BANGALORE RAPIDO RIDE SERVICES DATASET

Dataset Overview

This dataset contains comprehensive records of ride services offered by Rapido in Bangalore over a three month period. It includes various service types such as bike, bike lite, cab economy, auto, and parcel services. Each record provides extensive details about the ride, including the journey from a source to a destination, duration, distance, charges, and payment methods. This dataset is ideal for analyzing transportation patterns, service utilization, and ride economics in Bangalore.

Aim

The aim of this study is to analyze the operational and customer trends in Rapido's ride-hailing services in Bangalore, focusing on ride distribution, fare structure, ride durations, cancellation patterns, and payment preferences. This analysis seeks to identify factors influencing service performance, customer behavior, and potential areas for optimization.

Objective

The study aims to analyse the fare structure and identify trends in total fares across various service types. It seeks to investigate ride durations and the factors influencing travel time while exploring customer preferences for payment methods and their impact on ride completion. Additionally, the study aims to uncover correlations between distance, fare, duration, and ride status to provide deeper insights into service performance.

Dataset Structure

The dataset is organized in a tabular format with 50,000 rows and 13 columns. Each row represents a unique ride, and the columns provide detailed information about each ride.

services	date	time	ride_status	source	destination	duration	ride_id	distance	ride_charge	misc_charge	total_fare	Payment_method
cab economy	2024-07-15	08:30:40.542646	completed	Balagere Harbor	Harohalli Nagar	39	RD3161218751875354	27.21	764.83	31.51	796.34	Amazon Pay
auto	2024-07-05	23:36:51.542646	completed	Basavanagudi 3rd Block	Bikasipura 1st Stage	89	RD8171514284594096	34.03	314.83	49.52	364.35	Paytm
auto	2024-07-23	11:05:37.542646	cancelled	Babusapalya Cove	Kothaguda Terrace	25	RD9376481122237926	20.24				
cab economy	2024-06-24	08:45:10.542646	completed	Mahadevapura Mews	Kanakapura Arc	89	RD3676889143182765	31.17	484.73	15.84	500.57	QR scan
cab economy	2024-07-15	00:26:44.542646	completed	Ganganagar Cove	Basaveshwaranagar Colony	95	RD6639410275948084	27.21	663.50	14.13	677.63	Amazon Pay

Data Cleaning and Preparation

- Missing Value Handling

Null values in the dataset appear when the ride status is marked as 'cancelled.' Since these values are connected to the ride status and not due to errors or missing data, removing them would result in losing valuable context. Instead of dropping them, the null values are replaced with NaN using fillna( ) method.

- Formatting Date and Time

Converting date and time columns into the correct format is crucial in data preprocessing. It ensures that time-related data is properly handled, especially for time-series analysis or models needing date/time features.

- Feature Engineering

Added columns for “Day," “Month," and “Distance Category”.

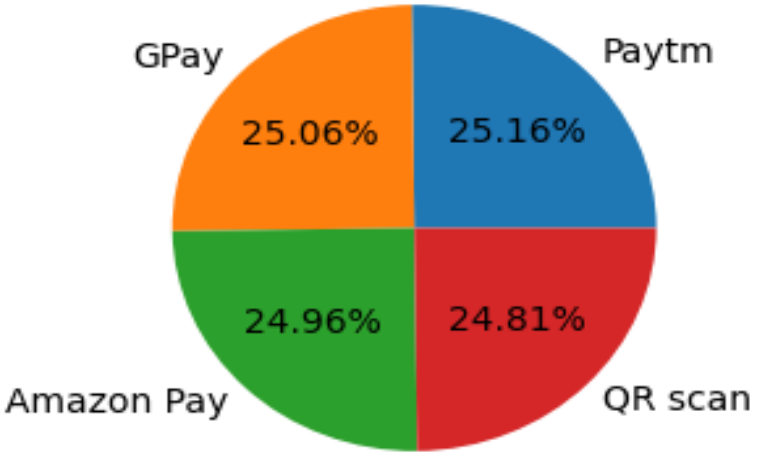
- Day: Extracts the day from a date/time column.
- Month: Extracts the month from a date/time column.
- Day Type: Categorizing Days as "Weekday" or "Weekend“.
- Distance Category: A categorization of distance values, which could be based on ranges like "Short," "Medium," or "Long."

These columns can help to provide more specific insights or make your data more accessible for analysis.

Analysis

Distribution of different Payment Methods

Paytm is the most popular payment method, followed closely by GPay, Amazon Pay, and QR scans, reflecting a strong customer preference for digital payment options. Paytm may have a user-friendly interface or offer special deals that keep users coming back.

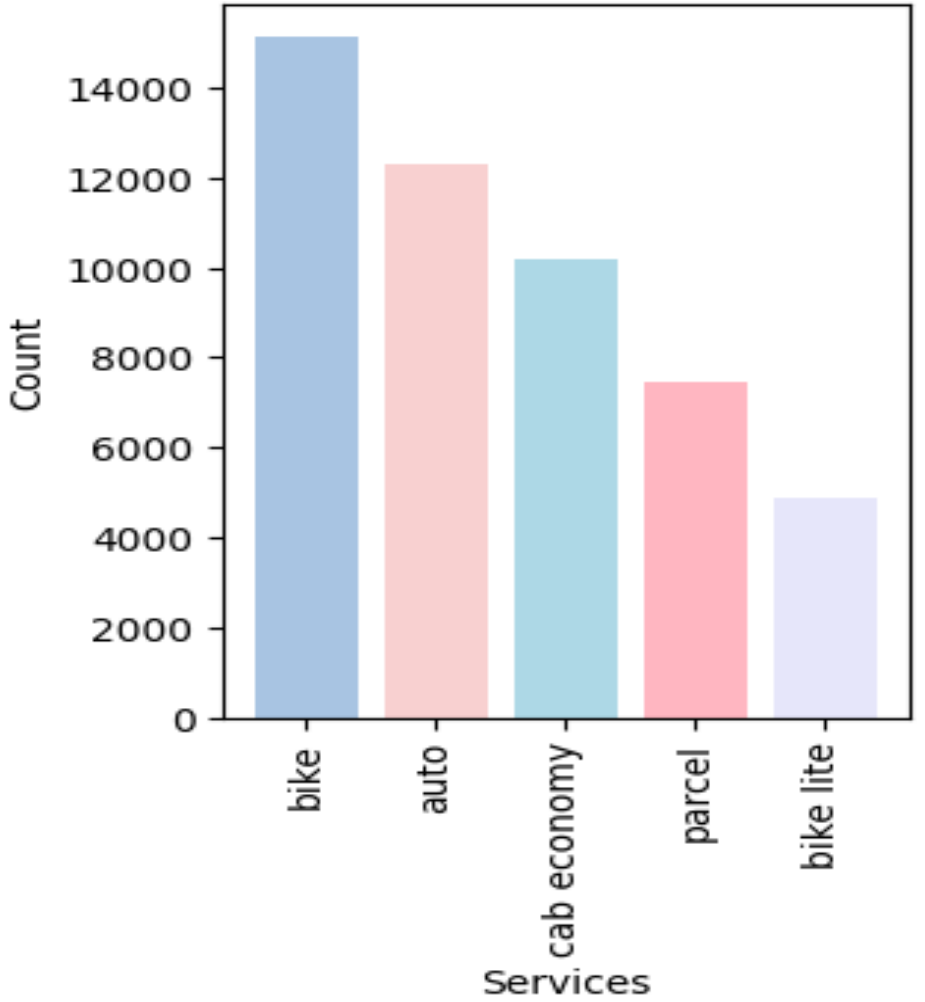


payment_method	Count
Paytm	11315
GPay	11268
Amazon Pay	11225
QR scan	11156

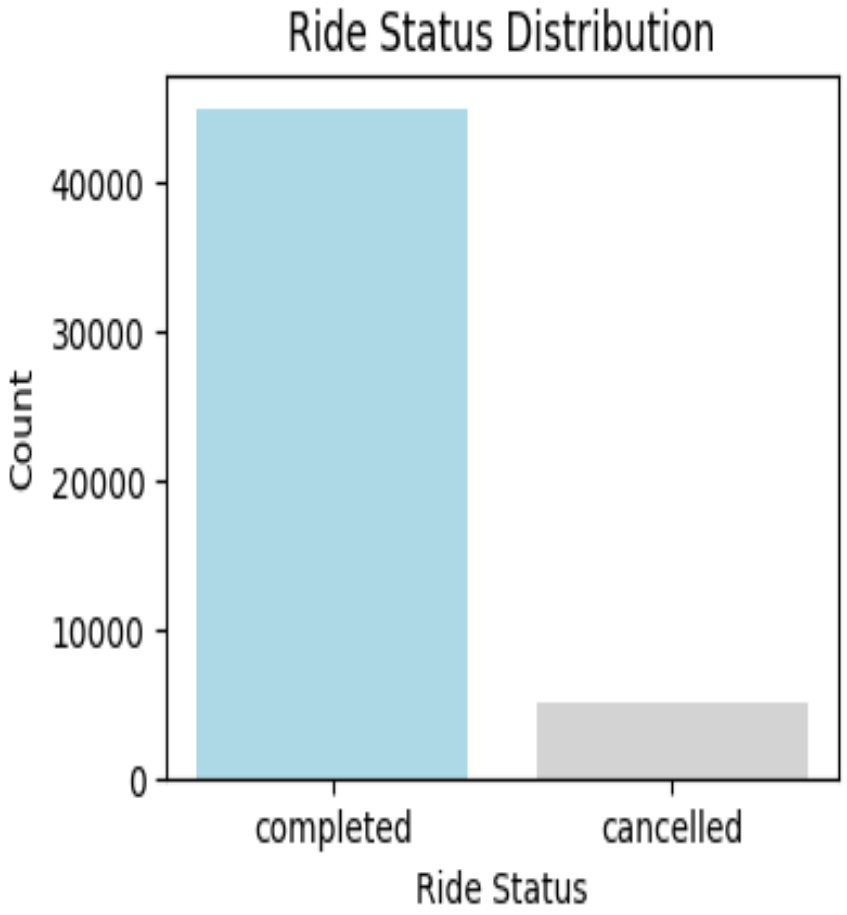
Service Distribution Visualization

Bike and Auto services are the most popular, likely because they're affordable and convenient for short to medium trips, while Bike Lite is less used, possibly due to limited availability or lower demand. People might prefer Auto services for longer trips as they offer more comfort and space.

Services	count
Bike	15128
Auto	12327
Cab Economy	10202
Parcel	7459
Bike Lite	4884



Visualization of Rides by Status



The majority of rides are completed, while a smaller portion is canceled. These cancellations could be due to ride availability or other factors that may need improvement.

ride_status	Count
completed	44964
cancelled	5036

Top 5 most common Source and Destination routes

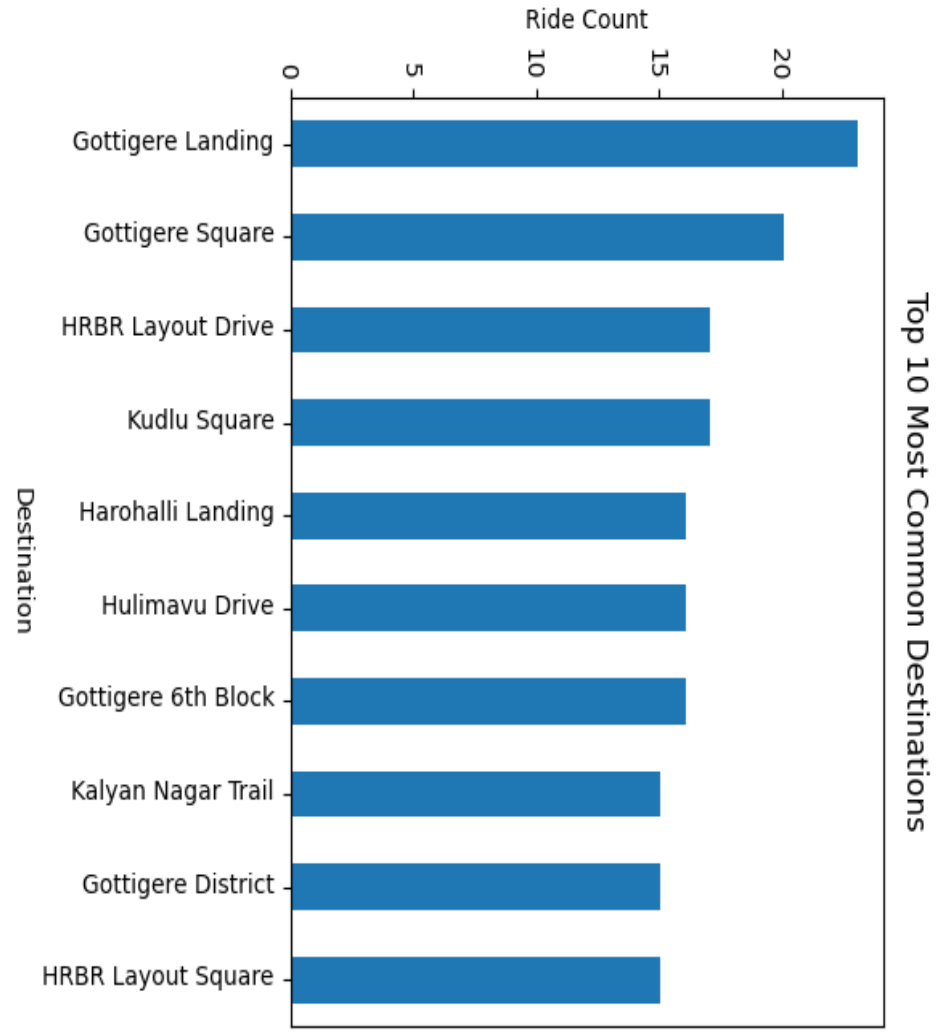
```
df.groupby(['source',
'destination']).size().nlargest().
reset_index()[['source',
'destination']]
```

This is the most common source and destination routes appear frequently. These popular routes could signify high-traffic areas or regions where services are in high demand, which could be targeted for better availability.

	Source	Destination
0	Anekal Woods	Yelahanka Complex
1	Basaveshwaranagar Place	Kalena Agrahara Fork
2	HSR Layout Crescent	Tavarekere View
3	Jayanagar Cut	Yelahanka Landing
4	Kundalahalli 6th Stage	Gottigere 5th Stage

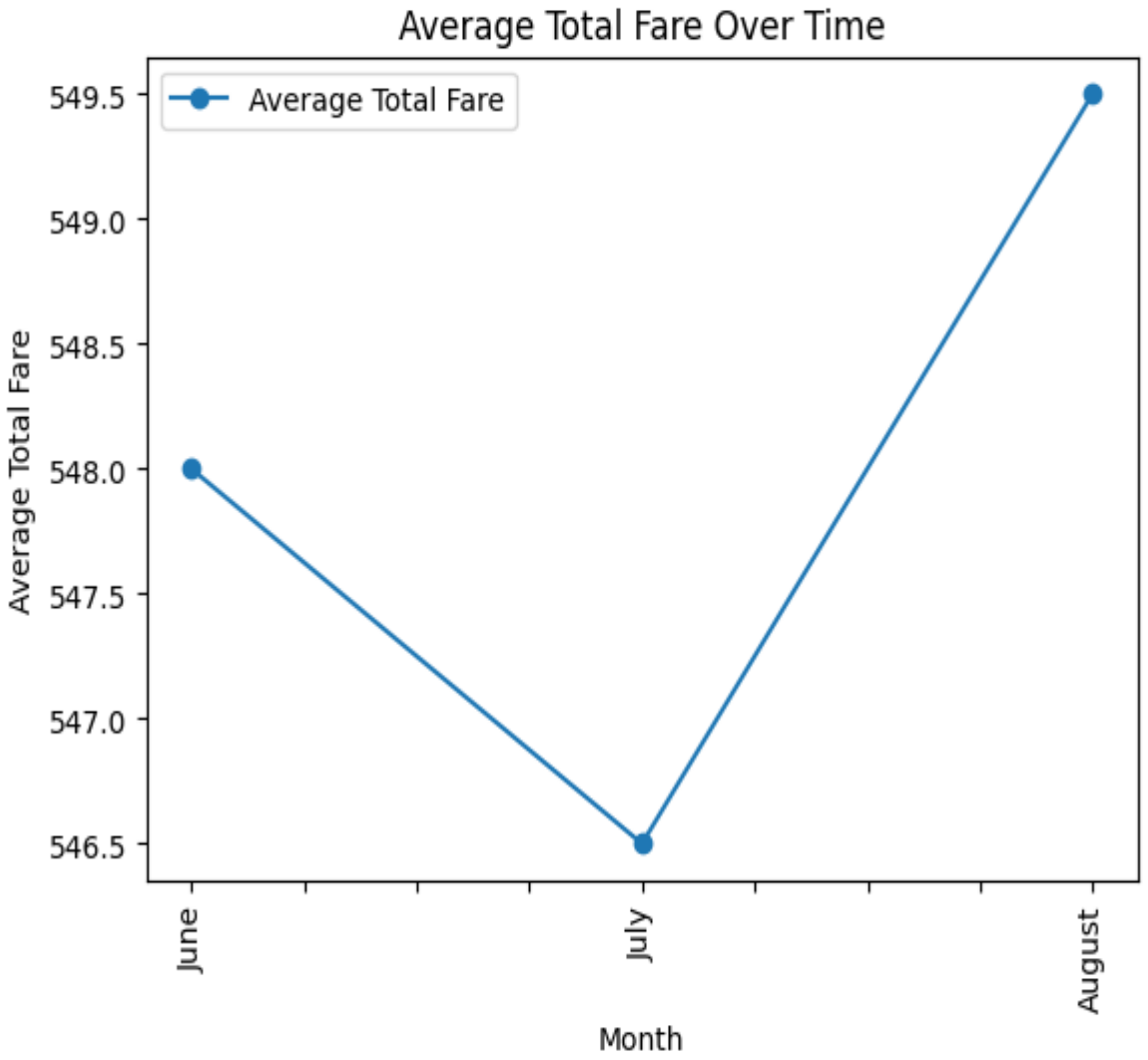
Visualization of top 10 most common Destinations

```
top_destinations =
df['destination'].value_counts().nlargest(10)

top_destinations.plot(kind='bar')

plt.title('Top 10 Most Common Destinations',pad=10)

plt.xlabel('Destination')

plt.ylabel('Ride Count')

plt.xticks(rotation=90)

plt.show()
```

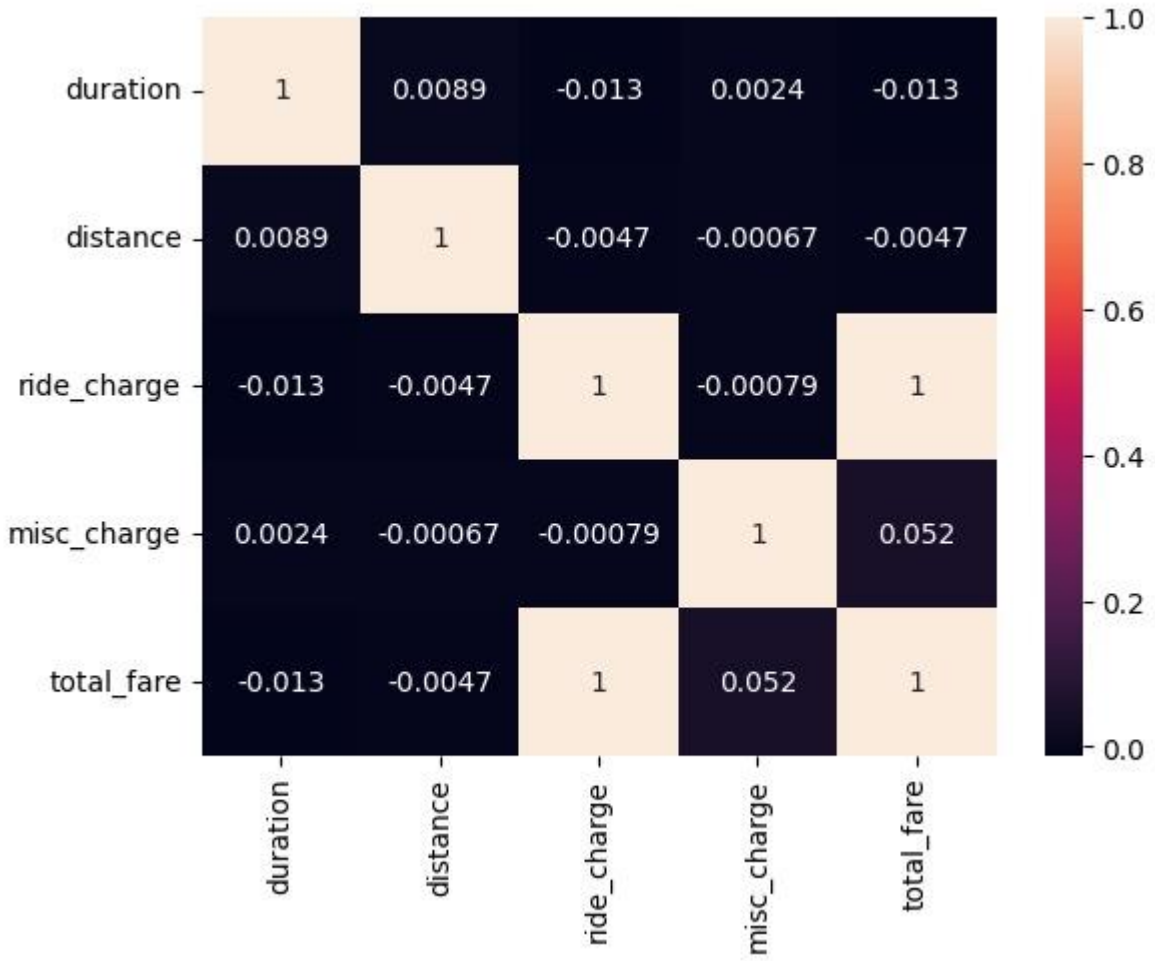


Average Fare Trend Over Time

Fare trends tend to change throughout the months, with noticeable peaks and valleys. These variations are often influenced by factors like seasonal demand, holidays, special events, or shifts in the economy. By analysing these trends, businesses can better understand when to adjust prices, helping to optimize revenue and meet customer demand more effectively.



Correlation Matrix



- Positive Correlations:

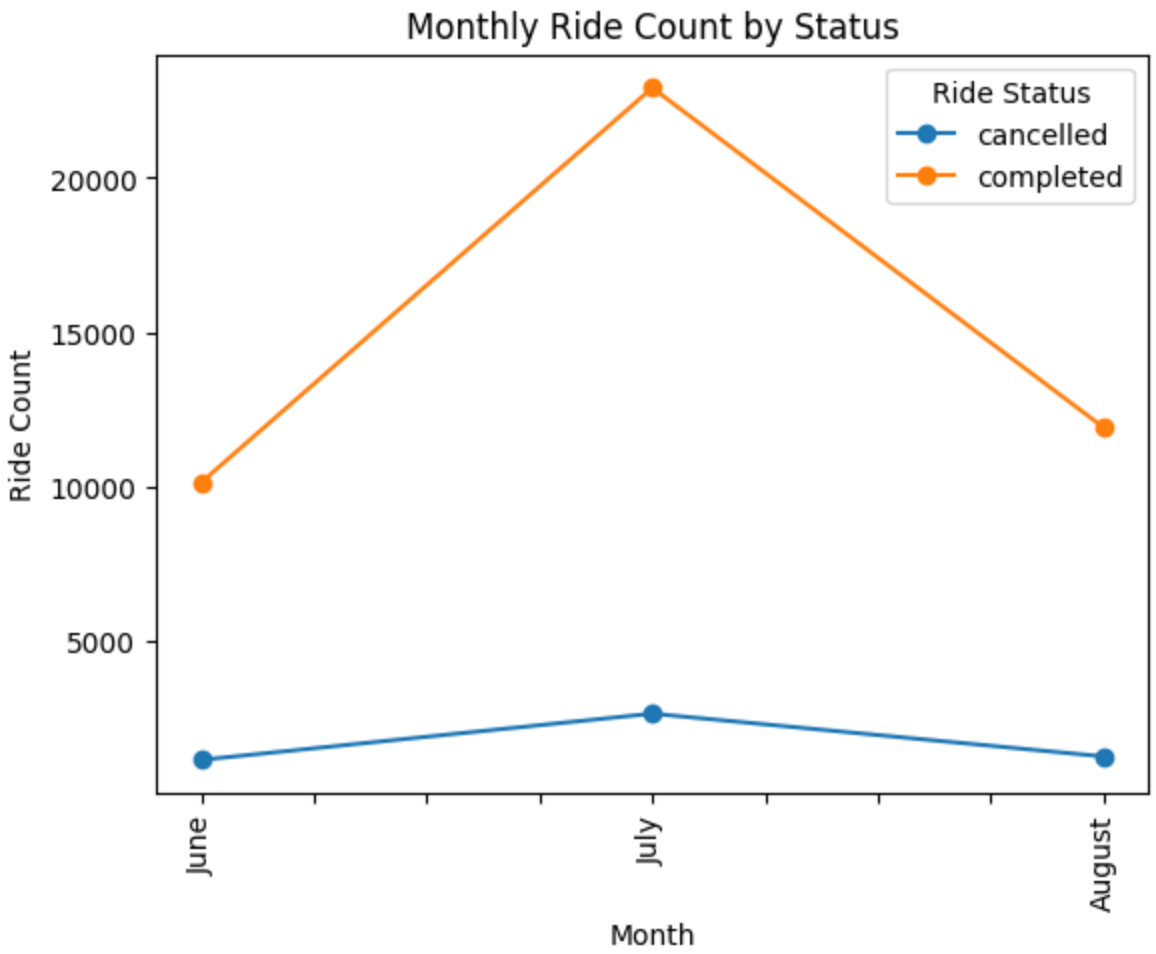
Total fare and ride charge have a perfect positive correlation (1), meaning the fare depends heavily on the ride charge. Miscellaneous charges show a slight positive correlation (0.052) with total fare.

- Negative Correlations:

Duration has a weak negative correlation with ride charge (-0.013) and total fare (-0.013), indicating minimal influence.

Monthly Ride Count by Status

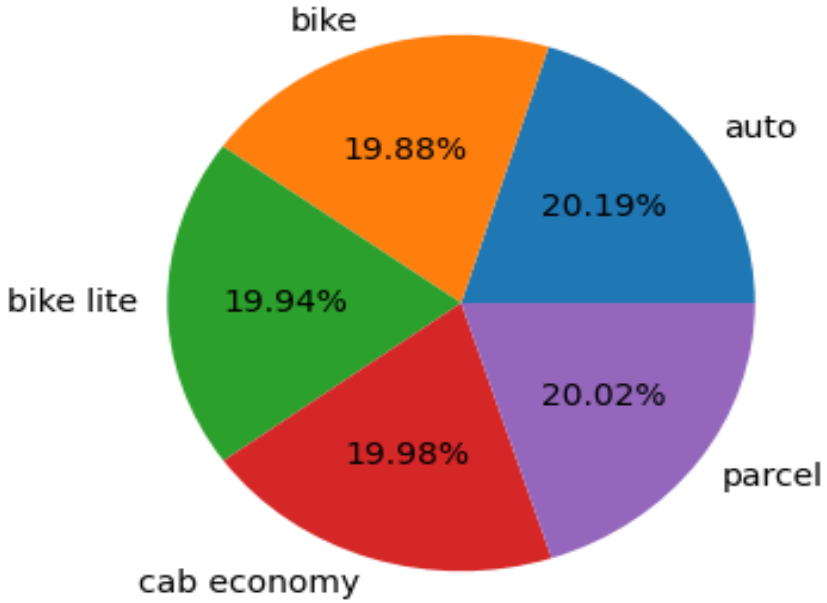
The graph shows that completed rides were highest in July, with June having fewer rides. Cancellations remained low overall, but there was a slight increase in July, probably due to the higher demand. This could be attributed to seasonal factors like school and college openings or rain.



Average Duration Per Service Type.

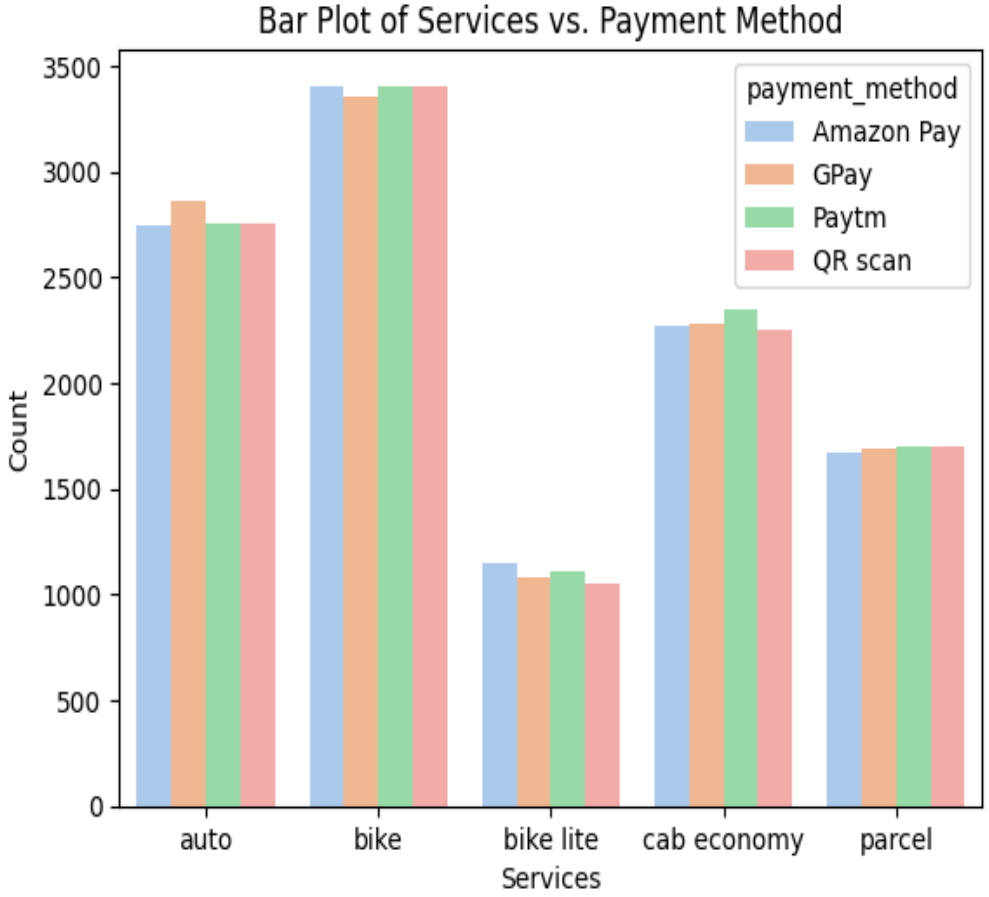
Most trip durations are similar, averaging around 60-65 minutes, showing that the services are generally consistent. If some services take longer or shorter, it might be due to factors like traffic, customer preferences, or the type of service, which could be looked into for improvement.

services	Duration
auto	64.916849
bike	63.925172
bike lite	64.117936
cab economy	64.247893
parcel	64.366001



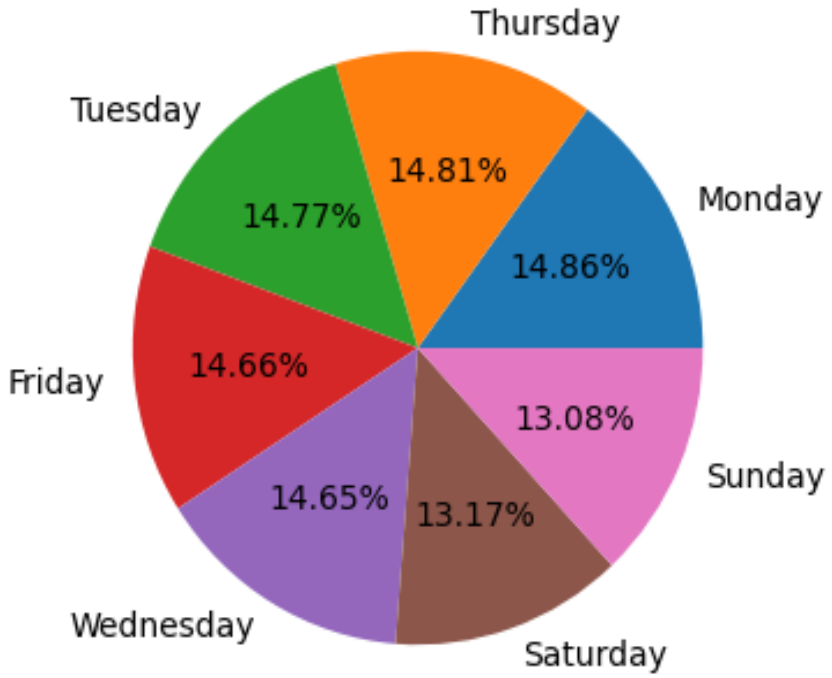
Payment Method Usage Trends

- Bike and auto services are the most popular, with high counts for all payment methods.
- Cab economy and parcel services have moderate usage, with counts that are slightly lower than bikes or autos.
- Bike lite is the least used service, regardless of the payment method.
- Across all services, the number of transactions is fairly balanced across the four payment methods, meaning no single payment method dominates.

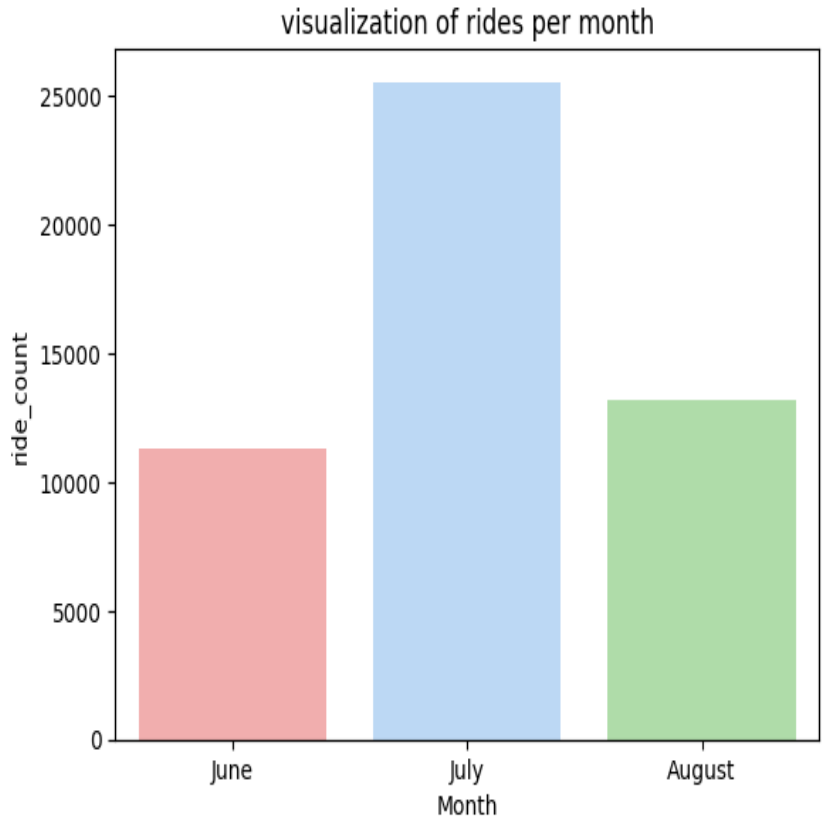


Day-wise Count

Day	Count
Monday	7432
Thursday	7404
Tuesday	7384
Friday	7329
Wednesday	7327
Saturday	6584
Sunday	6540



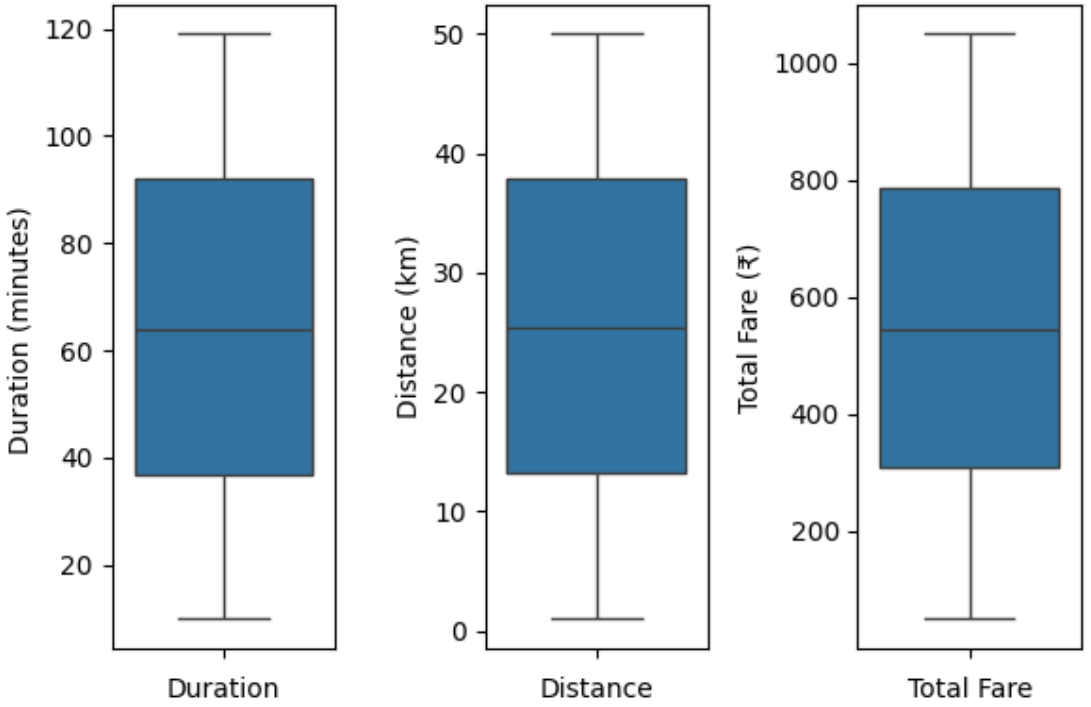
Visualization of Rides Per Month



The chart highlights ride activity over three months, with July being the most active month by a large margin. June had the least number of rides, while August showed a slight increase compared to June but didn't come close to July's high.

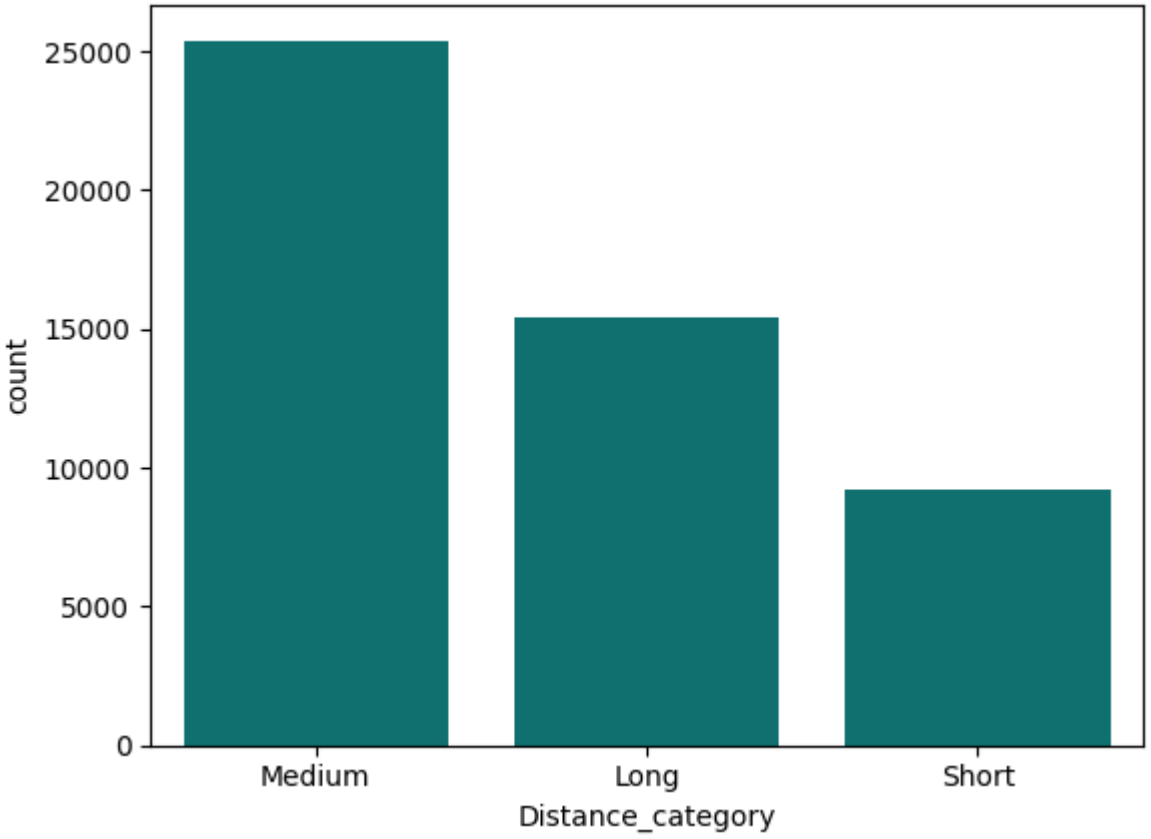
Comparison of Duration, Distance, and Fare Ranges

- Duration: Most trips between 40-90 minutes, with a median around 60 minutes.
- Distance: Trips typically cover 15-40 km, having a median of just above 25.
- Total Fare: Fares range from above 300-800 units, with an average value around 600 units.



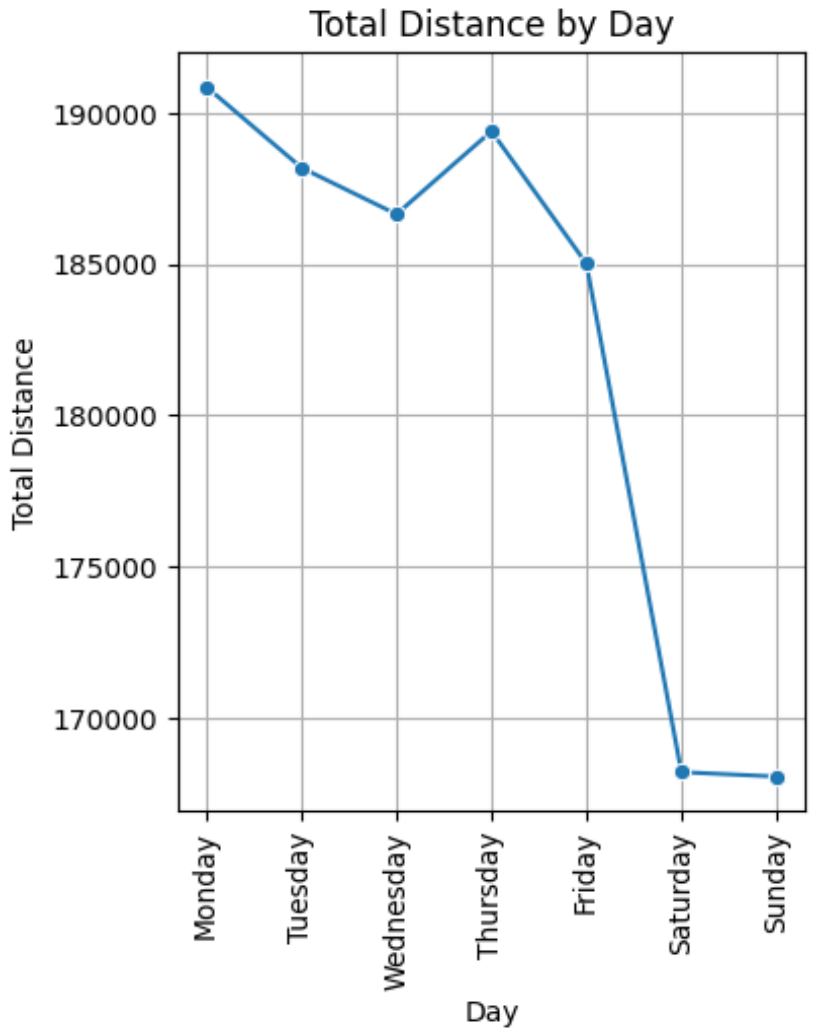
Distance Category

If the distance is less than or equal to 10, it returns “Short”, greater than 10 but less than or equal to 35 returns “Medium”, and greater than 35 return “Long”. From this analysis Medium category have highest count comparing to other.



Total Distance by Day

- Monday records the highest total distance, peaking above 190,000 units.
- Saturday and Sunday show the lowest total distances, dropping below 170,000 units, indicating a significant decline during the weekend.
- A sharp rise is observed on Thursday, bringing the total distance back above 185,000 units after the weekend dip.
- The total distance gradually declines again after Tuesday and Wednesday, maintaining values slightly below the peak days.



Key Findings

- Payment Methods: Paytm emerged as the most popular payment option, followed by GPay, Amazon Pay, and QR scans.
- Service Utilization: Bike services dominated the count, followed by autos, cab economy, and parcel services. Bike Lite had the least utilization.
- Ride Trends: Most rides occurred in July, with moderate activity in August and a dip in June. Monday and Thursday were the busiest days for rides.
- Common Routes : Popular routes included Anekal Woods to Yelahanka Complex and other city locations.
- Trip Durations : Average trip duration hovered around 60 minutes, with variations depending on service type.
- Fares and Distance : Average fares ranged between 300-800 units, while distances typically spanned 15-40 km.

Conclusion

- Weekend Specials :
  Offering discounts on weekends can encourage more people to ride on slower days like Saturdays and Sundays.
- Enhance Long-Distance Services :
  Focus on long-distance rides by introducing exclusive offers, rewards programs, or premium services to attract high-value customers and maximize revenue generation.
- Prioritize Key Locations :
  Ensure sufficient driver availability in high-demand areas during peak times to minimize wait times and maintain service quality.
- Encourage Rides During Quiet Times :
  Introducing special offers or bonuses for rides during quieter hours can attract more customers and keep drivers busy.