



Construcción de una sábana de datos para el desarrollo de un prototipo de analítica que permita la simulación y evaluación de los impactos financieros y de valor 360 en el portafolio, derivados de riesgos de transición en dos de los segmentos productivos, para una empresa en el sector de oil & gas.

Daniel Góngora García

Trabajo de Grado presentado para optar al Título de Ingeniero Industrial

Asesor

Pablo Andrés Maya Duque, Título académico más alto en Área de formación del asesor interno

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería Industrial

Medellín, Antioquia

2024

| | |
|----------------------------|---|
| Cita | (Góngora García, 2024) |
| Referencia | Góngora Garcia, D. (2024). <i>Construcción de una sábana de datos para el desarrollo de un prototipo de analítica que permita la simulación y evaluación de los impactos financieros y de valor 360 en el portafolio, derivados de riesgos de transición en dos de los segmentos productivos, para una empresa en el sector de Oil & Gas.</i> [Trabajo de grado profesional]. Universidad de Antioquia, Medellín, Colombia. |
| Estilo APA 7 (2020) | |



Centro de Documentación Ingeniería (CENDOI)

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

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda López

Decano/director: Jesús Francisco Vargas Bonilla.

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

Para Sebastián, Juliana y Angela, por que sin ellos no hubiera llegado hasta acá.

Agradecimientos

Agradecer primeramente a la universidad por guiarme a donde estoy hoy, agradezco también al equipo de trabajo en Accenture, Saulo, Lorena, Steven, Camilo, Simón y Valentina, por acogerme y enseñarme cosas nuevas cada día. Y finalmente agradezco el apoyo incondicional de mi familia, y las enseñanzas que me dejaron todas y cada una de las personas que conocí en este proceso.

Tabla de contenido

| | |
|-------------------------------------|----|
| Resumen | 8 |
| Abstract | 9 |
| Introducción | 10 |
| 1 Objetivos | 11 |
| 1.1 Objetivo general | 11 |
| 1.2 Objetivos específicos..... | 11 |
| 2 Marco teórico | 12 |
| 3 Metodología | 14 |
| 3.1 Primera Fase | 14 |
| 3.2 Segunda Fase | 14 |
| 3.3 Tercera Fase | 14 |
| 3.4 Cronograma de actividades | 15 |
| 4 Resultados y Análisis | 18 |
| 4.1 Extracción..... | 20 |
| 4.2 Transformación | 21 |
| 4.3 Carga | 23 |
| 5 Conclusiones | 24 |
| Referencias | 25 |
| Anexos..... | 26 |

Lista de tablas

| | |
|--|----|
| Tabla 1. Sprint planning de Ingeniería de Datos..... | 17 |
| Tabla 2. Librerías utilizadas en el proyecto y su funcionalidad..... | 20 |

Lista de ilustraciones

| | |
|---|----|
| Ilustración 1. Datos suministrados por el cliente contienen planeaciones a largo plazo y planeación y evaluación efectiva de proyectos..... | 18 |
| Ilustración 2. Arquitectura del proyecto..... | 19 |
| Ilustración 3. Tecnologías y plataformas utilizadas para abordar el proceso de ETL. | 19 |
| Ilustración 4. Snippet que extrae la información del archivo JSON y los guarda en una lista de Python..... | 20 |
| Ilustración 5. Snippet que convierte las listas obtenidas con el fragmento de la Ilustración 4 en columnas de un pandas DataFrame..... | 21 |
| Ilustración 6. Rubros correspondientes a cada segmento y los tipos de modelación requeridos para cubrir el alcance del proyecto..... | 21 |
| Ilustración 7. Snippet que representa la lógica de uno de los tipos de modelado. | 22 |
| Ilustración 8. Las sabanas de datos resultantes alojadas en DBFS. | 23 |

Siglas, acrónimos y abreviaturas

| | |
|--------------|--|
| CSV | Coma Separated Values |
| DBFS | Databricks File System |
| ETL | <i>Extract, Transform, Load</i> (Extraer, transformar, cargar) |
| PLP | Planeación a largo plazo |
| PEEP | Planeación y evaluación efectiva de proyectos |
| ACT | Accenture |
| Ltda. | Limitada. |
| UdeA | Universidad de Antioquia |
| STEPS | Stated Policies Scenario |
| APS | Announces Pledges Scenario |
| NZE | Net Zero Emissions at 2050 Scenario |

Resumen

La sábana de datos será el insumo de un prototipo analítico permitirá visualizar las implicaciones financieras y de valor 360 ocasionadas por la materialización de los daños físicos a los activos y su impacto en la operación. Esta visualización se habilitará para hasta 10 variables claves asociadas a los riesgos de transición (mercado, regulación, tecnología y reputación), además de generar sabanas ‘secundarias’ compuestas por agrupaciones de la sabana principal, con el objetivo de facilitar la visualización de los datos.

La construcción de esta sabana se realizará con unos insumos (Archivos de Excel) principales proporcionados por el cliente, que consumieron una serie de Jupyter Notebooks por medio de la librería pandas y por medio de algunos scripts en el lenguaje de programación Python generará la sabana en un nuevo archivo de Excel.

Palabras clave: escenarios climáticos, riesgos de transición asociados al clima, apetito de riesgo, exposición, impacto, datos.

Abstract

The data sheet will be the input for an analytical prototype that will allow visualizing the financial and 360 value implications caused by the materialization of physical damages to assets and their impact on operations. This visualization will be enabled for up to 10 key variables associated with transition risks (market, regulation, technology, and reputation), in addition to generating 'secondary' sheets composed of groupings of the main sheet, with the aim of facilitating data visualization.

The construction of this sheet will be carried out with main inputs (Excel files) provided by the client, which were processed using a series of Jupyter Notebooks through the pandas library, and by means of some scripts in the Python programming language, it will generate the sheet in a new Excel file.

Keywords: climate scenarios, climate-related transition risks, risk appetite, exposure, impact, data.

Introducción

Accenture Ltda es una empresa multinacional de consultoría estratégica, servicios tecnológicos y externalización (outsourcing), este proyecto se desarrolló junto con un equipo multidisciplinar de las áreas de Technology, Consulting y Applied Intelligence (AI).

Una sábana de datos es el registro organizado de la información disponible, para este caso particular tenemos diversas fuentes de datos suministradas por el cliente, correspondientes a dos de los tres sectores principales de la industria del gas y petróleo (upstream, midstream y downstream), donde el upstream se enfoca en la exploración, extracción y producción de petróleo y gas, el midstream se encarga del transporte y almacenamiento de estos recursos, y el downstream abarca la refinación, distribución y venta al por menor de productos derivados del petróleo y el gas. Cada uno de estos segmentos desempeña un papel crucial en la cadena de suministro de la industria del petróleo y el gas; a partir de estos datos se automatizará la construcción de la sábana de datos limitando esta a los segmentos del upstream y uno adicional manejado por el cliente que contiene los proyectos que tienen emisiones de carbono bajas o nulas.

El prototipo analítico permitirá evaluar las implicaciones financieras y de valor 360 de los riesgos de transición mediante una modelación dinámica de los flujos de caja, partiendo de 50 activos priorizados, sensibilizando hasta 10 variables clave, y considerando los tres escenarios climáticos de la IEA (International Energy Agency), junto con los ajustes pertinentes de acuerdo con la realidad del país.

La premisa del producto analítico se reduce a que se debe lograr un modelo flexible y escalable que permita evaluar las implicaciones financieras y de valor 360 para el resto de los activos pertenecientes a la empresa de una forma rápida, una vez se garantice los criterios de aceptación aplicados sobre los 50 activos priorizados.

1 Objetivos

1.1 Objetivo general

Desarrollar una sábana de datos confiable para el modelamiento de un prototipo de analítica que permita la simulación y evaluación de los impactos financieros y de valor 360 en el portafolio, derivados de los riesgos de transición asociados al cambio climático. Esto con el objetivo de ser un insumo para la planeación financiera, aportando indicadores financieros y de sostenibilidad¹.

1.2 Objetivos específicos

- Automatizar la generación de tablas relacionadas con los riesgos de transición.
- Consignar en un archivo .py las funciones utilizadas para poder replicar la generación de la sábana de datos.
- Garantizar la calidad y replicabilidad de la sábana de datos.

¹ La contribución y generación de operaciones responsables, seguras y eficientes, armonizando el relacionamiento con el medio ambiente y los grupos de interés; bajo un marco de gobierno transparente y ético, haciendo uso de la tecnología para el desarrollo de soluciones innovadoras a los retos actuales y futuros.

2 Marco teórico

La necesidad de adaptarse con anticipación a riesgos que en el pasado se consideraban impredecibles es cada vez más latente en las empresas, desde hace unos años el mundo está viviendo cambios de paradigma gigantes, a un ritmo acelerado, como lo sufrimos con la pandemia del COVID-19, el avance incontrolado de la inteligencia artificial, entre otras, esto hace que modelar los riesgos de transición, no es algo extraño o fuera de lugar, estar preparados para lo impredecible, se volvió una necesidad.

Como ejemplo podemos citar a Jawadi, F., Rozin, P., & Bourghelle, D. (2023), quienes relacionan el impacto ambiental del aumento emisiones de CO2 en países europeos, con la pandemia del COVID-19, tratando este último como un riesgo de transición, y como la predicción y preparación para estos podría llevar a un mejor control y prevención de los riesgos relacionados al cambio climático además de reducir las emisiones de dióxido de carbono.

Para las empresas del sector, la descarbonización más que una tendencia es una necesidad si buscan seguir subsistiendo, y este cambio claramente es un riesgo de transición, por lo que también se busca ser modelado para tener una preparación previa, Colombia no es el primer país que le apunta a la descarbonización, ni en el que se habla de riesgos de transición; Colenbrander, S., Vaze, P., Vikas, C., Ayer, S., Kumar, N., Vikas, N., & Burge, L. (2023), hablan del caso de la India, y como una descarbonización afectaría los mercados financieros de dicho país, ya que como mencionan estos autores, no se puede hablar de una industria libre de carbono sin una transformación profunda del del sistema financiero global, la relación directa entre estos dos frentes lo explican más claramente Salisu, A. A., Ndako, U. B., & Vo, X. V. (2023), ya que la adaptación temprana a los riesgos y la predicción de los mismos, tiene impacto directo en el precio del gas, el petróleo y la energía, los cambios en estos precios pueden afectar la inflación, la demanda del consumidor, la estabilidad de los mercados financieros y la economía en general, lo que los convierte en un factor crítico que los inversores, gobiernos y empresas deben tener en cuenta en sus decisiones financieras y políticas, otros autores como Riedl, D. (2021), relaciona estos riesgos de transición directamente con el avalúo de las empresas en la industria energética; también Ren, X., Li, J., He, F., & Lucey, B. (2023), estudian la relación entre la incertidumbre existente en

políticas climáticas, los mercados de energía tradicionales y los llamados ‘green markets’ o mercados de energías limpias.

Con este contexto, vale la pena recalcar que la IEA (International Energy Agency) ha definido 3 escenarios a la hora de hablar de energía y cambio climático, el primer escenario: STEPS (Stated Policies Scenario) “está diseñado para proporcionar una idea de la dirección predominante de la progresión del sistema energía” (IEA), el escenario STEPS, explora hacia donde podría ir el sistema energético sin la aplicación de políticas adicionales, asumiendo que los gobiernos no alcanzaran todos los objetivos anunciados.

El segundo escenario APS (Announces Pledges Scenario), “ilustra hasta qué punto las ambiciones y objetivos anunciados pueden lograr las reducciones de emisiones necesarias para lograr emisiones netas cero para 2050” (IEA), asume que todos estos objetivos se cumplan totalmente en el tiempo establecido (incluso si no hay políticas establecidas para cumplirlos).

Y el escenario NZE (Net Zero Emissions by 2050 Scenario), “es un escenario normativo que muestra un camino para que el sector energético alcance cero emisiones netas de carbono para el 2050” (IEA), considera un conjunto de medidas políticas hipotéticas que permitan reducir agresivamente las emisiones.

En conclusión, el estudio y modelado de los riesgos de transición representan un paso crucial hacia una transición energética responsable y sostenible. A medida que el mundo se enfrenta a desafíos ambientales y busca reducir su dependencia de los combustibles fósiles, entender y gestionar los riesgos asociados con la transición hacia fuentes de energía más limpias y sostenibles es esencial, este enfoque no solo es relevante para las empresas y la industria energética, sino también para gobiernos, inversores y la sociedad en su conjunto. La transición energética implica una reconfiguración significativa de la infraestructura, la inversión y las políticas, y conlleva tanto oportunidades como desafíos.

Es en este contexto donde el uso de herramientas analíticas desempeña un papel crucial, ya que permiten una evaluación más precisa y cuantitativa de los riesgos, ayudando a las partes interesadas a tomar decisiones informadas y diseñar estrategias efectivas para mitigar los impactos negativos, además, facilitan la identificación de oportunidades económicas y de inversión en la transición hacia una economía baja en carbono.

3 Metodología

Dando un contexto previo, la generación de la sabana de datos, está enmarcada en un proyecto macro, por lo que, al mencionar metodologías, se dará la claridad de la cual se refiere, ya que todos son aplicables a la hora de trabajar en el proyecto puntual de la generación de la sabana.

El enfoque metodológico del equipo de trabajo se estructuró en tres fases fundamentales para abordar eficientemente el proyecto. La primera etapa, denominada "Modelamiento", la segunda fase, denominada "Sensibilización", y finalmente la etapa de "Unión", las cuales serán descritas detalladamente a continuación:

3.1 Primera Fase

El proceso de modelamiento, primera fase del proyecto implicó una meticulosa extracción de datos de los archivos que consideramos insumos. Esta información se llevó hacia una sábana preliminar donde se llevaron a cabo cálculos intermedios destinados a la correcta estructuración y comprensión de los datos. El objetivo fue establecer una base sólida para el análisis siguiente.

3.2 Segunda Fase

En la fase de sensibilización, la atención se centró en la variación de precios de productos clave en respuesta a los diferentes escenarios climáticos previamente identificados en el marco teórico (STEPS, APS, NZE). Esta variabilidad directamente incidió en los rubros vinculados, generando así una línea de cifras específica para cada escenario. Esta aproximación permitió una comprensión más profunda de la influencia de las condiciones climáticas en los resultados financieros del proyecto.

3.3 Tercera Fase

La etapa de unión, última antes de la entrega final, requirió la consolidación de tres conjuntos de datos distintos, correspondientes a los sectores de Upstream, Downstream y Bajas Emisiones. Cada uno de estos segmentos había sido modelado y sensibilizado de manera única para capturar las particularidades de su operación y respuesta a los escenarios climáticos. La fusión de estos conjuntos de datos en un archivo total no solo consolidó la información, sino que también proporcionó una visión integral de la interrelación entre los distintos sectores, brindando así una comprensión holística del impacto del proyecto.

Durante la ejecución integral del proyecto, se implementaron metodologías ágiles, concretamente SCRUM, para gestionar eficazmente las complejidades inherentes a cada fase. Este enfoque ágil se orientó hacia la consecución de cuatro entregables claves para el proyecto macro previamente definido, si bien, por cuestiones de alcance de este proyecto, nos centraremos en los dos primeros. El primer entregable, denominado "Construcción", consistió en los notebooks responsables de la generación de las hojas de datos, asegurando así la representación efectiva de los datos extraídos. El segundo entregable, conocido como "Definición", comprendió

la documentación detallada de los códigos asociados al primer entregable, facilitando una transferencia de conocimientos fluida y completa con el cliente.

Para darle soporte a esta metodología ágil, se empleó la plataforma Azure DevOps. Al concluir cada sprint, se registraron de manera formal los entregables correspondientes a cada área, garantizando la transparencia y trazabilidad en el desarrollo del proyecto. Este proceso aseguró no solo la optimización de los procesos internos, sino también la capacidad de adaptación continua ante ajustes y cambios que surgieron en el curso del proyecto.

En el ámbito específico de la ingeniería de datos, encargada de la generación automatizada de la hoja de datos, nuestra elección recae en la plataforma Databricks. Este servicio, integrado en la nube, se emplea para simplificar y acelerar el procesamiento y análisis de datos a gran escala, estableciendo una conexión directa entre las ramas de desarrollo y el entorno de DevOps.

En cuanto a los lenguajes de programación, el principal es Python, utilizado para manipular datos en forma de DataFrames mediante la librería Pandas. Además, se incorpora de manera reducida Apache Spark para llevar a cabo transformaciones, uniones y exportaciones en DataFrames de mayor envergadura. Esta combinación estratégica de herramientas y tecnologías se alinea con la eficiencia y la escalabilidad requeridas para la manipulación de datos en el contexto del proyecto.

3.4 Cronograma de actividades

A continuación, se presenta el cronograma de actividades del área de ingeniería de Datos por sprint, donde el sprint 1 comienza el 3 de Julio de 2023 y cada sprint consta de dos semanas calendario. Por motivos de confidencialidad no se puede dar un cronograma detallado de todo el proyecto, sin embargo, el área de ingeniería de datos es la encargada de la generación de la sabana de datos.

| Sprint | Épica | Resumen |
|--------|---------------------|--|
| 1 | Ingeniería de datos | Generación de repositorio/git /Azure de datos |
| 1 | Ingeniería de datos | Primera aproximación la lectura y procesamiento de datos |
| 2 | Ingeniería de datos | Creacion de containers en Azure Data Lake y Carga manual de datos |
| 2 | Ingeniería de datos | Configuración de Databricks. |
| 2 | Ingeniería de datos | Definición de componentes de la sabana de datos. |
| 3 | Ingeniería de datos | Lectura datos |
| 3 | Ingeniería de datos | Creación de sabana de datos automática |
| 4 | Ingeniería de datos | Generar la sabana de datos para riesgos de transición del upstream |
| 4 | Ingeniería de datos | Entendimiento sábana de datos ESM e identificación de sinergias |
| 5 | Ingeniería de datos | Refinación sábana de datos automática |
| 5 | Ingeniería de datos | Creación de flujos de caja por proyecto (Upstream, CCUS y Bajas Emisiones) |
| 6 | Ingeniería de datos | Mejorar la Sabana de datos automática por escenario climático para el primer MVP |
| 6 | Ingeniería de datos | Generar una nueva versión optimizada de la sabana de datos. |
| 7 | Ingeniería de datos | Mejora en la Automatización de Procesos de la sabana de datos del Upstream y Downstream |
| 7 | Ingeniería de datos | Mejora de la sabana de datos del Upstream y Downstream e incorporación de nuevos Flujos de Caja. (Pruebas de Calidad de datos) |

| | | |
|----|---------------------|--|
| 7 | Ingeniería de datos | Optimización del Mapeo de Fuentes de Información, Documentación y Diagramación |
| 8 | Ingeniería de datos | Refinación de la sabana de datos del Upstream y Downstream |
| 8 | Ingeniería de datos | Creación de tablas basadas en la sabana de datos |
| 8 | Ingeniería de datos | Flujos de Caja y Actualización de Datos para Modelación de Riesgos de transición |
| 9 | Ingeniería de datos | Optimización Sabana Y aseguramiento de datos del MVP2 |
| 9 | Ingeniería de datos | Mejoras y ajustes del modelo de riesgos físicos |
| 9 | Ingeniería de datos | Estructuración de los modelos y mejora en el job de databricks |
| 10 | Ingeniería de datos | Garantizar la creación y avance de la documentación de riesgos físicos y riesgos de transición |
| 10 | Ingeniería de datos | Mejora de los Job modelo Clima, con integración de Power BI |
| 10 | Ingeniería de datos | Optimización de Sabana y Aseguramiento de Datos |
| 11 | Ingeniería de datos | Integración Eficiente Power BI y Databricks |
| 11 | Ingeniería de datos | Avance en Documentación de Riesgos Físicos y de Transición |
| 11 | Ingeniería de datos | Optimización y Aseguramiento de Datos |

Tabla 1. Sprint planning de Ingeniería de Datos

4 Resultados y Análisis

Antes de empezar, vale la pena aclarar que los resultados serán de visualización limitada y se mostrarán snippets² de código de referencia, ya que, por acuerdos de confidencialidad con el cliente, no se puede enseñar el trabajo realizado tal cual.

El entendimiento del negocio es esencial, ya que este proporciona la base para la identificación precisa de los problemas y objetivos comerciales clave que el proyecto busca abordar, por lo que las primeras semanas de trabajo el cliente se encargó de brindar al equipo de trabajo una contextualización y capacitación sobre la fuente de datos y el negocio, y por parte del equipo de trabajo ACT, se asignaron roles, se definió la arquitectura del proyecto (Ilustración 2) y se abordaron las fuentes de datos suministradas por el cliente (Ilustración 1).

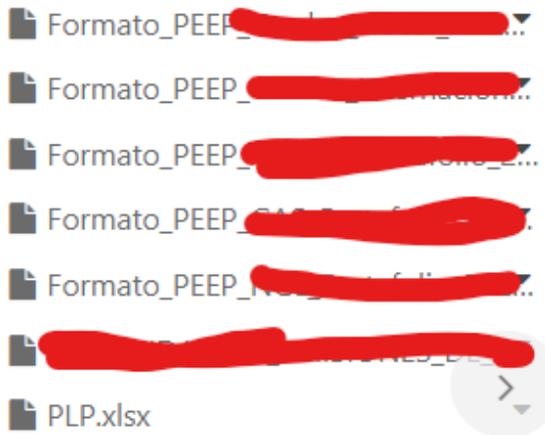


Ilustración 1. Datos suministrados por el cliente contienen planeaciones a largo plazo y planeación y evaluación efectiva de proyectos.

² En programación, un Snippet quiere decir un «fragmento de código» que puedes reutilizar en cualquier lugar que lo necesites y que suele solventar un problema en específico.

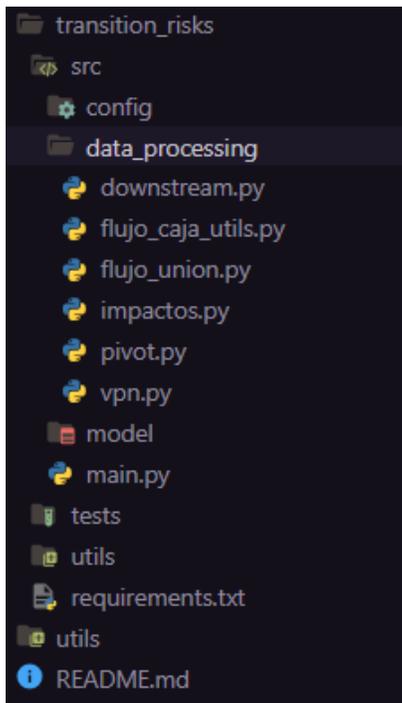


Ilustración 2. Arquitectura del proyecto

Una vez contextualizado el equipo se paso a definir el alcance del proyecto donde se identifico que la necesidad y el trabajo a realizar era similar a los procesos de ETL³, donde como se muestra en la ilustración 3 se estableció una forma de abordar cada uno de los procesos de la ETL.

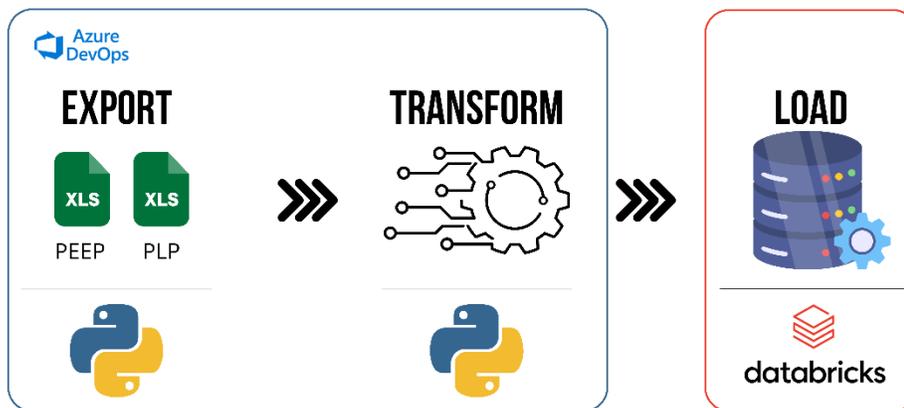


Ilustración 3. Tecnologías y plataformas utilizadas para abordar el proceso de ETL.

³ Extract, Transform and Load es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos.

Una vez definida la ruta de trabajo, se abordaron los datos, las librerías utilizadas para este trabajo fueron:

| <i>Librería</i> | <i>Funcionalidad</i> |
|-----------------|--|
| Pandas | Manipulación y análisis de datos |
| Spark | Framework utilizado para el análisis y manipulación de grandes volúmenes de datos. |
| Numpy | Crear y manipular vectores y matrices multidimensionales. |
| Unidecode | Estandarizar los tipos de codificación de los diferentes tipos de datos. |
| string | Realizar operaciones comunes con cadenas de texto. |

Tabla 2. Librerías utilizadas en el proyecto y su funcionalidad.

4.1 Extracción

Al empezar el proceso de extracción, se identificó un cambio de formato en los datos por lo que se decidió crear un paso intermedio en el que se almaceno la data de los archivos de Excel en formato JSON⁴ bajo el nombre “*nombre_json_config*” para su posterior procesamiento como dataframes. A continuación, se muestran algunos snippets de código de los procesos utilizados para la obtención de los datos.

```
rango_anios = self.nombre_json_config["parametros"]["rango_anio"]
precio_rubro = self.nombre_json_config["parametros"]["precio_rubro"]
precio_rubro_aps = self.nombre_json_config["parametros"]["precio_rubro_aps"]
precio_rubro_nze = self.nombre_json_config["parametros"]["precio_rubro_nze"]
precio_rubro_steps = self.nombre_json_config["parametros"]["precio_rubro_steps"]
```

Ilustración 4. Snippet que extrae la información del archivo JSON y los guarda en una lista de Python.

⁴ JavaScript Object Notation, es un formato ligero de manejo de datos.

```
data={"RANGO_ANIO" : años ,  
      "PRECIO_RUBRO" : precio_rubro,  
      "PRECIO_RUBRO_APS" : precio_rubro_aps,  
      "PRECIO_RUBRO_NZE" : precio_rubro_nze,  
      "PRECIO_RUBRO_STEPS" : precio_rubro_steps}  
  
df=pd.DataFrame( data)
```

Ilustración 5. Snippet que convierte las listas obtenidas con el fragmento de la Ilustración 4 en columnas de un pandas DataFrame

Con esto se obtiene una sabana preliminar que contiene toda la información obtenida de las fuentes de datos, con esto se inicia el proceso de transformación.

4.2 Transformación

El proceso de transformación consiste principalmente en la sensibilización de algunos rubros por cada uno de los escenarios climáticos ya mencionados; sin embargo, teniendo en cuenta el contexto del negocio, cada uno de los rubros debe ser sensibilizado de una manera diferente, encontrándonos en un escenario donde no podemos definir funciones replicables para todos los rubros, si no que debemos modelar cada rubro con sus características peculiares de acuerdo con los requerimientos del negocio, como lo muestra la ilustración 6.

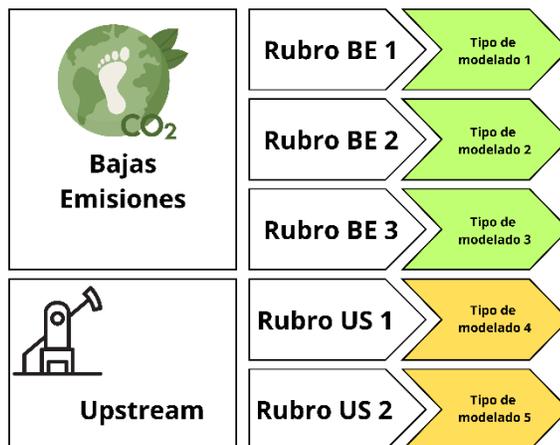


Ilustración 6. Rubros correspondientes a cada segmento y los tipos de modelación requeridos para cubrir el alcance del proyecto.

Debido a los acuerdos de confidencialidad antes mencionados solo hay posibilidad de ahondar en 2 de los 5 tipos de modelado, el primero tiene lugar cuando la sensibilización por escenario viene directamente del formato de PLP, en este caso, se aplica un proceso similar al mostrado en las ilustraciones 4 y 5, el segundo tipo de sensibilización se muestra en la ilustración 7, consiste en, después de la inicialización de las columnas de los escenario climáticos en 0 y con unas variaciones porcentuales guardadas en el diccionario *dict_esc_clima*, se varia cada rubro por año, siguiendo esta lógica: el primer año de todos los rubros es igual al del escenario base (traído directamente del PLP), a partir del segundo año y hasta el año final, se aplica la variación porcentual sobre el año inmediatamente anterior del rubro, en el escenario climático correspondiente.

```
dict_esc_clima = {2024:-5, 2025:-6, 2026:-7} #%
df_sabana_datos["PRECIO_RUBRO_ESC.CLIMA"] = 0
for i in range(len(df_sabana_datos)):
    año_actual = df_sabana_datos.loc[i, 'RANGO_ANIO']
    if año_actual == 2024:
        df_sabana_datos.loc[i, 'PRECIO_RUBRO_ESC.CLIMA'] = df_sabana_datos.loc[i, 'PRECIO_RUBRO'] +
            (dict_esc_clima[año_actual]/100 * df_sabana_datos.loc[i, 'PRECIO_RUBRO'])
    else:
        precio_anterior = df_sabana_datos.loc[i - 1, 'PRECIO_RUBRO_ESC.CLIMA']
        df_sabana_datos.loc[i, 'PRECIO_RUBRO_ESC.CLIMA'] = precio_anterior +
            (precio_anterior * dict_esc_clima[año_actual]/100)
```

Ilustración 7. Snippet que representa la lógica de uno de los tipos de modelado.

4.3 Carga

Una vez creadas las dos sabanas con las lógicas de modelado respectivas, se deben exportar los dataframes de pandas a archivos que puedan ser consumibles por el prototipo de analítica, en un inicio se propusieron archivos de Excel, sin embargo, debido al volumen de los datos y que el resultado de este proyecto es solo una etapa media del proyecto final, se tomó la decisión de exportar los dataframes como archivos CSV para reducir los tiempos de exportación. Estos archivos CSV se guardan en el sistema de archivos de Databricks (DBFS), para ser utilizados mas adelante en otro proyecto.



Ilustración 8. Las sabanas de datos resultantes alojadas en DBFS.

Por motivos de confidencialidad, estos archivos csv no se pueden enseñar, sin embargo, se maquetó una sabana de datos con datos dummy⁵ para mostrar su estructura por columnas; que puede ser encontrada en el anexo 1.

⁵ Los datos dummy son datos falsos que puedes utilizar para rellenar tablas y poder realizar tus pruebas.

5 Conclusiones

Al finalizar el presente proyecto, se identifican las siguientes conclusiones:

El entendimiento del negocio es esencial para abordar cualquier proyecto relacionado a datos, ya que expresan una realidad o un histórico de la compañía y esto se debe tener en cuenta a la hora de la manipulación o modificación de estos datos para asegurar que estén alineados al contexto del negocio.

La automatización de la sabana se considera exitosa, al ser validada por parte del cliente, los archivos de Python (.py) se encuentran en el ambiente de Azure DevOps suministrado por el cliente con su respectiva documentación.

Referencia

- Colenbrander, S., Vaze, P., Vikas, C., Ayer, S., Kumar, N., Vikas, N., & Burge, L. (2023). Low-carbon transition risks for India's financial system. *Global Environmental Change*, 78. <https://doi.org/10.1016/j.gloenvcha.2022.102634>
- Global Energy and Climate Model - Analysis - IEA. (s.f). IEA. <https://www.iea.org/reports/global-energy-and-climate-model>
- Green, J., Hadden, J., Hale, T., & Mahdavi, P. (2022). Transition, hedge, or resist? Understanding political and economic behavior toward decarbonization in the oil and gas industry. *Review of International Political Economy*, 29(6). <https://doi.org/10.1080/09692290.2021.19467081>
- Jawadi, F., Rozin, P., & Bourghelle, D. (2023). Insights into CO2 emissions in Europe in the context of COVID-19: A panel data analysis. *International Economics*, 173. <https://doi.org/10.1016/j.inteco.2022.11.006>
- Ren, X., Li, J., He, F., & Lucey, B. (2023). Impact of climate policy uncertainty on traditional energy and green markets: Evidence from time-varying granger tests. *Renewable and Sustainable Energy Reviews*, 173. <https://doi.org/10.1016/j.rser.2022.113058>
- Riedl, D. (2021). The magnitude of energy transition risk embedded in fossil fuel company valuations. *Heliyon*, 7(11). <https://doi.org/10.1016/j.heliyon.2021.e08400>
- Salisu, A. A., Ndako, U. B., & Vo, X. V. (2023). Transition risk, physical risk, and the realized volatility of oil and natural gas prices. *Resources Policy*, 81. <https://doi.org/10.1016/j.resourpol.2023.103383>
- Sobre Nuestra Empresa. (s/f). Accenture. <https://www.accenture.com/co-es/about/company-index>

Anexos

Anexo 1. Sabana de datos dummy.

<https://bit.ly/sabana-datos>