IMPLEMENTACIÓN DE UN SERVICIO DE TRANSPORTE EN LA CIUDAD DE NUEVA YORK

ANÁLISIS PRELIMINAR

OBJETIVO GENERAL

- Establecer la viabilidad del negocio

OBJETIVOS ESPECÍFICOS:

- Implementar un Data Lake on-premise para el negocio.
- Crear y automatizar dashboards para la toma de decisiones del negocio.
- Implementar un modelo de Machine Learning que ayude a predecir los momentos de alta demanda/retorno esperado por mes.
- Automatizar los procesos de ETL y procesamiento de datos.
- Determinar los recorridos más rentables.
- Determinar las horas de mayor demanda.
- Determinar la influencia que tienen las condiciones ambientales sobre la demanda.
- Determinar la cantidad mínima de taxis para suplir la demanda.
- Determinar cuáles son los distritos y zonas más rentables en la ciudad.
- Determinar el margen de ganancia diario.

MÉTRICAS A UTILIZAR

Se espera que las siguientes métricas sean suficientes para el análisis del negocio y dar cumplimiento a todos los objetivos planteados.

- Viajes inter e intra distrito
- Distritos con mayor/menor cantidad de viajes
- Distancia promedio de viaje
- Días de la semana y semana con mayor cantidad de viajes
- Margen de ganancia por milla recorrida
- Duración promedio de viaje
- Zonas TLC[1] con mayor/menor movimiento

ALCANCE DEL PROYECTO

El estudio pretende abarcar todos los objetivos específicos haciendo uso de las métricas propuestas, en conjunto con sets de datos correspondientes al clima en el mismo periodo de tiempo y datos adicionales tomados de la página oficial de la ciudad de Nueva York.

Dentro de los datos adicionales se espera complementar las métricas propuestas buscando la relación que las mismas pueden tener con datos de consumo de gasolina y su costo, de esta manera, el estudio tendrá más validez a la hora de proyectar las ganancias de manera más acertada.

El estudio tendrá en cuenta las variables que pueden afectar las ganancias teniendo en cuenta que la empresa cuenta con su propia flota de taxis.

FUERA DEL ALCANCE

Si bien el estudio parte del análisis de información correspondiente a los taxis amarillos para revisar el comportamiento de los clientes, no se estudiará a profundidad la demanda que tienen otras compañías y otras plataformas (Uber, Lyft, etc.)

Como el estudio tiene en cuenta que la flota es propia, no tendrá en cuenta las propinas dentro del estudio, estas se consideran un ingreso adicional para el conductor y no son relevantes para el estudio de viabilidad del negocio.

De igual manera, aunque existen datos desde donde se puede evaluar el impacto ambiental que puede tener introducir al mercado un número determinado de taxis, no se estudiarán estos datos ni los beneficios tributarios por hacer uso de tecnologías limpias.

Desde un punto de vista más técnico no se plantea la posibilidad de usar recursos en la nube para la implementación de la solución.

SOLUCIÓN PROPUESTA

Se propone la implementación de un Data Lake on-premise con pipelines ETL automatizados y dashboards que permitan ver con claridad los datos analizados, para la implementación de esta solución se propone el uso del siguiente stack tecnológico:

- MinIO
- Apache Airflow
- Python
- PostgreSQL
- Docker
- PowerBI

METODOLOGÍA DE TRABAJO

A partir de la presentación de este documento se espera completar el proyecto en tres semanas, para esto se trabajará en tres etapas.

La primera etapa comprenderá todo el trabajo correspondiente al diseño del modelo, pipelines para alimentar el Data Lake, automatización y validación de datos. Todo el proceso será además documentado.

Durante la segunda etapa se diseñarán los reportes y dashboards necesarios para presentar la información de manera adecuada para que los tomadores de decisiones tengan un panorama amplio y suficiente de la viabilidad del negocio.

En la etapa final se afinarán los detalles de visualización, documentación del proyecto y modelo predictivo, se ejecutan pruebas preliminares para validar toda la información modelada y se prepara el proyecto para la presentación final.

ENTREGABLES

Al finalizar el proyecto el cliente recibirá:

- Data Lake implementado con la información inicial debidamente cargada.
- Dashboards y Reportes para toma de decisiones
- Modelo predictivo
- Documentación relacionada con todas las etapas de proyecto

FUENTES

YELLOW CAB TLC
https://www1.nyc.gov/site/tlc/businesses/yellow-cab.page

Emissions from the Taxi and For-Hire Vehicle Transportation Sector in New York City
https://www1.nyc.gov/assets/dcas/downloads/pdf/fleet/Emissions-from-NYC-For-Hire-Vehicle-FHV-Industry.pdf

New York City Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)
https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPMR_PTE_Y35NY_DPG&f=M

---

[1] Zonas TLC son las subdivisiones correspondientes a cada Borough.