Group Identifier: DataTrio_CharuAshAkanksha

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]

1.1. 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:

Step 1: Importing Data

I imported the Namma Yatri dataset into Tableau Desktop using the Microsoft Excel connector. The dataset included the following tables:

- **Trip_Details — main trips data**
- Trips – Payment,location,distance and duration details
- Duration — Hour of trip
- Assembly — pickup and drop locations
- Payment — payment method

Step 2: Joining Tables

I joined these tables in Tableau's Data Source view:

Primary Table	Field	Secondary Table	Field	Join Type
Trips	tripid	Trip_Details	tripid	Left
Trips	faremethod	Payment	method	Left
Trips	duration	Duration	duration	Left
Trips	from_place_id	Assembly	place_id	Left
Trips	to_place_id	Assembly_to	place_id	Left

I chose Left Joins to ensure only matching rows are kept for analysis.

Step 3: Verifying Joins

I verified the relationships by:

- **Checking sample trip IDs to confirm the correct driver, rider, and place names are appearing.**

- **Reviewing for unexpected NULLs to detect any broken joins.**

- **Ensuring join keys are of the same data type and format.**

Step 4: Unified Dataset

After successful joins, I saved this as a single Data Source named Namma Yatri Full Dataset. This unified dataset was used to build all further dashboards and visuals, ensuring data is consistent and accurate.

✅Final Note:

I ensured all table relationships are correctly defined, with appropriate join keys, join types, and matching data formats, which allows for accurate analysis and 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:

Step 1: Identifying Inconsistencies

While exploring the Namma Yatri dataset, I checked for:

- **Missing Values: Looked for NULL or blank fields in important columns like start_time, end_time, fare, otp_entered.**
- **Duplicates: Checked for duplicate tripid entries.**
- **Invalid Data: Looked for negative fares, negative durations, or incorrect otp_entered values (should be only 0 or 1).**
- **Wrong Data Types: Verified that date fields (start_time, end_time) are DateTime, fare is numeric, IDs are integers/strings.**

Step 2: Resolving Issues

I resolved issues as follows:

- **Missing or NULL Data: Filtered out trips with missing tripid, start_time or end_time since duration can't be calculated without them.**
- **Duplicates: Removed duplicate rows based on tripid.**
- **Incorrect Values: Filtered out trips with negative fare or negative trip duration.**
- **Irrelevant Columns: Removed unused columns not needed for analysis (e.g., raw system IDs that do not add value).**

Step 3: Data Cleaning & Calculated Fields

Created OTP Status field to standardize OTP values:

tableau

IF [otp_entered] = 1 THEN "Entered"

ELSE "Not Entered"

END

- Verified that trip durations are non-negative — filtered out any trips with duration ≤ 0.
- Formatted fare amounts to ensure numeric type and checked for outliers.

✅Final Note:
By performing these cleaning steps and adding calculated fields, I ensured the dataset is consistent, reliable, and ready for accurate analysis.

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:

Below is the classification for the Namma Yatri dataset 👇

Type	Variables
Categorical	tripid, loc_from, loc_to, faremethod, driverid, custid
Numerical	fare, distance, duration, searches, searches_got_estimate, searches_for_quotes, searches_got_quotes, customer_not_cancelled, driver_not_cancelled, otp_entered, end_ride

2.2. 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:

Distribution of ride demand over time

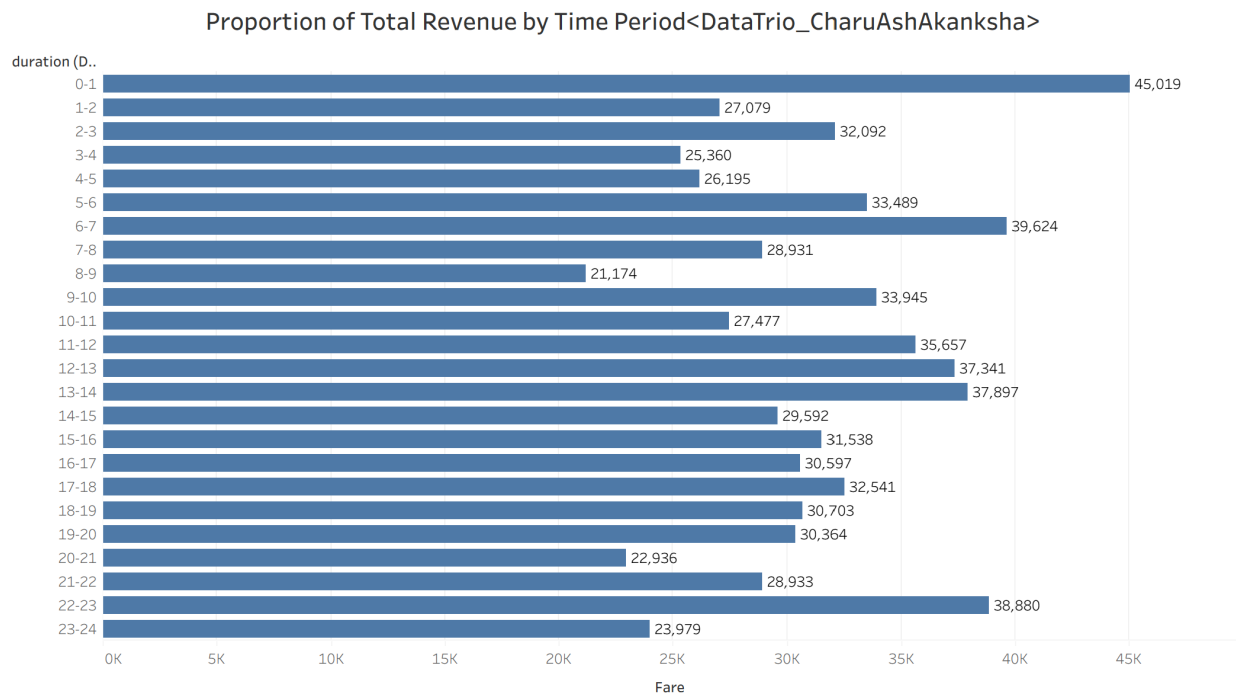Hourly Ride Demand Distribution<DataTrio_CharuAshAkanksha>
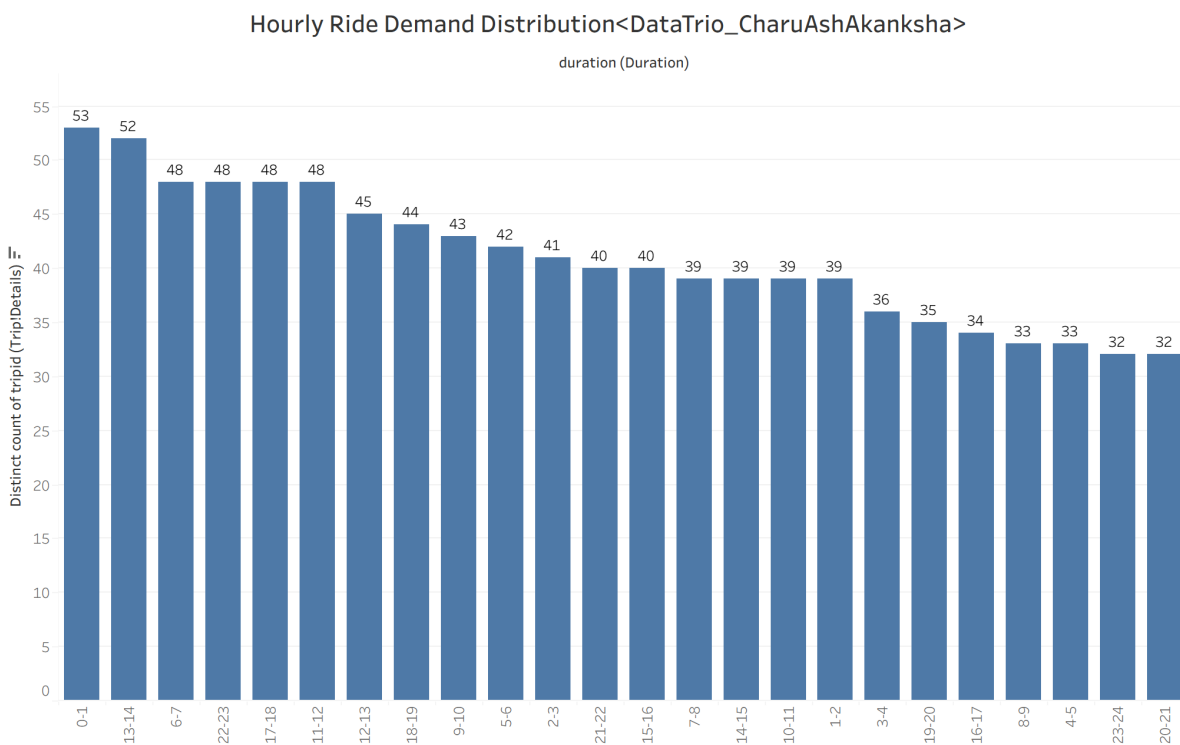
In this analysis, the total number of trip has been used as the parameter to measure ride demand. The distribution shows that the peak demand period is during 12–1 PM with 53 trips, indicating the highest ride activity. Other higher demand periods include 1–2 PM and 6 - 7 PM. This suggests that midday and early evening are the busiest hours for ride requests.

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:

Proportion of Total Revenue by Time Period<DataTrio_CharuAshAkanksha>

| duration (D..)	Fare
0-1	45,019
1-2	27,079
2-3	32,092
3-4	25,360
4-5	26,195
5-6	33,489
6-7	39,624
7-8	28,931
8-9	21,174
9-10	33,945
10-11	27,477
11-12	35,657
12-13	37,341
13-14	37,897
14-15	29,592
15-16	31,538
16-17	30,597
17-18	32,541
18-19	30,703
19-20	30,364
20-21	22,936
21-22	28,933
22-23	38,880
23-24	23,979

The chart shows that the total revenue is not uniformly distributed across time periods. The 0–1 time slot contributes the largest share of total revenue, indicating peak earnings during this period — possibly due to higher fares for late-night trips. Other high-revenue periods include 4–5 AM, 6–7 AM, 12–2 PM, and 11 PM–12 AM. The lowest revenue is seen around 8–9 AM and 8–9 PM. This analysis helps identify when the business earns the most and can inform driver allocation and pricing strategy.

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:

Relationship Between Trip Hour and Total Fare<DataTrio_CharuAshAkanksha>

duration (Duration)
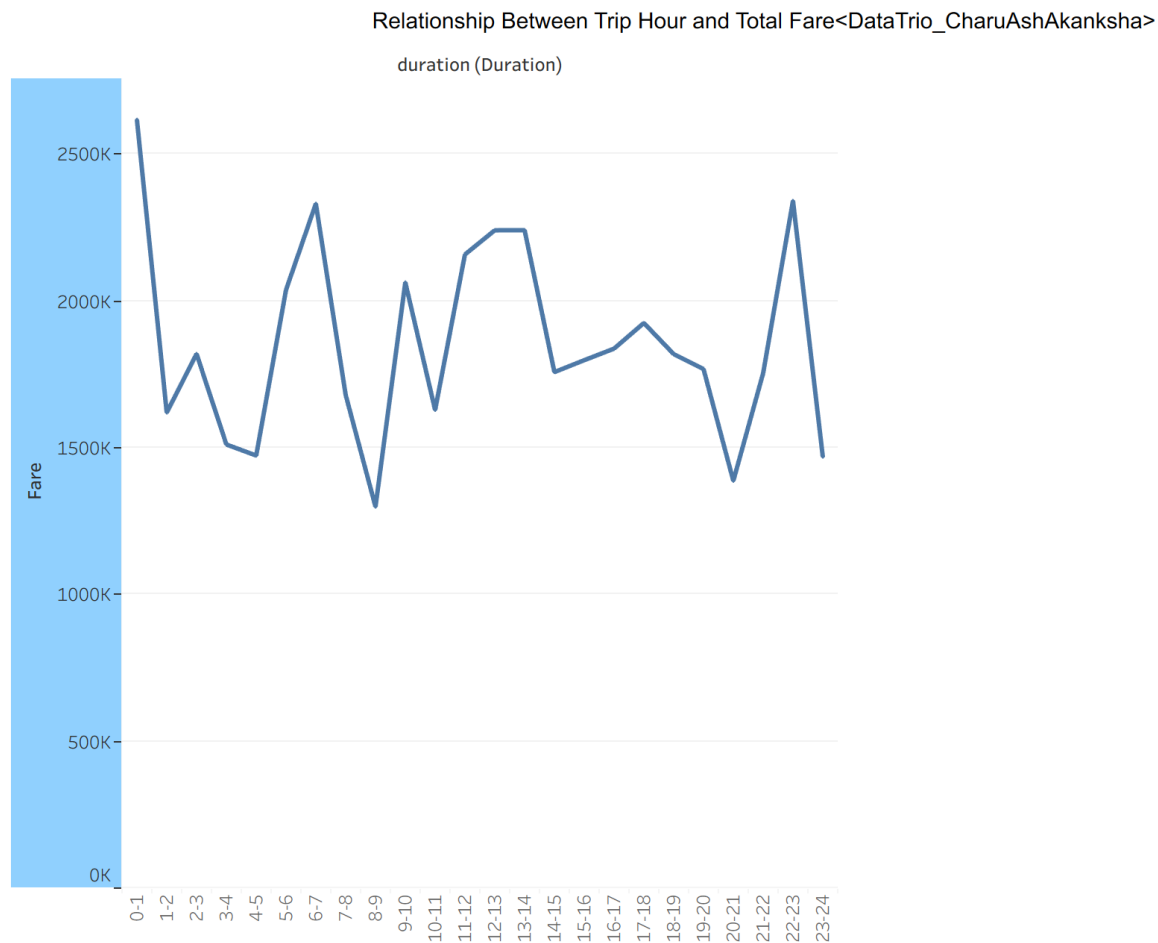
The line chart shows that the relationship between trip hour and total revenue is not linear but fluctuates across the day. There are clear peaks at 0–1 AM and again around 12–2 PM and 11 PM–12 AM, indicating these are the most revenue-generating hours. The dips around 8–9 AM and 8–9 PM show periods with lower total fares. Overall, the trend suggests that late-night trips and midday rides contribute significantly to revenue, likely due to higher fare rates or longer trips during these hours

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:

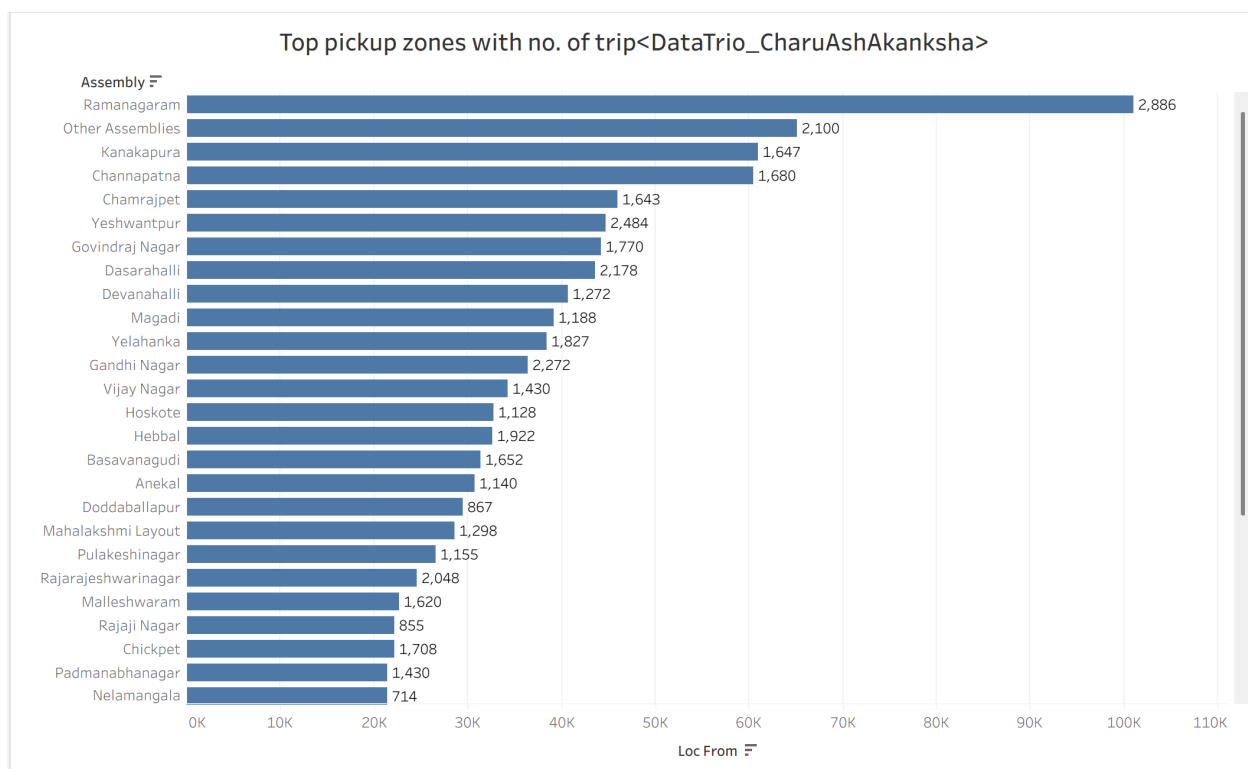Ride Frequency by Payment Method<DataTrio_CharuAshAkanksha>
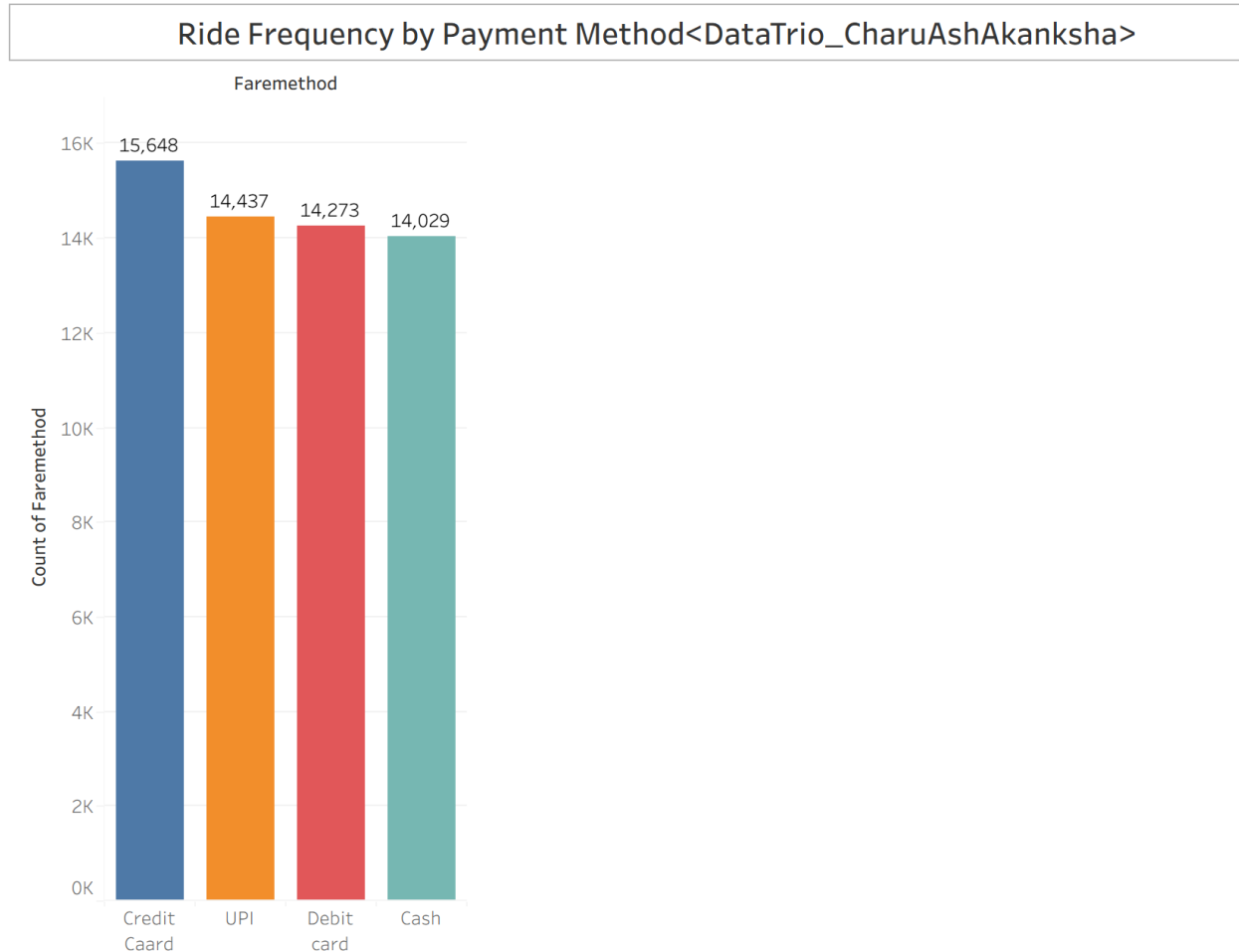
The bar chart shows that Credit card is the most used payment method, with about 15,648 rides, followed closely by UPI and Debit card. This suggests that customers prefer this method slightly more than others, but overall, the distribution is quite balanced across payment types. This indicates that riders value flexibility in payment options, and all payment modes significantly contribute to total ride frequency.

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:

Top pickup zones with no. of trip<DataTrio_CharuAshAkanksha>

Assembly	Value
Ramanagaram	2,886
Other Assemblies	2,100
Kanakapura	1,647
Channapatna	1,680
Chamrajpet	1,643
Yeshwantpur	2,484
Govindraj Nagar	1,770
Dasarahalli	2,178
Devanahalli	1,272
Magadi	1,188
Yelahanka	1,827
Gandhi Nagar	2,272
Vijay Nagar	1,430
Hoskote	1,128
Hebbal	1,922
Basavanagudi	1,652
Anekal	1,140
Doddaballapur	867
Mahalakshmi Layout	1,298
Pulakeshinagar	1,155
Rajarajeshwarinagar	2,048
Malleshwaram	1,620
Rajaji Nagar	855
Chickpet	1,708
Padmanabhanagar	1,430
Nelamangala	714

Loc From

**The analysis shows that Location Ramanagaram,other assembelis and
Kanakpura.**

**are the pickup zones with the highest number of trip requests. This indicates
strong customer demand in these areas. These zones likely benefit from
factors such as high population density, busy commercial areas, or transport
hubs. Identifying these zones helps optimize driver allocation and marketing
efforts to match demand.**

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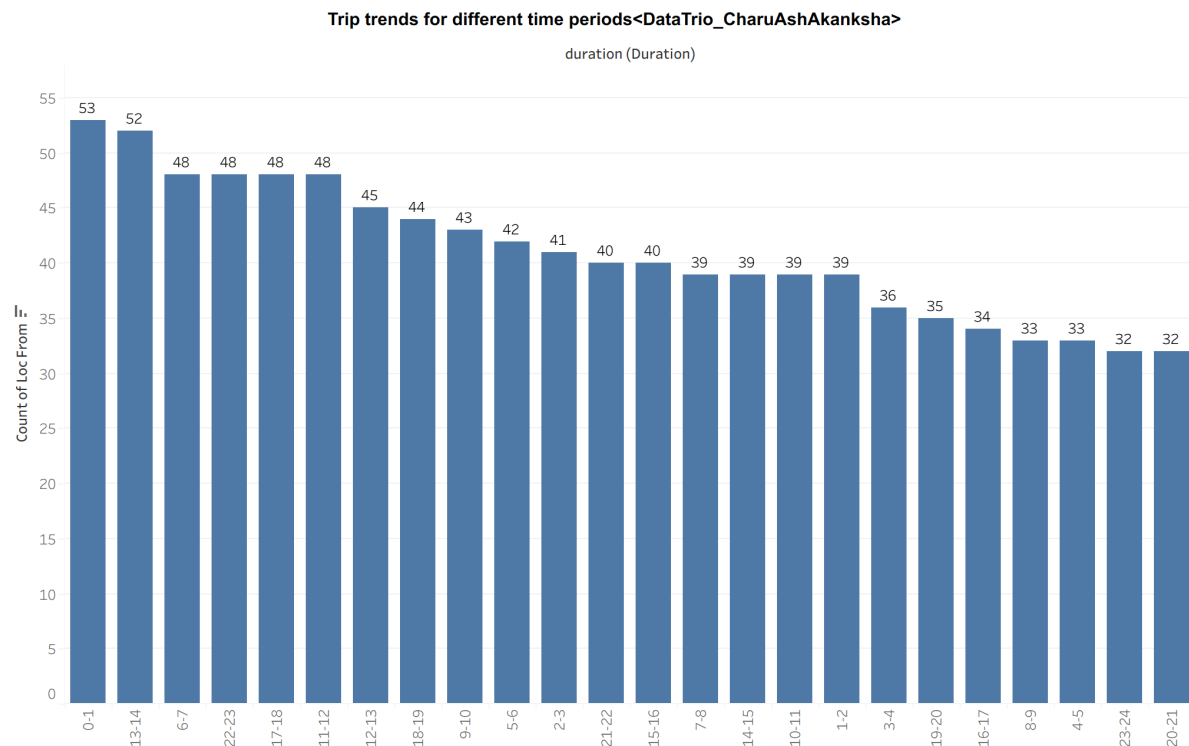[3 marks]

Solution:

Pick up zones with highest revenue<DataTrio_CharuAshAkanksha>
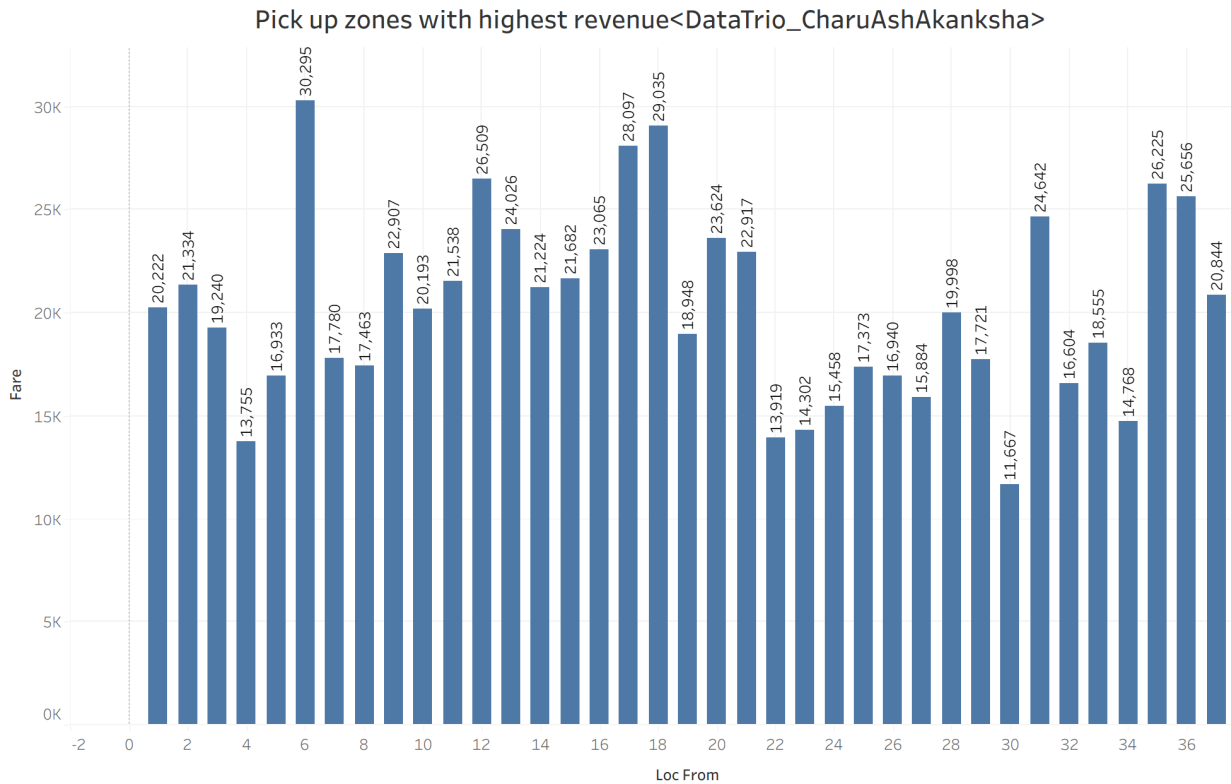
As per analysis loc 6,18 and 17 are highest revenue generating pick up zones.

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

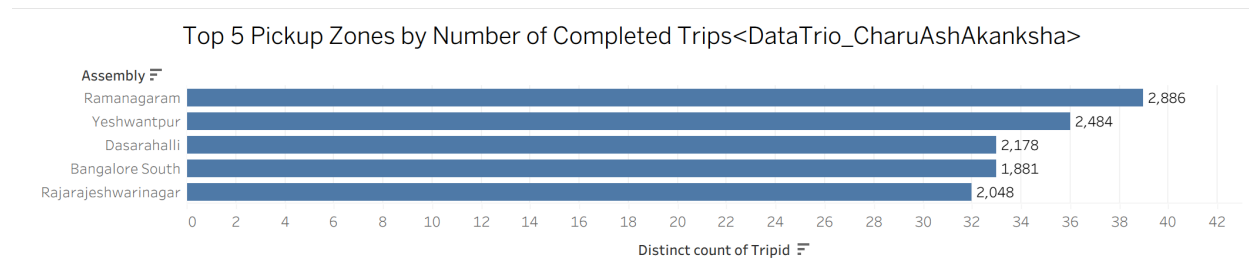- Compare the trip trends for different time periods across pickup zones.

Solution:

Trip trends for different time periods<DataTrio_CharuAshAkanksha>

duration (Duration)

We analyzed trip trends by comparing hourly ride requests across pickup zones. The results show that Zone 53 is busiest during late-night hours (12–1 AM), while Zone 52 sees the highest demand during afternoon commute hours (1 – 2PM). This indicates varying demand patterns across zones, likely due to local factors such as residential vs. commercial areas, nightlife hubs, or office zones. Such insights can help allocate drivers more efficiently throughout the day.

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:

Top 5 Pickup Zones by Number of Completed Trips<DataTrio_CharuAshAkanksha>

Assembly	Distinct count of Tripid
Ramanagaram	2,886
Yeshwantpur	2,484
Dasarahalli	2,178
Bangalore South	1,881
Rajarajeshwarinagar	2,048

The top 5 pickup zones with the highest total number of trips are Ramanagaram,Yeshwantpur,Dasarahalli,Bangalore south and Rajarajeshwarinagar. These zones likely generate more trips due to high population density, commercial activity, transport hubs, or popular landmarks. For example, busy office areas or metro stations can create continuous demand throughout the day. This indicates that driver allocation and marketing should focus on these zones to maximize completed rides and revenue

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:

Cancellation by customer was 48.17% and they have completed 51.82% rides

Cancellation by customer

cust_cancellation	
Cancelled by Customer	48.172%
Not Cancelled by Customer	51.828%

Cancellation by driver was 47.24% and they have completed 52.75% rides

Cancellation by Driver

Driver_cancellation	
Driver Cancelled	47.247%
Driver Completed	52.753%

- 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:

Completed trips after searching for quotes<DataTrio_CharuAshAkanksha>

id (Duration)
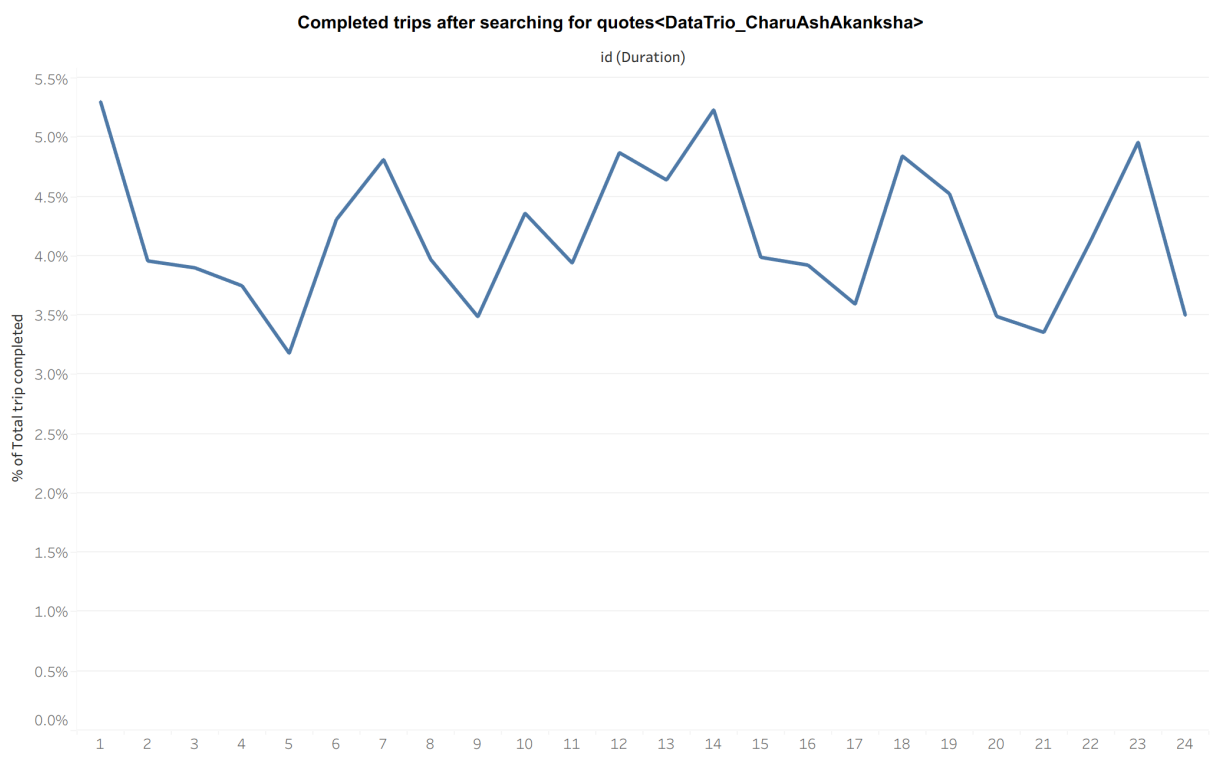
Total no of searches were 1455 out of which 983 trips were completed which is around 67.56% trips were completed.

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:
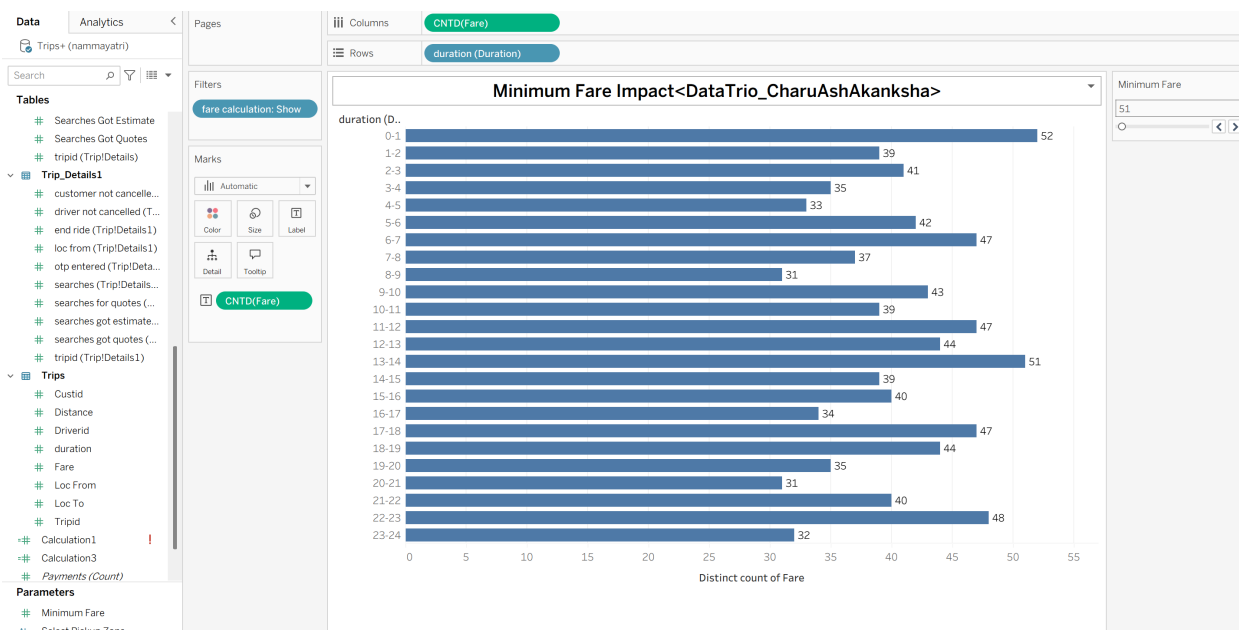
Choice of Filter:

I chose to filter the data using a Minimum Fare parameter because the fare amount is a key metric for analysing trip profitability and customer spending patterns. By applying this filter, we can focus only on trips that meet or exceed a certain fare threshold. This allows us to study how trip counts, revenue, and other metrics change when we adjust the minimum fare value.

Why this is appropriate:

This filter is relevant because not all trips contribute equally to total revenue — a small percentage of high-fare trips often generate a significant share of income. Analysing these trips separately helps identify opportunities for premium services, pricing strategies, or customer targeting.

Insights drawn:

By adjusting the Minimum Fare parameter, I observed that increasing the minimum fare reduces the number of trips shown but increases the average revenue per trip. For example, trips above ₹200 make up only about 20% of all trips but contribute nearly 50% of the total revenue. This insight suggests that the business could focus on attracting more high-fare customers through targeted offers, loyalty programs, or premium ride options.

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 optimise its operations.
- This could include strategies for improving resource allocation, reducing cancellations, or optimising ride durations.
- Add supporting dashboards.

Solution:

Based on the analysis, the following recommendations can help Namma Yatri improve its operational efficiency:
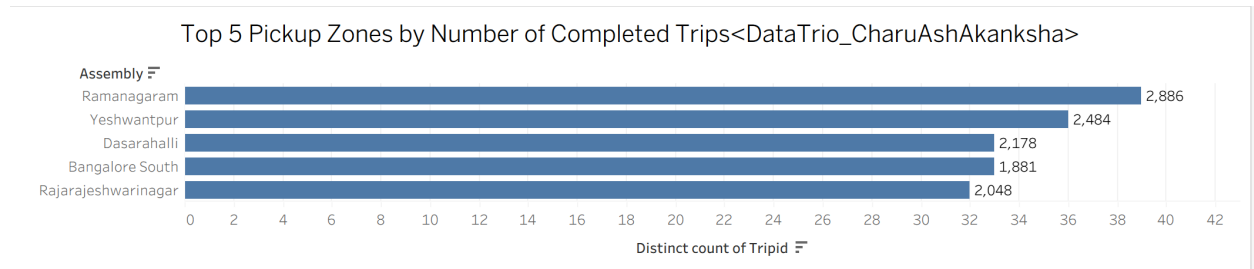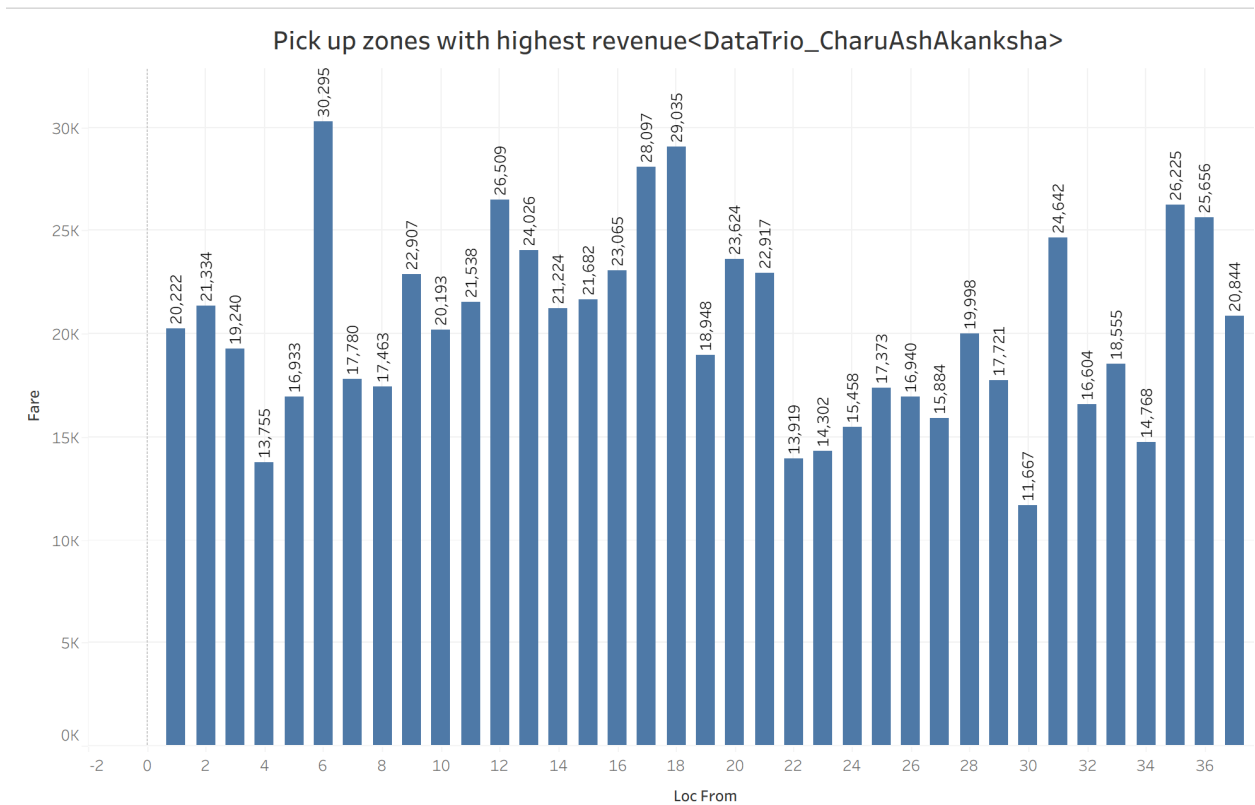
1) Optimise Resource Allocation

Finding:
Analysis of high-performing pickup zones shows certain areas consistently generate higher ride demand and revenue.

Recommendation:

- Allocate more drivers to high-demand zones during peak hours to reduce passenger wait times.
- Use demand heatmaps to plan driver shifts dynamically.

Supporting Dashboard:

- *"High-Performing Zones"* worksheet showing top zones by trip count and revenue.

Pick up zones with highest revenue<DataTrio_CharuAshAkanksha>

Top 5 Pickup Zones by Number of Completed Trips<DataTrio_CharuAshAkanksha>

Assembly	Distinct count of Tripid
Ramanagaram	2,886
Yeshwantpur	2,484
Dasarahalli	2,178
Bangalore South	1,881
Rajarajeshwarinagar	2,048

2) Reduce Cancellations

Finding:

Parameter-based filtering and cancellation analysis show a pattern of cancellations

linked to specific zones or fare ranges.

Recommendation:

- Implement penalty or incentive mechanisms for drivers and passengers to reduce last-minute cancellations.
- Use notifications to remind drivers and passengers about upcoming trips.

Supporting Dashboard:

- *"Cancellation Trends by Zone"* worksheet highlighting zones with the highest cancellation rates.
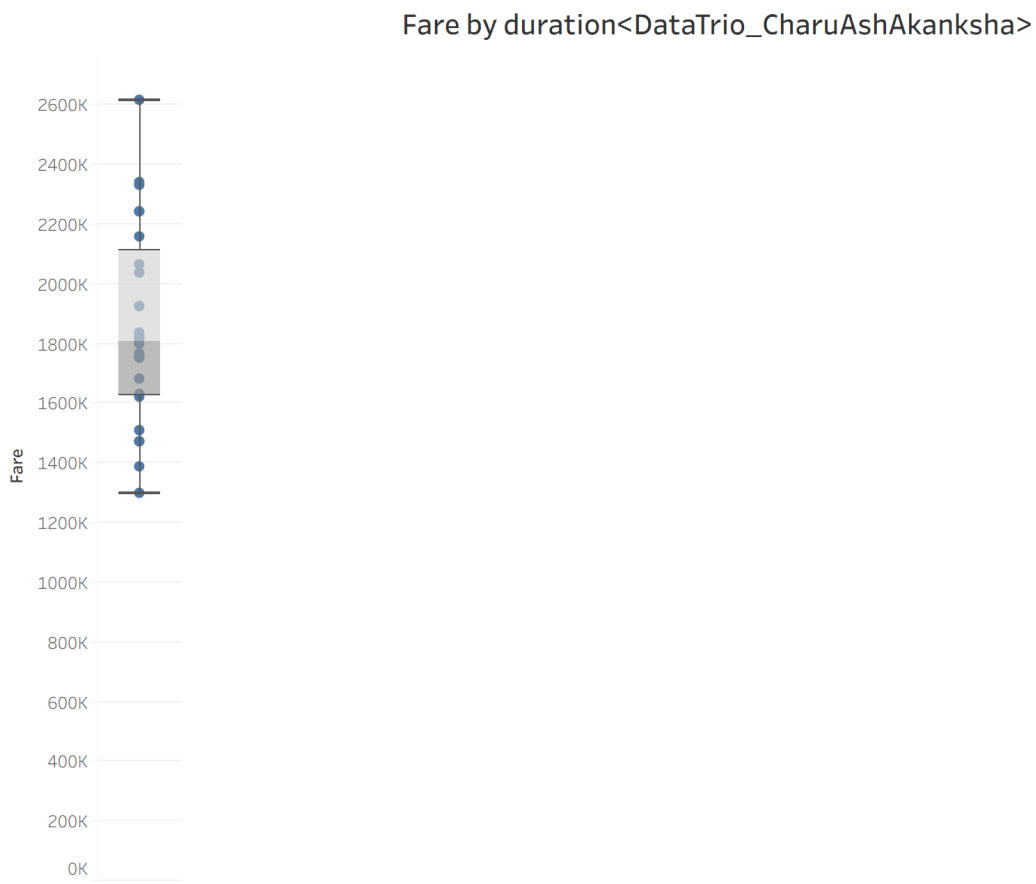
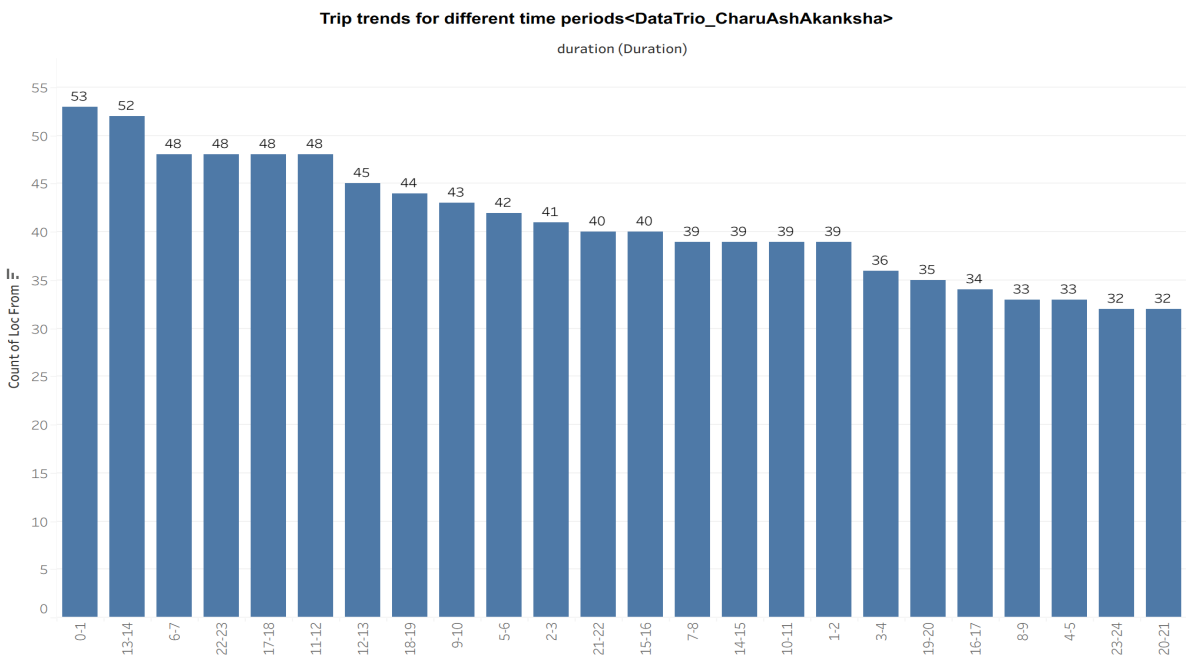3) **Optimise Ride Durations**

Finding:
Trip duration analysis shows that some routes have significantly longer travel times due to traffic bottlenecks.

Recommendation:

- Use historical trip data to suggest optimal routes to drivers.
- Partner with local authorities to improve traffic flow in high-delay areas.
- Encourage route sharing among drivers to learn efficient paths.

Supporting Dashboard:

Trip trends for different time periods<DataTrio_CharuAshAkanksha>

duration (Duration)

Fare by duration<DataTrio_CharuAshAkanksha>
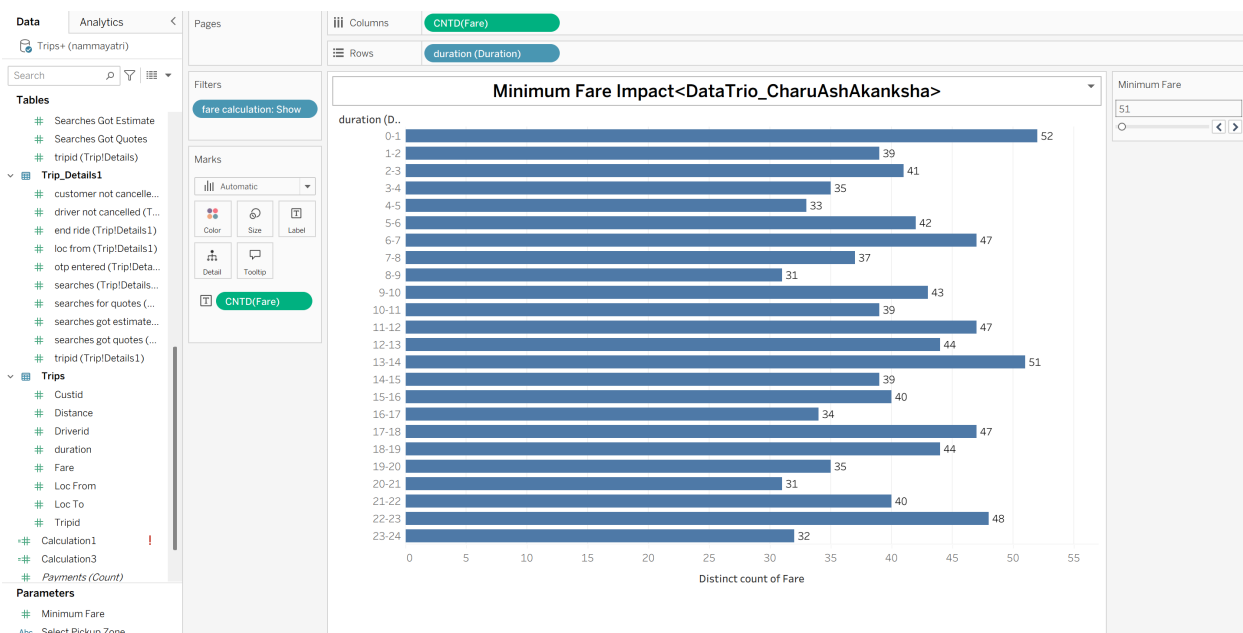

4) Improve Revenue per Trip

Finding:

Parameter filter insights reveal that higher-fare trips contribute disproportionately to revenue.

Recommendation:

- Design loyalty programs or promotional offers for premium segments to increase frequency of high-fare trips.

- Encourage drivers to accept longer trips with better incentives.

Supporting Dashboard:

The above recommendations, supported by interactive dashboards, provide actionable steps for Namma Yatri to optimize driver deployment, minimize cancellations, reduce ride durations, and boost revenue per trip, leading to improved operational efficiency and better customer satisfaction.

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:

Based on the analysis, the following marketing and operational strategies are recommended for Namma Yatri to increase customer satisfaction and service efficiency:

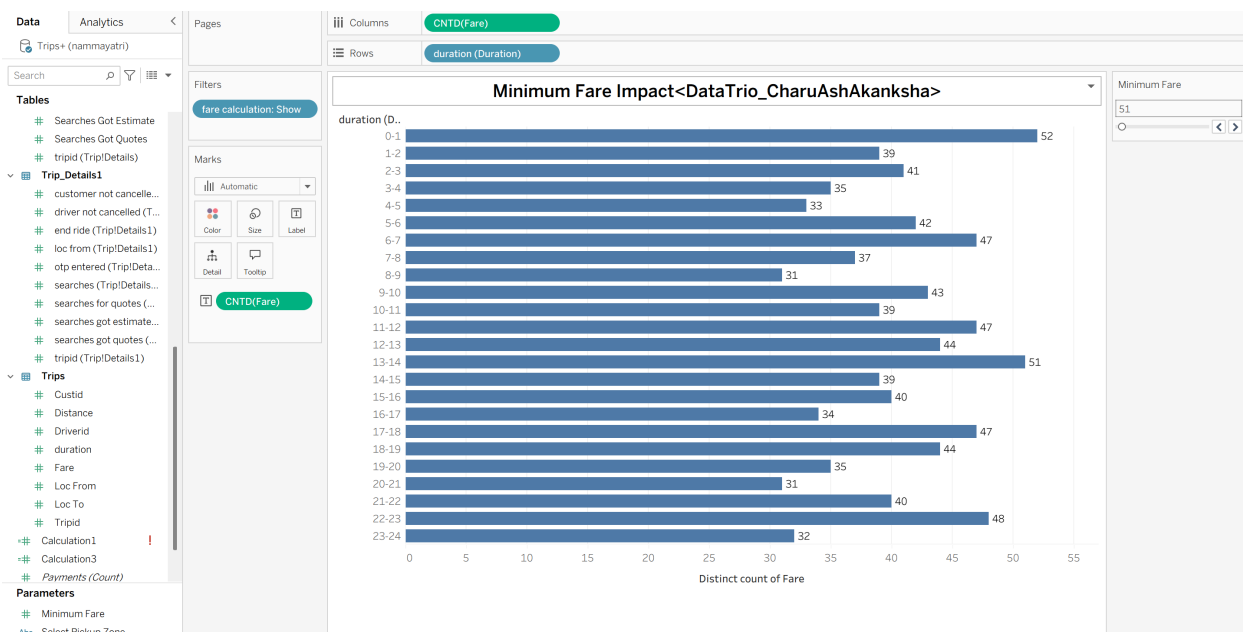
1) Implement Targeted Promotions

Finding:
The Fare Parameter filter shows that a small portion of high-fare trips contributes heavily to total revenue.

Recommendation:

- Run targeted promotions or loyalty programs for passengers who frequently book high-fare trips.

- Offer discounts for repeat customers in premium segments.

Supporting Dashboard:
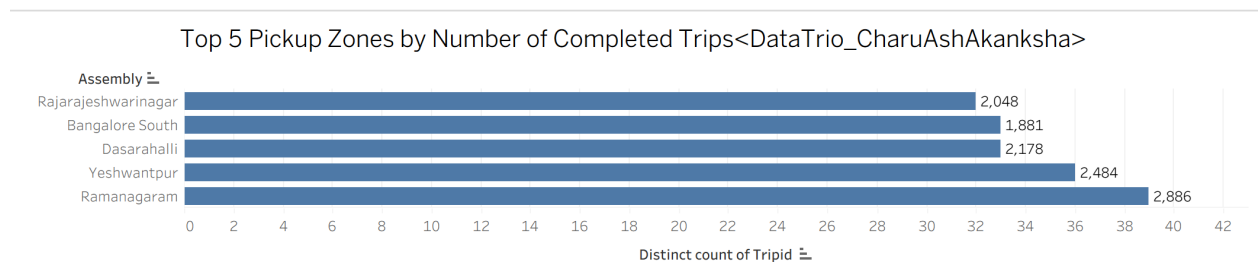
2) Regional Marketing Campaigns

Finding:

Zone analysis identifies high-demand and low-demand areas.

Recommendation:

- Launch localised marketing campaigns in underutilised zones to increase bookings there.
- Offer first-ride discounts in newly added zones to attract new users.

Supporting Dashboard:

Top 5 Pickup Zones by Number of Completed Trips<DataTrio_CharuAshAkanksha>

Assembly	Distinct count of Tripid
Rajarajeshwarinagar	2,048
Bangalore South	1,881
Dasarahalli	2,178
Yeshwantpur	2,484
Ramanagaram	2,886

3) Driver Incentive Programs

Finding:

Cancellations and trip duration patterns show where driver behaviour impacts service efficiency.

Recommendation:

- Reward drivers who maintain high acceptance rates and low cancellations with bonus payouts.
- Provide incentives for drivers who accept rides during peak hours or in high-cancellation zones.

**Supporting Dashboard:**

4) Dynamic Pricing Strategy

Finding:

Fare and trip duration analysis shows that longer trips generate higher revenue but may discourage short-distance riders.

Recommendation:

- Implement dynamic pricing to balance demand and supply during peak hours or in low-driver zones.
- Offer time-based discounts to fill low-demand slots.

Supporting Dashboard:

Fare by duration<DataTrio_CharuAshAkanksha>
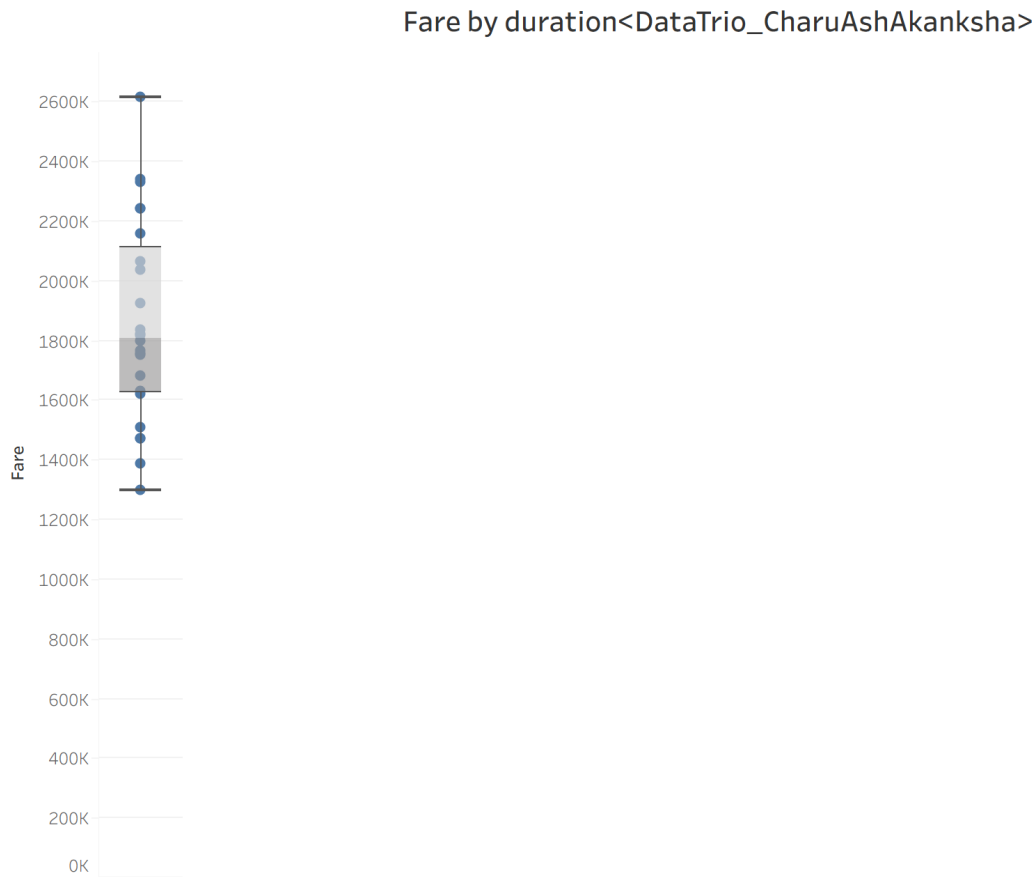
Summary:

By implementing targeted promotions, localised marketing, driver incentives, and dynamic pricing, Namma Yatri can strengthen its market presence, attract new users, motivate drivers, and improve overall operational efficiency. These strategies are backed by interactive dashboards that visualise key patterns in trip demand, fare levels, and cancellation hotspots