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

**Análisis del Proyecto**

Nueva York, conocida por su frenética actividad vial, se enfrenta a desafíos significativos en cuanto a congestión y contaminación. En respuesta, nuestro proyecto se centra en el sector del transporte de pasajeros, proponiendo soluciones innovadoras para mejorar la eficiencia, reducir costos y promover la sostenibilidad ambiental.

**Objetivos**

***1. Recopilar, depurar y organizar datos relevantes:***

- Recopilación exhaustiva de datos sobre el transporte de pasajeros en Nueva York.

- Depuración y limpieza de datos para garantizar su calidad y fiabilidad.

- Organización estructurada de la información para facilitar su análisis posterior.

***2. Desarrollar un Dashboard Interactivo para una mejor visualización y comprensión de los datos:***

- Creación de un Dashboard Interactivo que permita a los interesados visualizar análisis de valor basados en KPI clave.

- Diseño de una interfaz intuitiva y fácil de usar para facilitar la comprensión de los datos.

- Integración de herramientas de visualización avanzada para representar la información de manera efectiva.

***3. Evaluar y comparar los costos y beneficios de los vehículos de combustión y eléctricos:***

- Análisis detallado de los costos asociados a los vehículos de combustión y eléctricos.

- Comparación de los beneficios económicos y ambientales de ambos tipos de vehículos.

- Identificación de oportunidades para promover la adopción de vehículos eléctricos en el sector del transporte de pasajeros.

**Alcance**

Este primer sprint implica un análisis exhaustivo del transporte de pasajeros en Nueva York, utilizando diversas fuentes de datos y herramientas de visualización interactiva. Además, se aplicará un modelo de Machine Learning para identificar patrones de uso y optimizar las rutas para reducir el impacto ambiental y mejorar la eficiencia.

**KPIs Propuestos**

- Aumentar las ventas promedio en un 5% respecto al mes anterior.

- Incrementar los ingresos en un 10% respecto al semestre anterior.

- Reducir las emisiones de CO2 en un 40% anual en el servicio de taxis.

- Disminuir el gasto de combustible en un 10% anual.

**Stack Tecnológico**

Las herramientas seleccionadas incluyen:

- Python: Para análisis de datos y Machine Learning.

- Pandas, Numpy, Matplotlib: Manipulación y visualización de datos.

- Seaborn, Folium: Visualización de datos geoespaciales.

- Trello: Plataforma de gestión de tareas.

- PowerBI: Creación de Dashboards interactivos.

- MySQL Server: Gestión de bases de datos.

- Google Cloud: Almacenamiento y procesamiento de datos.

**Metodología de Trabajo**

Adoptaremos una metodología ágil tipo Scrum, utilizando Trello como herramienta central para la organización y seguimiento de tareas. Se llevarán a cabo reuniones diarias para revisar el progreso, identificar posibles obstáculos y ajustar el plan de trabajo según sea necesario.

**Diseño Detallado**

(PENDIENTE DE TERMINAR)

**Equipo de Trabajo - Roles y Responsabilidades**

**Analistas de Datos:** *Joyce Ruiz, Marco Delgado y Jovanni Escamilla.*

**Data Engineers:** *Joyce Ruiz, Jovanni Escamilla, Lucas Salzotto, Matias Barna.*

**Especialistas en Machine Learning:** *Matias Barna, Marco Delgado, Lucas Salzotto.*

Cronograma General de Actividades - Gantt

(PENDIENTE AGREGAR DIAGRAMA)

**Solución propuesta**

Deben detallar qué tareas harán para cumplir los objetivos de trabajo propuestos previamente y cómo lo harán (metodologías de trabajo, forma de organización, distribución de tareas, roles de cada uno dentro del equipo, etc). También, deben detallar qué productos surgirán de su trabajo y en qué etapa los presentarán, teniendo en cuenta los requerimientos generales (entregables esperados) para cada etapa del proyecto.

A su vez, deben realizar una estimación de tiempo para cada tarea, contemplando los tiempos de ejecución globales y los hitos previstos para cada semana; y plasmar esa estimación en un diagrama de Gantt.

Una parte muy importante de la solución propuesta, es con qué herramientas (stack tecnológico) van a realizar la arquitectura del proyecto. Para esto, lo que van a tener que hacer es seleccionar una pequeña porción de los datos que disponen y realizar un proceso de limpieza y transformación utilizando las herramientas que planean implementar. Esto les dará una idea de cómo funcionarán en el proyecto completo y les permitirá tener un mejor abordaje para futuras tareas. Hay que tener en cuenta que, como este ítem va a ser una presentación previa de lo que van a trabajar en el segundo sprint, el PO puede dar el OK o determinar cuál es el mejor camino para que tomen. Esto les va a permitir adelantar trabajo de la segunda semana, ya que no se va a tener que esperar hasta la segunda demo para verificar si la arquitectura cumple con los requisitos del PO.

Finalmente, como en Data es muy importante trabajar con datos de calidad, deberán incluir en su informe un análisis sobre los datos con los que van a trabajar (metadatos), detallándolos lo más posible: fuentes y confiabilidad de las mismas, qué representa cada columna de cada dataset, tipos de datos, método de adquisición, fecha de adquisición y última actualización, etc.