



# Analytica Data Solutions

TRANSFORMANDO DATOS EN DECISIONES

## Proyecto de Movilidad Sostenible NYC

### INTEGRANTES:

**María Marcela Balzarelli**

**Pablo Nahuel Barchiesi Ponce**

**Michael Williams Martinez Chinchilla**

**Jorgelina Paola Lujan Ramos**

# Tabla de Contenido

1. [Entendimiento de la situación propuesta](#)
2. [Objetivos del trabajo y del grupo](#)
3. [Alcance y fuera de alcance](#)
4. [Análisis preliminar de Calidad del dato](#)
5. [Kpi y métricas asociadas \(planteó\)](#)
6. [Solución Propuesta](#)
7. [Metodología Propuesta](#)
8. [Definición del stack tecnológico](#)
9. [Equipo de trabajo \(Roles y responsabilidades\)](#)
10. [Creación del repositorio en Github](#)



**Analytica Data Solutions**

TRANSFORMANDO DATOS EN DECISIONES

## Proyecto Movilidad Sostenible NYC

### 1. Entendimiento de la situación propuesta:

La empresa de servicios de transporte de pasajeros desea invertir en el sector con automóviles en la ciudad de Nueva York. Para ello, contrata a Analytica Data Solutions como Consultores externos para encontrar soluciones innovadoras, en donde se estudiará la relación entre estos medios de transporte particulares, la calidad del aire, la contaminación sonora y correlaciones climáticas, con el objetivo de considerar la posibilidad de implementar vehículos eléctricos en su flota. Se realizará un análisis preliminar del movimiento de los taxis en la ciudad para obtener un marco de referencia y tomar decisiones bien fundamentadas.

El proyecto implica recopilar, depurar y disponibilizar información relevante de diferentes fuentes para analizar la relación entre el transporte de pasajeros con automóviles en Nueva York, la calidad del aire, la contaminación sonora y correlaciones climáticas. A través de reportes, dashboards y el entrenamiento de un modelo de machine learning de clasificación, se resolverá un problema específico relacionado con los objetivos del proyecto.

### 2. Objetivos: específicos del trabajo y del grupo

#### **Específicos del trabajo:**

- Recopilar y depurar datos de diferentes fuentes para crear una base de datos (DataWarehouse).
- Realizar un análisis exploratorio de los datos para encontrar relaciones
- Crear un dashboard interactivo y visualmente atractivo que integre los resultados del análisis exploratorio de datos

- Entrenar y poner en producción un modelo de machine learning para resolver el problema de inversión en el sector.

#### **Objetivos del grupo:**

- Investigar y analizar los conjuntos de datos de taxis y viajes compartidos en Nueva York para identificar patrones y tendencias de movimiento.
- Entender la correlación entre los datos de viajes y las métricas ambientales con el fin de proporcionar una base sólida para las decisiones futuras relacionadas con la flota de vehículos.
- Proveer información relevante y confiable a la empresa para respaldar su toma de decisiones sobre la implementación de vehículos eléctricos en su flota.
- Contribuir a la visión de un futuro menos contaminado y ajustarse a las tendencias de mercado actuales.

### **3. Alcance y fuera de alcance**

#### **Alcance**

- La recopilación, limpieza y análisis de datos de taxis y viajes compartidos para extraer información relevante.
- El análisis exploratorio de datos para identificar patrones preliminares y tendencias en los datos recopilados.
- La correlación entre los datos de transporte y las métricas de calidad del aire, contaminación sonora y factores climáticos.
- Implementación de métricas e indicadores para evaluar diferentes aspectos del transporte de pasajeros y su relación con la contaminación ambiental y sonora.
- Se requerirá la incorporación de datasets adicionales que contengan información relevante y confiable.

#### **Fuera de alcance:**

- La implementación real de vehículos eléctricos en la flota de la empresa. El proyecto se enfoca en proporcionar información y análisis para respaldar esta decisión, pero la implementación concreta queda bajo responsabilidad de la empresa.
- Análisis exhaustivos de aspectos económicos, logísticos y regulatorios asociados con la transición a vehículos eléctricos.

### **4. Análisis preliminar de Calidad del dato**

Se realizó un análisis preliminar de los 8 dataset entregados por la empresa para observar la calidad de los datos, en el cual se cargaron, se obtuvo número de filas, columnas, las primeras 5 filas de cada dataset, nombre de columnas, cantidad de valores no nulos y tipo de datos.

Podemos observar el archivo en el repositorio en la carpeta sprint\_1 llamado el archivo Eda\_preliminar.

## 5. KPIs y métricas asociadas (planteó) :

### 1. KPI: Tarifa del servicio de transporte.

Objetivo: Demostrar la ganancia neta proveniente de tarifas en relación con los costos de operación, comparando los últimos dos años.

Análisis: Evaluación económica del servicio de transporte.

### 2. KPI: Nivel de contaminación sonora de autos a combustión vs. autos eléctricos.

Objetivo: Indicar el nivel porcentual de ruido generado por vehículos a combustión de un año respecto a otro, destacando las ventajas de los vehículos eléctricos en términos de reducción de la contaminación sonora.

Análisis: Comparar el nivel de ruido generado por vehículos a combustión y vehículos eléctricos durante el período actual.

### 3. KPI: Histórico de Emisión de CO2

Objetivo: Monitorear y analizar las emisiones de CO2 generadas por las operaciones de las empresas de transporte de vehículos eléctricos y de combustión en el último año en comparación con un período de referencia histórico.

Análisis: Comparar las métricas de emisiones del período actual con las del período de referencia histórico.

### 4. KPI: Precio de vehículos eléctricos vs. velocidad de carga.

Objetivo: Evaluar la relación entre el precio del vehículo eléctrico y la eficiencia y rapidez de su sistema de carga para comprender el valor que ofrecen en términos de carga rápida y autonomía.

Análisis: Evaluar la relación entre el precio del vehículo eléctrico y la eficiencia y rapidez de su sistema de carga para comprender el valor que ofrecen en términos de carga rápida y autonomía.

## 5. KPI: Días y horarios con mayor demanda de viajes

Objetivo: Identificar los días y horarios con mayor demanda de viajes para satisfacer las necesidades de los usuarios de manera más eficiente.

Análisis: Comprender las pautas de uso y los patrones de comportamiento de los usuarios en relación con el servicio de transporte.

## 6. KPI: Ciudad con mayor/menor cantidad de viajes o recorridos.

Objetivo: Obtener una visión clara de las áreas de la ciudad con la mayor y menor cantidad de viajes realizados.

Análisis: Proporciona información sobre las áreas de la ciudad con mayor y menor demanda de transporte.

# 6. Solución propuesta

Teniendo en cuenta que la empresa es la encargada de tomar la última decisión en administrar sus recursos, se propone los siguientes pasos para que pueda invertir en el sector de automóviles.

- Creación de dashboards y reportes visuales para presentar la información de manera clara y concisa a quienes toman decisiones en la empresa.
- Desarrollar un modelo de machine learning para predecir la viabilidad de implementar vehículos eléctricos en la flota.
- Generar un plan de acción sugerido de acuerdo a las interpretaciones del estudio, con medidas estratégicas y operativas si es necesario.

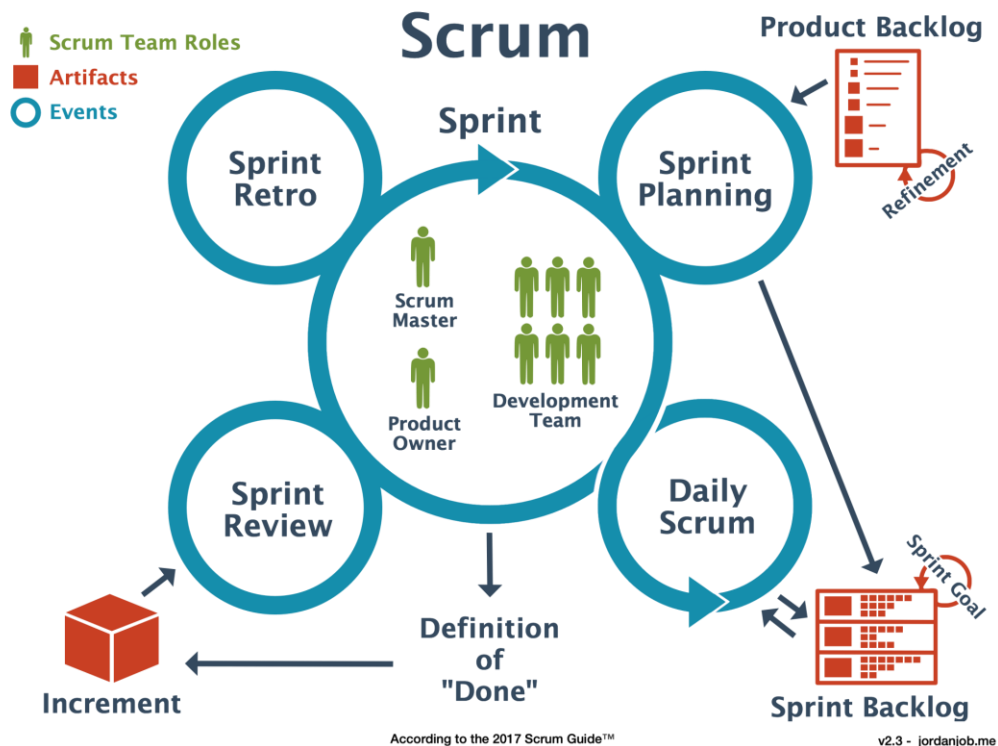
## Medidas estratégicas

- Realizar alianzas con fabricantes de vehículos eléctricos para adquirir una flota inicial de automóviles eléctricos.
- Establecer puntos de recarga eléctrica en ubicaciones clave de la ciudad para facilitar la transición a vehículos eléctricos.
- Promover campañas de concienciación entre los pasajeros y conductores sobre los beneficios ambientales y económicos de los vehículos eléctricos.

## Medidas operativas

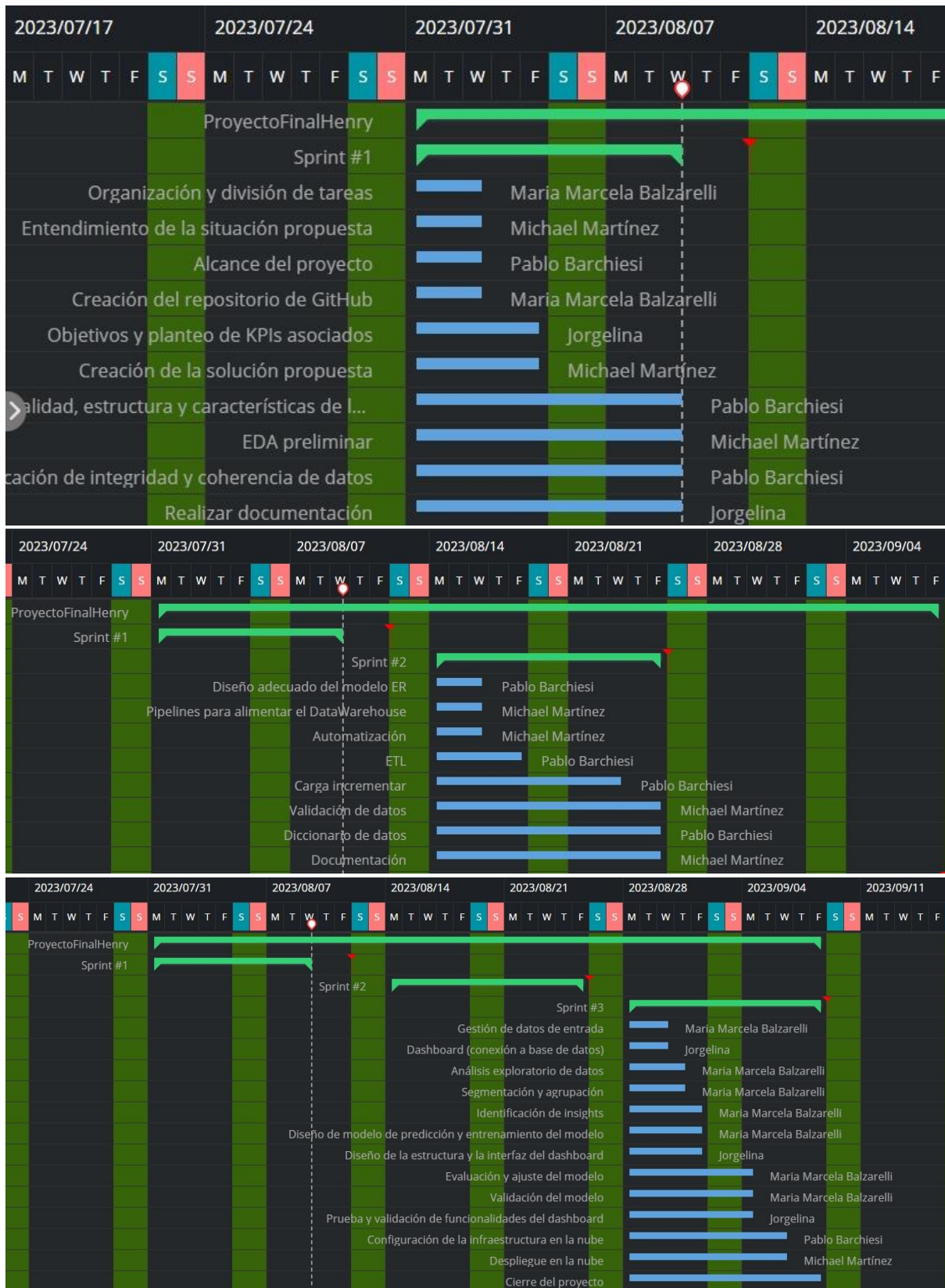
- Realizar un análisis de costos y beneficios para determinar la viabilidad económica de la transición a vehículos eléctricos.
- Monitorear regularmente la calidad del aire y la contaminación sonora en diferentes áreas de la ciudad para evaluar el impacto de los vehículos eléctricos.

## 7. Metodología propuesta



En este proyecto, estamos empleando la metodología ágil mediante el marco de trabajo Scrum utilizando el software de administración de proyectos Jira. Esto conlleva la incorporación de los valores y conceptos ágiles en nuestro enfoque de desarrollo. Estamos aplicando la estructura definida por Scrum en términos de roles, eventos, artefactos y reglas para la organización y gestión efectiva del trabajo colaborativo.

### Cronograma general (Gantt)



[https://marcebalzarelli-1690896078560.atlassian.net/plugins/servlet/ac/jp.ricksoft.plugins.wbsganttt-for-jira/main?ac.filter=GANTT\\_-10000](https://marcebalzarelli-1690896078560.atlassian.net/plugins/servlet/ac/jp.ricksoft.plugins.wbsganttt-for-jira/main?ac.filter=GANTT_-10000)

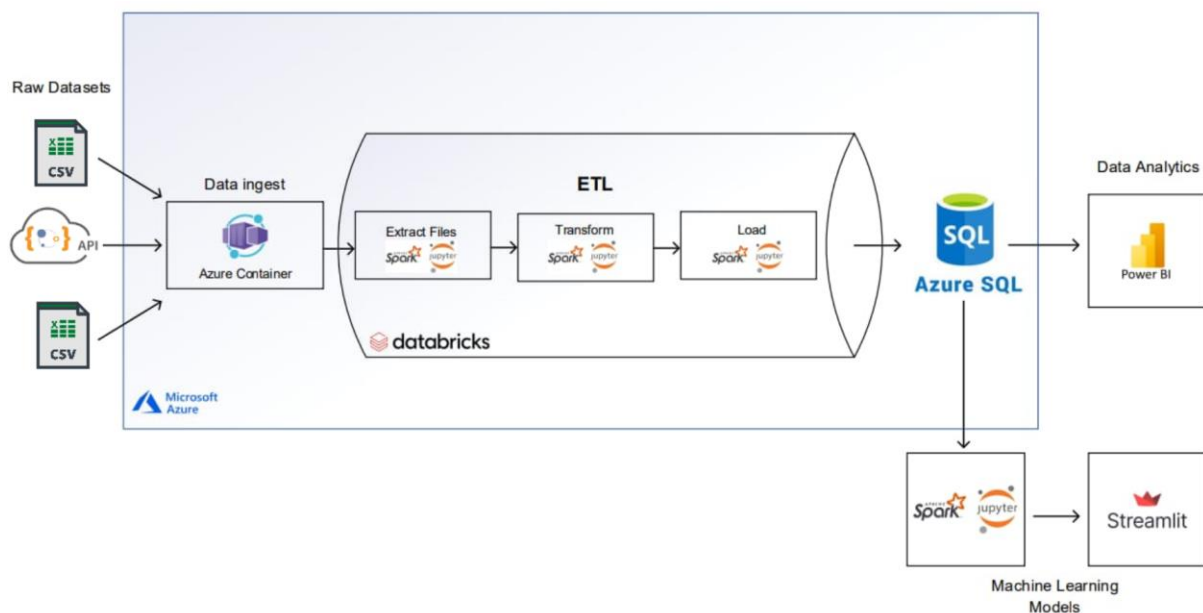


## 8. Stack tecnológico

Para llevar a cabo nuestro proyecto hemos seleccionado las siguientes tecnologías:

- **Trabajo diario:** python, google meet, github.
- **Ingeniería de datos:** Microsoft Azure, Python, sql.
- **Análisis y visualización de datos:** Power Bi, python.
- **Modelo de machine learning:** Python.
- **Gestión de proyectos:** Jira

### Solución data pipeline



En esta sección se estructurará el flujo del dato desde la recepción hasta la salida del ETL.

### Data Ingest

Los datos entregados por la empresa y extraídos por nuestro equipo mediante api y web scraping se descargaron y son almacenados de manera temporal en el localhost de nuestra máquina.

Dado que trabajaremos sobre el esquema de Microsoft Azure se creará un contenedor donde se almacenarán los datasets sin procesar en la nube. Para esto, fue necesario crear una cuenta de trabajo en el portal de Azure. En dicha cuenta se crea un grupo de recursos donde incluimos una cuenta de almacenamiento con un contenedor.

## Conexión con Databricks

Una vez almacenados los datasets en el contenedor de Azure se procede a realizar la conexión con Databricks, nuestro lugar de trabajo principal.

En el grupo de recursos previamente creado se añade un workspace de Databricks. Ahí se creará un clúster que permite computar nuestros datos (Single Node 10.4 LTS Apache Spark 14 GB Memory, 4 Cores), el criterio de selección es en base al alcance de nuestros recursos.

Dentro de Databricks creamos un Notebook y lo conectamos con el clúster. En dicho Notebook establecemos las variables necesarias para la conexión con el contenedor.

## ETL

se realizará todo el proceso de extracción, transformación y carga de los datos hacia el data warehouse

## Conexión con sql database

Creada la SQL Database de Azure se realizará la conexión con Databricks por medio del protocolo jdbc.

## Conexión con power bi

La conexión se realiza mediante el conector de Azure SQL Database de PowerBI. Se ingresan las credenciales del servidor de base de datos y se cargan los datos ya sea por Direct Query o Import Data.

## 9. Roles y responsabilidades

María Marcela Balzarelli - Data Science  
Pablo Nahuel Barchiesi Ponce - Data Engineer  
Michael Martinez Chinchilla - Data Engineer  
Jorgelina Paola Lujan Ramos - Data Analyst

	<b>Maria Balzarelli</b>	<b>Pablo Barchiesi</b>	<b>Michael Martinez</b>	<b>Jorgelina Ramos</b>
<b>arquitectura de datos</b>	Auxiliar	<b>líder</b>	<b>líder</b>	Auxiliar

<b>ingeniería de datos</b>	Auxiliar	<b>líder</b>	<b>líder</b>	Auxiliar
<b>análisis de datos</b>	Auxiliar	Auxiliar	Auxiliar	<b>líder</b>
<b>machine learning</b>	<b>líder</b>	Auxiliar	Auxiliar	Auxiliar

## 10. Creación del repositorio en Github

Para poder visualizar los trabajos de forma remota y conciliar la evolución de versiones, se adoptó la tecnología git , por lo tanto se creó un repositorio colaborativo en la plataforma github, donde se podrá crear, modificar y visualizar los contenidos que se irán requiriendo a lo largo del proyecto.

link: [https://github.com/marcebalzarelli/Proyecto\\_Movilidad\\_Sostenible\\_NYC.git](https://github.com/marcebalzarelli/Proyecto_Movilidad_Sostenible_NYC.git)