Uber Supply – Demand Gap Analysis
(Insights Report)

Prepared for :
Labmentix

Submitted by :
Samruddhi Sanap
(Data Analysis Intern)

Executive Summary

This report presents a comprehensive analysis of Uber's operational data to identify and understand the demand-supply gap within its ride-sharing service. Utilizing Python for data preprocessing, MySQL for detailed querying, and Excel for dynamic visualization, the analysis focuses on uncovering patterns related to request status (completed, cancelled, no cars available), hourly trends, pickup point specific issues, and driver performance.

The findings highlight significant challenges during peak hours, particularly for specific pickup points, leading to a substantial number of unfulfilled requests. Key recommendations are provided to mitigate these issues and improve service efficiency.

Project Objective

The primary objective of this project was to:

- Clean and preprocess raw Uber ride request data.
- Perform in-depth analysis using SQL queries to extract meaningful insights into demand-supply dynamics.
- Visualize these insights through an interactive dashboard in Excel.
- Identify critical areas of service inefficiency and provide data-driven recommendations.

Data Cleaning and Preprocessing (Python)

- **Initial Data Overview**

The raw dataset contained information on Uber ride requests, including timestamps (Request timestamp, Drop timestamp), Pickup point, Driver id, Trip id, and Status. Initial inspection revealed the need for data type conversions and feature engineering to facilitate analysis.

- **Preprocessing Steps**

Using Python (in a Google Colab environment) and the Pandas library, the following steps were performed to clean and prepare the data:

- **Missing Values:** Handled instances of missing Driver id and Drop timestamp for cancelled or 'no cars available' trips, which are expected in such scenarios.
- **Data Type Conversion:** Converted Request timestamp and Drop timestamp columns to appropriate datetime objects.
- **Feature Engineering:**
    - RequestHour was extracted from Request timestamp (0-23).
    - Trip Duration (minutes) was calculated for completed trips. Warnings about data truncation were noted during this process, indicating potential precision loss for very long decimal values.

    - Request Day of Week was derived from Request timestamp for daily pattern analysis.
- **Consistency Checks:** Ensured consistency in categorical fields like Pickup point and Status.
- **Output**

The cleaned and transformed data was saved as uber_data_cleaned.csv and used for subsequent SQL analysis.

Data Analysis and Querying (SQL - MySQL)

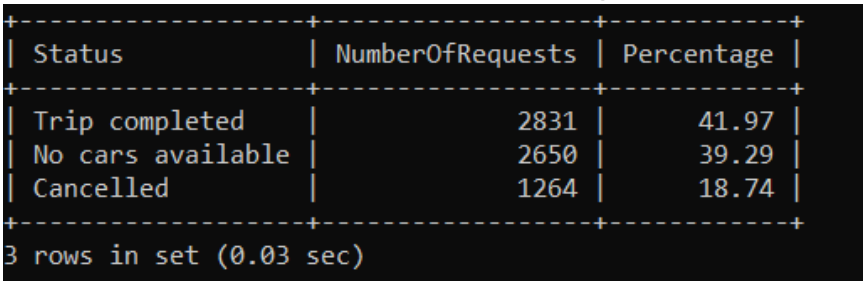
- **Database Setup and Data Import**

A MySQL database was set up, and the uber_data_cleaned.csv file was successfully imported into a table named uber_data_cleaned.

- **Key SQL Queries and Insights**

Various SQL queries were executed to extract specific insights from the cleaned data:

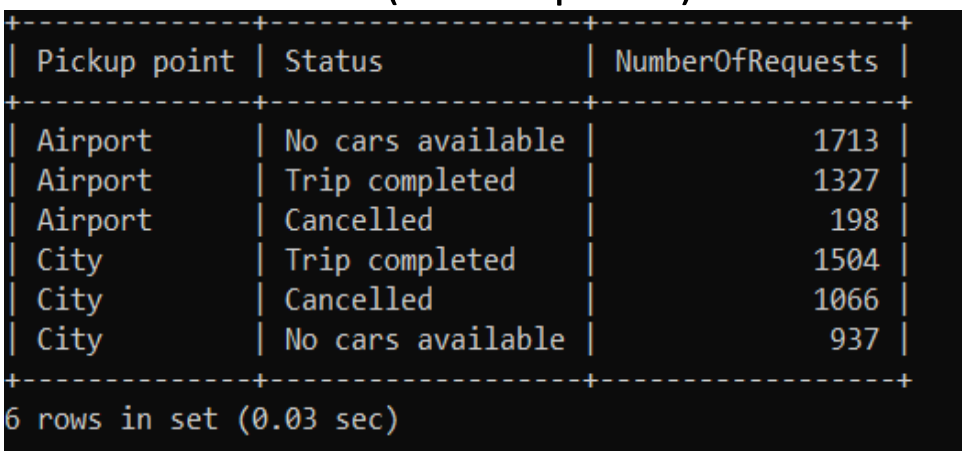
- **Query 1 (Enhanced by Query 5): Overall Request Status Distribution**
  - **Purpose:** To understand the high-level breakdown of all Uber requests by their final status (Completed, Cancelled, No cars available), including their proportional representation.
  - **Insight:** Provides a fundamental understanding of service fulfillment rates and the overall scale of unfulfilled demand across the entire dataset.
  - **Result:** Out of all requests, **41.97% were 'Trip completed' (2831 requests)**, **39.29% were 'No cars available' (2650 requests)**, and **18.74% were 'Cancelled' (1264 requests)**.



Status	NumberOfRequests	Percentage
Trip completed	2831	41.97
No cars available	2650	39.29
Cancelled	1264	18.74

3 rows in set (0.03 sec)

- **Query 2: Request Status Distribution by Pickup Point**
  - **Purpose:** To analyze how different statuses vary between 'Airport' and 'City' pickup points, revealing location-specific challenges.
  - **Insight:** Highlights distinct operational challenges and success rates specific to each major location, showing where each type of unfulfillment is more prevalent.
  - **Result:**
    - **Airport:** Most common status was **'No cars available' (1713 requests)**, followed by 'Trip completed' (1327 requests) and 'Cancelled' (198 requests).
    - **City:** Most common status was **'Trip completed' (1504 requests)**, followed by **'Cancelled' (1066 requests)** and 'No cars available' (937 requests).



Pickup point	Status	NumberOfRequests
Airport	No cars available	1713
Airport	Trip completed	1327
Airport	Cancelled	198
City	Trip completed	1504
City	Cancelled	1066
City	No cars available	937

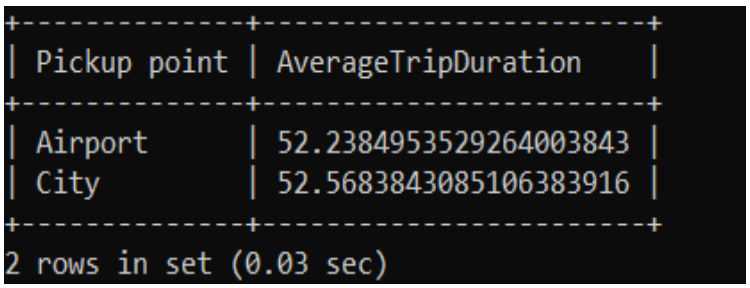
6 rows in set (0.03 sec)

- **Query 3: Overall Hourly Demand and Supply Patterns**
  - **Purpose:** To identify hourly trends in total requests, completed trips, cancellations, and "no cars available" instances across all locations.
  - **Insight:** Reveals peak demand times and corresponding service failures, indicating precisely when the supply-demand gap is most pronounced throughout the day.
  - **Result:** Analysis showed significant fluctuations. 'No cars available' requests sharply increased in the evening hours (e.g., from 17:00 to 22:00, reaching highs of **322 at 18:00** and **290 at 20:00**). 'Cancelled' requests were notable during morning peak hours (e.g., **176 at 05:00** and **175 at 09:00**).

RequestHour	TripCompleted	Cancelled	NoCarsAvailable	TotalRequests
0	40	3	56	99
1	25	4	56	85
2	37	5	57	99
3	34	2	56	92
4	78	51	74	203
5	185	176	84	445
6	167	145	86	398
7	174	169	63	406
8	155	178	90	423
9	173	175	83	431
10	116	62	65	243
11	115	15	41	171
12	121	19	44	184
13	89	18	53	160
14	88	11	37	136
15	102	21	48	171
16	91	22	46	159
17	151	35	232	418
18	164	24	322	510
19	166	24	283	473
20	161	41	290	492
21	142	42	265	449
22	154	12	138	304
23	103	10	81	194

24 rows in set (0.03 sec)

- **Query 4: Daily Request Status Summary**
  - **Purpose:** To summarize daily request counts, cancelled, and no-cars-available requests by the day of the week.
  - **Insight:** Reveals day-of-the-week patterns in demand and fulfillment challenges.
  - **Result:** The data provided daily breakdowns, for instance, **Thursday had 823 total requests** with 252 cancelled and 571 no cars available.



Pickup point	AverageTripDuration
Airport	52.2384953529264003843
City	52.5683843085106383916

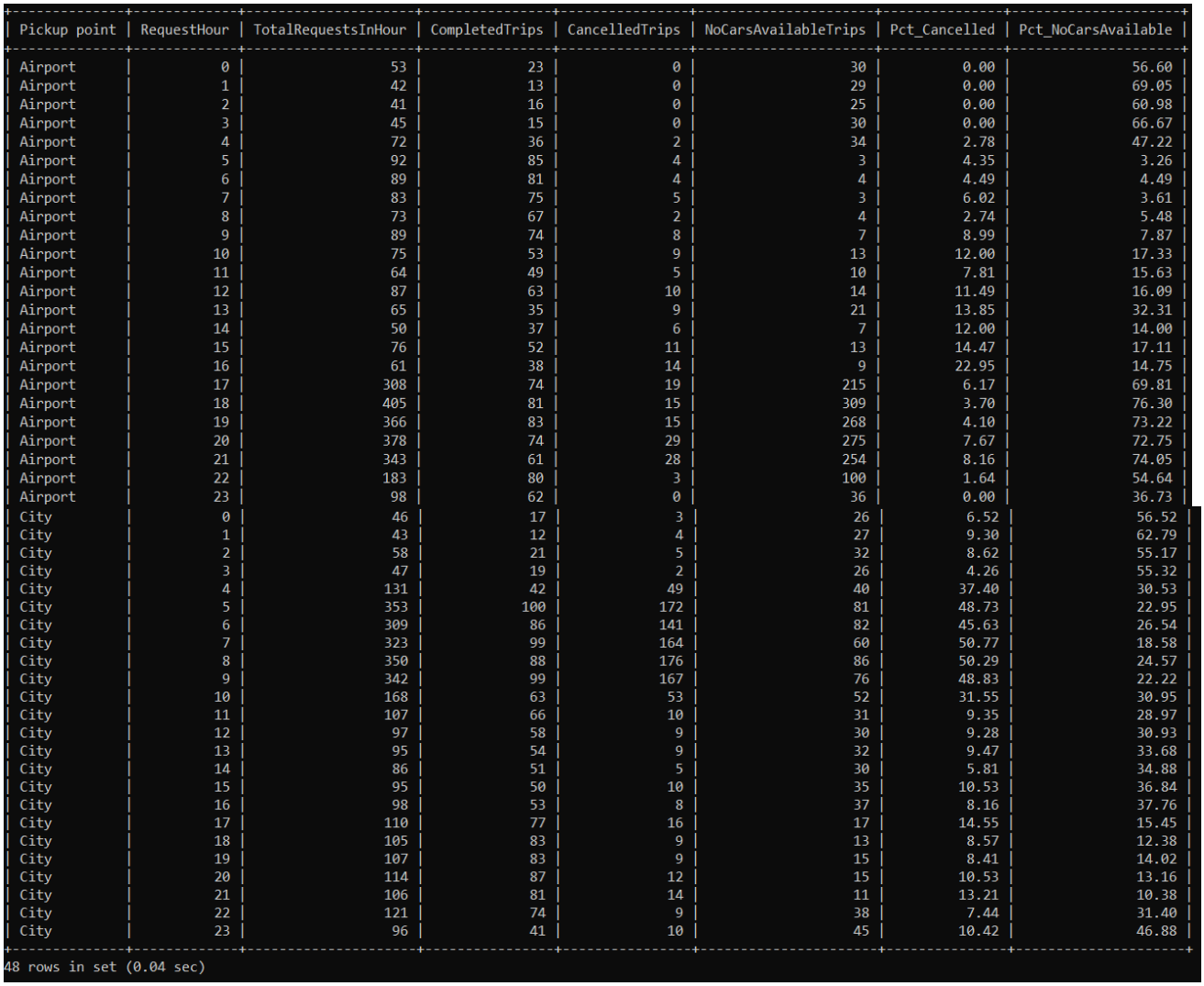
2 rows in set (0.03 sec)

- **Query 6: Average Trip Duration by Pickup Point**
  - **Purpose:** To compare the average duration of completed trips originating from 'Airport' vs. 'City'.
  - **Insight:** Helps understand typical travel times and potential differences in route complexities or service patterns for different pickup locations.
  - **Result:**
    - **Airport:** Average Trip Duration was approximately **52.24 minutes**.
    - **City:** Average Trip Duration was approximately **52.57 minutes**.
- This indicates very similar average trip durations for completed rides from both locations.

Request Day of Week	CancelledRequests	NoCarsAvailableRequests	TotalRequests
Thursday	252	571	823
Friday	240	580	820
Monday	262	504	766
Wednesday	270	490	760
Tuesday	240	505	745

5 rows in set (0.02 sec)

- **Query 7: Hourly Percentage of "No Cars Available" vs. "Cancelled" by Pickup Point**
    - **Purpose:** To drill down into the percentage of service failures (cancelled and no cars available) on an hourly basis, specific to each pickup point.
    - **Insight:** Pinpoints exact hours and locations where supply issues or rider cancellations are most prevalent, allowing for targeted, time-sensitive interventions.
    - **Result:** This query provided the granular percentages for both **'Pct_Cancelled' and 'Pct_NoCarsAvailable'** per hour per pickup point. This data was crucial for creating a dynamic visual in Excel to show these percentages.

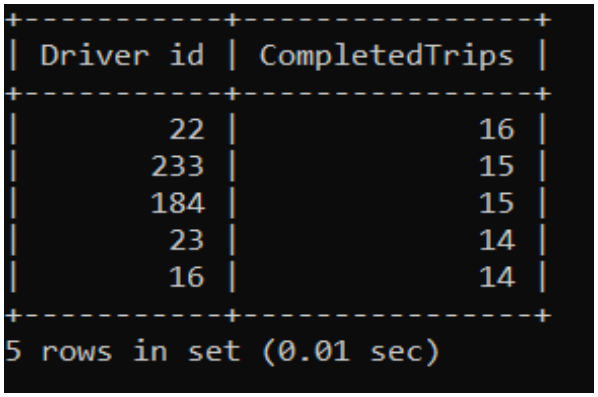


Pickup point	RequestHour	TotalRequestsInHour	CompletedTrips	CancelledTrips	NoCarsAvailableTrips	Pct_Cancelled	Pct_NoCarsAvailable
Airport	0	53	23	0	30	0.00	56.60
Airport	1	42	13	0	29	0.00	69.05
Airport	2	41	16	0	25	0.00	60.98
Airport	3	45	15	0	30	0.00	66.67
Airport	4	72	36	2	34	2.78	47.22
Airport	5	92	85	4	3	4.35	3.26
Airport	6	89	81	4	4	4.49	4.49
Airport	7	83	75	5	3	6.02	3.61
Airport	8	73	67	2	4	2.74	5.48
Airport	9	89	74	8	7	8.99	7.87
Airport	10	75	53	9	13	12.00	17.33
Airport	11	64	49	5	10	7.81	15.63
Airport	12	87	63	10	14	11.49	16.09
Airport	13	65	35	9	21	13.85	32.31
Airport	14	50	37	6	7	12.00	14.00
Airport	15	76	52	11	13	14.47	17.11
Airport	16	61	38	14	9	22.95	14.75
Airport	17	308	74	19	215	6.17	69.81
Airport	18	405	81	15	309	3.70	76.30
Airport	19	366	83	15	268	4.10	73.22
Airport	20	378	74	29	275	7.67	72.75
Airport	21	343	61	28	254	8.16	74.05
Airport	22	183	80	3	100	1.64	54.64
Airport	23	98	62	0	36	0.00	36.73
City	0	46	17	3	26	6.52	56.52
City	1	43	12	4	27	9.30	62.79
City	2	58	21	5	32	8.62	55.17
City	3	47	19	2	26	4.26	55.32
City	4	131	42	49	40	37.40	30.53
City	5	353	100	172	81	48.73	22.95
City	6	309	86	141	82	45.63	26.54
City	7	323	99	164	60	50.77	18.58
City	8	350	88	176	86	50.29	24.57
City	9	342	99	167	76	48.83	22.22
City	10	168	63	53	52	31.55	30.95
City	11	107	66	10	31	9.35	28.97
City	12	97	58	9	30	9.28	30.93
City	13	95	54	9	32	9.47	33.68
City	14	86	51	5	30	5.81	34.88
City	15	95	50	10	35	10.53	36.84
City	16	98	53	8	37	8.16	37.76
City	17	110	77	16	17	14.55	15.45
City	18	105	83	9	13	8.57	12.38
City	19	107	83	9	15	8.41	14.02
City	20	114	87	12	15	10.53	13.16
City	21	106	81	14	11	13.21	10.38
City	22	121	74	9	38	7.44	31.40
City	23	96	41	10	45	10.42	46.88

48 rows in set (0.04 sec)

- **Query 8: Top 5 Drivers by Completed Trips**
  - **Purpose:** To identify the drivers with the highest number of completed trips, highlighting individual contributions to service delivery.
  - **Insight:** Useful for performance review, recognizing top performers, and understanding the capacity contributions of highly active drivers.
  - **Result:** The top 5 drivers by completed trips were: **Driver 22 (16 trips)**, **Driver 233 (15**

trips), **Driver 184 (15 trips)**, **Driver 23 (14 trips)**, and **Driver 16 (14 trips)**.



Driver id	CompletedTrips
22	16
233	15
184	15
23	14
16	14

5 rows in set (0.01 sec)

Data Visualization and Dashboard (Excel)

- **Data Import and Preparation in Excel**

The results from the SQL queries were exported as CSV files using the SELECT ... INTO OUTFILE command. These CSV files were then imported into separate sheets in Microsoft Excel. Column headers, which are not included by default with INTO OUTFILE, were manually added to each sheet to ensure proper data interpretation and chart creation in Excel.

- **Dashboard Overview**

A comprehensive and interactive dashboard, named "**Uber Demand-Supply Gap Analysis Dashboard**," was created in Excel to visually represent the extracted insights. The dashboard facilitates easy exploration of the data.

- **Dashboard Visuals**

Chart 1: Overall Hourly Demand and Supply (All Pickups)

- **Type:** Combo Chart (Columns for Cancelled, NoCarsAvailable, TotalRequests; Line for TripCompleted).
- **Purpose:** Provides a holistic view of hourly demand and fulfillment across all locations.
- **Key Insight:** Clearly illustrates the significant gap between TotalRequests and TripCompleted during both morning (e.g., 5-10 AM) and evening (e.g., 5-9 PM) peak hours. The chart explicitly shows the rise in NoCarsAvailable particularly in the evening and Cancelled requests dominating the morning unfulfilled demand.

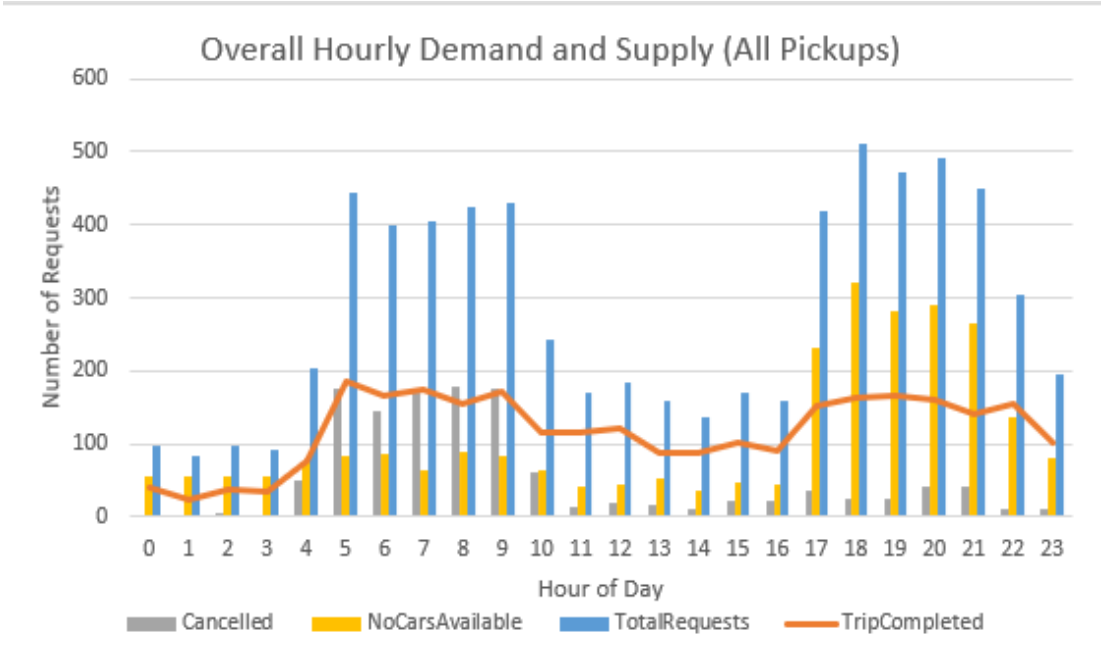


Chart 2: Uber Request Status by Pickup Point

- **Type:** Clustered Column Chart (generated via PivotChart).
- **Purpose:** Compares the distribution of request statuses specifically for Airport vs. City pickups.
- **Key Insight:** This chart vividly confirms the distinct problems: Airport experiences a much higher volume of No cars available requests (1713) compared to Trip completed (1327) and Cancelled (198). Conversely, City has a higher number of Trip completed (1504), but also a significant count of Cancelled requests (1066), which is higher than 'No cars available' (937) for the City.

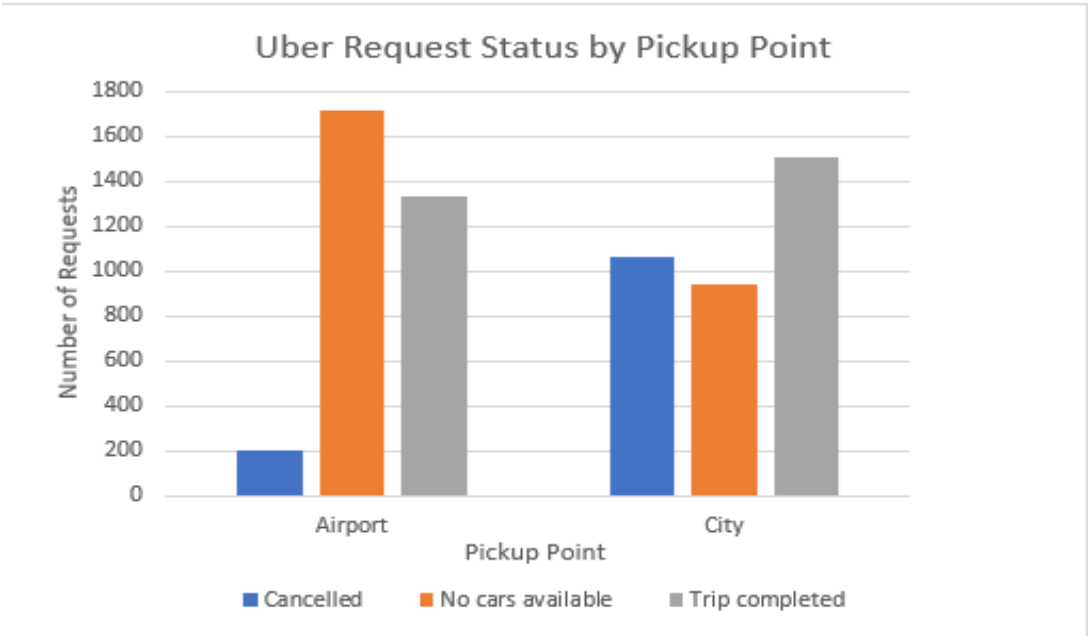


Chart 3: Hourly Percentage of Unfulfilled Requests by Pickup Point

- **Type:** Clustered Column Chart (generated via PivotChart with a dynamic 'Pickup Point' filter).
- **Purpose:** Offers a dynamic, granular view of the percentage of cancelled and no-cars-available trips throughout each hour, allowing specific analysis for Airport or City.
- **Key Insight:** When filtered for 'Airport', the chart would show very high percentages of 'No cars available' (approaching 100% in some evening hours, e.g., 18:00, 19:00). When filtered for 'City', it would show higher percentages of 'Cancelled' trips, especially during morning peaks.

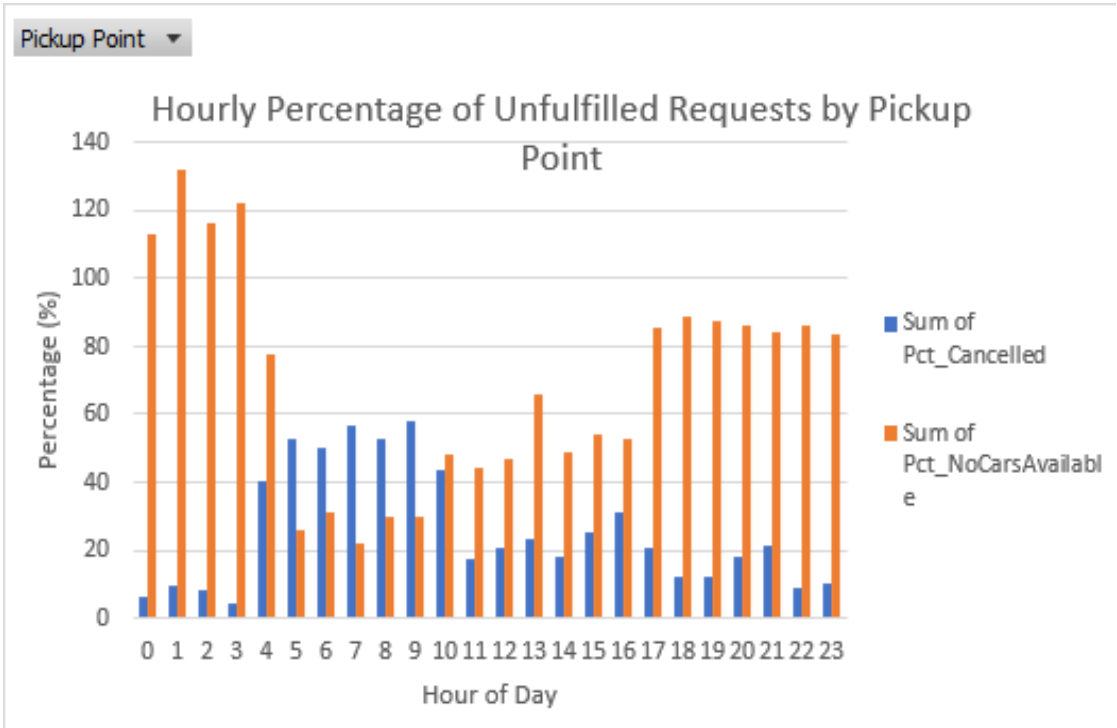
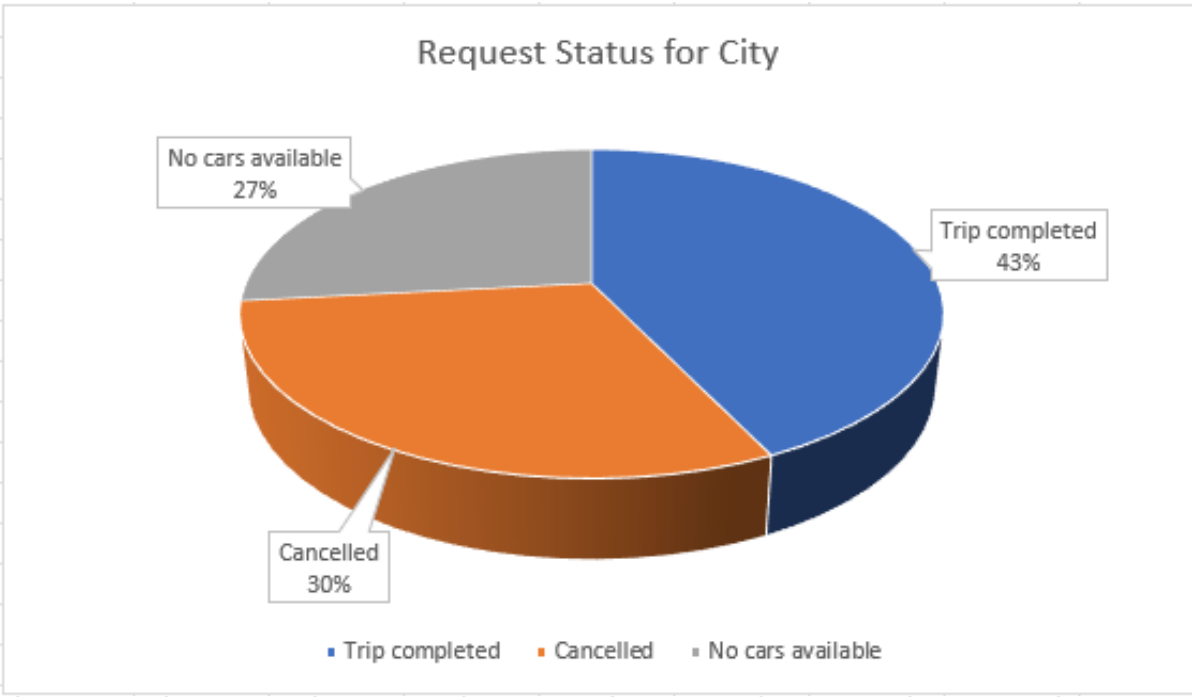


Chart 4 & 5 : Request Status for City & Request Status for Airport

- **Type:** Pie Charts.
- **Purpose:** Provide granular percentage breakdowns of request statuses for each pickup point individually, offering more detailed insights than the aggregated clustered column chart (Chart 2) for specific location analysis.
- **Key Insight:** The "Request Status for City" pie shows 'Trip completed' at 43%, 'Cancelled' at 30%, and 'No cars available' at 27%. The "Request Status for Airport" pie highlights 'No cars available' as the dominant issue at 53%, with 'Trip completed' at 41%, and 'Cancelled' at a much lower 6%. This vividly contrasts the primary unfulfillment reasons at each location.


Request Status for City
No cars available
27%
Trip completed
43%
Cancelled
30%
Trip completed
Cancelled
No cars available

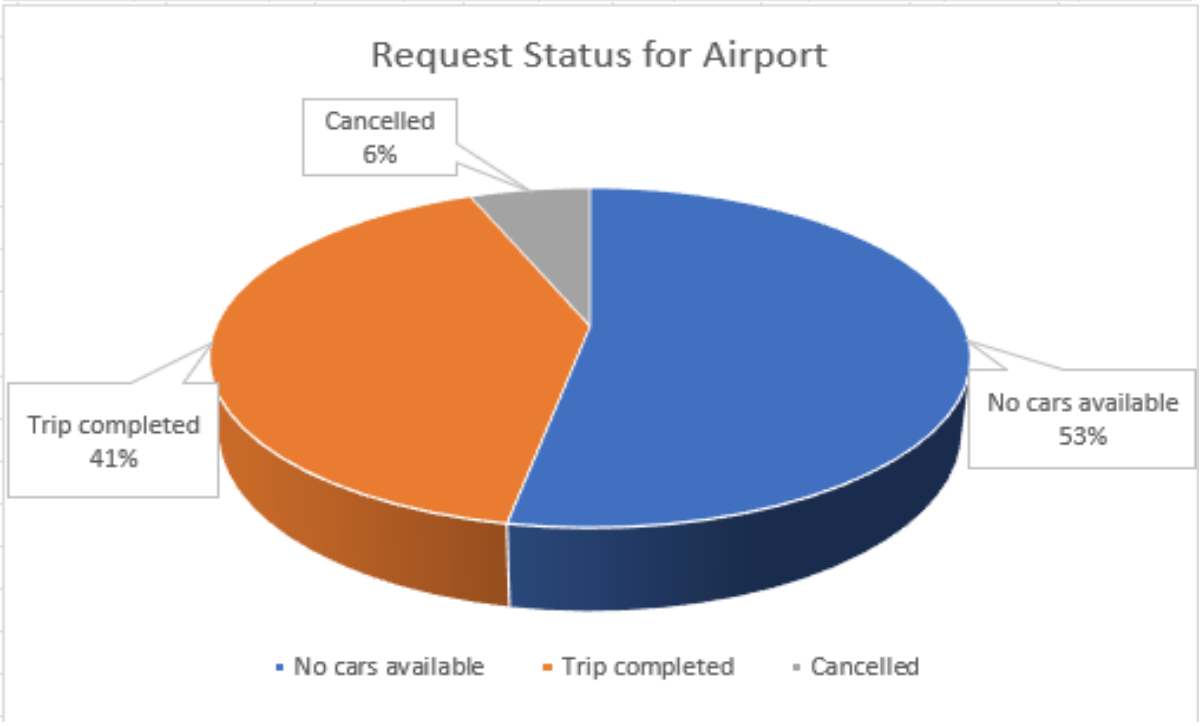
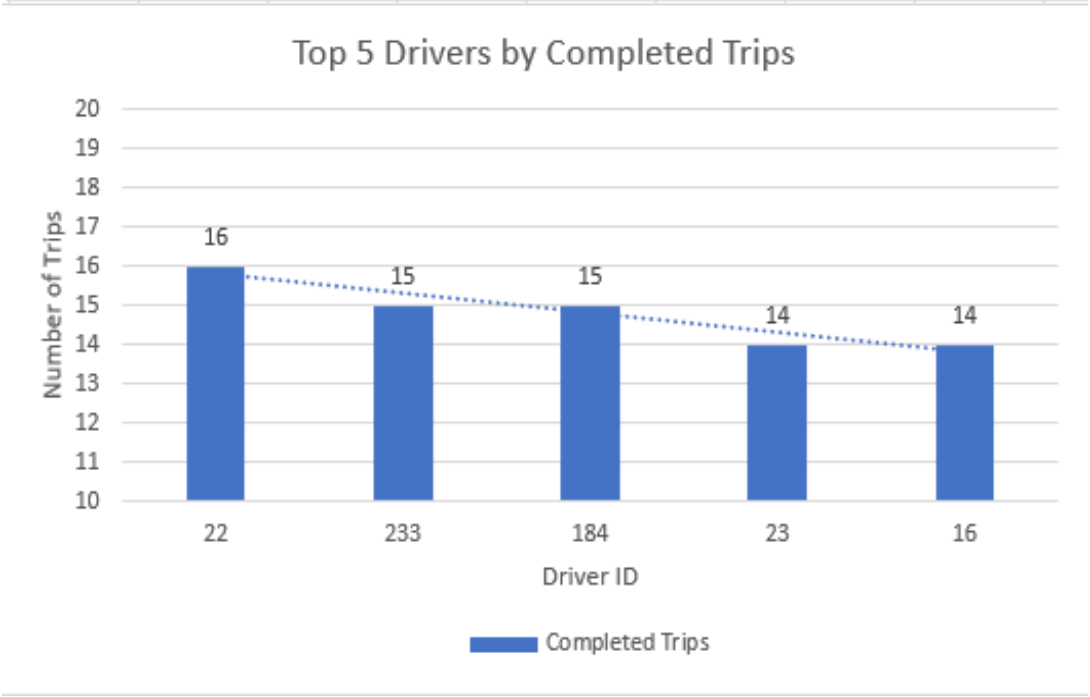

Request Status for Airport
Cancelled
6%
Trip completed
41%
No cars available
53%
No cars available
Trip completed
Cancelled

Chart 6: Top 5 Drivers by Completed Trips

- **Type:** Bar Chart .
- **Purpose:** Identifies and ranks the highest-performing drivers based on their number of completed trips, highlighting individual contributions to overall service delivery.
- **Key Insight:** Clearly displays the top drivers and their completed trip counts, e.g., Driver 22 with 16 trips, and Drivers 233 and 184 both with 15 trips.



Key Findings and Recommendations

Based on the comprehensive analysis, the following critical insights into Uber's demand-supply gap and operational challenges are identified, along with actionable recommendations:

- **Key Findings:**

- **Significant Overall Unfulfillment:** A substantial majority of requests (approximately **58%**) are unfulfilled (39% 'No cars available', 19% 'Cancelled'), indicating a major systemic inefficiency in service delivery.
- **Pronounced Peak Hour Demand-Supply Mismatch:** The periods of highest demand, particularly **morning (5 AM - 10 AM)** and **evening (5 PM - 9 PM)**, witness the largest discrepancies between total requests and completed trips.
- **Airport's Severe "No Cars Available" Crisis:** The Airport pickup point suffers critically from a lack of available cars, especially during the **evening peak hours (e.g., 6 PM onwards)**, leading to over half (53%) of its requests going unfulfilled due to car unavailability.
- **City's Predominant Cancellation Challenge:** The City pickup point experiences a **higher proportion of cancelled trips (30%)**, particularly during the **morning peak hours**. This suggests issues like long

rider wait times or drivers declining short-distance or unfavorable trips.

- **Consistent Trip Durations:** Despite differing demand patterns and unfulfillment reasons, the average trip duration for completed rides is remarkably similar for both Airport and City pickups, hovering around 52 minutes.
  - **Recommendations:**
- **Dynamic Pricing & Incentives:** Implement targeted surge pricing and attractive driver incentives for the **Airport during evening peak hours (5 PM - 9 PM)** to encourage driver repositioning and meet the acute demand.
- **Enhanced Driver Availability Forecasting:** Develop and utilize predictive models to accurately anticipate demand spikes and driver availability in specific zones (Airport vs. City) on an hourly basis. This allows for proactive driver positioning and resource allocation.
- **Targeted Driver Campaigns:** Launch specific campaigns or bonus programs to attract drivers to operate more frequently from the **Airport** during critical evening periods to directly address the chronic "No cars available" issue there.
- **Optimize Dispatch Algorithms for City:** Review and fine-tune dispatch algorithms for **City pickups**, especially during morning peaks, focusing on reducing rider wait times and minimizing driver

cancellations by optimizing route assignments and matching efficiency.

- **Improved Rider Communication:** Implement clearer communication to riders regarding potential wait times or car availability in specific areas during peak times. This can manage expectations, reduce frustration, and potentially decrease rider-initiated cancellations.
- **Overall Driver Pool Expansion:** Continue strategies for robust driver retention and recruitment to increase the overall supply, ensuring a healthier balance across all times and locations, and building resilience against demand surges.

Conclusion

This comprehensive analysis has successfully identified and quantified the critical factors contributing to Uber's demand-supply gap, moving beyond anecdotal observations to data-driven insights. The interactive Excel dashboard provides a clear, visual representation of these problems, pinpointing specific hours and locations where interventions are most needed. By implementing the suggested data-driven recommendations, Uber can significantly enhance service efficiency, reduce the number of unfulfilled requests, and ultimately improve overall customer and driver satisfaction, fostering a more reliable and effective ride-sharing ecosystem.