**### \*\*Resumen del Proyecto\*\***

El proyecto se enfoca en analizar datos de transporte, calidad del aire y contaminación sonora en Nueva York para evaluar la viabilidad de implementar vehículos eléctricos en la flota de una empresa de transporte. El análisis utiliza múltiples datasets, incluyendo información de viajes en taxis, datos climáticos, niveles de ruido, y emisiones de CO2, con el objetivo de tomar decisiones informadas sobre la adopción de vehículos eléctricos.

**### \*\*Lista de Trabajo y Tecnologías a Utilizar\*\***

1**. \*\*Recopilación y Depuración de Datos\*\***

**- \*\*Tareas\*\*:**

- Descargar y unificar datasets de diferentes fuentes (NYC TLC, APIs de clima, datasets de Kaggle, etc.).

- Limpiar y preparar los datos para análisis (tratar valores nulos, outliers, y formato de datos).

- \*\*Tecnologías\*\*:

- \*\*Python\*\*: Pandas para manipulación de datos, Requests para llamadas a APIs.

- \*\*SQLite/PostgreSQL\*\*: Para almacenar los datos de forma estructurada y permitir consultas eficientes.

2. **\*\*Construcción del Data Warehouse\*\***

- **\*\*Tareas\*\*:**

- Crear un Data Warehouse local utilizando una base de datos SQL.

- Integrar y normalizar los datos de diferentes fuentes.

- \*\*Tecnologías\*\*:

- \*\*SQLite/PostgreSQL\*\*: Para alojar el Data Warehouse localmente.

- \*\*SQLAlchemy\*\*: Para interactuar con la base de datos desde Python.

3. **\*\*Análisis Exploratorio de Datos (EDA)\*\***

**- \*\*Tareas\*\*:**

- Realizar análisis descriptivo de los datos para entender las relaciones entre variables.

- Visualizar datos geoespaciales y temporales para identificar patrones de viajes y contaminación.

- \*\*Tecnologías\*\*:

- \*\*Python\*\*: Matplotlib y Seaborn para visualización, GeoPandas y Folium para análisis geoespacial.

- \*\*Jupyter Notebook\*\*: Para desarrollar y documentar el EDA.

4. **\*\*Modelado Predictivo\*\***

**- \*\*Tareas\*\*:**

- Entrenar modelos de machine learning para predecir la demanda de taxis, emisiones de CO2, y niveles de ruido.

- Evaluar y ajustar los modelos para mejorar su precisión.

- \*\*Tecnologías\*\*:

- \*\*Python\*\*: Scikit-learn para machine learning, TensorFlow o PyTorch para modelos más complejos.

5. **\*\*Integración de Datos Climatológicos y Calidad del Aire\*\***

**- \*\*Tareas\*\*:**

- Integrar datos de calidad del aire y clima en el análisis para correlacionarlos con los patrones de viaje y emisiones.

- Utilizar datos en tiempo real para modelos predictivos y análisis de sensibilidad.

- \*\*Tecnologías\*\*:

- \*\*Python\*\*: Requests para obtener datos de APIs, Pandas para manipulación.

- \*\*SQLite/PostgreSQL\*\*: Para almacenar y consultar datos climatológicos.

6. **\*\*Análisis de Impacto y Estrategias de Implementación\*\***

**- \*\*Tareas\*\*:**

- Analizar el impacto de la introducción de vehículos eléctricos en la reducción de emisiones y ruido.

- Proponer estrategias de implementación y marketing basadas en microsegmentación.

- \*\*Tecnologías\*\*:

- \*\*Python\*\*: Para simulaciones y análisis estadístico.

- \*\*Jupyter Notebook\*\*: Para documentar los hallazgos y generar reportes.

7. **\*\*Documentación y Presentación de Resultados\*\***

**- \*\*Tareas\*\*:**

- Documentar todo el proceso y resultados en Jupyter Notebook.

- Crear un dashboard interactivo en Streamlit para visualizar los datos y resultados.

- \*\*Tecnologías\*\*:

- \*\*Jupyter Notebook\*\*: Para la creación de informes detallados.

- \*\*Streamlit\*\*: Para crear un dashboard interactivo y funcional.

**### \*\*Resumen de Tecnologías\*\***

- **\*\*Python\*\*:** Pandas, GeoPandas, Matplotlib, Seaborn, Scikit-learn, TensorFlow/PyTorch.

- **\*\*Bases de Datos\*\*:** SQLite/PostgreSQL.

- **\*\*Visualización Geoespacial\*\*:** Folium, GeoPandas.

- **\*\*Documentación\*\*:** Jupyter Notebook, Streamlit.

**### \*\*Print #1: Puesta en Marcha del Proyecto y Trabajo con Datos\*\***

**#### \*\*1. Entendimiento de la Situación Actual\*\***

**\*\*Contextualización\*\*:**

La empresa de transporte de pasajeros busca expandirse al sector de vehículos de pasajeros, con un enfoque en la adopción de vehículos eléctricos para reducir la contaminación y alinearse con las tendencias de sostenibilidad. La decisión requiere un análisis exhaustivo de datos de taxis y viajes compartidos en Nueva York, cruzándolos con datos de calidad del aire y contaminación sonora para evaluar la viabilidad y el impacto ambiental.

**\*\*Propuesta de Análisis/Soluciones\*\*:**

- \*\*Análisis de patrones de viaje\*\*: Evaluar la distribución espacial y temporal de los viajes en taxis.

- \*\*Impacto ambiental\*\*: Correlacionar los patrones de viaje con los niveles de contaminación del aire y ruido.

- \*\*Evaluación económica\*\*: Comparar los costos y beneficios de la adopción de vehículos eléctricos frente a los de combustión interna.

**#### \*\*2. Objetivos\*\***

- \*\*Desarrollar\*\* un Data Warehouse para centralizar y unificar los datos de transporte, calidad del aire y ruido.

- \*\*Realizar\*\* un análisis exploratorio de datos (EDA) para identificar patrones y correlaciones clave.

- \*\*Entrenar\*\* un modelo de machine learning para predecir la demanda de taxis y el impacto ambiental de diferentes tipos de vehículos.

- \*\*Evaluar\*\* la viabilidad económica y ambiental de la adopción de vehículos eléctricos.

**#### \*\*3. Alcance\*\***

**- \*\*Dentro del Alcance\*\*:**

- Análisis de datos de viajes en taxis y su relación con la calidad del aire y ruido.

- Desarrollo de un modelo predictivo para la demanda de taxis.

- Evaluación preliminar del impacto de vehículos eléctricos.

- \*\*Fuera del Alcance (Posibles continuidades)\*\*:

- Implementación de estrategias de marketing basadas en microsegmentación.

- Análisis exhaustivo de la depreciación de vehículos eléctricos vs. combustión interna.

**#### \*\*4. Objetivos y KPIs Asociados\*\***

- \*\*Objetivo 1\*\*: Reducir las emisiones de CO2 en un 15% mediante la adopción de vehículos eléctricos.

- \*\*KPI\*\*: Reducción porcentual de emisiones de CO2 en simulaciones basadas en datos históricos.

- \*\*Objetivo 2\*\*: Mejorar la calidad del aire en las zonas más transitadas de NYC.

- \*\*KPI\*\*: Reducción del índice de contaminación del aire en las zonas analizadas.

- \*\*Objetivo 3\*\*: Aumentar la eficiencia operativa mediante la optimización de rutas y tiempos de espera.

- \*\*KPI\*\*: Disminución del tiempo promedio de espera de los pasajeros.

**#### \*\*5. Repositorio en Github\*\***

- \*\*Repositorio Público\*\*: Crear un repositorio en GitHub para gestionar el código, la documentación y las versiones del proyecto.

- \*\*Estructura del Repositorio\*\*:

- \*\*Branches\*\*: Establecer ramas para cada miembro del equipo y cada funcionalidad principal.

- \*\*Colaboración\*\*: Utilizar pull requests para la revisión y aprobación de código.

**#### \*\*6. Solución Propuesta\*\***

- \*\*Metodologías de Trabajo\*\*:

- \*\*Scrum\*\*: Dividir el trabajo en sprints semanales con revisiones regulares.

- \*\*Roles\*\*:

- \*\*Data Engineer\*\*: Encargado de la integración y limpieza de datos.

- \*\*Data Scientist\*\*: Responsable del análisis y modelado predictivo.

- \*\*Data Analyst\*\*: Enfocado en la visualización y EDA.

- \*\*Project Manager\*\*: Gestión del proyecto y coordinación del equipo.

- \*\*Productos y Etapas\*\*:

- \*\*Semana 1\*\*: EDA inicial y creación del Data Warehouse.

- \*\*Semana 2\*\*: Desarrollo del modelo predictivo y análisis de impacto.

- \*\*Semana 3\*\*: Evaluación económica y generación de reportes.

- **\*\*Estimación de Tiempo\*\*:**

- Creación del Data Warehouse: 5 días.

- EDA: 3 días.

- Modelado Predictivo: 4 días.

- Evaluación y Documentación: 3 días.

- **\*\*Stack Tecnológico\*\*:**

- \*\*Base de Datos\*\*: SQLite/PostgreSQL para el almacenamiento y gestión de datos.

- \*\*Limpieza y Transformación\*\*: Python (Pandas, NumPy).

- \*\*Visualización\*\*: Matplotlib, Seaborn, Folium, GeoPandas.

- \*\*Modelado\*\*: Scikit-learn, TensorFlow.

- \*\*Documentación\*\*: Jupyter Notebook, Streamlit.

- \*\*Análisis Preliminar de Datos\*\*:

- \*\*Calidad de los Datos\*\*: Evaluar la precisión, completitud y consistencia de los datasets.

- \*\*Metadatos\*\*: Documentar la procedencia, tipo de datos, y última actualización de cada dataset utilizado.

**#### \*\*7. Hitos\*\***

- \*\*KPI’s\*\*: Definir 3 KPI’s clave para el proyecto.

- \*\*Alcance\*\*: Documentar el alcance del proyecto.

- \*\*EDA\*\*: Completar el análisis exploratorio de datos.

- \*\*Repositorio en Github\*\*: Implementar el repositorio y comenzar el versionado.

- \*\*Implementación del Stack Tecnológico\*\*: Configurar el entorno de trabajo.

- \*\*Metodología de Trabajo\*\*: Establecer la metodología de trabajo.

- \*\*Diseño Detallado\*\*: Completar el diseño de la solución.

- \*\*Roles y Responsabilidades\*\*: Definir claramente los roles del equipo.

- \*\*Cronograma General (Gantt)\*\*: Crear y seguir el cronograma de trabajo.

- \*\*Análisis de Calidad de Datos\*\*: Realizar un análisis preliminar de los datos.

- \*\*Entregables\*\*: Documentación, stack elegido, flujo de trabajo.