



Advisors on the Sustainable Economic Transition

## **Proyecto EcoTaxisNYC**

**Informe de Proyecto: Análisis Inicial y Propuesta de  
Solución**

**22 de marzo de 2024**

## Contenido

<b>Introducción</b>	3
<b>Objetivos</b>	4
<b>Alcance</b>	5
<b>KPI's</b>	7
<b>Modelo de Machine Learning</b>	10
<b>Solución propuesta y metodología de trabajo</b>	12
<b>Stack tecnológico</b>	15
<b>Datos a utilizar</b>	17

## **Introducción**

La ciudad de Nueva York, conocida por su constante actividad, enfrenta desafíos significativos en términos de contaminación ambiental, especialmente en lo que respecta a la calidad del aire y el ruido. A medida que sus casi 8,5 millones de personas se desplazan diariamente por las calles congestionadas de esta metrópolis, el impacto de las emisiones vehiculares y el ruido generado por el tráfico se ha convertido en una preocupación creciente tanto para los residentes, como para las autoridades gubernamentales y los defensores del medio ambiente por igual.

La proliferación de vehículos convencionales, impulsados por combustibles fósiles, ha contribuido en gran medida a la contaminación del aire en Nueva York, con emisiones de gases nocivos que afectan la salud pública y degradan la calidad del entorno urbano. Además, el constante zumbido de motores y el estruendo de bocinas resuenan en las calles, añadiendo un nivel de ruido que afecta la calidad de vida de quienes viven y trabajan en la ciudad.

Ante este panorama, surge una necesidad imperante de buscar soluciones innovadoras y sostenibles que mitiguen los impactos negativos del transporte urbano en el medio ambiente y la salud de la población. En este contexto, la introducción de una flota de taxis ecológicos emerge como una medida prometedora para abordar estos desafíos de manera integral.

La transición hacia vehículos eléctricos y otros modelos de transporte sostenible no solo ofrece la oportunidad de reducir las emisiones de gases de efecto invernadero y mejorar la calidad del aire, sino que también puede contribuir significativamente a la disminución del ruido en las calles de la ciudad. Al adoptar tecnologías más limpias y eficientes, se abre la posibilidad de transformar el panorama del transporte en Nueva York, promoviendo un entorno más saludable y habitable para todos sus residentes.

Advisors on the sustainable Economic Transition (ASET Company) es una empresa líder en asesoramiento estratégico que ayuda a las empresas a navegar y prosperar en la transición hacia una economía sostenible. Nuestro enfoque se basa en datos, brindando soluciones personalizadas respaldadas por análisis exhaustivos para impulsar la eficiencia, la innovación y el éxito a largo plazo en un mundo en evolución hacia la sostenibilidad.

En este contexto, ASET Advisors on the Sustainable Economic Transition se embarca en un proyecto ambicioso y visionario, con el objetivo de proporcionar análisis avanzado de datos y soluciones basadas en machine learning que respalden la introducción exitosa de una flota de taxis sostenibles en la ciudad de Nueva York. A través de este esfuerzo colaborativo, buscamos no solo ayudar a nuestro cliente a mejorar la movilidad urbana, sino también contribuir a la construcción de un futuro más sostenible y próspero para las generaciones venideras.

### **Staff:**

- Rafael Balestrini – Data Engineer
- Carlos Macea – Data Engineer
- Luis Rojas – Data Analytics
- Alter Caimi – Machine Learning Engineer

## **Objetivos**

Realizar un análisis avanzado de datos que proporcione insights significativos para respaldar la decisión de nuestro cliente en la introducción de una flota de taxis sostenibles en la ciudad de Nueva York. Este análisis abarcará una evaluación exhaustiva de diversos aspectos del mercado de transporte, incluyendo la demanda histórica y actual de servicios de taxis amarillos, taxis verdes y vehículos de alquiler de gran volumen. Además, se realizará un análisis geográfico para identificar áreas de alta demanda y oportunidades para una implementación eficiente de la flota sostenible.

Además, nuestro proyecto incluirá el desarrollo de un modelo de machine learning que prediga la demanda de pasajeros en el corto plazo y/o en tiempo real. Este modelo utilizará datos históricos de viajes y factores climáticos, para identificar los lugares donde la flota de taxis sostenibles puede competir con mayor probabilidad de éxito.

En resumen, nuestro objetivo es proporcionar a nuestro cliente una comprensión profunda y basada en datos del mercado de transporte en Nueva York, junto con herramientas predictivas que les permitan tomar decisiones estratégicas informadas sobre la introducción de una flota de taxis sostenibles.

## **Alcance**

### **1. Análisis Profundo de Datos de Servicios de Taxis (Estado de Mercado):**

Realizar un análisis detallado de la demanda de servicios de taxis amarillos, taxis verdes y vehículos de alquiler de gran volumen a lo largo del tiempo, identificando patrones estacionales, diarios y por hora.

Explorar la distribución geográfica de los servicios de taxis para identificar áreas de alta demanda y oportunidades para la implementación eficiente de una flota de taxis sostenibles.

Analizar la duración de los viajes, los puntos de inicio y finalización más comunes, así como los posibles cuellos de botella en la red de transporte.

Revisar los costos asociados a las tarifas por cada viaje.

### **2. Evaluación del Impacto Ambiental de la Flota de Taxis Sostenibles:**

Utilizar los datos disponibles sobre los diferentes tipos de vehículos y su eficiencia, así como los datos sobre la calidad del aire y el ruido, para evaluar el impacto potencial de una flota de taxis sostenibles en la reducción de emisiones y la mejora de la calidad ambiental. Esta evaluación se realizará por Distrito

### **3. Evaluación de la Viabilidad de la Flota de Taxis Sostenibles:**

Realizar un análisis comparativo de condiciones técnicas y costos entre los diferentes tipos de vehículos (eléctricos) para determinar los posibles tipos de vehículos que se deberían considerar en el momento de la compra de vehículos para la flota.

Identificar las tarifas promedio de servicio, impuestos, peajes, propinas, etc. Para estimar el costo del servicio.

### **4. Desarrollo de KPIs y Dashboard Personalizado:**

Identificar y definir KPIs específicos que sean relevantes para medir el éxito de la introducción de una flota de taxis sostenibles, como la reducción de emisiones y la eficiencia del servicio. (mínimo 3 KPI's).

Diseñar un dashboard personalizado que visualice estos KPIs así como también muestre el análisis de mercado, impacto ambiental y viabilidad de la introducción de la flota de taxis sostenibles. Para soportar la toma de decisiones informadas.

### **5. Modelo Predictivo:**

Desarrollar modelo predictivo utilizando técnicas de machine learning para prever la demanda futura de servicios de taxis y optimizar la asignación de recursos (flota de taxis sostenibles) en función de la demanda prevista y otros factores externos como el clima y eventos especiales.

### **6. Despliegue en la Nube y en la Web:**

Trasladar todos los componentes del proyecto, incluidos los datos, el código, el dashboard interactivo y el modelo de machine learning, a una plataforma en la nube para

su ejecución y acceso remoto. Lo que garantiza automatización, escalabilidad y disponibilidad.

Fuera de Alcance:

Análisis Económico de la Incorporación de Vehículos Alternativos vs Convencionales: No es necesario, no se considera como opción la incorporación mixta o completa de vehículos convencionales, dado que el propósito del cliente es el impacto positivo en la reducción de emisión del CO<sub>2</sub> en NYC. Además, el gobierno local tiene planes a corto, mediano y largo plazo para la sustitución de vehículos convencionales a eléctricos.

Análisis de la contaminación del aire o ruido por zonas: No se cuenta con información suficiente para esta discriminación en la información; aunque se hará lo posible “de ser necesario”, realizar estimaciones basadas en los tipos de vehículos que transitan por zona. Sin embargo, se considera que la agregación actual es suficiente para conocer el impacto de los vehículos en la contaminación ambiental.

Vehículos convencionales de alquiler: Este servicio no es representativo al ser de servicios especiales por lo general de lujo que al no incluirlo en el análisis, no genera impacto en el negocio particular del cliente.

## KPI's

- KPI\_1:

Definición: Variación porcentual de la contaminación sonora mensual con respecto al año anterior en NYC.

Fórmula:

$$\frac{CS_{\text{año\_anterior}} - CS_{\text{año\_actual}}}{CS_{\text{año\_anterior}}} \times 100$$

Siendo  $CS$  la cantidad total de sonidos de motores para el distrito en análisis.

Objetivo: reducción de un 10% con respecto al año anterior.

Alcance: Datos disponibles hasta septiembre de 2020. Datos discriminados por distrito.

- KPI\_2:

Definición: Variación porcentual de la contaminación del aire por km recorrido con respecto al año anterior por distrito en NYC.

Fórmula: Para este KPI definimos un indicador que combina la concentración ambiental de  $NO_2$ ,  $PM_{2.5}$  y  $O_3$ , y calculamos la variación de ese indicador.

$$\text{Indicador de calidad de aire} = ICA = k_1[NO_2] + k_2[O_3] + k_3[PM_{2.5}]$$

$ICA$  es una media ponderada. Los pesos de cada contaminante los asignaremos teniendo en cuenta la correlación de cada variable con las hospitalizaciones o muertes debidas a los mismos.

$$ICA_{km} = \frac{ICA}{km \text{ recorridos}} \times factor$$

$ICA_{km}$  es el Indicador de calidad de aire por kilómetro recorrido en un año, para un distritito determinado de NYC.  $factor$  es un factor de corrección de escala a determinar en el análisis, que dependerá del orden de magnitud de los kilómetros recorridos. Por ejemplo, puede ser igual a 100000.

Por lo tanto, la fórmula de la variación porcentual de  $ICA_{km}$  corresponde al KPI definido:

$$\frac{ICA_{km\_año\_anterior} - ICA_{km\_año\_actual}}{ICA_{km\_año\_anterior}} \times 100$$

Objetivo: reducción del ICA de un 10 % con respecto al año anterior

Alcance: datos anuales desde 2009 hasta 2021, discriminados por distrito.

- KPI\_3

Definición: Variación porcentual en las emisiones de  $CO_2$  por unidad de taxi en NYC.

Fórmula:

Para este KPI definimos primero la métrica ‘emisiones por unidad de taxi’:

$$E_{taxi} = \frac{Emisiones\ CO_2}{Cantidad\ de\ taxis}$$

Las emisiones de CO2 se miden en toneladas y se multiplica o divide por un factor de escala.

Luego, el KPI se define como la variación porcentual de esta métrica respecto al año anterior:

$$\frac{E_{taxi\_año\_anterior} - E_{taxi\_año\_actual}}{E_{taxi\_año\_anterior}} \times 100$$

Objetivo: Reducción de un 10 % las emisiones respecto del año anterior.

Alcance: Datos anuales de 2005 a 2021 no discriminados por distrito.

- KPI\_4

Definición: Variación porcentual trimestral del número de viajes con respecto al día de la semana de menor demanda en un distrito determinado de NYC.

Fórmula:

Definimos  $VT = \text{Viajes totales en el día de menor demanda}$

$$\frac{VT_{trimestre\ actual} - VT_{trimestre\ anterior}}{VT_{trimestre\ anterior}} \times 100$$

Objetivo: Aumentar en un 5% en un trimestre el número de viajes totales el día de la semana de menor demanda.

Alcance: Datos del total de viajes mensuales, discriminados por distrito.

- KPI\_5

Definición: Índice de Variación de la tasa de uso de vehículos con respecto del año anterior.

Fórmula: para calcular este KPI primero debemos definir la tasa de uso de vehículos.

Definimos la tasa de uso de vehículos como el número de horas de servicio dividido el total de horas disponibles, es decir, 24 horas por cada vehículo:

$$TUV = \frac{N^{\circ}\ total\ de\ horas\ de\ servicio}{24 \times N^{\circ}\ de\ veh\acute{iculos}}$$

Luego, la fórmula para el KPI es:

$$\frac{TUV_{año\ actual} - TUV_{año\ anterior}}{TUV_{año\ anterior}} \times 100$$



Objetivo: Mantenerse por encima del promedio del año anterior.

Alcance: calculado de forma anual, para todos los tipos de servicio y la ciudad completa.

## Modelo de Machine Learning

El modelo de machine learning propuesto tiene como objetivo predecir la demanda de taxis en la ciudad de Nueva York, discriminada por distrito, utilizando diversas variables temporales y meteorológicas como datos de entrada. A continuación, se detalla el proceso de diseño, implementación y evaluación de este modelo:

### Variables de Entrada (X):

Fecha

Hora (Franja horaria)

Temperatura (°C)

Lluvia (mm)

Nieve (cm)

Humedad relativa (porcentaje)

### Datos de Salida (Y):

La demanda de vehículos por distrito de Nueva York, es decir, la cantidad de viajes por franja horaria en cada distrito.

### Modelos a evaluar:

- Redes Neuronales Recurrentes (LSTM)
- ARIMA (Autoregressive Integrated Moving Average)
- Regresión Lineal Múltiple
- Regresión Polinomial

### Entrenamiento del Modelo:

El modelo será entrenado utilizando datos históricos de uno o dos años, que contienen todas las variables de entrada (X) y la variable de salida (Y), la cual es un vector de 5 componentes que indica la cantidad de viajes en cada distrito por fecha y franja horaria.

### Métricas de Evaluación:

Para evaluar el rendimiento del modelo, se utilizarán las siguientes métricas:

- Error Cuadrático Medio (MSE)
- Error Absoluto Medio (MAE)

### Implementación y Evaluación:

El modelo se implementará utilizando el lenguaje de programación Python y la biblioteca TensorFlow o PyTorch para las redes neuronales recurrentes (LSTM), así como las bibliotecas statsmodels y scikit-learn para ARIMA y regresiones lineales múltiples y polinomiales respectivamente. Se realizará una evaluación exhaustiva del modelo utilizando técnicas de validación cruzada y se ajustarán los hiperparámetros para maximizar su rendimiento predictivo.

Se espera que el modelo de machine learning propuesto sea capaz de proporcionar predicciones precisas de la demanda de taxis en Nueva York por distrito, lo que permitirá a la empresa tomar decisiones informadas y estratégicas para mejorar la eficiencia y la calidad del servicio en la ciudad. El uso de diferentes modelos y métricas de evaluación garantizará la robustez y la precisión de las predicciones obtenidas.

## Solución propuesta y metodología de trabajo

La solución propuesta consiste en realizar un análisis integral de los datos de servicios de taxis en la ciudad de Nueva York, abordando aspectos como la demanda histórica y geográfica, la duración de los viajes y los costos asociados. Además, se evaluará el impacto ambiental de una flota de taxis sostenibles por distrito, comparando condiciones técnicas y costos de vehículos eléctricos. Se desarrollarán KPIs específicos y un dashboard personalizado para medir el éxito de la introducción de la flota sostenible, junto con un modelo predictivo de machine learning para optimizar la asignación de recursos. Finalmente, se llevará a cabo el despliegue de todos los componentes del proyecto en una plataforma en la nube para garantizar su automatización, escalabilidad y disponibilidad.

La metodología de trabajo elegida para el proyecto es Scrum, una metodología ágil para la gestión de proyectos complejos, especialmente aquellos que requieren una entrega rápida de valor al cliente y se benefician de la colaboración y la comunicación entre los miembros del equipo.

Características:

Se divide el proyecto en espacios de tiempo cortos de trabajo (sprints):

- Sprint1: Del 11/3/24 al 22/3/24
- Sprint2: Del 23/3/24 al 12/4/24
- Sprint3: Del 13/4/24 al 26/4/24

En cada sprint se realizan las siguientes actividades organizadas en un cronograma Gantt:

<b>ASET Advisors on the Sustainable Economic Transition</b> AC = Alter Caimi RB = Rafael Balestrini CM = Carlos Macea LR = Luis Rojas			
Proyecto ECOTaxis NYC			
Actividad / Tarea	Inicio	Fin	Responsable
<b>Entender la situación actual</b>	11/3/2024	17/3/2024	
Elaborar el EDA de los datos	11/3/2024	17/3/2024	
Air Quality NYC	11/3/2024	17/3/2024	AC
CO2 Emissions NYC	11/3/2024	17/3/2024	AC
Historical weather API	11/3/2024	17/3/2024	AC
Taxi zone NYC	11/3/2024	17/3/2024	AC
TLC Trip Record data	11/3/2024	17/3/2024	LR
SONYC	11/3/2024	17/3/2024	RB
Electric Car Data	11/3/2024	17/3/2024	CM
Medallion Vehicles authorized	11/3/2024	17/3/2024	CM
<b>Elaborar la propuesta</b>	13/3/2024	20/3/2024	
Establecer los objetivos	13/3/2024	18/3/2024	CM
Delimitar el alcance	15/3/2024	19/3/2024	RB
Plantear los KPIs	17/3/2024	20/3/2024	LR
Plantear el modelo de Machine Learning	17/3/2024	20/3/2024	AC
<b>Crear el repositorio GitHub</b>	11/3/2024	12/3/2024	RB
<b>Detallar la solución propuesta</b>	13/3/2024	20/3/2024	

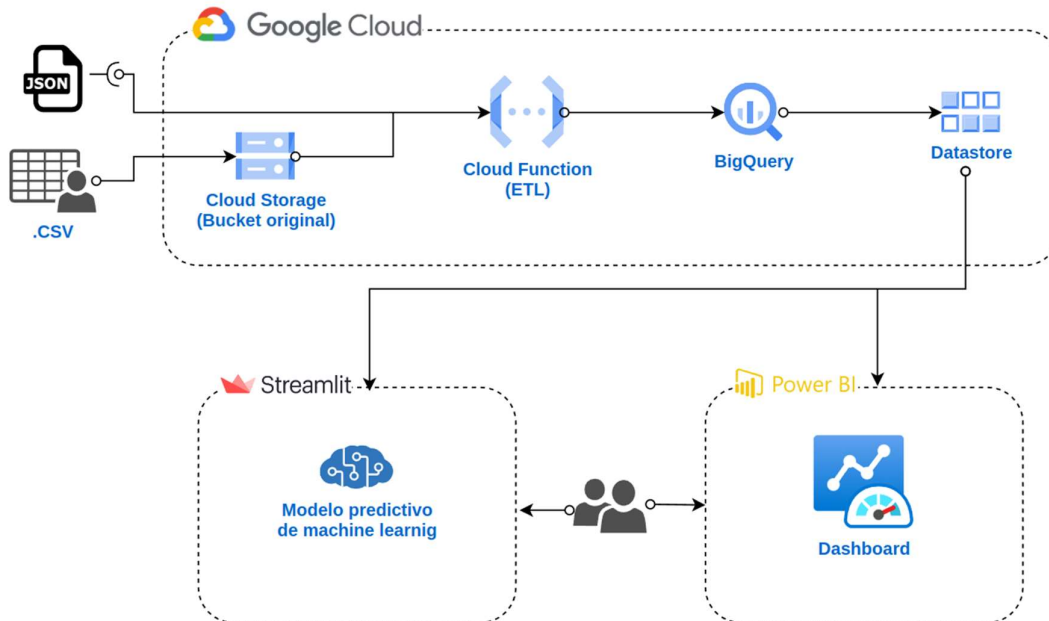
Definir roles y responsabilidades del equipo	13/3/2024	13/3/2024	CM
Establecer la metodología de trabajo	14/3/2024	16/3/2024	RB
Elaborar el cronograma general - Gantt	15/3/2024	16/3/2024	AC
Describir stack tecnológico	16/3/2024	18/3/2024	RB
Resumen de los EDA	19/3/2024	20/3/2024	LR
Flujo de trabajo	19/3/2024	20/3/2024	
<b>Análisis de metadatos</b>	19/3/2024	20/3/2024	LR
<b>Diseño Presentación Sprint 1</b>	18/3/2024	22/3/2024	LR
<b>ETL</b>	23/3/2024	26/3/2024	
Air Quality NYC	23/3/2024	26/3/2024	AC
CO2 Emissions NYC	23/3/2024	26/3/2024	AC
Historical weather API	23/3/2024	26/3/2024	AC
Taxi zone NYC	23/3/2024	26/3/2024	AC
TLC Trip Record data	23/3/2024	26/3/2024	LR
SONYC	23/3/2024	26/3/2024	RB
Electric Car Data	23/3/2024	26/3/2024	CM
Medallion Vehicles authorized	23/3/2024	26/3/2024	CM
<b>Implementación Estructura de Datos</b>	23/3/2024	3/4/2024	RB - CM
<b>Pipeline ETL automatizado</b>	27/3/2024	3/4/2024	RB - CM
<b>Pipelines para alimentar el DW</b>	27/3/2024	3/4/2024	RB - CM
<b>Validación de datos</b>	27/3/2024	3/4/2024	AC - LR
<b>Documentación</b>	4/4/2024	8/4/2024	
Diagrama ER detallado	4/4/2024	8/4/2024	RB
Diccionario de Datos	4/4/2024	8/4/2024	AC-RB-CM-LR
Workflow detallando tecnologías	4/4/2024	8/4/2024	AC - LR
<b>Análisis de datos de Muestra</b>	5/4/2024	10/4/2024	RB - CM
<b>MVP Dashboard</b>	7/4/2024	12/4/2024	AC - LR
<b>MVP Modelo ML</b>	7/4/2024	12/4/2024	AC - LR
<b>Diseño Presentación Sprint 2</b>	9/4/2024	12/4/2024	LR
<b>Diseño de Reportes / Dashboards</b>	13/4/2024	20/4/2024	
Desarrollo en Power BI	13/4/2024	16/4/2024	LR
Implementación de KPI's	13/4/2024	16/4/2024	LR
Implementación en Streamlit	16/4/2024	20/4/2024	LR - AC
<b>Análisis y comparación de Modelos de ML</b>	13/4/2024	20/4/2024	AC
ARIMA	13/4/2024	20/4/2024	AC
RNN (LSTM)	13/4/2024	20/4/2024	AC
Regresión Lineal Múltiple	13/4/2024	15/4/2024	AC
Regresión Polinomial	13/4/2024	15/4/2024	AC
<b>Modelo de ML en producción</b>	20/4/2024	26/4/2024	
ETL en la nube	20/4/2024	23/4/2024	RB - CM
Implementación en Streamlit	20/4/2024	23/4/2024	AC - LR
<b>Documentación</b>	23/4/2024	26/4/2024	
Sellección del Modelo + Feature Engineering	23/4/2024	26/4/2024	AC
Informe de Análisis	23/4/2024	26/4/2024	AC
<b>Video Presentación del Proyecto</b>	23/4/2024	26/4/2024	CM
<b>Diseño Presentación de Sprint 3</b>	23/4/2024	26/4/2024	LR

Herramientas de organización:

- GitHub como repositorio central y administrador de versiones de los archivos que conforman el proyecto.
- Presentación de Google como pizarra colaborativa para el equipo que lleva a cabo el proyecto.
- Hoja de cálculo de Google (Diagrama de Gantt) que hace las veces del tablero Scrum.
- Sprint Backlog: Lista de tareas que se completarán en cada sprint.
- Daily Scrum: Reunión los lunes, miércoles y viernes del equipo para sincronizar el trabajo.

## Stack tecnológico

Función en la nube para agregar CSV/JSON para actualizaciones del almacén de datos sin servidor



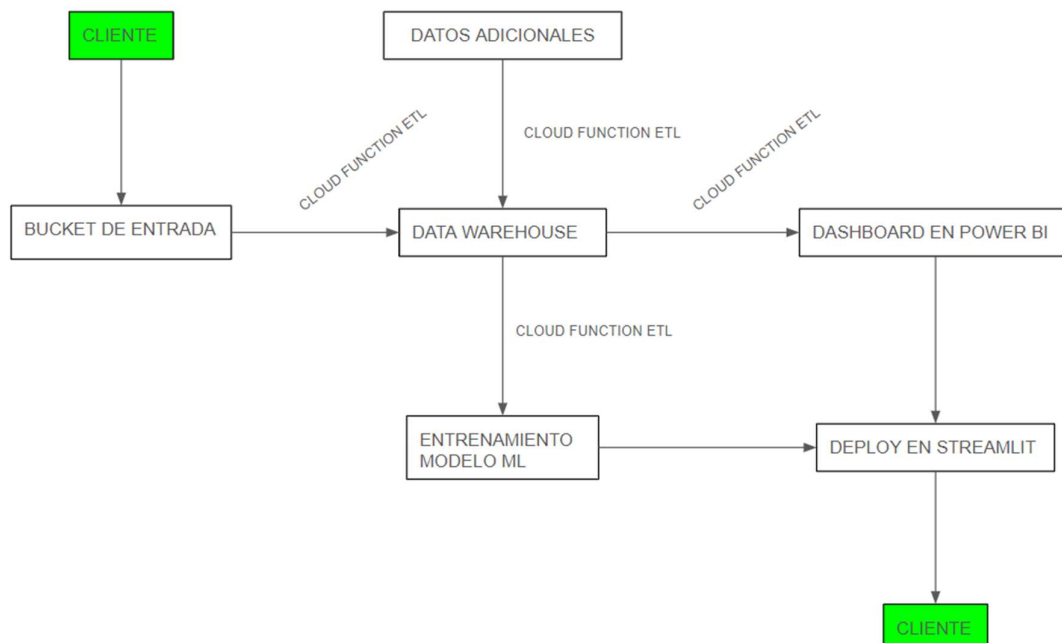
El proyecto se basará en un robusto conjunto de herramientas tecnológicas diseñadas para maximizar la eficiencia y la precisión en el análisis de datos y la implementación de soluciones de machine learning. A continuación, describimos cada una de las herramientas que conforman nuestro stack tecnológico:

- Python: Utilizaremos Python como lenguaje de programación principal debido a su amplia gama de bibliotecas especializadas en análisis de datos y machine learning, como pandas, NumPy, scikit-learn y TensorFlow.
- Jupyter Notebook: Emplearemos Jupyter Notebook como entorno interactivo para la ejecución de código y el análisis exploratorio de datos. Esta herramienta nos permite crear y compartir documentos que contienen código, visualizaciones y texto explicativo de manera integrada.
- Power BI: Para la creación de dashboards interactivos y visualizaciones dinámicas, utilizaremos Power BI. Esta herramienta nos permitirá visualizar de manera intuitiva los datos procesados y generar insights significativos para la toma de decisiones.

- Google Cloud Platform (GCP): Aprovecharemos la infraestructura escalable y confiable de GCP para almacenar, procesar y analizar grandes volúmenes de datos en la nube. Utilizaremos servicios como Cloud Storage para almacenar los datasets, Cloud Dataflow para realizar el procesamiento de datos en tiempo real y BigQuery como nuestro data warehouse en la nube.

- Streamlit: Para la creación de aplicaciones web interactivas de análisis de datos, utilizaremos Streamlit. Esta herramienta nos permitirá desplegar tanto el dashboard final generado en Power BI como el modelo de machine learning en producción de manera sencilla y eficiente.

El flujo de datos comenzará con la carga de los datasets por parte del cliente en un bucket de Cloud Storage en GCP. Luego, se realizará un proceso de extracción, transformación y carga (ETL) a través de una Cloud Function para combinar esta información con datos obtenidos de API's externas y almacenarla en nuestro data warehouse en BigQuery. Posteriormente, se ejecutarán ETL adicionales para cargar la información en Power BI y proporcionar los datos de entrenamiento necesarios para el modelo de machine learning. Finalmente, tanto el modelo como el dashboard serán publicados en Streamlit, permitiendo su acceso y visualización desde cualquier navegador web. Este enfoque garantiza la centralización, la automatización y la disponibilidad de los datos y las soluciones analíticas en todo momento.





## Datos a utilizar

### - Contaminación sonora en NYC:

Origen de los datos: Dataset almacenado en un .csv, “SONY Urban Sound Tagging.csv”. Disponible en <https://zenodo.org/records/3966543>

Contiene la cantidad de sonidos de motores pequeños, medianos y grandes, en NYC, por fecha, distrito y horario.

Campo	Tipo de dato	Descripción
borough	int64	Distrito de NYC
block	int64	El bloque de NYC en el que se encuentra el sensor. Dígitos del 2 al 6 en el sistema de números de parcelas de NYC de 10 dígitos conocido como BBL.
latitude	float64	Latitud
longitude	float64	Longitud
year	int64	Año
week	int64	Semana
day	int64	Día
hour	int64	Hora
1-1_small-sounding-engine_presence	int64	Presencia de sonido de motor pequeño
1-2_medium-sounding-engine_presence	int64	Presencia de sonido de motor mediano
1-3_large-sounding-engine_presence	int64	Presencia de sonido de motor grande

### - Servicios de taxis amarillos y verdes

Origen de datos: <https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page> descargados en formato parquet.

Contiene los datos de viajes de taxis amarillos, verdes y alquiler de gran volumen (horario de inicio, de finalización, ubicación de inicio, ubicación de finalización, importes abonados, etc). Cada archivo descargado contiene los datos de un tipo de servicio durante un mes de viajes.

Campo	Tipo de dato	Descripción
VendorID	int32	Un código que indica el proveedor de TPEP que proporcionó el registro. 1= Tecnologías móviles creativas, LLC; 2= VeriFone Inc.
tpep_pickup_datetime	datetime64[us]	La fecha y hora en que se activó el medidor.
tpep_dropoff_datetime	datetime64[us]	La fecha y hora en que se desconectó el medidor.

passenger_count	float64	Cantidad de pasajeros
trip_distance	float64	Distancia recorrida en millas reportada por el taxímetro
RatecodeID	float64	El código de tarifa final vigente al finalizar el viaje. 1= Tarifa estándar 2=JFK 3=Newark 4=Nasáu o Westchester 5=Tarifa negociada 6=viaje en grupo
store_and_fwd_flag	object	Esta bandera indica si el registro de viaje se almacenó en la memoria del vehículo antes de enviar al proveedor, también conocido como "almacenar y reenviar", porque el vehículo no tenía conexión con el servidor. Y = sí N= no
PULocationID	int32	Zona de Taxi TLC en la que se activó el taxímetro
DOLocationID	int32	Zona de Taxi TLC en la que se desactivó el taxímetro
payment_type	int64	Un código numérico que indica cómo se pagó el viaje. 1= tarjeta de crédito 2= Efectivo 3= Sin cargo 4= Disputa 5= Desconocido 6= Viaje anulado
fare_amount	float64	La tarifa de tiempo y distancia calculada por el taxímetro.
extra	float64	Extras y recargos varios. Actualmente, esto sólo incluye los cargos de \$0,50 y \$1 por hora pico y por noche.
mta_tax	float64	Impuesto MTA de \$0.50 que se activa automáticamente.
tip_amount	float64	Monto de propina.
tolls_amount	float64	Monto total de todos los peajes pagados durante el viaje
improvement_surcharge	float64	Recargo por mejora de \$0.30 al inicio del recorrido. El recargo por mejora comenzó a cobrarse en 2015.
total_amount	float64	Monto total cobrado a los pasajeros. No incluye propinas en efectivo
congestion_surcharge	float64	Monto total recaudado en el viaje por el recargo de congestión del estado de NY.
Airport_fee	float64	\$ 1,25 por recogida solo en los aeropuertos LaGuardia y JF Kennedy

- Servicios de vehículos de alquiler de gran volumen

Origen de los datos: archivo parquet descargable en: <https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>.

En NYC, el término "For-Hire Vehicles" (FHV) se refiere a vehículos de alquiler que brindan servicios de transporte de pasajeros a pedido, pero que no son taxis amarillos o verdes tradicionales. Cuando se refiere a servicios de gran volumen, se refiere a una categoría de licencia de TLC para negocios FHV que actualmente despacha o planea despachar más de 10,000 viajes FHV por día en NYC bajo una marca única, como Uber y Lyft.

Los FHV operan bajo diferentes regulaciones y modelos de negocio en comparación con los taxis tradicionales. Por lo general, los pasajeros solicitan un viaje a través de una aplicación móvil o por teléfono, y el vehículo de alquiler asignado recoge al pasajero en una ubicación específica y lo lleva a su destino.

Este sector del transporte ha experimentado un crecimiento significativo en los últimos años, proporcionando a los usuarios más opciones de movilidad y mayor flexibilidad en comparación con los servicios de taxi convencionales. Sin embargo, también ha generado debates sobre cuestiones de regulación, competencia y trabajo.

Campo	Tipo de dato	Descripción
Hvfhs_license_num	object	El número de licencia TLC del negocio de HVFHS. <ul style="list-style-type: none"> <li>• HV0002: Juno</li> <li>• HV0003: Uber</li> <li>• HV0004: Vía</li> <li>• HV0005: Lyft</li> </ul>
Dispatching_base_num	object	El número de licencia base TLC de la base que despachó el viaje.
Pickup_datetime	datetime64[us]	La fecha y hora de recogida del viaje.
DropOff_datetime	datetime64[us]	La fecha y hora de finalización del viaje.
PULocationID	int32	Zona de Taxis TLC en la que inició el viaje
DOLocationID	int32	Zona de Taxis TLC en la que finalizó el viaje
originating_base_num	object	Número de base de la base que recibió la solicitud de viaje original
request_datetime	datetime64[us]	fecha/hora en que el pasajero solicitó ser recogido
on_scene_datetime	datetime64[us]	fecha/hora en que el conductor llegó al lugar de recogida
trip_miles	float64	millas totales por viaje de pasajero
trip_time	int64	tiempo total en segundos para el viaje del pasajero
base_passenger_fare	float64	tarifa base de pasajero antes de peajes, propinas, impuestos y tarifas
tolls	float64	importe total de todos los peajes pagados en el viaje
bcf	float64	Monto total recaudado en viaje para Black Car Fund (BCF)
sales_tax	float64	monto total recaudado en el viaje por el impuesto sobre las ventas del estado de Nueva York
congestion_surcharge	float64	Monto total recaudado en el viaje por el recargo de congestión del estado de NY.

airport_fee	float64	\$2.50 para dejar y recoger en aeropuertos LaGuardia, Newark y John F. Kennedy
tips	float64	cantidad total de propinas recibidas del pasajero
driver_pay	float64	pago total del conductor (sin incluir peajes ni propinas y neto de comisión, recargos o impuestos)
shared_request_flag	object	¿Aceptó el pasajero un viaje compartido o conjunto, independientemente de si eran coincidentes? (sí/no)
shared_match_flag	object	¿El pasajero compartió el vehículo con otro pasajero reservado por separado en algún momento durante el viaje? (sí/no)
access_a_ride_flag	object	¿El viaje fue despachado en nombre de la Autoridad de Transporte Metropolitano (MTA)? (sí/no)
wav_request_flag	object	¿El pasajero solicitó un vehículo accesible para sillas de ruedas (WAV)? (sí/no)
wav_match_flag	object	¿El viaje se realizó en un vehículo accesible para sillas de ruedas (WAV)? (sí/no)

#### - Emisiones de CO<sub>2</sub> en NYC

Origen de los datos: archivo csv descargado de <https://climate.cityofnewyork.us/initiatives/nyc-greenhouse-gas-inventories/>.

Contiene las emisiones de CO<sub>2</sub> en NYC, clasificadas por sector y categoría, desde 2005 a 2022 inclusive. Los datos no están discriminados por distrito. Los datos fueron filtrados para quedarnos únicamente con las emisiones de sector transporte, y categoría "on-road", subcategoría autos de pasajeros.

Campo	Tipo de dato	Descripción
Año	string	Año de medición
Emisiones	float64	Emisiones de CO <sub>2</sub> en toneladas
Millas recorridas	float64	Millas recorridas por autos de pasajeros en NYC

#### - Medallion

Origen de los datos: csv descargado de [https://www.google.com/url?q=https://data.cityofnewyork.us/Transportation/Medallion-Vehicles-Authorized/rhe8-mgbb/about\\_data&sa=D&source=editors&ust=1711030103949688&usg=AOvVaw3i92dga7eM4-jXcTTMQ71J](https://www.google.com/url?q=https://data.cityofnewyork.us/Transportation/Medallion-Vehicles-Authorized/rhe8-mgbb/about_data&sa=D&source=editors&ust=1711030103949688&usg=AOvVaw3i92dga7eM4-jXcTTMQ71J).

Los "medallion" de los taxis en la ciudad de Nueva York son licencias especiales que otorgan a los propietarios de taxis el derecho exclusivo de recoger pasajeros en la calle (o "street hails"). Estas licencias funcionan como un permiso para operar un taxi en la ciudad y son emitidas por la Comisión de Taxis y Limusinas de la Ciudad de Nueva York (TLC, por sus siglas en inglés).

Este dataset incluye todos los vehículos con licencia de TLC que están al día y pueden conducir.

Campo	Tipo de dato	Descripción
License Number	int64	Número de Licencia del 'medallion' del taxi
Name	object	Nombre del Propietario de la licencia
Expiration Date	object	Fecha de Vencimiento
DMV License Plate Number	object	Número de Placa de Vehículo Registrado en el DMV
Vehicle VIN Number	object	Número VIN del Vehículo Registrado en el DMV
Vehicle Type	object	Tipo de Vehículo - Híbrido o Limusina
Model Year	int64	Año del Modelo del Taxi
Medallion Type	object	Tipo de licencia, ya sea Individual (independiente) o Múltiple (flota)
Agent Number	float64	Número de Agente Designado por TLC
Agent Name	object	Nombre del Agente de la licencia
Agent Telephone Number	object	Número de Teléfono del Agente de la licencia
Agent Website Address	float64	Dirección del Sitio Web del Agente de la licencia
Agent Address	object	Dirección Permanente del Agente de la licencia
Last Date Updated	object	Fecha y Hora de Creación del dataset
Last Time Updated	object	Hora de Creación del dataset

#### - Datos de autos eléctricos

Origen de los datos: dataset provisto en [https://drive.google.com/file/d/1znjBC1zmJ\\_s2nsLmxSlyLwkcZLFGEd/view?usp=drive\\_link](https://drive.google.com/file/d/1znjBC1zmJ_s2nsLmxSlyLwkcZLFGEd/view?usp=drive_link)

Este conjunto de datos sobre coches eléctricos incluye varias especificaciones como la marca, el modelo, la aceleración (en segundos), la velocidad máxima (Km/h), el alcance (Km), la eficiencia (Wh/Km), la velocidad de carga rápida (Km/h), si cuenta con carga rápida, el tipo de tren motriz, el tipo de enchufe, el estilo de carrocería, el segmento, el número de asientos y el precio en euros. Abarca 103 entradas con detalles completos sobre diferentes vehículos eléctricos disponibles en el mercado

Campo	Tipo de dato	Descripción
Brand	object	Marca
Model	object	Modelo
AccelSec	float64	Aceleración de 0 a 100 km/h en segundos
TopSpeed_KmH	int64	Velocidad máxima
Range_Km	int64	Rango en km
Efficiency_WhKm	int64	Eficiencia en watts por km
FastCharge_KmH	object	Velocidad de carga rápida
RapidCharge	object	Carga Rápida
PowerTrain	object	Tren de potencia
PlugType	object	Tipo de conector

BodyStyle	object	Tipo de carrocería
Segment	object	Segmentación por tamaño
Seats	int64	Asientos
PriceEuro	int64	Precio en euros

#### - Zonas de taxis

Origen de los datos: dataset provisto en [https://drive.google.com/file/d/1uKIP1StcjugMcM5oTywpAiJSGl4qtfr2/view?usp=drive\\_link](https://drive.google.com/file/d/1uKIP1StcjugMcM5oTywpAiJSGl4qtfr2/view?usp=drive_link)

Este conjunto de datos contiene las Zonas de Taxis de Nueva York, que corresponden a las zonas de recogida y entrega, o Identificadores de Ubicación (LocationIDs), incluidos en los Registros de Viajes de Taxis Amarillos, Verdes y de Vehículos de Alquiler con Conductor (FHV, por sus siglas en inglés) publicados en el sitio de Datos Abiertos de NYC.

Campo	Tipo de dato	Descripción
OBJECTID	int64	Columna de identificación única para cada entrada en el conjunto de datos.
Shape_Leng	float64	Representa la longitud de la forma geoespacial
Shape_Area	float64	Indica el área de la forma geoespacial.
zone	object	Nombre de la zona a la que pertenece cada registro.
LocationID	int64	Identificador de la zona
borough	object	Distrito de NYC
geometry	geometry	Columna vacía

#### - Calidad de aire:

Origen de los datos: dataset descargado de: <https://www.google.com/url?q=https://data.cityofnewyork.us/Environment/Air-Quality/c3uy-2p5r&sa=D&source=editors&ust=1711033064431771&usg=AOvVaw0m6W1L1ELHCQVln82jFrAH>.

El conjunto de datos contiene información sobre los datos de vigilancia de la calidad del aire de la ciudad de Nueva York.

Campo	Tipo de dato	Descripción
Unique ID	int64	Identificador de registro único
Indicator ID	int64	Identificador del tipo de valor medido en el tiempo y el espacio.
Name	object	Nombre del indicador
Measure	object	Indica cómo es medido el indicador
Measure Info	object	Información sobre la medida (por ejemplo, unidades)

Geo Type Name	object	Tipo de geografía; Por ejemplo, la Ciudad, los Distritos de Borough y Comunitarios son diferentes tipos de geografía.
Geo Join ID	int64	Identificador del área geográfica del vecindario, utilizado para unir archivos de geografía para crear mapas temáticos.
Geo Place Name	object	Nombre del vecindario
Time Period	object	Período de tiempo
Start_Date	object	Inicio del período
Data Value	float64	Valor del dato
Message	float64	Mensaje (columna vacía)

#### - Datos climáticos

Origen de los datos: API de condiciones climáticas históricas <https://open-meteo.com/en/docs/historical-weather-api>.

Los datos accedidos a través de la API son temperatura, humedad, lluvia y nieve, a nivel franja horaria, para NYC, desde el 01/01/2009 hasta 31/12/2023.

Campo	Tipo de dato	Descripción
date	datetime64[ns, UTC]	Fecha y hora de la medición
temperature_2m	float32	Temperatura del aire a 2 metros del suelo.
relative_humidity_2m	float32	Humedad relativa a 2 metros del suelo.
rain	float32	Sólo precipitaciones líquidas de la hora anterior, incluidos chubascos locales y lluvias de sistemas a gran escala. En mm.
snowfall	float32	Cantidad de nevada de la hora anterior en centímetros.