

NYC Taxi Analytics

STRATEGY-DRIVEN ANALYSIS USING GCP

BY

Prasad Watane

Tanuj Agrawal

Ojas Indulkar

Problem Statement

- Operating in a high-volume, competitive market, NYC taxi stakeholders currently struggle with limited visibility into key performance drivers, often relying on intuition rather than data. This leads to costly inefficiencies like "deadheading" (driving empty), suboptimal fleet utilization, and missed opportunities during peak demand. This project transforms millions of raw trip records into actionable intelligence, offering a comprehensive view of financial trends, traffic patterns, and passenger behaviors to maximize revenue and streamline operational logistics

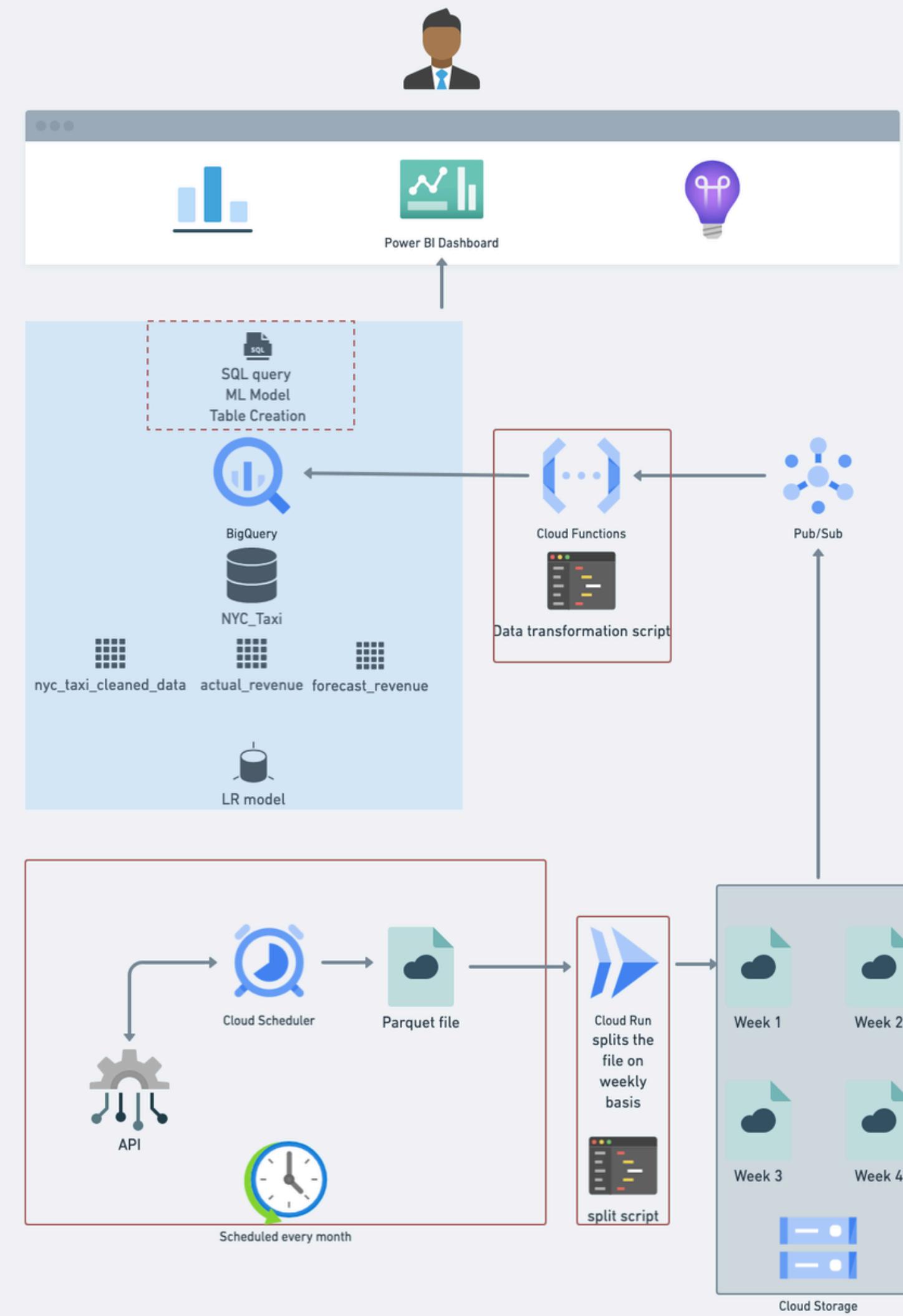
Dataset Description

- Source : NYC Taxi (TLC) Trip Record Data portal (official NYC.gov source) providing monthly trip files for Yellow/Green taxis (PARQUET), with trip timestamps/locations, distances, and itemized fare fields.
- Dataset includes both Yellow Taxi and Green Taxi trip records, published as separate monthly files on the NYC TLC portal.
- Taxi trip-level data with pickup/dropoff timestamps and pickup/dropoff zone IDs.
- Includes ride details (distance, passenger count) and payment info (payment type, rate code).
- Contains fare components (fare, tip, tolls, surcharges, total amount) for revenue forecasting.

Business Questions

- When is the absolute best time for drivers to be on the road?
- How much does traffic congestion hurt our speed and efficiency?
- Where are our most popular pickup spots?
- Do most customers pay with Cash or Card?
- Is our revenue coming from many cheap short trips or fewer expensive long trips?

Architecture



The End

THANK YOU FOR LISTENING