EXPLORATORY DATA ANALYSIS

NAMMA YATRI DATA SET

Merlin Treesa Benny

**DATA ANALYSIS REPORT ON NAMMA YATRI**

**Namma Yatri** is a ride-hailing app in India, designed to connect passengers directly with auto-rickshaw drivers without intermediaries. It offers affordable pricing, transparent fares, and reliable service by eliminating platform commissions.

**INTRODUCTION**

This report analyzes the **Namma Yatri** data set sourced from Kaggle, to uncover city-wise trends in ride frequency, pricing, and user behaviour. Utilizing various statistical techniques, visualization tools, exploring key metrics and patterns, it aims to provide actionable insights to enhance operational efficiency and improve the user experience across cities.

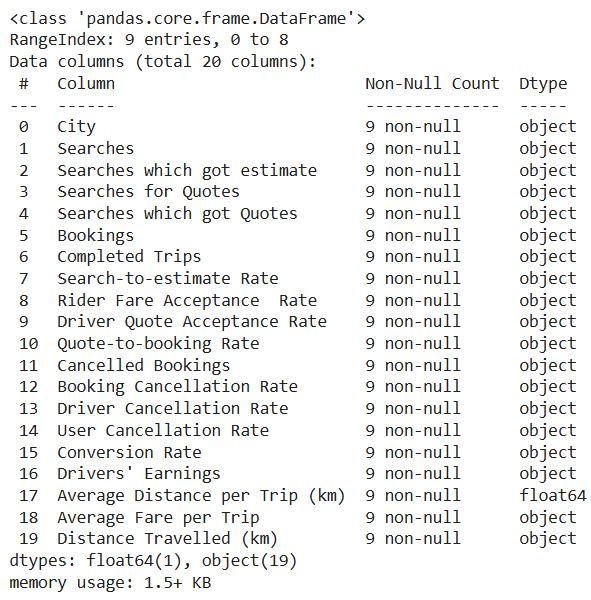
**GOALS**

1. Evaluate Platform Efficiency
2. Analyze Booking and Cancellation Trend
3. Assess Financial Performance
4. City-Wise Comparisons
5. Optimize Trip Efficiency

**DATA DESCRIPTION**

The data set contains 9 rows and 20 columns and the columns are:

* **City** – Cities of India (8 cities)
* **Searches** – Number of Searches
* **Searches which got estimate** - Number of searches that received fare estimates.
* **Searches which got Quotes**: Number of searches that received quotes from drivers.
* **Bookings**: Number of ride bookings made.
* **Completed Trips**: Number of trips that were successfully completed.
* **Search-to-estimate Rate**: Percentage of searches that resulted in fare estimates.
* **Rider Fare Acceptance Rate**: Percentage of fare estimates accepted by riders.
* **Driver Quote Acceptance Rate**: Percentage of quotes accepted by drivers.
* **Quote-to-booking Rate**: Percentage of quotes that led to bookings.
* **Cancelled Bookings**: Number of bookings that were cancelled.
* **Booking Cancellation Rate**: Percentage of bookings that were cancelled.
* **Driver Cancellation Rate**: Percentage of cancellations made by drivers.
* **User Cancellation Rate**: Percentage of cancellations made by users.
* **Conversion Rate**: Percentage of searches that converted into completed trips.
* **Drivers' Earnings**: Total earnings of drivers (in Rs).
* **Average Distance per Trip (km)**: Average distance covered per trip (in KM).
* **Average Fare per Trip**: Average fare earned per trip (in Rs).
* **Distance Travelled (km)**: Total distance travelled across all trips (in KM).



**DATA CLEANING AND PREPARATION**

1. Handling Missing Values:

* Verified that the dateset contains no null or missing values using methods like .isnull() and .dropna() or imputed missing data if necessary.

1. Checking for Duplicates:

* Checked for duplicate rows using .duplicated() ,this process confirmed that there was 0 duplicates ,this process is important as it reduces data redundancy and ensure accurate analysis.

1. Data Type Validation:

* Reviewed and confirmed that all columns have appropriate data types, converted object data type (Searches, Drivers' Earnings, Average Fare per Trip**…)** to numerical data type

1. Cleaning Text and Numerical Data:

* Removed unwanted characters such as ' , ', ‘%’ and "Rs" from relevant columns to standardize numerical data and make it analysis-ready.

1. Renaming Columns:

* Standardized column names for better readability and consistency, ensuring they are descriptive and follow a consistent naming convention.

1. Handling Outliers:

* Checked for extreme or inconsistent values in numerical columns. Used boxplot to check for outlier.

1. Feature Engineering:

* Created new columns or metrics (Earnings per Km, Earnings per trip) based on the existing data.

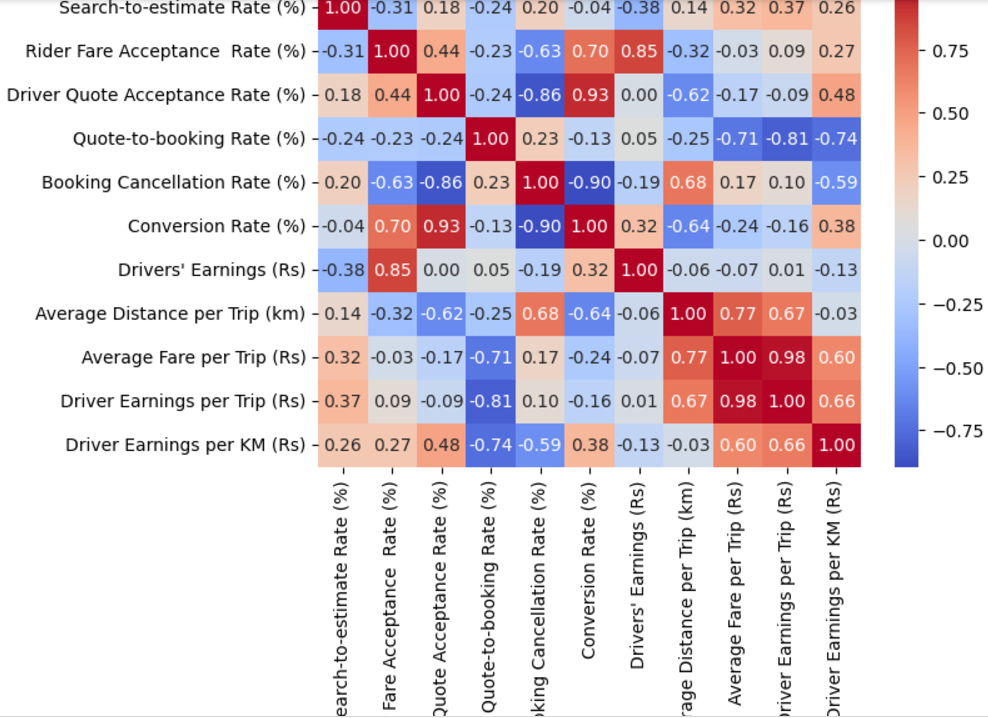
1. Data Filtering:

* Removed irrelevant rows or columns (e.g., Removed all\_total column) to focus on meaningful data.

**EXPLORATORY DATA ANALYSIS**

**Statistical Analysis of data set**

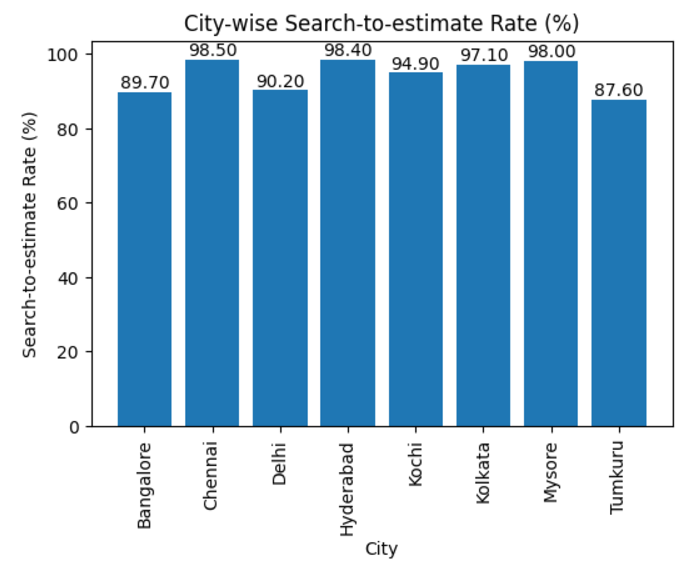
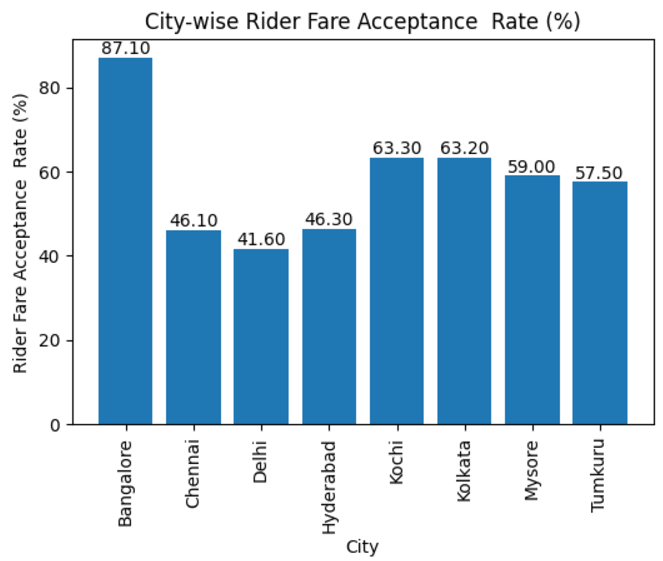
Correlation Heatmap for Namma Yatri Metrics

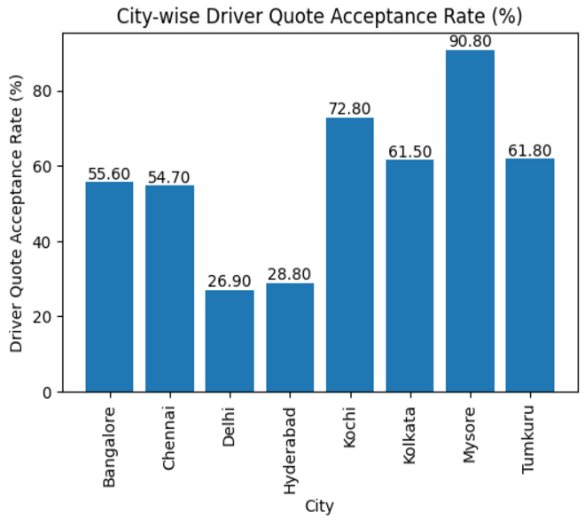
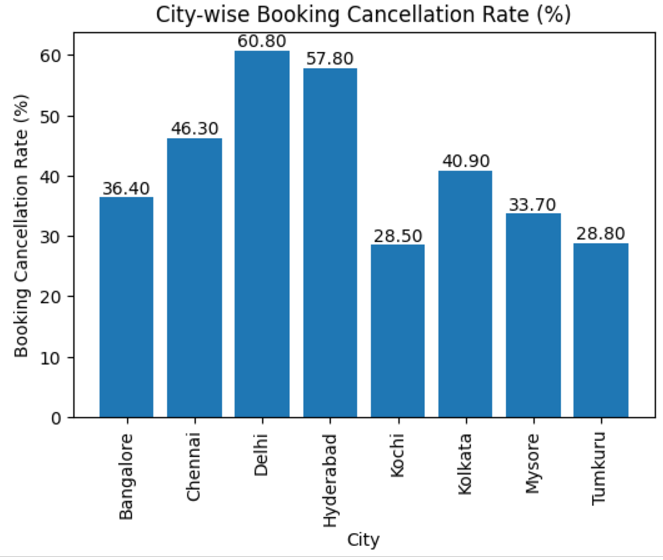


Insights: **Conversion Rate (%)** strongly correlates with **Driver Quote Acceptance Rate (%)** (0.93) and **Rider Fare Acceptance Rate (%)** (0.70), indicating that improving quote acceptance by both drivers and riders boosts conversions. **Booking Cancellation Rate (%)** is negatively correlated with **Driver Quote Acceptance Rate (%)** (-0.86) and **Conversion Rate (%)** (-0.90), suggesting that reducing cancellations significantly improves success rates.

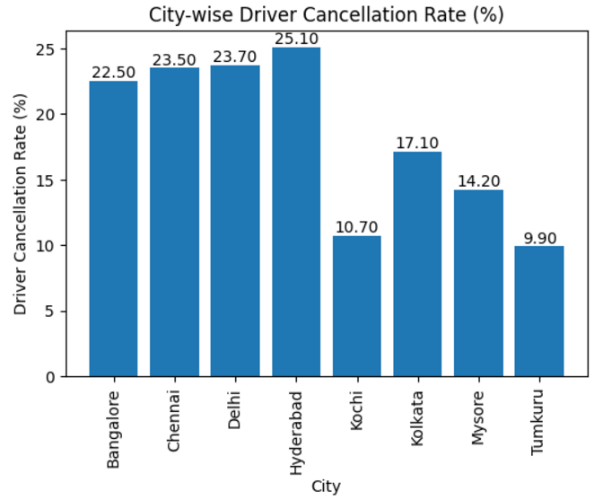
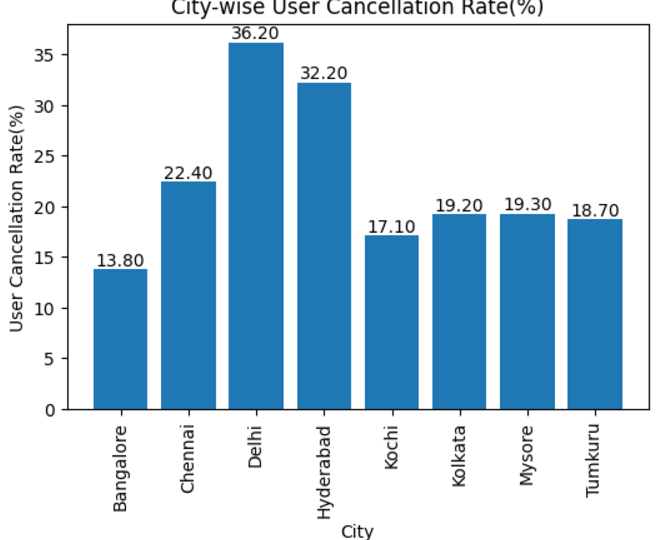
**Visualising city-wise distribution of**

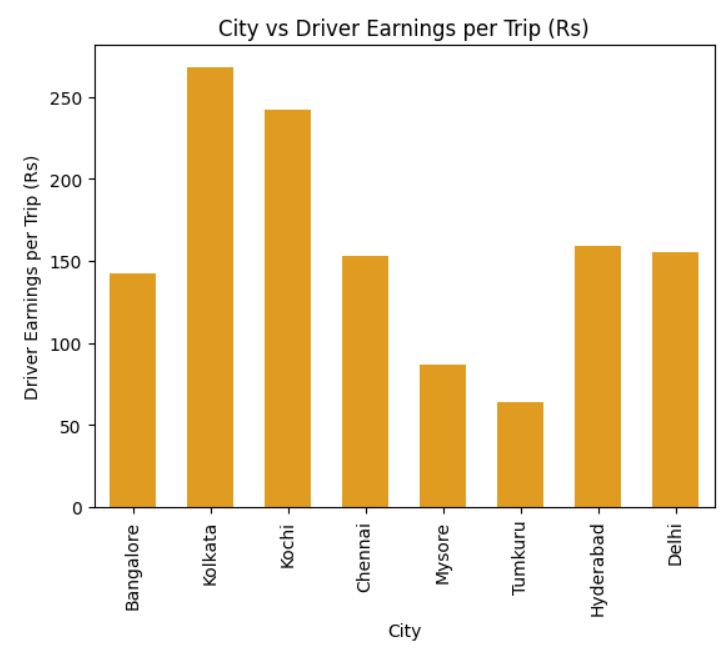
Search-to-estimate Rate (%) Rider Fare Acceptance Rate (%)

Driver Quote Acceptance Rate (%) Booking Cancellation Rate(%) 

Driver Cancellation Rate(%) User Cancellation Rate(%)

**Visualising Drivers' Earnings Per Trip**

Number of drivers according to city as of (2024)

Bangalore- 23,109

Kolkata- 3,888

Kochi - 1,656

Chennai- 2,419

Mysore- 361

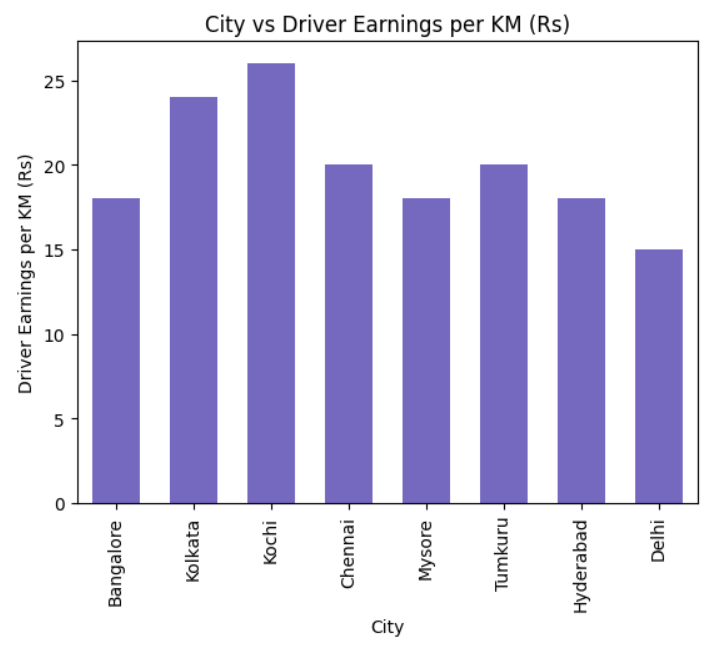
Tumkuru- 139

Hyderabad- 2,944

Insights:

* **Kolkata** and **Kochi**, despite having fewer drivers (3,888 and 1,656 respectively), show the highest earnings per trip, indicating potentially higher demand or longer trips in these cities.
* **Bangalore** has the most drivers (23,109) but lower earnings per trip compared to Kolkata and Kochi, suggesting a higher supply of drivers relative to demand.
* **Tumkuru** and **Mysore**, with the fewest drivers (139 and 361), also have the lowest earnings, potentially due to limited trip demand or shorter distances.

**Visualization of Drivers' Earnings Per KM Travelled**



Average Distance per Trip (km)

Bangalore- 7.54

Kolkata- 9.49

Kochi- 9.72

Chennai- 8.51

Mysore- 3.62

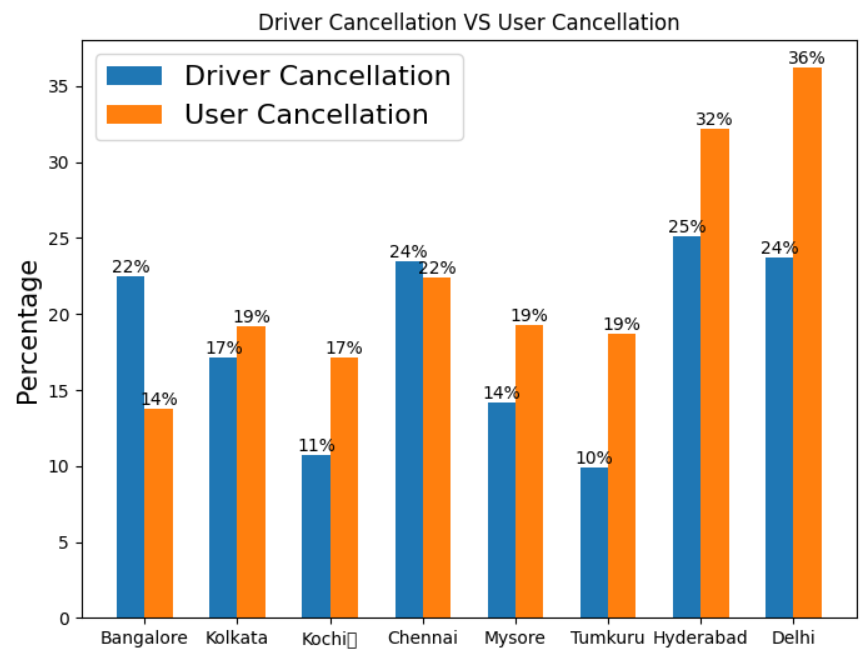
Tumkuru- 9.05

Hyderabad- 12.72

Insights

* **Hyderabad** has the longest trip distances (12.72 km), yet its earnings per km are moderate, suggesting a focus on higher total trip distances but not premium rates.
* **Mysore** has the lowest earnings per km and the shortest average trip distance (3.62 km), indicating fewer profitable opportunities.

**Comparison of Driver Cancellation VS User Cancellation Across Cities**

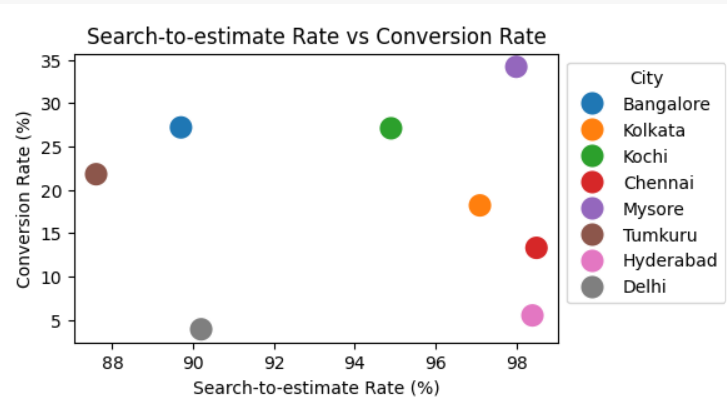


1,776,216 -Bangalore users, over 40 lakh users- Delhi,

over 14 lakh users - Kolkata, 65,000 users -Hyderabad

Insight: Delhi has the highest driver cancellation rate ,as well as user cancellation rate

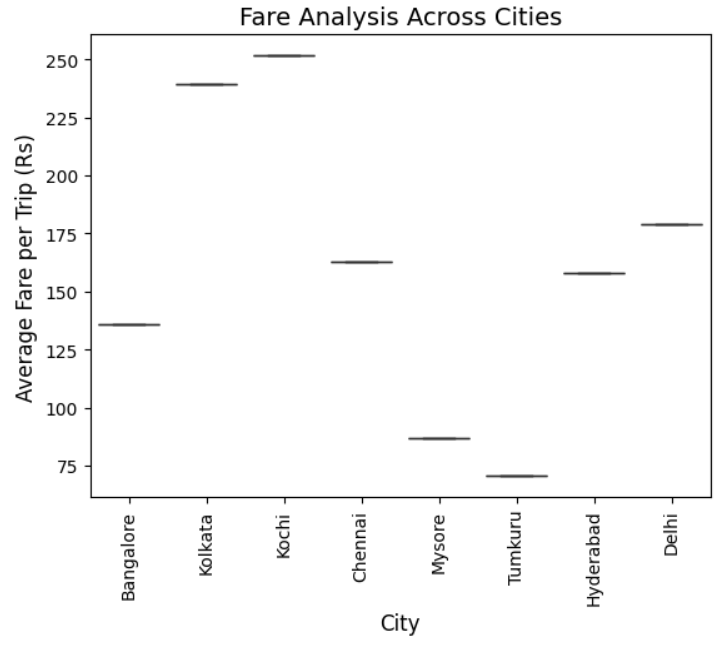
**Scatter plot of Search to estimate with conversion rate**



Insights:

* Cities with a higher Search-to-estimate Rate tend to have a higher Conversion Rate, but there are exceptions like Hyderabad and Delhi.
* Cities like Kochi and Mysore have a good balance between Search-to-estimate Rate and Conversion Rate.

**Boxplot of Fare analysis across cities**



Insights:There is a significant variation in the average fare per trip across the different cities.

* Kochi and Bangalore have the highest average fares, while Tumkuru and Mysore have the lowest.

**CONCLUSION**

The exploratory analysis of the Namma Yatri dataset has provided valuable insights.

Key findings from the analysis include:

1. **Evaluate Platform Efficiency:**
   * Strong correlations between Conversion Rate and both Driver and Rider Quote Acceptance Rates (0.93 and 0.70, respectively) highlight the importance of enhancing quote acceptance for platform success.
   * Cities with a higher Search-to-Estimate Rate generally exhibit higher Conversion Rates, though exceptions like Hyderabad and Delhi suggest room for optimization.
2. **Analyze Booking and Cancellation Trends:**
   * High cancellation rates, particularly in Delhi, significantly hinder efficiency. A negative correlation between cancellations and Conversion Rate (-0.90) emphasizes the need to reduce both driver and user cancellations to improve booking success.
3. **Assess Financial Performance:**
   * Kolkata and Kochi demonstrate the highest earnings per trip, driven by potentially higher demand or longer trips.
   * Bangalore, despite its large driver base, faces lower earnings per trip, indicating a supply-demand imbalance.
   * Tumkuru and Mysore, with fewer drivers, report lower earnings, likely due to limited demand or shorter trip distances.
4. **City-Wise Comparisons:**
   * Cities like Kochi and Mysore maintain a good balance between Search-to-Estimate Rate and Conversion Rate, serving as benchmarks for operational efficiency.
   * Significant variations in average fares across cities reveal differing dynamics, with Kochi and Bangalore at the higher end and Tumkuru and Mysore at the lower end.
5. **Optimize Trip Efficiency:**
   * Longer average trip distances in cities like Hyderabad and Kochi suggest opportunities to streamline trip lengths while maintaining driver profitability.
   * Cities with low average trip distances, such as Mysore, may focus on increasing trip demand to enhance driver earnings.

These findings provide actionable insights to improve platform performance, optimize operations, and ensure financial sustainability across cities.