Methodology Report:
Visualization & Analysis on Namma Yatri Data

By – Zayed Syed, GSS Vinay and Srushti Keskar

Include your visualisations, analysis, results, insights, and outcomes.

Explain your methodology and approach to the tasks. Add your conclusions to the sections.

Table 1: Data Description

Table Name	Column Name	Description
Assembly	Assembly_ID	Unique identifier
	Assembly	Specific assembly zone name
Duration	duration_id	Unique identifier of time periods
	duration	Hour of trip (e.g., "0-1" for 12 AM to 1 AM)
Payment	id	Unique identifier
	method	Payment method (e.g., Cash, UPI, Credit Card)
Trip Details	tripid	Unique identifier of trips
	loc_from	Source Location code
	searches	Trip request count
	searches_got_estimate	Got an estimated price (1 = user gets an estimate, 0 = does not get an estimate)
	searches_for_quotes	Searched for drivers after estimate (1 - searched, 0 - not searched)
	searches_got_quotes	Got quotes (1 = Driver allotted, 0 = not allotted)
	customer_not_cancelled	Whether customer cancelled or not (1 = Not cancelled)
	driver_not_cancelled	Whether driver cancelled or not (1 = Not cancelled)
	otp_entered	(1 = OTP entered, 0 = not entered)
	end_ride	Whether ride was completed (1 = Completed)
Trips	tripid	Links to Trip Details
	faremethod	Payment method ID, links to Payment table
	fare	Fare amount
	loc_from	Location ID of source
	loc_to	Location ID of destination, links to Assembly table
	driverid	Driver ID
	custid	Customer ID
	distance	Distance in KM from source to destination
	duration	Unique identifier of time periods like duration_id

Points to Note:

1. **Without this methodology document, the other parts of your case study will not be evaluated.**

2. This assignment is different from the ones you have solved before. Make sure that you treat this case study as a storytelling exercise and not an analysis/visualisation one. This will help you be better prepared for the presentations.

3. Once you are done with the analysis and visualisations, there will be many insights at your hand. Make sure that you map the right visuals and takeaways with the right audience since some of these insights might be relevant to one group but not to the other group.

4. **DO NOT** change the text or numbering of any task, as it may cause problems with grading. Write your solutions to a task in the space provided below the respective task.

Tasks to be performed

- Present the overall approach of the analysis.

- Mention the problem statement and the analysis approach briefly.

- To solve a task, you have to create relevant visualisations and derive appropriate insights from the visualisations.

- Add all the plots, insights, calculated field commands, results and outcomes for a task with proper numbering and sequence in the report.

- The scores for all tasks (except conclusions) comprise both analysis work in the visualisation tool and its outcome in the report.

- You will be awarded a score for a task only if the Tableau/PowerBI analysis is correct and is included in the report along with the subsequent insights.

- Finally, draw conclusions based on the analysis.

Scoring:

Report Total Marks: 70
Sections: 3 sections (10 marks + 40 marks + 20 marks)

Analysis and Visualisation

1. Data Preparation   [10 Marks]

1.1. Import and Join Tables Correctly [5 Mark]

- Import the Namma Yatri dataset into Tableau/Power BI.
- Ensure that you correctly join all tables to create a unified dataset for analysis.
- Verify the relationships between different tables and confirm that data from various sources is properly aligned for accurate insights.

Solution:

I have successfully imported the Namma Yatri dataset into Power BI and established a unified data model by creating the following relationships between tables:

- ***Trips (tripid)** is linked to **Trip_Details (tripid)** as a one-to-one relationship.*
- ***Trips (faremethod)** connects to **Payment (id)** in a many-to-one relationship.*
- ***Trips (duration)** maps to **Duration (id)** as a many-to-one relationship.*
- ***Trips (loc_to)** relates to **Assembly (ID)** in a many-to-one relationship.*

These connections were built within the Model View by linking the appropriate keys. All relationships are currently active and visually confirmed to maintain referential integrity. This structured data model enables seamless, cross-table analysis on factors such as ride location, payment types, and trip durations, ensuring consistency and accuracy in insights.

1.2. Find and Resolve Inconsistencies [5 Marks]

- Identify and resolve any inconsistencies or issues in the dataset that might affect the analysis.
- Clean the data to ensure it is structured properly for analysis, removing any irrelevant, duplicate, or erroneous entries.
- While performing the analysis, create calculated fields as needed to ensure the accuracy and relevance of the insights.

Solution:

After thoroughly reviewing all tables in the dataset, I found the data to be clean, organized, and ready for analysis. The following checks were performed:

- Key fields such as tripid, fare, and duration were examined for missing or null values, but none were found.

- Primary key columns were checked for duplicate records, and no duplicates were identified.

- There were no incorrect values or outliers like negative fares or durations.

- Data types were consistent across related fields in different tables.

Overall, the dataset demonstrated strong integrity across all linked tables, eliminating the need for data cleaning or correction. This provides a solid foundation for producing accurate and trustworthy insights.

2. Exploratory Data Analysis   [40 Marks]

2.1.   Classify Variables into Categorical and Numerical [2 Marks]

● Classify all the variables in the dataset into numerical and categorical types.

Solution:

A comprehensive review of all dataset tables revealed that the data is well-organized, clean, and suitable for analysis. The following validations were conducted:

- Critical fields including tripid, fare, and duration were inspected for any missing or null values, with none detected.

- Checks for duplicate entries in primary key columns confirmed the absence of duplicates.

- The dataset contained no invalid values or anomalies such as negative fares or durations.

- Data types were verified to be consistent across related columns in different tables.

The dataset maintained strong referential integrity across all linked tables, requiring no further cleaning or adjustments. This ensures the reliability and accuracy of any subsequent analysis

2.2.   Analyze Ride Demand Over Time [3 Marks]

● Explore the distribution of ride demand over time, including trends across different periods.
● Identify the peak demand periods. Choose an appropriate parameter for demand based on your own understanding.
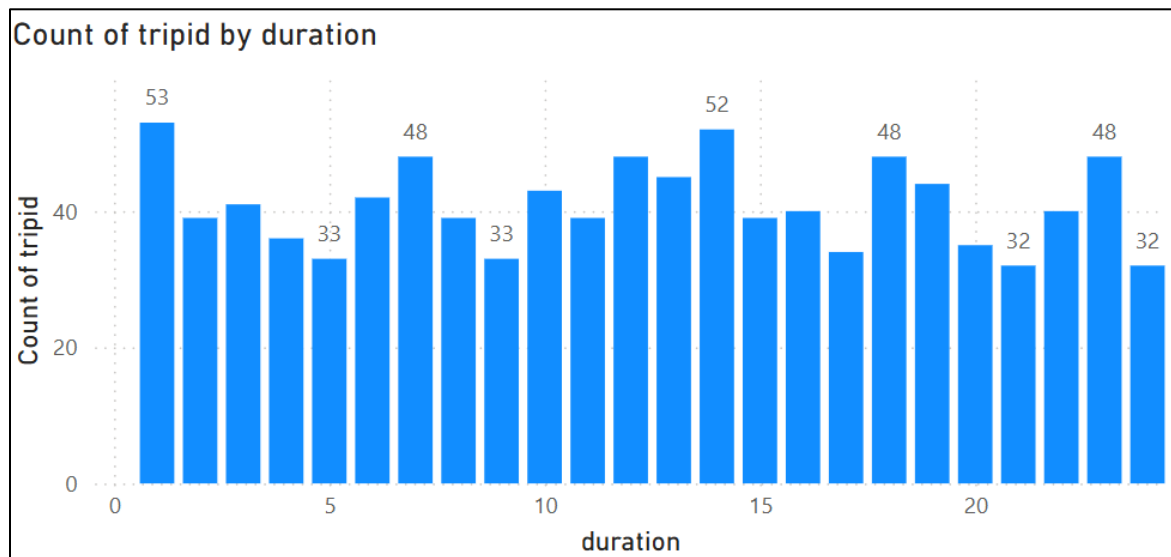
Solution:

To examine ride demand trends over time, I created a chart plotting the count of trips (tripid) against hourly time intervals (duration). This visualization displays how trip volume is distributed across the 24 hours of a day.

- **Observation:** The greatest ride demand occurs between midnight and 1 AM (0-1 hour interval), followed by a moderately high but steady demand from 1 AM to 6 AM. After this period, ride requests gradually decrease throughout the

day, reaching the lowest levels between 10 PM and midnight (22:00 - 24:00).

- **Trend:** The data indicates that most rides are requested late at night and in the early morning, which may be linked to limited public transportation or night shift commuters. Demand consistently falls during daytime and evening hours.

- **Peak Demand Period:** The highest concentration of rides is seen in the midnight to 1 AM window.

Count of tripid by duration

2.3. Proportion of Total Revenue from Different Time Periods [3 Marks]

- Calculate the proportion of revenue generated during different time periods and visualise how it contributes to total revenue.

Solution:

To calculate the proportion of revenue generated across different time periods, I used the fare column from the Trips table and grouped the data by duration (hourly slots).
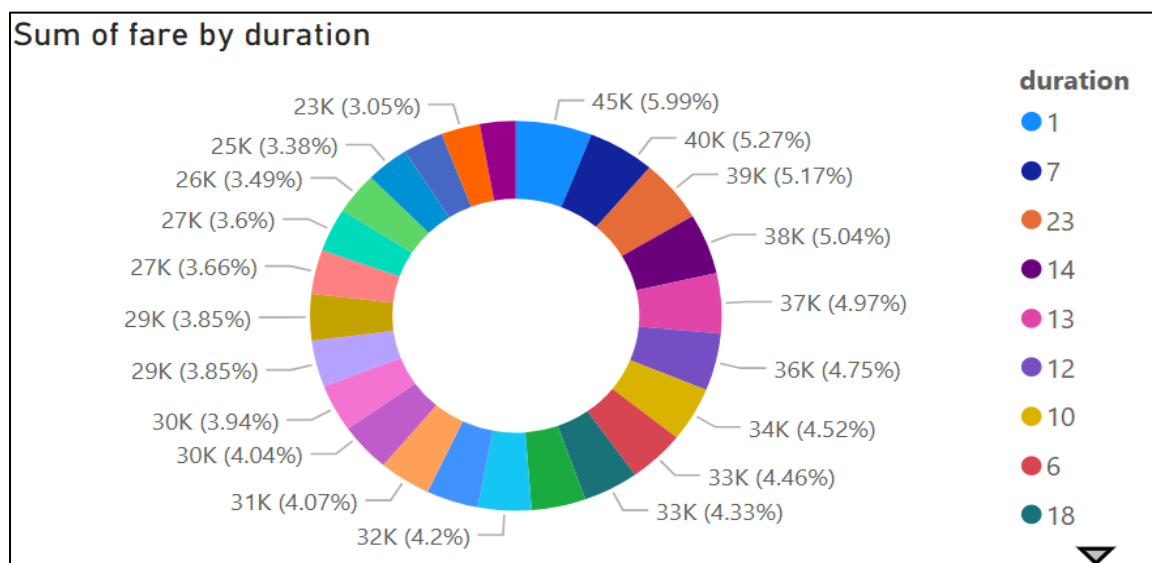
- A Donut chart was created in Power BI with:

  o **Axis:** duration

  o **Values:** Sum of fare

- The total fare was summed for each hour to illustrate the revenue contribution of each time period.

Observations:

- The highest revenue occurred during the 0–1 AM time slot, matching the peak ride demand period.
- Revenue steadily decreases throughout the day, reflecting a similar pattern to ride demand.
- The smallest revenue share was found between 22–24 hours.

Conclusion:

- There is a clear correlation between ride demand and revenue by hour. Time periods with higher trip volumes naturally generate more revenue.
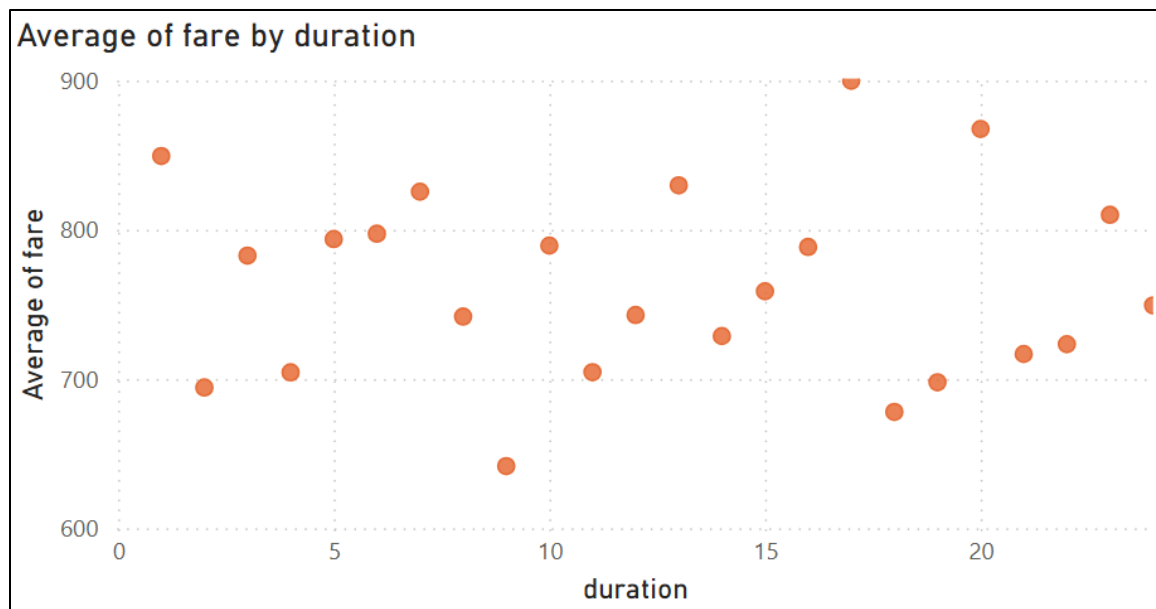

Sum of fare by duration

2.4. Explore the Relationship Between Trip Hour and Revenue [3 Marks]

- Investigate the correlation between trip hour and total fare.
- Explain any trends or patterns that emerge.

Solution:

A scatter chart was created to examine the relationship between ride duration and average fare

Average of fare by duration

From the chart, it is observed that the total fare generally increases as ride duration increases, although the relationship is not perfectly linear. Some shorter rides have higher average fares, likely due to factors like minimum fare rules, high demand surcharges, or peak-time pricing.
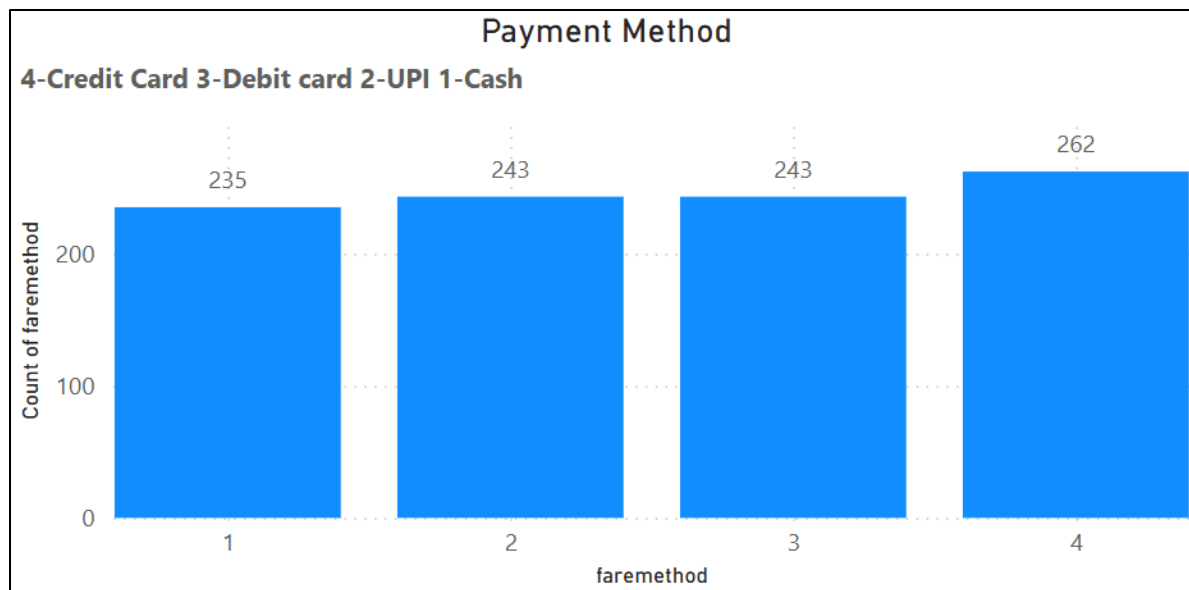
Overall, the chart indicates a moderate positive correlation—longer rides typically cost more on average, but fare amounts are also affected by other variables such as distance traveled and the time of day.

2.5. Examine the Popularity of Different Payment Methods [3 Marks]

- Analyse the distribution of various payment methods used by customers.
- Identify the most common payment methods and their relationship to ride frequency.

Solution:

The analysis of payment method usage reveals that credit cards are the most commonly used option, accounting for 262 trips. This is followed by UPI, cash, and debit cards. These findings indicate a clear preference for digital payment methods, particularly credit cards and UPI. The frequency of rides appears to be positively influenced by the accessibility and convenience of digital payment options.

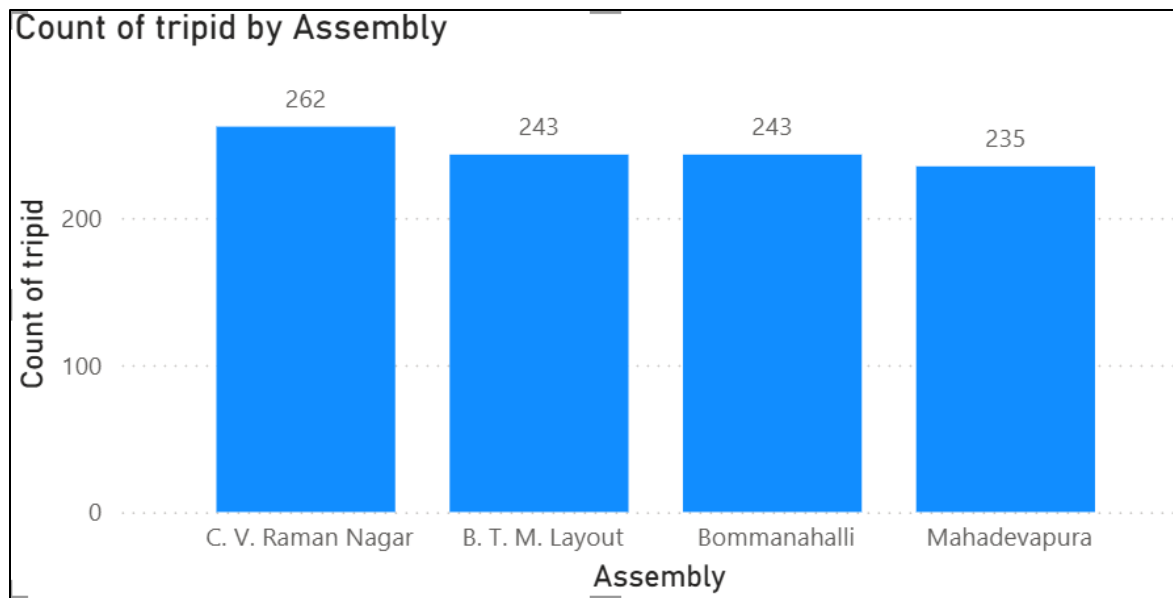## 2.6. Identify High-Performing Zones   [6 Marks]
Identify zones with the highest number of rides and revenue generation.
Analyse factors contributing to their performance:

- **2.6.1. Rides:** Identify pickup zones with the highest number of trip requests. **[3 marks]**

Solution:

1. Create   Relationships:
   • Connect Trips[loc_from] to Assembly[Assembly_ID] to associate location IDs with their corresponding assembly zone names.

2. Create   Clustered   Column   Chart:
   • Set the X-axis to Assembly from the Assembly table (zone names).
   • Set the Y-axis to tripid from the Trips table, using a Count aggregation.

3. Sort   the   Chart:
   • Click on the visual's ellipsis (...), select Sort by Trip Count, and choose Sort descending.

This visualization shows the number of trips originating from each pickup zone, with zones like C.V Raman Nagar recording the highest trip count of 262.

● **2.6.2. Revenue**: Identify pickup zones generating the highest revenue. **[3 marks]**

Solution:

I created a clustered column chart in Power BI to visualize the total revenue generated from each pickup zone using the fields "Assembly" and "Sum of fare."

Steps followed:

1. Imported the dataset into Power BI.
2. Chose the clustered column chart from the Visualizations pane.
3. Placed the *Assembly* field on the X-axis to represent pickup zones.
4. Placed the *fare* field on the Y-axis and applied the *Sum* aggregation to represent total revenue.

Outcome:

• **C.V Raman Nagar** recorded the highest revenue, with a total fare of ₹197941.

2.7.   Analyse Ride Time Periods Across Zones [4 Marks]

● Compare the trip trends for different time periods across pickup zones.

Solution:

To analyze how trips are distributed across different parts of the day, I created a custom column named Time_Period in the Duration table. Since the duration values were text ranges like "13-14", only the starting hour was extracted for binning into time segments.

DAX Formula Used:

DAX

```
Time_Period =
VAR hourText = LEFT(Duration[duration], FIND("-", Duration[duration]) - 1)
VAR hour = VALUE(hourText)
RETURN
SWITCH(TRUE(),
hour >= 6 && hour < 12, "Morning",
hour >= 12 && hour < 17, "Afternoon",
hour >= 17 && hour < 21, "Evening",
"Night"
```
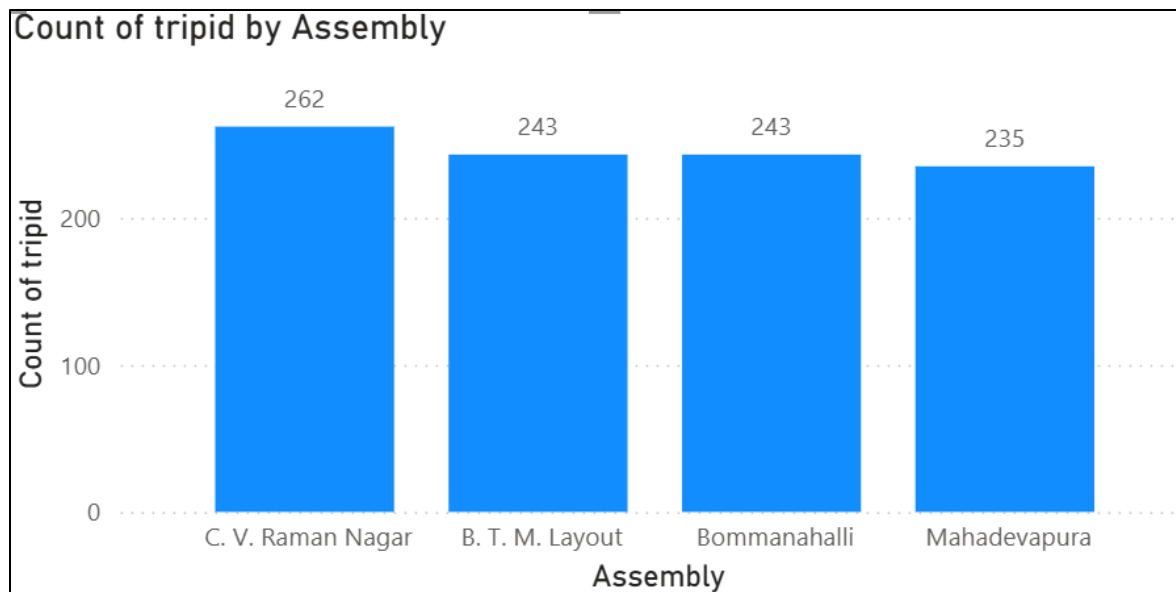
2.8.   Top Zones with Highest Trip Volume [3 Marks]
● Identify the top 5 pickup zones with the highest total number of completed trips.
● Analyse factors contributing to the higher number of trips.
Solution:

An analysis of completed trips reveals that Hoskote,

C.V. Raman Nagar, B.T.M. Layout, Bomanahalli and Mahadevapura are the top four pickup zones with the highest number of completed trips.

Count of tripid by Assembly

Count of tripid by Assembly:
- C. V. Raman Nagar: 262
- B. T. M. Layout: 243
- Bommanahalli: 243
- Mahadevapura: 235

2.9.   Basic Analytical Tasks [8 Marks]
- **2.9.1**
  What are the percentages of cancellations and successful rides by both driver and customer? **[3 marks]**

Solution:

DataSource:

Used the Trip_Details table, which contains columns:

customer_not_cancelled (1 = not cancelled, 0 = cancelled)
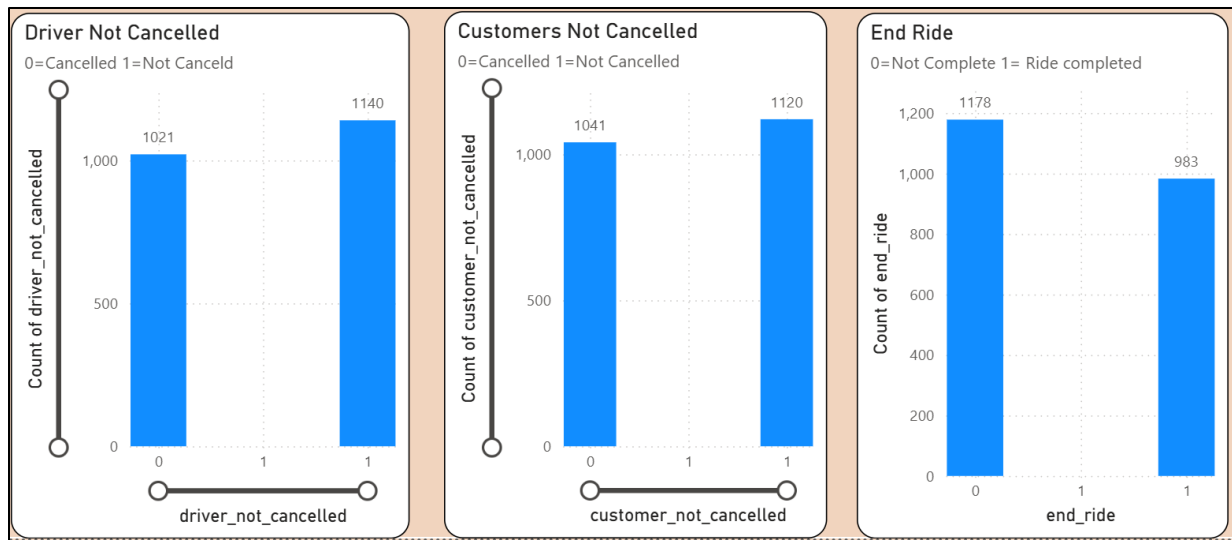
driver_not_cancelled (1 = not cancelled, 0 = cancelled)

end_ride (1 = ride completed, 0 = not completed)

Read the % distribution directly from the chart.

Final Answer:

● Cancelled by Driver: 1021
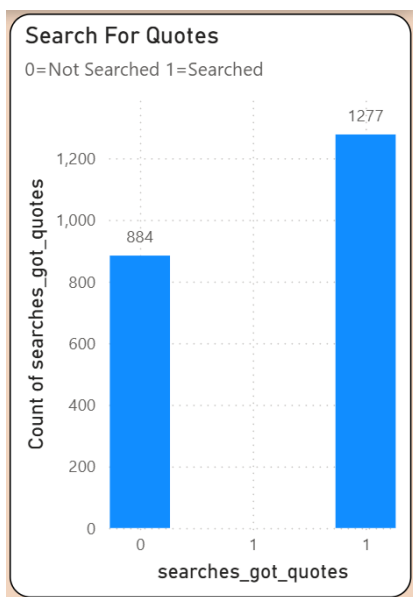
● Cancelled by Customer: 1041

●SuccessfulRide: 983

- **2.9.2**
  Analyse the percentage of people who completed trips after searching for quotes. Visualise the variation of this ratio by time periods. **[5 marks]**

Solution:

The final chart the users who searched for quotes completed the trip, indicating a high conversion rate or that the data may only include completed trips.

2.10. Create a Parameter and Use Filters [5 Marks]

● Create a parameter and use it as a filter on an appropriate subset of the data to interactively analyse and visualise different subsets of the data.
● Explain your choice of filter and insights drawn from this step.

Solution:

Added a slicer visual connected to the parameter to interactively adjust the fare range

This filter helped isolate the revenue contribution from low-fare vs high-fare trips based on Locations

3. Conclusion   [20 Marks]

3.1. Recommendations for Operational Efficiency   [10 Marks]

● Based on your findings from the analysis, provide recommendations on how Namma Yatri can optimize its operations.
● This could include strategies for improving resource allocation, reducing cancellations, or optimizing ride durations.
● Add supporting dashboards.

Solution:

1. Minimizing Trip Cancellations

● Namma Yatri should introduce a penalty and reward system for both drivers and users to discourage frequent cancellations and promote reliability.
● The app's ride-matching algorithm should be enhanced to consider previous cancellation behaviour, driver responsiveness, and location proximity.
● Accurate estimated time of arrival (ETA) notifications should be sent to users to manage expectations and reduce last-minute cancellations.

2. Enhancing Route Efficiency

● The platform should integrate real-time traffic data to dynamically suggest the fastest available routes for drivers.
● Time-based pricing should be introduced to spread demand more evenly

throughout the day and reduce congestion during peak hours.
● Frequently travelled routes should be optimized using historical data to pre-plan the most efficient paths.

3. Strategic Driver Allocation

● Namma Yatri should implement dynamic driver repositioning strategies by using predictive models to forecast ride demand across time slots and zones.
● Drivers should receive targeted surge-based incentives in high-demand and low-supply areas to ensure better supply-demand alignment.
● Driver shifts should be scheduled in a more data-driven way to ensure adequate driver availability during peak hours.

4. Empowering Driver Performance

● Drivers should be provided with performance dashboards that give insights into their completed rides, earnings per hour, and cancellation rates.
● The company should organize periodic training programs to improve driver behavior, efficiency in routing, and customer service.
● A monthly rewards program should be launched to recognize top-performing drivers and encourage consistent service.

5. Strengthening Customer Loyalty

● Push notifications and timely alerts should be used to inform users about driver delays, estimated pick-up time, and alternate ride options.
● Loyalty programs or reward points should be introduced to retain frequent users and reward them for consistent app usage without cancellations.
● Regular collection and analysis of user feedback should be done to identify pain points and implement targeted improvements.

3.2. Marketing and Operational Strategy Improvements   [10 Marks]
   ● Suggest improvements to Namma Yatri's marketing or operational strategies based on your analysis.
   ● Recommendations could involve promotional efforts, driver incentives, or regional targeting to increase customer satisfaction and service efficiency.
   ● Add supporting dashboards.

Solution:

1. Drive Engagement in Low-Traffic Areas and Hours

Introduce special deals such as cashback or limited-time discounts during early morning hours or in areas with minimal demand. These offers can help stimulate bookings, leading to better vehicle utilization and expanded reach.

2. Reward Consistent and Reliable Drivers

Implement a reward system that offers financial incentives or recognition to drivers with high ride completion and low cancellation metrics. This motivates better service standards and ensures more dependable rides for users.

3. Run Targeted Campaigns in Problematic Zones

Focus marketing and awareness efforts in zones with high cancellation rates. Personalized ads and local engagement can rebuild trust and encourage users to re-engage with the platform in those regions.

4. Improve Efficiency with Smart Driver Placement

Enhance ride experience by strategically placing drivers during peak times based on demand forecasts. Optimized routing and proactive deployment will reduce delays and boost user satisfaction.

5. Encourage Cashless Transactions with Incentives

Promote UPI and card payments by offering exclusive perks. This reduces dependency on cash, speeds up operations, and creates a smoother, more modern user experience

.