Data Analysis Report

on

Food Time Dataset BIAnalysis

Prepared by

Saad Bin Manjur Adit

Contents

Dataset

File name: Food_Time_Data_Set_BIAnalysis.csv
Format: CSV

A Brief Information of Dataset:

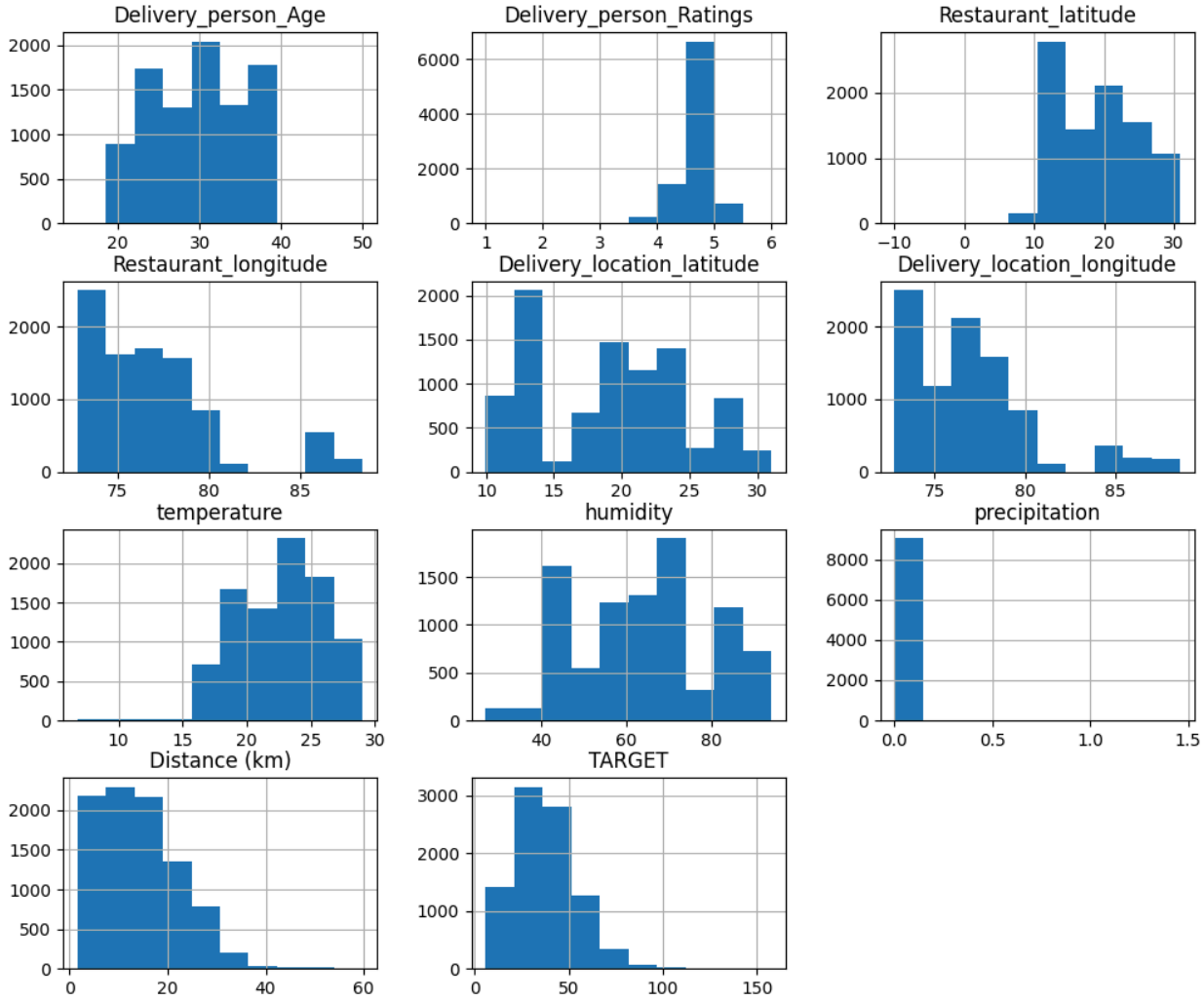
```
RangeIndex: 10001 entries, 0 to 10000
Data columns (total 18 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   ID                           10000 non-null  object
 1   Delivery_person_ID           10000 non-null  object
 2   Delivery_person_Age          10000 non-null  float64
 3   Delivery_person_Ratings      10000 non-null  float64
 4   Restaurant_latitude          10000 non-null  float64
 5   Restaurant_longitude         10000 non-null  float64
 6   Delivery_location_latitude   10000 non-null  float64
 7   Delivery_location_longitude  10000 non-null  float64
 8   Type_of_order                10000 non-null  object
 9   Type_of_vehicle              10000 non-null  object
 10  temperature                  9995 non-null   float64
 11  humidity                     9995 non-null   float64
 12  precipitation                9995 non-null   float64
 13  weather_description          9995 non-null   object
 14  Unnamed: 14                  0 non-null      float64
 15  Traffic_Level                9085 non-null   object
 16  Distance (km)                9080 non-null   float64
 17  TARGET                       9459 non-null   object
dtypes: float64(11), object(7)
```

After Cleaning:

```
RangeIndex: 9087 entries, 0 to 9086
Data columns (total 17 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   ID                           9087 non-null   object
 1   Delivery_person_ID           9087 non-null   object
 2   Delivery_person_Age          9087 non-null   float64
 3   Delivery_person_Ratings      9087 non-null   float64
 4   Restaurant_latitude          9087 non-null   float64
 5   Restaurant_longitude         9087 non-null   float64
 6   Delivery_location_latitude   9087 non-null   float64
 7   Delivery_location_longitude  9087 non-null   float64
 8   Type_of_order                9087 non-null   object
 9   Type_of_vehicle              9087 non-null   object
 10  temperature                  9087 non-null   float64
 11  humidity                     9087 non-null   float64
 12  precipitation                9087 non-null   float64
 13  weather_description          9082 non-null   object
 14  Traffic_Level                9085 non-null   object
 15  Distance (km)                9087 non-null   float64
 16  TARGET                       9087 non-null   float64
dtypes: float64(11), object(6)
```

Data Analysis

Distribution of Numerical Features



The above data distribution shows the delivery persons' age is between 20-40 and most of them are 4.5+ rated.
Restaurant and delivery location latitude and longitude data distribution is irrelevant here.
Temperature is between 15-30 and the maximum frequency is at 23-24 on the other hand humidity has the highest frequency at 60-70 and precipitation data is not accurate here where it stays less than 0.5.

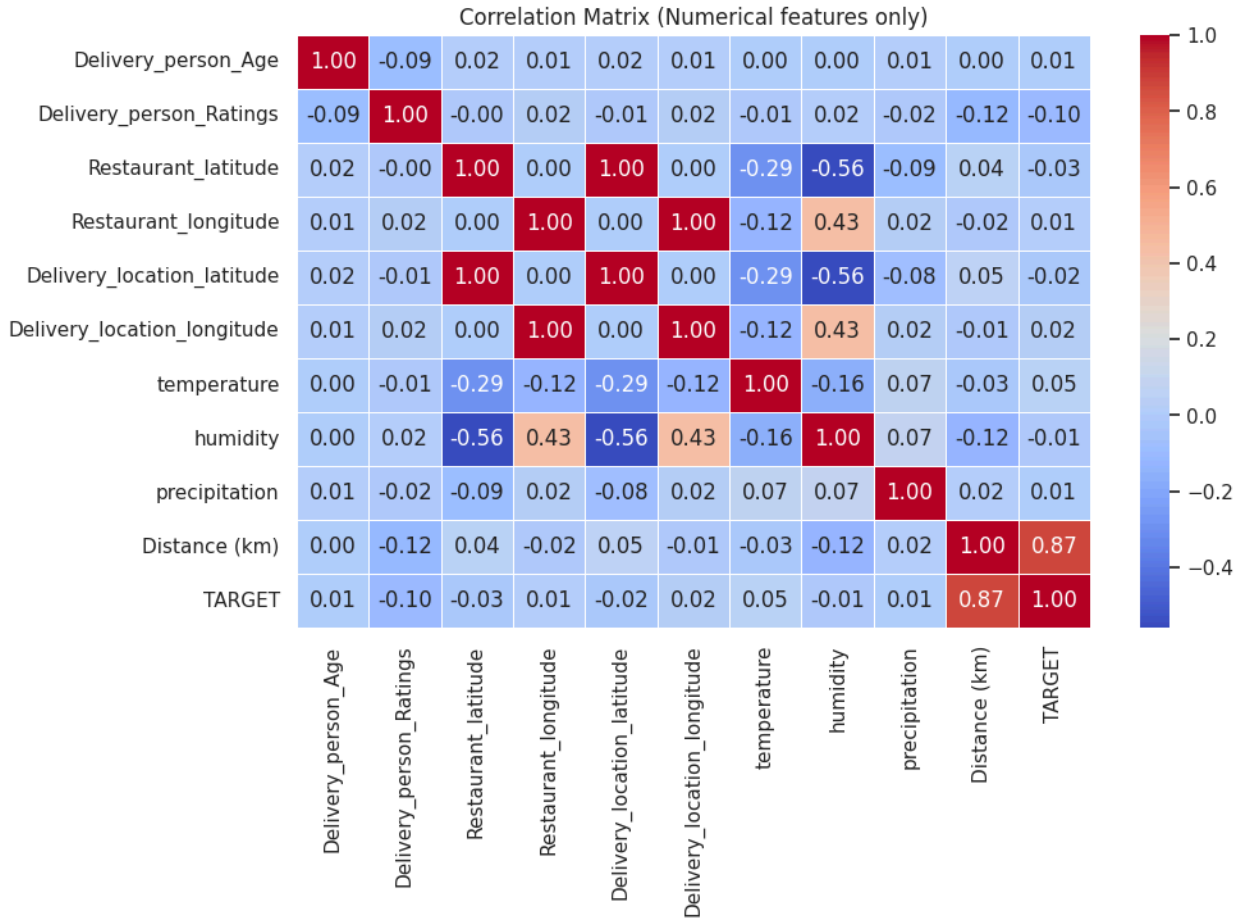
Distance (km) is between 0-35 has the maximum frequency. The target is the delivery time where the range is 5-150 minutes (probably because the unit is not mentioned). The most frequent delivery time is 25-50 minutes.

Unique Values of All Features

ID	9084
Delivery_person_ID	1135
Delivery_person_Age	22
Delivery_person_Ratings	28
Restaurant_latitude	389
Restaurant_longitude	388
Delivery_location_latitude	3394
Delivery_location_longitude	3394
Type_of_order	4
Type_of_vehicle	4
temperature	1062
humidity	65
precipitation	25
weather_description	11
Traffic_Level	5
Distance (km)	2344
TARGET	3389

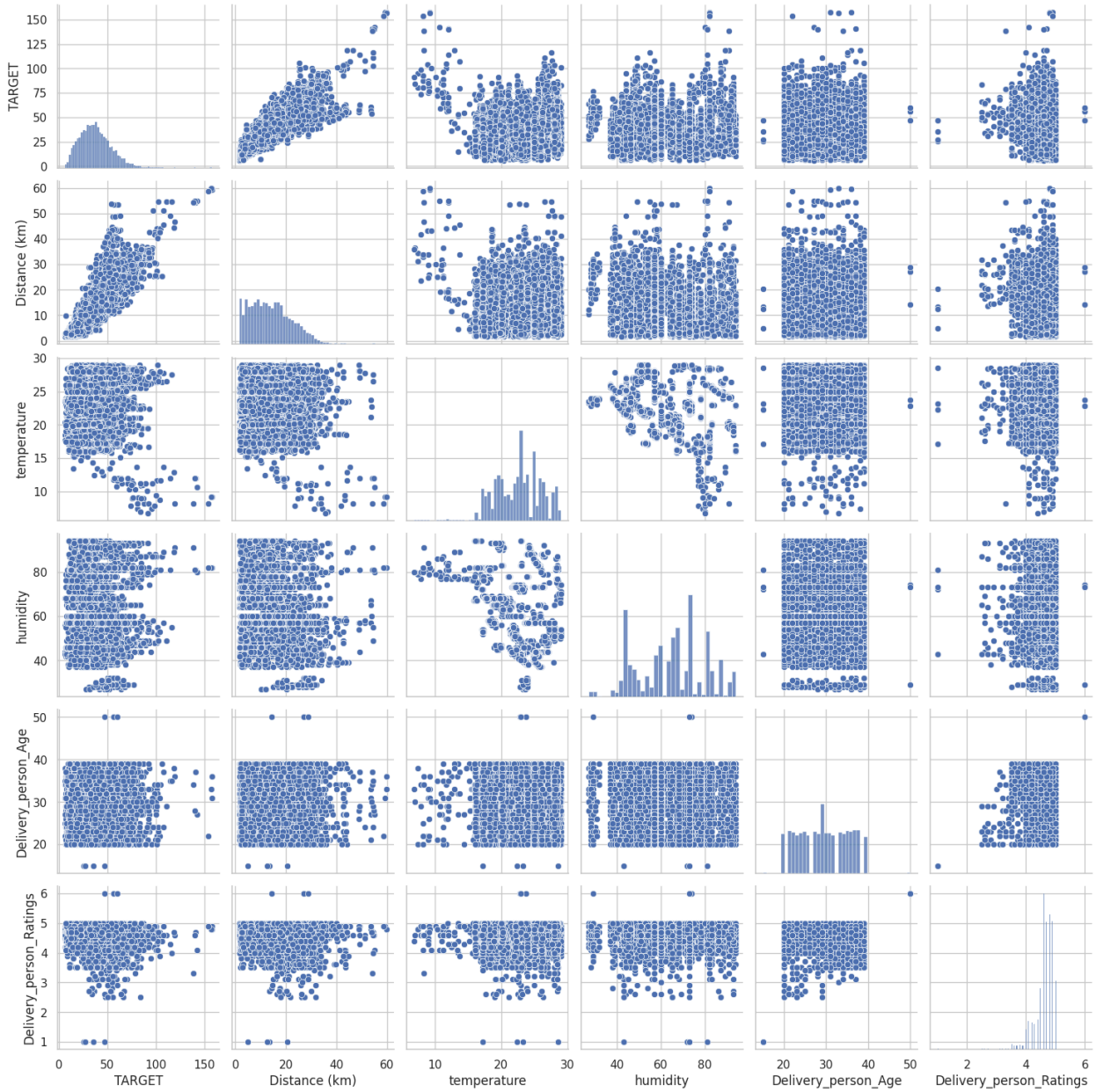
Visualization

Correlation Matrix



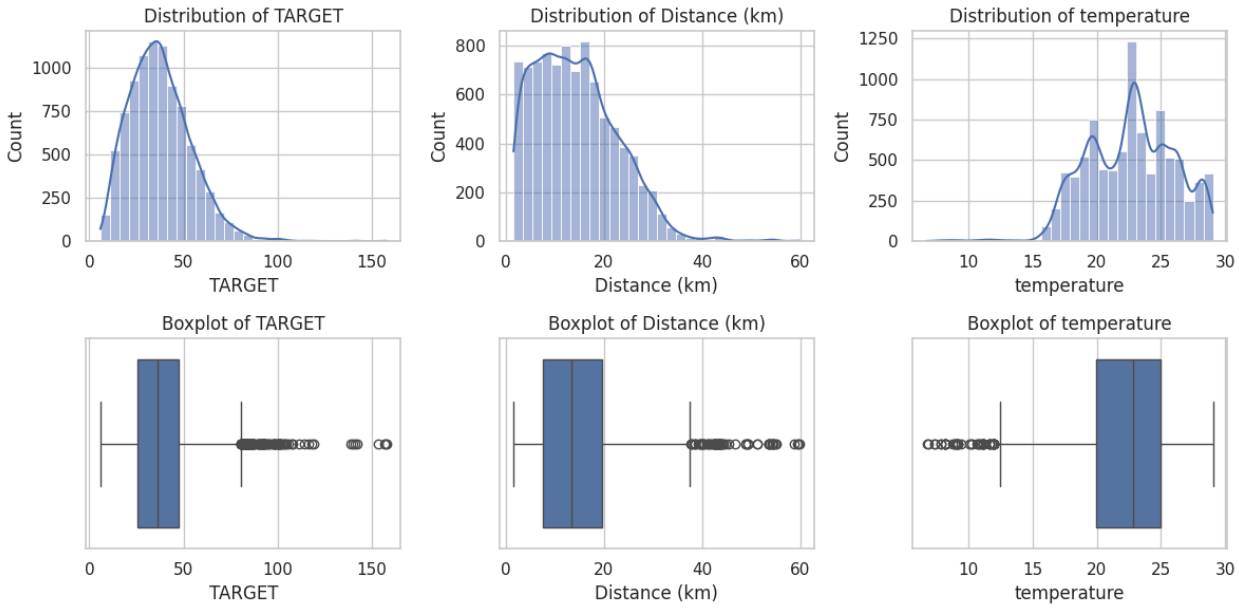
At the correlation matrix, we see there are 11 pairs of features correlated to each other where 4 pairs are close to 0.5 or -0.5 and 3 are close to 1.

Pairplot of Correlated Features



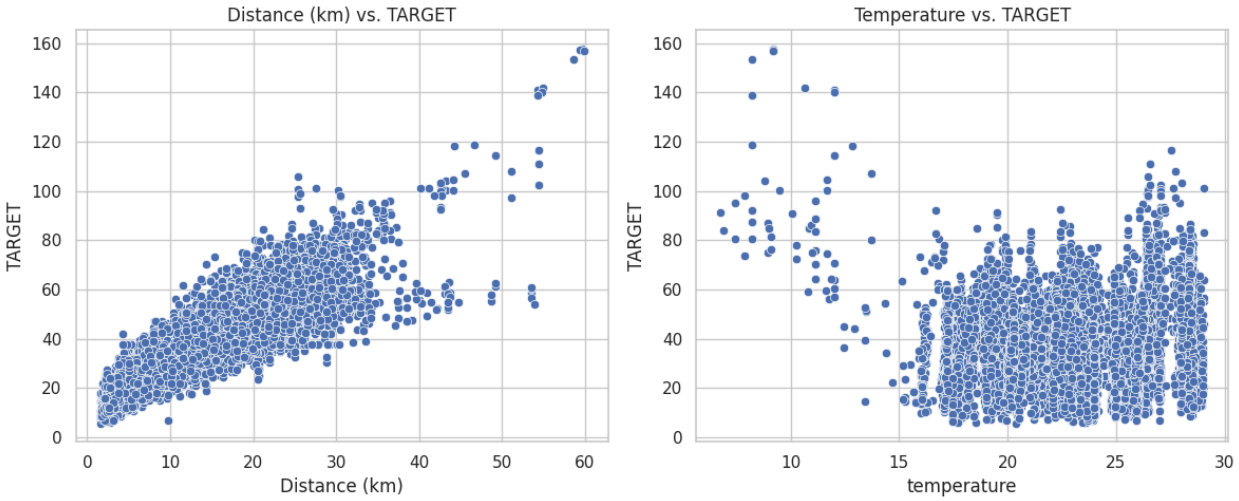
In the pair plot, we see that Target and Distance are positively correlated and on the other hand temperature and humidity are negatively correlated.

Histograms & Boxplots for Target, Distance, and Temperature



The distribution histogram plot shows that most data is distributed point and the box plot shows that data has outliers. And maximum data of Target stays below 50, for distances it is below 20 km, and temperatures 20-25 degree Celsius.

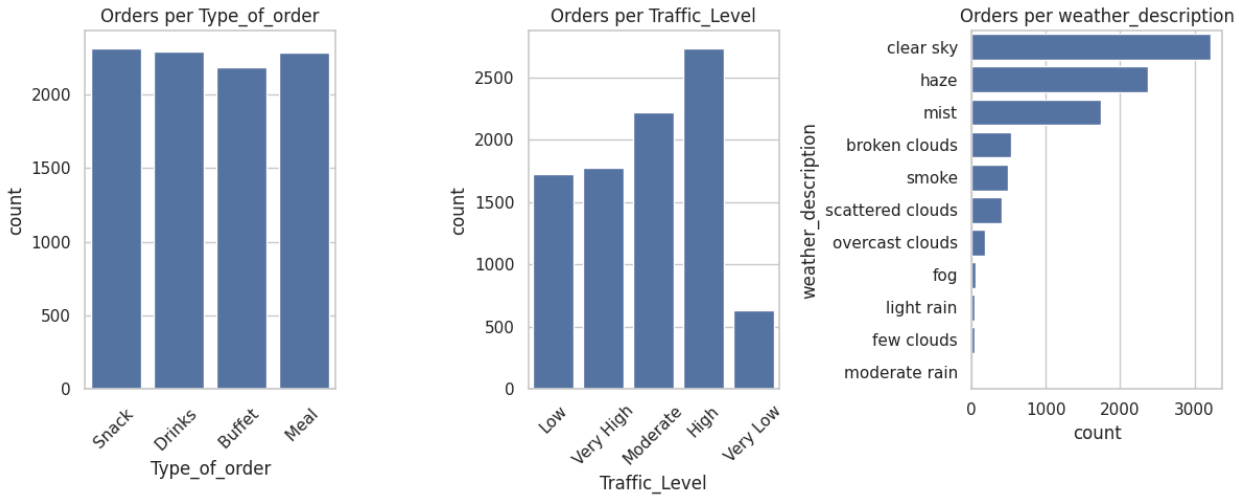
Scatterplot of Distance vs. Target & Temperature vs. Target



We have already seen that there is a good positive relationship between the Target (Delivery Time) and Distance (km) of the delivery location (left plot). It has outliers, at 60 km we have few points.

The temperature has an impact on the delivery time (right plot) we can see that when the temperature is less than 15 degree the delivery time (Target) increases. Also, it is applicable to temperatures more than 25 degree.

Frequency Count of Categorical Features



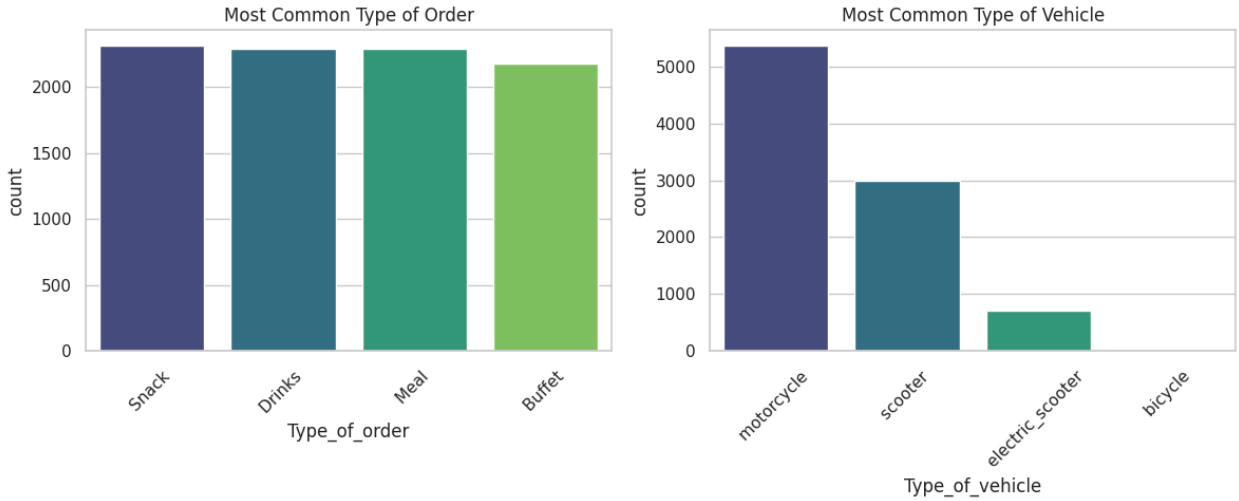
Above three plots we see that:

- Snack-type food orders have been placed in maximum time, then Drinks and Meal.
- The traffic level maximum frequency is High, then Moderate.
- The most frequent weather type is a clear sky, haze, and mist.

Insight

Common Type of Order and Vehicle

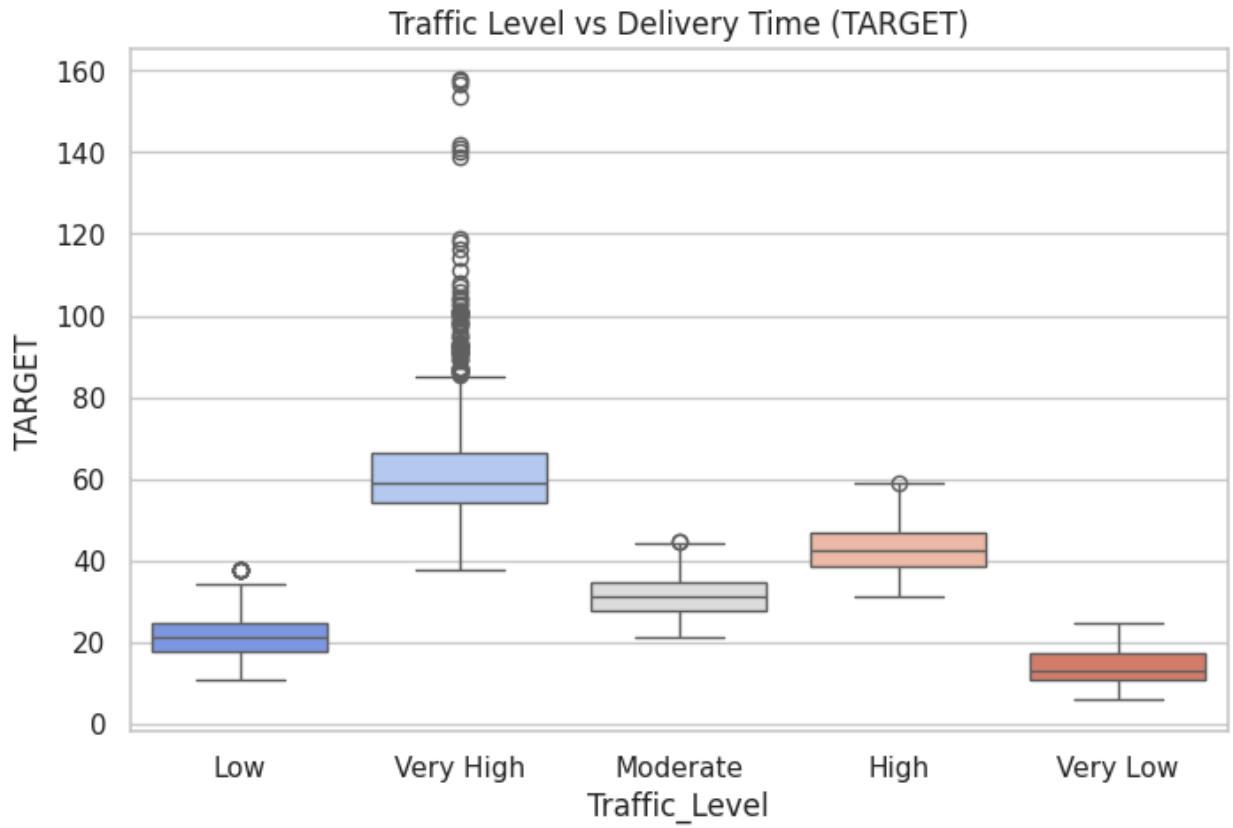
What are the most common Type_of_order and Type_of_vehicle?



The most common type of order is Snack and the most common type of vehicle used by delivery persons is Motorcycle.

Relationship between Target and Traffic

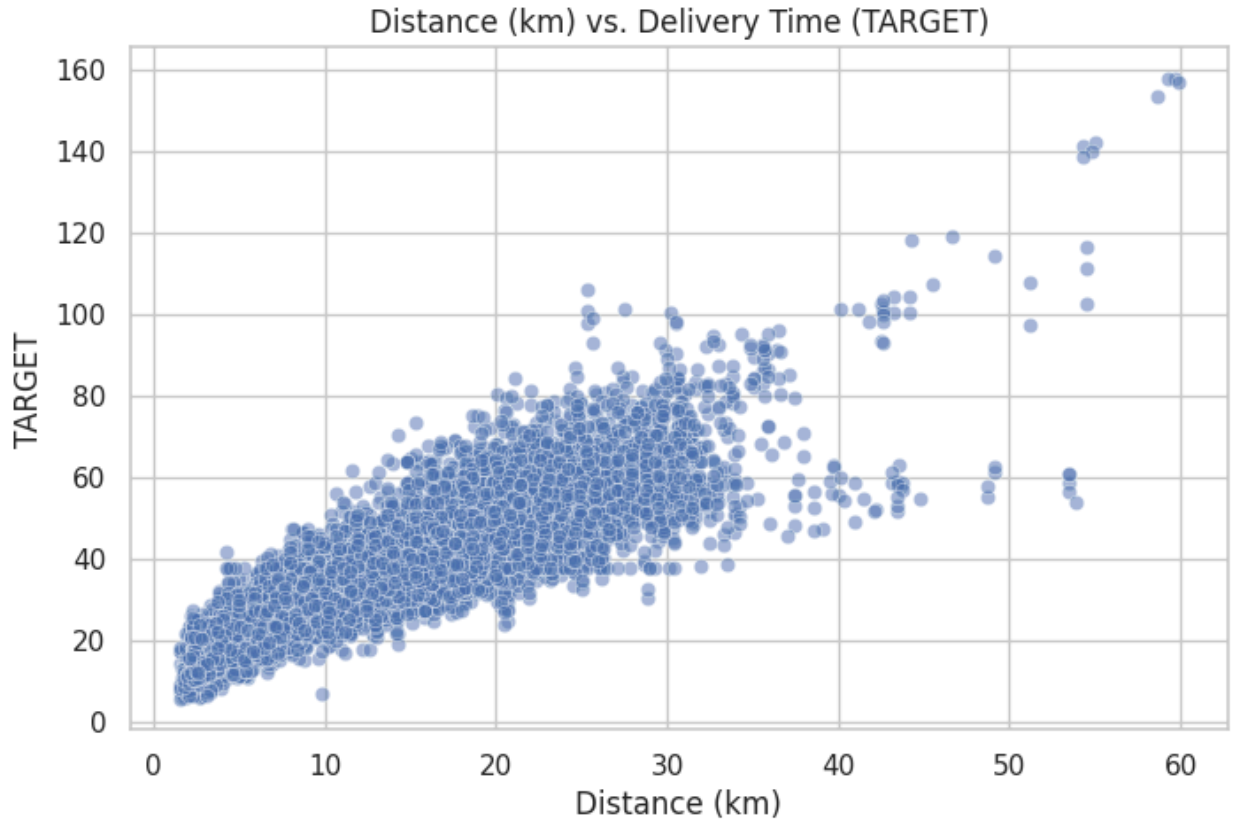
How does Traffic_Level affect TARGET (delivery time)?



From the above box plot when traffic is very high the delivery time increases to above 140 minutes and the median of the delivery time of that traffic status is close to 60 minutes. The traffic level has a very close relationship with delivery time which is clearly visible in the above figure.

Relationship between Distance and Target

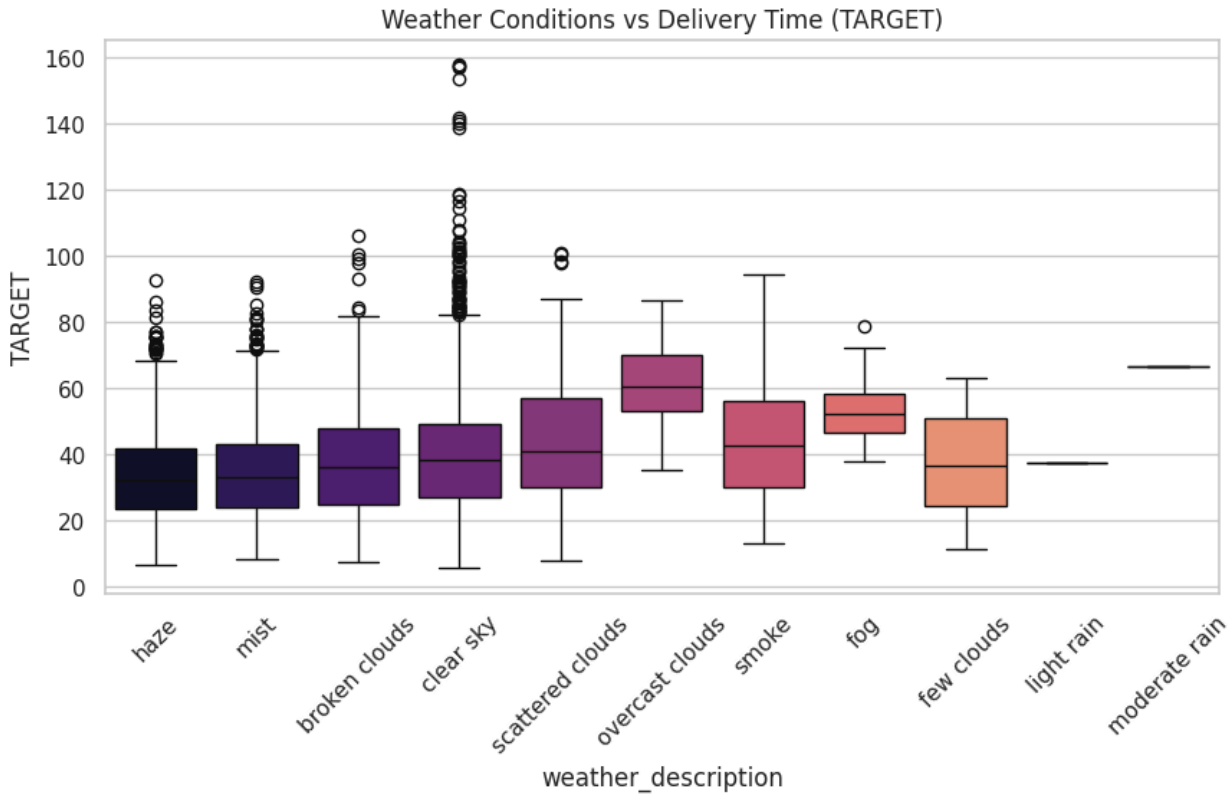
Can we see a correlation between Distance (km) and TARGET?



Also, it is quite true that when the distance becomes larger the delivery time gets delayed. So that is shown from the above-scattered plot. The most frequent distance is between 10-30 km where delivery time is between 20-80 minutes.

Impact of Weather on Delivery Time

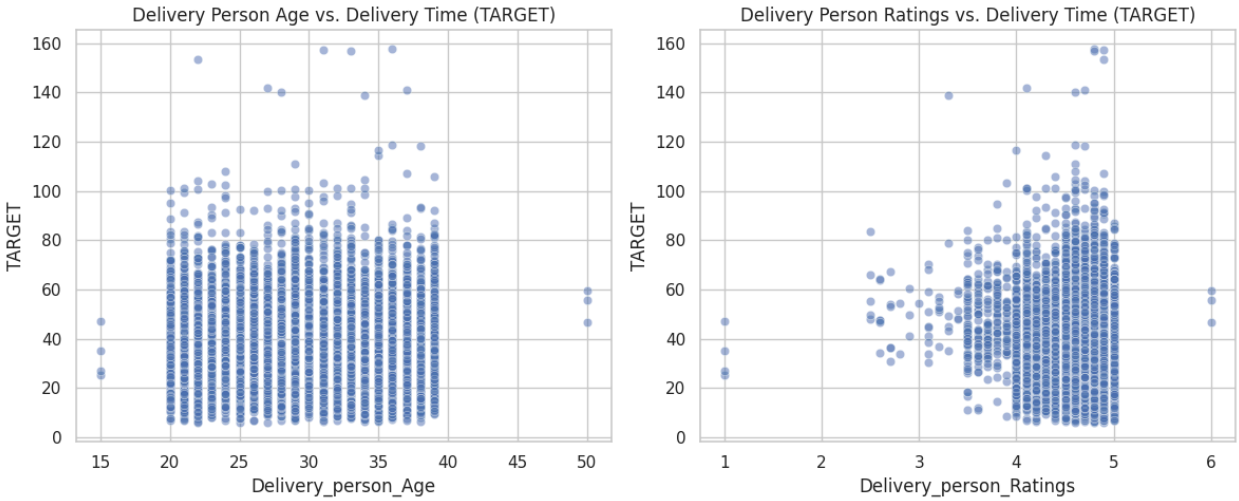
Does weather (weather_description) have on impact delivery times?



At weather type, there is significant changes that can be seen at overcast clouds where the median value is 60 minutes delivery time. Clear sky has a median of 40 minutes and a lot of outliers there is visible on the other hand light rain and moderate rain have a very low amount of samples presented in the dataset.

Influence of Delivery Person Age and Rating on Target

Is there any influence of Delivery_person_Age or Delivery_person_Ratings on TARGET?



We also checked the influence of delivery persons' age and rating on delivery time, but unfortunately, our data presented no good correlation between them. We can only see the patterns that age are in the range of 20-40 years where ratings are 4-5 stars rated and on both parts, the delivery time is 10-80 minutes.