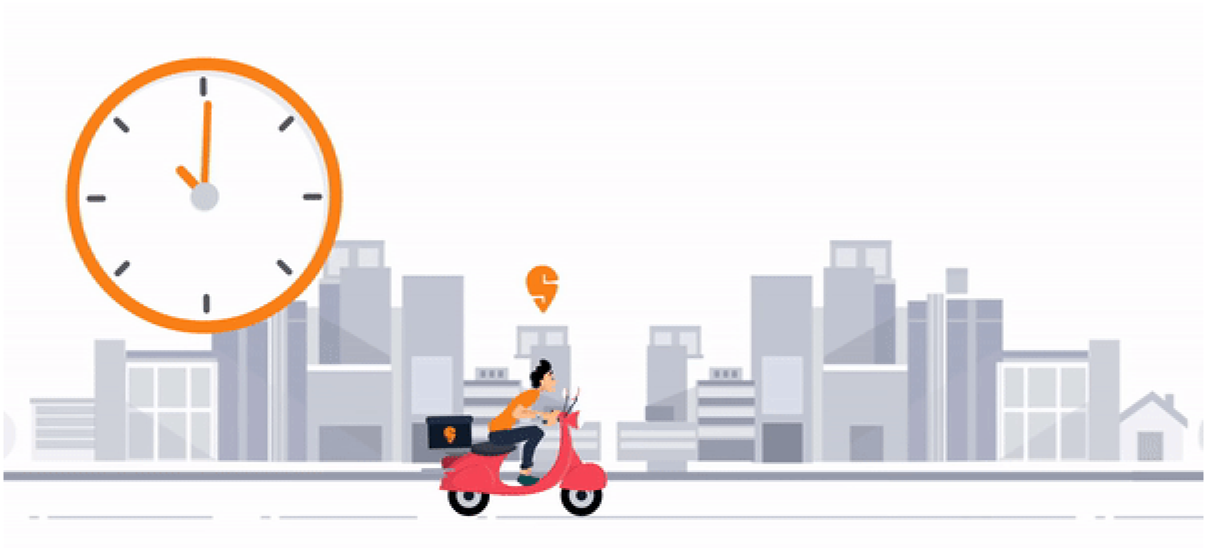
ANALYTICAL REPORT

Swiggy Delivery Challenge-Avoiding Rejects

🎯 Problem Statement:

To solve the hyperlocal delivery challenge of
“ Assigning the right delivery partners to the right set of orders at the right time”

Swiggy delivery partners can reject an order if they wish to. However, rejection of order increases the delivery time for the customer, hence **we want to avoid rejects**.



🏷️ Here are two sample datasets:

Assignment: Each row corresponds to an order assignment to the delivery partner.

Unique Identifier for Assignment

Column name	Description
ORDER_ID	Unique Identifier for the ORDER
DE_ID	Unique Identifier for the DE
ASSIGNMENT_START_TIME	Start Time of the Assignment
ASSIGNMENT_END_TIME	End Time of the Assignment
reject_ind	Whether this assignment was rejected
reject_type	Reject Type of this Assignment
PLACED_TIME	Order Placed Time
DELIVERED_TIME	Order Delivered Time
LASTMILE_DISTANCE	Distance to travel in Last Mile (from Restaurant to Customer)
FIRSTMILE_DISTANCE	Distance to travel in First Mile (from DE Assignment Location to Restaurant)
LAST_MILE_TIME_PREDICTED	Time prediction for the last mile
PAYOUT_MADE_TO_DE	Actual payout made to DE for this order

Column name	Description
NUM_PING_COUNT_LAST10MIN	# of pings received from DE device in last 10 minutes
LAST_PING_TIME_LAST10MIN	time of last the ping received from DE device (within last 10 minutes)
CUSTOMER_ZONE	Zone ID for the customer
CUSTOMER_LAT	Coordinates of the customer
CUSTOMER_LAT	Coordinates of the customer

Delivery Partners: Each row corresponds to a delivery partner.

Unique Identifier for Delivery Partner

Column name	Description
DE_ID	Unique Identifier for the DE
SHIFT_END_TIME	Shift end time for DE (in HH: MM)
DE_HOME_LAT	Home Location coordinate for the DE
DE_HOME_LNG	Home Location coordinate for the DE
DE_JOINING_DATE	Joining date of the DE
DE_ZONE_ID	Zone ID for the DE

▼ 🏷️ Additional details:

- Every 2 minutes, all the non-assigned orders are input into the assignment module. For each order, a set of DEs are evaluated, and the order is finally assigned one of these DEs.
- Delivery partners are allowed one reject per day, beyond which they are penalized.
- No payout is made to DE if he rejects an order
- Every instance of a DE reject, is stored in the production tables, as a unique entry.

🏷️ Approach ✍🏼

- ☐ Exploratory Data Analysis to visualize the datasets by the means of graphs to gain overall insights.
- ☐ Understanding the correlations between different data values from the graphs obtained.
- ☐ In-depth analysis of the datasets by the use of seaborn and matplotlib functions (Heatmap, cluster map, pair plot, bar plot, scatter plot, etc.)
- ☐ Cleaning the datasets to reduce the data redundancy and removing the null values by the use of Data Mining.
- ☐ Training ML models to predict the rejection rate on the basis of previous fields' values.
- ☐ Understanding the accuracy of several models like Logistic Regression, Linear Regression, and Linear Discriminant and determining their fit score for the datasets.

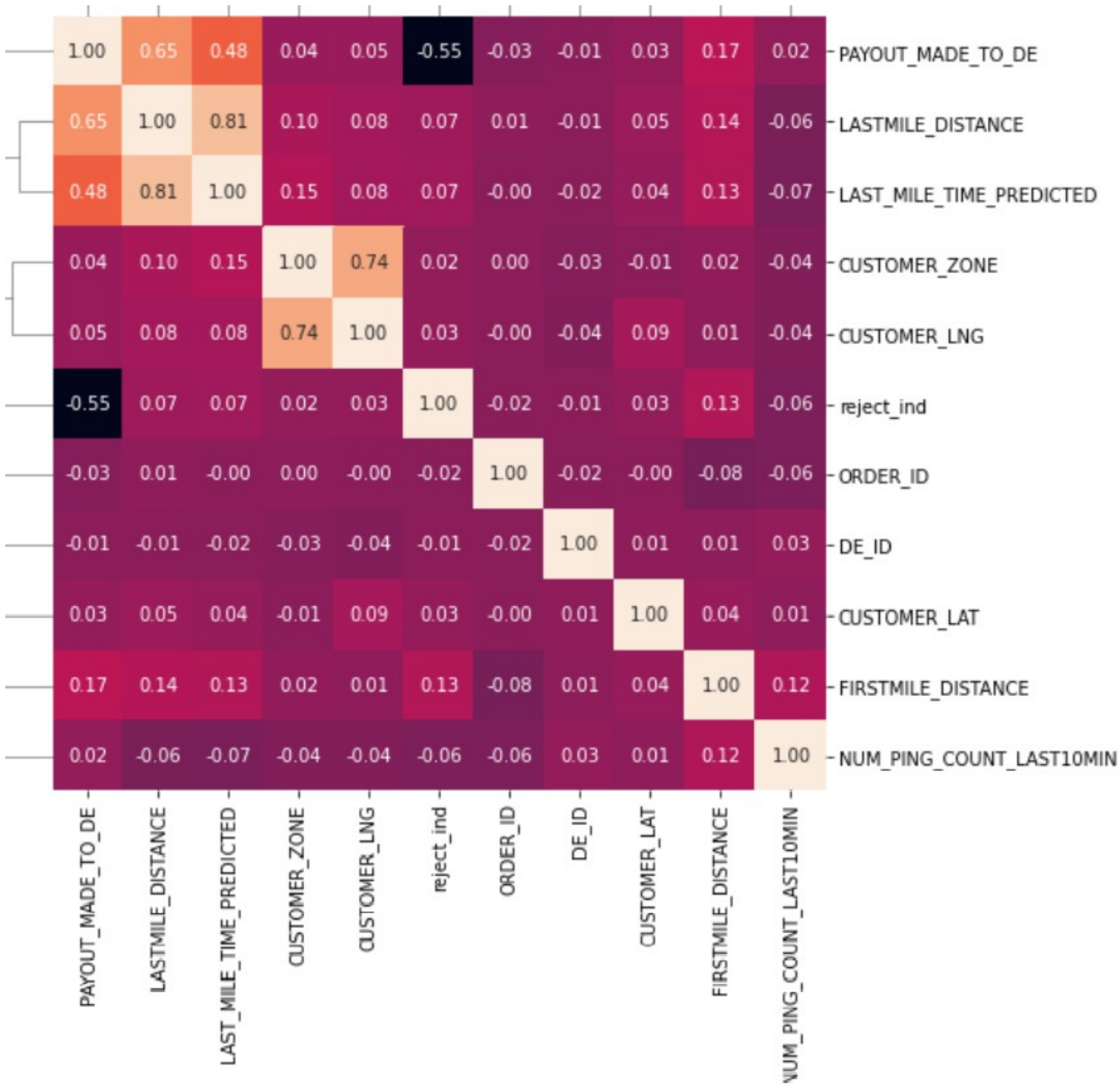
🏷️ Key Insights of the Datasets ⚡

- **The number of rows:** 132394
- **The number of columns:** 17
- **Data types of Identifiers:**

```
ORDER_ID                      int64
DE_ID                         int64
ASSIGNMENT_START_TIME        object
ASSIGNMENT_END_TIME          object
reject_ind                    int64
reject_type                  object
PLACED_TIME                  object
DELIVERED_TIME               object
LASTMILE_DISTANCE           float64
FIRSTMILE_DISTANCE          float64
LAST_MILE_TIME_PREDICTED    float64
PAYOUT_MADE_TO_DE           float64
NUM_PING_COUNT_LAST10MIN      int64
LAST_PING_TIME_LAST10MIN     object
CUSTOMER_ZONE               float64
CUSTOMER_LAT                float64
CUSTOMER_LNG                float64
dtype: object
```

🏷️ Outcomes of the Exploratory Data Analysis 🧿

Generic correlations b/w the attributes by means of Heat Map:

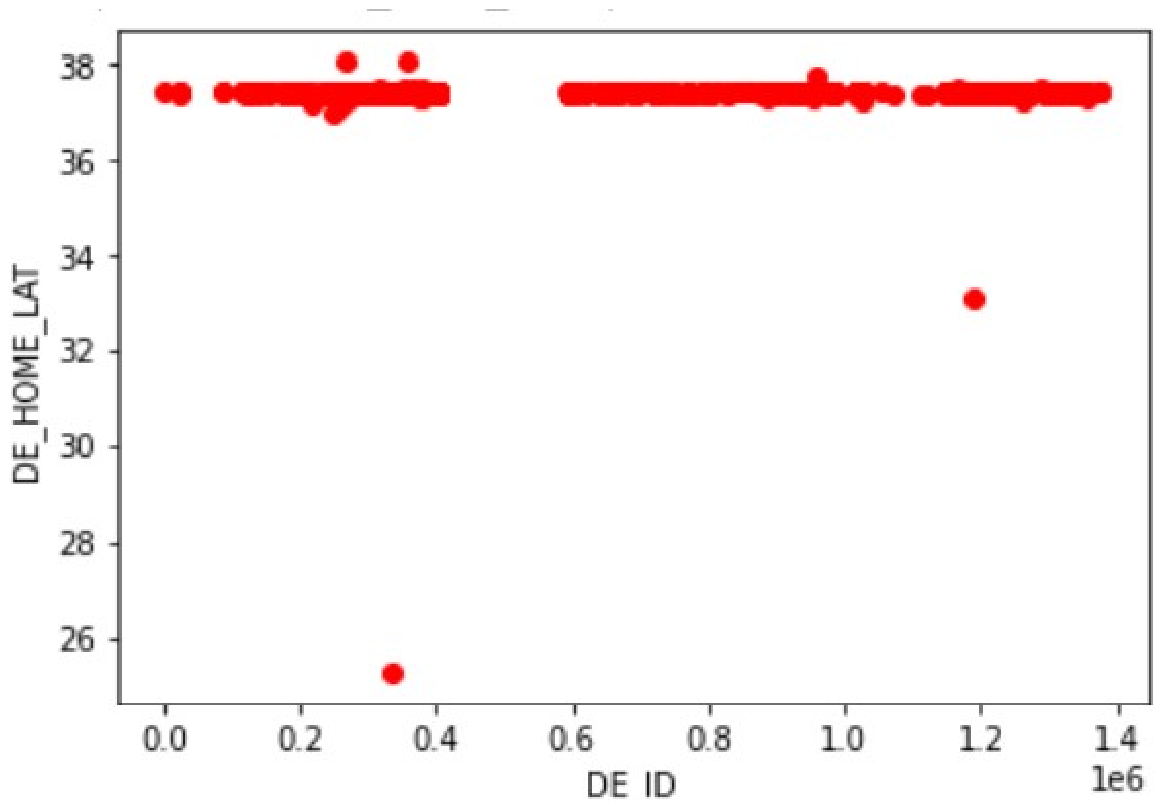


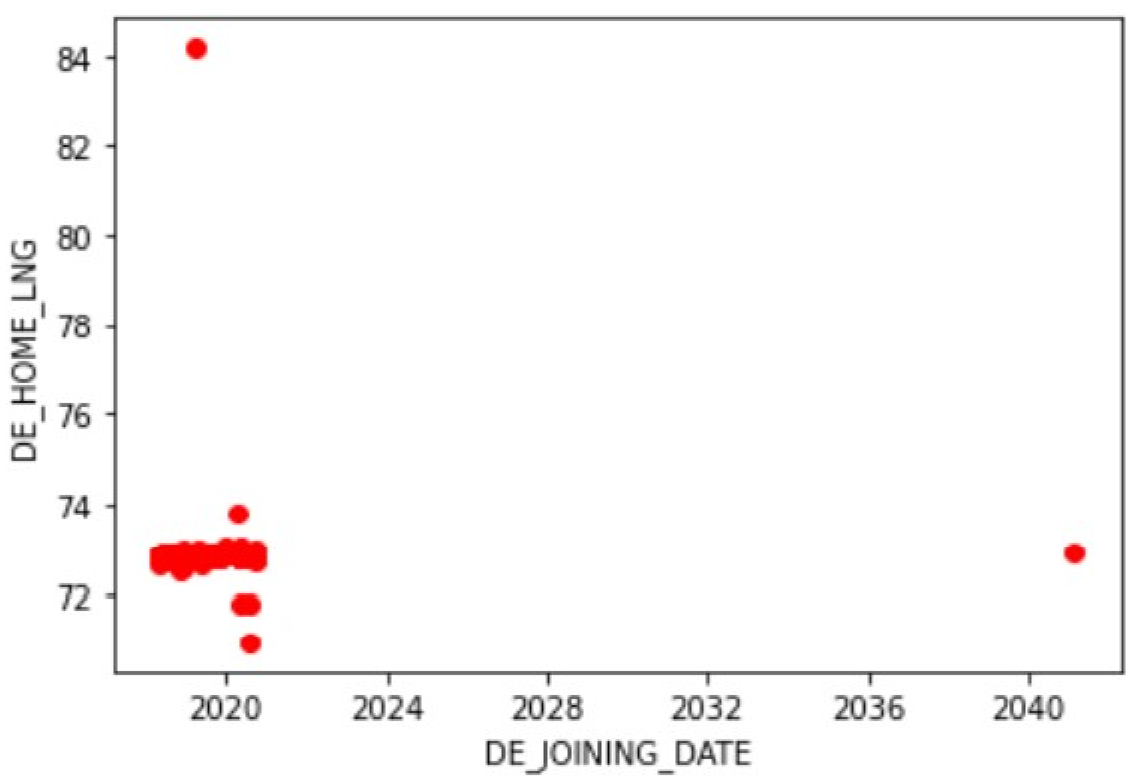
- There is a very obvious high negative relation b/w the payout made to the DE and the rejection rate.
- Positive correlation b/w first-mile distance and reject_ind. This shows that rejection increases with the increase in the first-mile distance.
- There is a negative correlation between b/w num_ping_count and reject_ind. This shows that the more updated service always leads to fewer cancellations.
- Negative correlation b/w last_mile dist and reject_ind. This shows that more is the distance traveled from the restaurant to the customer less is the rejection rate.

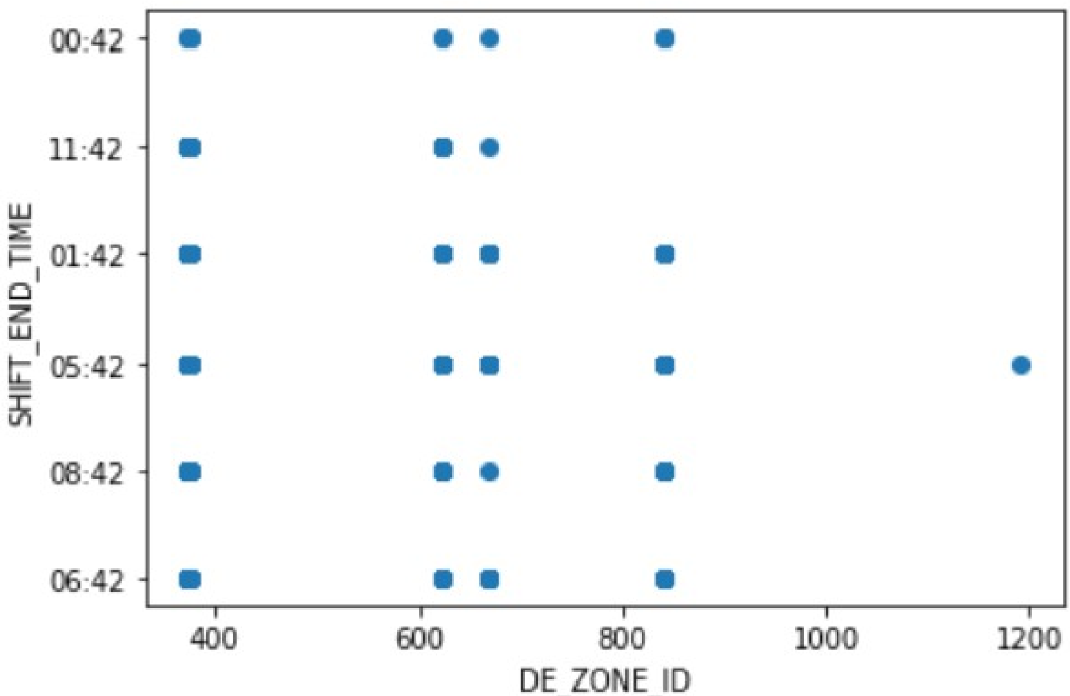
🏷️ Various ML Models used and their accuracy:

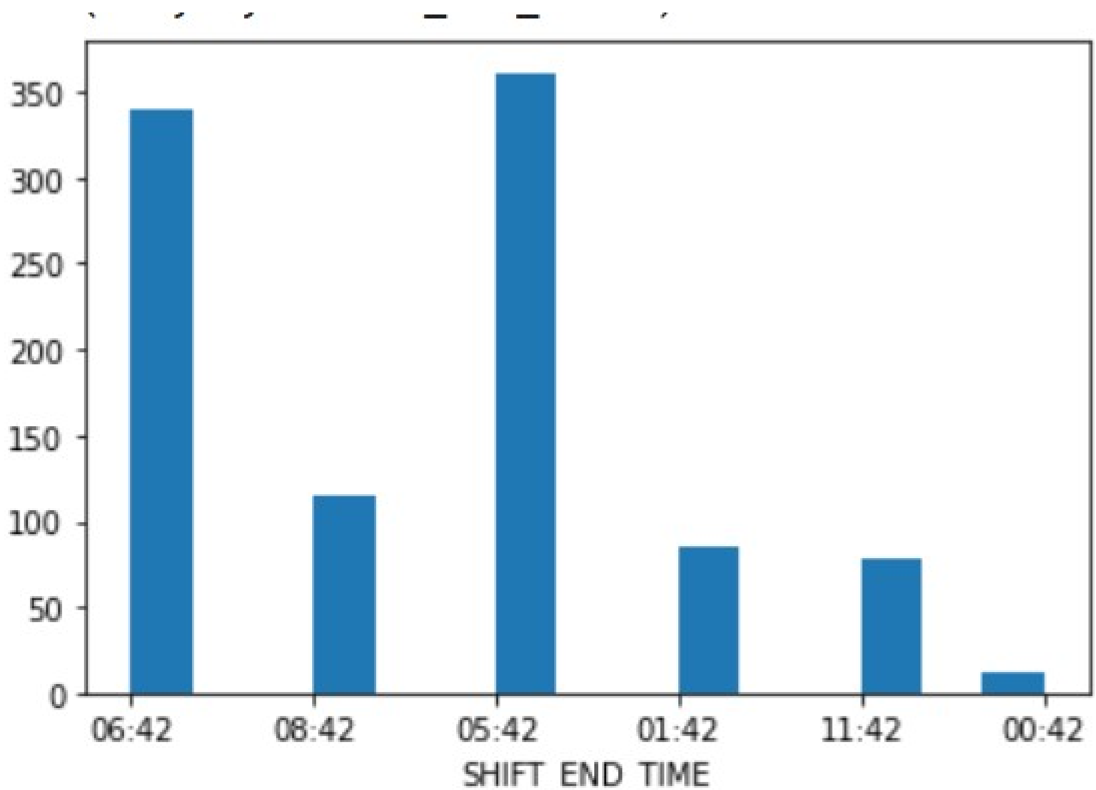
- The model used: **Linear regression**
    - The accuracy obtained: 46 %
- The model used: **Logistic Regression**
    - The accuracy obtained: 96.6%
- The model used: **Linear Discriminant Analysis**
    - The accuracy obtained: 93.78%
- R2 score for the model: 0.9660261910652639
    - (This tells us the relationship of the movements between the dependent and the independent variables)
    - A higher value of R2 in the model tells us that the proposed model is perfect for working on variability in data.

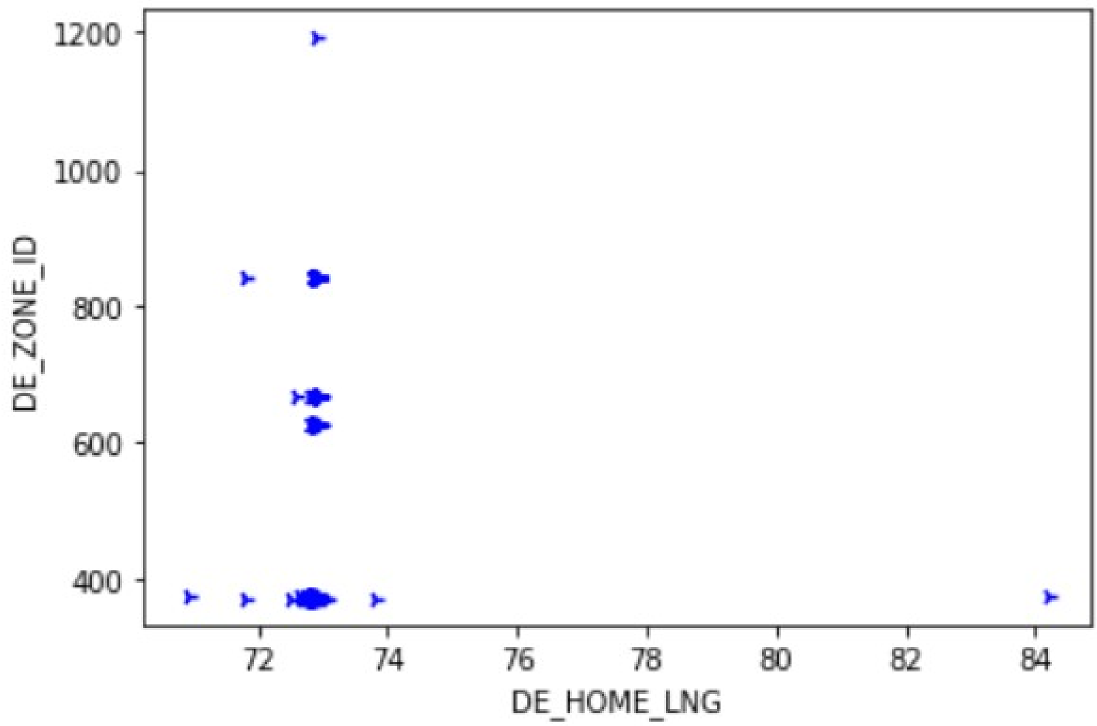
🏷️ Overall Analysis 📈


DE_HOME_LAT
38
36
34
32
30
28
26
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1e6
DE_ID


DE_HOME_LNG
84
82
80
78
76
74
72
2020
2024
2028
2032
2036
2040
DE_JOINING_DATE


00:42
11:42
01:42
05:42
08:42
06:42
SHIFT_END_TIME
400
600
800
1000
1200
DE_ZONE_ID


350
300
250
200
150
100
50
0
06:42
08:42
05:42
01:42
11:42
00:42
SHIFT_END_TIME



🏷️ Recommendation and suggestions

- On analysis, most of the delivery partners' longitude and latitudes is lying in the range of **72-74 and 36-38 degrees.** If the system automates the order in this range that will reduce the possibility of rejection.
- Most delivery partner **shift ends at 5:42 PM** which reduces the workforce and results in maximum order rejections.
- Traveling distances for the delivery partners should be reduced to ensure a lesser value of the rejection rate.
- More amounts should be leveraged in the form of payouts or tips for the delivery partners in order to ensure strong retention.
- We should try to recruit more people so that we become distance specific and the delivery partner does not have to travel large distances.
- The interface of the app should be made more location specific in order to ensure that the deliveries are only set up for nearby joints. Distances are a major hampering factor and should be minimized as much as possible.

- There are many reasons for a DE to reject an order, like payouts made, distances, odd timings, etc. The company should try to minimize these factors and work upon the reduction of rejection rates by understanding employee psychology. The preferences of the employees can be understood on the basis of open talks, conducting surveys, one-to-one interactions with employees, etc.

🏷️ Exploratory Data Analysis Code | Link

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/ce14caea-30dd-43e9-9d68-cab132a4bfca/Swiggy_Dataset_EDA_-_Colaboratory.pdf

🏷️ Model Code | Link

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/1731e87e-4b6a-499a-90e0-c74d6dddac65/Swiggy_Delivery_Dataset_Models_-_Colaboratory.pdf

🏷️ Resources

- Data Visualization : Link1 | Link2 | Link3 | Link4
- Linear Regression: Link1 | Link2
- Basics: Link1