Predictivo:*

- **Cuantitativa:** Implementación de algoritmos de aprendizaje automático para construir un modelo predictivo adaptado a las particularidades de Bancolombia.

4. *Técnicas de Clustering:*

- **Cuantitativa:** Aplicación de algoritmos de clustering para segmentar a los empleados en grupos homogéneos según su riesgo de rotación y características relevantes.

5. *Diseño de Estrategias de Retención:*

- **Cualitativa y Cuantitativa:** Integración de resultados cuantitativos con análisis cualitativos para diseñar estrategias de retención personalizadas, considerando la cultura organizativa y necesidades específicas.

6. *Presentación Ejecutiva:*

- **Cualitativa:** Preparación de una presentación ejecutiva basada en los resultados cuantitativos y cualitativos para comunicar efectivamente los hallazgos y estrategias propuestas.

Fases

1. ***Recolección y Limpieza de Datos:*** Recopilación de datos demográficos, salariales, años en la empresa, respuestas a encuestas internas y movimientos de crecimiento de empleados.
2. ***Análisis Exploratorio de Datos:*** Aplicación de técnicas estadísticas para identificar patrones y correlaciones en los datos.
3. ***Desarrollo del modelo Predictivo:*** Implementación de algoritmos de aprendizaje automático adaptados a las características específicas de Bancolombia.
4. ***Técnicas de Clustering:*** Aplicación de algoritmos de clustering para segmentar a los empleados en grupos homogéneos.
5. ***Diseño de Estrategias de Retención:*** Integración de resultados cuantitativos y análisis cualitativos para diseñar estrategias personalizadas.
6. ***Presentación Ejecutiva:*** Preparación de una presentación ejecutiva basada en los resultados obtenidos, estrategias propuestas y recomendaciones para la toma de decisiones en la Vicepresidencia de Talento y Cultura en la Gerencia de Estrategia y Talento en Bancolombia.

Cronograma de Actividades

Se describen las diferentes actividades a realizar según la metodología, indicando el tiempo programado para su ejecución.

Duración del Proyecto: 5 meses

Mes 1: Preparación y Recolección de Datos

Semana 1-2: Planificación Inicial

- Definición detallada de objetivos y alcance.
- Revisión de la metodología propuesta.

Semana 3-4: Identificación de Fuentes de Datos y Permisos

- Identificación de fuentes de datos relevantes.
- Gestión de permisos y accesos necesarios.

Mes 2: Análisis Exploratorio de Datos (EDA) y Limpieza

Semana 1-2: Recopilación y Limpieza de Datos

- Recopilación de datos demográficos, salariales, encuestas internas y movimientos de crecimiento.
- Limpieza de datos para abordar valores atípicos y datos faltantes.

Semana 3-4: Análisis Exploratorio de Datos (EDA)

- Aplicación de técnicas estadísticas para explorar patrones y correlaciones.

Mes 3: Desarrollo del modelo Predictivo y Técnicas de Clustering

Semana 1-2: Desarrollo del modelo Predictivo

- Implementación de algoritmos de aprendizaje automático adaptados a Bancolombia.

Semana 3-4: Aplicación de Técnicas de Clustering

- Utilización de algoritmos de clustering para segmentar empleados según su riesgo de rotación.

Mes 4: Diseño de Estrategias de Retención y Presentación Ejecutiva

Semana 1-2: Diseño de Estrategias de Retención

- Integración de resultados cuantitativos y análisis cualitativos para diseñar estrategias personalizadas.

Semana 3-4: Preparación de Presentación Ejecutiva

- Desarrollo de una presentación ejecutiva detallando resultados y estrategias propuestas.

Mes 5: Revisión Final y Entrega del Proyecto

Semana 1-2: Revisión Final y Ajustes

- Revisión exhaustiva de resultados y estrategias.
- Ajustes finales según retroalimentación.

Semana 3-4: Entrega del Proyecto y Presentación Final

- Entrega del informe final y resultados.
- Presentación final ante la Gerencia de Estrategia y Talento en Bancolombia.

4 Resultados

Extracción, transformación y carga

A continuación, se detalla el proceso de clusterización realizado en Bancolombia. Se comenzó con la base principal, en este caso es la planta, obtenida mediante consultas de SQL. Esta base, correspondiente a la planta de la empresa, se compone de registros quincenales, de los cuales se seleccionó el registro máximo de cada mes para evitar duplicidades. Se creó una variable ID única concatenando el número de identificación de cada empleado con el año y mes de ingestión de la tabla, la cual se utilizó como llave para conectar con otras tablas. Se seleccionaron variables demográficas como: género, edad, generación, estado civil y antigüedad, así como variables del cargo como: antigüedad en el cargo, nivel del cargo, modalidad de trabajo y régimen laboral.

La sábana de datos de la implementación del modelo abarca desde 2022 hasta abril de 2024, excluyendo becarios, aprendices y empleados con contratos a término definido. Posteriormente, se aplicaron diversas transformaciones, como recalcular los rangos de edad y antigüedad, la creación de la variable de cantidad de cargos anteriores, la corrección de errores en la antigüedad del cargo, etc. Había una variable con datos nulos, se completaron los datos faltantes utilizando la moda de registros similares.

Luego se describió la tabla de desempeño, donde se consideró únicamente la valoración global del desempeño, medida anualmente por los jefes. La tabla de ausentismo registró únicamente los días de ausentismo por diversas causas laborales. Para las vacaciones, se consideraron únicamente las vacaciones pendientes como indicador de equilibrio trabajo-vida.

En cuanto a la tabla de nivel educativo, se construyó a partir de los niveles educativos desde primaria hasta doctorado. La movilidad se abordó considerando únicamente el medio de transporte utilizado por los empleados. Respecto a la tabla de familiares, se tomaron variables relacionadas con hijos menores y mayores de 18 años. La encuesta "Y tú, ¿qué dices?" se resumió en el promedio de todas las respuestas, mientras que en la tabla de reconocimientos se consideraron los días de reconocimiento y el uso de "Happy Friday". En cuanto a los beneficios, se seleccionó el saldo promedio del crédito de los empleados y se verificó si contaban con la modalidad de ahorro Mutuocolombia. Esta modalidad consiste, a grandes rasgos, en que el banco proporciona un 50% adicional sobre los ahorros del empleado. La tabla de vivienda se simplificó considerando la situación habitacional y el tipo de vivienda. La formación se cuantificó a partir del total de cursos terminados acumulados. Finalmente, la tabla de horas extras se centró en el promedio de horas extras por mes.

Se unieron las 13 tablas resultantes, eliminando duplicados y nulos, y rellenando valores faltantes utilizando distintas estrategias según el tipo de dato. La base final contiene 645,790 registros con 34 columnas. (Tabla 1)

Tabla 1: Descripción de las columnas que fueron tomadas en cuenta

Variable	Descripción	Categorías
Id	Id empleado + año de ingestión + mes de ingestión	No aplica (variable única)
N_personal	Identificación del empleado	No aplica (variable única)
Ingestion_date	Fecha de ingestión de los datos	No aplica (variable única)
Genero	Género del empleado	Masculino, femenino
Edad	Edad del empleado	No aplica (variable numérica)
Generación	Generación a la que pertenece el empleado	Baby boomers, generación x, millennials, centennials
Estado_civil	Estado civil del empleado	Cas., sol., u.lib., div., sep., viu.
Antigüedad	Antigüedad del empleado en la empresa	No aplica (variable numérica)
Fecha_ingreso	Fecha de ingreso del empleado	No aplica (variable única)

Variable	Descripción	Categorías
Nivel_cargo	Nivel del cargo del empleado	Operativo, profesional, jefe, estratégico medio, alta dirección
Modalidad_trabajo	Modalidad de trabajo del empleado	Presencial, flexiwork, suplementario, móvil, autónomo, móvil administrativo
Division_personal	División regional del personal	Antioquia, Bogotá y sabana, sur, centro, caribe
Area_personal	Área personal: si el empleado está en relación o colectividad sindical	Convencionados (co), no convencionados (co), en transición (co)
Unidad_organizativa	Unidad organizativa del empleado	No aplica (variable única)
Cargo	Cargo del empleado	No aplica (variable única)
Cantidad_cargos_anteriores	Cantidad de cargos anteriores del empleado	0, 1, 2 o más
Antiguedad_cargo	Antigüedad en el cargo actual	No aplica (variable numérica)
Desempeño_valoracion_global	Valoración global de desempeño: se realiza anualmente y escala de 0 a 4	No aplica (variable numérica)
Ausentismo_dias_naturales	Días naturales de ausentismo del empleado	No aplica (variable numérica)
Vacaciones_dias_pendientes	Días de vacaciones pendientes	No aplica (variable numérica)
Escolaridad_max_nivel_texto	Nivel máximo de escolaridad alcanzado (texto)	No aplica (variable única)
Escolaridad_max_nivel_numero	Nivel máximo de escolaridad alcanzado (número)	2.0, 3.0, 4.0, 5.0, 6.0, 7.0
Movilidad_medio_transporte	Medio de transporte utilizado	Automóvil, transporte público masivo, moto, caminata, plataformas de transporte, ninguno, bicicleta
Familiares_hijos_menores_18	Cantidad de hijos menores de 18 años	0, 1, 2, >=3
Familiares_hijos_mayores_18	Cantidad de hijos mayores de 18 años	0, 1, 2, >=3
Promedio_ytqd	Promedio de la encuesta de autopercepción "y tú qué dices": se realiza anualmente	No aplica (variable numérica)
Reconocimiento_toma_dias	Días de reconocimiento tomados	0, 0.5-1.5, 2-3, >=4
Reconocimiento_happy_friday	Participación en happy friday	No, si
Credito_saldo_promedio_millones	Saldo promedio del crédito en millones	0, 1-10m, 10-20m, 20-100m, más de 100m
Ahorros_es_mutuocolombia	Afiliación a Mutuocolombia	No, si
Vivienda_tipo	Tipo de vivienda	Propia, arrendada, familiar
Vivienda_composicion_familiar	Composición familiar de la vivienda	Vivo con pareja/familiares/amigos y compartimos gastos, vivo con

Variable	Descripción	Categorías
		hijos/familiares/amigos que dependen de mí, vivo solo(a)
Formacion_cursos_terminados_acum	Cursos de formación terminados acumulados	No aplica (variable numérica)
Horas_extras_promedio	Promedio de horas extras trabajadas	No aplica (variable numérica)
Motivo_retiro	Motivo del retiro del empleado	No aplica (variable única)
Abandona	Indica si el empleado abandonó la empresa	0 (no), 1 (sí)

Después de unir las 13 tablas resultantes, se procedió a eliminar duplicados y manejar los valores nulos presentes en la base de datos final. Se aplicaron distintas estrategias para rellenar estos valores faltantes, teniendo en cuenta la naturaleza de los datos y el contexto del análisis

- **Forward Fill y Backfill por Año (Relleno hacia delante y hacia atrás):** Esta técnica se utilizó para columnas como desempeño y promedio. Se optó por este enfoque debido a que estas variables, como la valoración del desempeño y los promedios, se evalúan una vez al año y tienden a mantenerse constantes durante ese período. Se rellenaron los valores nulos con el último valor conocido dentro del mismo año para cada empleado, utilizando tanto forwardfill como backfill. Esto garantiza la consistencia temporal y evita discontinuidades en los registros anuales.
- **Forward Fill (Relleno hacia delante) basado en la estabilidad:** Para variables como vacaciones, escolaridad, movilidad, familiares, ahorros_es_mutuocolombia, vivienda y formación, se realizó un análisis de la estabilidad de los datos. Se asumió que, una vez registrada la información de un empleado, es poco probable que estas características cambien significativamente hasta que se registre una nueva información. Por lo tanto, se decidió rellenar los valores nulos con forward fill, manteniendo la coherencia y la integridad de los datos.
- **Asignación de cero:** En el caso de variables como ausentismo, horas extras, vacaciones, crédito_saldo_promedio_millones, horas extras y formación, donde la ausencia de registros podría indicar la falta de eventos relevantes en esos períodos, se asignaron valores de cero a los valores nulos. Esta decisión se tomó con el

objetivo de evitar distorsiones en el análisis y garantizar la precisión de las métricas relacionadas con estas variables.

- **Asignación de "NO"**: Para las columnas `beneficio_es_mutuocolombia`, `reconocimiento_happy_friday` y `reconocimiento_toma_dias`, se llenaron los valores nulos con "NO". Esto indica la ausencia de beneficios o reconocimientos en esos períodos específicos.
- **Eliminación de registros nulos**: Finalmente, se eliminaron los registros que tenían valores nulos en las columnas `desempeño_valoracion_global` y `promedio_ytqd` para asegurar la integridad del análisis en esas métricas específicas.

La combinación de estas técnicas de relleno de nulos permitió obtener una base de datos final coherente y completa, lista para el análisis exploratorio de datos y la posterior clusterización. Al aplicar estas estrategias de manera deliberada y fundamentada en el contexto específico del estudio, se aseguró que los resultados obtenidos sean confiables y representativos de la realidad de los empleados de Bancolombia.

Análisis de Motivos de Retiro Excluidos

Motivos Excluidos: Se omitieron motivos como "Retiro por pensión", "Retiro por fallecimiento", "Retiro por justa causa", entre otros, debido a la claridad existente sobre el motivo de renuncia. Estos motivos ya tienen causas evidentes y no requieren investigación adicional.

Motivos Considerados: Se incluyen retiros por otras causas, sin justa causa, retiro voluntario y voluntario con arreglo mutuo, que ofrecen perspectivas cruciales para comprender la dinámica de rotación en Bancolombia.

Enfoque en Investigación Profunda: Se planea investigar más a fondo los retiros sin justa causa, retiro voluntario y retiro voluntario con arreglo mutuo. Se quiere investigar a ciencia cierta qué pasó con estos empleados, ya que estos casos podrían revelar factores subyacentes que influyen en estas decisiones, fortaleciendo el análisis predictivo y las decisiones estratégicas para mejorar la retención y satisfacción laboral.

Cabe aclarar que para realizar este proyecto se hace una suposición importante. Debido al desequilibrio en el número de registros mensuales de empleados, hay una gran desigualdad de clases: solo existe un registro para cada empleado que abandona en el mes de su salida, frente a múltiples registros mensuales para aquellos que siguen activos. Para compensar esta disparidad, se propone una solución: considerar a los empleados que abandonaron en un mes determinado y etiquetar todo su recorrido desde su ingreso al banco con "ABANDONA". De esta manera, se puede lograr un equilibrio de clases de manera lógica, siguiendo esta hipótesis. Además, tiene sentido ver a ese empleado desde que entra a la empresa como un "ABANDONO", dado que en algún punto futuro dejará la organización según nuestros datos históricos.

Preparación de los datos

Se ha desarrollado un proceso para limpiar el conjunto de datos, centrado en eliminar registros que presenten valores 0 o nulos en múltiples columnas específicas, siempre y cuando estos registros estén asociados con una etiqueta de motivo de retiro "ACTIVO". Inicialmente, se identifican aquellos registros que contienen exclusivamente valores 0 o nulos en todas las columnas designadas para su verificación. Posteriormente, se localizan registros que presentan al menos tres valores 0 en estas columnas. Los índices de estos registros seleccionados se combinan y se eliminan del conjunto de datos principal.

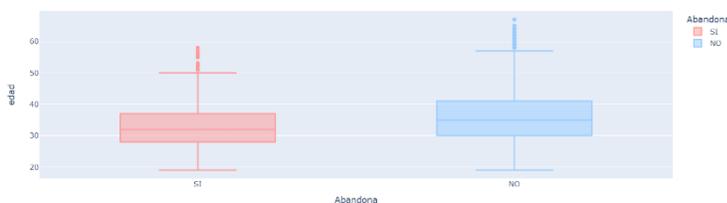
Adicionalmente, se aborda la necesidad de reducir la presencia de registros con valores de 0.0 en la columna 'credito_saldo_promedio_millones'. Para esto, se eliminan de manera aleatoria la mitad de estos registros, utilizando un divisor ajustable según sea necesario para alcanzar el objetivo deseado. Este proceso garantiza la integridad y la consistencia de los datos, preparándolos adecuadamente para análisis y modelado posterior.

Análisis exploratorio de los datos

El análisis muestra que las variables con correlaciones más fuertes y negativas con el abandono de empleados son la formación acumulada en cursos (-0.402456), la antigüedad en la empresa (-0.253586), la edad (-0.219379) y el saldo promedio en créditos (-0.243713). Esto sugiere

que los empleados que reciben más formación tienen mayor antigüedad y edad, y gestionan mejor sus finanzas, son menos propensos a abandonar la empresa. En contraste, factores como días de vacaciones pendientes, tamaño del personal y horas extras muestran correlaciones muy bajas, indicando poca influencia en la decisión de abandonar. Estos hallazgos subrayan la importancia de invertir en la formación y desarrollo continuo de los empleados como estrategia clave para mejorar la retención.

En los análisis categóricos, se confirmaron algunos de estos hallazgos más relevantes. Por ejemplo, cómo vemos en la Figura 1 la

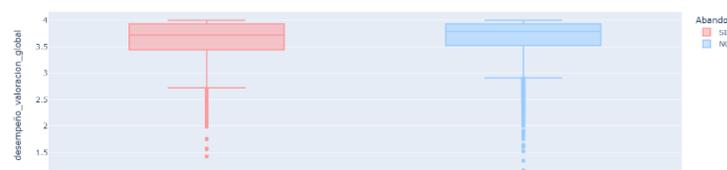


media de edad de los empleados que abandonan es de 32 años, en comparación con una media de 35 años para aquellos que no abandonan.

Figura 1: Boxplot abandona vs edad

Además, se observó en otros boxplots similares que los empleados solteros mostraron una mayor propensión a abandonar, seguidos por los separados, unión libre y luego los casados. En términos de nivel de cargo, se destacó que los profesionales demostraron una tendencia más alta de abandono en comparación con los operativos. Además, se encontró que los empleados no convencionales tienden a abandonar más, lo que sugiere un posible impacto de la afiliación sindical en la retención del personal.

Los boxplots proporcionaron una visualización adicional de estos hallazgos. Aunque no se



observó una diferencia clara en la valoración de desempeño entre los que abandonan y los que no (Figura2),

Figura 2: Boxplot de abandona vs desempeño global

se encontró que los empleados que no abandonan obtienen créditos de mayor valor (Figura 3) y una mayor cantidad de cursos acumulados. Por otro lado, aquellos que abandonan

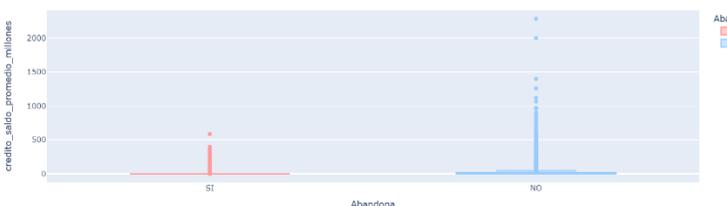


Figura 3: Boxplot de abandona vs días pendientes vacaciones

tienden a trabajar menos horas extras en promedio.



Figura 4: Boxplot de abandona vs cursos terminados acumulados

terminados. Esta diferencia es notable.

En conclusión, el análisis exploratorio sugiere varias hipótesis sobre el abandono de empleados en Bancolombia que deberán ser verificadas en etapas posteriores del estudio. Las correlaciones observadas indican que una mayor formación acumulada, mayor antigüedad, mayor edad y una mejor gestión financiera podrían estar asociadas con una menor propensión a abandonar la empresa. Además, los empleados más jóvenes, solteros, profesionales y no sindicalizados parecen tener una mayor tendencia a abandonar. Estos hallazgos preliminares serán evaluados mediante un modelo más avanzado para confirmar si estas variables tienen una relación significativa con la retención de empleados.

Análisis temporal de ingresos vs retiros

La Figura 5 muestra la evolución mensual de los ingresos y retiros de empleados en Bancolombia

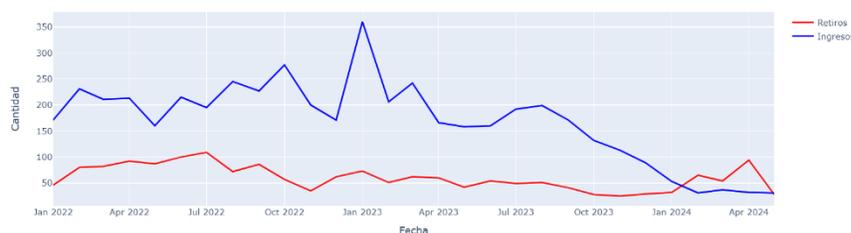


Figura 5: Ingresos vs Retiros desde 2022 hasta abril 2024

durante el período de enero 2022 a abril 2024. Se observa una tendencia general de mayores ingresos en comparación con los retiros, aunque ambos indicadores muestran variaciones significativas en diferentes momentos del período.

Ingresos

- **Enero 2022 - abril 2022:** Los ingresos comienzan con un aumento significativo, alcanzando aproximadamente 200 empleados en abril de 2022.
- **Octubre 2022 - enero 2023:** Se registra un pico notable en enero de 2023, con alrededor de 350 empleados ingresando, lo que podría estar relacionado con una expansión o un proyecto de gran envergadura.
- **Mayo 2023 - abril 2024:** A partir de mayo de 2023, los ingresos muestran una tendencia decreciente, con algunos incrementos menores, pero terminan en abril de 2024 con un valor cercano a cero.

Retiros

- **Enero 2022 - octubre 2022:** Los retiros muestran un aumento gradual, alcanzando un pico en octubre de 2022 con alrededor de 120 empleados, posiblemente debido a una reorganización interna.
- **Noviembre 2022 - abril 2023:** Tras el pico de octubre, los retiros se estabilizan alrededor de 50 empleados por mes, con pequeñas fluctuaciones.
- **Marzo 2024:** Se observa un incremento temporal en los retiros, alcanzando aproximadamente 50 empleados, seguido de una rápida disminución.

Factores Externos Influenciantes en 2022

Impacto de la Entrada de Competidores Fintech: Un experto con amplia experiencia laboral en Bancolombia sugirió que la alta rotación de personal en 2022 coincidió con la entrada de competidores Fintech en el mercado, como Nu Bank y Rappi. Estos nuevos actores demandaron muchos cargos estratégicos, como ingenieros de software y analistas, ofreciendo sueldos más competitivos que la media del sector. Esta situación probablemente contribuyó al aumento de retiros observados en ese año, ya que empleados clave fueron atraídos por mejores ofertas laborales en las Fintech.

Puntos Clave

- **Pico en Ingresos (enero 2023):** El ingreso masivo de empleados en enero de 2023 sugiere períodos de contratación intensiva, posiblemente relacionados con proyectos estratégicos o expansiones.
- **Pico en Retiros (octubre 2022):** El aumento en los retiros en octubre de 2022 podría estar relacionado con la atracción de empleados por competidores Fintech, lo que refleja un mercado laboral más competitivo.
- **Tendencia Decreciente (2024):** La tendencia hacia la disminución en los ingresos y retiros a principios de 2024 indica una estabilización en la plantilla de empleados.

Modelo supervisado

Analizando la gráfica de la evolución de los ingresos y retiros del personal de Bancolombia, observamos picos significativos en ambos durante el período de 2022, coincidiendo con la entrada de competidores Fintech en el mercado. Este análisis inicial nos lleva a la necesidad de profundizar en las variables que pueden influir en estos movimientos de personal. Para ello, hemos utilizado un modelo supervisado XGBoost, que nos permitirá identificar las variables más importantes que explican estos comportamientos.

Una vez seleccionadas estas variables relevantes con XGBoost, procederemos a aplicar un modelo no supervisado, específicamente K-means, para agrupar los datos y entender mejor los patrones subyacentes en la rotación del personal. A continuación, describimos la metodología utilizada para la selección de variables:

Metodología para la Selección de Variables con XGBoost

- **Normalización y Codificación de Variables:** Antes de aplicar el modelo XGBoost, las variables se normalizaron utilizando StandardScaler y se codificaron las variables categóricas. La normalización asegura que todas las variables tengan la misma escala, mientras que la codificación permite que el modelo procese adecuadamente las variables no numéricas.
- **División del Conjunto de Datos:** El conjunto de datos se dividió en un conjunto de entrenamiento (train) y uno de validación (validation) para evaluar la capacidad del modelo en datos no vistos.

- **Validación Cruzada con 10 Folds:** Se aplicó una validación cruzada con 10 folds para asegurar que los resultados del modelo sean robustos y no dependan de una sola partición del conjunto de datos. La validación cruzada permite promediar los resultados obtenidos en cada fold, proporcionando un puntaje más confiable del desempeño del modelo.
- **Entrenamiento del Modelo XGBoost:** El modelo XGBoost se entrenó utilizando las 34 variables originales. Durante el entrenamiento, se calcularon las importancias de las variables, seleccionando aquellas que más contribuían al rendimiento del modelo.
- **Resultados del XGBoost:** Se decidió no ajustar hiperparámetros debido al excelente rendimiento obtenido por el modelo. Los puntajes de Validación Cruzada fueron consistentes: 0.9804, 0.9801, 0.9797, 0.9804, 0.9727, 0.9815, 0.9752, 0.9867, 0.9829, con un promedio de 0.9798, demostrando alta consistencia y generalización del modelo en diferentes particiones del conjunto de datos. En el Conjunto de Validación, el modelo muestra una precisión de 0.96 para la Clase 0 (empleados que no abandonan) y 0.99 para la Clase 1 (empleados que abandonan), con un recall de 0.99 para la Clase 0 y 0.96 para la Clase 1. Esto se traduce en un F1-Score de 0.98 para ambas clases, indicando un equilibrio entre precisión y recall. La exactitud global es de 0.98, reflejando un alto nivel de predicciones correctas en general. El análisis macro y ponderado de las métricas de precisión, recall y F1-Score también arroja un promedio de 0.98, subrayando la consistencia del modelo en situaciones de clases desbalanceadas.

En resumen, el modelo XGBoost ha demostrado ser robusto y altamente efectivo, con excelentes métricas en ambas clases y una alta capacidad para identificar correctamente casos de empleados que no abandonan y casos de empleados que abandonan. Los resultados de Validación Cruzada validan su rendimiento consistente en diferentes particiones de datos. Estas características confirman su idoneidad para la selección de variables importantes y respaldan su aplicación en el siguiente paso del análisis con el algoritmo K-means para explorar patrones subyacentes más detalladamente.

- 1. Selección de Variables Importantes:** Las variables más importantes se identificaron y se seleccionaron 25 para su uso en el modelo no supervisado. Las características seleccionadas fueron: credito_saldo_promedio_millones, formacion_cursos_terminados_acum, ausentismo_dias_naturales, horas_extras_promedio, desempeño_valoracion_global, antiguedad, ahorros_es_mutuocolombia_NO, edad, antiguedad_cargo, promedio_ytqd, vacaciones_dias_pendientes, nivel_cargo_PROFESIONAL, division_personal_ANTIOQUIA, area_personal_No Convencionad (CO), nivel_cargo_OPERATIVO, division_personal_BOGOTA Y SABANA, modalidad_trabajo_PRESENCIAL, estado_civil_sol., genero_Femenino, modalidad_trabajo_FLEXIWORK, division_personal_CENTRO, reconocimiento_happy_friday_NO, familiares_hijos_mayores_18_NO, modalidad_trabajo_SUPLEMENTARIO, vivienda_tipo_Propia.

A parte de la obtención de variables que serán enviadas al K-Means se crea un modelo para verificar por índice si el empleado tiende a irse o no, por ejemplo, este empleado tiene una proporción alta a abandonar, y nos muestra porque su promedio es más alto, recordemos que no deben tomarse los valores porque están normalizados, pero si podemos verificar que variables hacen que ese empleado sea propenso o no al abandono

Modelo no supervisado K-Means

Preparación de los datos

Variables: En esta etapa, se eliminaron variables que no aportan significativamente al modelo. Variables como "Motivo de Retiro" esta es la variable que etiqueta los retiros que queremos indagar y queremos que busque grupos sin sesgo. Las variables "Cargo" y "Unidad Organizativa" tienen demasiadas categorías, lo que introduce ruido y complejidad innecesaria. Además, se eliminaron variables con baja variabilidad que no ofrecen información útil para el análisis.

Eliminación de Outliers: Se aplicó el Z-Score para identificar y eliminar outliers, utilizando un umbral de 3 desviaciones estándar.

Selección del Número Óptimo de Clústers y Análisis de Resultados: Después de eliminar los outliers del conjunto de datos, determinamos el número óptimo de clusters utilizando el método del codo y el coeficiente de silueta. Sin embargo, estos métodos no siempre proporcionan una respuesta clara, por lo que se desarrolló una función que realiza 10 iteraciones para cada posible número de clústers, guardando los centroides, el promedio del porcentaje de varianza explicada y la desviación estándar del proceso.

Resultados de la Función para la Selección de Clústers

La siguiente tabla muestra los resultados de las 10 iteraciones:

Tabla 2: Promedios 10 iteraciones por 10 clústers para encontrar el clúster óptimo

Clústers	Promedio	Varianza Explicada	Desviación	Varianza Explicada	Diferencia
2	0.004		0.000		0.004
3	0.014		0.000		0.014
4	0.033		0.001		0.031
5	0.069		0.006		0.064
6	0.124		0.001		0.123
7	0.199		0.006		0.193
8	0.284		0.015		0.269
9	0.399		0.016		0.383
10	0.530		0.010		0.520

Aunque tener 10 clústers ofrece un mejor rendimiento en términos de porcentaje de varianza explicada, no es práctico debido a la complejidad de interpretar tantos clústers. Buscamos un número de clústers que equilibre rendimiento y simplicidad, y observamos que 6 clusters explican más del 10% de la varianza, lo cual es una base sólida para análisis posteriores.

Análisis Visual y Selección Final de Clústers

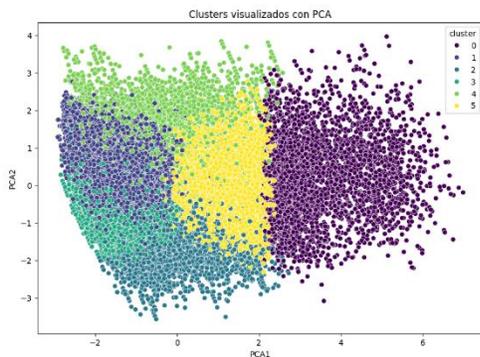


Figura 6: PCA con etiquetas de KMeans 6 Clusters

con el K-Means,

Después de ejecutar el K-means con 6 clústers y visualizar los resultados utilizando un PCA de 2 dimensiones (que explica el 42% de la varianza), notamos que los grupos no se forman claramente y se traslapan (Figura 6). Este porcentaje de varianza (42%) es significativo para visualizar clústers, ya que representa una gran parte de la información total. Sin embargo, al usar PCA para tomar decisiones sobre los clústers, esta varianza sería insuficiente. Por eso es por lo que tomaremos decisiones y estrategias no con el PCA sino

Decidimos probar con diferentes números de clústers (5 y 4) y evaluarlos visualmente.

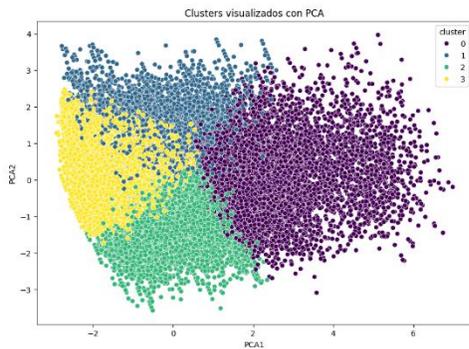
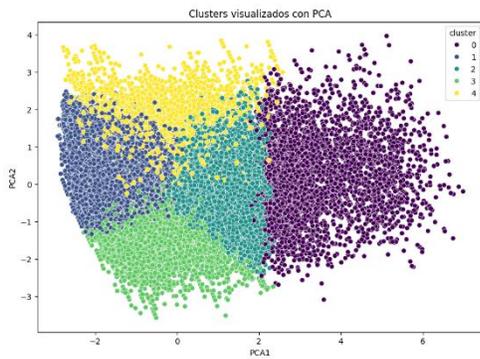


Figura 7: PCA con etiquetas de KMeans 4 Clusters

tanto, optamos por utilizar 5 clústers para el análisis final

Observamos en la Figura 7 que con 4 clústers los grupos eran más definidos, pero aún existían traslapes considerables. Finalmente, en la Figura 8, vimos que con 5 clústers se obtenían los mejores resultados visuales, con grupos más definidos y coherentes. Además, este enfoque no solo proporcionaba clústers más definidos, sino que también ofrecía un equilibrio adecuado y coincidía con los cinco grupos identificados en una investigación cualitativa previa, realizada mediante una encuesta en el banco. Por lo



Se utilizó el algoritmo K-Means para realizar la clusterización inicial con 5 clústers, aplicando promedio de 10 iteraciones para aumentar la estabilidad y fiabilidad de los resultados mediante 10 iteraciones, recalculando los clústers en cada iteración y promediando los resultados para obtener clusters más robustos.

Figura 8: PCA con etiquetas de KMeans 5 Clusters

5 Análisis

A continuación, se presenta el análisis de clústers aplicado a datos estandarizados de empleados dentro de una organización. Los resultados se presentan en una escala relativa de bajo, medio y alto, permitiendo identificar patrones y características distintivas de cada grupo. No se presentan los valores exactos, sino una interpretación escalada que facilita entender las tendencias y necesidades de cada clúster.

Clúster 0: "Estabilidad y Experiencia" (Bajo Abandono)

Descripción: Este clúster se caracteriza por empleados con alta antigüedad tanto en el cargo como en la empresa, y una edad relativamente mayor. Aunque el reconocimiento Happy Friday es alto, hay una percepción positiva hacia la formación y estabilidad laboral. La modalidad de trabajo presencial es común, y muchos empleados están casados y en niveles operativos.

Características Clave:

- Alto:
 - Antigüedad en el Cargo: 1.9255
 - Antigüedad: 1.9160
 - Edad: 1.6727
 - Reconocimiento Happy Friday No: 0.8552
- Moderado:
 - Formación y Cursos Terminados Acumulados: 0.8797
 - Modalidad de Trabajo Presencial: 0.6272
 - Estado Civil Casado: 0.5555
 - Nivel de Cargo Operativo: 0.4934
 - Nivel de Cargo Profesional: 0.2719
- Bajo:
 - Abandona: 0.1629
 - Horas Extras Promedio: -0.2108
 - Ahorros en Mutuocolombia No: 0.1198
 - Vacaciones Días Pendientes: 0.1195

- Modalidad de Trabajo Suplementario: 0.1105
- Crédito Saldo Promedio: 0.0639
- Desempeño Valoración Global: 0.0395
- Movilidad Medio Transporte Moto: 0.0293
- Ausentismo Días Naturales: 0.0138
- Promedio YTQD: 0.0127

Clúster 1: "Operativos Jóvenes y No Reconocidos" (Medio Abandono)

Descripción: Este clúster agrupa principalmente a empleados jóvenes con baja antigüedad en la empresa y en sus cargos. Tienen una alta presencia en roles operativos y un alto porcentaje de solteros. El reconocimiento Happy Friday es muy bajo, lo que sugiere una posible insatisfacción en este aspecto. La modalidad de trabajo es mayoritariamente presencial y tienen pocos días de vacaciones pendientes.

Características Clave:

- Alto:
 - Reconocimiento Happy Friday No: 0.9269
 - Modalidad de Trabajo Presencial: 0.7061
 - Estado Civil Soltero: 0.7387
 - Nivel de Cargo Operativo: 0.6989
- Moderado:
 - Abandona: 0.5418
 - Ahorros en Mutuocolombia No: 0.5012
 - Promedio YTQD: 0.3522
 - Nivel de Cargo Profesional: 0.2766
 - Desempeño Valoración Global: 0.2086
- Bajo:
 - Antigüedad: -0.7947
 - Edad: -0.7486
 - Antigüedad en el Cargo: -0.7263
 - Vacaciones Días Pendientes: -0.6931

- Formación y Cursos Terminados Acumulados: -0.3281
- Estado Civil Casado: 0.1139
- Crédito Saldo Promedio: -0.1073
- Horas Extras Promedio: 0.0870
- Ausentismo Días Naturales: -0.0828
- Movilidad Medio Transporte Moto: 0.0698
- Modalidad de Trabajo Suplementario: 0.0565
- Vivienda Propia: 0.0467
- Nivel de Cargo jefe: 0.0210

Clúster 2: "Profesionales Solteros y Con Baja Antigüedad" (Alto Abandono)

Descripción: Este grupo está compuesto mayoritariamente por profesionales jóvenes con poca antigüedad tanto en el cargo como en la empresa. Son solteros y tienen pocos días de vacaciones pendientes. El reconocimiento Happy Friday es bajo y tienen una tendencia alta a abandonar la empresa.

Características Clave:

- Alto:
 - Vacaciones Días Pendientes: 0.9305
 - Reconocimiento Happy Friday No: 0.8802
 - Estado Civil Soltero: 0.6965
 - Nivel de Cargo Profesional: 0.6284
- Moderado:
 - Abandona: 0.6458
 - Ahorros en Mutuocolombia No: 0.5048
 - Desempeño Valoración Global: 0.4808
 - Modalidad de Trabajo Presencial: 0.2570
 - Nivel de Cargo Operativo: 0.2553
- Bajo:
 - Antigüedad en el Cargo: -0.4401
 - Antigüedad: -0.4089

- Edad: -0.4007
- Promedio YTQD: -0.2192
- Horas Extras Promedio: -0.2152
- Modalidad de Trabajo Suplementario: 0.1927
- Estado Civil Casado: 0.1761
- Formación y Cursos Terminados Acumulados: -0.1287
- Nivel de Cargo jefe: 0.1020
- Ausentismo Días Naturales: -0.0901
- Vivienda Propia: 0.0688
- Crédito Saldo Promedio: -0.0453
- Movilidad Medio Transporte Moto: 0.0337

Clúster 3: "Operativos Solteros y No Reconocidos" (Alto Abandono)

Descripción: Este clúster incluye empleados en roles operativos con baja antigüedad, principalmente solteros y con una valoración negativa del desempeño. El reconocimiento Happy Friday es muy bajo, lo que sugiere una insatisfacción considerable. Tienen una alta tendencia a abandonar la empresa.

Características Clave:

- **Alto:**
 - Reconocimiento Happy Friday No: 0.9515
 - Modalidad de Trabajo Presencial: 0.8766
 - Nivel de Cargo Operativo: 0.9042
- **Moderado:**
 - Estado Civil Soltero: 0.5656
 - Ahorros en Mutuocolombia No: 0.4438
- **Bajo:**
 - Desempeño Valoración Global: -1.7060
 - Abandona: 0.6486
 - Promedio YTQD: -0.2804
 - Antigüedad: -0.2006

- Estado Civil Casado: 0.1973
- Antigüedad en el Cargo: -0.1953
- Edad: -0.1902
- Crédito Saldo Promedio: -0.1596
- Horas Extras Promedio: -0.0903
- Nivel de Cargo Profesional: 0.0769
- Vivienda Propia: 0.0763
- Ausentismo Días Naturales: -0.0611
- Movilidad Medio Transporte Moto: 0.0447
- Modalidad de Trabajo Suplementario: 0.0183
- Nivel de Cargo jefe: 0.0179
- Vacaciones Días Pendientes: 0.0114

Clúster 4: "Operativos con Alta Antigüedad y Moderado Reconocimiento" (Medio Abandono)

Descripción: Este grupo está compuesto por empleados operativos con alta antigüedad y moderado reconocimiento Happy Friday. Son mayoritariamente casados y tienen una edad y antigüedad superiores a la media. Tienen una valoración de desempeño positiva y una baja cantidad de días de vacaciones pendientes.

Características Clave:

- **Alto:**
 - Reconocimiento Happy Friday No: 0.9059
 - Nivel de Cargo Operativo: 0.6655
 - Modalidad de Trabajo Presencial: 0.6213
- **Moderado:**
 - Edad: 0.5651
 - Antigüedad: 0.5189
 - Estado Civil Casado: 0.5048
 - Antigüedad en el Cargo: 0.4739
 - Abandona: 0.4253

- Ahorros en Mutuocolombia No: 0.3580
- Desempeño Valoración Global: 0.3102
- Estado Civil Soltero: 0.2856
- Nivel de Cargo Profesional: 0.2308
- Modalidad de Trabajo Suplementario: 0.1497
- Vivienda Propia: 0.1265
- Nivel de Cargo jefe: 0.0900
- **Bajo:**
 - Horas Extras Promedio: -0.1182
 - Crédito Saldo Promedio: -0.0946
 - Formación y Cursos Terminados Acumulados: 0.0481
 - Movilidad Medio Transporte Moto: 0.0372
 - Vacaciones Días Pendientes: -0.0310
 - Ausentismo Días Naturales: -0.0303
 - Promedio YTQD: 0.0138

Estrategias específicas por clúster

Clúster 0: "Veteranos Estables y Moderadamente Reconocidos" (Bajo Abandono)

Puntos Fuertes:

- Alta antigüedad en el cargo y la empresa, edad avanzada.
- Formación continua con varios cursos completados.
- Reconocimiento moderado, indicando una percepción positiva.

Estrategias:

- Reforzar la estabilidad laboral y programas de reconocimiento continuos.
- Ofrecer oportunidades de mentoría para aprovechar su experiencia.

Clúster 1: "Jóvenes Operativos No Reconocidos" (Alto Abandono)

Puntos Fuertes:

- Alta presencia de empleados solteros y en roles operativos.
- Trabajo presencial y falta de reconocimiento Happy Friday.

Estrategias:

- Intensificar programas de reconocimiento.
- Implementar oportunidades de desarrollo profesional y formación.
- Revisar y adaptar beneficios para apoyar a empleados jóvenes.

Clúster 2: "Profesionales Solteros con Alta Percepción de Desarrollo" (Medio Abandono)

Puntos Fuertes:

- Alta cantidad de días de vacaciones pendientes y reconocimiento moderado.
- Solteros y en roles profesionales con percepción positiva de desarrollo.

Estrategias:

- Considerar opciones de flexibilidad laboral.
- Ofrecer más beneficios personalizados, como días de vacaciones y planes de ahorro.

Clúster 3: "Operativos Solteros y No Reconocidos" (Alto Abandono)

Puntos Fuertes:

- Alta presencia en roles operativos, solteros y trabajo presencial.
- Baja valoración de desempeño y falta de reconocimiento.

Estrategias:

- Implementar programas de reconocimiento continuos.
- Ofrecer formación y desarrollo profesional.

- Considerar opciones de flexibilidad laboral.
- Revisar beneficios para satisfacer mejor las necesidades de empleados solteros.

Clúster 4: "Operativos con Alta Antigüedad y Moderado Reconocimiento" (Medio Abandono)

Puntos Fuertes:

- Alta antigüedad y edad, mayormente casados.
- Moderado reconocimiento y percepción positiva de desempeño.

Estrategias:

- Reforzar la formación continua y programas de reconocimiento.
- Potenciar programas de bienestar y balance vida-trabajo.
- Ofrecer beneficios personalizados, considerando sus necesidades específicas.

Conclusiones del análisis

El análisis de los clústers de empleados revela diversas características y necesidades que requieren estrategias específicas para reducir la probabilidad de abandono y mejorar la satisfacción laboral. A continuación, se presentan conclusiones generales y estrategias aplicables a cada clúster:

1. **Importancia del Reconocimiento y Motivación:** La falta de reconocimiento es una variable alta en varios clústers (Clústers 1, 2, 3 y 4). Implementar programas de reconocimiento continuos y significativos es crucial. Esto puede incluir el Happy Friday y otros incentivos que valoren el esfuerzo diario.
2. **Desarrollo Profesional y Formación Continua:** Ofrecer oportunidades claras de desarrollo profesional es vital, especialmente en clústers con baja percepción de desarrollo (Clústers 1 y 3). Programas de formación, mentorías y planes de carrera pueden mejorar la percepción del desarrollo y reducir la intención de abandono.
3. **Flexibilidad Laboral:** La flexibilidad laboral es una estrategia efectiva para mejorar la satisfacción y el equilibrio entre la vida laboral y personal. Esta es

particularmente relevante para clústers con empleados jóvenes y en modalidad presencial (Clúster 2 y Clúster 3).

4. **Beneficios y Compensaciones Personalizadas:** Revisar y adaptar los beneficios para satisfacer mejor las necesidades de cada segmento es esencial. Esto incluye considerar beneficios específicos como días de vacaciones adicionales, incentivos financieros y planes de ahorro. (Clúster 1, Clúster 2, Clúster 3, Clúster 4)
5. **Comunicación y Retroalimentación Efectiva:** Mejorar los canales de comunicación y la retroalimentación es crucial para fortalecer el compromiso y la motivación de los empleados, especialmente en clústers con alta probabilidad de abandono (Clúster 2 y Clúster 3).

Conclusiones

1. El análisis exhaustivo de estos datos ha permitido identificar patrones claros y diferencias significativas entre los distintos grupos de empleados. Estos patrones son fundamentales para entender los factores que influyen en la satisfacción laboral y la probabilidad de abandono.
2. El modelo predictivo desarrollado ha demostrado ser eficaz en la identificación de empleados con alto riesgo de rotación. Utilizando variables clave como antigüedad, nivel de satisfacción y reconocimiento, el modelo puede prever con precisión la probabilidad de abandono, proporcionando una herramienta valiosa para la toma de decisiones.
3. La segmentación de empleados mediante técnicas de clustering ha permitido crear grupos homogéneos que reflejan las diferentes necesidades y características de los empleados. Esto facilita la personalización de estrategias de retención y mejora la efectividad de las intervenciones.
4. Las estrategias de retención diseñadas, basadas en los resultados del modelo predictivo y la segmentación, son específicas y alineadas con la cultura organizativa de Bancolombia. Estas estrategias abordan las necesidades particulares de cada clúster, mejorando la probabilidad de éxito en la retención de talento.
5. La comunicación clara y efectiva de los resultados y estrategias propuestas ha sido crucial para asegurar la comprensión y aceptación por parte del equipo analítico de Talento y Cultura. Esta comunicación ha facilitado la toma de decisiones informadas y ha promovido la implementación de las estrategias diseñadas.

Recomendaciones

- Integrar un clúster vinculado al perfil del personal para diseñar estrategias altamente personalizadas y específicas. Esto permitirá adaptar las acciones y decisiones a las características únicas de cada empleado, maximizando así su eficiencia y satisfacción.
- Desarrollar un sistema de recomendación para el área de solicitudes de puesto. Este sistema proporcionará puntos de vista detallados sobre cada candidato, basándose en el resultado del modelo aplicado a empleados similares. De esta manera, se facilitará la selección de los candidatos más adecuados.
- Crear un sistema de recomendación que se ejecute semestralmente, analizando las variables de los empleados actuales. Este sistema identificará a aquellos empleados que están en riesgo de abandonar la empresa y sugerirá estrategias específicas para retenerlos, reduciendo así la rotación de personal y promoviendo un entorno laboral más estable y comprometido.

Referencias

- Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology*, 63(1), 1-18.
- American Psychological Association [APA]. (2020). *Publication Manual of the American Psychological Association* (7a ed.). American Psychological Association.
- Becker, B. E., Huselid, M. A., & Ulrich, D. (2001). *The HR Scorecard: Linking People, Strategy, and Performance*. Harvard Business Review Press.
- Bock, L. (2015). *Work Rules!: Insights from Inside Google That Will Transform How You Live and Lead*. Twelve.
- Davenport, T. H., Harris, J. G., & Shapiro, J. (2010). Competing on talent analytics. *Harvard Business Review*, 88(10), 52-58.
- Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 26(3), 463-488.
- Kumar, V., & Ravi, V. (2007). Predicting the behavior of employees using data mining techniques. *Expert Systems with Applications*, 32(1), 106-113.
- March, J. G., & Simon, H. A. (1958). *Organizations*. Wiley.
- Stone, R. J., & Ramsay, M. (2004). *The recruitment process: One company's approach*. Prentice Hall.
- Zhang, Z., Zhao, S., & Qi, X. (2018). Predicting employee turnover using machine learning techniques. *Human Resource Management*, 57(4), 889-905. <https://doi.org/10.1002/hrm.21892>