NammaYatri Trip, Revenue & User Behavior Analysis

1. Project Overview

This capstone project analyzes NammaYatri's trip, revenue, and user behavior data using Python-based EDA, SQL queries, and Power BI dashboards. The goal is to uncover actionable insights that improve operational efficiency, user experience, and revenue optimization.

2. Dataset Summary

NammaYatri creates total 5 datasets, these are Duration, Assembly, Payment, Trip_Details, and Trips.

(i)Duration

- Rows: 24
- Columns: 2
- Key Features: ID, Duration
- Missing Data: 0

(ii)Assembly

- Rows: 37
- Columns: 2
- Key Features: ID, Assembly
- Missing Data: 0

(iii)Payment

- Rows: 4
- Columns: 2
- Key Features: ID, Method
- Missing Data: 0

(iv)Trip_Details

- Rows: 2161
- Columns: 10

- Key Features: trip_id, loc_from, searches, searches_got_estimate, searches_for_quotes, searches_got_quotes, customer_not_cancelled, driver_not_cancelled, otp_entered, end_ride
- Missing Data: 0

(v)Trips

- Rows: 983
- Columns: 9
- Key Features: trip_id, faremethod, fare, loc_from, loc_to, driverid, custid, distance, duration
- Missing Data: 0

3. Exploratory Data Analysis using Python

To identify patterns and inefficiencies in trip data using Python libraries (Pandas, Seaborn, Matplotlib).

1. **KEY FINDINGS**

- **Conversion Funnel Leakage**: 14% drop-off between fare estimate and quote request.
- **Weak Fare-Distance Correlation**: Pearson correlation of 0.01 suggests fare is influenced by other factors.
- **Geographic Hotspots**: Assemblies like Champmartin and Vijay Nagar show longer average trip distances.
- **Payment Method Nuances**: Credit card and cash trips span wider fare and distance ranges.
- **Trip Completion & OTP Entry**: Suboptimal OTP entry rates impact ride success and data reliability.

2. **STRATEGIC RECOMMENDATIONS**

- Streamline the estimate-to-quote UX to reduce drop-offs.
- Enhance pricing models using multi-factor inputs.
- Tailor promotions based on payment behavior.
- Investigate causes of failed OTP entries and incomplete trips.

4. Data Analysis using SQL

Schema: nammayatri_new with tables trips, trip_details, payment, assembly, duration

CATEGORY 1: REVENUE & PRICING INSIGHTS

1. **Total Revenue by Payment Method:** Calculated the total fare collected for each payment method

	method	total_revenue
▶	upi	183624
	credit card	197941
	debit card	184676
	cash	185102

2. **Average Fare per Area:** Found the Assembly area (loc_from) with highest average fare per trip

Result Grid | Filter Rows:

	assembly_name	avg_fare
▶	Byatarayanapura	919.1053

3. **High-Value Trips:** Found the total number of trips where the fare was greater than 500 and the payment method was 'Credit Card'.

	high_value_trips
▶	178

4. **Fare per Distance Efficiency:** Calculated the average fare-per-distance (fare / distance) for each payment method.

	method	avg_fare_per_km
▶	upi	85.83840494
	credit card	82.71785687
	debit card	84.18514568
	cash	87.51262511

5. **Earnings by Duration:** Found the duration range (from the Duration table, e.g., "0-1", "1-2") generates the highest total revenue

	duration	total_revenue
▶	0-1	45019

CATEGORY 2: DRIVER & CUSTOMER PERFORMANCE

6. **Top 5 Earners:** Identified the top 5 drivers (driverid) based on total earnings.

	driverid	total_earnings
▶	12	36787
	8	30101
	21	29787
	24	28870
	30	28853

7. **Frequent Travelers:** Found the top 5 customers (custid) who have taken the most number of rides.

	custid	ride_count
▶	22	20
	92	18
	29	17
	43	17
	47	16

8. **Driver Utilization:** Listed the drivers who have completed more than 50 trips.

	driverid	trip_count

9. **Single Trip High Spenders:** Identified the customers who have taken exactly 1 trip, but that trip cost more than 750.

	custid	fare

10. **Driver Average Distance:** Calculated the average distance traveled per trip for each driver, sorted in descending order.

	driverid	avg_distance
▶	30	16.2703
	28	15.4571
	14	15.4483
	10	15.3548
	26	15.1667
	1	15.1282
	24	15.0952
	18	15.0690
	22	15.0303
	29	14.9118
	17	14.8667
	15	14.8125
	27	14.7500
	13	14.6757
	7	14.6400
	21	14.5750
	16	14.5429
	19	14.4667
	4	14.3056
	12	14.2174

Result 10

CATEGORY 3: OPERATIONAL FUNNEL METRICS

11. **Search to Estimate Rate:** For each assembly area, calculated the percentage of searches that resulted in an estimate (searches_got_estimate / searches).

	assembly_area	estimate_rate_pct
▶	Gandhi Nagar	56.34
	Vijay Nagar	63.64
	Yeshwantpur	57.97
	Chamrajpet	64.15
	Basavanagudi	57.63
	Kanakapura	59.02
	Mahalakshmi Layout	54.24
	Govindraj Nagar	67.80
	Devanahalli	58.49
	Yelahanka	68.25
	Rajaji Nagar	46.67
	Bommanahalli	63.79
	Anekal	55.00
	Hoskote	59.57
	Chickpet	55.74
	Dasarahalli	59.09
	Ramanagaram	62.16
	Rajarajeshwarinagar	54.69
	C. V. Raman Nagar	53.13
	Bangalore South	75.44

12. **Driver Cancellation Hotspots:** Found the assembly area with the highest number of driver cancellations (where driver_not_cancelled = 0)

	assembly_area	driver_cancellations
▶	Mahadevapura	43

13. **Customer Cancellation Hotspots:** Found the assembly area with the highest number of customer cancellations (where customer_not_cancelled = 0)

	assembly_area	customer_cancellations
▶	C. V. Raman Nagar	40

14. **OTP Drop-offs:** Found the total number of trips where an OTP was entered (otp_entered = 1) but the ride was not completed (end_ride = 0).

	otp_dropoffs
▶	0

15. **End-to-End Conversion:** Calculated the overall conversion rate from 'search' to 'end_ride' (Total Completed Rides / Total Searches) for the entire dataset.

	conversion_rate_pct
▶	45.49

CATEGORY 4: LOCATION & ROUTE ANALYSIS

16. **Most Popular Pickup Locations:** Listed the top 3 Assembly names that generated the most trip requests

	assembly_area	total_searches
▶	Mahadevapura	75
	Ramanagaram	74
	Gandhi Nagar	71

17. **Top Routes:** Identified the most frequent route (combination of loc_from and loc_to).

	loc_from	loc_to	route_count
▶	16	21	5

18. **Long Distance Connections:** Found the pair of Assembly areas (loc_from and loc_to) with the longest average distance.

	loc_from	loc_to	avg_distance
▶	28	9	27.0000

19. **Demand vs. Supply (Quotes):** Found the Assembly area with the lowest "Quote Acceptance" rate (ratio of searches_got_quotes to searches_for_quotes)

	assembly_area	quote_acceptance_rate
▶	Rajaji Nagar	75.00

20. **Payment Preference by Location:** For a specific area (e.g., 'Hebbal'), found is the most commonly used payment method

	method	usage_count
▶	credit card	13

5. Power BI Dashboard

To visualize trip performance, revenue, and customer behavior for strategic decision-making.

1. **DASHBOARD HIGHLIGHTS**

- **Summary Metrics**:
    - Total Searches: 2,161
    - Completed Trips: 983
    - Completion Rate: 45.49%
    - Total Revenue: ₹751K
    - Average Fare: ₹764
    - Fare per KM: ₹53.11
- **Trip Booking Funnel**:
    - Key drop-offs from estimate to quote and OTP entry.
- **Revenue by Payment Method**:
    - Cash leads slightly; all methods show balanced contribution.
- **Cancellation Metrics**:
    - Customer: 48.17%
    - Driver: 47.25%
- **Hourly Trip Distribution**:
    - Peak hours: 33–53 trips/hour
- **Top 10 Demand Locations**:
    - Mahadevapura, Ramanagaram, Gandhi Nagar top the list.

2. **STRATEGIC VALUE**

- Real-time KPI monitoring
- Funnel optimization
- Revenue targeting
- Location-based forecasting

6. Business Recommendations

- Improve the user journey by optimizing the fare estimate to quote request flow to reduce drop-offs and increase conversion.
- Refine pricing strategies by incorporating multiple factors beyond distance to better reflect trip costs and customer willingness to pay.
- Deploy drivers strategically in geographic hotspots identified through trip data to meet demand efficiently and reduce wait times.
- Customize promotions and payment options based on user payment behavior to enhance customer retention and satisfaction.
- Address operational issues such as OTP entry failures and cancellation hotspots through targeted interventions and UX improvements.
- Leverage Power BI dashboards for continuous monitoring of KPIs and to enable data-driven decision-making across departments.

This integrated analysis using Python, SQL, and Power BI reveals key operational and behavioral insights. By addressing conversion bottlenecks, pricing inefficiencies, and location-based demand, NammaYatri can enhance user satisfaction, optimize revenue, and strengthen its market position.