Eric Nunes
enunes1@asu.edu

Executive Summary:

The data used for analysis is customer information from a food delivery company. I cannot share the data but can provide examples and insights from the data.
The problem formation was an open-ended discussion based on the given data sample. A variety of problems can be formulated like customer segmentation, predictive modeling for number of orders. During the exploratory data analysis, I instituted that the data can be used for understanding customer behaviors, which opens a variety of new problem set like Anomaly detection, Revenue Churn rate analysis, viral marketing (like giveaway coupons, offers), customer engagement issues, so on and so forth.

During the analysis, I have touched base on few customer behavior problem set.

Customer behavior Analysis:

For any organization, understanding the customer behavior is one of the major problems. Understanding customer behavior is a superset of problems like anomaly detection, customer engagement problem, marketing strategy modeling, customer churn rate analysis, revenue churn rate analysis, fraud analysis and so on.

In this analysis, I worked on Anomaly detection for customer monitoring, marketing strategy modeling and revenue churn rate analysis for anomalies on certain assumptions. The assumptions were made because of data restrictions- the numbers of orders are uniformly distributed for period of 3 weeks (i.e. last 4 weeks – last 7 days). Analyzing individual customers can be a painful task; if the behavior slightly deviates then it will be detected as an anomaly. Hence, we use kmeans clustering to analyze group behavior.

- **Approach:**
  We can classify the whole process into below criteria:
  - **Exploratory Data Analysis:** I did univariate, bivariate analysis on the feature set first. Feature engineering included:
    - Orders before last week
    - Orders in last 3 weeks before last week
    - Average orders in per week in last 3 weeks before last week
    - Amount before last week
    - Amount in last 3 weeks before last week
    - Average amount in per week in last 3 weeks before last week
    - Divided First time order and Recent time order to Year, Month, Date, Hour, Minute

Below are some figures explaining the distribution and correlation of few variables. Figure 1 explains the distribution of the variables. Figure 2 explains the correlation among variables. Figure 3 shows the final feature set and their null values analysis.

Figure 1: Distribution of variables

```
In [119]: Analyze_correlation_continuous(data, '# of Orders in last 7 days')
```

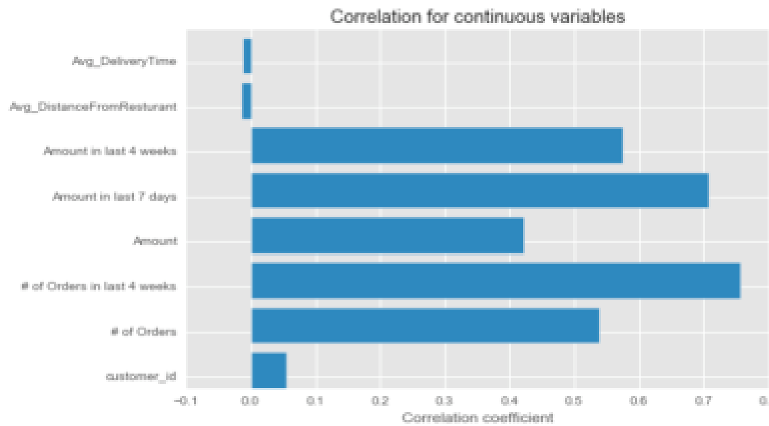

Figure 2: Correlation of variables with #orders in last 7 days

```
In [15]: _1, _2, _3 = Analyze_null_values(data)

customer_id                                        0
First Time                                         0
Recent Time                                        0
# of Orders                                        0
# of Orders in last 7 days                         0
# of Orders in last 4 weeks                        0
Amount                                             0
Amount in last 7 days                              0
Amount in last 4 weeks                             0
Avg_DistanceFromResturant                          0
Avg_DeliveryTime                                   0
# of order before last week                        0
# of Orders in last 3 weeks before last week       0
# of Avg Orders in a week                          0
Amount in last 3 weeks before last week            0
Avg Amount in a week                               0
Amount before last week                            0
First_year                                         0
First_month                                        0
First_date                                         0
First_hour                                         0
First_minute                                       0
Recent_year                                        0
Recent_month                                       0
Recent_date                                        0
Recent_hour                                        0
Recent_minute                                      0
dtype: int64
```

Figure 3: Final Feature set and their null value analysis

- **Key Intuition:** The key intuition here is to get the baseline for customer behavior using the average historical weekly data. Once the baseline is established using kmeans clustering, the last 7 days or last 1 week is compared against the baseline to detect the anomalies. Once the anomalies are detected it can be used for marketing strategy modeling and churn rate analysis.

- **Model:** I trained kmeans clustering on averaged 3-week's data to get the baseline cluster centroids. Once the centroids are computed, we use them in clustering the last week's data. If the cluster for a customer has changed that could be taken as an anomaly for customer behavior.
  Later we use the anomalies to understand the users who didn't order or ordered a lot. This set of users can be used for sending coupons or giveaway offers. We can work on maintaining customer engagement techniques as well.
  Post that, we analyze the revenue gained and churned revenue for these anomalies. We can use the anomalies to track the weekly profit/Loss and revenue churn rate.

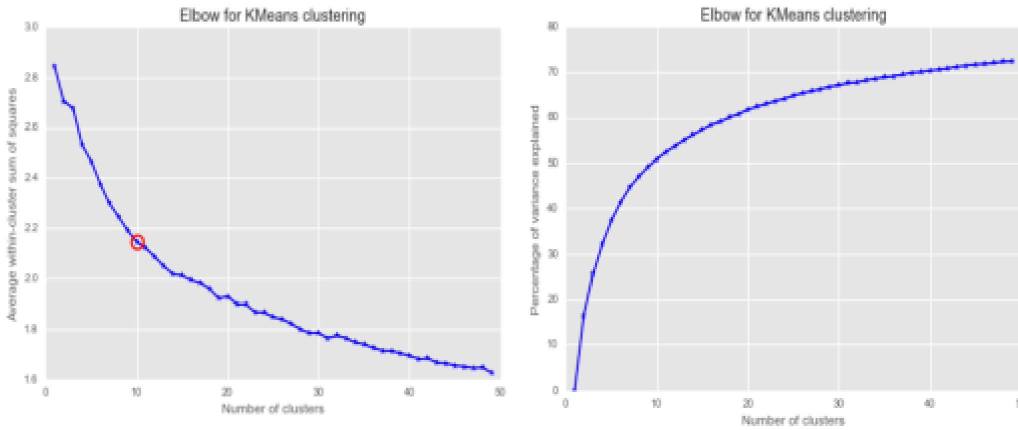

Figure 4: Selecting the number of clusters (k) for kmeans

The below figures explain the analysis of the anomalies.

```
In [72]: check = test.loc[test['# of Orders in last 7 days'] < test['# of Avg Orders in a week']]
         check.shape[0]

Out[72]: 292

In [73]: check1 = test.loc[test['# of Orders in last 7 days'] > test['# of Avg Orders in a week']]
         check1.shape[0]

Out[73]: 314

In [70]: check['Lost_amt'] = check['Avg Amount in a week'] - check['Amount in last 7 days']
         check1['Gained_amt'] = check1['Amount in last 7 days'] - check1['Avg Amount in a week']
         revenue_churn_rate = np.mean(np.where(check['Avg Amount in a week'] < 1, check['Avg Amount in a week'], check['Lost_amt

In [67]: PnL = sum(check1['Gained_amt'])-sum(check['Lost_amt'])

In [69]: "Total Profit/Loss is: ", PnL
Out[69]: ('Total Profit/Loss is: ', 71285.999999999971)
```

Figure 5 a) #Customers who lost interest, b) #Customers who gained interest and c) Revenue Churn Rate d) Total Profit/Loss from the anomalies