## 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 servicios de transporte de pasajeros, que actualmente se encuentra operando en el sector de micros de media y larga distancia, está interesada en invertir en el sector de transporte de pasajeros con automóviles. Con una visión de un futuro menos contaminado y ajustarse a las tendencias de mercado actuales, quieren corroborar la relación entre estos medios de transporte particulares y la calidad del aire, como también la contaminación sonora, para estudiar la posibilidad de implementar vehículos eléctricos a su flota.

Sin embargo, debido a que sería una unidad de negocio nueva, se pretende hacer un análisis preliminar del movimiento de los taxis en la ciudad de Nueva York, para poder obtener un marco de referencia y poder tomar decisiones bien fundamentadas.

**Transformación del Transporte en Nueva York:**

En la ciudad de Nueva York, los servicios de taxis y viajes compartidos han evolucionado para convertirse en alternativas populares al transporte público y al alquiler de automóviles. La conveniencia y accesibilidad ofrecidas por estas plataformas han generado un flujo constante de datos en tiempo real que abarcan desde la ubicación del vehículo hasta la calificación del conductor.

**Cambio en los Patrones de Trabajo:**

En el contexto de un mundo postpandémico, hemos observado una notable transformación en los patrones laborales, con un aumento significativo en la flexibilidad de horarios. Esta flexibilidad ha permitido a las personas reevaluar y adaptar sus rutinas diarias, incluyendo sus viajes diarios. En este escenario, surge la oportunidad de ofrecer a los usuarios una experiencia de viaje más personalizada y alineada con sus preferencias individuales.

La flexibilización de los horarios laborales ha generado un cambio fundamental en la manera en que las personas planifican sus actividades diarias. Ahora, los usuarios tienen la capacidad de organizar sus viajes de una manera que no solo se ajuste a sus responsabilidades laborales, sino que también considere otros factores que influyen en su bienestar, como la calidad del aire y la contaminación sonora.

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

La empresa de servicios de transporte de pasajeros, especializada en micros de media y larga distancia, busca expandir su operación hacia el transporte de pasajeros en automóviles. Este cambio estratégico tiene como objetivo principal la reducción de la huella ambiental y la adaptación a las tendencias de mercado que favorecen soluciones de transporte más sostenibles. La empresa aspira a comprender la relación entre el uso de vehículos particulares y la calidad del aire, así como la contaminación sonora, para evaluar la viabilidad de incorporar vehículos eléctricos a su flota.

**Necesidad de Análisis Preliminar:**

Dado que esta expansión representa una nueva unidad de negocio, se plantea la necesidad de realizar un análisis preliminar del movimiento de taxis en la ciudad de Nueva York. Este análisis servirá como marco de referencia para la toma de decisiones informadas sobre la implementación de vehículos eléctricos en la flota de la empresa. Para abordar este desafío, nuestro equipo ha sido contratado con el propósito de aprovechar los datos disponibles, provenientes tanto de fuentes proporcionadas como de otras incorporadas, para ofrecer insights valiosos y respaldar la toma de decisiones estratégicas.

**Desafíos y Oportunidades:**

El proyecto enfrenta desafíos relacionados con la complejidad de los datos, la diversidad de las fuentes y la necesidad de encontrar patrones significativos en un contexto de cambio climático acelerado. Sin embargo, también presenta oportunidades para aprovechar tecnologías avanzadas, como el Machine Learning, para extraer conocimientos profundos de los datos y proporcionar recomendaciones valiosas que alinean la estrategia de la empresa con las metas de sostenibilidad y las expectativas del mercado.

## **OBJETIVO GENERAL**

El objetivo general de este proyecto es proporcionar a la empresa de transporte de pasajeros la información necesaria para evaluar la viabilidad del negocio de transporte de pasajeros con automóviles, considerando la calidad del aire y la contaminación sonora en la ciudad de Nueva York.

## **OBJETIVOS ESPECÍFICOS**

1. Estructurar un Data Warehouse que integre datos de servicios de taxis y viajes compartidos con información adicional, asegurando calidad y coherencia.
2. Implementar al menos dos tipos diferentes de extracción de datos, como datos estáticos y llamadas a API, para construir una base de datos robusta.
3. Desarrollar un reporte dinámico que destaque métricas clave y tendencias identificadas en el análisis de datos.
4. Seleccionar, entrenar y validar un modelo de machine learning que prediga la demanda de taxis en función de variables relevantes.
5. Implementar el modelo en un entorno de producción para proporcionar estimaciones de demanda en tiempo real.
6. Integrar datos adicionales de servicios de viajes de otras plataformas, contaminación del aire y emisión de CO2.

## **ALCANCE**

El proyecto se centrará en el análisis de datos para comprender la relación entre el uso de vehículos particulares y la calidad del aire, para evaluar la viabilidad de incorporar vehículos eléctricos a la flota de la empresa.

El alcance del proyecto se limitará al análisis de taxis amarillos y verdes en la ciudad de Nueva York durante un año (septiembre 2022-septiemxbre 2023)

El alcance del proyecto se centrará en los siguientes elementos:

**Demanda de taxis**

* Los patrones de demanda de taxis a lo largo del tiempo
* Los factores que afectan a la demanda de taxis, como la hora del día, el día de la semana, el clima y los eventos especiales

**Impacto ambiental de los taxis**

* Las emisiones de gases de efecto invernadero de los taxis
* El ruido producido por los taxis
* La relación entre la demanda de taxis y el impacto ambiental

**Evaluación de la viabilidad de los vehículos eléctricos**

* Los costos de los vehículos eléctricos
* La disponibilidad de infraestructura de carga

**Modelos de machine learning**

* Modelo de predicción de la demanda de taxis
* Modelo de recomendación: Viaja + limpio y rápido

## **OBJETIVOS Y KPIS ASOCIADOS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **KPI** | **Nombre** | **Fórmula** | **Objetivo** | **Meta** | **Tiempo** |
| 1 | Cantidad de taxis disponibles para satisfacer la demanda | (Número de taxis disponibles / Flota de taxis) | Satisfacer la demanda de taxis | > 90% | Mensual |
| 2 | Margen de ganancia de los recorridos más rentables | (Ingresos de los recorridos más rentables - Costos de los recorridos más rentables) / Ingresos de los recorridos más rentables | Mejorar la rentabilidad | > 20% | Mensual |
| 3 | Margen de ganancia de los distritos y zonas más rentables | (Ingresos de los distritos y zonas más rentables - Costos de los distritos y zonas más rentables) / Ingresos de los distritos y zonas más rentables | Mejorar la rentabilidad | > 30% | Mensual |
| 4 | Reducción de emisiones | (EmisionesCO2\_emitidas\_por\_taxi\_combustion- emisionesCO2\_emitidas\_por\_taxi\_electrico)/mes | Reducción de emisiones |  | Mensual |
| 5 | Reducción de combustible | (Costo\_de\_combustible\_de\_combustible\_fosil **-**Costo\_de\_combustible\_electrico)/mes | Reducción de combustible |  | Mensual |
| 6 | %Rates | (Rates positivos anuales / Rates totales anuales)\*100 | Satisfacción del usuario |  | Anual |

## **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**
2. **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 en Analytical Insights se han definido cinco roles fundamentales para proceso de análisis:

**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**: Romina Capurro

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 y Edwin Torre

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

**Data Engineer**: Juan y Patricio

Encargado 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**

Es esencial establecer cronogramas realistas y coherentes con la disponibilidad de recursos y la complejidad de las tareas para la organización efectiva del proyecto. Por esta razón, se ha elaborado un diagrama de Gantt que facilita la monitorización del progreso del proyecto.

Diagrama de Gantt

Adicionalmente, como se detalla en la sección 6.3 'Metodología de trabajo', el proyecto se administrará utilizando la metodología ágil SCRUM. Bajo este enfoque, las actividades del proyecto se han dividido en tres sprints. Los detalles específicos de esta división se pueden encontrar en el ANEXO al final de este documento.

## **Análisis de los datos a trabajar**

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:

---------------

-----------------------

--------------------

-------------------

La totalidad del código y los notebooks de los EDA preliminares se encuentran en el siguiente link: **Exploración Inicial**.

## **Repositorio Github**

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

**Repositorio**