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]

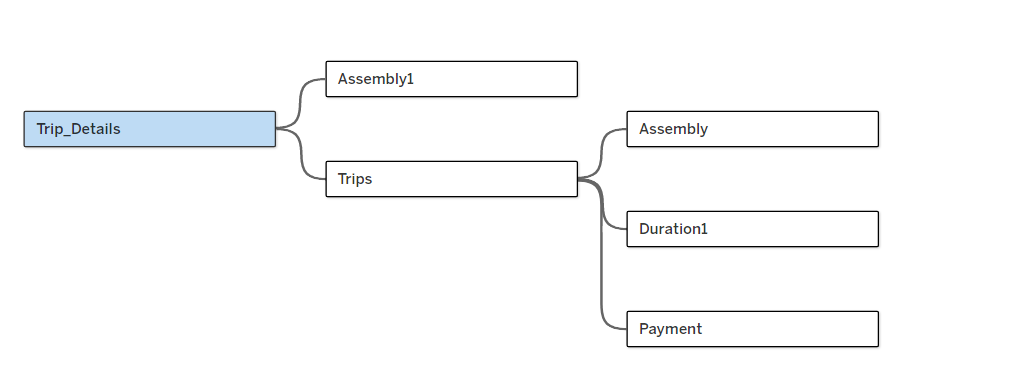
1. Import and Join Tables Correctly [5 Mark]
   * Import the Namma Yatri dataset into Tableau/Power BI.
   * Ensure that you correctly join all tables to create a unified dataset for analysis.
   * Verify the relationships between different tables and confirm that data from various sources is properly aligned for accurate insights.

*Solution:*

We imported all the tables in tableau and established the relationship as follows :-

|  |  |  |  |
| --- | --- | --- | --- |
| Table 1 | Table 2 | Relational coloumn | Relationship type |
| Trips | Assembly | loc\_from, loc\_to & ID | Many to one |
| Trips | Payment | Duration & ID | Many to one |
| Trips | Duration | Fare method & ID | Many to one |
| Trips | Trip\_details | Trip\_id & Trip\_id | One to Many |
| Trip\_details | Assembly | loc\_from & ID | Many to one |

A screenshot of a computer

AI-generated content may be incorrect.

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:*

1. We had hide the columns which were not required for analysis such as driver\_id , Cust\_id.
2. We created Calculated field like same Location Trip to filter out those rows for which loc\_from & loc\_to are same and having more distance & fare, these rows are classified as erroneous entries.
3. We created calculated field as cancelled trip which marks trip as cancelled when either driver cancels or customer cancels the trip.
4. We changed datatype of categorical data from interger to string because after conversion during analysis by default it will show as count not sum.

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:*

|  |  |  |
| --- | --- | --- |
| Table Name | Column Name | Categorical / Numerical |
| Assembly | Assembly\_ID | Categorical |
| Assembly | Categorical |
| Duration | duration\_id | Categorical |
| duration | Categorical |
| Payment | id | Categorical |
| method | Categorical |
| Trip Details | tripid | Categorical |
| loc\_from | Categorical |
| searches | Categorical |
| searches\_got\_estimate | Categorical |
| searches\_for\_quotes | Categorical |
| searches\_got\_quotes | Categorical |
| customer\_not\_cancelled | Categorical |
| driver\_not\_cancelled | Categorical |
| otp\_entered | Categorical |
| end\_ride | Categorical |
| Trips | tripid | Categorical |
| faremethod | Categorical |
| fare | Numerical |
| loc\_from | Categorical |
| loc\_to | Categorical |
| driverid | Categorical |
| custid | Categorical |
| distance | Numerical |
| duration | Categorical |

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:*

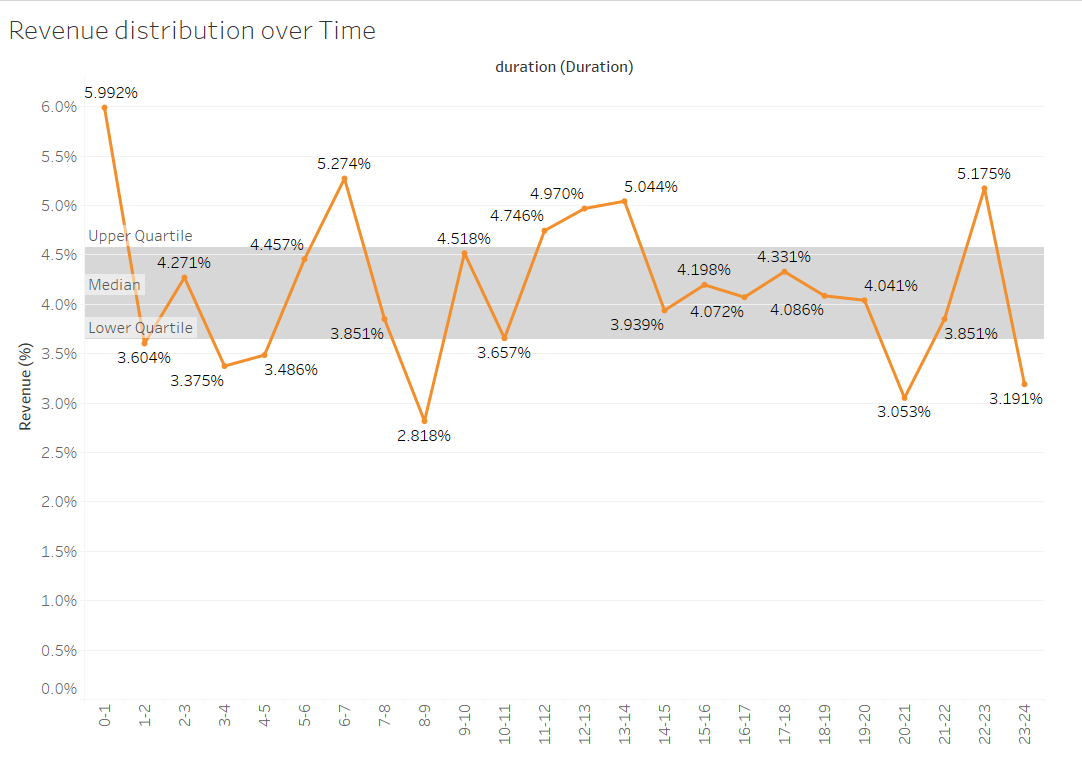
A graph with orange lines and numbers

AI-generated content may be incorrect.

Ride demands are normally distributed over time as per trend line. We considered count of trip\_id from trip table as demand since these are completed rides, we won’t be able use trip\_id from trip\_details table because there we don’t have data of the duration for that.

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

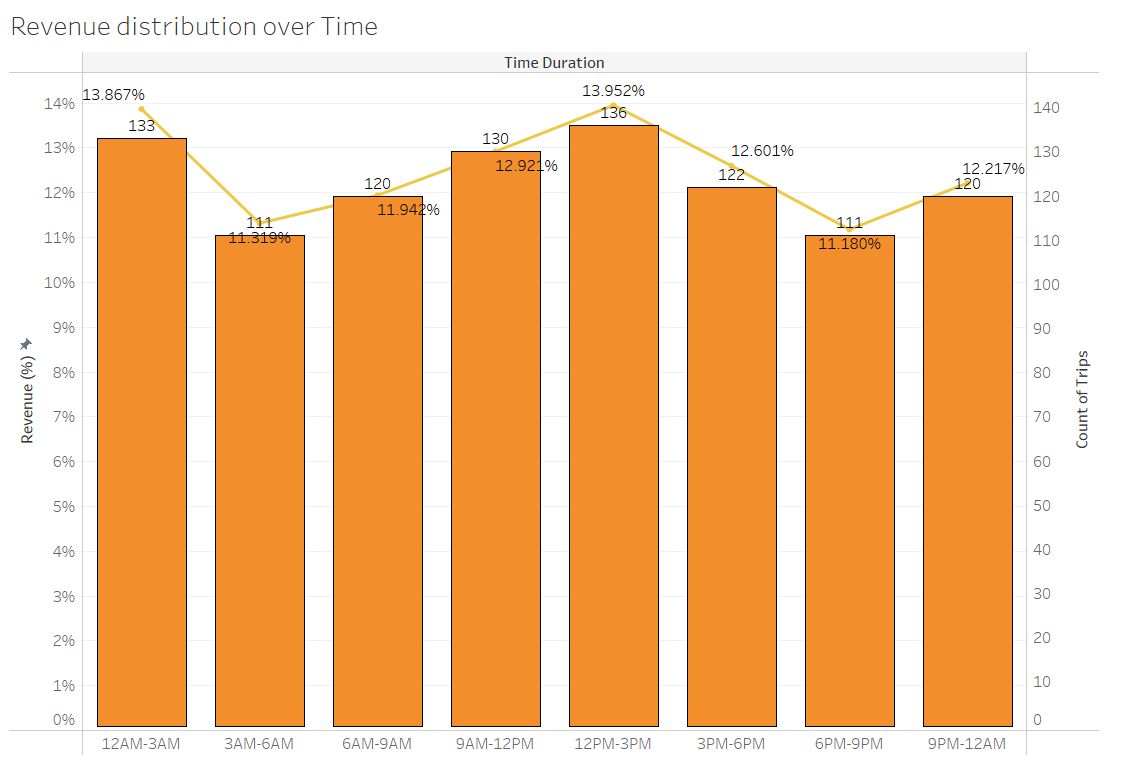
*Solution:*



We created Line plot of Revenue Generation (Fare attribute from Trip table) and duration and marked quartile which shows distribution of revenue of different time periods and shown % of total contribution for each time zone.

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:*

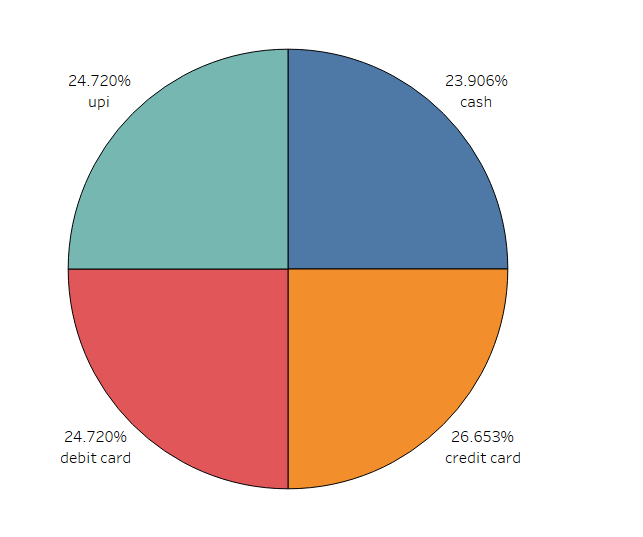


* + - 1. AS we can see from above chart revenue decreases as count of trip decreases which is natural
      2. we can see during peak hours; revenue is not at peak in above chart but at an average level.
      3. Revenue is at peak during noon time (12PM-3PM) & mid night time(12AM-3AM).

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

*Solution:*

As we can see from the below chart, Credit card payment method is used by the 26.65% of the customers which is most of them all payment methods



.

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:*

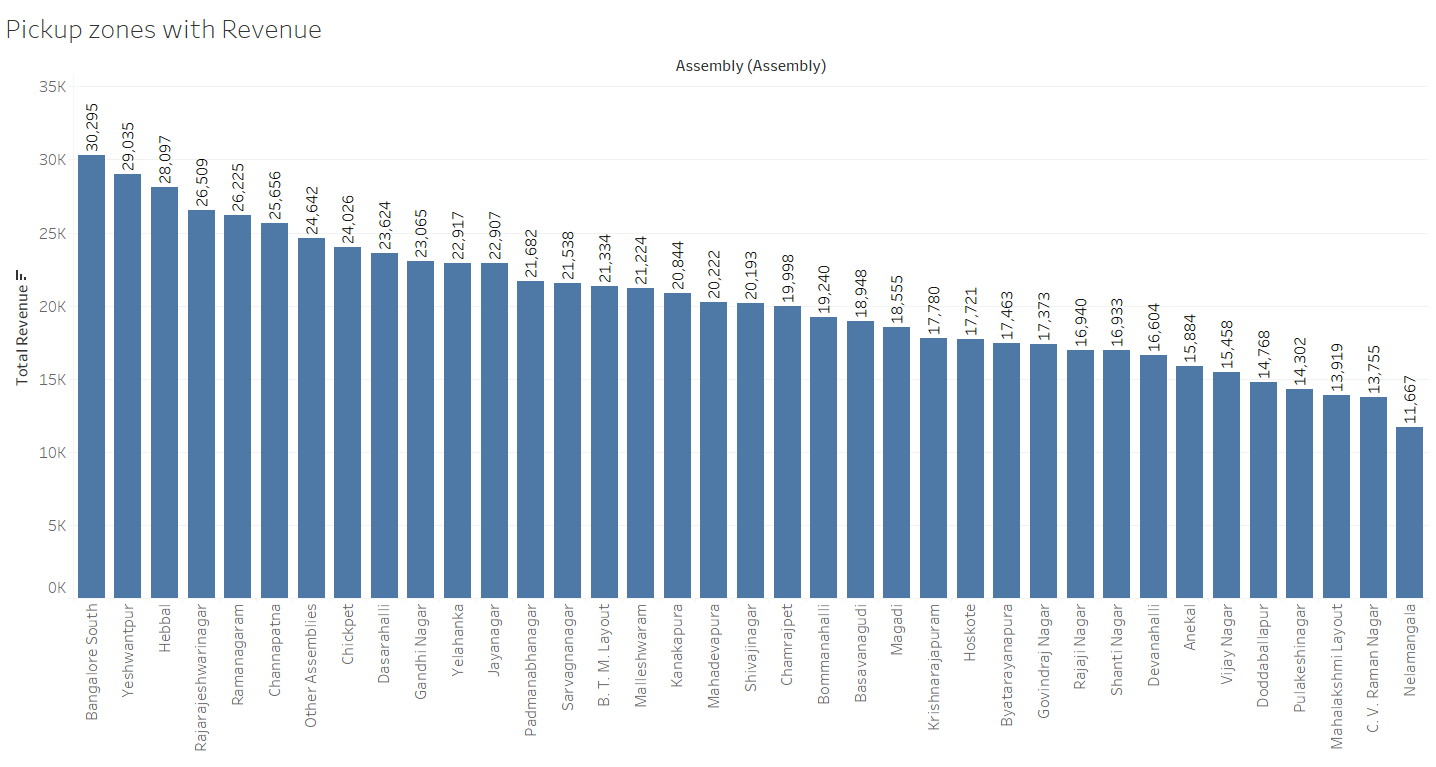
A graph of blue lines

AI-generated content may be incorrect.

Highest number of pickup request are generated from Mahadevapura i.e. 75 no’s from overall trips.

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

*Solution:*

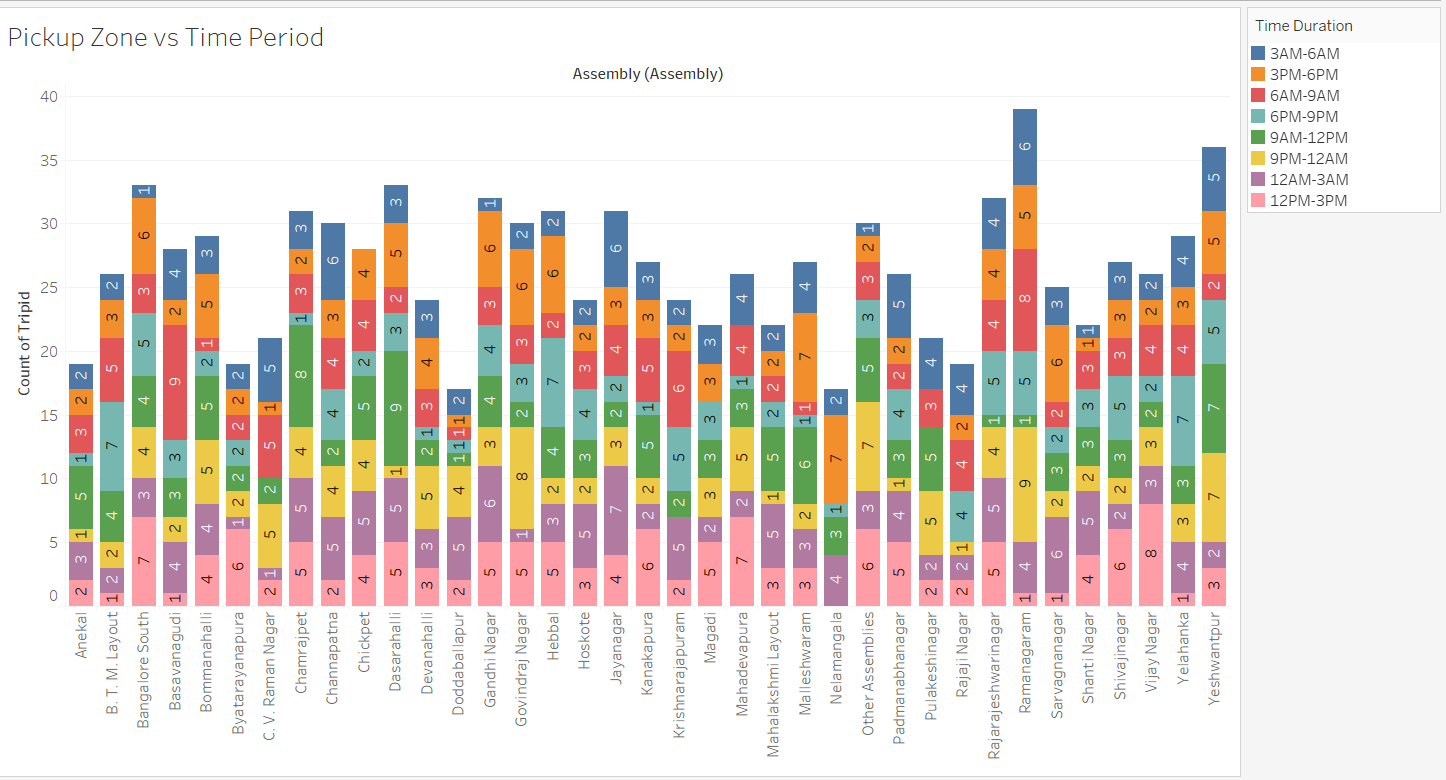


The highest revenue generating pickup zone is Bangalore south i.e. 30,225 Rs.

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

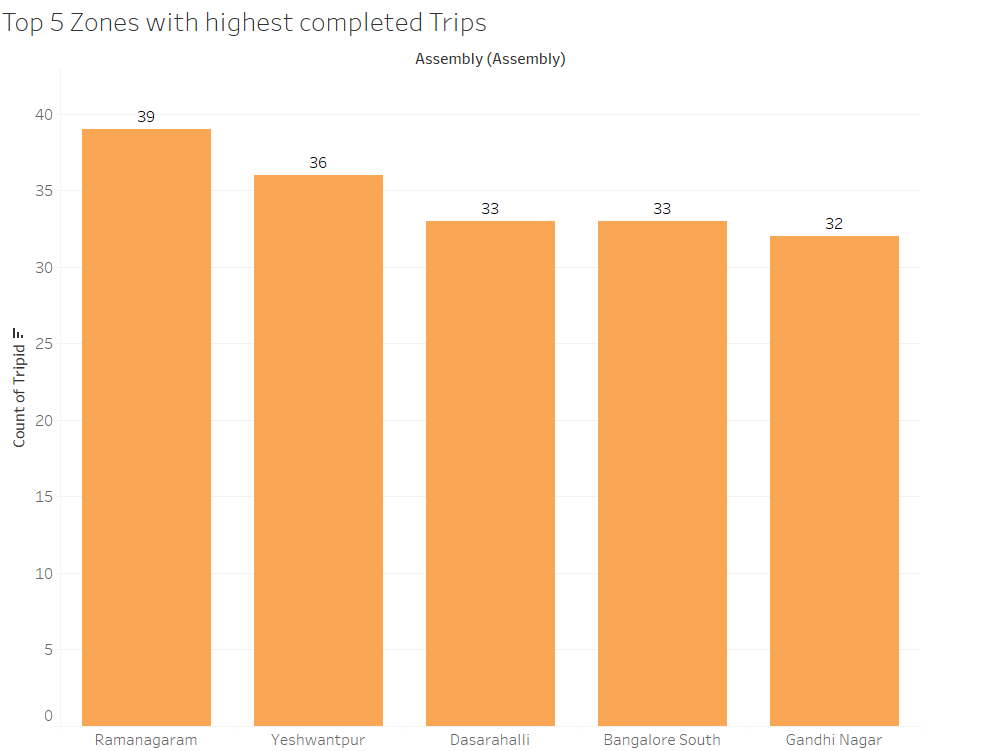
*Solution:*

From the below attached chart we can see trip trend across different time periods across pickup zones.



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

*Solution:*



Factors for top 5 pickup zones with highest no of trips.

* + - 1. Minimum 80% of people got an Estimate for the Ride.
      2. Minimum 56% of people got Driver allotted for the Ride
      3. Minimum 45% of customers are ending the ride

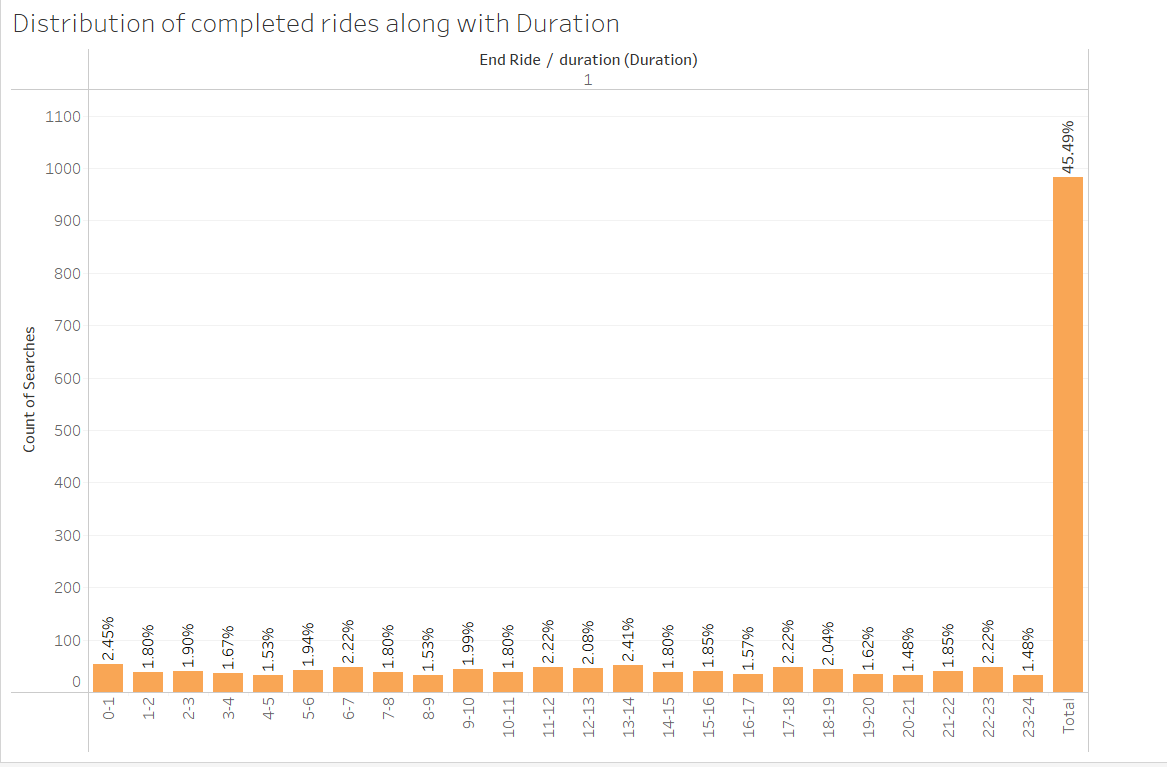
1. Basic Analytical Tasks [8 Marks]
   * 2.9.1   
     What are the percentages of cancellations and successful rides by both driver and customer? [3 marks]

*Solution:*

|  |  |  |
| --- | --- | --- |
|  | For Drivers | For Customer |
| % of successful Rides | 52.75% | 51.83% |
| % of Cancelled Rides | 47.25% | 48.17% |

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

*Solution:*



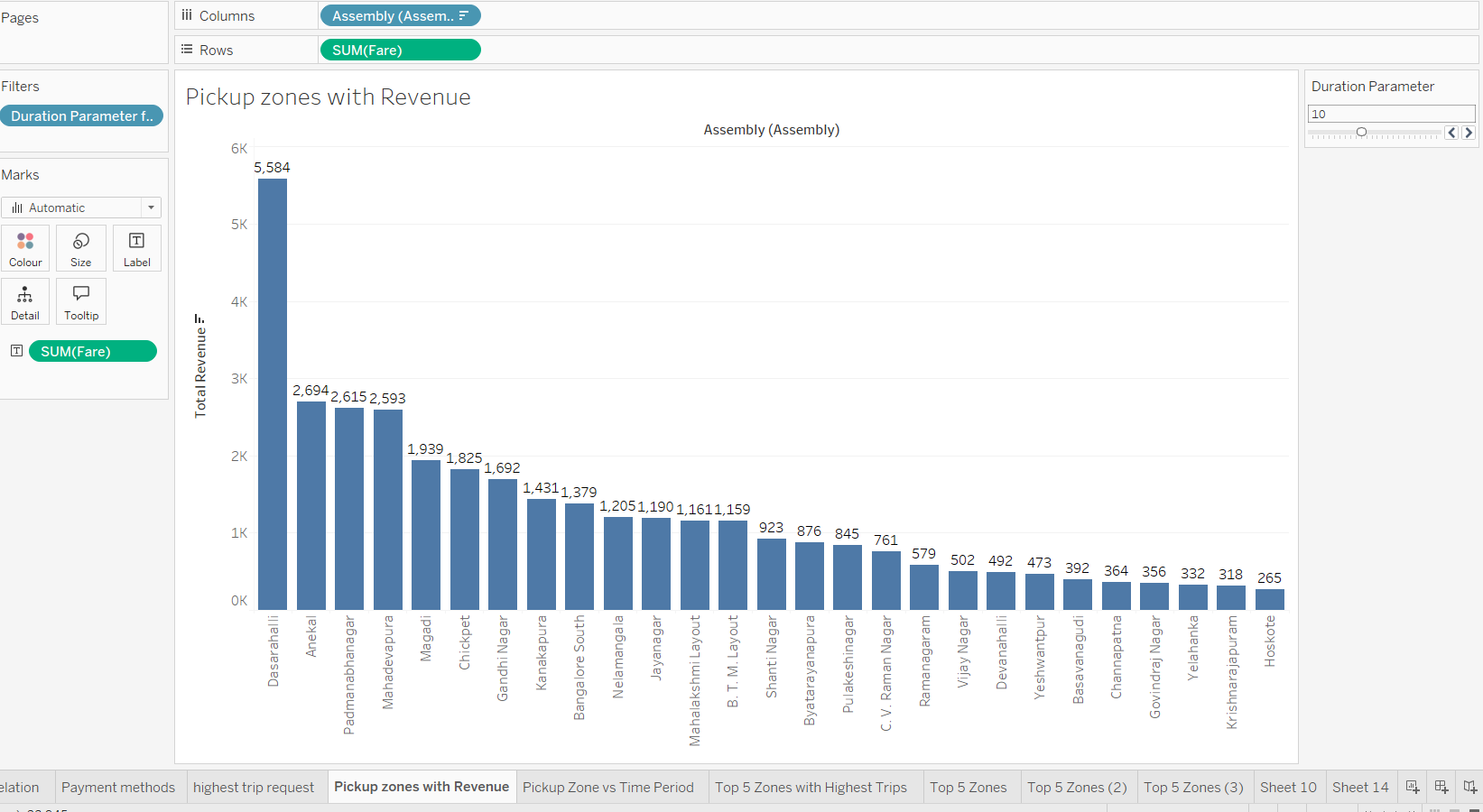
1. Create a Parameter and Use Filters [5 Marks]
   * Create a parameter and use it as a filter on an appropriate subset of the data to interactively analyse and visualise different subsets of the data.
   * Explain your choice of filter and insights drawn from this step.

*Solution:*

We created Duration Parameter and Duration parameter filter as calculated field, by which we can able to change the Graph by duration Parameter.

We choose this calculate hourly revenue across different locations.

Insights:- Maximum hourly revenue generated from Dasarahalli pickup location between 9-10AM.



### 3. Conclusion [20 Marks]

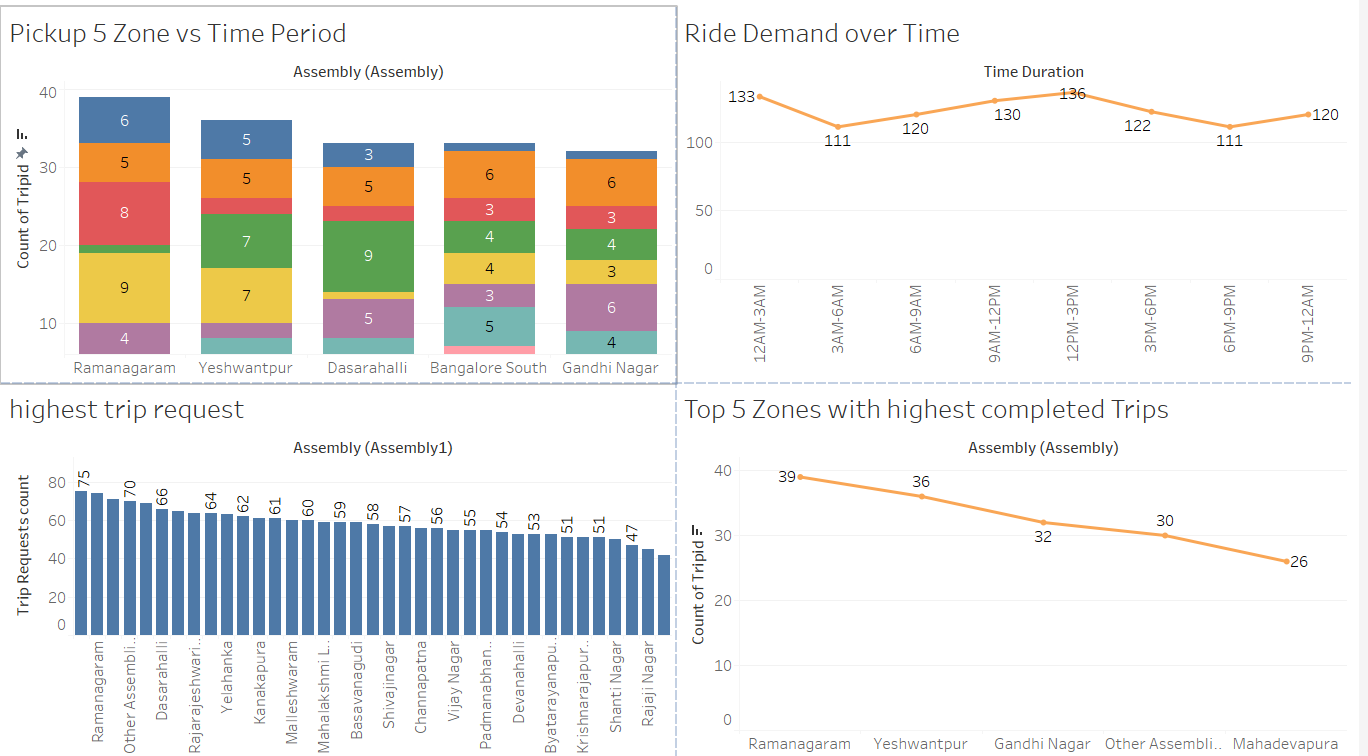
1. Recommendations for Operational Efficiency [10 Marks]
   * Based on your findings from the analysis, provide recommendations on how Namma Yatri can optimise its operations.
   * This could include strategies for improving resource allocation, reducing cancellations, or optimizing ride durations.
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

*Solution:*

Recommendations:-

* + - 1. Not having peak hours rides as peaks, increase the no of drivers during peak hours like 6AM-9AM & 3PM-6PM.
      2. Focus more on the requested top five pick up zones to allocate the Drivers efficiently for increasing the completed ride counts.
      3. To reduce cancellation, need to allocate drivers with top 5 locations in their respective highest pickup duration.

Dashboard as follows:-



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:*

Improvements:-

* + - 1. For more no of completed rides, incentives can be given to driver so that they will not cancel rides.
      2. Since Max hourly revenue is generated during 12AM-1AM & 12PM-1PM, Namma yatri has to focus between 5PM-7PM to increase revenue during peak hours.
      3. Location wise bangalore south is highest revenue generated pickup location so namma yatri has to focus more on this location.

Dashboard as follows:-

* + - 1. 