Bangalore

Uber Ride Analysis

by **Debabrata Palit**

SELF INTRODUCTION

Hello everyone,

My name is **Debabrata Palit.**
An aspiring Data Analyst with recent internship experiences at KultureHire and Cognifyz Technologies.

debabratapalit98@gmail.com

/debabrata-palit03

/Debabrata-palit

ABOUT UBER

Uber Technologies Inc., founded in March 2009, is an American multinational transportation company that provides ride-hailing services, courier services, food delivery, and freight transport. It operates in approximately 70 countries and 10,500 cities worldwide. It is the largest ridesharing company worldwide with over 150 million monthly active users and 6 million active drivers and couriers.

PROJECT OBJECTIVE

This project aims to analyze **Uber ride data** from **Bangalore** (**January 2024**) to extract meaningful business insights. By leveraging **PostgreSQL** for data storage and querying, the analysis focuses on key aspects such as:

- Identifying peak ride times and high-demand locations.
- Understanding cancellation patterns (driver vs. customer-initiated).
- Examining revenue distribution across vehicle types and locations.
- Evaluating customer satisfaction through ride ratings.
- Improving operational efficiency by analyzing turnaround times.

DATASETS OVERVIEW

booking_details.csv

	A	B	C	D
1	Date	Time	Booking ID	Booking Status
2	1/1/2024	0:00:00	BKG000001	Success
3	1/1/2024	0:00:00	BKG000002	Success
4	1/1/2024	0:00:00	BKG000003	Success
5	1/1/2024	0:00:00	BKG000004	Success
6	1/1/2024	0:00:00	BKG000005	Cancelled by Driver
7	1/1/2024	0:00:00	BKG000006	Cancelled by Driver
8	1/1/2024	0:00:00	BKG000007	Success
9	1/1/2024	0:00:00	BKG000008	Incomplete
10	1/1/2024	0:00:00	BKG000009	Success
11	1/1/2024	0:00:00	BKG000010	Success
12	1/1/2024	0:00:00	BKG000011	Success
13	1/1/2024	0:00:00	BKG000012	Success
14	1/1/2024	0:01:00	BKG000013	Cancelled by Driver
15	1/1/2024	0:01:00	BKG000014	Success
16	1/1/2024	0:01:00	BKG000015	Success
17	1/1/2024	0:01:00	BKG000016	Success
18	1/1/2024	0:01:00	BKG000017	Success
19	1/1/2024	0:02:00	BKG000018	Success
20	1/1/2024	0:02:00	BKG000019	Incomplete
21	1/1/2024	0:02:00	BKG000020	Success
22	1/1/2024	0:02:00	BKG000021	Success
23	1/1/2024	0:02:00	BKG000022	Success
24	1/1/2024	0:02:00	BKG000023	Success
25	1/1/2024	0:03:00	BKG000024	Success

ride_details.csv

	A	B	C	D	E	F	G	H	I	J
1	Booking ID	Vehicle Type	Pickup Location	Drop Location	Avg VTAT	Avg CTAT	Price	Payment Method	Ride Distance	Customer Ratings
2	BKG000001	UberXL	Vidyaranyapura	Cox Town	4.5	3.2	271.1	UPI	25.12	2.6
3	BKG000002	Premier	Koramangala	Koramangala	12.19	11.13	1332.2	Card	14.34	2.4
4	BKG000003	eBike	RT Nagar	Richmond Town	3.76	3.34	1444.3	Card	12.29	4.9
5	BKG000004	Moto	BTM Layout	Hosur Road	11.28	5.58	1497.07	Wallet	10.96	3.3
6	BKG000005	Uber Go	Yelahanka	Kammanahalli	5.06	9.56	0		0	
7	BKG000006	Moto	KR Puram	HSR Layout	6.45	4.37	0		0	
8	BKG000007	Moto	MG Road	Vijayanagar	6.39	11.86	1060.65	Wallet	21.19	4.6
9	BKG000008	Uber Go	Marathahalli	Richmond Town	8.45	7.63	0		0	
10	BKG000009	Non-AC Taxi	Bannerghatta Road	Malleswaram	5.12	10.41	1825.17	Card	22.95	1.4
11	BKG000010	Moto	Yeshwanthpur	Kammanahalli	6.13	11.92	1694.59	Card		
12	BKG000011	Moto	Mahadevapura	Cox Town	14.41	7.57	476.22	UPI		
13	BKG000012	eBike	Kudlu Gate	Ulsoor	9.19	4.27	1694.37	Cash		
14	BKG000013	Premier	Cox Town	Kammanahalli	11.16	3.89	0			
15	BKG000014	Moto	MG Road	Kengeri	11.01	10.86	1850.88	Cash		
16	BKG000015	eBike	Bommanahalli	Malleswaram	14.65	6.08	257.37	UPI		
17	BKG000016	Go Sedan	Electronic City	Vidyaranyapura	14.15	5.02	970.11	UPI		
18	BKG000017	Uber Go	Vijayanagar	Bannerghatta Road	4.99	10.16	681.09	UPI		
19	BKG000018	UberXL	Mahadevapura	Electronic City	14.76	9.09	679.3	Card		
20	BKG000019	Non-AC Taxi	Malleswaram	Hennur	13.39	2.62	0			
21	BKG000020	Go Sedan	MG Road	Hosur Road	14.81	2.66	803.32	Card		
22	BKG000021	Uber Go	Basavanagudi	Koramangala	13.57	8.8	1393.23	UPI		
23	BKG000022	Premier	Marathahalli	Hennur	4.36	8.78	1811.08	Card		
24	BKG000023	Go Sedan	Mahadevapura	Koramangala	9.67	9.12	206.04	Cash		
25	BKG000024	UberXL	Devanahalli	HSR Layout	6.86	10.28	1171.54	UPI		

unsuccessful_rides.csv

	A	B	C	D
1	Booking ID	Cancelled Rides by Customer Reason	Cancelled Rides by Driver Reason	Incomplete Ride Reason
2	BKG000002			Vehicle Breakdown
3	BKG000010	Change of plans		
4	BKG000011			Vehicle Breakdown
5	BKG000013		More than permitted people in there	
6	BKG000014		Customer related issue	
7	BKG000015		Customer related issue	
8	BKG000019		Customer related issue	
9	BKG000020	AC is not working		
10	BKG000021			Other Issue
11	BKG000022		Customer related issue	
12	BKG000025		The customer was coughing/sick	
13	BKG000027	Driver asked to cancel		
14	BKG000029			Other Issue
15	BKG000031		Personal & Car related issues	
16	BKG000035		More than permitted people in there	
17	BKG000038		More than permitted people in there	
18	BKG000039	AC is not working		
19	BKG000043		More than permitted people in there	
20	BKG000044		More than permitted people in there	
21	BKG000048			Other Issue
22	BKG000050			Customer Demand

KEY OBSERVATIONS

- *Each dataset contains 200,000 records.*
- *The datasets include booking details, ride metadata, and reasons for cancellations or incomplete rides.*
- *There are 7 unique vehicle types and 51 unique pickup and drop locations.*

COLUMNS OVERVIEW

- **booking_details**
  - **Date**: The date of the ride.
  - **Time**: The time of the ride.
  - **Booking ID**: Unique identifier for each booking.
  - **Booking Status**: Status of the ride (e.g., Successful, Cancelled, Incomplete).

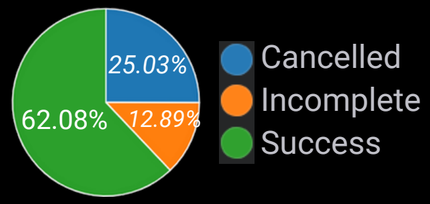
- **ride_details**
  - **Booking ID**: Unique identifier for each booking.
  - **Vehicle Type**: Type of vehicle used for the ride.
  - **Pickup Location**: Location where the ride started.
  - **Drop Location**: Location where the ride ended.
  - **Avg VTAT (Vehicle Turnaround Time)**: Average time taken by the vehicle to start the ride.
  - **Avg CTAT (Customer Turnaround Time)**: Average time taken by the customer to board the vehicle.
  - **Price**: The monetary value of the booking.
  - **Payment Method**: Payment method used (e.g., UPI, Wallet, Cash).
  - **Ride Distance**: Distance covered during the ride.
  - **Customer Ratings**: Ratings given by the customer.

- **unsuccessful_rides**
  - **Booking ID**: Unique identifier for each booking.
  - **Cancelled Rides by Customer Reason**: Reason for cancellation by the customer.
  - **Cancelled Rides by Driver Reason**: Reason for cancellation by the driver.
  - **Incomplete Ride Reason**: Reason for an incomplete ride.

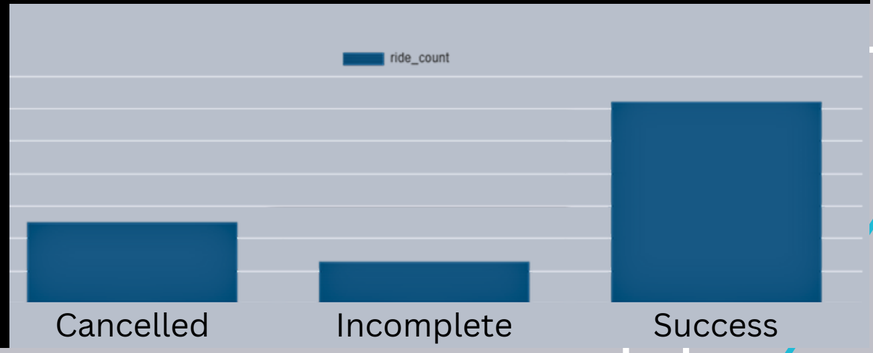
DATA ANALYSIS

Understanding the Distribution of Booking Status

```
SELECT booking_status,
   COUNT(booking_status) AS ride_count,
   ROUND((COUNT(booking_status) * 1.0
   /(SELECT COUNT(*) FROM booking_details)) * 100, 2) AS rides_percentage
FROM booking_details
WHERE booking_status IN ('Success', 'Incomplete')
GROUP BY booking_status
UNION
SELECT 'Cancelled' AS booking_status,
   COUNT(booking_status) AS ride_count,
   ROUND((COUNT(booking_status) * 1.0
   /(SELECT COUNT(*) FROM booking_details)) * 100, 2) AS rides_percentage
FROM booking_details
WHERE booking_status LIKE ('Cancelled%')
```



booking_status character varying	ride_count bigint	rides_percentage numeric
Cancelled	50064	25.03
Incomplete	25772	12.89
Success	124164	62.08



Cancellation Reason Breakdown

(a) Cancelled Ride Types

```
SELECT booking_status,
 COUNT(booking_status) AS ride_count
FROM booking_details
WHERE booking_status NOT IN ('Success', 'Incomplete')
GROUP BY booking_status
```

(b) Rides Cancelled by Customer

```
SELECT booking_status,
 COUNT(booking_status) AS ride_count
FROM booking_details
WHERE booking_status NOT IN ('Success', 'Incomplete')
GROUP BY booking_status
```

(c) Rides Cancelled by Driver

```
SELECT cancelled_rides_by_driver_reason,
   COUNT(*) AS driver_cancel_count
FROM unsuccessful_rides
WHERE cancelled_rides_by_driver_reason IS NOT NULL
GROUP BY cancelled_rides_by_driver_reason
ORDER BY driver_cancel_count DESC
```

Cancellation Reason Breakdown

booking_status character varying (50)	ride_count bigint
Cancelled by Customer	14191
Cancelled by Driver	35873

cancelled_rides_by_customer_reason character varying (100)	customer_cancel_count bigint
AC is not working	2927
Driver asked to cancel	2863
Change of plans	2838
Driver is not moving towards pickup location	2795
Wrong Address	2768

cancelled_rides_by_driver_reason character varying (100)	driver_cancel_count bigint
More than permitted people in there	9029
The customer was coughing/sick	9013
Customer related issue	8989
Personal & Car related issues	8842

Incomplete Ride Reason Breakdown

```
SELECT incomplete_ride_reason,
   COUNT(*) AS incomplete_ride_count
FROM unsuccessful_rides
WHERE incomplete_ride_reason IS NOT NULL
GROUP BY incomplete_ride_reason
ORDER BY incomplete_ride_count DESC
```

incomplete_ride_reason character varying (100)	incomplete_ride_count bigint
Customer Demand	8620
Other Issue	8601
Vehicle Breakdown	8551

Vehicle-Type Breakdown

```
SELECT vehicle_type, COUNT(*) as ride_count
FROM ride_details
GROUP BY vehicle_type
```

vehicle_type character varying (20)	ride_count bigint
eBike	28558
Go Sedan	28717
Moto	28478
Non-AC Taxi	28573
Premier	28685
Uber Go	28163
UberXL	28826

Avg VTAT and Avg CTAT for Each Vehicle

```
SELECT vehicle_type, ROUND(AVG(avg_vtat), 2) as
avg_vtat, ROUND(AVG(avg_ctat), 2) as avg_ctat
FROM ride_details
GROUP BY vehicle_type
```

vehicle_type character varying (20)	avg_vtat numeric	avg_ctat numeric
eBike	9.02	6.99
Go Sedan	9.01	7.01
Moto	8.96	6.99
Non-AC Taxi	8.98	6.99
Premier	9.00	7.00
Uber Go	9.02	7.00
UberXL	9.01	7.00

Top Pickup Locations

```
SELECT pickup_location, COUNT(*) as ride_count
FROM ride_details
GROUP BY pickup_location
ORDER BY ride_count desc
LIMIT 5
```

pickup_location character varying (50)	ride_count bigint
Hebbal	4062
Hormavu	4054
Bellandur	4048
Vidyaranyapura	4043
Kudlu Gate	4040

Top Drop Locations

```
SELECT drop_location, COUNT(*) as ride_count
FROM ride_details
GROUP BY drop_location
ORDER BY ride_count desc
LIMIT 5
```

drop_location character varying (50)	ride_count bigint
Richmond Town	4056
Kammanahalli	4050
Magadi Road	4001
Bannerghatta Road	3995
Hormavu	3994

Average Booking Value by Vehicle Type

```
SELECT vehicle_type, round(avg(price), 2) as
avg_booking_value
FROM ride_details
GROUP BY vehicle_type
ORDER BY avg_booking_value desc
```

vehicle_type character varying (20)	avg_booking_value numeric
Uber Go	639.79
Moto	637.88
eBike	636.67
Premier	636.31
Non-AC Taxi	633.60
Go Sedan	630.49
UberXL	629.34

Payment Method Popularity

```
SELECT r.payment_method, COUNT(*) AS ride_count
FROM ride_details r
JOIN booking_details b USING (booking_id)
WHERE b.booking_status = 'Success'
GROUP BY r.payment_method
ORDER BY ride_count DESC
```

payment_method character varying (20)	ride_count bigint
UPI	31352
Wallet	31056
Card	31000
Cash	30756

Customer Rating Analysis

(a) Vehicles with Most Number of High Ratings (4 or more)

```
SELECT r.vehicle_type,COUNT(*) as
high_rated_ride_count FROM ride_details r
JOIN booking_details b USING(booking_id)
WHERE b.booking_status = 'Success'
   AND r.customer_ratings >= 4
GROUP BY r.vehicle_type
ORDER BY high_rated_ride_count desc
```

vehicle_type character varying (20)	high_rated_ride_count bigint
Moto	8827
eBike	8682
Uber Go	8490
Premier	4701
UberXL	4673
Non-AC Taxi	4646
Go Sedan	4639

(b) Vehicles with Most Number of Below Average Ratings

```
SELECT r.vehicle_type,COUNT(*) as
below_avg_rated_ride_count FROM ride_details r
JOIN booking_details b USING(booking_id)
WHERE b.booking_status = 'Success'
  AND r.customer_ratings < (SELECT
AVG(customer_ratings) FROM ride_details)
GROUP BY r.vehicle_type
ORDER BY below_avg_rated_ride_count desc
```

vehicle_type character varying (20)	below_avg_rated_ride_count bigint
Go Sedan	10479
Premier	10442
UberXL	10410
Non-AC Taxi	10363
eBike	4213
Uber Go	4171
Moto	4142

Customer Rating Analysis

(c) Average Customer Ratings by Vehicle Type

```
SELECT r.vehicle_type, ROUND(AVG(r.customer_ratings),1) AS avg_customer_ratings
FROM ride_details r
JOIN booking_details b USING(booking_id)
WHERE b.booking_status = 'Success'
GROUP BY r.vehicle_type
ORDER BY avg_customer_ratings desc
```

vehicle_type character varying (20)	avg_customer_ratings numeric
Moto	3.8
Uber Go	3.8
eBike	3.8
UberXL	3.0
Non-AC Taxi	3.0
Go Sedan	3.0
Premier	3.0

Revenue Analysis

Revenue from Top Pickup Locations

```
SELECT pickup_location,
   SUM(price) as total_revenue
FROM ride_details
GROUP BY pickup_location
ORDER BY total_revenue DESC
LIMIT 5
```

pickup_location character varying (50)	total_revenue numeric
Hebbal	2646743.20
Bommanahalli	2602199.23
Basaveshwaranagar	2586896.55
Hormavu	2586671.96
Mahadevapura	2584782.34

Average Revenue Gained from Each Vehicle-Type

```
SELECT vehicle_type,
   ROUND(AVG(price),2) as avg_revenue
FROM ride_details
GROUP BY vehicle_type
ORDER BY avg_revenue DESC
```

vehicle_type character varying (20)	avg_revenue numeric
Uber Go	639.79
Moto	637.88
eBike	636.67
Premier	636.31
Non-AC Taxi	633.60
Go Sedan	630.49
UberXL	629.34

Revenue Per Day

```
SELECT ROUND(SUM(r.price)/COUNT(DISTINCT b.date),2)
   AS revenue_per_day
FROM ride_details r
JOIN booking_details b USING (booking_id)
```

revenue_per_day numeric
4095793.73

Successful Rides Per Day

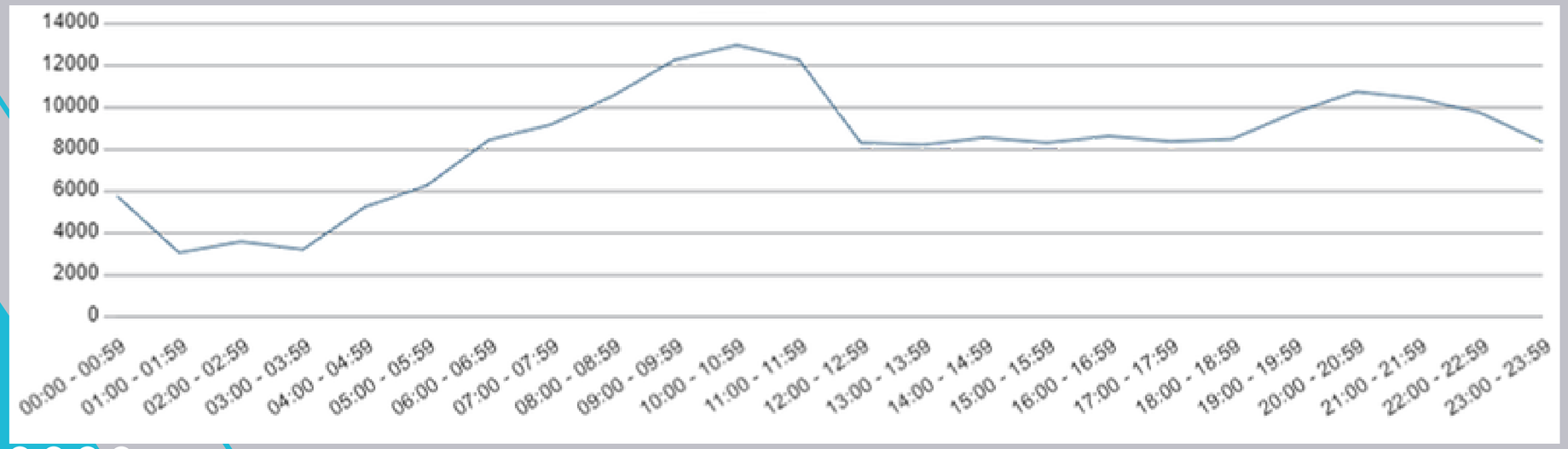
```
SELECT ROUND(COUNT(*)* 1.0 / COUNT(DISTINCT dates))
   AS successful_rides_per_day
FROM ride_details
WHERE booking_status = 'successful';
```

successful_rides_per_day numeric
4005.3

Hourly Ride Analysis

```
SELECT
 CONCAT(
  LPAD(EXTRACT(HOUR FROM time)::TEXT, 2,'0'),
  ':00 - ',  LPAD(EXTRACT(HOUR FROM time)::
  TEXT, 2, '0'), ':59') AS time_range,
  COUNT(*) as ride_count
FROM booking_details
GROUP BY EXTRACT(HOUR FROM time)
ORDER BY ride_count DESC
```

time_range text	ride_count bigint
00:00 - 00:59	5733
01:00 - 01:59	3030
02:00 - 02:59	3554
03:00 - 03:59	3181
04:00 - 04:59	5214
05:00 - 05:59	6242
06:00 - 06:59	8411
07:00 - 07:59	9148
08:00 - 08:59	10521
09:00 - 09:59	12234
10:00 - 10:59	12937
11:00 - 11:59	12255
12:00 - 12:59	8287
13:00 - 13:59	8176
14:00 - 14:59	8531
15:00 - 15:59	8282
16:00 - 16:59	8597
17:00 - 17:59	8345
18:00 - 18:59	8464
19:00 - 19:59	9728
20:00 - 20:59	10722
21:00 - 21:59	10390
22:00 - 22:59	9713
23:00 - 23:59	8305



INSIGHTS

- The most common cancellation reasons are related to passenger count, sickness, and general customer issues.
- Customer demand, miscellaneous issues, and vehicle breakdowns are major contributors to incomplete rides.
- All vehicle types have nearly equal ride distribution, with UberXL having the highest ride count.
- VTAT and CTAT times are very similar across all vehicle types.
- Uber Go has the highest average booking value, while UberXL has the lowest.
- Moto and eBike receive the most high-rated rides, while Go Sedan and Premier receive the least.
- Go Sedan, Premier, and UberXL have the highest below-average ratings.
- Hebbal, and Bommanahalli generate the highest revenue (over ₹2.6M).
- UPI is the most used payment method, followed closely by Wallet, Card, and Cash.
- Ride counts are consistently high from 8:00-11:59 AM, likely due to work/school commutes. Ride demand spikes again around 7:00-10:59 PM, likely due to people returning home from work.
- Late-night rides decline steadily.

RECOMMENDATIONS

- Reduce cancellations by offering incentives for drivers and ensuring better customer service to minimize incomplete rides.
- Improve vehicle maintenance programs and enhance customer service training to address demand-related issues.
- Consider promotional discounts or targeted marketing to push underperforming vehicle types.
- Encourage more eBike and Moto rides for short-distance travel by offering incentives.
- As Hebbal, Hormavu, and Bellandur have the highest ride counts, deploy more drivers in these areas during peak hours to reduce wait times.
- As Richmond Town and Kammanahalli are the most common drop locations, implement dynamic pricing and optimize routing for these locations to improve efficiency.
- Promote UberXL by offering discounts or bundling services for larger groups.
- Improve customer experience for Go Sedan and Premier by training drivers on better customer service.
- Investigate service complaints for these vehicle types and improve service quality for the lower-rated vehicle to balance customer satisfaction.
- Focus on marketing efforts and driver availability in high revenue generating pickup locations to maximize profitability.
- Increase driver availability in the morning and evening to match peak demand and reduce wait times.
- Introduce dynamic pricing during rush hours to optimize revenue and manage high demand.

THANK YOU!