Methodology Report:

Visualisation & Analysis on Namma Yatri Data

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

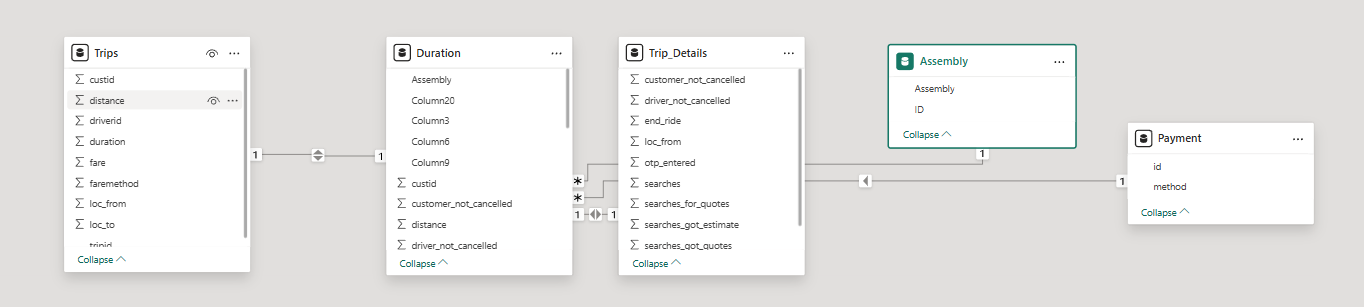
Sections: 3 sections (10 marks + 40 marks + 20 marks)

## Analysis and Visualisation

### 1. Data Preparation [10 Marks]

1. 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:*



1. 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:*

**Upon careful inspection of all the tables in the dataset, no major inconsistencies were found. The data is clean, well-structured, and ready for analysis. I checked for:**

● **Missing or null values in key fields such as tripid, fare, duration, etc., and found none.**

● **Duplicate entries, especially in primary key columns, but no duplicates were detected.**

● **Incorrect or outlier values such as negative fares or durations were not present in the data.**

● **Consistent data types across related columns in different tables.**

### 2. Exploratory Data Analysis [40 Marks]

1. Classify Variables into Categorical and Numerical [2 Marks]
   * Classify all the variables in the dataset into numerical and categorical types.

*Solution:*

**Numerical Variables:**

**• duration (from Trips, Duration)**

**•distance (from Trips)**

**• searches, searches\_for\_quotes, searches\_got\_estimate,**

**searches\_got\_quotes (from Trip\_Details)**

**Categorical Variables:**

**• custid (Trips)**

**• driverid (Trips)**

**• faremethod (Trips)**

**• loc\_from, loc\_to (Trips)**

**• tripid (Key in multiple tables, categorical for joining)**

**• Assembly (Assembly table – area names like Mahadevapura, etc.)**

**• method (Payment method)**

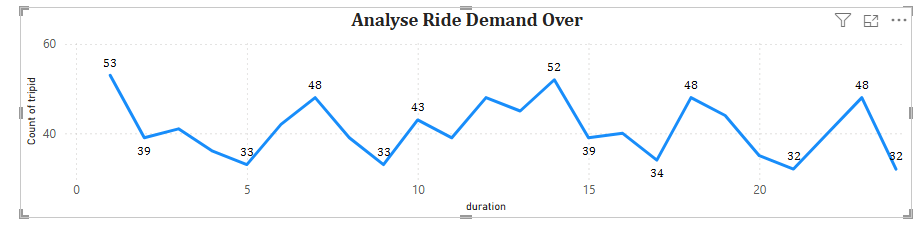
**• customer\_not\_cancelled, driver\_not\_cancelled, otp\_entered (Trip\_Details –**

**binary categories)**

**Variables were classified based on their data types and role in analysis. Numerical variables are those representing measurable quantities, while categorical variables represent labels, groupings, or identifiers.**

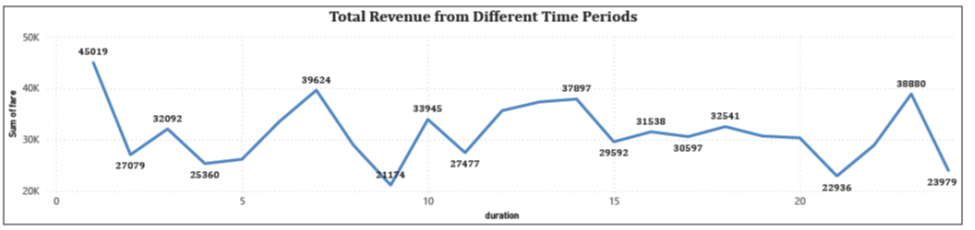
1. Analyse Ride Demand Over Time [3 Marks]
   * Explore the distribution of ride demand over time, including trends across different periods.
   * Identify the peak demand periods. Choose an appropriate parameter for demand based on your own understanding.

*Solution:*



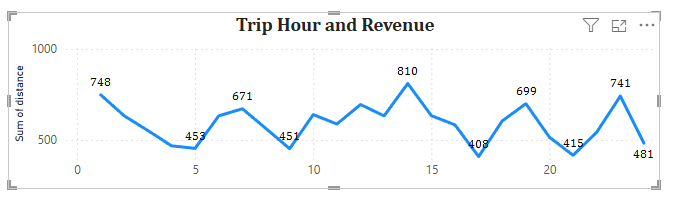
1. Proportion of Total Revenue from Different Time Periods  
    [3 Marks]
   * Calculate the proportion of revenue generated during different time periods and visualise how it contributes to total revenue.

*Solution:*

****

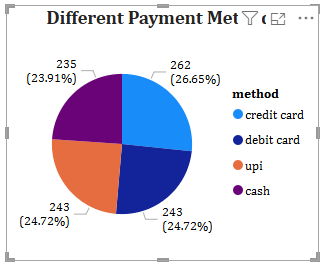
1. Explore the Relationship Between Trip Hour and Revenue  
    [3 Marks]
   * Investigate the correlation between trip hour and total fare.
   * Explain any trends or patterns that emerge.

*Solution:*



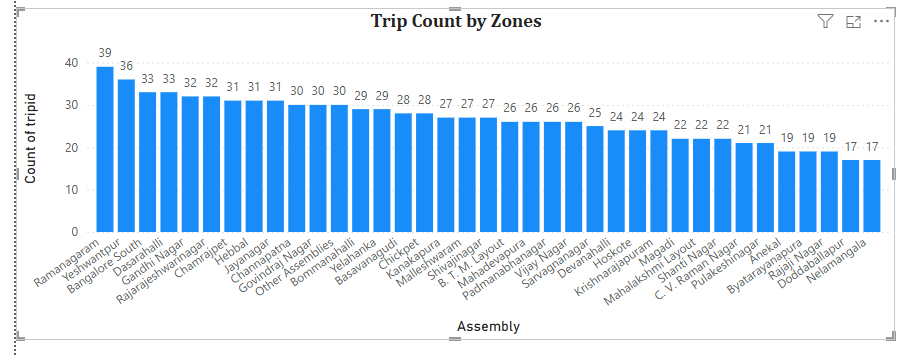
1. Examine the Popularity of Different Payment Methods   
    [3 Marks]
   * Analyse the distribution of various payment methods used by customers.
   * Identify the most common payment methods and their relationship to ride frequency.

*Solution:*



1. 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]

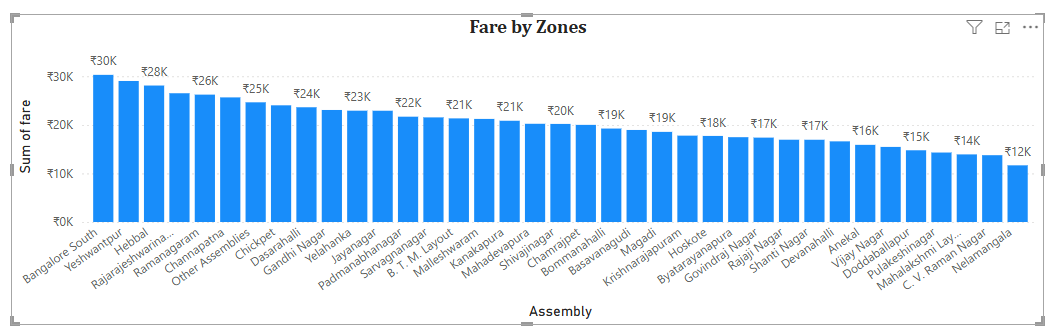
*Solution:*



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

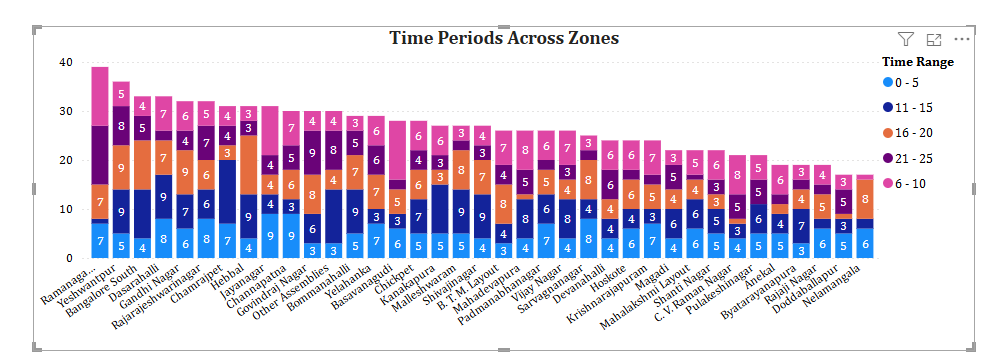
*Solution:*

**.**



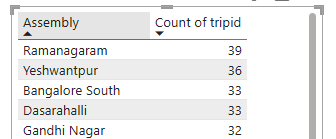
1. Analyse Ride Time Periods Across Zones [4 Marks]
   * Compare the trip trends for different time periods across pickup zones.

*Solution:*



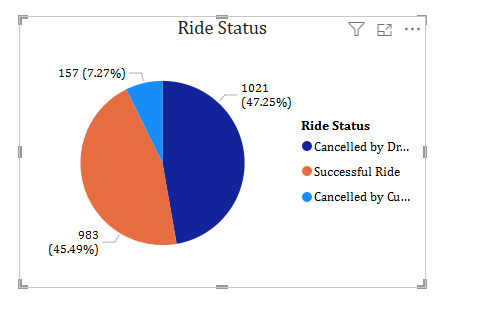
1. Top Zones with Highest Trip Volume [3 Marks]
   * Identify the top 5 pickup zones with the highest total number of completed trips.
   * Analyse factors contributing to the higher number of trips.

*Solution:*



1. 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:*



* + 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:*



1. 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:*

**Time Range and period.**

### 3. Conclusion [20 Marks]

1. Recommendations for Operational Efficiency [10 Marks]
   * Based on your findings from the analysis, provide recommendations on how Namma Yatri can optimise its operations.
   * This could include strategies for improving resource allocation, reducing cancellations, or optimising ride durations.
   * Add supporting dashboards.

*Solution:*

**1. Reducing Ride 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 behavior, 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. Optimizing Ride Durations**

● 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.Improving Resource 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. Enhancing Driver Productivity**

● 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. Boosting User Retention through Experience**

● 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.

1. 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.Promotional efforts in low-demand zones and off-peak hours**

● Offer discounts or cashback during early mornings and in low-demand areas.

● Increases ride requests and improves driver utilization.

**2. Incentivize high-performing drivers**

● Give bonuses to drivers with high ride completion rates and low cancellation

rates.

● Improves service reliability and reduces cancellation instances.

**3. Target high-cancellation zones with marketing campaigns**

● Run localized ads and awareness drives in areas with more cancellations.

● Builds trust and increases service usage in underperforming zones.

**4. Optimize routing and driver deployment during peak hours**

Adjust driver distribution and routes to reduce ride delays and improve

efficiency.

Enhances punctuality and user satisfaction.

**5. Promote digital payment options**

● Encourage UPI and card payments with exclusive offers.

● Reduces cash handling and improves operational efficiency