Data Analytics Track – Stage 2A Task: HNG Ride SQL Business Analysis

Business Scenario

You’ve joined HNG Ride, a mid-sized transportation company operating in North America. Management wants to review performance for the period mid-2021 through Dec 2024 to understand how operations have evolved and where to improve. The data provided by the IT team is messy with duplicates, inconsistent city names, invalid fares, and missing payments

As a data analyst, your task is to clean the data, answer key business questions, and prepare a short insight summary that management can use for decision-making.

Dataset: The dataset for HNG Ride can be found here

Instructions

Data Cleaning & Preparation

Address complex issues such as missing values, duplicates, and inconsistent formatting the datasets.

Validate data integrity by checking for incorrect fares, invalid driver ratings, or misaligned timestamps.

Focus on the defined analysis window (June 2021 – December 2024) to ensure relevant insights.

Business Questions

Using the cleaned data, write queries to answer the following questions:

Find the top 10 longest rides (by distance), including driver name, rider name, pickup/dropoff cities, and payment method.

How many riders who signed up in 2021 still took rides in 2024?

Compare quarterly revenue between 2021, 2022, 2023, and 2024. Which quarter had the biggest YoY growth?

For each driver, calculate their average monthly rides since signup. Who are the top 5 drivers with the highest consistency (most rides per active month)?

Calculate the cancellation rate per city and identify which city had the highest cancellation rate?

Identify riders who have taken more than 10 rides but never paid with cash.

Find the top 3 drivers in each city by total revenue earned between June 2021 and Dec 2024. If a driver has multiple cities, count revenue where they picked up passengers in that city.

Management wants to know the top 10 drivers that are qualified to receive bonuses using the criteria below;

at least 30 rides completed,

an average rating ≥ 4.5, and

a cancellation rate under 5%.

Insights & Report

After completing the data cleaning and analysis, prepare a short business report in PDF format that includes screenshots of your SQL queries and their results for each of the business questions.

Deliverables

A Google Drive link that contains;

8 .sql files, one for each question in the task and the files should be named in the format: 1.sql, 2.sql, and so on, up to 8.sql.

A structured PDF document summarizing the objective, queries, and findings.

Hint: Queries should be on completed rides. Criteria for completed rides is amount in payments table > 0