



# ANALYSIS REPORT

*TINY CODERS*  
**IITG - CASCADE CUP 22**



# CONTENTS

---

**01**

INTRODUCTION

**02**

DATA STATISTICS

**03**

CANCELLED ORDERS

**04**

WHAT DAY IS IT?

**05**

FIRST AND LAST MILES

**06**

RIGHT-SKEWED

**07**

CALL DATA

**08**

POWER BI DASHBOARD

**09**

CONCLUSION

## INTRODUCTION

---

**SHADOW FAX - INDIA'S LARGEST HYPERLOCAL DELIVERY PLATFORM** IS A SINGLE, TECH-DRIVEN, ALL-ACCESS PLATFORM SUPPORTING DELIVERY PARTNERS PAN INDIA BY PROVIDING OPPORTUNITIES FOR EARNING AND DELIVERING ACROSS ALL COMPANIES.

**PROBLEM STATEMENT** : TO PRODUCE DETAILED DATA ANALYSIS REPORT ON THE RIDERS DATASET PROVIDED BY SHADOW FAX DELIVERY COMPANY AND TO GET INSIGHTS FROM THE DATA WHICH HELPS THE COMPANY TO IMPROVE ITS POLICIES OF WORK.

THE FOLLOWING REPORT IS THE SUMMARY OF THE COMPANY'S DATA.

# DATA STATISTICS

## TRAIN DATA

THE TRAIN DATA CONSISTS OF 20 COLUMNS IN TOTAL, OUT OF WHICH 11 ARE NUMERICAL AND REMAINING 9 ARE CATEGORICAL.

THE DATASET CONSISTS OF 450000 RECORDS, AND HAS 20.2% OF CELLS OCCUPIED BY MISSING VALUES.

SOME STANDARD RESULTS FROM THE NUMERICAL COLUMNS CAN BE FOUND FROM THE ADJACENT TABLE.

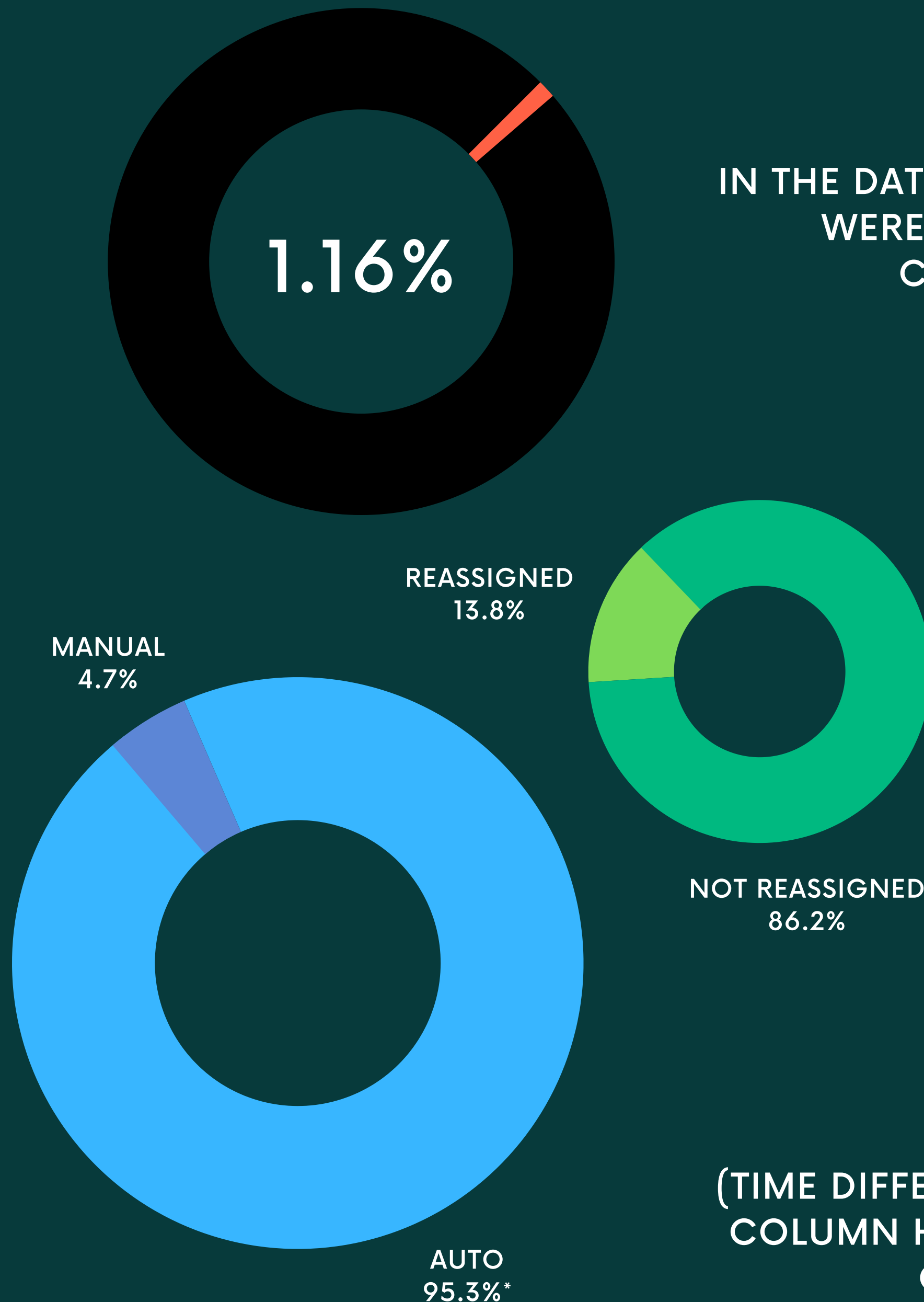
## CALL DATA

THE CALL DATA CONSISTS OF 5 COLUMNS IN TOTAL, OUT OF WHICH 3 ARE NUMERICAL AND REMAINING 2 ARE CATEGORICAL.

THE DATASET CONSISTS OF 562624 RECORDS, AND HAS 18.7% OF CELLS OCCUPIED BY MISSING VALUES.

VARIABLE	MEAN	STD	MIN	MAX
order_id	369143.1	131146.9	118350	594842
rider_id	7763.2	5592.9	0	21566
first_mile_distance	1.23	0.85	0.0001	42.04
last_mile_distance	2.97	1.884	0	22.41
alloted_orders	104.62	90.14	1	567
delivered_orders	103.95	89.64	1	562
cancelled	0.0116	0.1071	0	1
undelivered_orders	0.764	1.0665	0	9
lifetime_order_count	853.64	1502.98	0	30469
reassigned_order	1	0	1	1
session_time	220.475	176.714	0	1298.97

## CANCELLED VS UNCANCELLED



# CANCELLED ORDERS

IN THE DATASET, OUT OF 450000 ORDERS PLACED, 5218 ORDERS WERE CANCELLED. FROM THIS WE CAN SAY THAT THE CANCELLATION RATE IS SIGNIFICANTLY LOW.

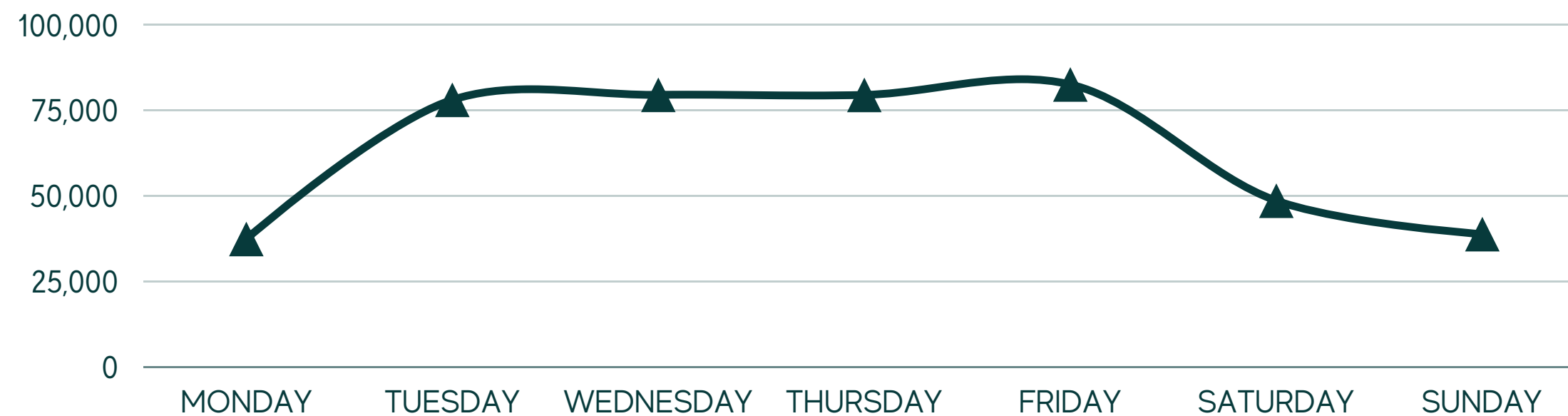
PERCENTAGE OF REASSIGNED ORDERS IN CANCELLED ORDERS IS 13.8% WHEREAS IN THE UNCANCELLED ORDERS POOL. THE PERCENTAGE FALLS TO 2.93% ↘

PERCENTAGE OF AUTOMATICALLY REASSIGNED ORDERS IN CANCELLED ORDERS IS 95.3% BUT THE PERCENTAGE GOES AS FAR AS 97.49% IN THE UNCANCELLED ORDERS POOL.

AFTER PRE-PROCESSING, TIME DELAY IN ACCEPTING THE ORDER (TIME DIFFERENCE BETWEEN ALOTTED TIME AND ACCEPTED TIME) COLUMN HAS SHOWN HIGHER CORRELATION WITH CANCELLED COLUMN COMPARED TO OTHER COLUMNS.

# WHAT DAY IS IT?

## UNCANCELLED ORDERS



**MONDAY**

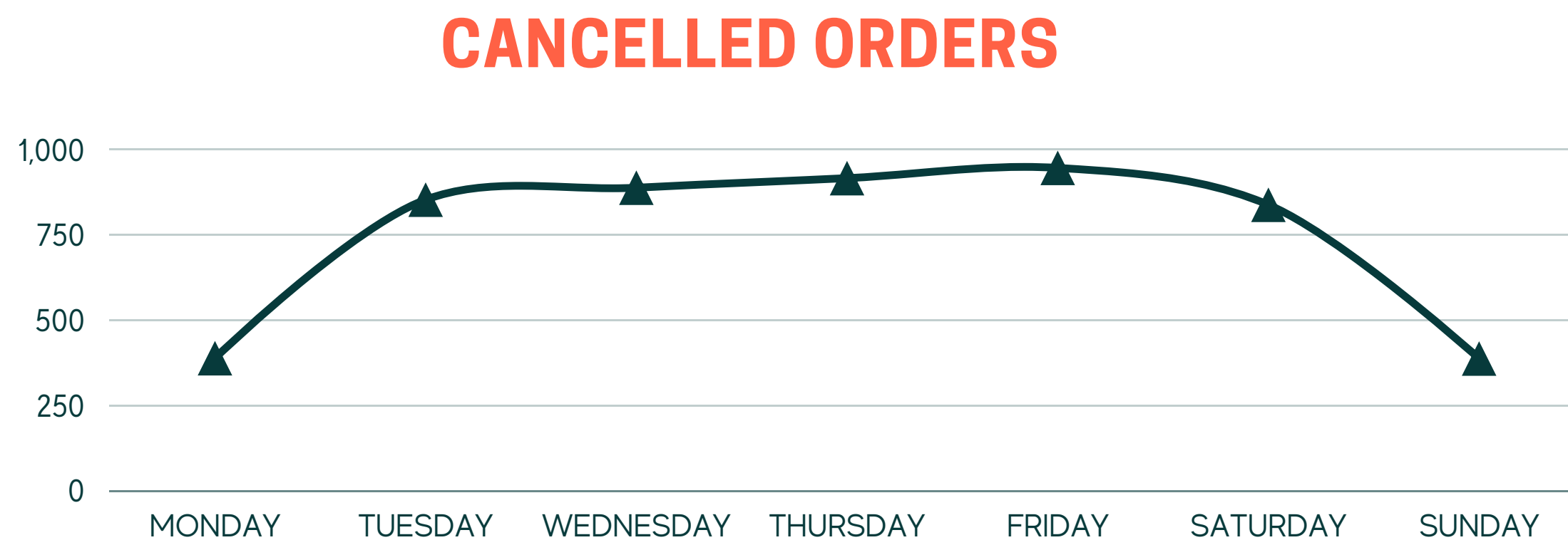
HAS SEEN LEAST NUMBER OF UNCANCELLED ORDERS

**FRIDAY**

HAS SEEN HIGHEST NUMBER OF CANCELLATIONS

**FRIDAY**

HAS HIGHEST CANCELLED TO UNCANCELLED RATIO  
(0.01723)



EVEN THOUGH **MONDAY** HAS LEAST NUMBER OF UNCANCELLED ORDERS THE **LOWEST** CANCELLED TO UNCANCELLED RATIO **0.00998** IS SEEN ON **SUNDAY**, THIS IS DUE TO THE VARIATION IN TOTAL NUMBER OF ORDERS PLACED ON THESE RESPECTIVE DAYS.

# FIRST AND LAST MILES

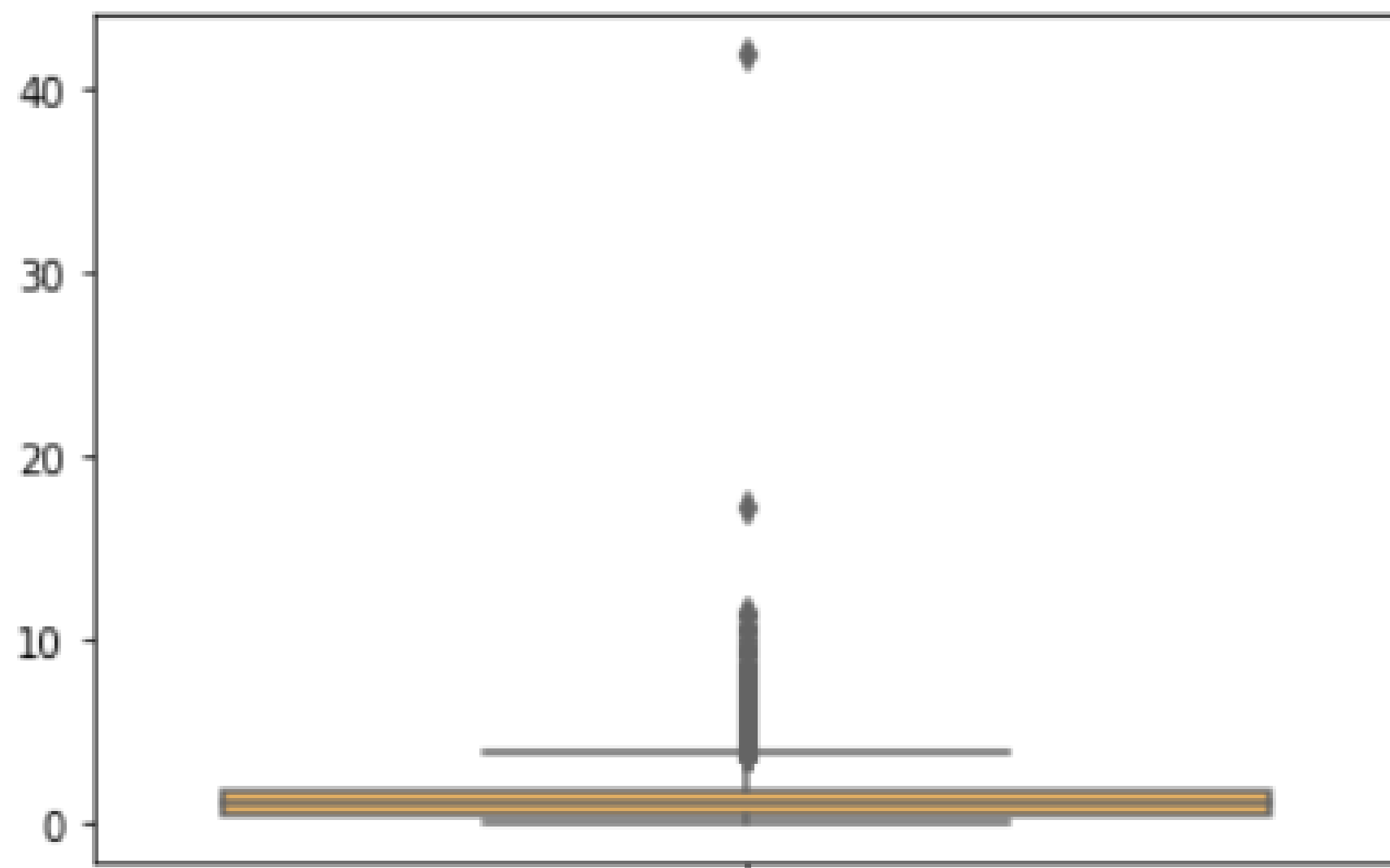
---

THE **FIRST MILE DISTANCE** FOR MOST OF THE ORDERS IS LESS THAN **10 MILES**. THIS INDICATES THAT IN MOST OF THE CASES THE ORDER ASSIGNMENT IS DONE IN SUCH A WAY THAT RIDER'S LOCATION IS NEAR TO PICKUP LOCATION.

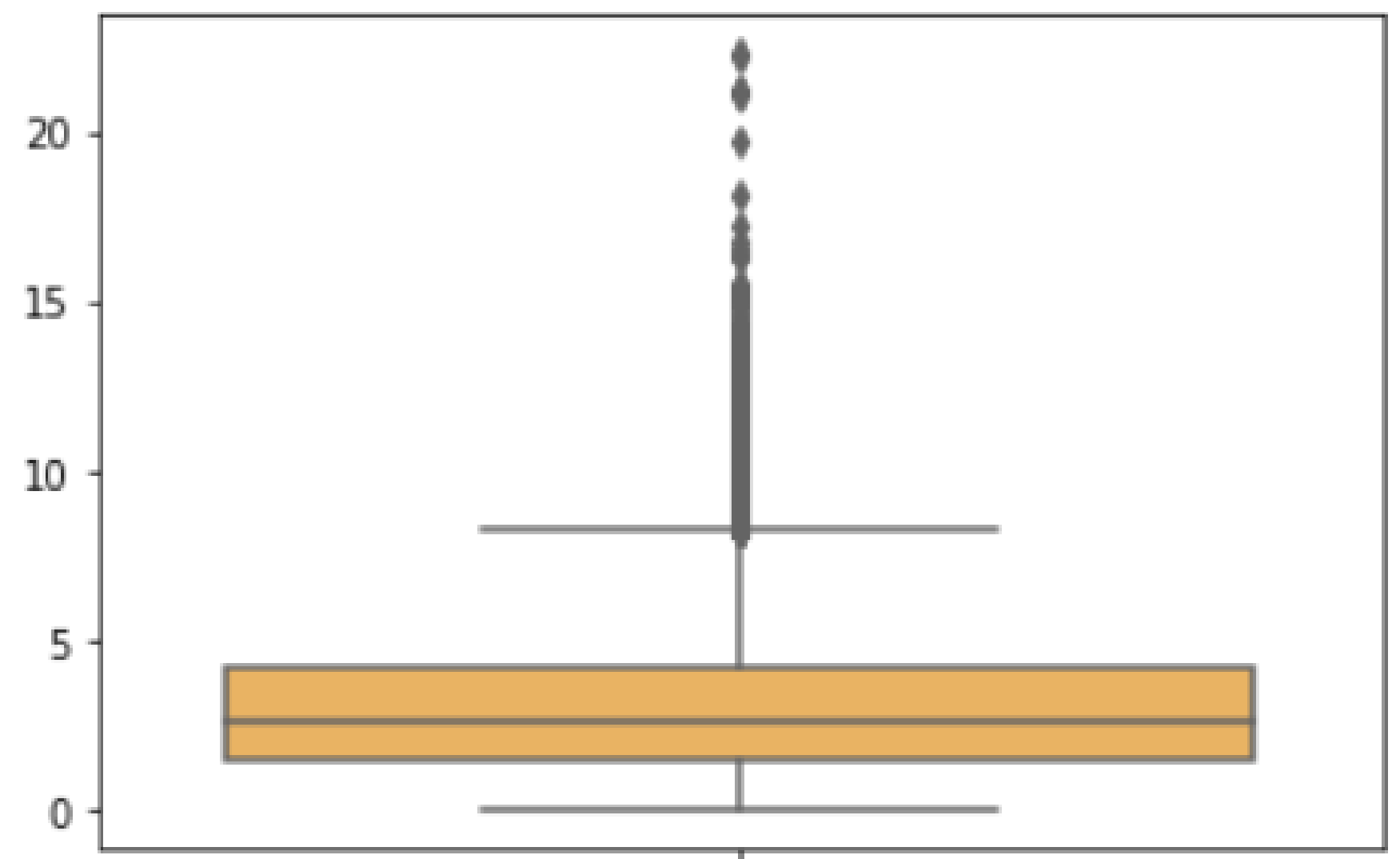
SIMILARLY, WE CAN SEE THAT **LAST MILE DISTANCE** FOR MOST OF THE ORDERS IS UNDER **10 MILES**. THIS INDICATES THAT THE DELIVERY LOCATION TOO IS NEAR TO PICKUP LOCATION OF THE ORDER.

DUE TO LESS MILE DISTANCE, THE TIME TO DELIVER THE ORDER IS REDUCED. HENCE THERE IS VERY **LOW CHANCE** FOR THE ORDER TO GET CANCELLED. BUT THE **OUTLIERS** IN THE BELOW BOXPLOTS REPRESENT HIGH DISTANCES, WE'LL DISCUSS ABOUT THEM IN THE NEXT SLIDE.

FIRST MILE DISTANCE



LAST MILE DISTANCE





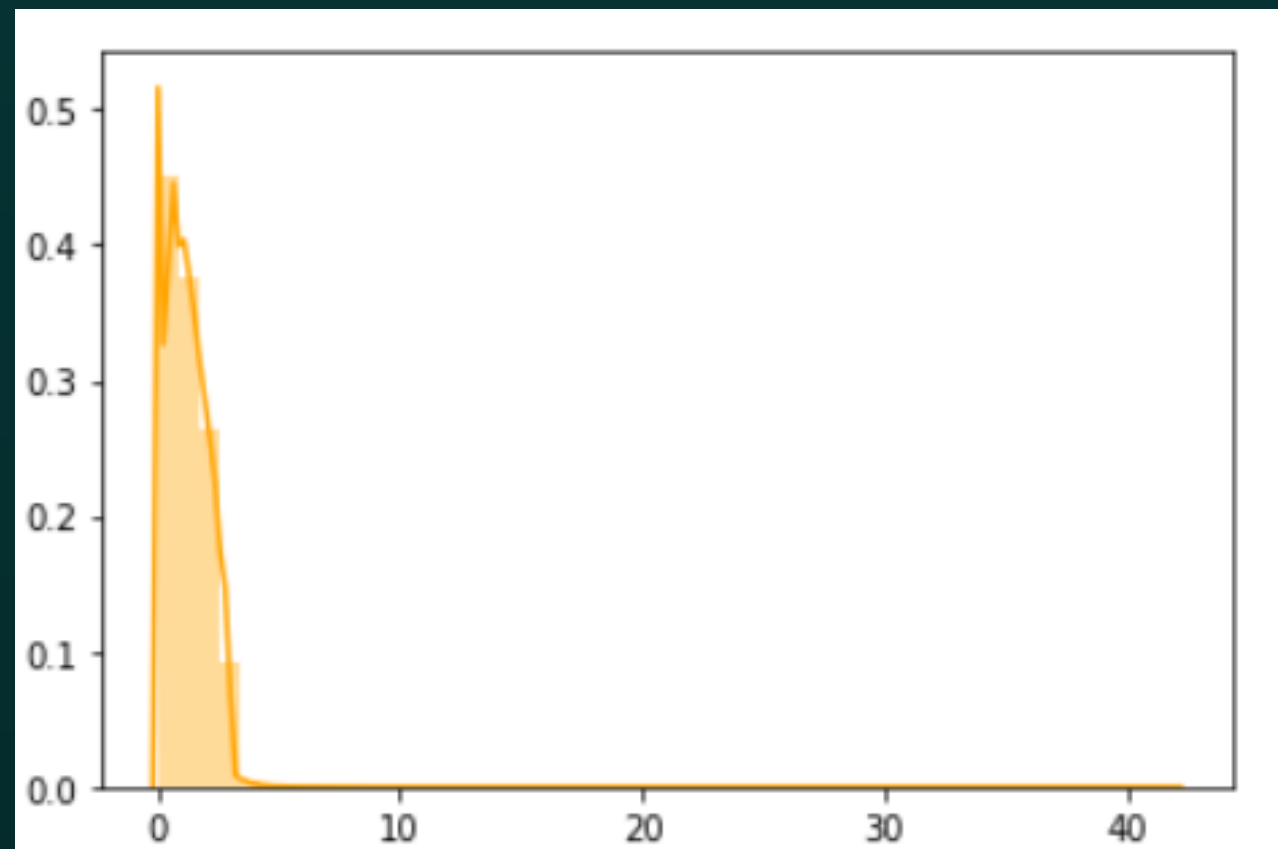
# RIGHT-SKEWED

ALL THE NUMERIC COLUMNS PROVIDED IN THE DATASET ARE **RIGHT-SKEWED**. THIS TELLS US THAT MOST OF THE DATA POINTS ARE IN THE LOWER BOUNDARY OF THE DATASET.

IT ALSO TELLS US THAT OUR DATA HAS OUTLIERS. THESE OUTLIERS ARE THE EXTREME CASES WHERE DISTANCE THAT HAS TO BE TRAVELLED BY DRIVER IS **HIGH**.

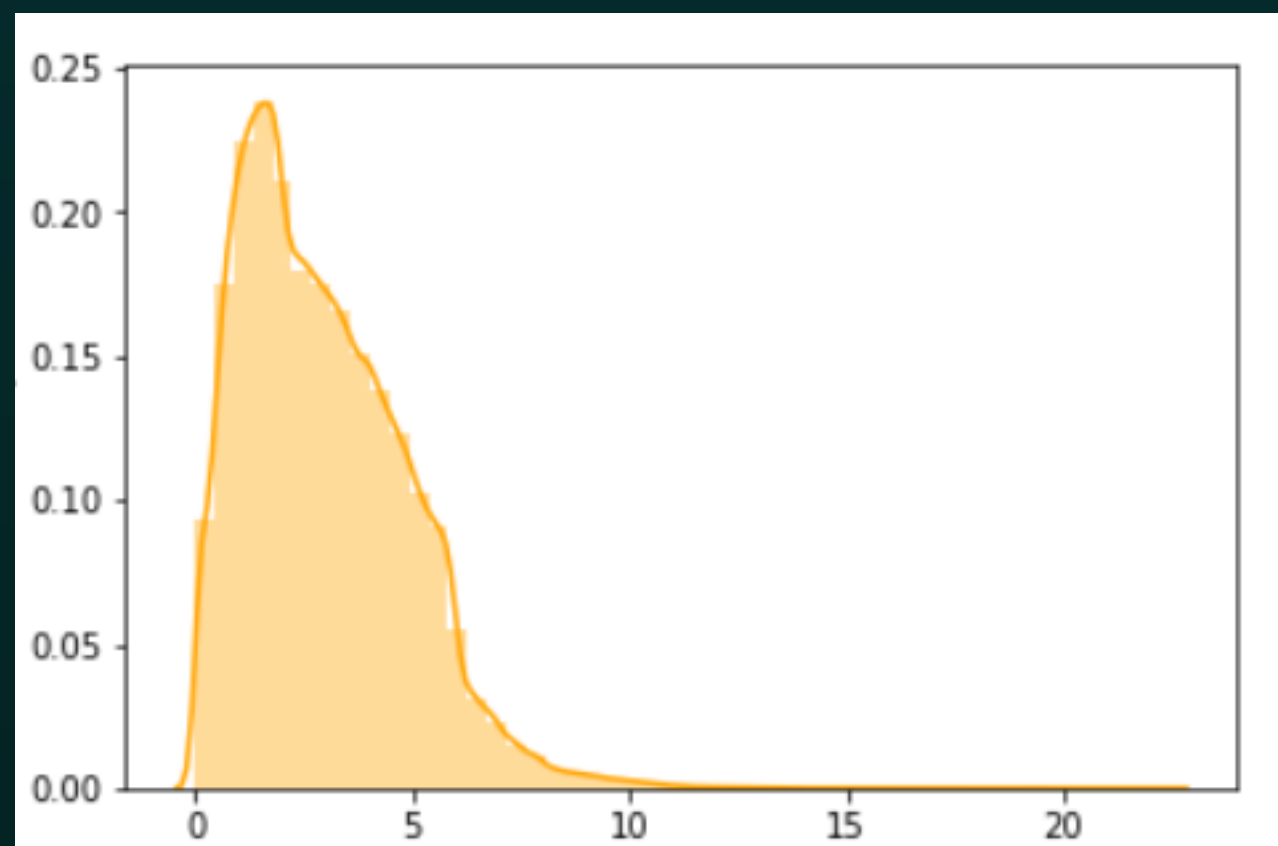
THE FIRST MILE AND LAST MILE DISTANCES ARE MORE IN SOME CASES, RESULTING THE **SESSION TIME** OF THOSE CASES TO BE **HIGH** AND ALSO **INCREASING** THE CHANCE OF CANCELLATION DUE TO HIGH TRAVELLING DISTANCE.

DENSITY



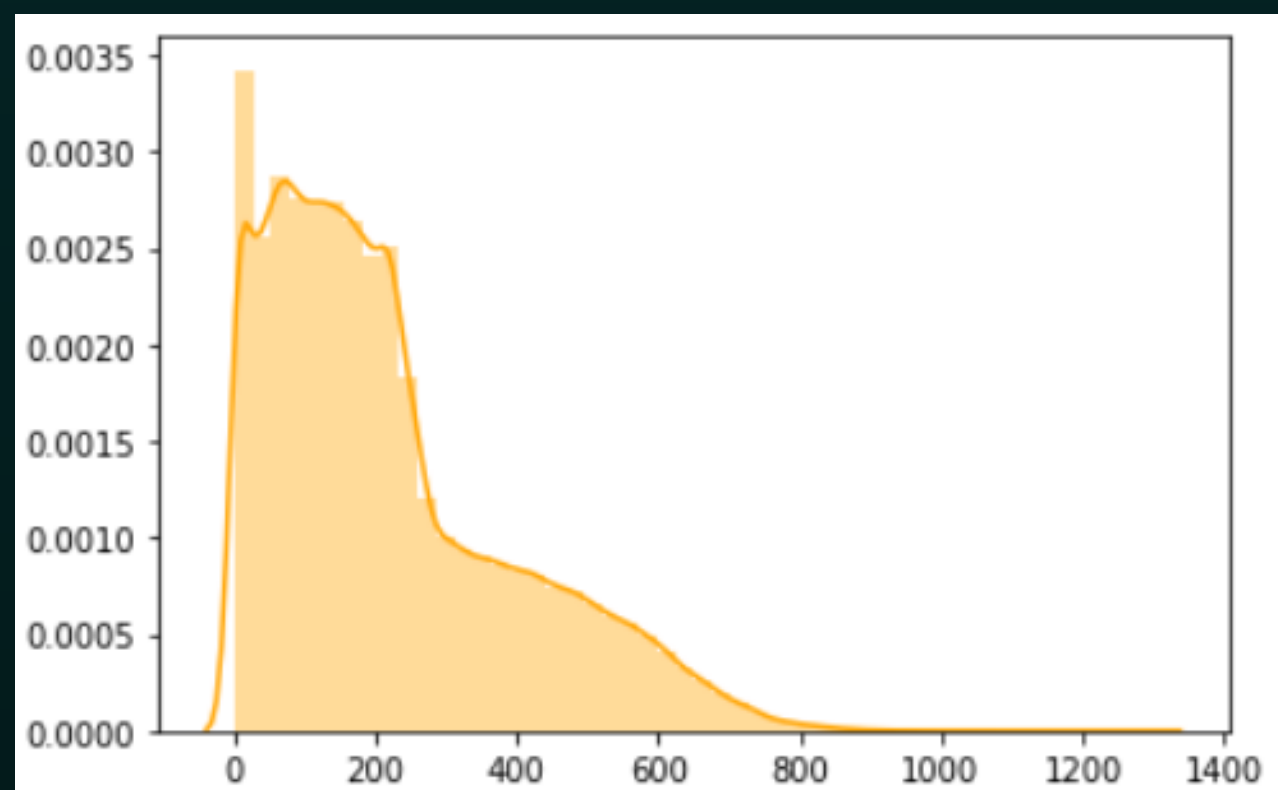
FIRST MILE DISTANCE

DENSITY



LAST MILE DISTANCE

DENSITY

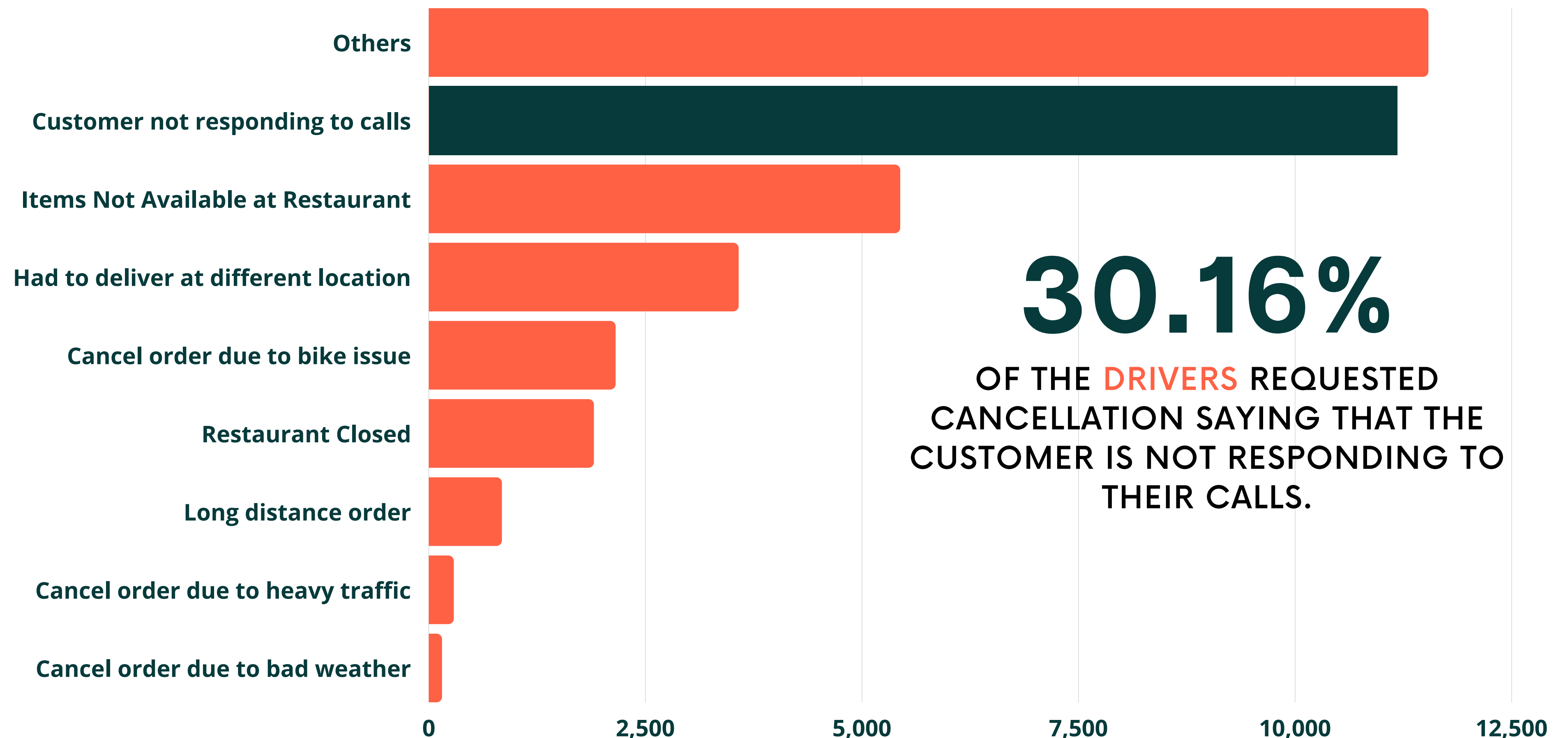


SESSION TIME

# CALL DATA

WE CAN SEE THAT MOST OF THE RIDERS ie., **11178** OF THEM HAS REQUESTED TO CANCEL THE ORDER SINCE THE **CUSTOMER WAS NOT ANSWERING THEIR CALLS.**

THERE ARE ONLY FEW CASES ie., **150** OF THEM, WHERE THE RIDER REQUESTED TO CANCEL THE ORDER **DUE TO BAD WEATHER CONDITIONS.**





WE CAN SEE THAT **AVERAGE SESSION TIME IN JANUARY** IS MORE WHEN COMPARED TO AVERAGE SESSION TIME IN **FEBRUARY**.

**AVERAGE FIRST MILE DISTANCE IS LESS THAN AVERAGE LAST MILE DISTANCE.** THIS INDICATES THAT RIDERS LOCATION IS NEARER TO PICKUP LOCATION WHEN COMPARED TO DELIVERY LOCATION

# POWER BI DASHBOARD

WE CAN SEE THAT THE RIDER WITH ID **190** HAS HIGHEST NUMBER OF DELIVERED ORDERS.

HE DELIVERED **562 ORDERS** IN THE LAST **30 DAYS**.

2716

IS THE ID OF THE RIDER WHO HAS **HIGHEST MAXIMUM LIFE TIME ORDER COUNT**

3476

IS THE ID OF THE RIDER WHO HAS **HIGHEST UNDELIVERED ORDER COUNT**

## SHADOW FAX - ANALYSIS REPORT

TEAM TINY CODERS

2716

Best Rider Id

3476

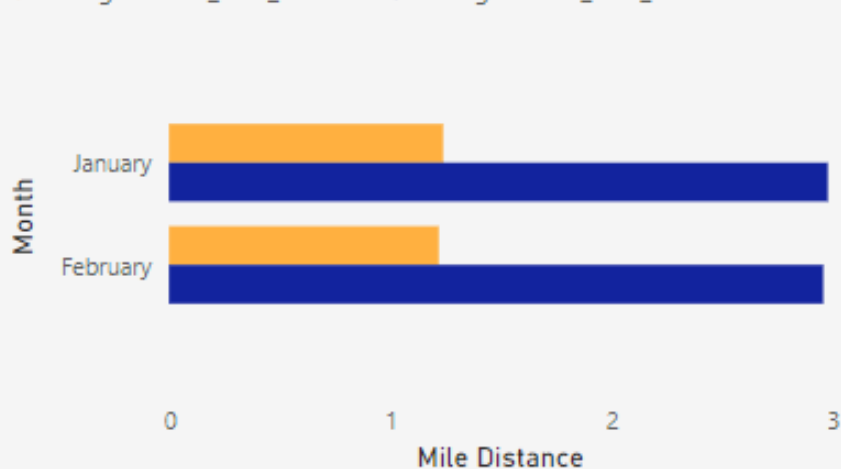
Least Rider Id

5218

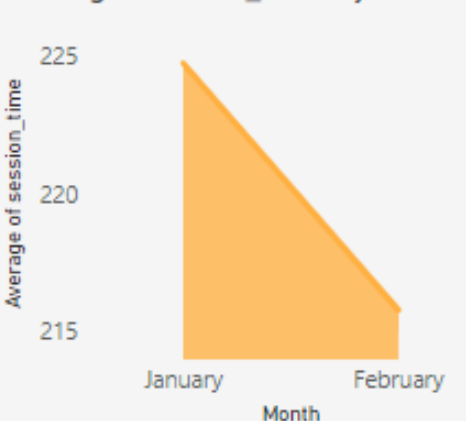
Cancelled Orders

Mile\_distance by Month

● Average of first\_mile\_distance ● Average of last\_mile\_distance



Average session\_time by Month



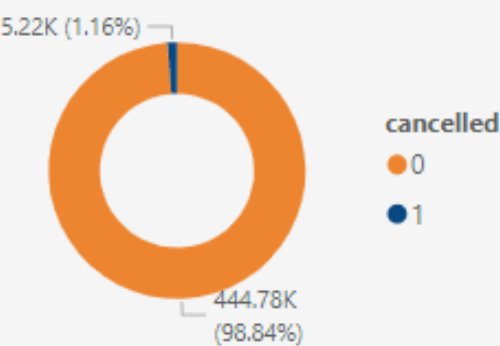
Top 10 Riders

rider_id	Max of lifetime_order_count
2716	30469
3771	27972
3581	26810
8424	24933
2382	24022
4087	23626
3391	22581
4311	22033
10528	21995
17959	21938

Bottom 10 Riders

rider_id	Max of undelivered_orders
3476	9.00
7273	9.00
15005	9.00
558	8.00
2012	8.00
2471	8.00
1266	7.00
1962	7.00
2318	7.00
18328	7.00

Delivered\_orders by rider\_id



# CONCLUSION

---

BASED ON THE ABOVE INSIGHTS, WE CAN SAY THAT **REASSIGNED ORDERS** ARE HIGHLY LIKELY TO BE CANCELLED AGAIN, WHEN COMPARED TO ORDERS WHICH ARE **NOT REASSIGNED**.

COMPANY CAN EXPECT HIGH NUMBER OF CANCELLATIONS ON **FRIDAY**.

WHEN ASSIGNMENT OF ORDER IS DONE, IF THE **LAST MILE DISTANCE** IS HIGH, THEN THERE'S A HIGH CHANCE OF THE DRIVER REQUESTING TO CANCEL THE ORDER.

ALLOTING ORDERS TO THE RIDERS WITH LESS **LIFE TIME ORDER COUNT** COULD RESULT IN CANCELLATION OF THE ORDER.

IF THE DRIVERS WITH HIGH NUMBER OF **UNDELIVERED ORDERS** ARE ASSIGNED ORDERS THEN PROBABILITY OF CANCELLATION IS HIGH.



# APPENDIX

---

**NOTE** - ABOVE INSIGHTS ARE A MIX OF INSIGHTS DERIVED FROM DATASET BOTH AFTER AND BEFORE PRE-PROCESSING AND DATA CLEANING, KINDLY REFER ROUND 2 NOTEBOOK FOR ANY ASSUMPTIONS IF NEEDED.

FOR PANDAS PROFILE REPORT AND POWER BI DASHBOARD FILES CLICK [\*\*HERE\*\*](#).