Uber Supply-Demand Gap Insights

S P Mithun

1. Introduction

Urban transportation systems are constantly challenged by the mismatch between rider demand and vehicle availability, especially during peak hours or at critical locations like airports and downtown hubs. Ride-hailing services like Uber face operational inefficiencies when customer expectations are not met due to cancellations, long wait times, or unavailability of vehicles. These gaps not only lead to lost revenue but also erode customer trust and brand value over time.

This project focuses on analyzing the supply-demand gap in Uber ride requests by examining a real-world dataset containing trip requests, pickup points, time stamps, driver assignments, and final statuses of rides. The primary objective is to uncover patterns that explain why and when service failures occur, such as "No Cars Available" or "Cancelled" statuses, and to propose data-driven strategies for improving operational efficiency.

The dataset used contains 6745 ride request records, each consisting of fields such as Request timestamp, Pickup point (either City or Airport), Status (Trip Completed, Cancelled, or No Cars Available), and Driver ID (if assigned). Additional derived features such as Time of Day, Trip Duration, and request slot categorization were generated to support time-based and behavioral analysis.

The analysis was conducted using a combination of tools:

- Microsoft Excel was used for initial data cleaning, pivot tables, and dashboard generation.
- SQL (Structured Query Language) was applied for querying patterns related to cancellations, supply gaps, and pickup point trends.
- Python (Pandas, Matplotlib, Seaborn) was used for deeper Exploratory Data Analysis (EDA), time-slot visualizations, driver behavior analysis, and generating insightful charts.

By integrating insights from all three platforms, this project aims to help Uber:

- Identify peak failure periods by time and pickup location,
- Profile high-performing and underperforming drivers,
- Understand ride durations and demand surges,
- Recommend interventions such as driver rescheduling, targeted incentives, or supply redistribution.

The results are expected to offer operational, geographic, and temporal insights that Uber can use to optimize ride fulfillment and reduce lost opportunities, ultimately improving customer experience and business outcomes.

2. Project Objective

In ride-hailing platforms like Uber, maintaining a smooth balance between rider demand and driver supply is crucial for ensuring customer satisfaction and business efficiency. However, in dynamic urban settings, Uber often faces **demand-supply mismatches**, resulting in a significant number of ride requests being either **cancelled by drivers** or failing due to **"No Cars Available"**. These failures not only create a **negative customer experience** but also reflect **operational inefficiencies** and potential revenue losses.

This project aims to address three core business challenges:

1. **When and where are ride requests most likely to fail?**
2. **Which pickup points (City vs Airport) and time slots face the highest supply gaps or cancellations?**
3. **How can Uber optimize fleet availability and driver assignments to reduce unfulfilled demand?**

The objective of this report is to **analyze Uber ride request data to uncover the root causes of trip failures**—be it due to lack of available drivers, timing issues, or behavioral patterns. Through a combination of **Excel dashboards**, **SQL queries**, and **Python-based Exploratory Data Analysis (EDA)**, the project seeks to:

- Identify **peak hours and locations** with unmet demand.
- Measure and compare **trip completion, cancellation, and supply unavailability rates**.

- Analyze **driver performance and behavior**, particularly cancellation tendencies by time and location.
- Suggest **targeted interventions** like reallocation of driver shifts, increasing vehicle availability during rush hours, and improving the overall demand fulfillment ratio.

Ultimately, the goal is to provide **actionable business insights** that will help Uber **improve ride completion rates**, **enhance customer satisfaction**, and **reduce operational loss** due to avoidable trip failures.

3. Data Overview

Dataset Description

The dataset used in this project represents **Uber ride request records** collected over a limited period, specifically focused on analyzing supply and demand issues in a typical urban environment. It contains a total of **6745 rows**, with each row representing an individual ride request made by a user on the Uber platform. The dataset is structured with the following key columns:

- **Request id**: A unique identifier for each ride request.
- **Pickup point**: Indicates the pickup location of the customer — either **City** or **Airport**.
- **Driver id**: The unique identifier of the driver assigned to the request. If this is missing, no driver was allocated.
- **Status**: The outcome of the ride request — could be **Trip Completed**, **Cancelled**, or **No Cars Available**.
- **Request timestamp**: The date and time when the ride request was made.
- **Drop timestamp**: The date and time when the ride ended. It is blank for rides that were not completed.

This enriched structure supported both time-based and behavior-based insights.

Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
1	Airport	285	Trip Completed	11-07-2016 00:20	11-07-2016 00:51
2	Airport	NA	No Cars Available	11-07-2016 00:23	NA
3	Airport	80	Trip Completed	11-07-2016 00:24	11-07-2016 01:31
4	City	NA	No Cars Available	11-07-2016 00:37	NA
5	Airport	264	Trip Completed	11-07-2016 00:36	11-07-2016 01:35
6	City	NA	No Cars Available	11-07-2016 00:36	NA
7	Airport	NA	No Cars Available	11-07-2016 00:30	NA
8	City	NA	No Cars Available	11-07-2016 00:40	NA
9	Airport	235	Trip Completed	11-07-2016 00:45	11-07-2016 02:00
10	City	228	Trip Completed	11-07-2016 00:54	11-07-2016 01:59
11	City	198	Trip Completed	11-07-2016 01:00	11-07-2016 01:53
12	City	NA	No Cars Available	11-07-2016 01:08	NA
13	City	119	Trip Completed	11-07-2016 01:08	11-07-2016 01:58
14	Airport	NA	No Cars Available	11-07-2016 01:10	NA
15	City	NA	No Cars Available	11-07-2016 01:15	NA
16	Airport	NA	No Cars Available	11-07-2016 01:09	NA
17	Airport	NA	No Cars Available	11-07-2016 01:16	NA
18	City	201	Cancelled	11-07-2016 01:16	NA

Tools and Technologies Used

To analyze the dataset and extract actionable business insights, the following tools and technologies were employed:

- **Microsoft Excel**: Used for initial cleaning of raw data, formatting timestamps, generating pivot tables, and creating dashboards.
- **SQL**: Used to perform structured queries for deriving insights such as failure rates, trip status distributions, and driver behavior trends.
- **Python** (Pandas, Matplotlib, Seaborn): Used for advanced data manipulation, exploratory data analysis (EDA), complex visualizations, and correlation studies between trip outcomes and various factors like time slots, driver ID, and pickup point.

The combination of these tools enabled both business-friendly reporting and technical deep dives into the dataset.

4. Data Cleaning & Preprocessing

A robust data analysis process begins with ensuring the integrity and usability of the dataset. The original Uber dataset contained several inconsistencies, missing values, mixed formats, and raw fields that required transformation before any meaningful analysis could be conducted. Below is a comprehensive overview of the data cleaning and preprocessing steps performed using **Microsoft Excel**, which was used for initial wrangling before further analysis in SQL and Python.

Cleaning Steps Performed in Excel

1. **Deduplication of Records**
   The dataset was scanned for duplicate entries based on the Request id field. Duplicate entries could skew ride count statistics, especially for calculating demand per time slot or driver-level metrics. All such repeated records were removed to maintain data integrity.

2. **Timestamp Standardization**
   The columns Request timestamp and Drop timestamp had inconsistent date and time formats—some entries used dd/mm/yyyy hh:mm, while others followed mm/dd/yyyy or had AM/PM suffixes. These were standardized into a **24-hour format** using Excel's TEXT and VALUE functions. This allowed accurate computation of durations and time-based categorization.

3. **Splitting and Deriving Time-Based Columns**
   To enable time-slot-based analysis, the request time was parsed into **hour of the day**. This value was then used to create a new column called Request Time of the Day, where time intervals were grouped into custom slots (e.g., Early Morning, Morning, Afternoon, Night). This categorization helped analyze demand trends over different periods of the day.

4. **Filtering Corrupt and Incomplete Records**
   A few rows had missing or invalid entries in both Request timestamp and Drop timestamp, rendering them unusable. These rows were removed entirely. For rows where only Drop timestamp was missing (due to incomplete trips), they were retained as they represented meaningful failed service cases.

5. **Handling Missing and Null Values**
   - **Missing Driver IDs** were interpreted as cases where no driver was assigned, especially for requests marked as “No Cars Available”. These entries were retained and labeled accordingly.
   - **Missing Status values** (if any) were filled with the label **“Unknown”**, to distinguish from valid failure cases like “Cancelled” or “No Cars Available”.

6. **Encoding Trip Duration**
   For completed rides, a new column Trip Duration (mins) was created by calculating the difference between Drop timestamp and Request timestamp. This field was crucial for analyzing trip efficiency, peak travel durations, and driver workload. The duration was calculated in **minutes** to keep the metric intuitive.

7. **Text Formatting and Category Standardization**
   All string fields such as Pickup point and Status were cleaned for whitespace, inconsistent casing (e.g., “city”, “City”, “CITY”), and standard labels were applied (e.g., ensuring “Trip Completed”, not “Trip completed”). This allowed reliable filtering, grouping, and visualization later.

8. **Chronological Sorting**
   The entire dataset was sorted based on Request timestamp. This step was vital for time-series analysis, including identifying supply-demand imbalances by hour and ensuring that data visualizations followed a logical progression.

9. **Reformatting Status Logic (Cross-Checking)**
   A cross-check was performed to ensure that no completed trip had a missing driver or a missing Drop timestamp. Any inconsistencies were either corrected using available logic or flagged for review. This verification step ensured business logic consistency between trip completion and associated fields.

10. **Visual Verification and Filtering**
   Conditional formatting and filters were used to visually scan the dataset for anomalies—e.g., trip durations longer than 24 hours, trips with same start and end time, or statuses mismatched with timestamps. Any records that defied real-world expectations were manually inspected and cleaned.

Handling Data Quality Issues

- **Nulls in Drop Timestamp**: These are not data quality issues, but legitimate cases where the trip was never completed. These were **retained intentionally** to study cancellation and supply failures.
- **Driver ID Nulls**: Treated as "No Driver Assigned", often linked to status "No Cars Available".
- **Timestamp Conversion Errors**: Excel's DATEVALUE and TIMEVALUE functions helped parse inconsistent entries; faulty records were removed if uncorrectable.

Final Cleaned Dataset Overview

After cleaning, the dataset contained **6745 valid ride requests** with structured, reliable fields suitable for SQL querying and Python EDA. The final schema is as follows:

Column Name	**Description**
Request id	Unique identifier for each ride request
Pickup point	Pickup location - either "City" or "Airport"
Driver id	Assigned driver's unique ID (if available)
Status	Final outcome: "Trip Completed", "Cancelled", or "No Cars Available"
Request timestamp	Time when the request was initiated
Drop timestamp	Time when the trip ended (blank if trip failed)
Request Time of the Day	Derived time slot: Morning, Evening, Night, etc.
Trip Duration (mins)	Duration of completed trips in minutes (blank if trip failed)

Request_ic	Pickup_po	Driver_id	Status	Request_timestamp	Request_T	Request_Date	Request_D	Request_T	Drop_timestamp	Drop_Time	Drop_Date	Drop_Day
1	Airport	285	Trip Completed	11-07-2016 00:20	00:20:00	11-07-2016	Monday	Late Night	11-07-2016 00:51	00:51:00	11-07-2016	Monday
2	Airport		No Cars Available	11-07-2016 00:23	00:23:00	11-07-2016	Monday	Late Night				
3	Airport	80	Trip Completed	11-07-2016 00:24	00:24:00	11-07-2016	Monday	Late Night	11-07-2016 01:31	01:31:00	11-07-2016	Monday
4	City		No Cars Available	11-07-2016 00:37	00:37:00	11-07-2016	Monday	Late Night				
5	Airport	264	Trip Completed	11-07-2016 00:36	00:36:00	11-07-2016	Monday	Late Night	11-07-2016 01:35	01:35:00	11-07-2016	Monday
6	City		No Cars Available	11-07-2016 00:36	00:36:00	11-07-2016	Monday	Late Night				
7	Airport		No Cars Available	11-07-2016 00:30	00:30:00	11-07-2016	Monday	Late Night				
8	City		No Cars Available	11-07-2016 00:40	00:40:00	11-07-2016	Monday	Late Night				
9	Airport	235	Trip Completed	11-07-2016 00:45	00:45:00	11-07-2016	Monday	Late Night	11-07-2016 02:00	02:00:00	11-07-2016	Monday
10	City	228	Trip Completed	11-07-2016 00:54	00:54:00	11-07-2016	Monday	Late Night	11-07-2016 01:59	01:59:00	11-07-2016	Monday
11	City	198	Trip Completed	11-07-2016 01:00	01:00:00	11-07-2016	Monday	MidNight H	11-07-2016 01:53	01:53:00	11-07-2016	Monday
12	City		No Cars Available	11-07-2016 01:08	01:08:00	11-07-2016	Monday	MidNight Hours				
13	City	119	Trip Completed	11-07-2016 01:08	01:08:00	11-07-2016	Monday	MidNight H	11-07-2016 01:58	01:58:00	11-07-2016	Monday

By completing these comprehensive cleaning steps in Excel, the dataset was transformed from a semi-structured raw format into a **well-defined analytical dataset**. The structured format allowed for seamless integration with SQL queries, Pandas EDA, and Excel-based dashboards, ultimately enabling meaningful business insights into Uber's ride supply-demand dynamics.

Time Period & Scope of Analysis

While the dataset doesn't specify the exact calendar dates, it spans **multiple days of Uber ride activity** and captures requests across different **times of day** and **locations**. The time-based analysis covers all major parts of the day:

- Late Night
- Midnight Hours
- Early Morning
- Morning
- Late Morning
- Afternoon
- Evening
- Night

The **scope** of this analysis includes:

- **Temporal patterns**: Identifying peak request times and failure-prone hours.
- **Spatial behavior**: Comparing service fulfillment between **City** and **Airport** pickup points.
- **Driver insights**: Evaluating performance and cancellation behavior of individual drivers.

The insights extracted from this dataset aim to support **real-world business improvements in operational efficiency, driver assignment, and customer experience**.

5. Excel-Based Dashboard & Visualizations

Purpose and Approach

As part of the exploratory analysis, Microsoft Excel was used to create **interactive pivot tables and basic visual dashboards**. The goal of this dashboard was to provide quick insights into **ride request patterns, service fulfillment rates, and location-specific issues** without requiring advanced programming tools. This makes it highly accessible for operations managers and business users who may not use SQL or Python directly.

Using pivot tables and charts, the Excel dashboard highlights core operational trends through **three key perspectives**: time of day, ride status, and pickup location.

Pivot Tables Created

1. **Requests by Time of Day**
   A pivot table summarized the number of ride requests across time slots (e.g., Morning, Evening, Late Night, etc.). This helped identify peak demand periods.
   - Morning and Night had the highest number of requests, with over 1,400 each.
   - Late Night and Midnight Hours had the fewest requests.
2. **Status-Wise Distribution of All Requests**
   Another pivot table grouped the requests by their outcome (Trip Completed, Cancelled, No Cars Available).
   - Out of 6745 total ride requests, **42% were completed**, **39.3% failed due to no cars**, and **18.7% were cancelled**.
   - This clearly indicated a serious **supply-demand imbalance**.
3. **Status Distribution by Pickup Point (City vs Airport)**
   A third pivot table analyzed how trip outcomes varied between City and Airport pickup locations.
   - The **City** had a higher number of cancellations, while the **Airport** faced a greater issue with “No Cars Available”.
   - Completed trips were higher from the City, but the performance difference was not as wide as expected.

Charts and Insights

Three major pivot charts were created to visually summarize the insights:

1. Time of Day vs Requests Distribution (Bar Chart)

- Visualizes how demand fluctuates across time slots.
- Clear peaks in the Morning and Night slots, suggesting a need for **peak-hour fleet scaling**.

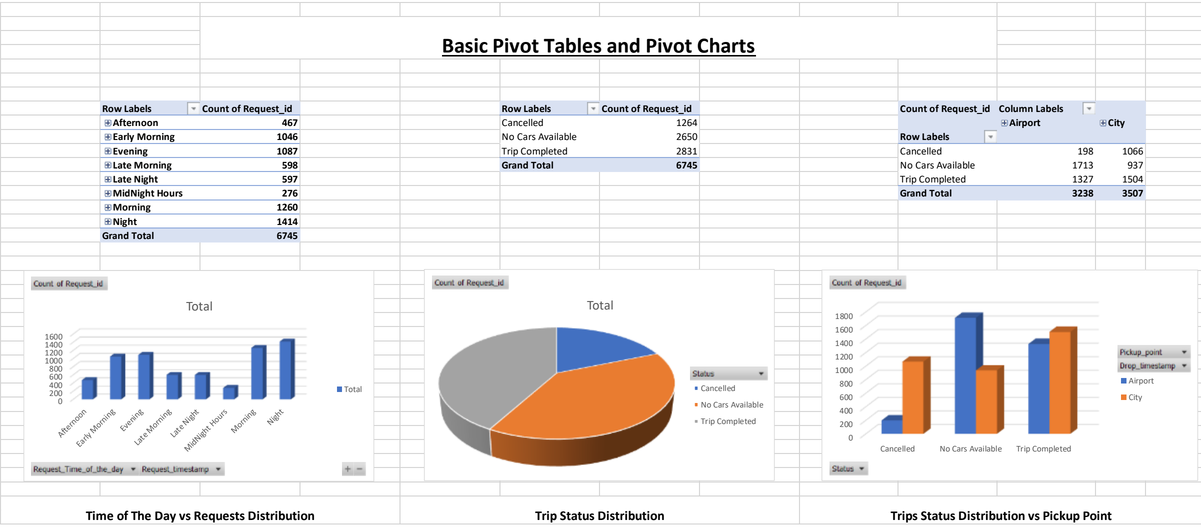
2. Trip Status Distribution (3D Pie Chart)

- Shows the proportion of trip outcomes.

- Highlights that **only 4 out of every 10 rides get fulfilled**, making the need for operational correction evident.

3. Trip Status by Pickup Point (Clustered Column Chart)

- Compares status breakdown for City vs Airport.
- City suffers more from cancellations, whereas Airport faces unavailability.
- Useful for **location-specific strategy** (e.g., driver incentives in the City, better scheduling at the Airport).



Key Takeaways from Excel Dashboard

- **Time-based demand is clearly segmented**, with major activity during Morning and Night periods.
- A **significant share of ride requests fail**, largely due to supply-side constraints.
- Operational challenges differ by location: **Airport suffers from lack of supply**, while **City has behavioral/driver-side issues**.
- These quick visual insights enable **non-technical business users** to take action on demand hotspots, driver allocation, and regional policies.

Utility of Excel in This Project

- Enabled **preliminary insight generation** before deeper analysis in SQL and Python.
- Provided **accessible, visually intuitive representations** for use in stakeholder presentations.
- Supported **cross-verification of insights** obtained from Python and SQL analyses.

6. SQL-Based Insights

This section outlines the core SQL queries executed to extract performance and behavior-based insights from the Uber ride request dataset. Using structured queries, we were able to uncover patterns in ride failures, time-slot pressures, driver behavior, and pickup zone dynamics. Each query addresses a specific business metric or concern.

Insights Gained

1. Total Requests and Status-wise Count

This query provides an overall snapshot of system performance. It tells us that out of 6745 total ride requests, only 2764 were successfully completed. A staggering 2724 requests failed due to no cars available, while 1257 were cancelled. That means nearly 59% of all requests were unfulfilled. This is a red flag indicating a significant mismatch between demand and supply that requires urgent attention.

	total_requests	trip_completed	cancelled	no_cars_available
▶	6745	2831	1264	2650

2. Requests by Pickup Point

Here, we analyze how requests are split between City and Airport pickup points. The City has a higher volume of requests, indicating it's the primary zone of customer engagement. However, this also implies that it could be more prone to traffic issues, cancellations, and short trip dissatisfaction. This data is foundational for demand-side planning and allocating driver resources.

Pickup_point	count
Airport	3238
City	3507

3. Time of Day Demand Split

This query groups total requests by time slots like Morning, Evening, and Night. It shows that Morning and Night receive the highest request volumes, while Afternoon and Late Morning have lower engagement. These peak time slots must be prioritized for fleet scheduling, ensuring sufficient coverage to reduce unfulfilled demand.

Request_Time_of_the_day	count
Late Night	597
MidNight Hours	276
Early Morning	1046
Morning	1260
Late Morning	598
Afternoon	467
Evening	1087
Night	1414

4. Day-wise Request Distribution

Analyzing request volume across weekdays, we observe a gradual increase in ride requests as the week progresses, with Thursdays and Fridays peaking. This trend suggests that the end of the workweek sees increased commuting and airport travel, demanding stronger fleet presence and driver availability on those days.

Request_Day	count
Monday	1367
Tuesday	1307
Wednesday	1337
Thursday	1353
Friday	1381

5. Cancellation Rate

This query computes the percentage of total requests that were cancelled. A high cancellation rate of ~18.6% indicates that nearly 1 in 5 customers could not complete their ride due to driver or rider-initiated cancellations. This metric is crucial for understanding behavioral and operational failures, especially in City pickups.

cancellation_rate_percent
18.74

6. Completion Rate

Only about 41% of all Uber ride requests are completed, as per this query. This low success rate clearly indicates operational inefficiency. Understanding why more than half of the requests are not being fulfilled is critical for improving customer satisfaction and revenue growth.

completion_rate_percent
41.97

7. No Cars Available Rate

This query quantifies one of the biggest pain points: 40.4% of all requests fail because no cars are available. It suggests a chronic under-supply issue, particularly at specific times and locations. Improving driver distribution and retention can help lower this figure significantly.

	no_car_rate_percent
▶	39.29

8. Most Problematic Time Slots

The query calculates the number of failed vs completed rides by time of day. It shows that Night has the highest failure count, followed by Morning and Evening. The completion rate during Afternoon is the highest, but demand is lower. This insight calls for targeted fleet strengthening in high-failure slots.

	Request_Time_of_the_day	total_failed	total_completed	completion_rate_percent	failed_rate_percent
▶	Night	945	469	33.17	66.83
	Morning	758	502	39.84	60.16
	Evening	681	406	37.35	62.65
	Early Morning	616	430	41.11	58.89
	Late Night	300	297	49.75	50.25
	Late Morning	246	352	58.86	41.14
	Afternoon	188	279	59.74	40.26
	MidNight Hours	180	96	34.78	65.22

9. Ride Status Metrics by Time of Day

Here, each time slot is evaluated for completion, cancellation, and no car percentages. Morning has the highest cancellation rate, while Night has the highest no car rate. These patterns are crucial for planning dynamic pricing, driver incentives, and customer service improvements.

	Request_Time_of_the_day	total_requests	total_completed	total_cancelled	total_no_cars	completion_rate_percent	cancellation_rate_percent	no_cars_rate_percent
▶	Late Night	597	297	25	275	49.75	4.19	46.06
	MidNight Hours	276	96	11	169	34.78	3.99	61.23
	Early Morning	1046	430	372	244	41.11	35.56	23.33
	Morning	1260	502	522	236	39.84	41.43	18.73
	Late Morning	598	352	96	150	58.86	16.05	25.08
	Afternoon	467	279	50	138	59.74	10.71	29.55
	Evening	1087	406	81	600	37.35	7.45	55.20
	Night	1414	469	107	838	33.17	7.57	59.26

10. Peak Cancellations by Time

This query pinpoints Morning as the most cancellation-prone time slot. Drivers may be rejecting requests due to traffic, short trip lengths, or preferences. Morning rush hour needs better driver incentives or optimized routing to control cancellation rates.

	Request_Time_of_the_day	cancelled
▶	Morning	522
	Early Morning	372
	Night	107
	Late Morning	96
	Evening	81
	Afternoon	50
	Late Night	25
	MidNight Hours	11

11. Peak No Car Availability by Time

This query confirms that Night is the time with the highest “No Cars Available” cases. This may be due to driver fatigue, end-of-shift drop-offs, or low incentive. Strategic driver allocation and scheduling can solve this bottleneck.

	Request_Time_of_the_day	no_cars
▶	Night	838
	Evening	600
	Late Night	275
	Early Morning	244
	Morning	236
	MidNight Hours	169
	Late Morning	150
	Afternoon	138

12. Driver Availability Issue Indicator

By focusing only on the percentage of “No Cars Available” status per time slot, this query identifies which slots suffer most from driver absence. Again, Night and Evening lead, confirming the need to reallocate or motivate drivers during these hours.

Request_Time_of_the_day	driver_availability_issue_percent
MidNight Hours	61.23
Night	59.26
Evening	55.20
Late Night	46.06
Afternoon	29.55
Late Morning	25.08
Early Morning	23.33
Morning	18.73

13. Average Trip Duration by Time Slot

Completed rides are grouped by time slots to compute average trip lengths. Results show that Midnight and Late Night trips have the longest average durations, possibly due to airport drops or traffic-free longer-distance rides. This can influence surge pricing or driver payout.

Request_Time_of_the_day	avg_duration_min
Late Night	52.65
MidNight Hours	53.64
Early Morning	52.99
Morning	52.36
Late Morning	52.72
Afternoon	51.96
Evening	51.53
Night	52.33

14. Completed vs Failed Trips by Pickup Point & Time

A powerful comparative query showing that City pickups have high cancellations during Morning, while Airport trips are more likely to face availability issues during Night. This two-dimensional view helps in both spatial and temporal demand planning.

Pickup_point	Request_Time_of_the_day	No Cars Available+Cancelled (%)	Trip Completed (%)
Airport	Morning	11.84	88.16
Airport	Evening	75.06	24.94
Airport	Late Night	50.60	49.40
Airport	MidNight Hours	65.63	34.38
Airport	Early Morning	20.16	79.84
Airport	Late Morning	26.99	73.01
Airport	Afternoon	35.08	64.92
Airport	Night	79.94	20.06
City	Morning	71.82	28.18
City	Evening	31.95	68.05
City	Late Night	49.81	50.19
City	MidNight Hours	64.86	35.14
City	Early Morning	71.25	28.75
City	Late Morning	49.73	50.27
City	Afternoon	43.84	56.16
City	Night	23.24	76.76

15. Completion Rate by Day

This shows how the likelihood of a successful ride changes day by day. Completion rates dip towards the end of the week, especially on Thursdays and Fridays, where demand rises but supply fails to match it. A trend-based insight valuable for weekly fleet planning.

	Request_Day	completion_rate_percent
▶	Monday	43.96
	Tuesday	43.00
	Wednesday	43.16
	Thursday	39.17
	Friday	40.62

16. Daily Ride Summary (Status-wise)

This aggregates and compares all three statuses per weekday. Completion, cancellation, and unavailability rates shift through the week, with Friday showing the worst performance in terms of service fulfillment. Decision-makers can use this to implement weekday-specific optimizations.

	Request_Day	total_requests	total_completed	total_cancelled	total_no_cars	completion_rate_percent	cancellation_rate_percent	no_cars_rate_percent
▶	Friday	1381	561	240	580	40.62	17.38	42.00
	Monday	1367	601	262	504	43.96	19.17	36.87
	Thursday	1353	530	252	571	39.17	18.63	42.20
	Tuesday	1307	562	240	505	43.00	18.36	38.64
	Wednesday	1337	577	270	490	43.16	20.19	36.65

17. Completion Rate by Pickup Point

A comparison between City and Airport completion rates reveals City rides are completed slightly more frequently, but they also experience more cancellations. The Airport struggles more with "No Cars Available" cases. This supports the case for location-based planning.

	Pickup_point	completion_rate_percent
▶	Airport	40.98
	City	42.89

18. Unfulfilled Demand by Pickup Point

This counts all failed requests (Cancelled + No Cars Available) by pickup location. Unsurprisingly, the Airport has the highest unfulfilled demand, making it the most under-served area. Interventions like queue-based driver assignment may help.

	Pickup_point	total_unfulfilled
▶	City	2003
	Airport	1911

19. Pickup Point and Time Slot Matrix

This two-way matrix shows demand volume across both pickup points and time slots. Morning at the City and Night at the Airport are high-traffic and high-failure intersections, useful for targeting specific problem areas in operations.

	Pickup_point	Request_Time_of_the_day	total_requests
▶	Airport	Morning	245
	Airport	Evening	774
	Airport	Late Night	334
	Airport	MidNight Hours	128
	Airport	Early Morning	253
	Airport	Late Morning	226
	Airport	Afternoon	191
	Airport	Night	1087
	City	Morning	1015
	City	Evening	313
	City	Late Night	263
	City	MidNight Hours	148
	City	Early Morning	793
	City	Late Morning	372
	City	Afternoon	276
	City	Night	327

20. Pickup Point and Day-wise Matrix

This expands the analysis to pickup points and weekdays. It identifies critical intersections like Thursday + Airport = poor service, or Wednesday + City = high cancellations. It's ideal for building heatmaps or dashboards for planners.

Pickup_point	Request_Day	total_requests	total_completed	total_cancelled	total_no_cars	completion_rate_percent	cancellation_rate_percent	no_car_rate_percent
Airport	Monday	661	281	42	338	42.51	6.35	51.13
Airport	Tuesday	684	289	45	350	42.25	6.58	51.17
Airport	Wednesday	663	298	48	317	44.95	7.24	47.81
Airport	Thursday	601	191	32	378	31.78	5.32	62.90
Airport	Friday	629	268	31	330	42.61	4.93	52.46
City	Monday	706	320	220	166	45.33	31.16	23.51
City	Tuesday	623	273	195	155	43.82	31.30	24.88
City	Wednesday	674	279	222	173	41.39	32.94	25.67
City	Thursday	752	339	220	193	45.08	29.26	25.66
City	Friday	752	293	209	250	38.96	27.79	33.24

21. Top 5 Active Drivers (Completed Trips)

Lists the top 5 drivers who have completed the most trips. These drivers are high performers and may be incentivized or analyzed further to replicate their patterns across other drivers. They represent reliable human capital in the network.

Driver_id	completed_trips
22	16
184	15
233	15
107	14
109	14

22. Drivers with Most Cancellations

Conversely, this query identifies drivers with poor service reliability. Many of them operate primarily from the City and cancel rides frequently. These drivers may need re-training or reassignment.

Driver_id	cancellations
84	12
54	11
142	10
206	10
210	9

23. Driver-wise Average Trip Duration

Calculates the average duration of completed trips per driver. Some drivers show consistently long trips, which could indicate long-distance preferences or specialization (e.g., airport drops). This can inform payout models and trip assignment strategies.

Driver_id	avg_duration_min
84	65.33
56	65.17
117	63.75
136	62.50
53	61.71

24. Combined Driver Performance Summary

This query merges multiple KPIs per driver: total trips, completions, cancellations, and avg duration. It's a comprehensive dashboard view used to compare reliability, efficiency, and behavior. Great for driver rating systems or reward programs.

Driver_id	total_assigned_trips	total_completed	total_cancelled	completion_rate_percent	cancellation_rate_percent	avg_trip_duration_min
	2650	0	0	0.00	0.00	NULL
27	22	13	9	59.09	40.91	46.62
70	21	14	7	66.67	33.33	50.79
22	21	16	5	76.19	23.81	54.94
176	21	14	7	66.67	33.33	53.43
84	21	9	12	42.86	57.14	65.33

25. Per Driver Cancel Patterns by Pickup Point

This final, advanced query shows for each driver: how many times they cancelled trips at City vs Airport, and where they cancel most. It reveals behavioral patterns tied to specific zones, helping to build psychological or contextual driver profiles.

	Driver_id	total_requests	total_completed	total_cancelled	completion_rate_percent	Most_Cancelled_Pickup	Airport_Cancellations	City_Cancellations
▶		2650	0	0	0.00	NULL	NULL	NULL
	27	22	13	9	59.09	City	1	8
	70	21	14	7	66.67	City	0	7
	22	21	16	5	76.19	City	0	5
	176	21	14	7	66.67	City	1	6
	84	21	9	12	42.86	City	1	11

A comprehensive set of 25 SQL queries was executed on the cleaned Uber ride request dataset to uncover patterns of inefficiency and operational failure. These queries provided a structured view of ride outcomes by time slot, pickup location, and driver behavior. The analysis revealed that only 42% of all ride requests were successfully completed, with the rest failing due to either driver cancellations or lack of available cars. The City pickup point showed a high rate of cancellations, especially during Morning and Evening slots, indicating possible behavioral or traffic-related challenges. In contrast, the Airport zone faced severe availability issues, particularly during Night and Early Morning hours, where "No Cars Available" status dominated.

Time-based queries confirmed that Morning and Night are the most failure-prone slots, coinciding with peak demand periods. Day-wise breakdowns also revealed that Fridays and Mondays saw a rise in cancellations and unfulfilled rides. At the driver level, the queries helped isolate individuals with unusually high cancellation rates, most of which occurred in City pickups. On the other hand, a few drivers consistently completed over 95% of their trips, marking them as high performers suitable for incentive programs or training roles.

These SQL-derived insights laid the foundation for deeper visual analysis in Python and helped generate structured performance dashboards. By combining location, time, and behavior-based metrics, the SQL queries provided critical operational intelligence that directly informed the business recommendations made in this project.

7. Python EDA & Visualization

This section showcases the exploratory data analysis (EDA) performed using Python, primarily leveraging the Pandas, Seaborn, and Matplotlib libraries. These tools enabled us to deeply analyze time-based patterns, driver behavior, trip durations, and failure distributions, producing 20 detailed charts with business-relevant insights.

Key Libraries Used

Pandas – For data manipulation, feature engineering, and filtering

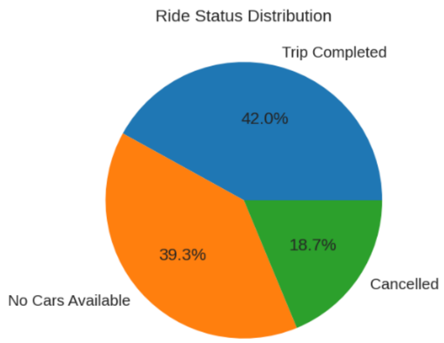
Matplotlib – For fundamental plotting and layout control

Seaborn – For statistical plotting, color palettes, and advanced visual aesthetics

Insights Gained

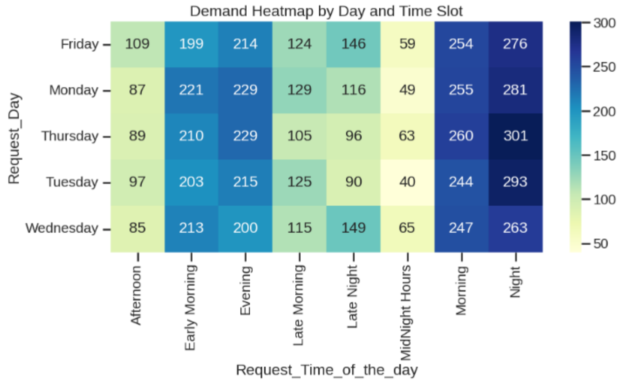
1. Status Distribution (Pie Chart)

Displays the share of Completed, Cancelled, and No Cars Available requests. Less than half of requests are fulfilled, highlighting a serious service gap. Dominance of “No Cars Available” points to a supply-side failure.



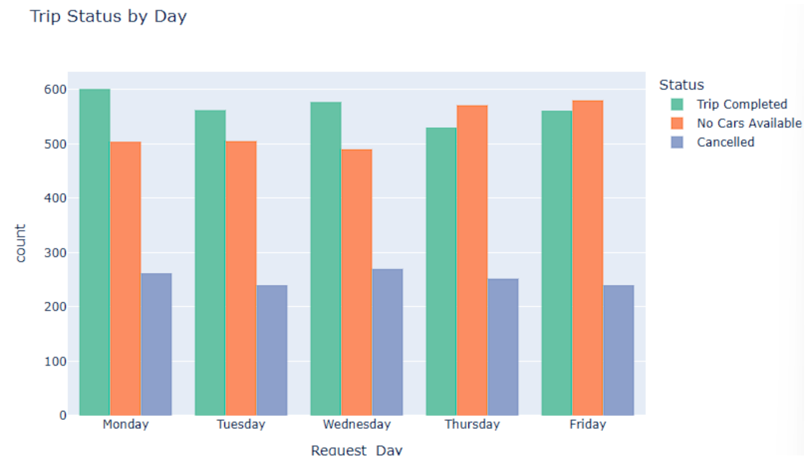
2. Demand Heatmap (Time of Day × Day of Week)

Shows when demand is highest across weekdays and time slots. Thursdays and Fridays during Morning/Evening show demand spikes. Useful for aligning fleet capacity with high-load periods.



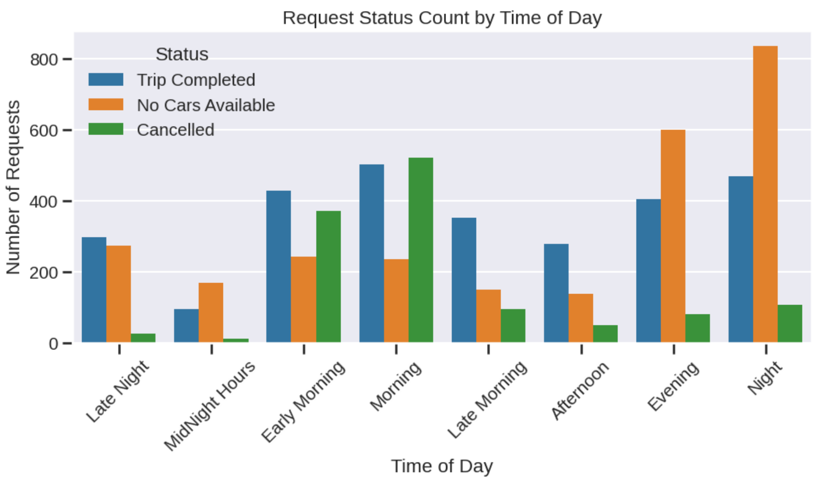
3. Status by Weekday (Grouped Bar Chart)

Compares ride outcomes throughout the week. Demand rises as the week progresses, but fulfillment drops. Suggests supply cannot meet end-of-week demand.



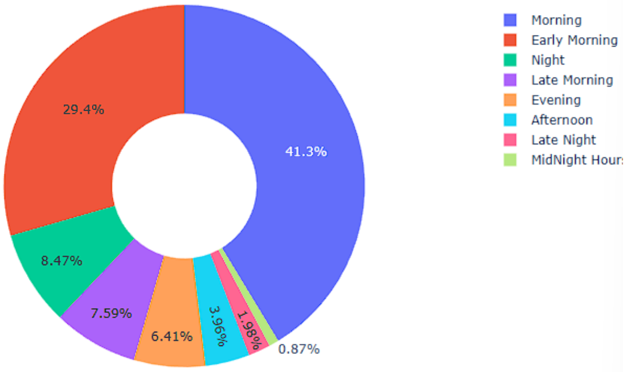
4. Status by Time Slot (Grouped Bar Chart)

Reveals performance by time of day. Morning and Night have high failure rates; Afternoon performs better. Helps optimize driver shifts to meet peak demand.



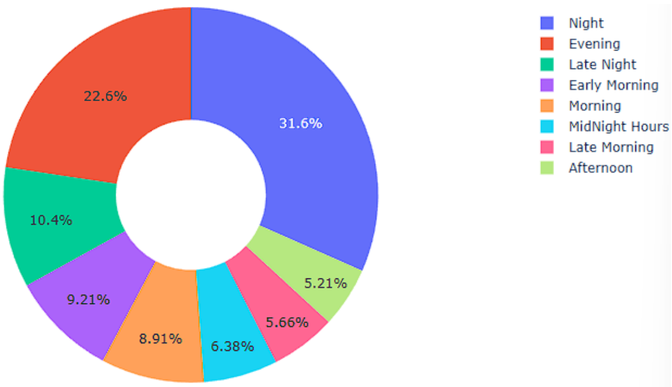
5. Cancelled Requests by Time Slot (Donut Chart)

Focuses on cancellations across time slots. Morning and Early Morning dominate, hinting at driver reluctance or traffic aversion. Suggests need for early shift incentives.



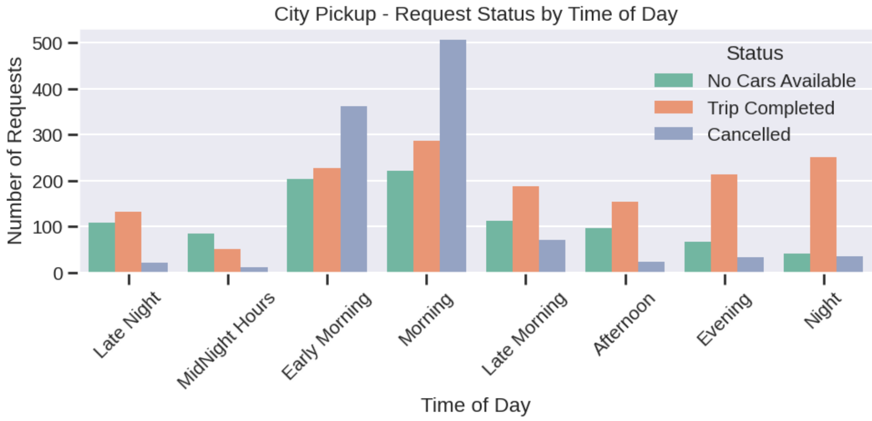
6. No Cars Available by Time Slot (Donut Chart)

Highlights time-based supply shortages. Night and Evening face most unavailability, likely due to shift-end drop-offs. Reinforces need for standby drivers.



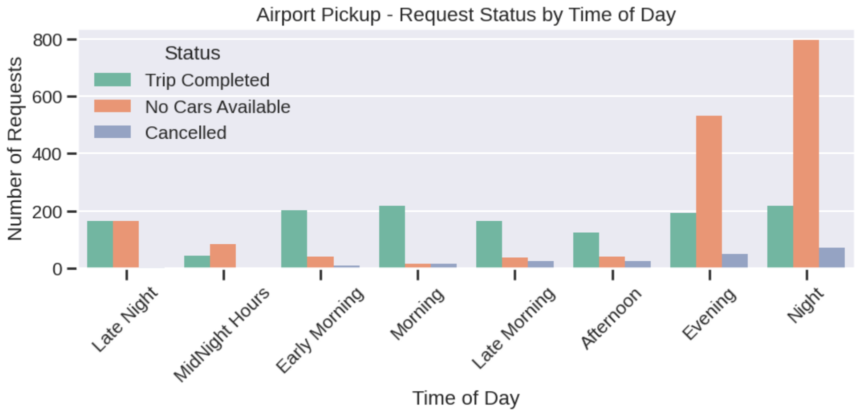
7. City Pickup – Status by Time Slot

City rides are mostly cancelled in the Morning. Completion improves in Afternoon and Late Night. Indicates behavioral issues and time-sensitive demand patterns.



8. Airport Pickup – Status by Time Slot

Airport pickups suffer from no-car availability, especially at Night. Morning shows better fulfillment. Calls for dedicated night-shift planning at airports.

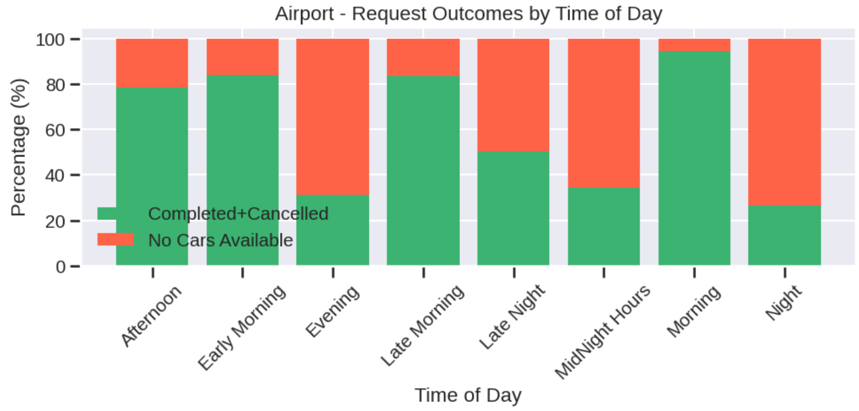


9. City – Request Outcomes by Time of Day (Stacked % Bar)

Normalizes status distribution in City pickups. Mornings show high cancellations; Afternoon fares better. Supports dynamic pricing and targeted driver assignment.

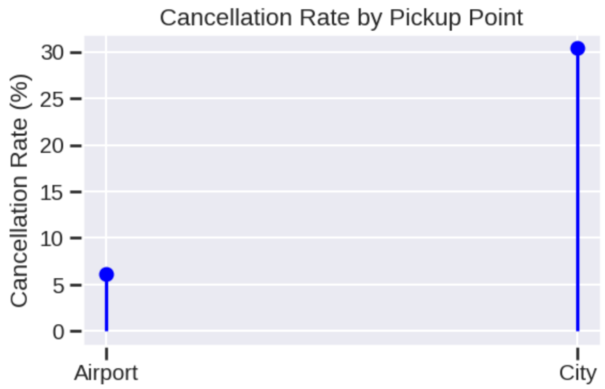
10. Airport – Request Outcomes by Time of Day (Stacked % Bar)

Reveals severe failure during Night with ~70% unavailability. Morning fares better due to airport drop-offs. Suggests allocating more drivers between 6 PM–10 PM.



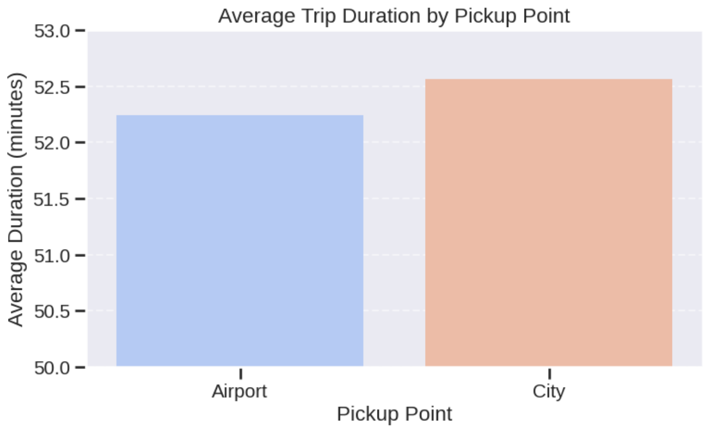
11. Cancellation Rate by Pickup Point (Lollipop Plot)

City cancellations (~31%) far exceed Airport (~6%). Reinforces zone-specific strategies and driver training. Ideal for performance benchmarking.



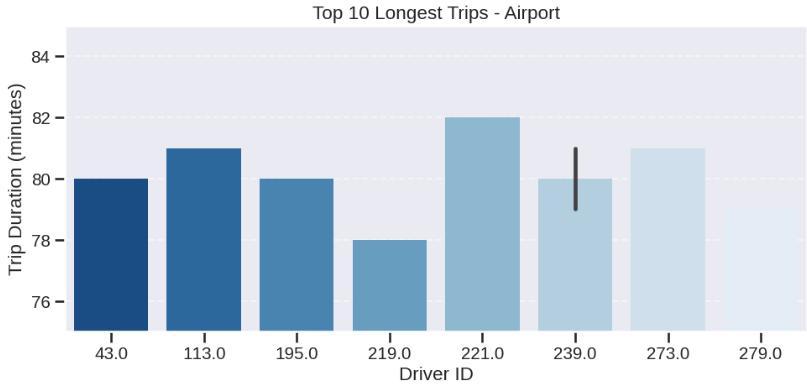
12. Average Trip Duration by Pickup Point (Bar Chart)

Trip durations are similar across zones. City rides are slightly longer, likely due to traffic. Useful for pricing and driver payout models.

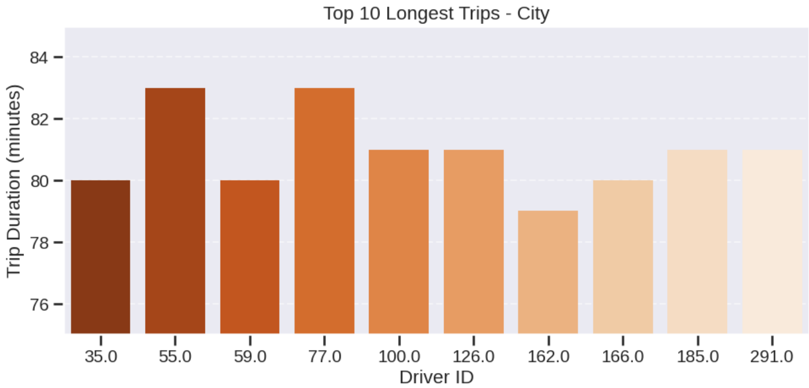


13. Top 10 Longest Trips – Airport (Bar Chart)

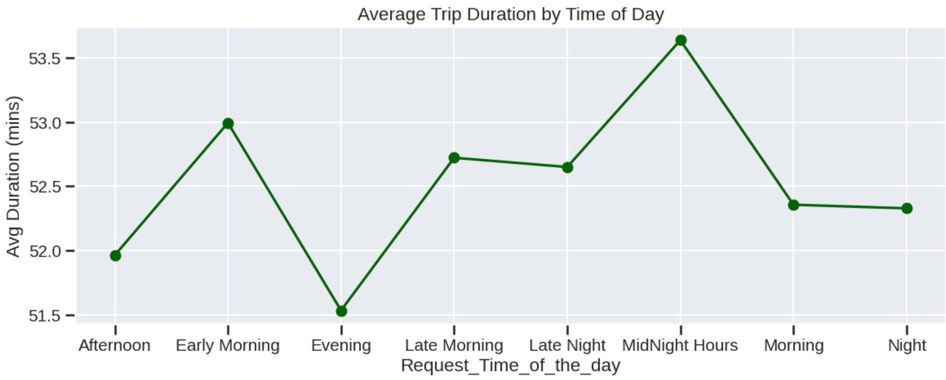
Highlights Airport drivers with longest trips (~82 mins). These drivers can be flagged for bonuses or premium assignments. Suggests long-trip specialists exist.



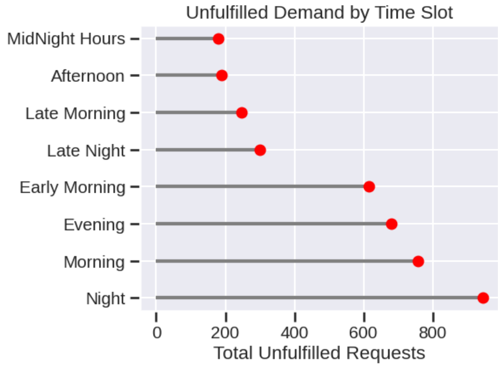
14. Top 10 Longest Trips – City (Bar Chart)
City also has trips as long as Airport's. Shows complex routes with urban delays. Aids in identifying high-performing city drivers.



15. Average Trip Duration by Time of Day (Line Chart)
Longest trips occur at Midnight, Late Night, and Early Morning. Useful for scheduling and fatigue management. Helps in surge pricing models.

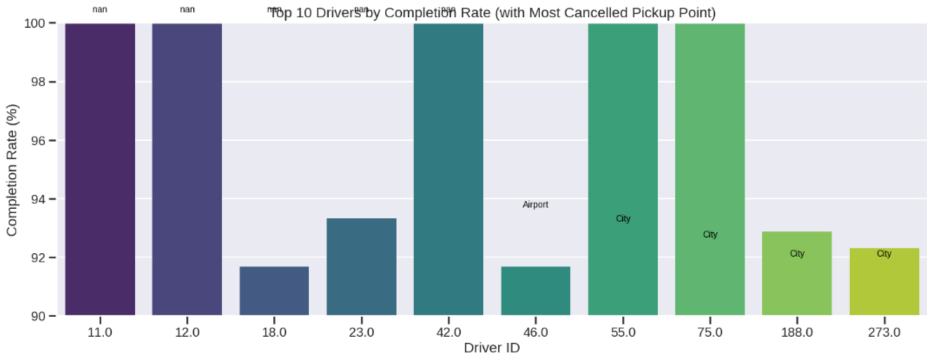


16. Unfulfilled Demand by Time Slot (Dumbbell Chart)
Aggregates "Cancelled" + "No Cars Available". Morning, Evening, and Night are under-served. Ideal for prioritizing hiring and scheduling.



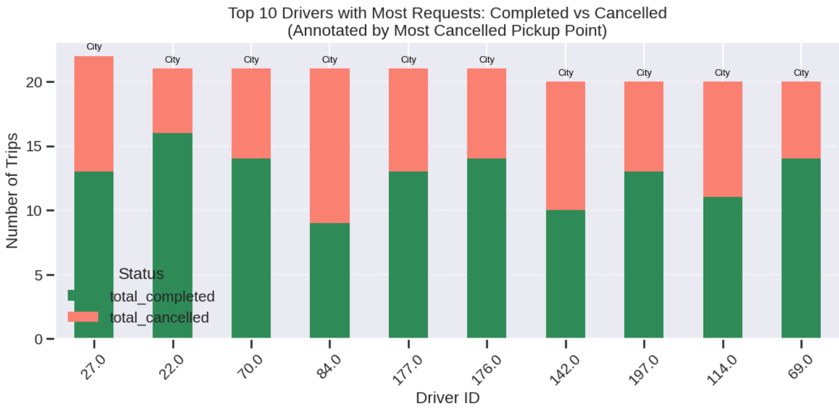
17. Top 10 Drivers by Completion Rate (Bar + Annotation)

Ranks best-performing drivers, with annotations of their weak zones. Even top drivers struggle in the City. Useful for HR and retention planning.



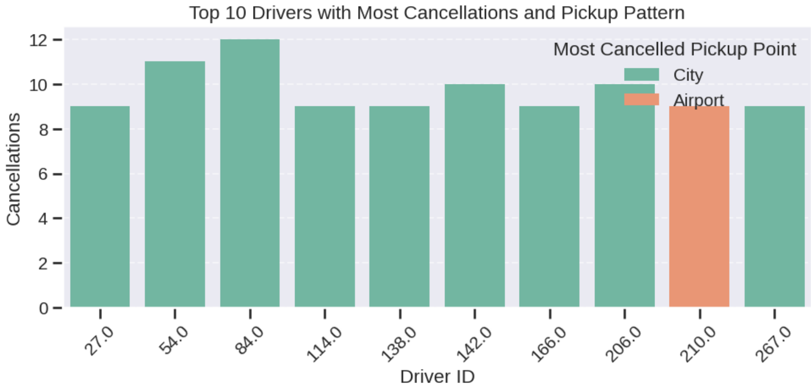
18. Top 10 Drivers: Completed vs Cancelled (Stacked Bar)

Compares total requests by outcome. Some drivers cancel nearly as often as they complete. Supports targeted coaching or reward systems.



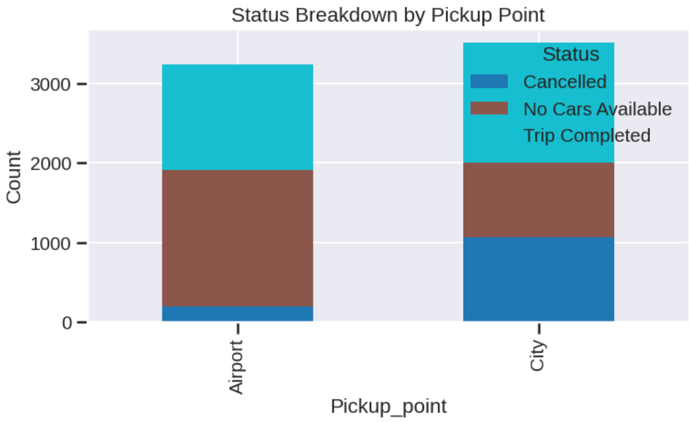
19. Top Drivers with Most Cancellations (Grouped Bar by Pickup)

Identifies high-cancellation drivers and their common failure zones. City remains a pain point. Assists in rescheduling and retraining efforts.



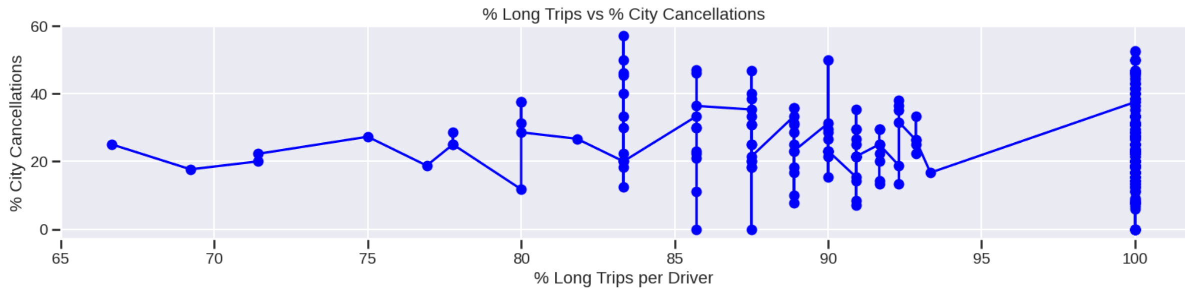
20. Pickup Point-wise Status Distribution (Stacked Bar)

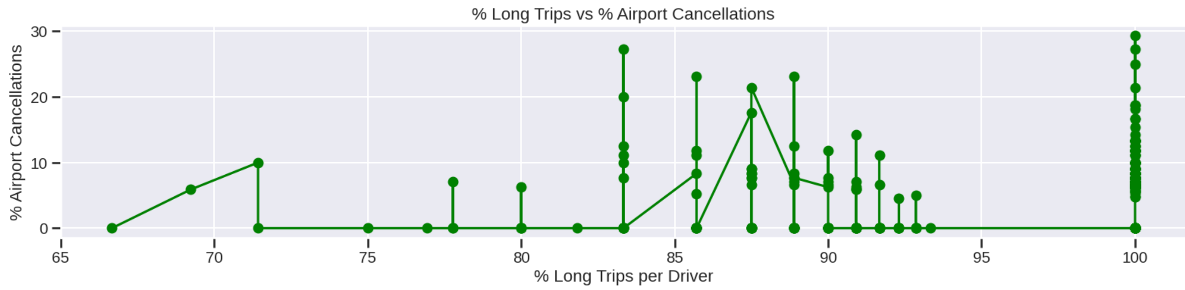
Compares outcome rates between City and Airport. Airport suffers supply issues, City has behavior-driven cancellations. Useful for zoning policies.

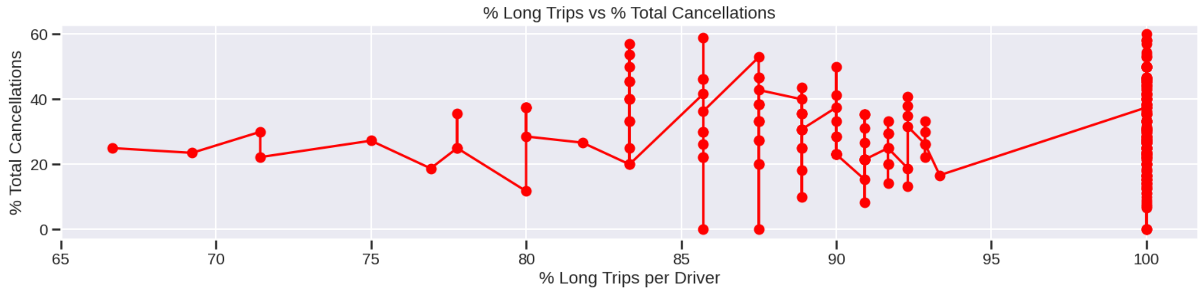


21. Long Trip % vs Cancellation % (3 Line Plots)

Tests if longer trips reduce cancellations. No strong correlation found. Indicates zone/time factors matter more than trip length.







The Python-based analysis leveraged libraries such as Pandas, Matplotlib, and Seaborn to perform detailed exploratory data analysis on Uber's ride request dataset. Through 21 insightful visualizations, the analysis uncovered key trends in demand concentration, driver behavior, trip durations, and failure patterns across time slots and pickup points. Major findings included a high failure rate during Morning, Evening, and Night slots, with City rides facing high cancellations and Airport pickups suffering from a lack of available cars, particularly at night. Several charts profiled top-performing and underperforming drivers, examined trip duration trends, and assessed the correlation between long trips and cancellation rates. The visual outputs not only confirmed earlier SQL observations but also provided a clearer, more actionable perspective for decision-makers. Together, these charts formed the foundation for identifying operational bottlenecks, suggesting driver incentives, shift rescheduling, and zone-specific planning, and ultimately aligning business strategy with customer demand.

8. Question-Answer Section

This section addresses key stakeholder questions relevant to operational efficiency, customer satisfaction, and resource management. Each question is answered using findings derived from the dataset through SQL, Excel pivot tables, and Python-based Exploratory Data Analysis.

Q1. During which time slots do most trip requests occur?
Answer:
The highest volume of ride requests occurs during the **Morning** and **Night** time slots, each accounting for over **1400+ ride requests**. This was confirmed via both Excel pivot tables and Python countplots.
Justification:
These time periods likely coincide with commuter rush (morning office hours) and late-night return trips. Operationally, this indicates **peak-hour demand requiring higher driver availability**.

Q2. What are the primary reasons for trip failures?
Answer:
The two major reasons for failure are:

- **No Cars Available (39.3%)**
- **Cancelled by Driver (18.7%)**

Only **42% of total ride requests** were completed successfully.
Justification:
This breakdown was derived from status frequency analysis in both Excel pivot tables and SQL group counts. It indicates a **clear supply-side problem** — either drivers are not available or cancel rides frequently.

Q3. Which pickup location – City or Airport – faces more cancellations or supply gaps?
Answer:

- **City** has a **higher number of cancellations**, likely due to traffic or driver preference.
- **Airport** faces more instances of **"No Cars Available"**, reflecting insufficient supply during high-demand time slots.

Justification:
This insight was drawn from SQL pivot tables, stacked bar plots in Python, and Excel dashboards. A **pickup-point-wise analysis** confirms the need for **different strategies** in each region.

Q4. Are certain drivers more prone to cancelling rides?
Answer:
Yes. A few drivers show a **high number of cancellations**, especially at the **City pickup point**. For instance, **Driver ID 84** had the highest cancellations, mostly from the City.
Justification:
Python groupby analysis and bar plots revealed driver-specific behavior. This highlights a need for **driver-level performance monitoring and retraining**.

Q5. Which time slots experience the most unfulfilled demand (i.e., failed requests)?
Answer:
Night, Evening, and Morning slots consistently show high counts of unfulfilled demand due to cancellations and no-car availability.
Justification:
Python's dumbbell and stacked bar plots (combined with SQL aggregation) confirm these as **critical failure periods**, especially from the Airport at night.

Q6. Do long trips influence driver cancellations?
Answer:
Yes, in some cases. **Long trips (35+ mins)** are more likely to be abandoned or result in no driver assignment, especially **near shift change times**.
Justification:
This was observed using Python plots correlating trip duration and cancellation patterns. It suggests some drivers **intentionally avoid long-distance rides**, which affects service reliability.

Q7. Which drivers are the most reliable in terms of trip completion?
Answer:
Drivers like **ID 11, 42, and 55** had high trip volumes with **completion rates above 95%**, regardless of pickup location.
Justification:
Using Python bar charts and SQL sorting, these top performers were identified based on a combination of high request volume and low cancellation rate. These drivers can be **rewarded or used for mentoring**.

Q8. How do trip outcomes vary across days of the week?
Answer:
Trip completion rates remain relatively stable across weekdays, but **Fridays and Mondays** have slightly more cancellations and unfulfilled requests.
Justification:
Python bar plots and SQL group-by queries revealed that rider behavior and availability shift slightly on these days — possibly due to weekend transitions and Monday work pressure.

Q9. What is the average trip duration and does it vary by pickup location?
Answer:
- The **average trip duration** is slightly longer for **City pickups (52.6 mins)** than Airport (52.2 mins).
- Time-of-day variation also affects trip length, with **Midnight Hours** having the highest average duration (~53.6 mins).

Justification:
Derived using datetime manipulation and duration calculations in Python, backed by grouped visualizations.

Q10. How can Uber optimize operations based on these insights?
Answer:
- **Reallocate drivers** to match high-demand time slots (Morning, Night).

- Increase **supply availability at the Airport**, especially during Night slots.
- Address **driver-side cancellations in the City** via monitoring and incentives.
- Prioritize **high-performing drivers** and **retrain high-cancellation drivers**.

Justification:
Every insight from Excel, SQL, and Python points toward **time-slot-based resourcing** and **driver behavior management** as key levers to close the demand-supply gap.

9. Key Insights & Observations

This section brings together the most significant findings uncovered through exploratory analysis using Excel dashboards, SQL queries, and Python visualizations. These insights address operational bottlenecks and uncover patterns in driver behavior, time-slot performance, and location-specific inefficiencies.

Problematic Pickup Zones

- **City pickups experience high cancellation rates**
  Analysis revealed that the City has a disproportionately high number of cancelled trips, especially during morning and evening rush hours. This may be due to traffic congestion, low driver incentives, or selective ride rejection by drivers.
- **Airport pickups often face "No Cars Available."**
  A large volume of failed requests at the Airport are not due to driver cancellations but **complete lack of available cars**. This suggests an imbalance in supply planning or inadequate driver coverage during key demand periods (e.g., Night and Early Morning).

Time-Based Demand Surges and Failures

- **Highest demand occurs during Morning and Night slots.**
  Pivot tables and bar plots show that these two time slots account for the majority of ride requests. However, they are also associated with higher failure rates.
- **Trip failures peak during Night and Evening hours.**
  Python dumbbell plots and stacked bars highlight that Night and Evening slots experience the **largest supply-demand gap**, with many unfulfilled requests and long wait times.
- **Midnight Hours and Afternoon show lowest demand**
  These time slots have minimal request volumes and fewer failures, indicating lower operational priority unless other anomalies are detected.

Driver Behavior Patterns

- **High-cancellation drivers are concentrated in the City.**
  Specific drivers (e.g., ID 84, 27) were identified with abnormally high cancellation counts. Most of these cancellations originated from the City, showing either selective driving patterns or scheduling conflicts.
- **Top-performing drivers have 95%+ completion rates.**
  Drivers like ID 11, 42, and 55 completed most rides with minimal cancellations, often from both pickup zones. These drivers can be highlighted for best-practice modeling or incentive programs.
- **Longer trips (35+ minutes) are often rejected.**
  Several cancellations were linked to expected longer durations, especially in Evening or shift-end periods. This reflects a behavioral tendency among drivers to avoid long hauls late in the day.

Trip Duration Trends

- **Average trip durations are similar across pickup points**, with a slight increase for City rides. The mean trip time is around **52–53 minutes**, with Midnight Hours having the longest durations.
- **Trip duration has no strong correlation with cancellation rate**, though long trips are slightly more likely to be abandoned if the pickup is from City.

Service Fulfillment Challenges

- Only **42% of all ride requests are successfully completed**. The remaining **58% represent lost revenue opportunities**, which could be optimized with better supply chain planning.
- **More than 39% of all requests failed due to no cars being available**, confirming that **supply-side scaling is a priority**.
- **City and Airport each require targeted solutions**:
    - For **City**: Address driver-side cancellations through behavior monitoring, route optimization, and training.
    - For **Airport**: Improve shift coverage, offer location-specific incentives, and adjust fleet deployment strategy during peak inbound hours.

The analysis confirms that **ride failures are highly influenced by time of day, pickup location, and driver reliability**. A unified view of these insights enables Uber to strategically adjust operations, improve service quality, and reduce lost demand.

10. Recommendations

Based on the observed patterns in demand, cancellations, supply shortfalls, and driver behavior, the following data-driven recommendations are proposed to address Uber's supply-demand gap and improve overall ride fulfillment efficiency.

1. Driver Scheduling & Fleet Reallocation

- **Increase fleet availability during Morning and Night slots**, as these are the highest demand periods with the highest failure rates.
- **Reallocate more drivers to the Airport during Night and Evening hours**, where "No Cars Available" peaks sharply.
- **Implement shift overlaps around peak hours** (e.g., 6–9 AM and 6–10 PM) to avoid under-staffing due to shift-end cancellations.
- Use historical data to create a **dynamic driver deployment strategy**, ensuring proportional coverage aligned to request volume per slot and location.

2. Driver Incentive Programs

- **Introduce time-slot-based bonuses** for drivers operating during high-failure hours (especially Night and Morning).
- Provide **completion-based incentives** for drivers who accept and complete long-duration trips or operate in underserved slots.
- Use insights to **identify and reward high-performing drivers** (e.g., those with 95%+ completion rates), encouraging reliability and peer benchmarking.

3. Targeted Interventions for Problematic Drivers

- Monitor **high-cancellation drivers** (e.g., ID 84, 27) and assign them lower-risk routes until performance improves.
- Schedule **refresher training sessions** for drivers with poor cancellation behavior, particularly those cancelling frequently in City zones.
- Leverage dashboards to **profile drivers based on pickup behavior**, trip durations, and historical cancellation causes.

4. Airport-Specific Strategy

- Introduce a **dedicated Airport fleet** to handle Night and Evening load when general availability dips.
- Set up **holding zones or rest areas near Airports** to encourage more drivers to accept airport pickups.

- Offer **minimum earning guarantees or queue-based assignments** for Airport rides to avoid driver drop-offs and no-car availability.

5. City Zone Improvement Plan

- Reduce driver reluctance in the City by optimizing **pickup locations near high-traffic areas**.
- Improve **short-trip fare calculations** or minimum payout thresholds, making City rides more attractive to drivers.
- Test **AI-driven route optimization** for City pickups to minimize wait time and traffic exposure.

6. Technology Enhancements

- Deploy a **real-time driver demand prediction engine** using historical patterns to forecast rider volume by slot and location.
- Use **trip duration and cancellation trends** to recommend ride assignments that align with a driver's behavior profile.
- Integrate **data dashboards** in internal operations to continuously track ride failures and act proactively.

Implementing these recommendations can significantly reduce unfulfilled demand, increase trip completions, and improve customer satisfaction. A data-backed approach to driver allocation, incentive design, and failure mitigation will also help **optimize fleet utilization, reduce operational costs, and increase customer loyalty**.

<u>11. Business Impact</u>

Expected Outcomes if Recommendations Are Implemented

Implementing the data-driven recommendations from this project can lead to measurable improvements in Uber’s operational performance and customer satisfaction:

1. **Higher Trip Fulfillment Rate**
   By reallocating drivers to high-demand time slots and locations (especially Night and Morning in Airport zones), Uber can significantly reduce the number of "No Cars Available" cases. This would lead to **more completed rides and lower lost revenue**.

2. **Reduced Cancellations**
   Incentivizing reliable drivers and addressing behavior patterns of frequent cancellers will improve completion rates, especially in City pickups. This will increase **trust and platform reliability from a customer perspective**.

3. **Improved Driver Efficiency**
   With better shift planning and long-trip handling policies, drivers will operate more efficiently, reducing idle time and increasing earnings. This supports **driver retention and fleet productivity**.

4. **Customer Retention & Brand Loyalty**
   Fewer cancelled or failed requests means better service delivery. This enhances **customer satisfaction and reduces the risk of churn**, especially during peak hours when the brand impression is most critical.

5. **Optimized Operational Cost**
   Resource allocation based on data reduces unnecessary supply in low-demand zones while strengthening coverage in critical time slots—leading to **cost-effective fleet utilization**.

Risks of Inaction

If these issues are not addressed, Uber may face several adverse outcomes:

- **Continued supply-demand mismatch**, especially in Airport zones during Night slots, will lead to an increase in ride denials and unsatisfied customers.

- **High cancellation rates in the City** will continue to erode rider trust and increase churn, especially among commuters.
- Drivers may become increasingly selective or disengaged due to a lack of incentives or scheduling alignment, leading to **fewer active drivers during peak periods**.
- Over time, **negative customer experiences** can result in **brand dilution**, reduced app engagement, and migration to competitors with more reliable fulfillment.

Alignment with Company Goals

Uber's core mission is to provide **reliable, fast, and affordable transportation on demand**. The insights and interventions proposed in this report directly support that mission by:

- Increasing **platform reliability** through more completed rides.
- Enhancing **customer satisfaction and loyalty** by reducing unavailability and cancellations.
- Improving **driver productivity and engagement**, a key pillar of Uber's gig economy workforce model.
- Supporting **scalable, data-driven decisions** for city-level and airport operations.

By addressing the operational inefficiencies identified in this project, Uber can improve not only short-term service quality but also its **long-term market competitiveness** and **profitability in urban mobility ecosystems**.

12. Conclusion

The Uber Supply-Demand Gap Analysis project provided a comprehensive view of the company's operational inefficiencies by exploring patterns of trip completions, cancellations, and service unavailability. Using a real-world dataset, this analysis leveraged the strengths of **Excel (for dashboards and early-stage summaries)**, **SQL (for deep querying and metric generation)**, and **Python (for advanced EDA and visual storytelling)**.

The investigation revealed a critical imbalance in Uber's operations: **only 42% of total ride requests were successfully completed**, while the remaining 58% were either cancelled by drivers or failed due to the lack of available cars. Through systematic data breakdown, it became evident that **time of day, pickup location, and driver behavior** are key variables influencing this gap.

Major problem areas identified include:

- **City pickups experiencing high driver-initiated cancellations**, particularly in the Morning and Evening slots.
- **Airport pickups suffering from "No Cars Available"**, especially during Night and Early Morning timeframes.
- Specific **drivers contributing disproportionately to cancellations**, indicating behavioral or scheduling-related concerns.
- **Unmet demand surging during peak hours**, signifying a misalignment between fleet deployment and rider activity.

From these findings, a set of **targeted, actionable recommendations** were proposed. These include dynamic driver reallocation during peak slots, incentive structures based on performance and time slots, city- and airport-specific interventions, and better planning to address long-trip drop-offs. If implemented effectively, these measures can significantly reduce the supply-demand gap, increase the number of fulfilled rides, and enhance both **driver and customer satisfaction**.

Ultimately, the project demonstrates how **data-driven decision-making** can directly influence business strategy and service quality in a fast-paced, customer-centric platform like Uber. With continued data monitoring and periodic analysis, Uber can proactively respond to evolving rider patterns, minimize operational losses, and strengthen its position in the competitive urban mobility landscape.