Uber Data Analysis Case Study

Project Overview

The goal of this project was to analyse a two-week dataset of Uber trip requests and driver supply to answer 11 key business questions. The dataset included hourly data on app users ("Eyeballs"), zero-availability events ("Zeroes"), completed trips, requests, and unique drivers logged in.

The analysis involved data cleaning, aggregation, time series calculations, and visualization to uncover patterns and insights that could inform operational decisions such as driver scheduling and resource allocation.

Question 1: Which date had the most completed trips during the two week period?

Thought Process:

- Group data by date and sum completed trips.
- Identify the date with the maximum total.
- This reveals peak demand days, critical for resource planning.
-

Question 2: What was the highest number of completed trips within a 24 hour period?

Thought Process:

- Calculate rolling sums of completed trips over 24 consecutive hours.
- Find the maximum rolling sum to identify peak daily trip volume.
- Helps understand daily capacity and demand spikes.

Question 3: Which hour of the day had the most requests during the two week period?

Thought Process:

- Aggregate requests by hour (0-23) across all days.
- The hour with the highest total requests indicates peak demand times.
- Useful for scheduling drivers during busiest hours.

Question 4: What percentage of all zeroes occurred on weekend (Friday 5pm to Sunday 3am)?

Thought Process:

- Define weekend hours as Friday 17:00 to Sunday 03:00.
- Sum zeroes during these hours and divide by total zeroes.
- This shows if availability issues are concentrated on weekends, guiding driver deployment.

Question 5: What is the weighted average ratio of completed trips per driver during the two week period?

Thought Process:

- Calculate ratio of completed trips to unique drivers for each hour.
- Weight each ratio by the number of trips or requests to reflect volume.
- Provides a more accurate overall efficiency metric accounting for demand variation.

Question 6: When are the busiest 8 consecutive hours in terms of unique requests?

Thought Process:

- Sum unique requests in rolling 8-hour windows starting every 8 hours (shifts).
- Identify the shift with the highest total requests.
- Supports drafting driver schedules aligned with demand peaks.

Question 7: True or False: Driver supply always increases when demand increases?

Thought Process:

- Visualize and compare time series of driver supply and requests.
- Check if supply consistently rises with demand or if there are mismatches.
- Important for understanding responsiveness of driver availability.

Question 8: In which 72hour period is the ratio of Zeroes to Eyeballs the highest?

Thought Process:

- Calculate rolling 72-hour sums of zeroes and eyeballs.
- Compute ratio and find the period with the highest value.
- Identifies periods of greatest supply shortage relative to demand.

Question 9: If adding 5 drivers to a single hour daily, which hour should it be?

Thought Process:

- Calculate ratio of eyeballs to unique drivers by hour.
- The hour with the highest ratio indicates greatest unmet demand per driver.
- Adding drivers at this hour maximizes impact on availability.

Question 10: True or False: There is exactly two weeks of data?

Thought Process:

- Check date range from earliest to latest date.
- Confirm if it spans exactly 14 days.
- Validates dataset completeness for analysis.

Question 11: Which time might make the most sense to consider a true "end day" instead of midnight?

Thought Process:

- Visualize supply and demand by hour across all days.
- Identify hour with natural minimums in both metrics.
- This hour can be used as a logical day boundary for operational reporting.

Summary

This project combined data cleaning, aggregation, rolling calculations, and visualization to extract actionable insights from Uber trip data. The analysis helped identify peak demand periods, supply-demand mismatches, and optimal driver scheduling windows. These insights can guide operational decisions to improve rider experience and driver utilization.

1. What problem were you trying to solve?

The primary problem was to analyze Uber's trip request and driver supply data over a two week period to understand patterns in demand and supply. Specifically, the goal was to identify peak demand times, supply shortages, and operational inefficiencies by answering 11 targeted questions. This analysis aimed to support better driver scheduling, improve rider experience by reducing zero-availability events, and optimize resource allocation.

Real-World Use:
These insights can help Uber improve rider satisfaction by reducing wait times and zero-availability events, optimize driver earnings by aligning shifts with demand, and enhance overall platform efficiency through data-driven operational decisions.

Uber Data Analysis – Summary / Outcomes

- We found the day and hour when the most trips were completed and when the most ride requests happened.

- Driver supply doesn't always match rider demand, causing some times when no cars are available.

- Many zero-availability events happen on weekends, especially from Friday evening to Sunday morning.

- The busiest 8-hour shifts were identified to help plan driver work schedules better.

- Adding 5 extra drivers during a specific hour each day can reduce wait times the most.

- The data covers exactly two weeks, and we found a better time than midnight to mark the end of the day for operations.

These insights help Uber schedule drivers smarter, reduce wait times, and improve rider experience.