Deep Dive into Rapido Bangalore:

A Data-Driven Analysis

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.*

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 dataset

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 I	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

ROWS=50000, COLUMNS=13

Data Cleaning

```
UPDATE rapidorides SET ride_charge = 0,
misc_charge = 0, total_fare = 0 WHERE ride_status
= 'cancelled';
```

The SQL query executto handle 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.

Feature Engineering

```
ALTER TABLE rapidorides ADD COLUMN distance_category TEXT; UPDATE rapidorides SET distance_category = CASE WHEN distance < 5 THEN 'short' WHEN distance BETWEEN 5 AND 15 THEN 'medium' ELSE 'long' END;
```

Categorized rides into 'short', 'medium', and 'long' based on distance for a more granular analysis of ride behavior.

ANALYSIS

Calculating the total revenue and average fare for each service type (bike, auto, cab, etc.). It helps to identify the most profitable services and understand their revenue potential.

```
SELECT services,SUM(total_fare) AS total_revenue,AVG(total_fare) AS avg_fare
FROM rapidorides WHERE payment_method != 'cancelled' GROUP BY services order by total_revenue desc;
-- bike 7432783.949999977   547.8575919510561
-- auto 6099731.319999989   548.83312218823
```

Analysing the revenue and average fare based on distance categories (short, medium, long). It helps to identify the most profitable distance segments and optimize pricing strategies.

```
SELECT distance_category,SUM(total_fare) AS total_revenue, AVG(total_fare) AS avg_revenue
FROM rapidorides WHERE payment_method != 'cancelled' GROUP BY distance_category order by total_revenue desc;
-- long 17559047.229999907  546.6531935493884
-- medium   5064196.179999989   550.1571080934264
-- short    1989739.6400000001  546.9322814733371
```

The top 10 busiest hours of the day based on the number of rides. It helps to optimize resource allocation and staffing during peak hours.

```
SELECT HOUR(time) AS ride_hour,COUNT(*) AS ride_count
FROM rapidorides WHERE payment_method != 'cancelled'
GROUP BY HOUR(time) ORDER BY ride_count DESC limit 10;
```

Analyse the number of rides on each day of the week. It helps to identify peak days and adjust operations accordingly.

```
SELECT dayname(`date`) AS ride_day,COUNT(*) AS ride_count FROM rapidorides
WHERE payment_method != 'cancelled'
GROUP BY ride_day ORDER BY ride_count DESC;
# monday is most with 6679, tuesday-6659
```

Analyse the usage of different payment methods. It helps to understand customer preferences and optimize payment options.

```
SELECT payment_method,COUNT(*) AS method_count FROM rapidorides
WHERE payment_method != 'cancelled' GROUP BY payment_method ORDER BY method_count DESC;
# paytm-11315
```

Popular Locations:By identifying these popular locations, Rapido can optimize driver allocation and route planning to improve efficiency and reduce wait times.

```
SELECT source, COUNT(*) AS pickup_count FROM rapidorides
GROUP BY source ORDER BY pickup_count desc;
#most:Kothanur Landing  23

SELECT destination, COUNT(*) AS destination_count FROM rapidorides
GROUP BY destination ORDER BY destination_count desc;
#most:Gottigere Landing  23
```

Calculates the percentage of completed and canceled rides. By analyzing this data, identify trends in ride success and take steps to improve completion rates.

```
SELECT ride_status, COUNT(*) AS count,(COUNT(*) * 100.0 / (SELECT COUNT(*) FROM rapidorides)) AS percentage
FROM rapidorides GROUP BY ride_status;
-- completed    44964    89.92800
-- cancelled    5036     10.07200
```

Compare the performance of different services (auto, cab, bike, etc.) in terms of average fare, duration, and distance. By analyzing these metrics, identify trends, optimize pricing strategies, and improve service quality.

```
SELECT services,AVG(total_fare) AS avg_fare, AVG(duration) AS avg_duration,AVG(distance) AS avg_distance
FROM rapidorides WHERE payment_method != 'cancelled' GROUP BY services;
```

KEY INSIGHTS

- **Ride Status & Cancellation**: Out of 50,000 rides, 89.93% are completed successfully, while only 10.07% are canceled, indicating a high success rate.
- **Revenue by Service Type**: Bike services lead in revenue generation at ₹7,432,784, with Auto and Cab Economy closely following, showing these services' profitability.
- **Revenue by Distance**: Long-distance rides generate the highest revenue with an average fare around ₹547, making them the most valuable trips.
- **Peak Ride Times**: Rides peak during commuting hours, 8–11 AM and 5–9 PM, highlighting times of high demand.
- **Day-wise Demand**: Monday and Tuesday have the highest number of rides, while demand drops on Saturday and Sunday.
- **Payment Preferences**: Digital payments, especially Paytm and QR Scan, are the most common methods, indicating a strong trend toward cashless transactions.
- **Popular Locations**: The most common pickup point is Kothanur Landing, and the most frequent drop-off is Gottigere Landing, suggesting areas to optimize driver availability.
- **Auto vs Cab Services**: Auto and Cab Economy services show similar average fare and duration, offering flexible options for customers.

CONCLUSION

- **Focus on Long-Distance Rides**: Since long-distance rides generate higher revenue, targeted promotions or loyalty rewards could maximize profitability in this category.
- **Optimize Demand Balance**: Offering discounts during non-peak hours could help smooth out demand and improve service utilization.
- **Enhance Digital Payments**: Expanding partnerships with digital wallets like Paytm and QR Scan can further boost user convenience and satisfaction.
- **Driver Availability at Key Locations**: Prioritizing driver allocation at popular pickup and drop-off points during peak hours will ensure smoother service and reduce wait times.
- **Weekend Promotions**: Implementing special discounts on weekends could help increase ridership on typically low-demand days like Saturday and Sunday.