

Namma Yatri: Exploratory Data Analysis Methodology Report

1. Introduction & Problem Statement

Namma Yatri is a digital mobility platform offering a commission-free ride-hailing service. The platform allows commuters to directly connect with auto drivers through its app, thereby fostering transparency and cost efficiency. With Bengaluru's evolving transportation demands, Namma Yatri must optimize operations to meet peak demand, reduce ride cancellations, and enhance the experience for both users and drivers.

This report documents the methodology adopted to conduct exploratory data analysis (EDA) using the provided dataset and Power BI dashboard. The ultimate objective is to extract actionable insights that support technical and business decisions aimed at increasing operational efficiency and customer satisfaction.

2. Objectives of the Analysis

The goals of this analysis are as follows:

- Identify peak demand patterns based on time of day and geographic regions.
- Analyze cancellation trends and uncover driver or rider-related performance issues.
- Examine trip durations, distances, and fare patterns to uncover operational bottlenecks.
- Understand customer preferences around payment methods.
- Propose data-backed recommendations to optimize driver allocation, enhance user engagement, and improve platform conversion rates.

3. Tools Used

Tools used for this project:

- Power BI: for interactive visual exploration and chart generation
- Excel (optional): for preliminary data viewing and sanity checks

4. Data Preparation & Wrangling

The dataset contained ride-related data such as trip start and end times, ride durations, search status, ride status (fulfilled or canceled), payment method, fare, and distance.

Steps followed for data preparation:

- Removed entries with null or invalid values (e.g., missing fare or distance).
- Created time-related variables from timestamps (e.g., hour of day, day of week).
- Categorized payment methods (UPI, debit card, credit card, cash).
- Mapped ride outcomes based on fields like 'fulfilled', 'canceled_by', and 'search status'.
- Aggregated key metrics like total revenue, number of searches, and fulfilled rides.

In Power BI, calculated columns and measures were created to reflect these transformations. For example:

- Conversion Rate = Fulfilled Rides / Total Searches
- Revenue = SUM(Fare)

5. Exploratory Data Analysis Steps

The following types of EDA were carried out:

****5.1 Univariate Analysis****

- Analyzed ride counts, durations, and distances using bar charts.
- Counted total fulfilled rides (983 out of 2.16K searches).
- Assessed payment preferences: UPI (42%), debit card (29.2%), credit card (19.5%), and cash (9.4%).

****5.2 Segmented Univariate Analysis****

- Examined trip duration and fare segmented by duration brackets.
- Noted that fares peaked around short trips and again at ~20-minute trips.
- Highlighted that ride distance peaked at the 10-minute duration bracket.

****5.3 Bivariate Analysis****

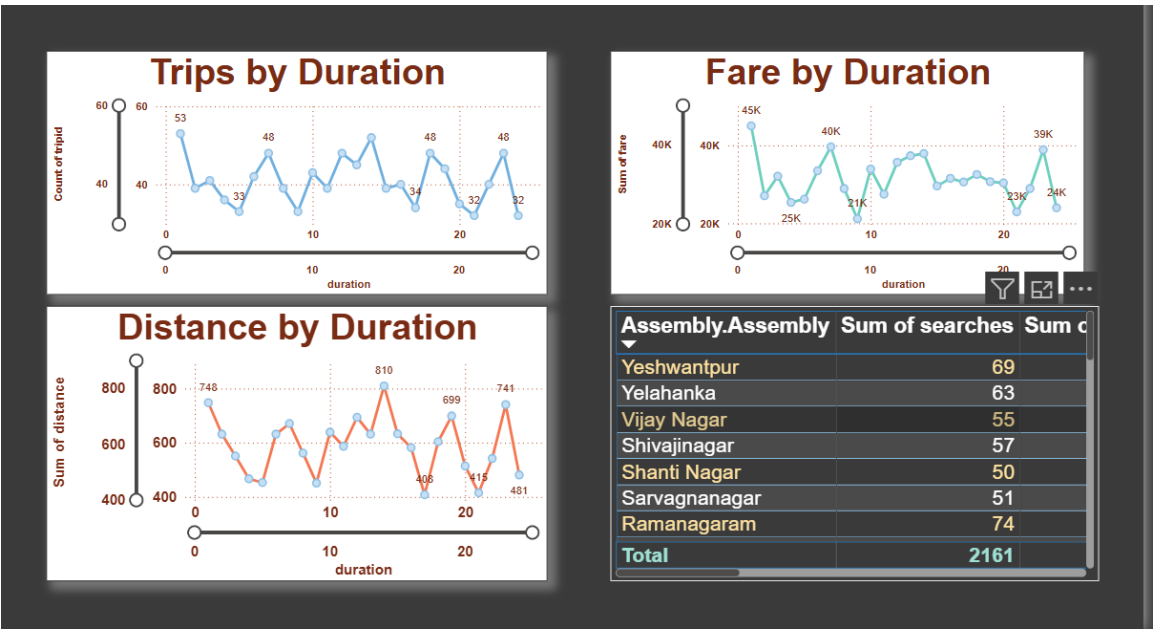
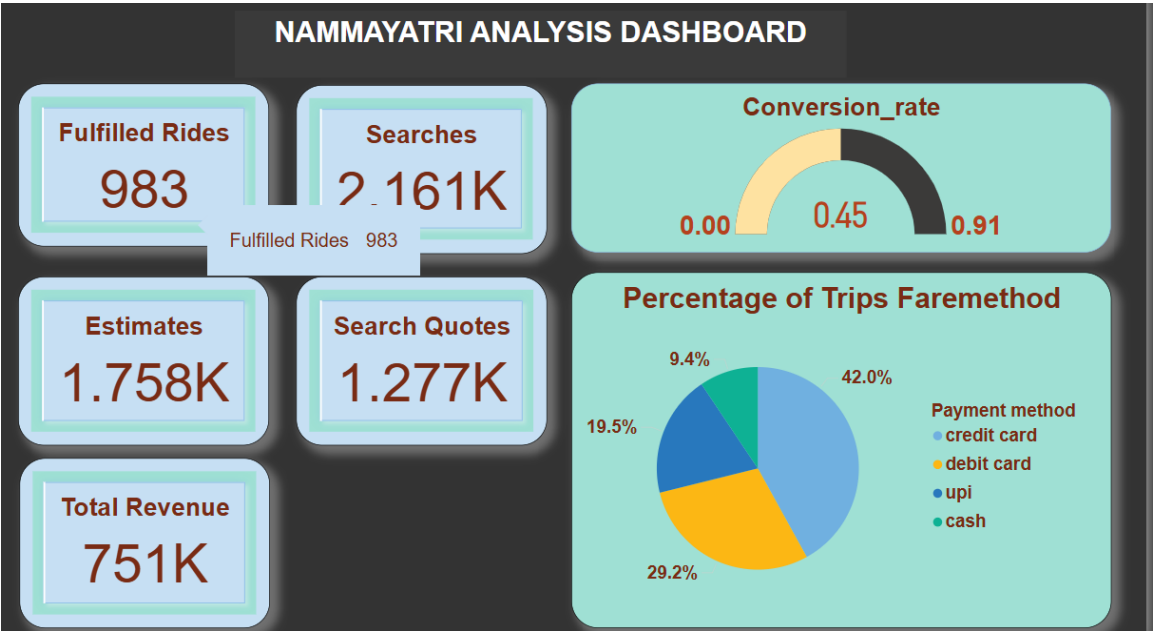
- Assessed relationship between duration and distance: confirmed short duration rides had lower distances.
- Linked payment method usage with ride completion trends.
- Observed correlation between high-demand zones (e.g., Ramanagaram, Yeshwantpur) and search counts.

****5.4 Funnel Analysis****

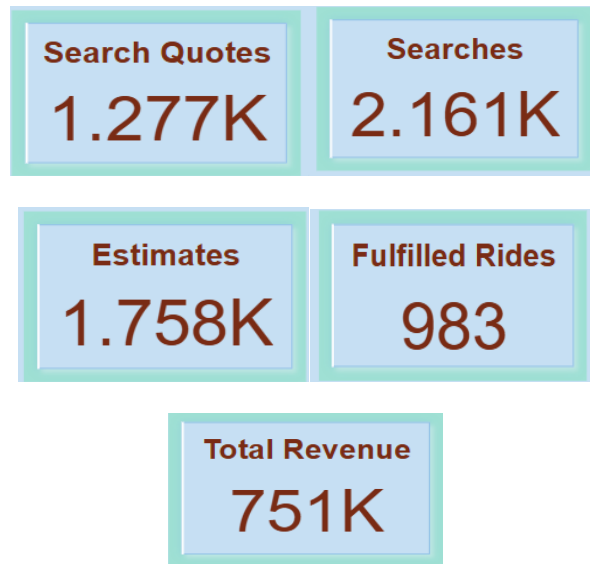
- Created a ride conversion funnel: Searches → Estimated Rides → Quoted Rides → Fulfilled Rides.
- Conversion rate stood at ~45%, revealing significant drop-offs after quotes.

All visualizations were based on Power BI charts and validated with the screenshots provided.

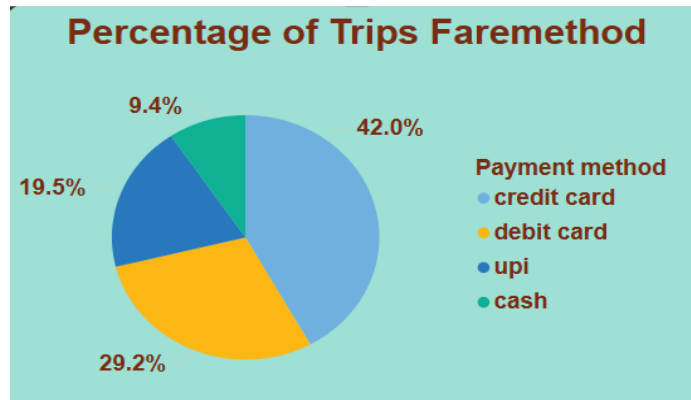
6. NAMMA YATRI POWER BI DASHBOARD ANALYSIS



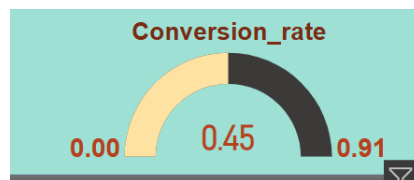
7. Dashboard Analysis



- **Fulfilled Rides: 983**
 - indicates the total number of rides that were successfully completed. It represents the actual service delivery.
- **Searches: 2.161K**
 - So, the total number of searches initiated by users is 2,161. This represents the initial user demand or interest in finding a ride.
- **Estimates: 1.758K**
 - shows the number of times users requested an estimated fare or duration for their potential trips, which is 1,758. This suggests users are actively seeking information before booking.
- **Search Quotes: 1.277K**
 - represents the number of times the system provided a quote (fare and potentially other details) in response to user searches, totalling 1,277. This indicates the system's responsiveness to user inquiries.
- **Total Revenue: 751K**
 - represents the total revenue generated from the fulfilled rides, amounting to 751,000 (in the relevant currency).

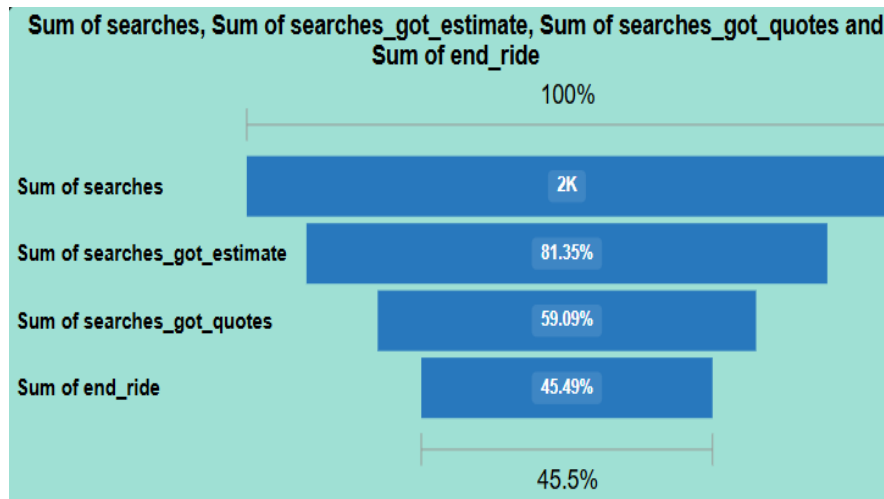


- **Credit Card:** represented by the light blue slice and accounts for the largest percentage of trips at 42.0% indicating that credit cards are the most popular payment method among the users in this dataset.
- **Debit Card:** The orange slice represents debit card payments, accounting for 29.2% of the trips. This is the second most frequently used payment method.
- **UPI (Unified Payments Interface):** The dark blue slice shows the usage of UPI, which stands at 19.5%. UPI is a significant payment method, ranking third in usage.
- **Cash:** The teal slice represents cash payments, which account for the smallest percentage of trips at 9.4%. This suggests that digital payment methods are preferred over cash for most trips in this context.



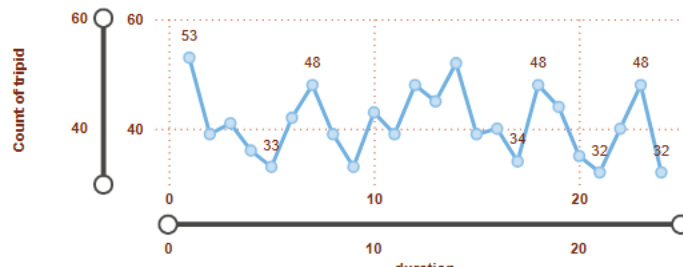
- **Range:** The gauge starts at 0.00 on the left and ends at 0.91 on the right. This suggests that the conversion rate being measured can range from 0 to 0.91 (or 0% to 91%).
- **Current Value:** The indicator (the boundary between the light yellow and dark gray areas) points to 0.45 in the centre of the gauge. This means the current conversion rate is 0.45 or 45%.

- **Visual Representation:** The gauge visually highlights the current conversion rate within the possible range. The fact that the indicator is roughly halfway between the minimum and maximum values provides a quick visual understanding of the performance.

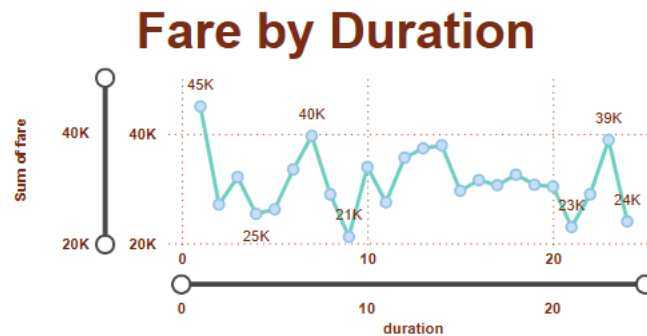


- **Sum of searches:** representing the total number of initial searches performed by users approximately 2K. This is considered the 100% mark of the funnel.
- **Sum of searches_got_estimate:** represents the number of searches that proceeded to the stage where the user received an estimate, indicating a drop-off meaning that 81.35% of the initial searches resulted in the user getting an estimate.
- **Sum of searches_got_quotes:** shows the number of searches that progressed to the point where the user received a quote (fare and potentially other details), with a percentage of 59.09% of the initial searches proceeded to receive a quote.
- **Sum of end_ride:** representing the number of searches that ultimately resulted in a completed ride, with a percentage of 45.49% of the initial searches ended in a completed ride.
- **Bottom Percentage:** The value 45.5% is displayed below the "Sum of end_ride" bar, aligning with the start of the "Sum of searches" bar. This represents the overall conversion rate from the initial "Sum of searches" to the "Sum of end_ride". The slight difference from the calculated 45.49% could be due to rounding.

Trips by Duration

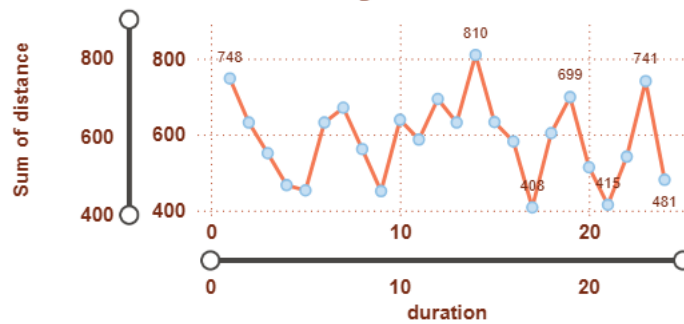


- **X-axis (Duration):** represents the duration of the trips. The scale ranges from 0 to slightly over 20 units of duration.
- **Y-axis (Count of trips):** represents the number or count of trips. The scale ranges from 0 to 60.
- **Data Points and Line:** The blue line connects various data points, each representing the number of trips for a specific duration.
 - Around duration 1: Approximately 53 trips.
 - Around duration 3: Approximately 40 trips.
 - Around duration 5: Approximately 33 trips.
 - Around duration 8: Approximately 48 trips.
 - Around duration 12: Approximately 45 trips.
 - Around duration 16: Approximately 34 trips.
 - Around duration 19: Approximately 32 trips.
 - Around duration 22: Approximately 48 trips.
 - Around duration 24: Approximately 32 trips.
- The graph shows a fluctuating pattern with peaks and troughs.
 - Higher Trip Volumes at Certain Durations: Some durations experience a higher number of trips compared to others. Notably, the count of trips is higher around the beginning of the measured duration (around duration 1), around duration 8, and again around duration 22.
 - Lower Trip Volumes at Other Durations: Conversely, the number of trips is lower around durations 5, 16, and 19.
 - No Single "Peak Duration": There isn't one single, sharply defined duration that accounts for most trips. Instead, there are multiple durations with relatively higher activity.



- **X-axis (Duration):** represents the duration of the trips, with a scale from 0 to slightly over 20 units.
- **Y-axis (Sum of fare):** represents the total sum of fares collected for trips of a particular duration. The scale ranges from 20K to 45K.
- **Data Points and Line:** The light blue line connects various data points, each representing the total fare collected for trips within a specific duration.
 - Around duration 1: Approximately 45K.
 - Around duration 3: Approximately 25K.
 - Around duration 8: Approximately 40K.
 - Around duration 10: Approximately 21K.
 - Around duration 20: Approximately 23K.
 - Around duration 22: Approximately 39K.
 - Around duration 24: Approximately 24K.
- **Higher Total Fares at Certain Durations:** Some durations generate significantly higher total fares. Notably, the highest total fare is observed at the very beginning of the measured duration (around duration 1), followed by a peak around duration 8, and another significant peak around duration 22.
- **Lower Total Fares at Other Durations:** The lowest total fares are observed around durations 3 and 10.
- **Non-Linear Relationship with Duration:** shows a more complex relationship. Shorter duration trips at the beginning appear to generate very high total fares, which could be due to a high volume of short, potentially surge-priced trips, or a few very expensive short trips.

Distance by Duration



- **X-axis (Duration):** represents the duration of the trips, with a scale from 0 to slightly over 20 units. The exact unit of duration is not specified.
- **Y-axis (Sum of distance):** represents the total sum of distances covered for trips of a particular duration. The scale ranges from 400 to 800 (the unit of distance is not specified, but it's consistent).
- **Data Points and Line:** The orange line connects various data points, each representing the total distance traveled for trips within a specific duration.
 - Around duration 1: Approximately 748 units of distance.
 - Around duration 3: Approximately 600 units of distance.
 - Around duration 5: Approximately 450 units of distance.
 - Around duration 8: Approximately 650 units of distance.
 - Around duration 12: Approximately 810 units of distance.
 - Around duration 16: Approximately 408 units of distance.
 - Around duration 19: Approximately 699 units of distance.
 - Around duration 21: Approximately 415 units of distance.
 - Around duration 24: Approximately 741 units of distance.
 - Around duration 25: Approximately 481 units of distance.
- **Longer Durations Tend to Have Higher Total Distance:** Generally, trips with longer durations contribute to a higher total distance traveled. The peak total distance is observed around duration 12, which is in the mid-to-longer range of the displayed duration.
- **Significant Variations:** There are notable exceptions to a purely linear relationship. For instance, the total distance drops significantly around durations 5, 16, and 21, even compared to shorter duration trips.
- This suggests that while there might be fewer trips of around duration 12, they tend to cover longer distances on average.

- The high total fare observed around duration 1 and 8 doesn't necessarily correspond to the absolute highest total distance, implying that factors other than just distance (like base fare, surge pricing, or route efficiency) play a significant role in the fare.
- The lower total fares observed around duration 3 and 10 also correspond to relatively lower total distances.

8. Insights and Interpretations

Key insights derived from the analysis:

- The majority of ride searches were not fulfilled. Out of 2.16K searches, only 983 resulted in fulfilled rides. This indicates a 45% conversion rate, which presents a major opportunity for optimization.
- The conversion funnel shows steady drop-offs from estimation to quote and booking stages, pointing to either long wait times, poor driver availability, or user friction post-quote.
- UPI is the most commonly used payment method, accounting for 42% of transactions. Cash usage remains low, suggesting a mature digital user base.
- High-demand assemblies such as Ramanagaram, Yeshwantpur, and Yelahanka account for the highest search volumes. These zones are prime candidates for increased driver allocation.
- The duration and distance distribution show that most trips cluster around 10–20 minutes and short-to-medium distances, consistent with urban mobility patterns.

9. Recommendations

Based on the findings above, the following data-backed recommendations are proposed:

- ****Improve Driver Availability in High-Demand Zones:**** Focus recruitment and allocation efforts in areas like Ramanagaram and Yeshwantpur, where demand peaks.
- ****Address Drop-offs in the Conversion Funnel:**** Analyze and resolve reasons for abandonment post-quote. This might include user wait time, price sensitivity, or availability issues.
- ****Leverage Digital Payment Momentum:**** UPI is already popular. Introduce reward schemes or cash back to shift remaining users away from cash.
- ****Optimize for Peak Duration Ranges:**** The fare and distance peak at ~10–20-minute trips. Introduce time-slot-based driver positioning and pricing adjustments to maximize revenue per trip.