RAPIDO DRIVE ANALYSIS:
Using Pandas:
Deep Dive into Rapido Bangalore

Introduction to the Dataset:

The dataset used in this analysis is the Rapido Bangalore dataset, sourced from Kaggle, covering two months of ride data. It includes details such as ride duration, fare charges, payment methods, and ride statuses, along with source and destination locations.

Dataset Columns:

The Rapido dataset contains several features that provide valuable insights into ride behavior and service usage

- **Services**: Type of service chosen (e.g., "Bike," "Auto," "Parcel").
- **Total_Fare**: Total ride fare, including base and extra charges.
- **Duration**: Time taken for the ride (in minutes).
- **Rider_ID**: Unique identifier for the rider.
- **Distance**: Distance traveled during the ride.
- **Ride_Status**: Whether the ride was completed or canceled.
- **Ride_Charge**: The base charge for the ride.
- **Payment_Method**: Payment mode used (e.g., Paytm).
- **Source**: Pickup location.
- **Destination**: Drop-off location.
- **Miscellaneous_Charge**: Extra charges for the ride.
- **Date and Time**: The date and time of the ride

Dataset Rows:
After cleaning, the total number of rows in the Rapido dataset is **50,000**. This includes details of completed and canceled rides. Null or missing values in the dataset (e.g., for canceled rides) have been handled by imputing zeros or marking them appropriately.

Aim: To highlight key insights and recommendations to optimize business strategies, improve operational efficiency, and enhance customer satisfaction.

Objective: A data-driven analysis of Rapido Bangalore's operations, highlighting key insights and recommendations to optimize service delivery, enhance customer satisfaction, and drive business growth.

EDA-Exploratory Data Analysis
Steps in EDA:

- Data collection
- Understanding the data
- Data cleaning
- Feature engineering
- Univariate analysis
- Bivariate analysis
- Multivariate analysis
- Conclusion and Analysis

Overview about data

The Rapido dataset, sourced from Kaggle, contains 50,000 rows of data from approximately two months of Rapido rides in Bangalore. It includes both completed and canceled rides with 13 columns: services, date, time, ride_status, source, destination, duration, ride_id, distance, ride_charge, misc_charge, total_fare, and payment_method. Missing values, especially for canceled rides, have been handled by imputing zeros or marking them appropriately.

services	date	time	ride_status	source	destination	duration	ride_id	distance	ride_charge	misc_charge	total_fare	payment_method
cab economy	2024-07-15	8:30:41	completed	Balagere Harbor	Harohalli Nagar	39	RD31612187518753	27.21	764.83	31.51	796.34	Amazon Pay
auto	2024-07-05	23:36:52	completed	Basavanagudi 3rd Bl	Bikasipura 1st Stage	89	RD81715142845940	34.03	314.83	49.52	364.35	Paytm
auto	2024-07-23	11:05:38	cancelled	Babusapalya Cove	Kothaguda Terrace	25	RD93764811222379	20.24				nan
cab economy	2024-06-24	8:45:11	completed	Mahadevapura Mew	Kanakapura Arc	89	RD36768891431827	31.17	484.73	15.84	500.57	QR scan
cab economy	2024-07-15	0:26:45	completed	Ganganagar Cove	Basaveshwaranagar	95	RD66394102759480	27.21	663.5	14.13	677.63	Amazon Pay
auto	2024-07-02	1:28:30	completed	HSR Layout Area	JP Nagar Viewpoint	18	RD59222054864419	33.69	456.73	25.19	481.92	QR scan
cab economy	2024-07-23	20:55:29	completed	Arekere Heights	Dooravani Nagar Poi	85	RD95570993968884	20.44	836.39	14.95	851.34	GPay
parcel	2024-07-18	13:38:34	completed	Electronic City Villa	Ganganagar Station	89	RD94738739535256	35.31	724.76	31.34	756.1	QR scan
parcel	2024-08-08	7:59:54	completed	Mysore Road Lane	Billekahalli 6th Bloc	72	RD12855660121672	45.99	641.55	21.48	663.03	Amazon Pay
bike lite	2024-07-10	12:30:08	completed	Kundalahalli Alley	RT Nagar 5th Block	94	RD72386280941420	44.43	571.5	47.63	619.13	QR scan
auto	2024-08-03	16:57:15	completed	Hosur Road Mews	Nagawara Layout	80	RD09884314035185	5.25	356.07	33.18	389.25	QR scan
cab economy	2024-07-08	17:46:16	completed	Bikasipura Close	Kadugodi Park	42	RD17462985177331	18.53	920.85	44.03	964.88	Paytm
cab economy	2024-08-10	13:46:08	completed	Mahadevapura Trac	Whitefield Cut	68	RD73415047756837	5.11	595.8	23.72	619.52	Paytm
parcel	2024-08-15	22:59:05	cancelled	Bhadrappa Layout V	Subramanyapura Ro	82	RD02566801183667	29.83				nan
bike	2024-07-22	13:07:17	cancelled	Frazer Town Works	Hebbal Kempapura	44	RD71150594028065	49.92				nan
bike	2024-07-24	5:35:39	completed	Nagawara Dam	Dommasandra Color	42	RD23696468127295	39.3	256.39	31.98	288.37	Paytm
bike	2024-07-26	18:53:04	completed	Sonnenahalli Layou	Kothanur Loop	95	RD80479898090024	21.39	62.83	23.51	86.34	Paytm
auto	2024-06-28	22:55:09	completed	Hulimavu Cutting	HRBR Layout Valley	37	RD35925713304851	48.12	811.27	41.32	852.59	Amazon Pay
auto	2024-08-08	11:59:05	completed	Billekahalli Cove	Hennur Road 1st Blo	42	RD88176734558719	2.36	188.04	46.52	234.56	GPay
parcel	2024-08-08	0:57:03	completed	Banaswadi District	Sadashiva Nagar Co	50	RD13858430527524	37.5	473.87	13.24	487.11	Amazon Pay
parcel	2024-07-22	10:04:39	completed	Hebbal Kempapura	Ramnagar 1st Stage	104	RD72089065313846	3.04	617.35	33.34	650.69	Paytm
auto	2024-07-03	1:01:30	completed	Rajarajeshwari Naga	ITI Layout Close	11	RD06546321588797	9.52	306.79	44.68	351.47	GPay
cab economy	2024-07-08	1:16:17	completed	Kudlu Gate Fields	Subramanyapura Pa	117	RD22391777321424	37.93	853.42	39.57	892.99	GPay
auto	2024-07-01	7:18:37	completed	Byatarayanapura Pla	Koramangala 6th Bl	46	RD11170357606527	21.6	907.37	20.55	927.92	GPay
bike	2024-06-21	6:23:05	completed	Whitefield Hills	MG Road Hills	49	RD42268110798580	12.8	131.29	29.98	161.27	GPay

Data Cleaning

Handling Null Values for Canceled Rides by setting the ride_charge, misc_charge, and total_fare to 0 for canceled rides, to ensure data consistency and accuracy for further analysis.

df.loc[df['ride_status'] == 'canceled', ['ride_charge', 'misc_charge', 'total_fare']] = 0 df.loc[df['ride_status'] == 'canceled', 'payment_method'] = 'canceled'

df.loc[df['ride_status'] == 'canceled', ['ride_charge', 'misc_charge', 'total_fare']] = 0

df.loc[df['ride_status'] == 'canceled', 'payment_method'] = 'canceled'

The dataset was thoroughly examined for duplicates and using boxplot found that there were no outliers. No duplicate records or significant outliers were identified during the data cleaning process.

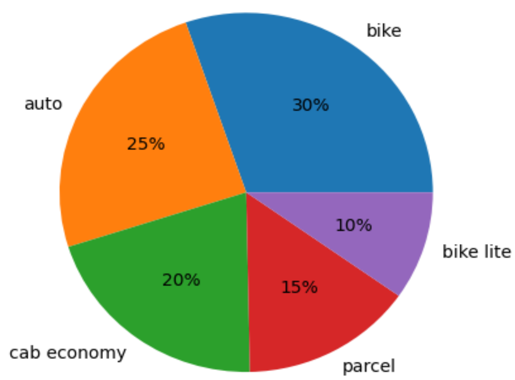
Feature Engineering:

Categorized rides into 'short', 'medium', and 'long' based on distance for a more granular analysis of ride behavior. Also added a feature hour and day name of the drive

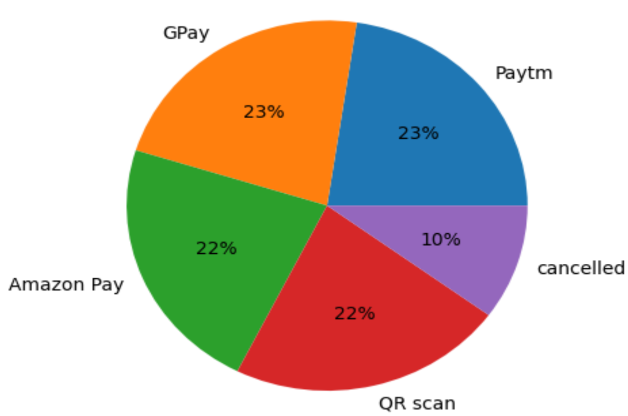
```
def cate(i):
  if i['ride_status'] == 'canceled':
   return 'Canceled'
elif i['distance'] < 5:
 return 'Short'
elif i['distance'] >= 5 and i['distance'] <= 15:
  return 'Medium'
else: return 'Long'

df['distance_category'] = df.apply(cate, axis=1)
```

ANALYSIS:

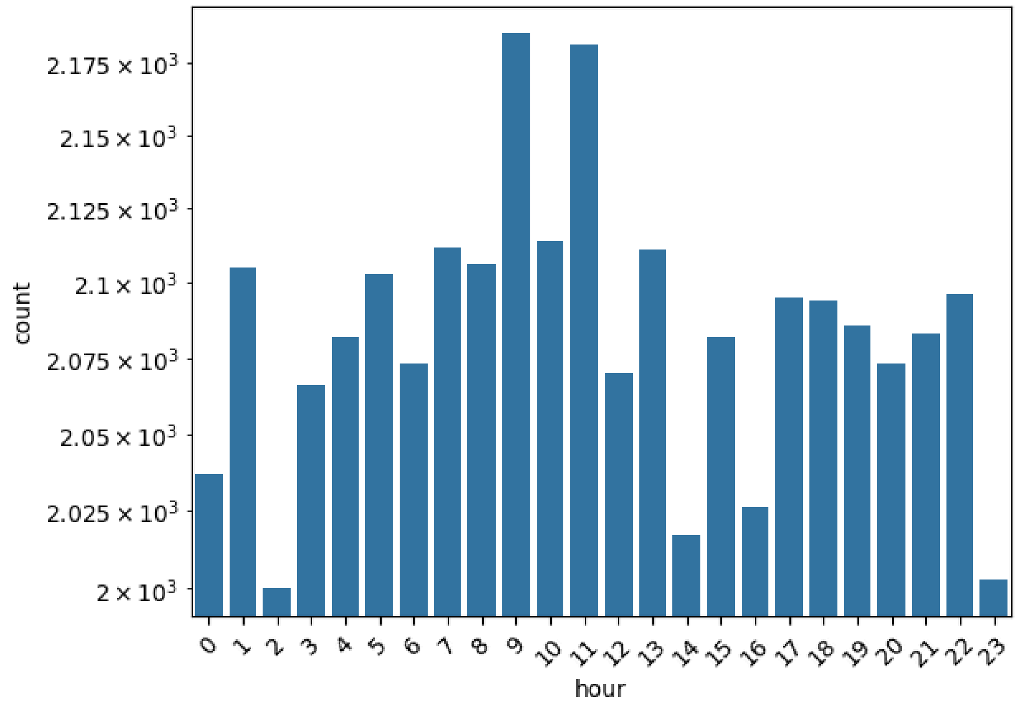
- **Total Fare**: Average ₹492.26; Max ₹1048.8 (Parcel service); 50,000 rides.
- **Charges**: Ride charge ₹23.49M; Misc charges ₹1.12M (₹0–₹50, avg ₹22.45).
- **Ride Duration**: Ranges 10–119 mins; Avg 64.32 mins.
- **Ride Success**: 89.93% completed, 10.07% canceled.
- **Service Preferences**: Most used - Bike (30%); Least used - Bike Lite (10%).



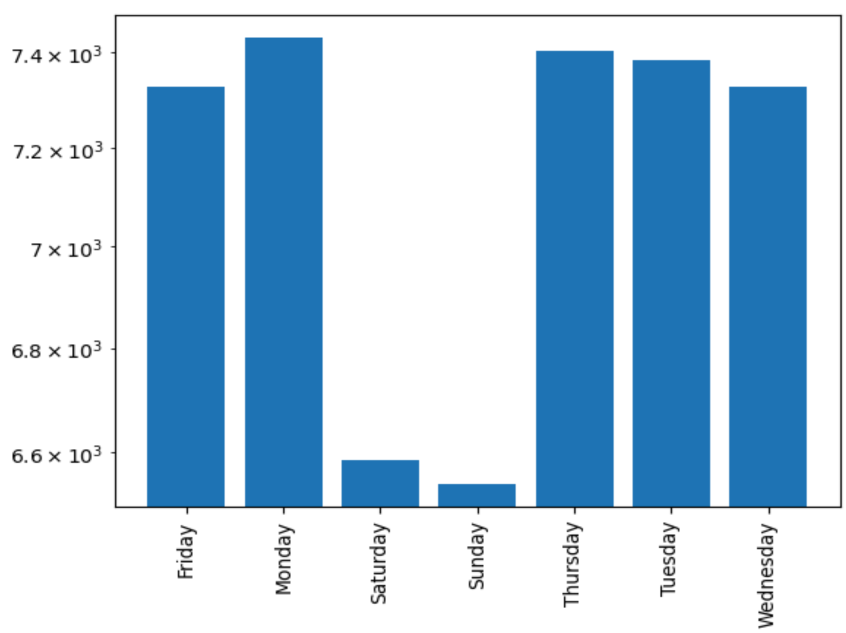
- **Payment Method**: Paytm leads with 11,315 transactions, generating ₹6.2M revenue (avg ₹548.05).



- **Peak Hours**: Morning (8–11 AM) and evening (5–9 PM).



- **Preferred Days**: Monday (7,432 rides) and Thursday (7,404); Least - Sunday (6,540) and Saturday (6,584).



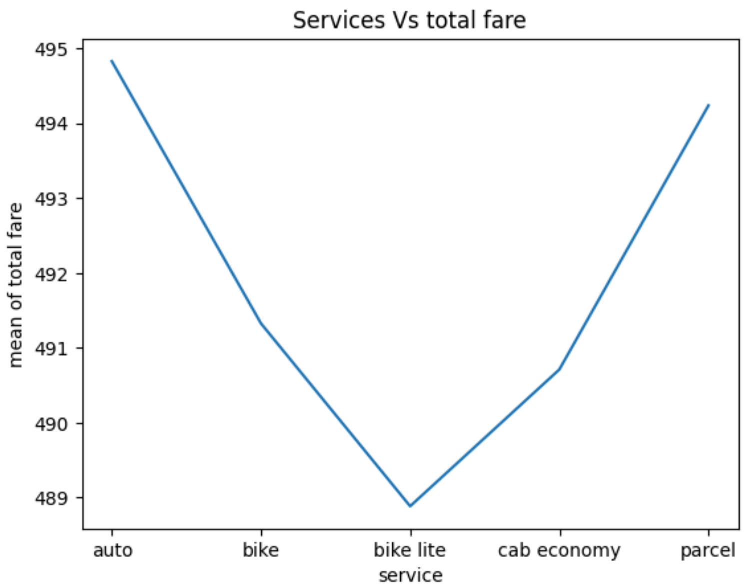
- **Top Locations**: Kothanur Landing (23 pickups); Gottigere Landing (23 drop-offs).

source	source
Kothanur Landing	23
Banaswadi Landing	20
Vijayanagar Square	20
HRBR Layout Landing	19
Ramamurthy Nagar Landing	18
Banaswadi Square	17
HRBR Layout Square	16
Kalyan Nagar Close	15
Chokkanahalli Landing	15
Kothanur 6th Block	15

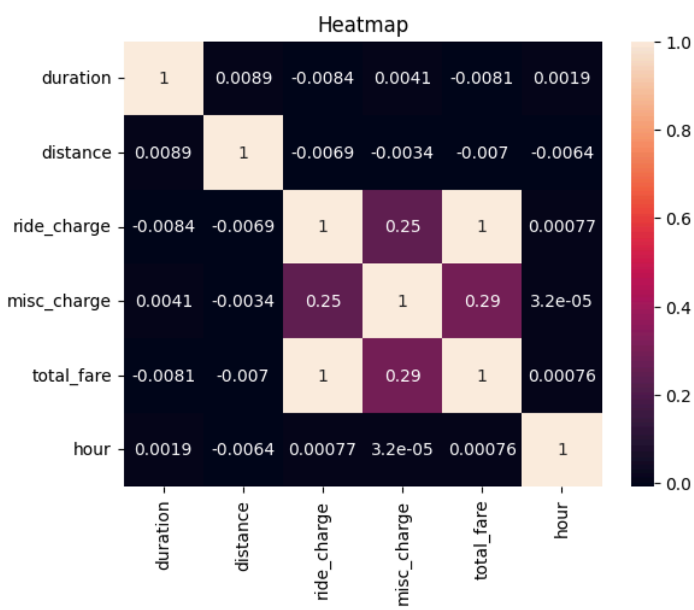
destination	
Gottigere Landing	23
Gottigere Square	20
Kudlu Square	17
HRBR Layout Drive	17
Harohalli Landing	16
Gottigere 6th Block	16
Hulimavu Drive	16
Hulimavu Square	15
Ramamurthy Nagar 6th Block	15
Gottigere District	15

- **Service Performance**:
  - Highest average fare: Auto.
  - Longest average duration: Auto.
  - Longest average distance: Bike Lite.

- **Revenue by Distance**: Long-distance rides earned ₹17.56M from 32,121 rides (avg ₹546.65).
- **Fare Insights**: Auto and Parcel have the highest mean fares; Bike Lite the lowest.



- **Correlations**: Ride charge and miscellaneous charges are strongly related.



KEY INSIGHTS

- **Ride Status & Cancellation:** 89.93% of rides are successfully completed, with only 10.07% canceled, showcasing a high success rate.
- **Revenue by Service Type:** Bike services lead with ₹7,432,784 in revenue, followed by Auto and Cab Economy, highlighting their profitability.
- **Revenue by Distance:** Long-distance rides are the most valuable, with an average fare of ₹547.
- **Peak Ride Times:** Commuting hours (8–11 AM and 5–9 PM) see the highest ride demand.
- **Day-wise Demand:** Mondays and Tuesdays have the most rides, while weekends see lower demand.
- **Payment Preferences:** Digital payments, particularly Paytm and QR Scan, dominate, reflecting a cashless trend.
- **Popular Locations:** Kothanur Landing is the top pickup spot, while Gottigere Landing is the most frequent drop-off location.
- **Auto vs Cab Services:** Auto and Cab Economy services have similar average fares and durations, providing flexibility for customers.

CONCLUSION

- **Capitalize on Long-Distance Rides**: Introducing targeted promotions or loyalty programs for long-distance rides can further enhance revenue from this high-value category.
- **Balance Demand During Off-Peak Hours**: Providing discounts during non-peak periods can help distribute demand evenly and boost service efficiency.
- **Strengthen Digital Payment Options**: Expanding collaborations with platforms like Paytm and QR Scan can enhance payment convenience and improve customer satisfaction.
- **Improve Driver Allocation**: Ensuring driver availability at high-demand pickup and drop-off locations during peak hours can reduce wait times and optimize operations.
- **Boost Weekend Ridership**: Launching special weekend discounts can encourage higher usage on low-demand days, such as Saturdays and Sundays.