Jahez Application - ML-Based Approach To Minimize the Delays in Accepted Orders to Provide Better Customer Experience.

Prepared by:

Anfal Almohaimeed

Hajer Alhoqail

Norah Alamri

Nouf Almuwaisheer

Renad Alahmadi

Zainab Alhejji

20/11/2022

Content

1. Introduction

Saudi’s 2030 Vision

The Kingdom’s high dependency on oil, volatility in the global oil prices, and inefficiencies in the economy, have led the government to focus on reducing the reliance of its economy on the Oil and Gas sector. The envisioned transition and growth are underpinned by various initiatives of Vision 2030 and the 13 Vision Realization programs. Three of these programs either directly or indirectly impact the Online Food Aggregator, Quick Commerce, and Cloud Kitchen businesses. The Kingdom is expected to spend over SAR 260 billion starting in 2016 to support the National Transformation Program (“NTP”) initiatives. The program focuses on developing the digital economy through initiatives aimed at supporting the leadership of local digital companies, promoting e-commerce, raising digital awareness among citizens and expatriates as well as enhancing the spread and speed of the internet across the Kingdom of Saudi Arabia.

Food Delivery Applications

The concept of food delivery apps in Saudi Arabia has spread like wildfire with an exponential increase in home deliveries seen during the COVID-19 lockdowns. While in the past, delivering food at home might have meant long phone calls and waiting times, now it’s nothing more than a few taps on your mobile phone. Food delivery apps in Saudi Arabia have never been as popular as they are now.

Jahez Application

This is one of those dominant Saudi platforms that specialize in delivering restaurant orders online. Their mission is to provide a unique customer experience through fast delivery, delivery time scheduling, order tracking, user-friendly application, multiple payment options, continuous service developments, and a variety of restaurants.

Problem Statement

In this project we will focus on the delivery service and how can optimize the process of it. When we investigated the data and started the analysis process, we had couple of questions in mind. Those questions are related to factors that could cause the delays into two sections.

First, internal factors:

- Are there districts that have a high density of customers that can't be served quickly by the existing restaurants?
- Are there restaurants that have slower service compared to the other restaurants?
- Do restaurants give priority to their in-place customers over Jahez's customers?

Second, external factors:

- Do the delays happen during rush hour in Riyadh and if so, can we minimize them?
- Is there a strong correlation between Riyadh rush hours and Jahez's number of orders?
- Can we possibly have a variable value for the customers' radius depending on the rush hour?

So, the main objective of the case study is to build a classification model that can better predict if we should acceptance the order of our customers with the goal of minimizing the delays that could happen.

Problem Objectives

- Identifying rush hours in Riyadh.
- Identifying districts that cause delays to the orders.
- Identifying orders that exceeded 25 km distance.
- Set criteria for the accepted orders during rush hours to improve Jahez's service.

2. Data Review

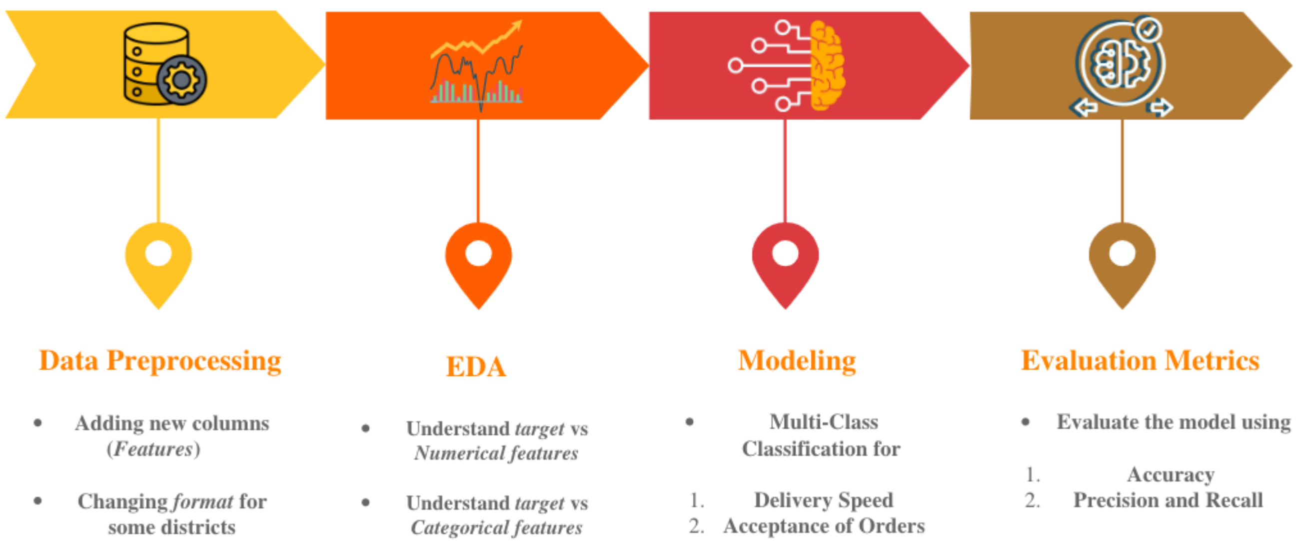
Jahez Dataset which was used in this case study has been taken from Jahez Company. It includes the orders in the first two weeks of March 2021. At the start of the project, we are going to have a clear and only one data frame to study. After applying feature engendering to create a new variable. We have 11 variables and 50001 observations on the dataset.

Variables and Explanations

Variable	Meaning
Ordered Date	Date of order
Day Name	Day name of order
Hour	Time of hour of order
Minute	Time of minute of order
Restaurant District	Name of restaurant district
Customer District	Name of customer district
Distance	Distance between restaurant and customer
Time Spent to Deliver	Time spent to deliver for order

Rush Hours	The time of order is in rush hours yes or no
Delivery Speed	The delivery speed is normal or fast or late
Accept Order	The acceptance of the order is delivered or not deliver or deliver with speed normal or deliver with speed late

3. Plan Overview



4. Data Preprocessing

In the preprocessing stage, we processed categorical columns in two columns ' Customer districts' and ‘Restaurant districts' through remove the word 'district' from these columns. And change the writing formats of districts, from small to capital letters as shown in the code below.

```
#Create function to remove District word from all rows
import re
def remove_district(text):
     regex = re.compile('(\s*)District(\s*)')
     text = regex.sub('', text)
     return text
```

```
df[["restaurant_District", "customer_District"]] =
df[["restaurant_District","customer_District"]].replace('alsalam', "Al salam");
```

In the feature engineering stage, we create three new columns: First, to represent Riyadh's rush hours [1]. Second, to categorize the order delivered into three classes. Third, to decide whether to deliver an order or not.

```
def Rush_Hours(value):
  if value > 15 and value < 23:
        return "Yes"
  else:
        return "No"
df['Rush_Hours'] = df['hour'].map(Rush_Hours)
```

.

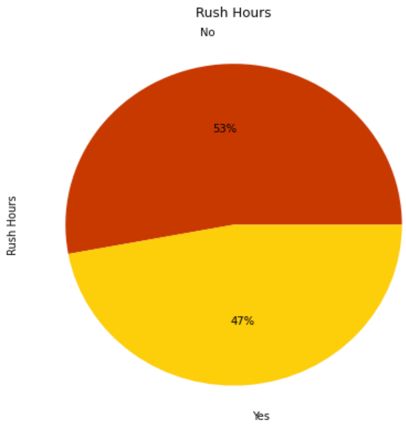
Also, we create Accept Order column: since all the orders were accepted in the dataset, we wanted to put a cutting criteria or strategy based on the observations we had. We put in mind several factors, first is the rush hour. Second is distance and lastly is the time. Our solution suggests that we have dynamic/variable radius for our customers. In rush hours we will cover only 15 km, in normal hours we will cover 25 km. However, in all situations we should not exceeds 60 mins, so we don’t lose the freshness factor. We care about the quality of food but also the availability of more restaurants if we can provide in the application in normal hours.

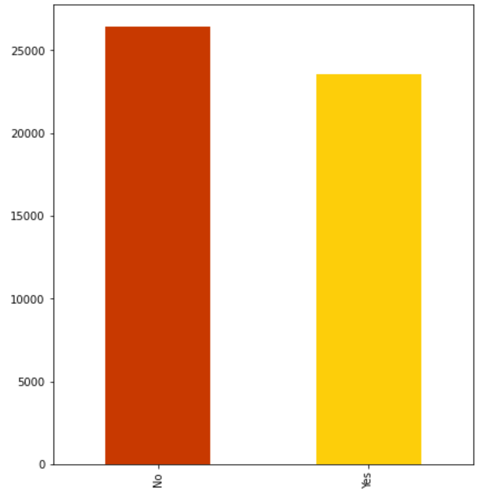
Distance	15 Km	25 Km	15 Km	25 Km
Time to deliver	41 minutes	60 minutes	24 minutes	41 minutes
Order Status	Normal Deliver	Late Deliver	Fast Deliver	Normal Deliver

```
for dis,rh ,td in zip(df_.distance, df_.Rush_Hours , df_.time_spent_to_deliver):
      if dis <= 15 and (rh == 'Yes' or rh =='No'):
         deliverOrder.append("Deliver")
      elif (25 >= dis > 15) and rh == 'Yes':
         if (td * 2 < trafic_factor):
            deliverOrder.append("Deliver Speed: Normal")
         else:
           deliverOrder.append("Deliver Speed: Late")
      elif (25 >= dis > 15) and rh == 'No':
         deliverOrder.append("Deliver")
```

5. Data Exploration

EDA1 - Pie chart and count plot for rush hours



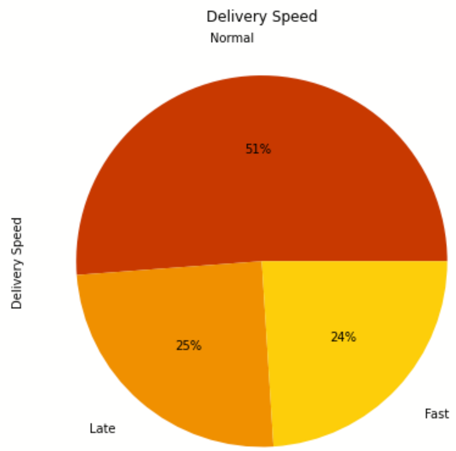


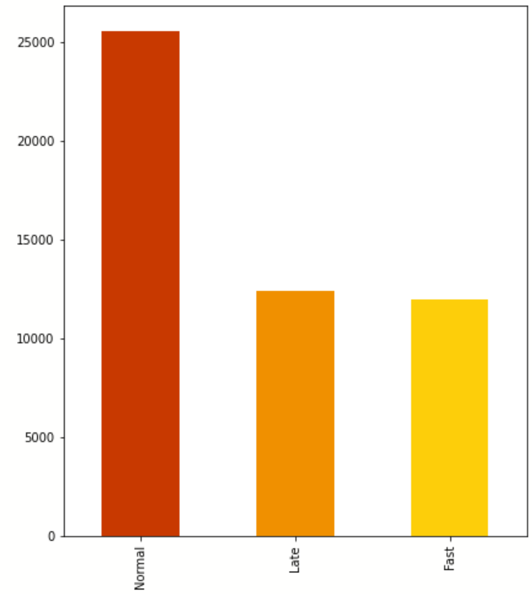
Insight:

We can see rush hours data distribution is almost equal, having 47% of hours as rush hours and 53% of hours as not rush hours.
hence,

- The number of accepted orders in rush hour: 23546
- The number of accepted orders in not rush hour: 26455

EDA2 - Pie chart and count plot for delivery speed



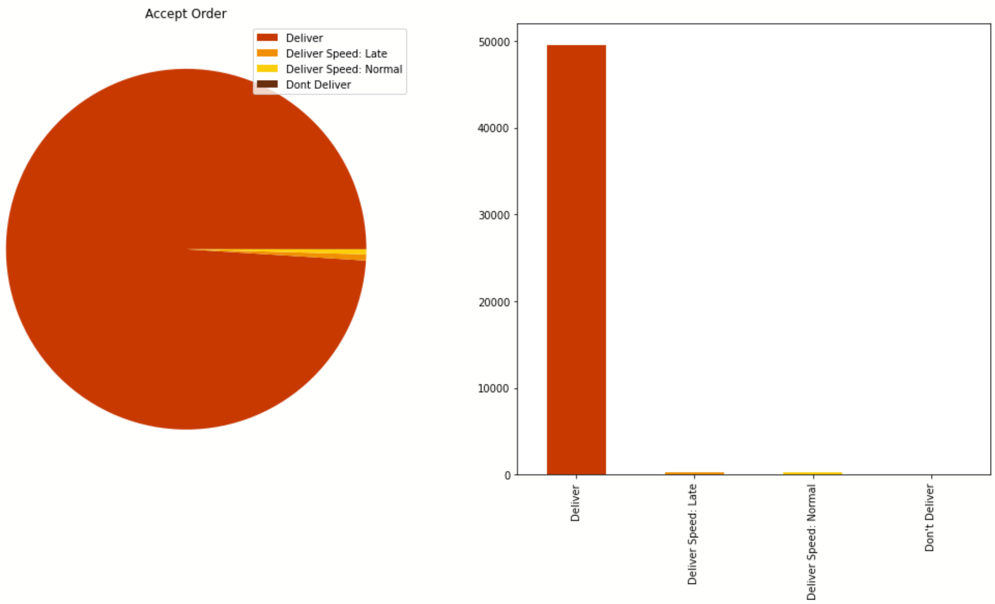


Insight:

The pie chart above represents delivery speed where speed is one of three (fast, normal, and late). We can see that majority of orders (51%) were delivered in a reasonable time (within 25 - 40 minutes).

- 25% of orders were delivered late (took more than 40 minutes to deliver).
- 24% of orders were delivered fast (in less than 24 minutes).

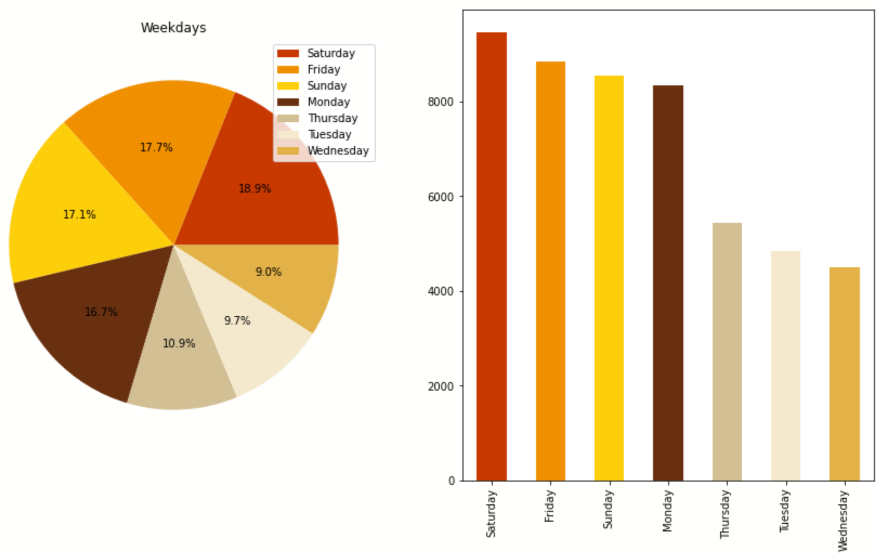
EDA3 - Pie chart and count plot for Accept order



Insight:

The pie chart above represents order acceptance, where it is one of four options (deliver, deliver speed: Normal, deliver speed: Late, or don't deliver). We can see that majority of orders (98%) were delivered as they should. 0.05% and 0.04% of orders were dependent on certain conditions (they will eventually be delivered but either late or within a normal time). And only 0.00008% of orders should be rejected.

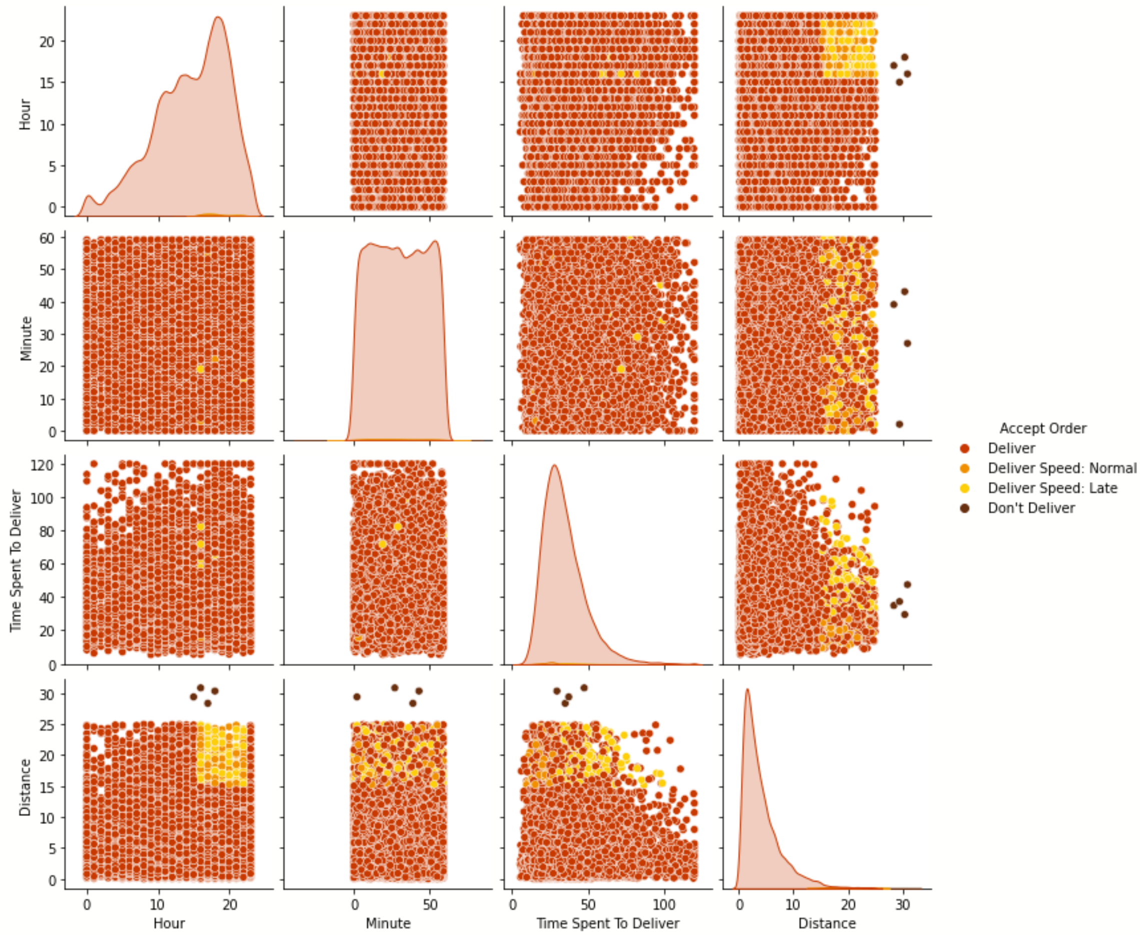
EDA4 - Pie chart and count plot for weekdays



Insight:

From the bar plot above, we can see that Saturday, Friday, and Sunday are the days people order the most, this may be due to various reasons such as people being busy on Saturday and Sunday. Unlike the days when people order the least where it could be because they are more likely to go out to eat instead of ordering online.

EDA5 - Pair plot for Accept Order



Insight:

From the pair plot above, we can see that the category 'Don't Deliver' represent the outliers in all the plots. Another interesting observation from the pair plot is the upper little square on the right of the (Distance - Hour) plot which showcases that the accepted orders which arrived either late or within a normal time, were delivered in such a range due to the distance being too far or the time being too late in the day.

6. Building Classification Models

Random Forest Classify Model

```
[ ] RFC = train_using_RFC(x_train_sc, y_train)
```

```
[ ] y_pred_RF = prediction(x_test_sc, RFC)

    Predicted values:
    [1 1 1 ... 1 1 1]
```

Performance Results

```
Confusion Matrix:  [[    0    3    0    0]
 [    0 9888    0    0]
 [    0   52    0    0]
 [    0   58    0    0]]
Accuracy :  98.87011298870114
Report :               precision    recall  f1-score   support

           0       0.00      0.00      0.00         3
           1       0.99      1.00      0.99      9888
           2       0.00      0.00      0.00        52
           3       0.00      0.00      0.00        58

    accuracy                           0.99     10001
   macro avg       0.25      0.25      0.25     10001
weighted avg       0.98      0.99      0.98     10001
```

7. Results

In the building models part, where our target here is to classify whether an order should be delivered, not delivered, should be delivered but notify the user it will arrive late, or should be delivered but notify the user it will arrive in a normal(expected) time period. we have worked on 6 algorithms Random Forest Classifier, Support Vector Classifier, Logistic Regression,KNN Classifier and XGB Classifier. For model accuracies and sensitivities, we can say, were insanely high, but that is only because there is an obvious bias in the target column.

8. Conclusion and Future Work

This study examined and analyzed the data of the order Jahez application in Saudi Arabia in March 2021. Utilized a predictive machine learning model which depends on a classification algorithm Random Forest Model and Accuracy performance was utilized for evaluation. The data analysis and order acceptance strategy in these study is help stakeholders to take proactive decisions and provide models that contribute to forecasting the future and making good decisions.

In the future proposal, the results of the current study recommend studying the factors that affect rush hours like the Riyadh Season. In case the restaurant is within more than 25 km of distance, we need to make sure to keep the quality of food. In the modeling phase, we need more features and a variety of data to avoid the model bias problem.

References

[1]. https://www.tomtom.com/traffic-index/riyadh-traffic/