

# CASCADE CUP 2022

## Round 3



A Data Analysis Report

on

**“Rider-Driven Cancellation Problem”**

By

**Team :** Blaze\_Warriors

Vivek Rai  
Deepanshu Raj



# INDEX

→ Introduction	3
→ Train Data Analysis	4
◆ Cancellation Statistics	4
• Overall Cancellation Statistics	4
• Cancellation w.r.t Order Dates	4
• Cancellation w.r.t Order Time	5
◆ Order Specific Analysis	5
• Cancellation w.r.t First Mile Distance	5
• Cancellation w.r.t Last Mile Distance	6
◆ Rider Specific Analysis	6
• Allotted Order Statistics	6
• Cancellation w.r.t Undelivered Order Value	6
• Cancellation w.r.t Rider's Order number for the Day	7
◆ Time Delta Analysis	7
• Cancellation w.r.t Time Delta 1	8
• Time Delta ~45 mins Bins	8
• Cancellation w.r.t Time Delta 4	9
→ Call Data Analysis	9
◆ Call Data's Cancellation Distribution	9
◆ Reason Contributions for Order Cancellation	10
→ Conclusion	10



## Introduction

Shadowfax is India's largest hyperlocal delivery platform. Their aim is to build the fastest and most reliable logistics network by empowering a million micro entrepreneurs through technology to deliver anything, anywhere. It also helps in delivering food orders from clients such as Swiggy and Zomato to customers. In the process of delivering many times riders cancel the delivery which impacts company growth and value between customers.

The data provided to us by the company consists of two files :

### 1. **train\_data.csv** :

This file contains one row for each order created in a specific time-frame. Some points about this data are -

- a. It includes 449999 unique orders over 12 days from 26th January, 2021 to 6th February, 2021.
- b. Total number of distinct riders delivering the orders is 19537.
- c. 13753 out of total orders are reassigned orders.

### 2. **call\_data.csv** :

This file contains one row for each call placed by the rider (whether to the customer or to our call centre). Some points about this data are -

- a. It includes 562624 calls placed from 20327 riders to customers or call centres for 384228 orders.
- b. 525563 out of total calls were placed to customers and 37061 were placed to support / call centres.

The Data Analysis part is divided into two sections :

- a. Train Data Analysis
- b. Call Data Analysis

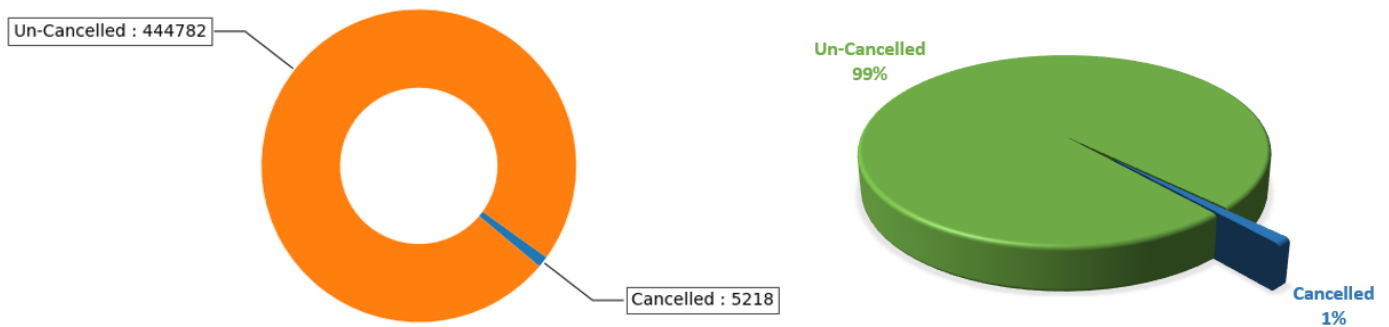


## Train Data Analysis



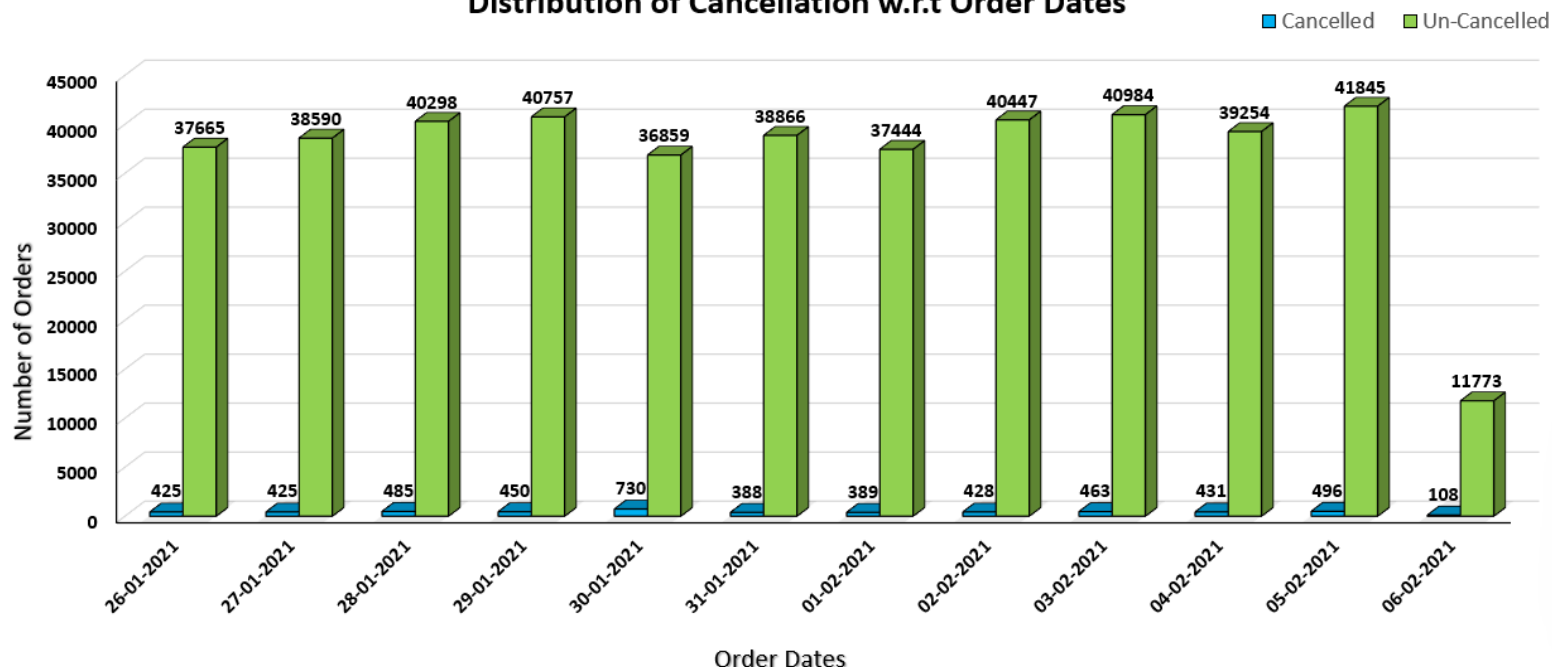
### Cancellation Statistics

#### OVERALL CANCELLATION STATISTICS



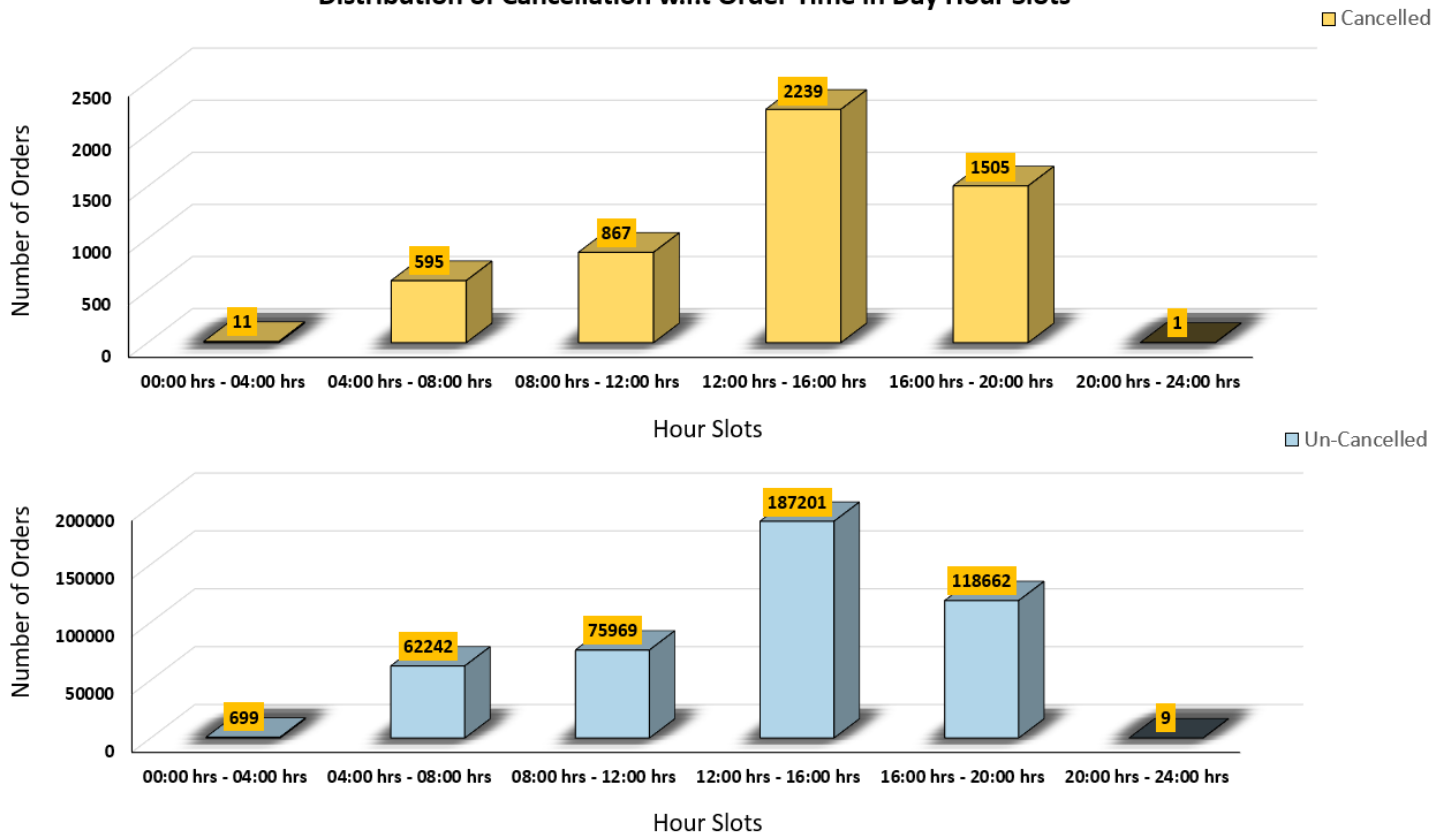
As can be seen in the above Pie Chart; the Rider-Driven Cancellation Problem's Dataset, is skewed to the ratio of 1:100 (approx.), in favour of Cancelled Orders. This can be interpreted as : **1 out of every 100 orders get cancelled due to the Rider-Driven Cancellation Problem.**

#### Distribution of Cancellation w.r.t Order Dates



Above is the Distribution of Orders across all the data points in the dataset, divided in bins of Order date on which the order was placed. The distribution of both the classes remained almost equal throughout for the amount of data given. **This can be interpreted as the cancellation issue remains the same throughout the service time, irrespective of the month-end or days of the week.**

Distribution of Cancellation w.r.t Order Time in Day Hour Slots

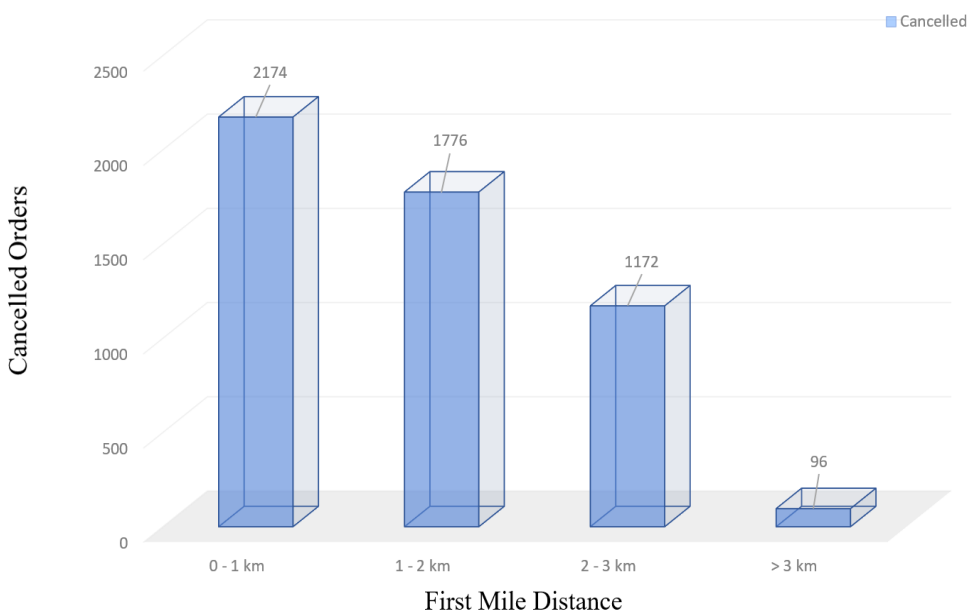


Above is the distribution of orders across the 24 hours time slot of the Day. As can be seen from the above distribution, Order frequency is at its peak during the 12:00 pm - 4:00 pm & 4:00 pm - 8:00 pm time slots of the day. **To make the service face less cancellation, Rider(s) availability frequency/count can be increased from 12:00 pm - 8:00 pm.**



## Order Specific Analysis

Number of Cancelled Orders w.r.t First Mile Distance



The chart to the left is the distribution of Cancelled orders w.r.t the First Mile distance (Distance of Rider from the Pick-up Location).

Following can be inferred from the Chart :

→ First mile distance distribution with order cancellation does not clearly resemble that, the distance between Rider's location and Order's pickup location has any relevance on the Order Cancellation.

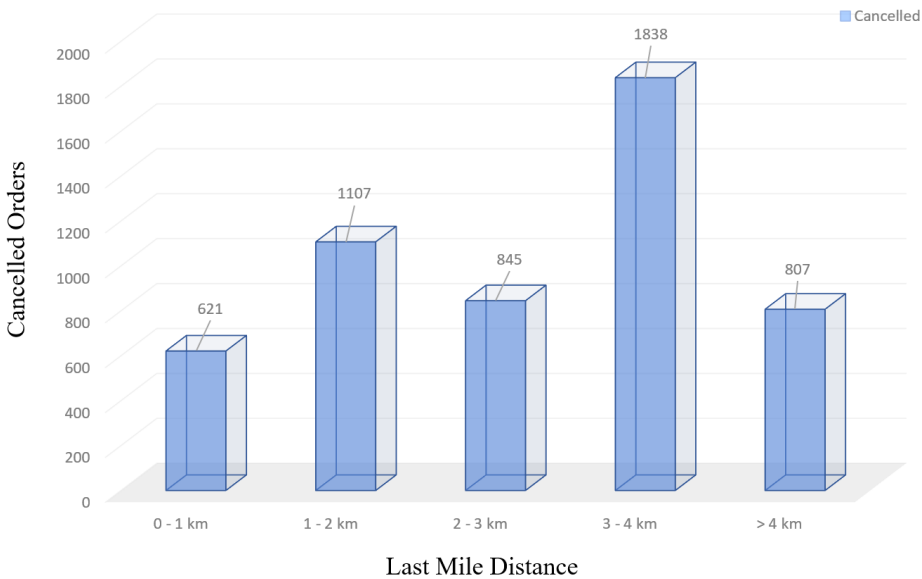
→ For the same reason, Last Mile Distance( Distance of delivery location from pickup location), can be crucial in identifying Rider-Driven Cancellation Problem better.

The Chart on the right, is the distribution of cancelled order w.r.t the Last Mile Distance (Distance of delivery Location from the pick-up location).

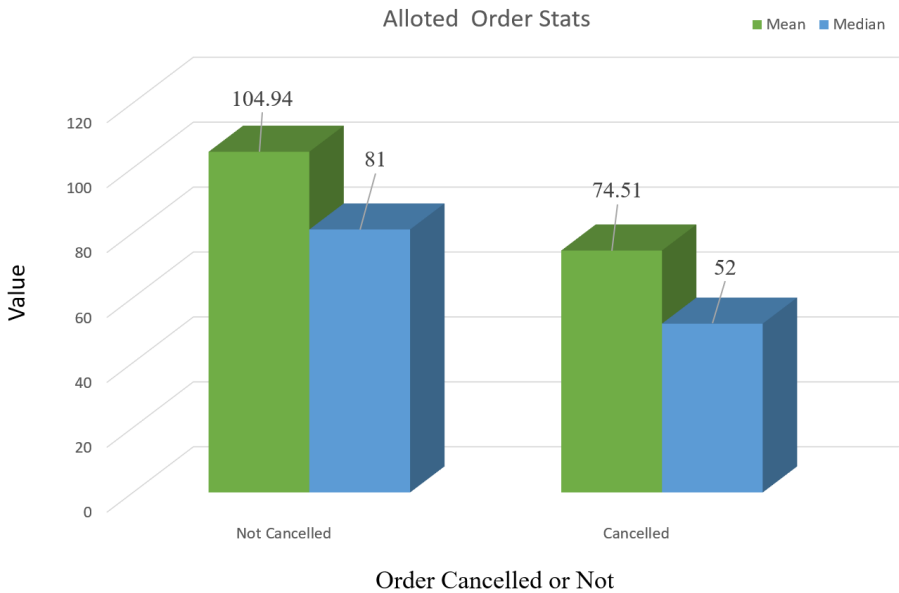
As can be clearly seen from the chart, **50%, (~2645) of the total cancelled Orders got cancelled due to a rider driven problem where the last mile distance of order was  $\geq 3$  km.**

A solution to this could be to **incentivise the rider, for order where Last mile distance  $\geq 3$ km. Adding up perks to the rider's bucket, might help solve the problem of rider-driven cancellation for long-distance orders.**

Number of Cancelled Orders w.r.t Last Mile Distance



### Rider Specific Analysis

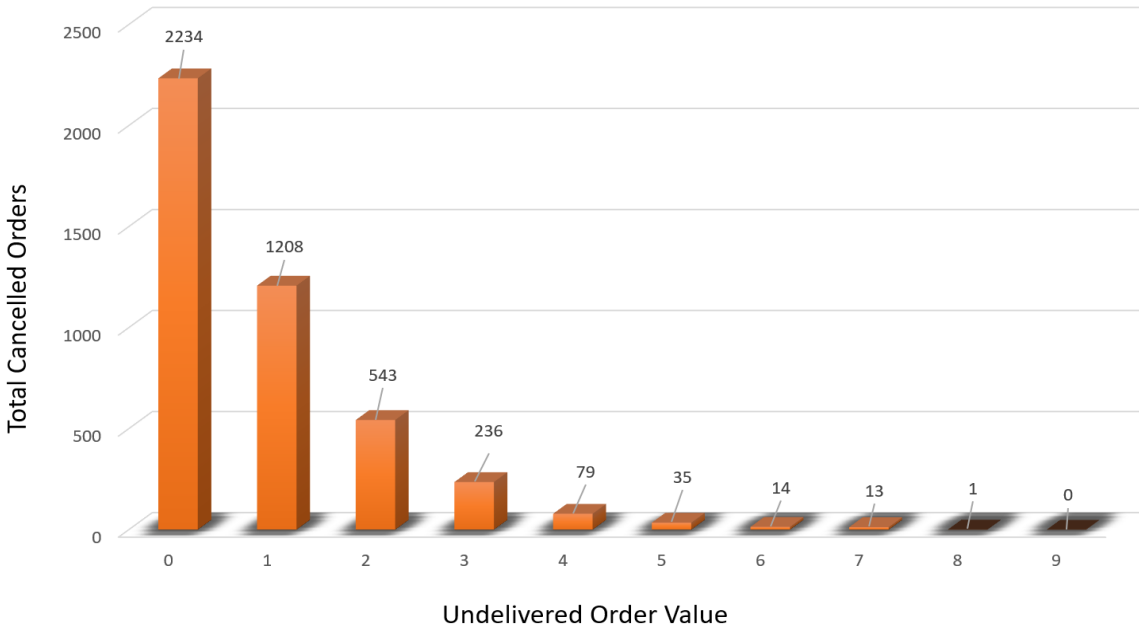


The chart on the left, describes the mean and median number in the Sequence of Number of Allotted Order per Rider given in the Dataset w.r.t Order Cancellation.

From the chart, it's obvious that **both mean and median of Alloted orders for Riders whose order(s) don't get cancelled > those whose order gets cancelled.**

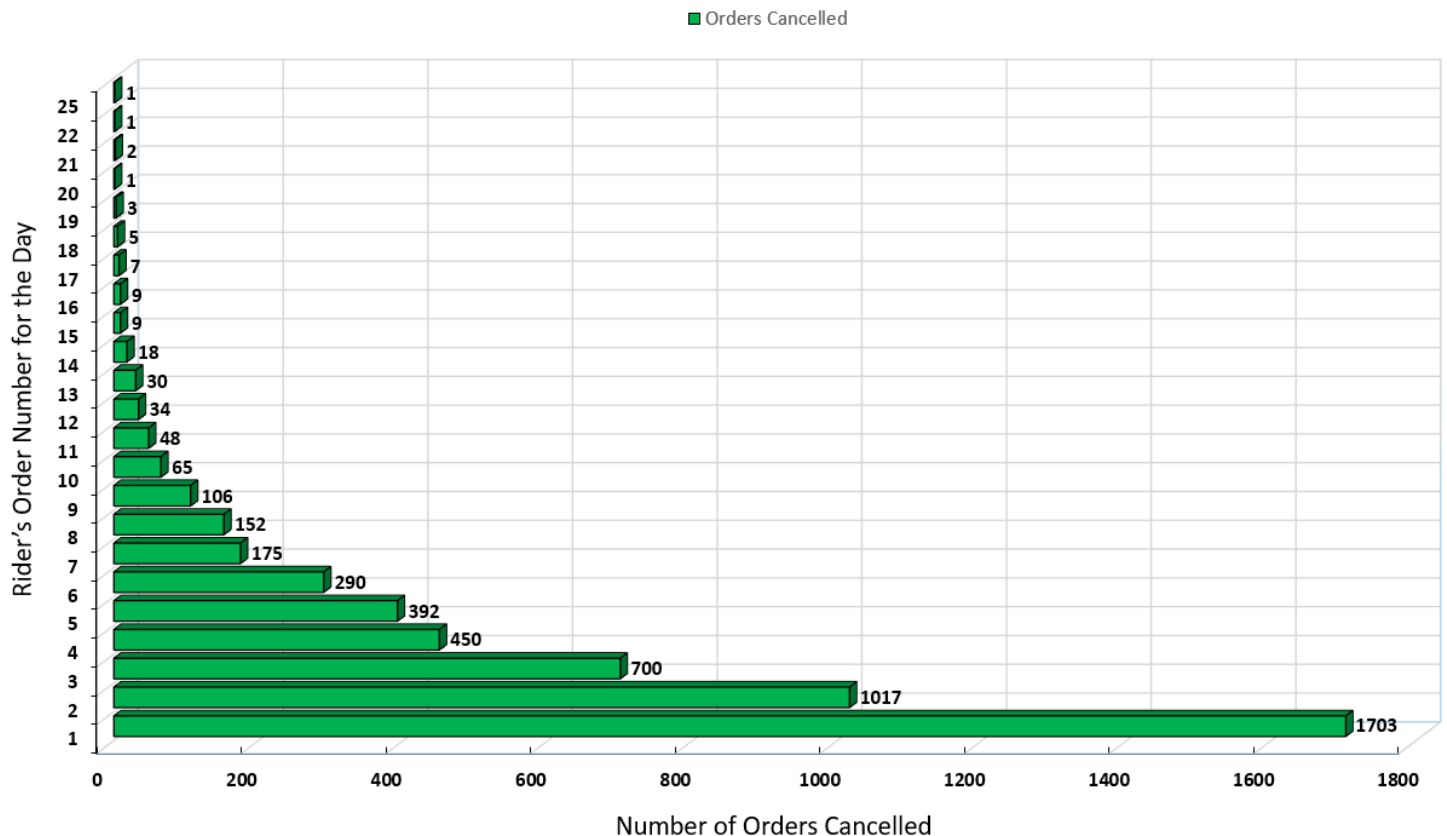
**Hence, it can be inferred that a Rider with less number of Allotted orders, is more likely to cancel the Order.**

Number of Cancelled Orders w.r.t. Undelivered Order Value



The above chart represents the number of cancelled order w.r.t undelivered order count of a Rider. From the above chart, it can be inferred that, **the rider with less number of undelivered order count, is more likely to cancel the order**. A solution to this could be, **create a streak-based model with rewards for those riders, who have the highest streak of no-undelivered order; on weekly and/or monthly basis**.

**Order Cancellation Distribution w.r.t Order Number For the Day**



Above chart is the **distribution of cancelled orders w.r.t the Order number of the day, for the rider that cancelled the order**. The chart was prepared to **correlate the workload on the rider with the orders getting cancelled by the same**. As an inference, it can be clearly seen that **workload does not make the rider cancel the order**.



## Time Delta Analysis

Time Delta is an additional feature which is derived from time features in the delivery service.

Time Features are :

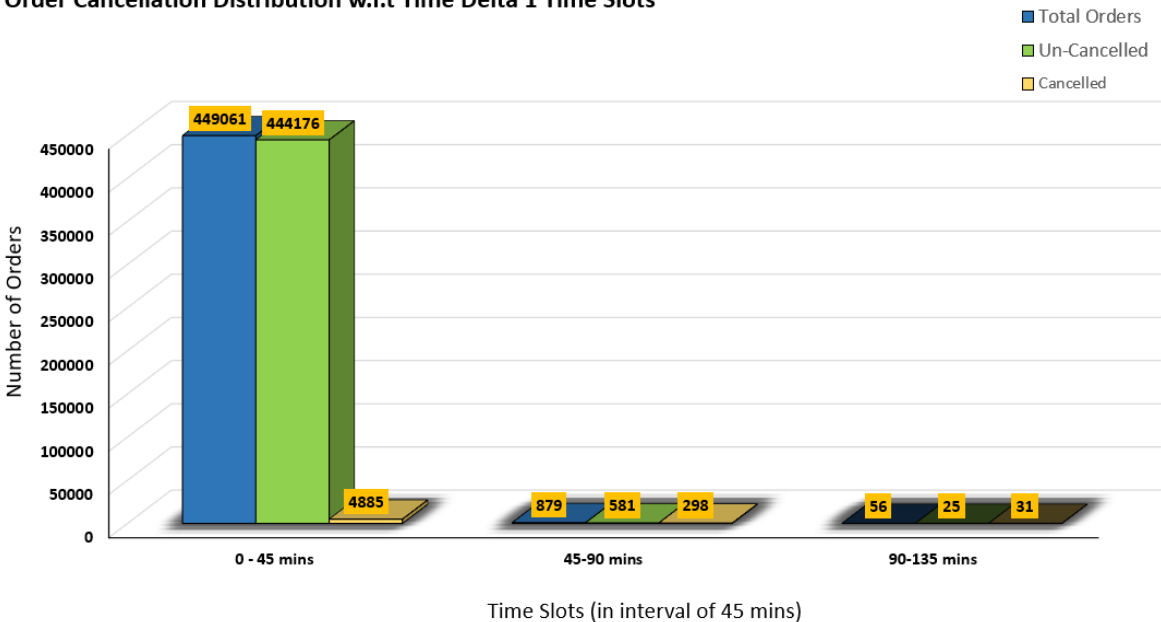
- Order Time
- Allot Time
- Accept Time
- Pickup Time

**Time Delta(s) are the specific difference taken between these delivery time service features.**

- **Time Delta 1** = Allot Time - Order Time
- **Time Delta 2** = Accept Time - Allot Time
- **Time Delta 3** = Pickup Time - Accept Time
- **Time Delta 4** = Pickup Time - Order Time

**Time Delta 1 :** Infers how long it takes the Service Provider to Allot an order to a Rider.  
**Time Delta 4 :** Infers the total time gap between Order time and Rider picking it up from the pickup location.

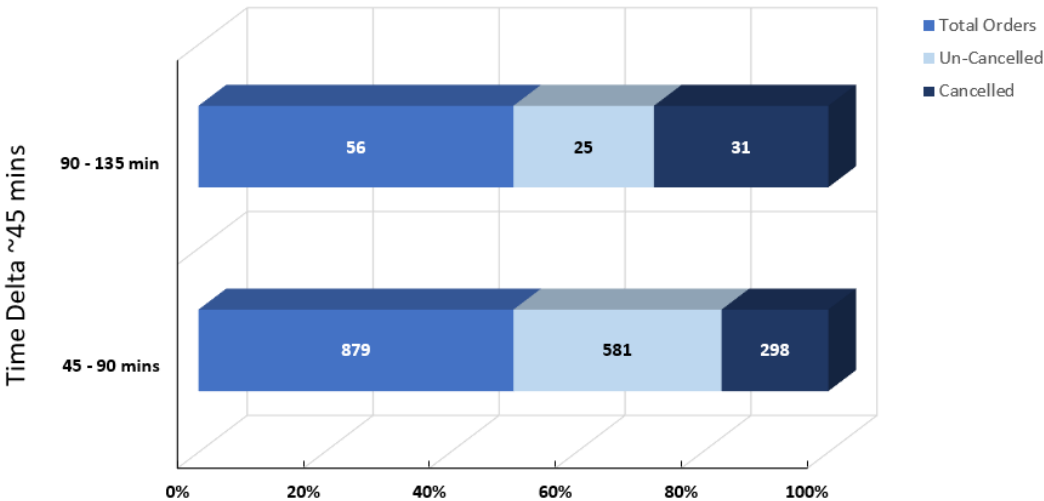
Order Cancellation Distribution w.r.t Time Delta 1 Time Slots



The plot to the left is the distribution of Orders (Total, Un-Cancelled & Cancelled) w.r.t Time Delta 1 in the time intervals of 45 mins. **Most of the Orders are allotted a rider within 45 mins duration by the Restaurant/service provider.** Hence, to draw an insight on cancellation's correlation with Delay in Rider Allotment, we need to zoom in on the 45-90 mins & 90-135 mins bins.

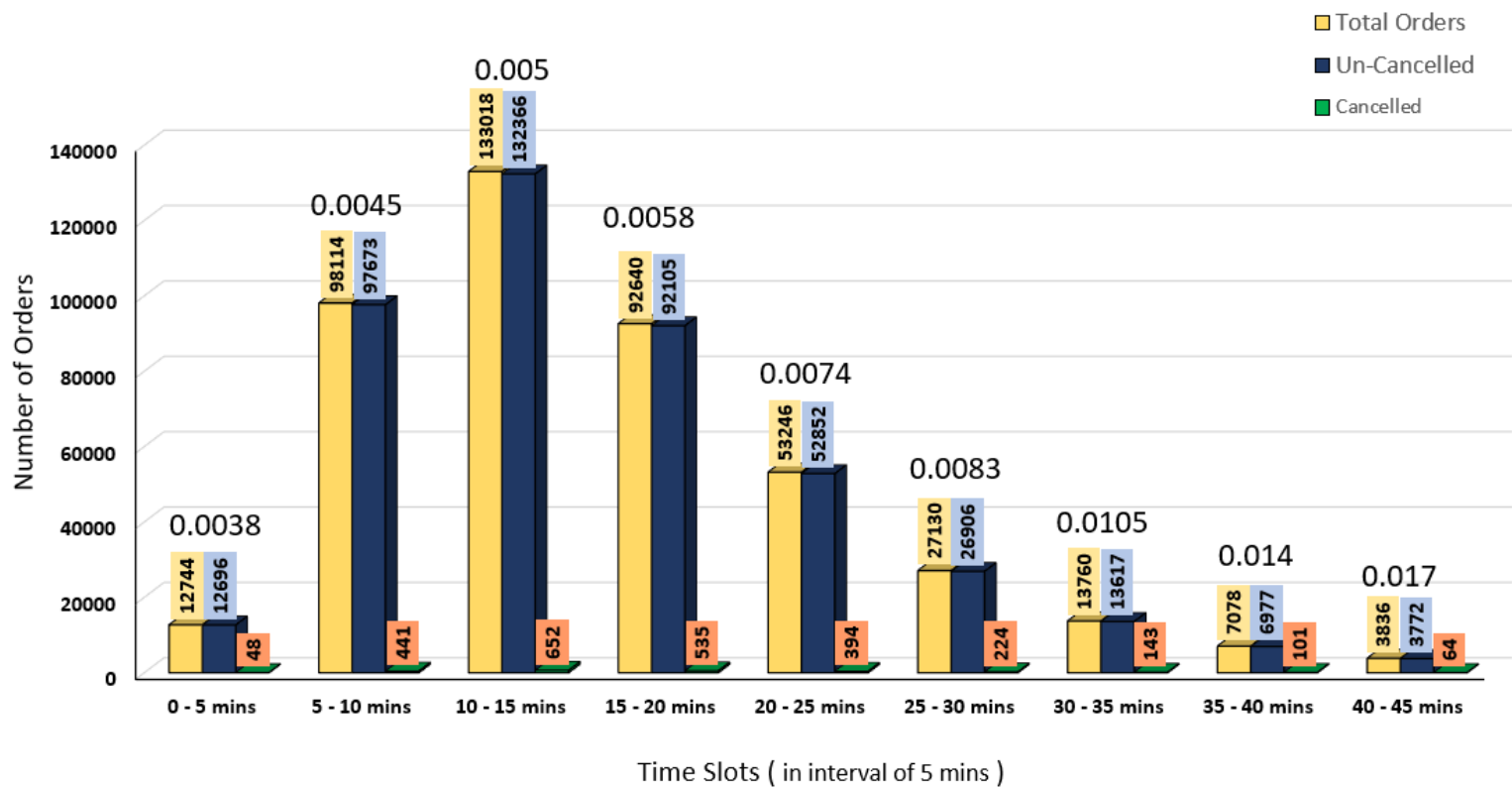
Order Cancellation Distribution w.r.t Time Delta 1 Time Slots

As can be seen in the chart on the right; **As the delay in the Rider allotment to an order increases; the Ratio of Cancelled to Un-Cancelled order goes up from 0.51 298:581 to 1.24 (31:25), which clearly indicates that, the larger the delay in Rider allotment to an order, the more likely the order is to get cancelled by the Rider.**





## Order Cancellation Distribution w.r.t Time Delta 4 Time Slots



Above is the distribution of Orders(Total, Un-Cancelled & Cancelled) w.r.t Time Delta 4 in the time intervals of 5 mins. Time Delta 4 represents the Time delay between the Order Time and Time at which the Rider Picks up the orders from Restaurant/Service Provider. **The floats mentioned above the bars is the Ratio of Cancelled to Un-Cancelled Orders in the corresponding time slot.** As can be clearly seen, with increase in Time delay; this Ratio goes up. As a matter of fact, **it is 4.5x more likely that an order will get cancelled by a rider, if the time delay between Order time and pickup time  $\geq$  40 mins than when the delay is  $\leq$  5 mins.**



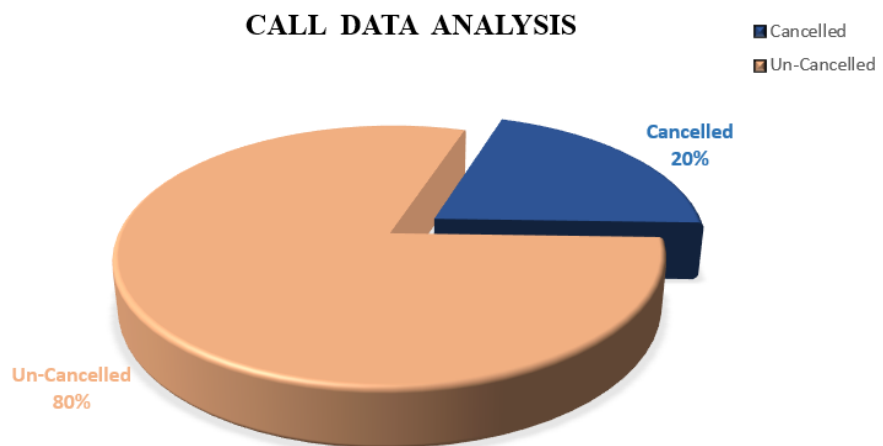
## Call Data Analysis



### Call Data's Cancellation Distribution

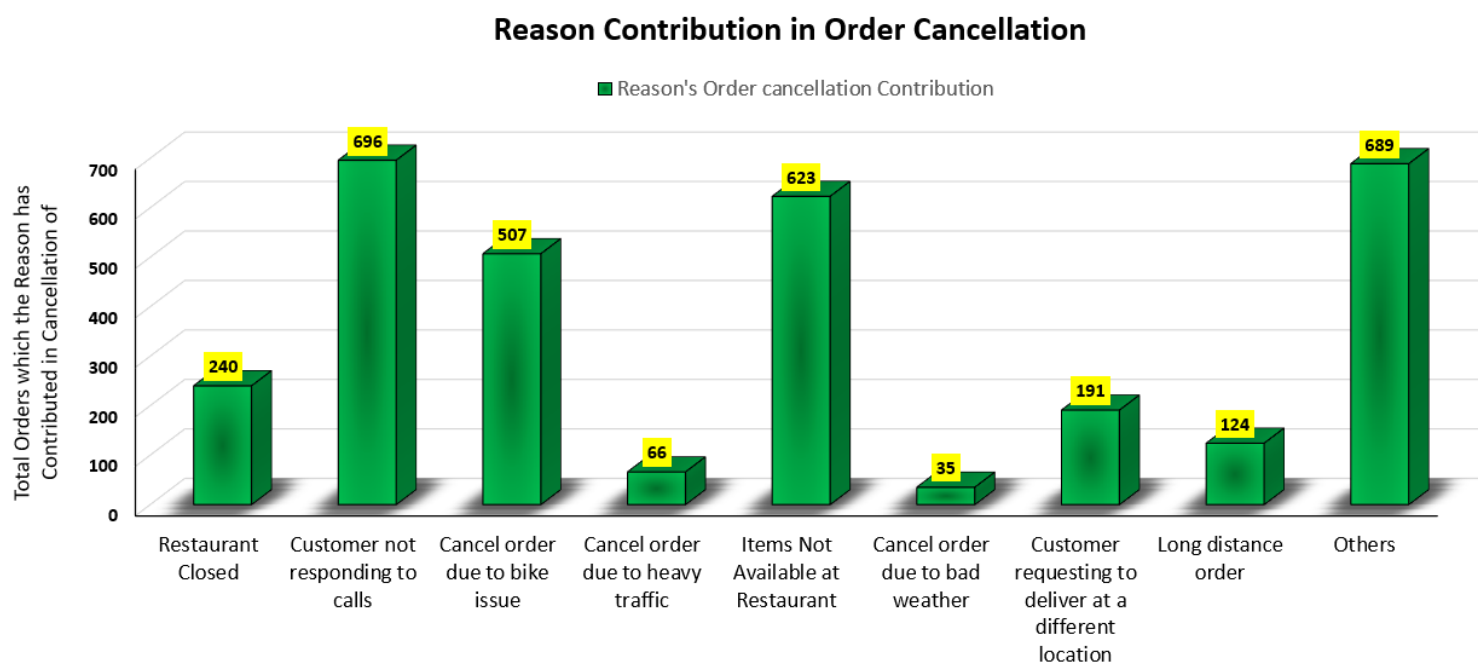
From the data available, and for the orders that are in sync between **call\_data** and **train\_data**'s, "order\_id" attribute; following are the stats :

1. Call\_data present & Cancelled  
20% ~ (2685 instances)
2. Call\_data present & Un-cancelled  
80% ~(10424 instances)





## Reason Contributions for Order Cancellation



Above is a Distribution Chart for the Order Cancellation Reason given as Feedback/Call\_data.

The Top 3 Reasons given to cancel the Order are :

- **Customer not responding to calls. [ 21.94% ]**
- **Items Not Available at Restaurant. [ 19.64% ]**
- **Cancel Order due to bike issue. [ 16% ]**

It can be inferred that weather, traffic, customer issues and long distance orders do contribute to the orders getting cancelled. Apart from the aforementioned Reasons, **Items unavailability and Restaurant's Closure contributes to 27.2% Order Cancellation Cases, which is a major Setback.**



## Conclusion

The Rider-Driven Cancellation Data has been analysed back and forth, and all the relevant features, that we thought were meaningful and significant w.r.t the Cancellation of an order has been put up in their respective section of analysis; along with visual representations and textual inferences and/or suggestions (if we could think of any), to curtail the net order cancellation number.

Hope, our work helps solve the real-world problem!!

**Thank You**