## DOCUMENTACIÓN DE LA PROPUESTA DEL PROYECTO

# **"EcoRide NYC: Transformando la Movilidad Urbana en Nueva York"**

## **INTRODUCCIÓN**

En el dinámico escenario de la ciudad de Nueva York, donde la movilidad urbana se transforma incesantemente, nace "**EcoRide NYC: Transformando la Movilidad Urbana en Nueva York**". En respuesta al cambio impulsado por servicios de taxis y viajes compartidos, una empresa de transporte de pasajeros se embarca en una iniciativa audaz. Con la visión de un mañana más limpio y alineado a las tendencias del mercado, la empresa explora la conexión entre el transporte en automóviles privados y la calidad del aire, evaluando la viabilidad de integrar vehículos eléctricos.

Nuestro equipo asume la misión de analizar minuciosamente los datos generados por estos servicios en Nueva York. Este análisis no solo proporcionará un marco esencial para la toma de decisiones, sino que también pavimentará el camino hacia un transporte más verde y eficiente. **EcoRide NYC** encarna no solo un proyecto, sino un compromiso con la transformación sostenible, utilizando la ciencia de datos para guiar decisiones estratégicas en el complejo entorno del transporte urbano. A través de este análisis, exploramos patrones y conexiones para allanar el camino hacia un futuro más sostenible en el corazón de la Gran Manzana. ¡Bienvenidos a **EcoRide NYC**, donde conducimos hacia un futuro más verde y sostenible del transporte!

## **ENTENDIMIENTO DE LA SITUACIÓN ACTUAL**

Una empresa de micros de media y larga distancia planea expandir hacia el transporte de pasajeros en automóviles, explorando vehículos eléctricos para alinearse con tendencias de mercado y reducir la contaminación. Buscan analizar el movimiento de taxis en Nueva York para tomar decisiones informadas.

**Transformación del Transporte en Nueva York**:

Los taxis y viajes compartidos en Nueva York, alternativas populares al transporte público, ofrecen datos valiosos sobre la ubicación del vehículo y la calificación del conductor, reflejando cambios en los patrones de viaje postpandemia y la flexibilidad laboral.

**Motivación de la Empresa de Transporte de Pasajeros**:

La empresa busca adaptarse a un enfoque más sostenible y reducir la huella ambiental, evaluando la incorporación de vehículos eléctricos y su impacto en la calidad del aire y la contaminación sonora.

**Necesidad de Análisis Preliminar**:

Se realizarán un análisis preliminar del tráfico de taxis en Nueva York para guiar la expansión hacia el transporte en automóviles, utilizando análisis de datos para extraer insights y apoyar decisiones estratégicas.

**Desafíos y Oportunidades**:

El proyecto enfrenta retos en el análisis de datos complejos y diversos, pero también presenta oportunidades para aplicar tecnologías avanzadas y alinear la estrategia empresarial con la sostenibilidad y demandas del mercado.

## **OBJETIVO GENERAL**

Desarrollar un análisis integral de datos y aplicar técnicas de Machine Learning para evaluar la viabilidad y el impacto de la incorporación de vehículos eléctricos en la flota de una empresa de transporte de pasajeros en la ciudad de Nueva York, orientado a mejorar la sostenibilidad y eficiencia del transporte urbano.

## **OBJETIVOS ESPECÍFICOS**

1. **Desarrollar un análisis detallado del mercado de taxis y viajes compartidos en la ciudad de Nueva York**.

Realizar un estudio exhaustivo sobre el mercado actual de taxis y viajes compartidos en Nueva York, enfocándose en identificar tendencias, patrones de demanda, preferencias de los clientes, y la estructura competitiva. Este análisis ayudará a comprender mejor el entorno del mercado y a identificar oportunidades y desafíos.

1. **Realizar un Análisis de Datos para Evaluar el Impacto Ambiental y Económico de Introducir Vehículos Eléctricos:**

Emplear técnicas de análisis de datos para evaluar de manera exhaustiva el impacto de integrar vehículos eléctricos en la flota existente. Este análisis debería incluir la comparación de emisiones de CO2, consumo energético, y costos operativos entre vehículos eléctricos y convencionales. Además, se puede usar machine learning para modelar escenarios futuros basados en diferentes niveles de adopción de vehículos eléctricos y predecir los efectos a largo plazo en términos de sostenibilidad ambiental y rentabilidad económica**.**

1. **Desarrollar Modelos Predictivos de Machine Learning para Optimizar la Operación y Demanda del Servicio de Taxis**:

Utilizar técnicas de machine learning para analizar grandes conjuntos de datos relacionados con el tráfico, las tendencias de viaje y las preferencias de los usuarios. El objetivo sería desarrollar modelos predictivos que puedan anticipar patrones de demanda, optimizar rutas y horarios, y mejorar la asignación de recursos. Estos modelos ayudarían a la empresa a tomar decisiones basadas en datos para mejorar la eficiencia operativa y la experiencia del cliente.

## **ALCANCE**

El proyecto se enfocará en un análisis integral de datos y la aplicación de técnicas de Machine Learning para evaluar la incorporación de vehículos eléctricos en la flota de una empresa de transporte de pasajeros en Nueva York, buscando mejorar la sostenibilidad y eficiencia del transporte urbano.

**Periodo de Análisis**:

El análisis se centrará en los datos recopilados de taxis amarillos y verdes en la ciudad de Nueva York durante un año, de septiembre de 2022 a septiembre de 2023.

**Elementos Clave del Análisis**:

1. **Demanda de Taxis:**

- Análisis de patrones de demanda de taxis a lo largo del tiempo.

- Evaluación de factores que influyen en la demanda, incluyendo hora del día, día de la semana, condiciones climáticas y eventos especiales.

2. **Impacto Ambiental de los Taxis**:

- Estudio de las emisiones de gases de efecto invernadero y el ruido generado por los taxis.

- Análisis de la relación entre la demanda de taxis y su impacto ambiental.

3. **Evaluación de la Viabilidad de Vehículos Eléctricos:**

- Análisis de costos y beneficios de los vehículos eléctricos comparados con los convencionales.

- Evaluación de la disponibilidad y requerimientos de la infraestructura de carga.

4. **Desarrollo de Modelos de Machine Learning**:

- Creación de un modelo predictivo para la demanda de taxis, optimizando la asignación de recursos y la eficiencia operativa.

- Desarrollo de un modelo de recomendación para viajes más limpios y eficientes, alineado con la sostenibilidad y preferencias de los clientes.

**Limitaciones y Delimitaciones:**

1. El proyecto se limitará al análisis de datos disponibles de taxis amarillos y verdes, así como a la información sobre vehículos eléctricos y su infraestructura correspondiente en la ciudad de Nueva York.
2. Se considerarán únicamente los datos y tendencias dentro del periodo especificado, reconociendo que cambios en políticas o tendencias del mercado pueden influir en futuras proyecciones.

## **OBJETIVOS Y KPIS ASOCIADOS**

1. **Tasa de Cambio en la Demanda de Taxis:**

**Objetivo**: Medir el cambio porcentual en la demanda de taxis mes a mes.

**Fórmula**: *(DemandaActual − DemandaAnterior) / DemandaAnterior × 100*

**Meta**: Lograr un crecimiento mensual constante en la demanda de taxis de al menos un 5%, indicando una expansión positiva del mercado.

1. **Reducción Porcentual de Emisiones de CO2:**

**Objetivo:** Calcular la reducción potencial de CO2 al implementar vehículos eléctricos.

**Fórmula:** *(EmisionesCO2vehiculoConvencional − EmisionesCO2vehiculoElectrico) / EmisionesCO2vehiculoConvencional × 100*

**Meta**: Alcanzar una reducción del 30% anual en las emisiones de CO2 por kilómetro con la introducción de vehículos eléctricos, en comparación con los vehículos convencionales.

1. **Costo Promedio por Kilómetro de Vehículos Eléctricos vs. Convencionales:**

**Objetivo**: Comparar los costos operativos de vehículos eléctricos y convencionales.

**Fórmula**: *Costo total de operación (mantenimiento, energía/combustible) / Kilómetros recorridos*

**Meta***:* Reducir el costo operativo por kilómetro de vehículos eléctricos en un 20% anual en comparación con los vehículos convencionales.

1. **Índice de Disponibilidad de Infraestructura de Carga:**

**Objetivo**: Medir la disponibilidad de infraestructura de carga en relación con el número de vehículos eléctricos en la flota.

**Fórmula**: *Número de estaciones de carga accesibles / Número total de vehículos eléctricos en la flota*

**Meta**: Alcanzar una proporción de al menos una estación de carga por cada cinco vehículos eléctricos en la flota, asegurando una infraestructura adecuada para las necesidades operativas.

1. **Porcentaje de Crecimiento en la Base de Usuarios de Servicios de Taxi:**

**Objetivo**: Medir el crecimiento en el número de usuarios de servicios de taxi.

**Fórmula**: (*NumerodeUsuariosalfinaldelPeriodo* − *NumerodeUsuariosaliniciodelPeriodo*) / *NumerodeUsuariosaliniciodelPeriodo*

**Meta**: Lograr un incremento del 10% en la base de usuarios de servicios de taxi al final del período de análisis, reflejando una exitosa captación y retención de clientes.

## **SOLUCIÓN PROPUESTA**

La solución propuesta para el presente proyecto tiene un enfoque multifacético que integre el análisis de datos y las técnicas de Machine Learning para facilitar la transición hacia una flota de transporte más sostenible y eficiente en la ciudad de Nueva York. Basándonos en los objetivos específicos, la solución podría estructurarse de la siguiente manera:

1. **METODOLOGÍA DE TRABAJO**

**Fase 1: Recopilación de datos**

En esta fase, se recopilarán los datos necesarios para el análisis. Los datos se obtendrán de diversas fuentes, incluyendo:

Datos de servicios de taxis y viajes compartidos, como la ubicación, la hora y la fecha de los viajes.

Datos adicionales, como las condiciones climáticas, la contaminación sonora y la contaminación del aire.

La recopilación de datos se llevará a cabo utilizando una combinación de métodos, incluyendo:

Extracción de datos estáticos, como los datos históricos de los servicios de taxis y viajes compartidos.

Llamadas a API, para obtener datos en tiempo real de los servicios de taxis y viajes compartidos.

**Fase 2: Preparación de datos**

En esta fase, los datos recopilados se prepararán para el análisis. Esto implicará tareas como:

Limpieza de datos, para eliminar errores y valores atípicos.

Normalización de datos, para asegurar que los datos estén en el mismo formato.

Integración de datos, para combinar los datos de diferentes fuentes.

**Fase 3: Análisis de datos**

En esta fase, se analizarán los datos preparados para identificar patrones y tendencias. El análisis se centrará en las siguientes áreas:

Demanda de taxis, para comprender los factores que afectan a la demanda de taxis en la ciudad de Nueva York.

Impacto ambiental de los taxis, para evaluar el impacto de los taxis en la calidad del aire y la contaminación sonora.

**Fase 4: Desarrollo de modelos**

En esta fase, se desarrollarán modelos de machine learning para predecir la demanda de taxis y evaluar el impacto ambiental de los taxis. Los modelos se entrenarán utilizando los datos preparados en la fase 3.

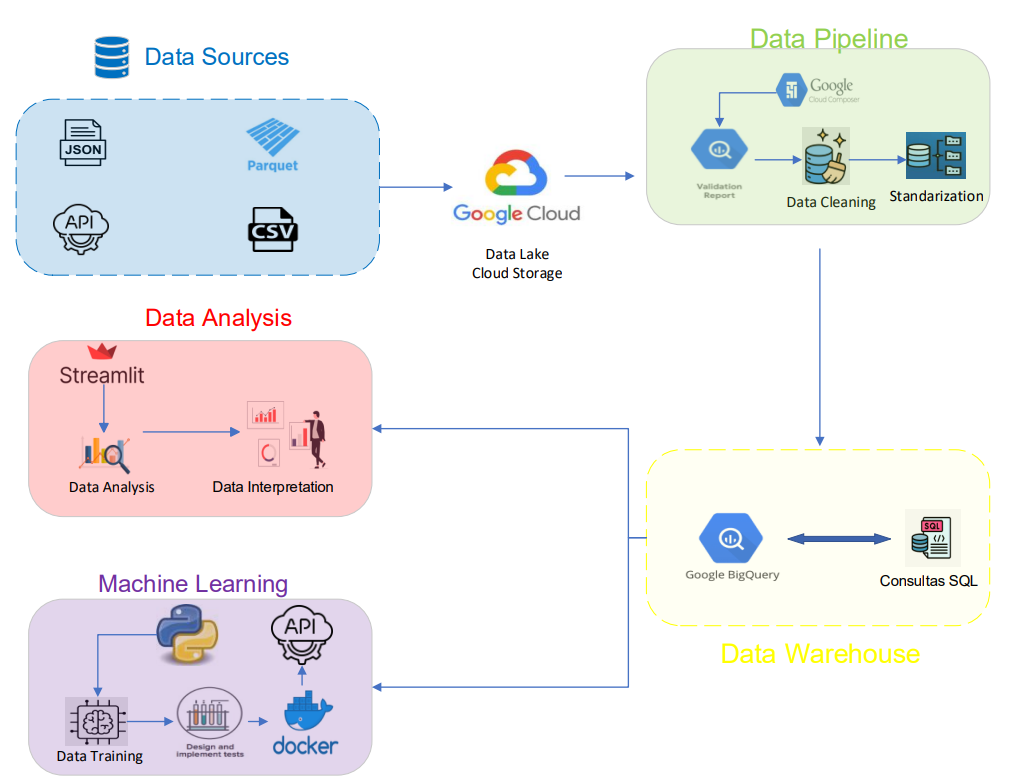
**Fase 5: Implementación de modelos**

En esta fase, los modelos desarrollados se implementarán en un entorno de producción para proporcionar estimaciones de demanda y análisis de impacto ambiental en tiempo real.

**Fase 6: Comunicación de resultados**

En esta fase, se comunicarán los resultados del proyecto a la empresa de transporte de pasajeros. Los resultados se presentarán en un informe que destacará las métricas clave y tendencias identificadas en el análisis de datos.

1. **IMPLEMENTACIÓN STACK TECNOLÓGICO SPRING 1**

****

1. **METODOLOGÍA DE TRABAJO**

El proyecto usará la metodología ágil SCRUM para la gestión de las responsabilidades y tareas necesarias para la construcción y entrega del proyecto.

1. **ROLES Y RESPONSABILIDADES**

En un proyecto de análisis de datos es necesario establecer diferentes ocupaciones que ayuden a la organización de tareas y el flujo de trabajo, es por eso que se han definido cinco roles fundamentales para este fin:

**Product Manager**: Edwin Torre

Encargado de gestionar el proyecto definiendo el problema y sus posibles soluciones, dirigiendo cada una de las tareas establecidas en el equipo.

**Program Manager**: Heidi Seltzer

Encargado de la coordinación y planificación estratégica de todo el proceso, haciendo un seguimiento del desarrollo y rendimiento en cada una de las asignaciones.

**Data Analyst**: Romina Capurro y Romina Escudero

Encargadas de encontrar información relevante en la base de datos según el objetivo del proyecto, a través de técnicas estadísticas.

**Data Scientist**: Heidi Seltzer y Edwin Torre

Encargados de hacer análisis profundos de los datos, desarrollando modelos predictivos para resolver problemas complejos.

**Data Engineer**: Juan Carlos Vásquez y Patricio Manuel Marzi

Encargados de diseñar arquitecturas e infraestructuras de bases de datos automatizadas para la optimización del sistema, haciendo disponible su uso, asegurando que los datos guarden coherencia y sean recuperables en caso de algún inconveniente a la hora de trabajarlos.

1. **DISEÑO DETALLADO**

Establecer cronogramas realistas y alineados con la disponibilidad de recursos y la complejidad de las tareas es un componente esencial para garantizar la organización efectiva del proyecto. Con este propósito, hemos creado un diagrama de Gantt que simplifica el seguimiento del avance del proyecto.

Adicionalmente, según se explica en la sección 6.3 titulada 'Metodología de Trabajo', hemos optado por administrar el proyecto utilizando la metodología ágil SCRUM. En este enfoque, hemos dividido las actividades del proyecto en tres sprints. Puede encontrar información detallada sobre esta división en el siguiente enlace:

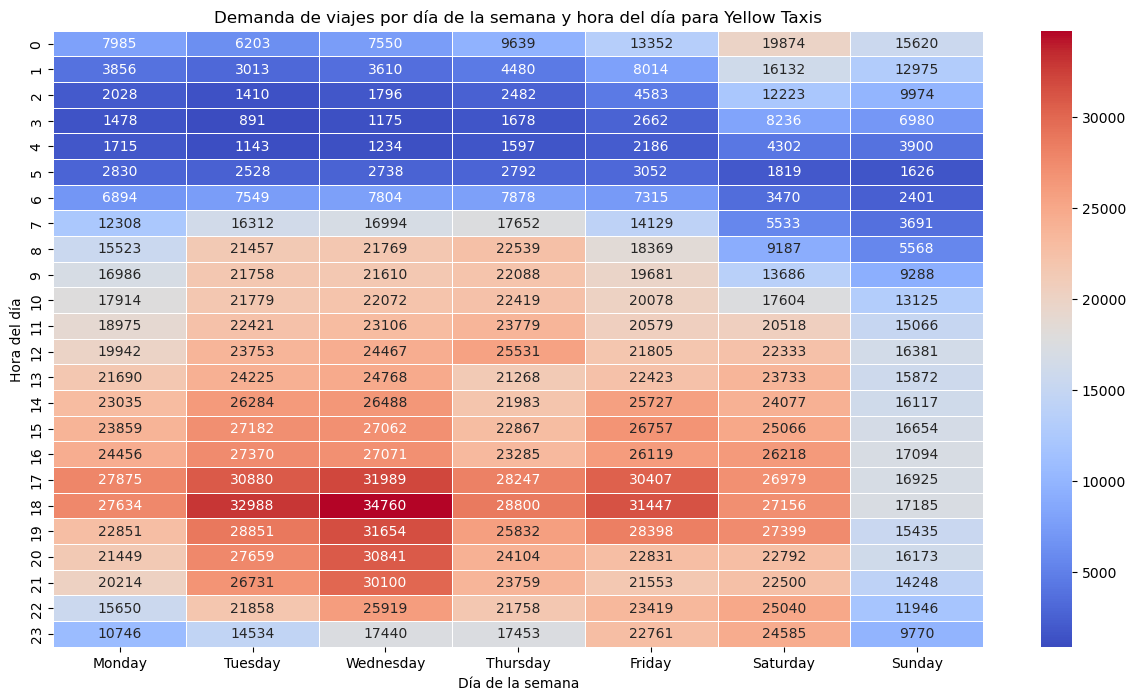
[Diagrama de Gantt](https://app.ganttpro.com/#/project/1698180333629/gantt)

1. **ANÁLISIS EXPLORATORIO DE LOS DATOS (EDA)**

El entendimiento de los datos es fundamental para poder determinar la problemática y la solución propuesta, es por ello que se realizó un análisis preliminar de los datos donde se realizaron los siguientes procedimientos:

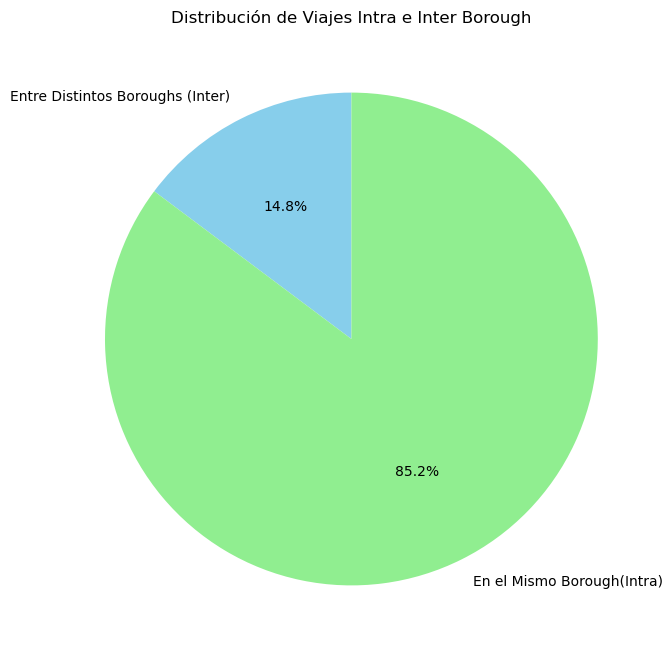
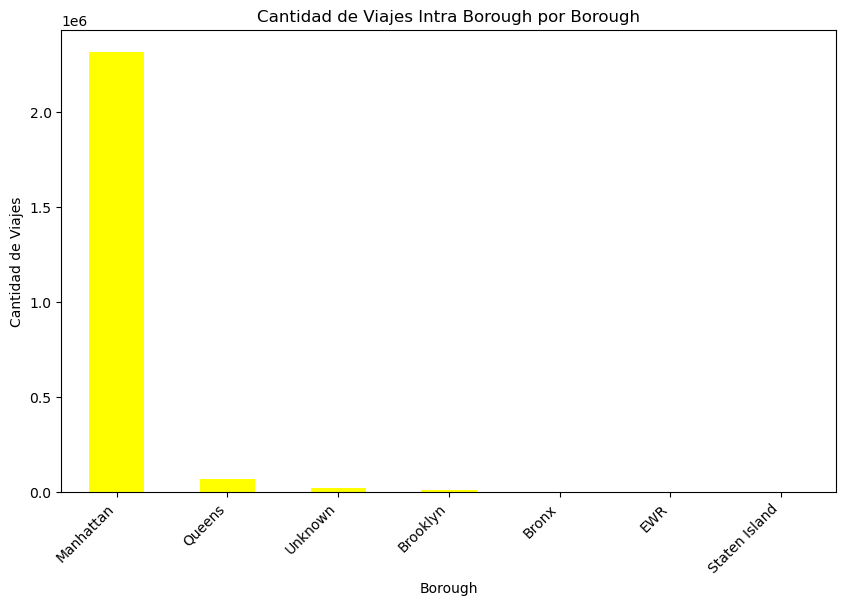
* Análisis preliminar de datos de viajes de Taxis Amarillos y Verdes en NYC.
* Análisis preliminar de emisiones CO2 de Taxis Amarillos.
* Análisis preliminar de estaciones de carga eléctrica y de combustibles alternativos.
* Análisis preliminar de contaminación del aire.
* Análisis preliminar de la calidad del aire en la ciudad de NYC.
* Análisis preliminar del clima en la ciudad de NYC.
* Análisis preliminar de la demanda de Taxis Amarillos y Verdes en NYC.
* Análisis preliminar para modelo que determine el mejor momento para viajar considerando la baja demanda de taxis.

**6.6.1.** **ANÁLISIS Preliminar**

**Demanda de viaje de Taxis Amarillos en NYC por horario y día**

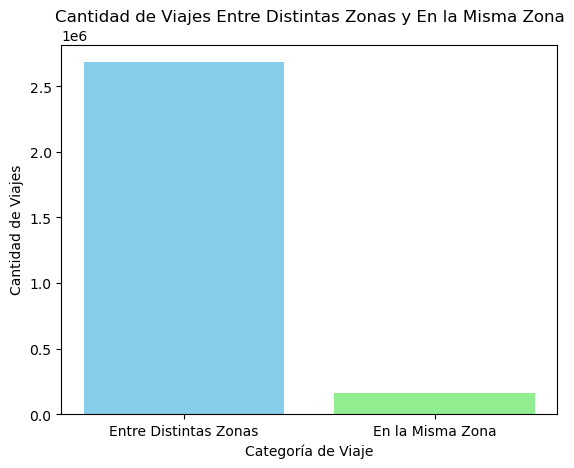
La demanda de viajes de Taxis amarillos muestra cuales son los horario y días con mayor cantidad de viajes. En este caso de martes a viernes de 17 a 19 hs.

**Demanda de Taxis amarillos por Borough (viajes en el mismo Borough o entre distintos Borough)**

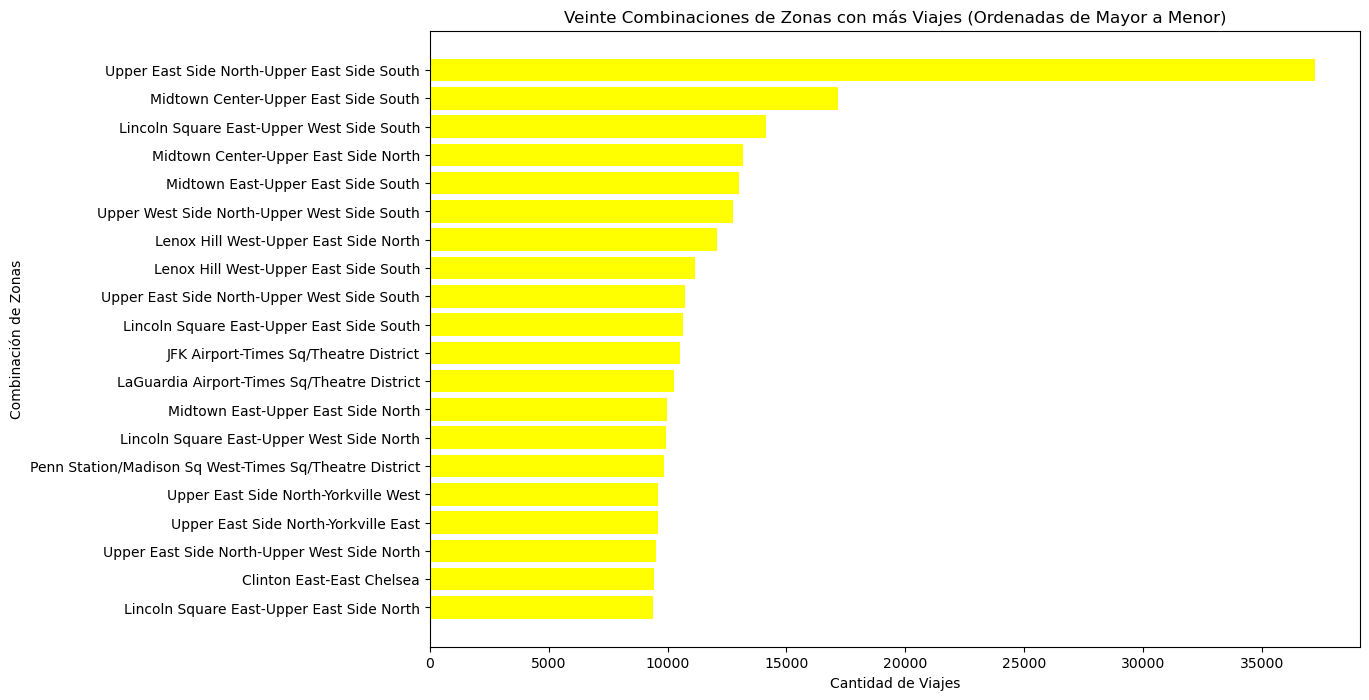
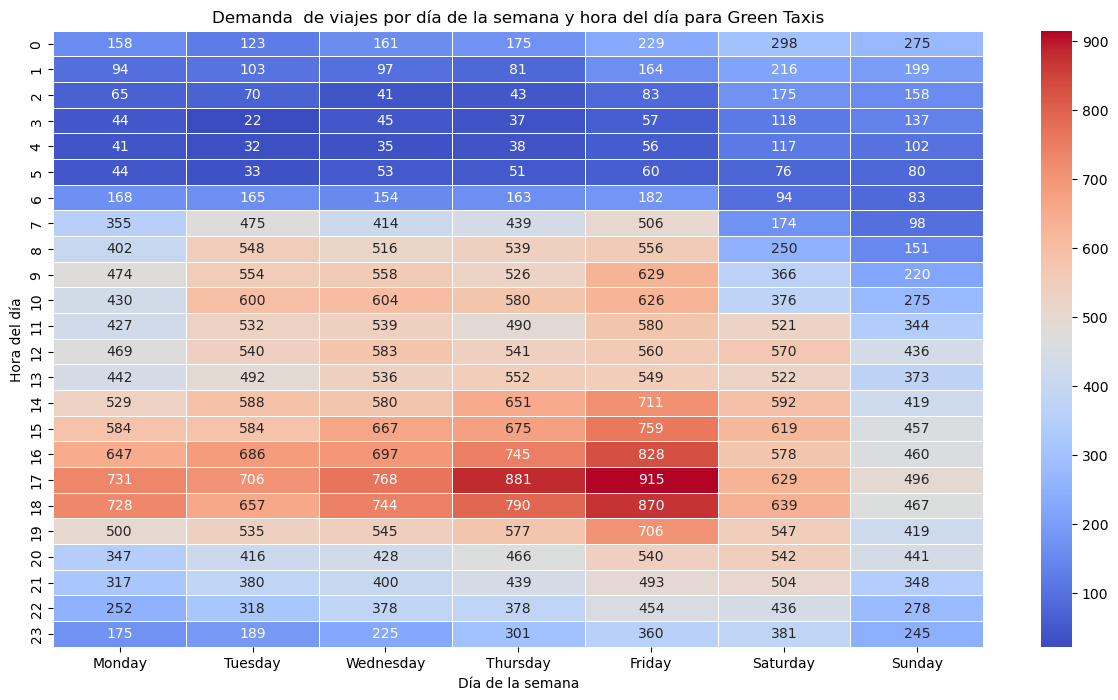
**** 

Los taxis amarillos hacen viajes principalemte dentro de un mismo Borough y este es Manhattan.

**Demanda de Taxis amarillos por Zonas**

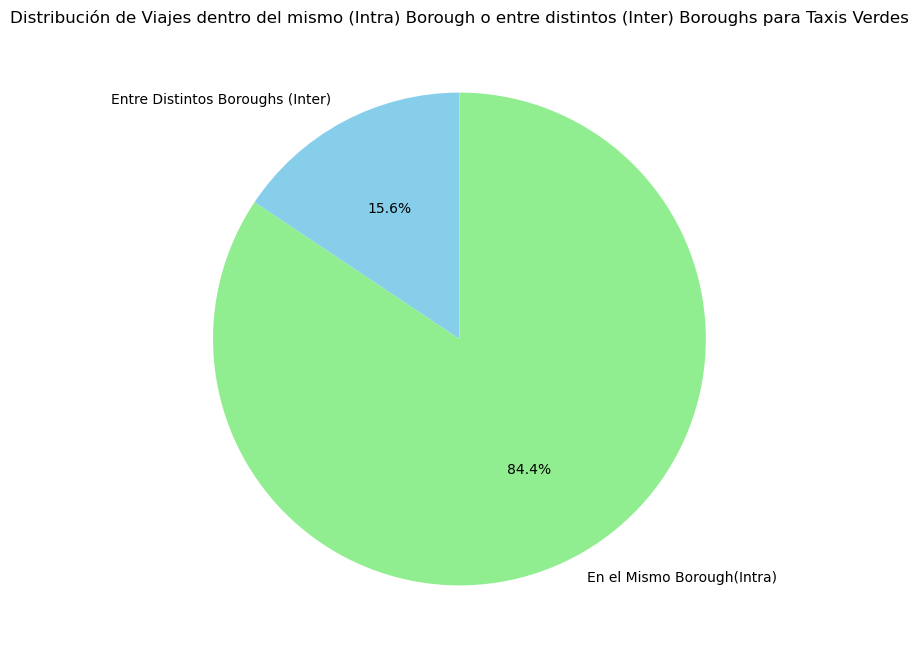


Los taxis en su gran mayoría inician el viaje en una zona y dejan a los pasajeros en otra zona.

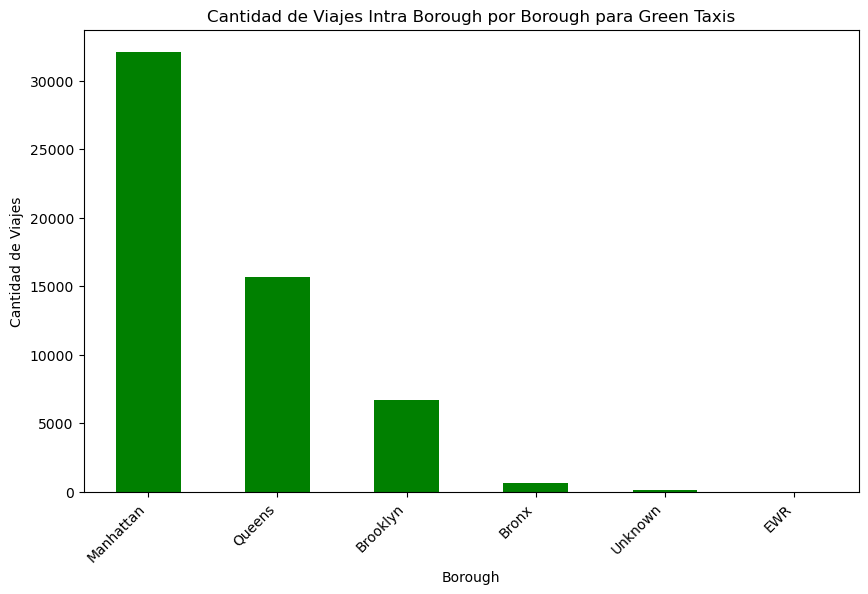
**20 Combinaciones más usuales de zonas de viaje para taxis amarillosDemanda para Taxis verdes por días y horarios**

La demanda de taxis verdes se concentra entre jueves y viernes de 16 a 19 hs.

**Demanda de Taxis verdes por cantidad de viajes entre distintos Borough o dentro del mismo Borough**

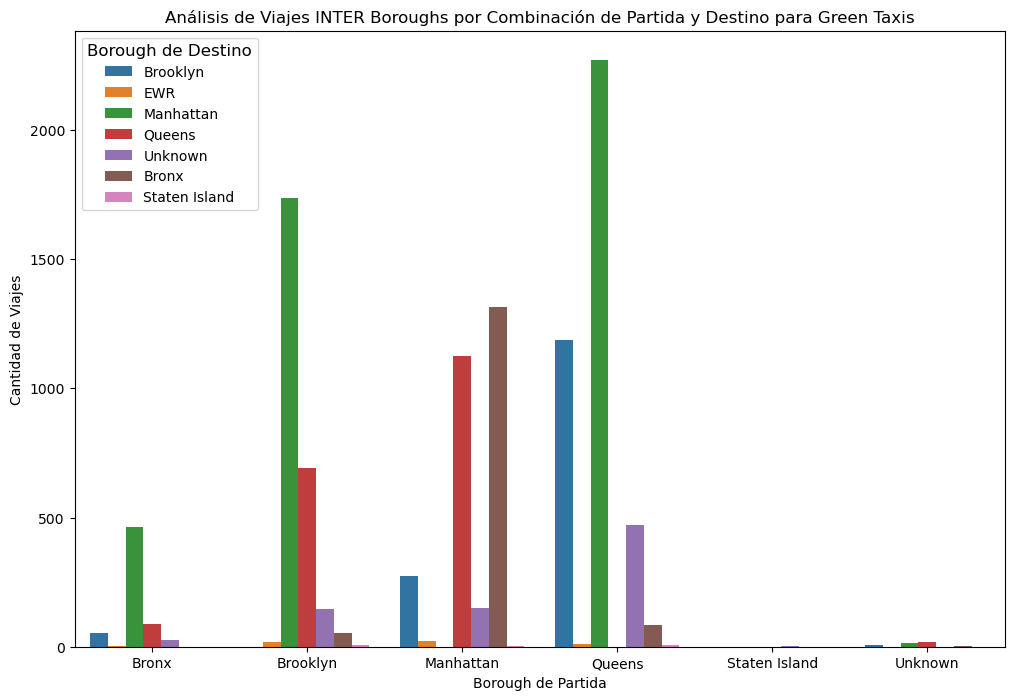
****

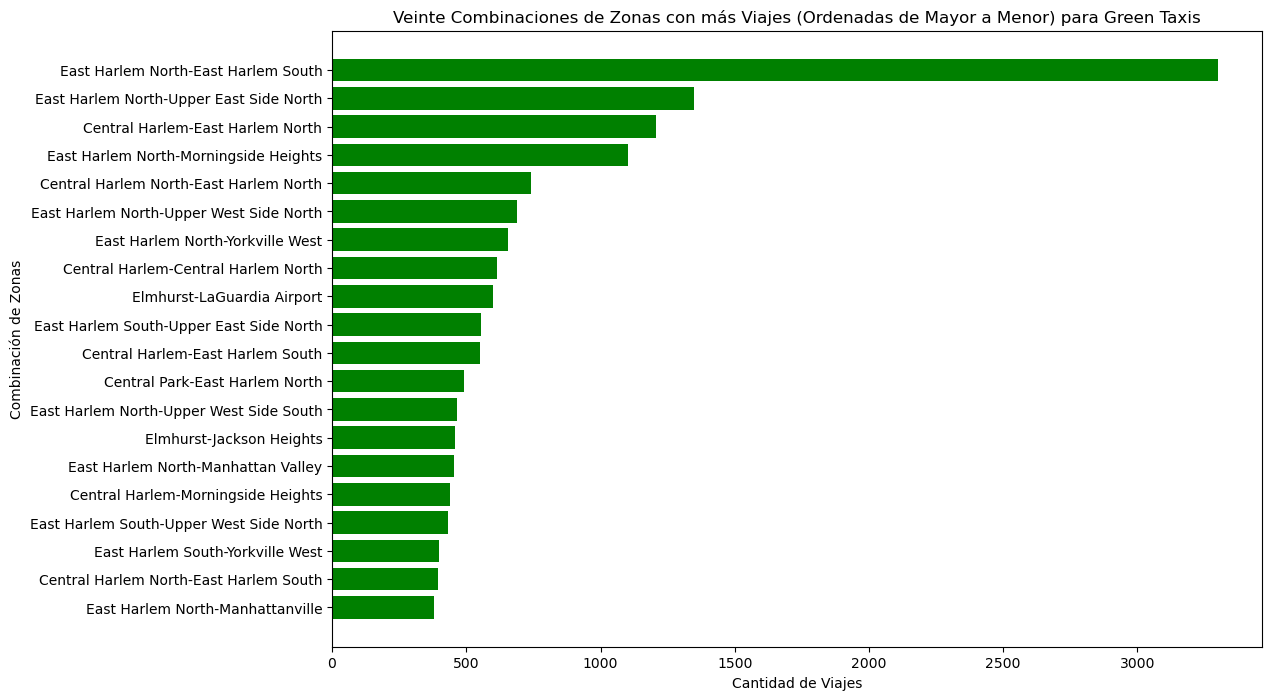
En general los viajes se realizan dentro de un mismo Borough

****

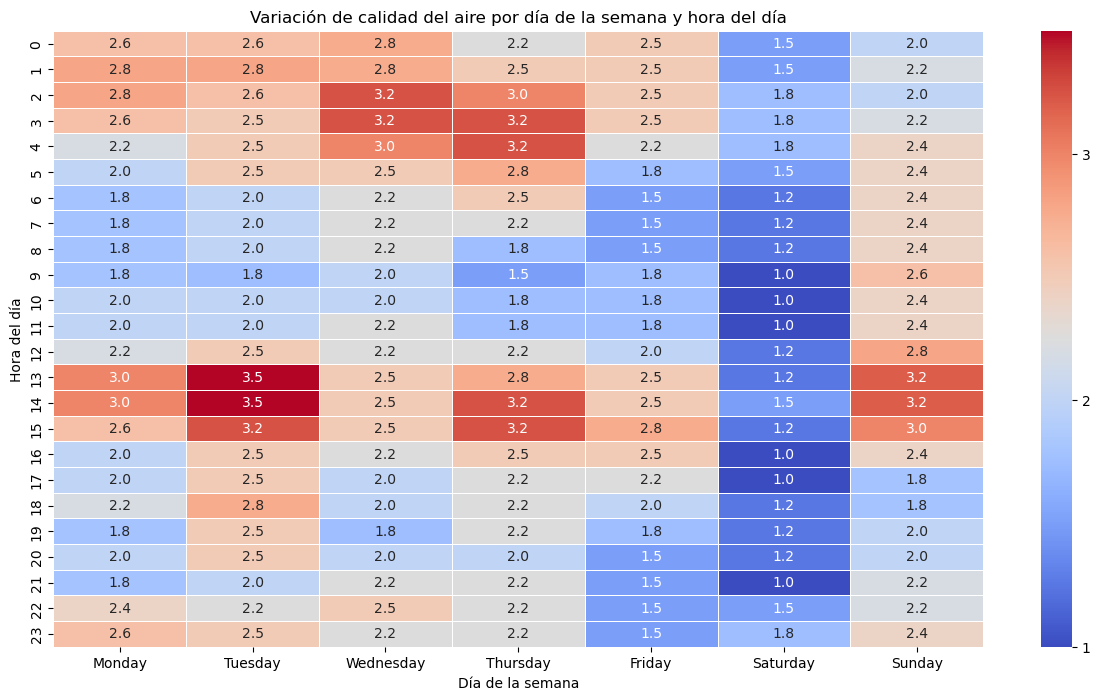
Los viajes que se realizan dentro de un mismo Borough son principalmente en Manhattan sigue siendo los más abundante aunque más equilibrado que taxis amarillos, los taxis verdes realizan viaje dentro de Queens, Brooklyn y Bronx.

**Demanda de viajes de taxis verdes entre distintos Borough**

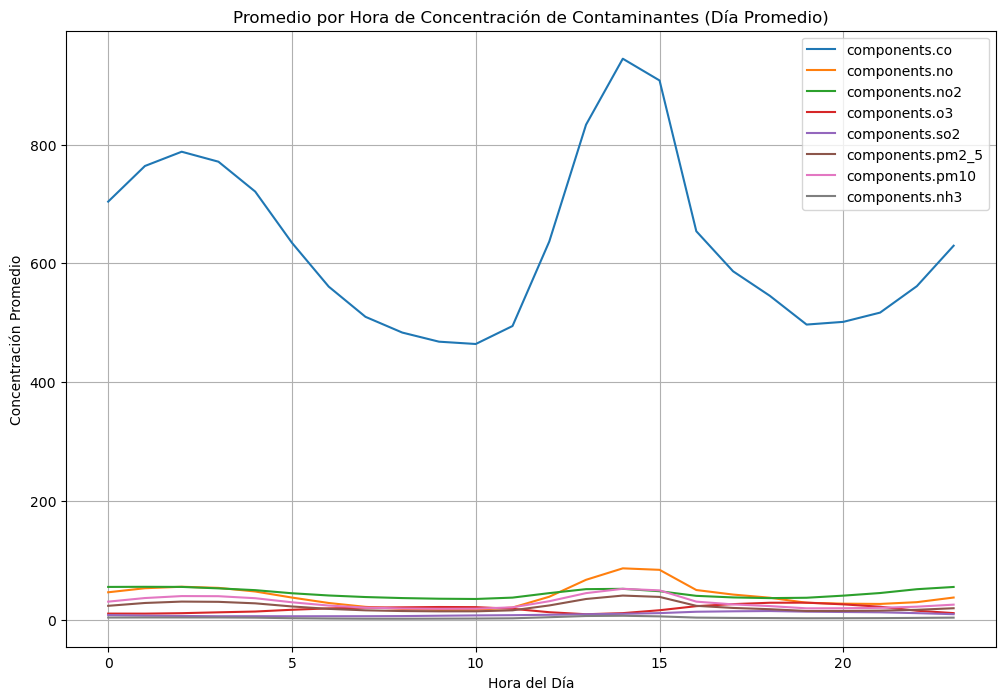
****Los principales viajes inter Borough son de Queens a Manhattan y de Brooklyn a Manhattan para taxis verdes . Le Siguen de Manhattan a Bronx y Manhattan a Queens.

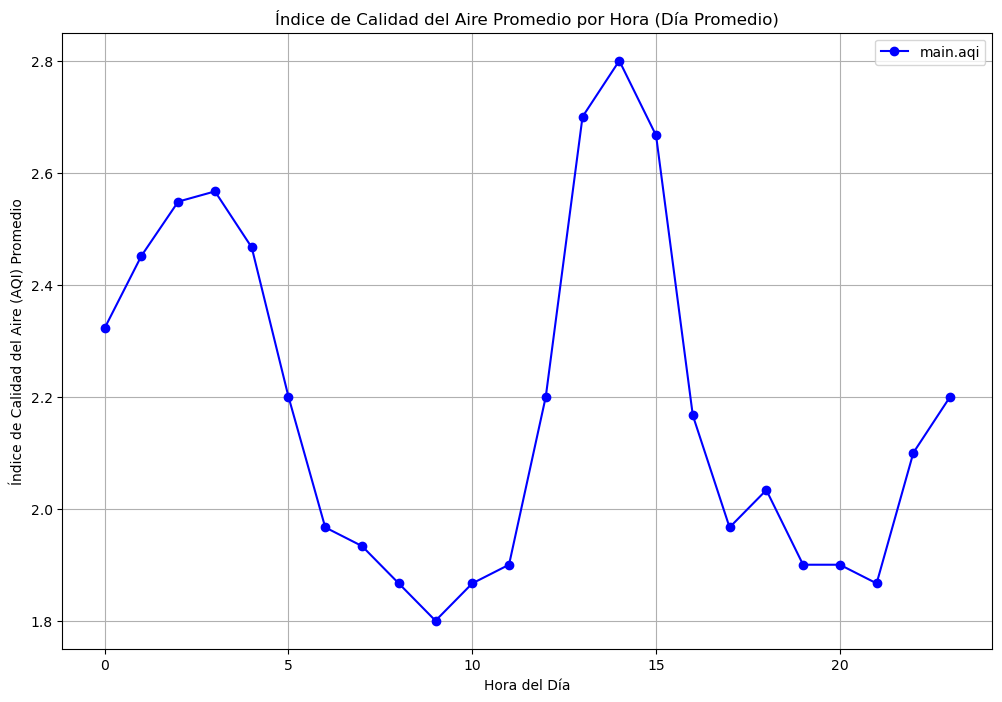
**Las 20 combinaciones de Zonas con más viajes para taxis verdes son:**

**Contaminación del aire NYC**



Los horarios y días de contaminación aérea tiene dos picos de peores calidades y los picos máximos son entre las las 13 hs y las 15 hs.





La totalidad del código y los notebooks de los EDA preliminares se encuentran en el siguiente link: [**Exploración Inicial**](https://github.com/Chinaskidev/Taxis-NY/tree/main/EDA_Preliminar).

1. **REPOSITORIO GITHUB**

El repositorio, readme y esquematización del proyecto se encuentra en el siguiente link:

[**Repositorio**](https://github.com/Chinaskidev/Taxis-NY)

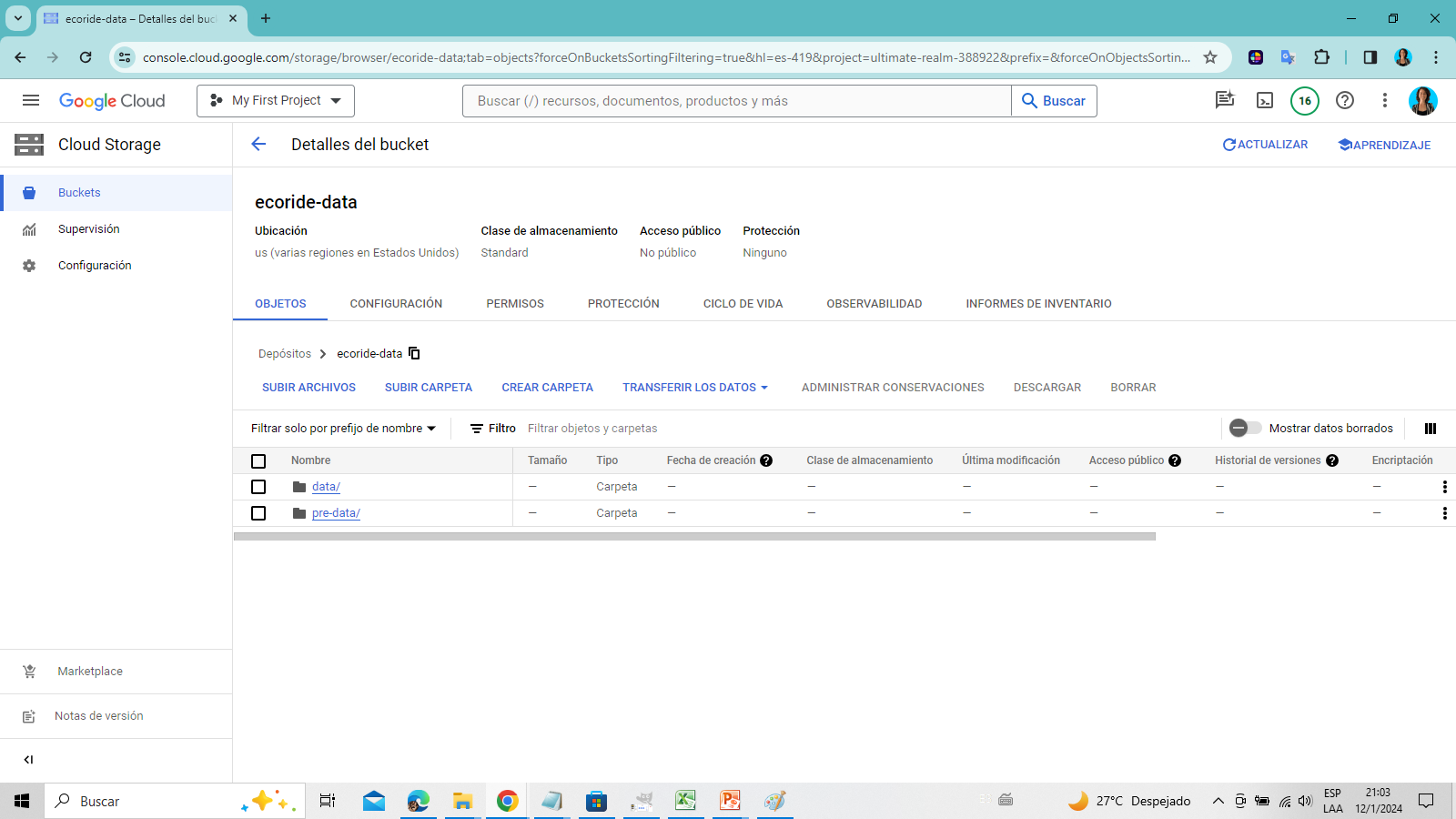
**7. SEGUNDO SPRINT**

**7.1 ARQUITECTURA DE DATOS**

La arquitectura de datos que se propone, utiliza Google Cloud Storage como Data Lake para almacenar datos en bruto y Google BigQuery como Data Warehouse para el análisis estructurado. Además, la implementación en la nube se realiza utilizando las herramientas y servicios disponibles en Google Cloud Platform para garantizar una gestión eficiente y escalable de la infraestructura.

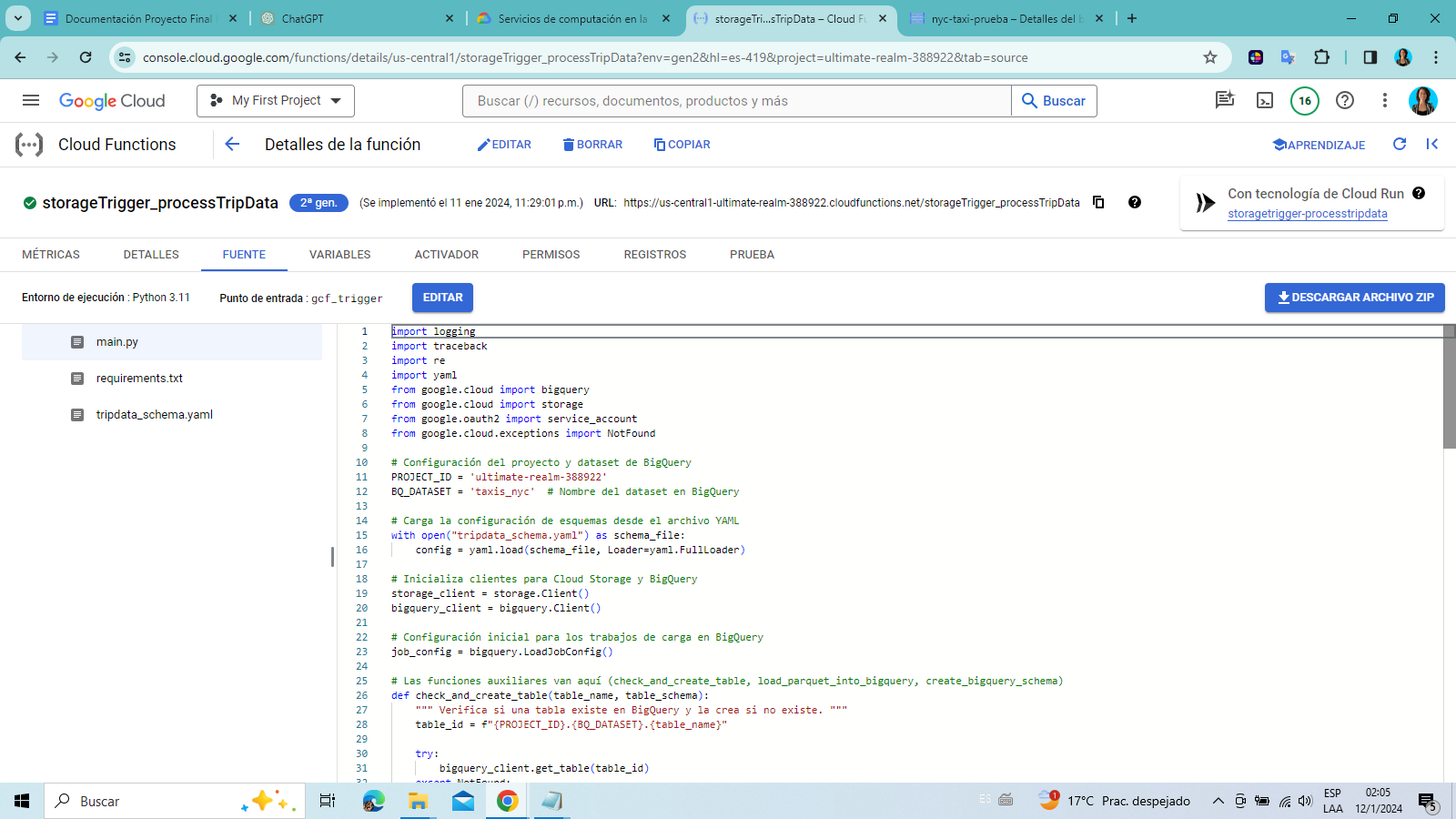
**7.2. PIPELINE ETL AUTOMATIZADO**

**7.2.1. Extracción**

* **Extracción de Datos TLC:**
  + Se extrae información de la TLC de la ciudad de Nueva York. Esta información incluye datos relacionados con servicios de taxi amarillos, taxis verdes y autos de alquiler.
* **Formato de Almacenamiento:**
  + Los datos extraídos se encuentran en formato Parquet y se almacenan en este mismo formato. Parquet es un formato de almacenamiento columnar eficiente que es particularmente adecuado para conjuntos de datos grandes.
* **En Google Cloud Storage (Bucket ) se crea el Datalake**
  + Se crea un bucket en Google Cloud Storage llamado 'ecoride-data'.
  + Dentro de la carpeta data, se encuentra tres carpetas: taxis amarillos, verdes y autos de alquiler. En estas carpetas es donde hay que subir los archivos descargados de la TLC de NYC con formato parquet.
  + 
* **7.2.2. Transformación**
  + Una vez que se cargaron los datos en la carpeta correspondiente se dispara de forma automática una función Trigger, llamada [storageTrigger\_cleanTripData](https://console.cloud.google.com/functions/details/us-central1/storageTrigger_cleanTripData?env=gen2&hl=es-419&project=ultimate-realm-388922) , que desencadena un proceso de ETL. Para ello llama internamente una función llamada clean\_data que verifica outliers, duplicados, nulos y normaliza el nombre y tipo de cada columna. Al finalizar el proceso los archivos limpios se almacenan, dentro de la carpeta pre-data del bucket ‘ecoride-data’. Los buckets de Google Cloud Storage proporcionan un almacenamiento escalable y duradero en la nube.

**7.2.2. Carga**

La transformación de datos es una parte crucial del proceso de preparación de datos. En este paso a través de la implementación de una Cloud Function llamada, “[storageTrigger\_processTripData](https://console.cloud.google.com/functions/details/us-central1/storageTrigger_processTripData?env=gen2&hl=es-419&project=ultimate-realm-388922)”, se activa en respuesta a un archivo subido a Google Cloud Storage. La función realiza varias acciones relacionadas con la carga y transformación de datos en BigQuery.

* **Configuración Inicial:**
  + Se configuran las credenciales y la información del proyecto, así como el nombre del dataset en BigQuery.
* **Carga de Configuración de Esquema:**
  + Se carga la configuración de esquema desde un archivo YAML llamado "tripdata\_schema.yaml".
* **Inicialización de Clientes de Cloud Storage y BigQuery:**
  + Se inicializan los clientes de Cloud Storage y BigQuery utilizando las bibliotecas proporcionadas por Google Cloud.
* **Funciones Auxiliares:**
  + Se definen varias funciones auxiliares, como check\_and\_create\_table, load\_parquet\_into\_bigquery, y create\_bigquery\_schema. Estas funciones realizan tareas específicas relacionadas con la creación de tablas en BigQuery, carga de datos desde Parquet en Cloud Storage a BigQuery, y conversión de esquemas YAML a esquemas de BigQuery, respectivamente.
* **Función Principal (gcf\_trigger)**:
  + Esta función se activa cuando se sube un archivo a Google Cloud Storage.
  + Verifica el nombre del archivo y determina en qué carpeta específica se encuentra (pre-data/taxis-amarillos/, pre-data/taxis-verdes/, o pre-data/taxis-alquiler/).
  + Luego, itera sobre la configuración de tablas definida en el archivo YAML y, si encuentra una coincidencia en el nombre del archivo, crea la tabla en BigQuery y carga los datos desde el archivo Parquet.
* **Manejo de Errores:**
  + La función incluye un manejo básico de errores que imprime información de seguimiento en caso de cualquier excepción.
* 

**8.1. DATAWHEREHOUSE**

La tabla correspondiente a los datos transformados se crea en BigQuery utilizando la estructura y esquema definidos durante la transformación.

**Consulta SQL para Generar Tablas Consumibles:**

* + Se utilizan consultas SQL para generar tablas específicas que contienen la información necesaria para los KPIs planteados. Esto puede implicar la agregación, filtrado y manipulación de datos según los requisitos de los KPIs.

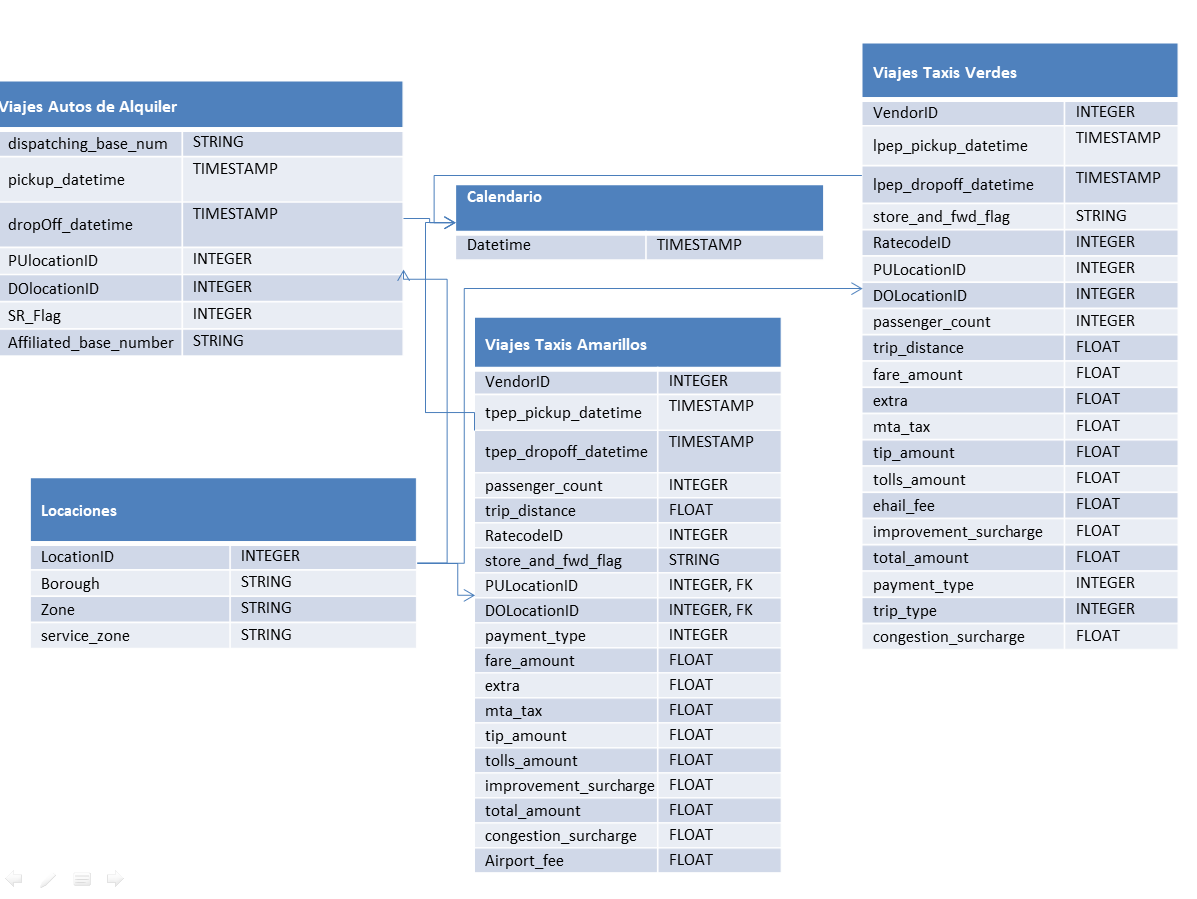
**Exportación a Looker Studio:**

* + Después de haber preparado las tablas consumibles con los datos necesarios, se exportan a Looker Studio. Looker es una plataforma de inteligencia comercial que permite crear visualizaciones y dashboards interactivos.

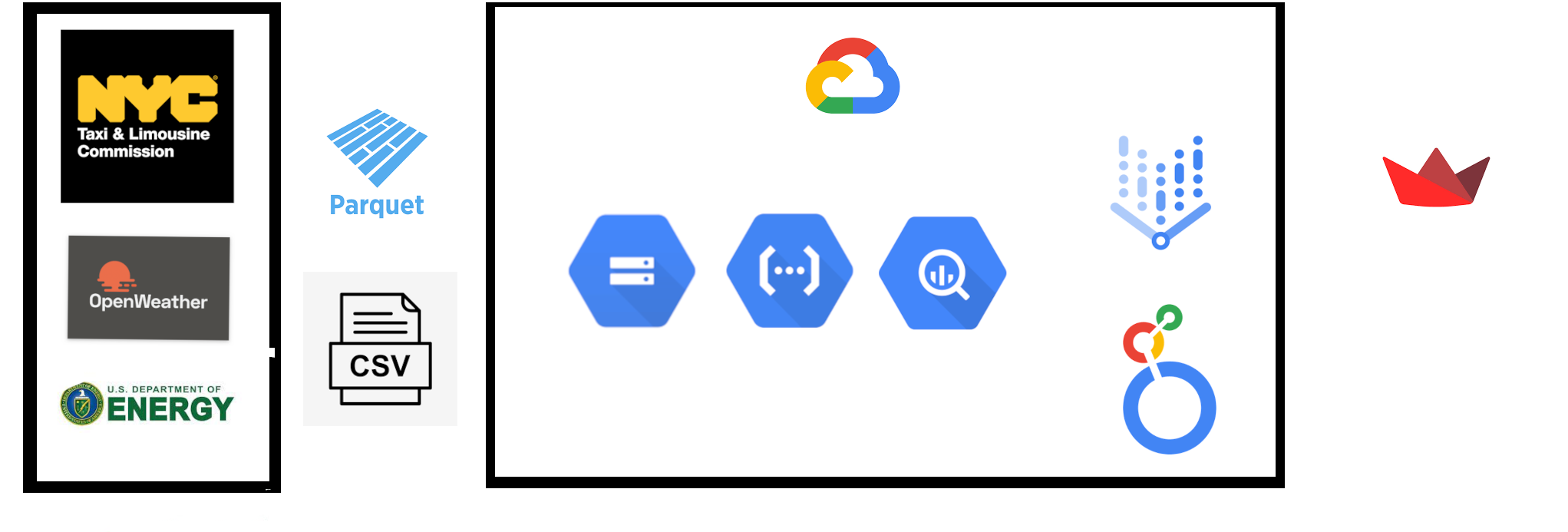
**Creación de Dashboards de KPIs:**

* + Utilizando Looker Studio, se diseñan y crean dashboards interactivos que presentan los KPIs planteados. Estos dashboards proporcionan una interfaz fácil de usar para explorar y analizar los datos en función de los indicadores clave de rendimiento.

**8.2 DIAGRAMA ER DETALLADO (TABLAS, PK, FK Y TIPO DE DATO)**

****

**8.3 Pipeline, detallando tecnologías**

****

### **8.4. MVP/ PRUEBA DE CONCEPTO DE MACHINE LEARNING:**

### **Análisis de Demanda de Taxis en Nueva York**

### Previsión de demanda

Las funciones proporcionadas de análisis basado en datos y análisis predictivo con tecnología de inteligencia artificial ayudan a estimar la demanda de los usuarios al siguiente nivel. La función de minería de datos de la IA recupera datos históricos relacionados con la demanda de taxis de los usuarios en un rango particular. Además, estos datos se analizan en función del análisis predictivo que ofrece predicciones de demanda muy precisas. Teniendo en cuenta datos tan refinados, los operadores empresariales ponen a disposición los recursos de manera eficiente.

El sistema de apoyo al despacho de taxis predice la demanda del servicio de taxi para cada Borough. El sistema utiliza datos procedentes de registros de servicios de taxi junto con varios otros factores que afectan la demanda de taxis en una región en particular, como las condiciones climáticas. Con la ayuda de la tecnología de inteligencia artificial, el sistema pronostica el número de taxis ocupados en NYC.

Para lograr esto se propone evaluar distintos modelos de aprendizaje y medir su rendimiento de predicción.

.**Objetivo:**

Sugerir un modelo de aprendizaje automático para mejorar la rentabilidad del negocio de taxis en Nueva York.

Utilizando los datos de viajes de taxis, podríamos desarrollar un modelo para predecir la demanda de taxis en diferentes zonas y horarios. Esto ayudaría a distribuir los taxis de manera más eficiente, reduciendo los tiempos de espera y maximizando los ingresos.

**Metodología:**

En general, el código realiza un análisis completo de datos de viajes en taxi, preprocesa los datos, entrena varios modelos de regresión y evalúa su rendimiento, concluyendo con una recomendación basada en el rendimiento observado. Se explica paso a paso los modelos de aprendizaje:

* Se importan las bibliotecas necesarias para el análisis de datos y modelado, como Pandas para manipulación de datos, NumPy para operaciones numéricas, Matplotlib para visualización y varias clases de modelos y métricas de Scikit-Learn para aprendizaje automático.
* Se cargan dos conjuntos de datos: uno relacionado con los viajes en taxi (taxi\_data) y otro con información sobre los distritos (borough\_data)
* Se realiza el preprocesamiento de los datos, incluyendo la conversión de fechas, mapeo de ubicaciones a distritos, y creación de variables dummy (one-hot encoding). Luego, los datos se dividen en conjuntos de entrenamiento y prueba
* Se definen varios modelos de regresión (Lineal, Árbol de Decisión, Bosque Aleatorio y Gradient Boosting). Luego, se utiliza una función (train\_and\_evaluate) para entrenar estos modelos y evaluar su rendimiento.
* Se crea una visualización que compara las predicciones de los modelos con los valores reales.
* Se genera una tabla que resume el rendimiento de cada modelo en términos de MSE, RMSE y R².
* Se proporciona una interpretación de los resultados, destacando el rendimiento de cada modelo y haciendo una recomendación basada en este análisis.

**Despliegue en Streamlit**

Se desarrolla una aplicación web de Streamlit que permite a los usuarios obtener de manera interactiva predicciones de la demanda de viajes en taxi para diferentes distritos (boroughs) de la Ciudad de Nueva York, basándose en un modelo de aprendizaje automático desplegado.

### **8.5. MVP/ PRUEBA DE CONCEPTO DE DASHBOARD**

**Replanteamiento de los KPIs**

**1. \*\*Eficiencia de la Flota Eléctrica:\*\***

- \*\*Objetivo:\*\* Evaluar la eficiencia de la flota de vehículos eléctricos en comparación con los vehículos convencionales.

- \*\*Fórmula:\*\* (Kilómetros recorridos por vehículo eléctrico / Consumo energético total de vehículos eléctricos) / (Kilómetros recorridos por vehículo convencional / Consumo de combustible total de vehículos convencionales)

- \*\*Meta:\*\* Alcanzar una eficiencia un 15% mayor en la flota de vehículos eléctricos en comparación con la flota de vehículos convencionales.

**2. \*\*Satisfacción del Usuario con Vehículos Eléctricos:\*\***

- \*\*Objetivo:\*\* Medir la satisfacción del usuario con los viajes en vehículos eléctricos.

- \*\*Fórmula:\*\* (Número de calificaciones positivas para vehículos eléctricos / Total de calificaciones para vehículos eléctricos) \* 100

- \*\*Meta:\*\* Mantener una satisfacción del usuario con vehículos eléctricos superior al 80%.

**3. \*\*Eficiencia en la Asignación de Recursos:\*\***

- \*\*Objetivo:\*\* Evaluar qué tan eficientemente se asignan los recursos en función de la demanda prevista.

- \*\*Fórmula:\*\* (Número real de taxis disponibles / Número estimado de taxis necesarios según la demanda) \* 100

- \*\*Meta:\*\* Lograr una eficiencia del 90% en la asignación de recursos.

**4. \*\*Tiempo Promedio de Recarga de Vehículos Eléctricos:\*\***

- \*\*Objetivo:\*\* Medir el tiempo promedio que los vehículos eléctricos pasan en el proceso de recarga.

- \*\*Fórmula:\*\* Suma de tiempo de recarga de todos los vehículos eléctricos / Número total de recargas

- \*\*Meta:\*\* Reducir el tiempo promedio de recarga a menos de 30 minutos.

**Estas métricas adicionales permitirán evaluar aspectos específicos de la implementación de vehículos eléctricos y su impacto en la eficiencia, satisfacción del usuario y sostenibilidad ambiental en el contexto del proyecto "EcoRide NYC".**

### 