July 2020

JUMPMEN 23-
NEW MARKET ANALYSIS

1. DATA INTEGRITY

2. POPULARITY ANALYSIS THROUGH VISUALIZATION

3. HEATMAP ANALYSIS

INTRODUCTION

Jumpman23 is an on-demand delivery platform connecting “Jumpmen” and customers purchasing a variety of goods. Jumpman23 will send Jumpmen to merchants to purchase and pickup any items requested by the customer. Whenever possible, Jumpman23 will order the requested items ahead to save the Jumpmen time. Each time a Jumpman23 delivery is completed, a record is saved to the Jumpman23 database that contains information about that delivery. Jumpman23 is growing fast and has just launched in its newest market -- New York City.

The csv file has a total of 18 columns and 5983 rows.

DATA INTEGRITY

MISSING VALUES

The following columns have missing values

```
place_category                        883
item_name                            1230
item_quantity                        1230
item_category_name                   1230
how_long_it_took_to_order            2945
when_the_Jumpman_arrived_at_pickup    550
when_the_Jumpman_left_pickup          550
```

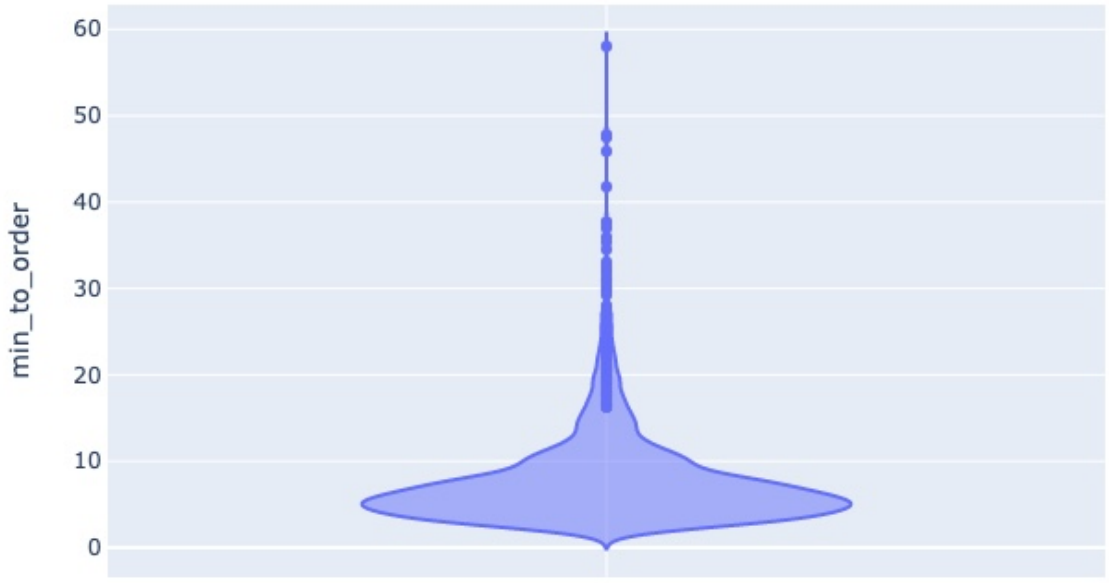
The categorical columns are imputed with most common values to replace missing values.

DUPLICATE RECORDS

There were a total of 16 duplicate rows which was identified and deleted using drop_duplicates() method.

OUTLIER ANALYSIS

The column how_long_it_took_to_order was converted to minutes to identify any orders took longer than average order time. Plotted with a violin plot

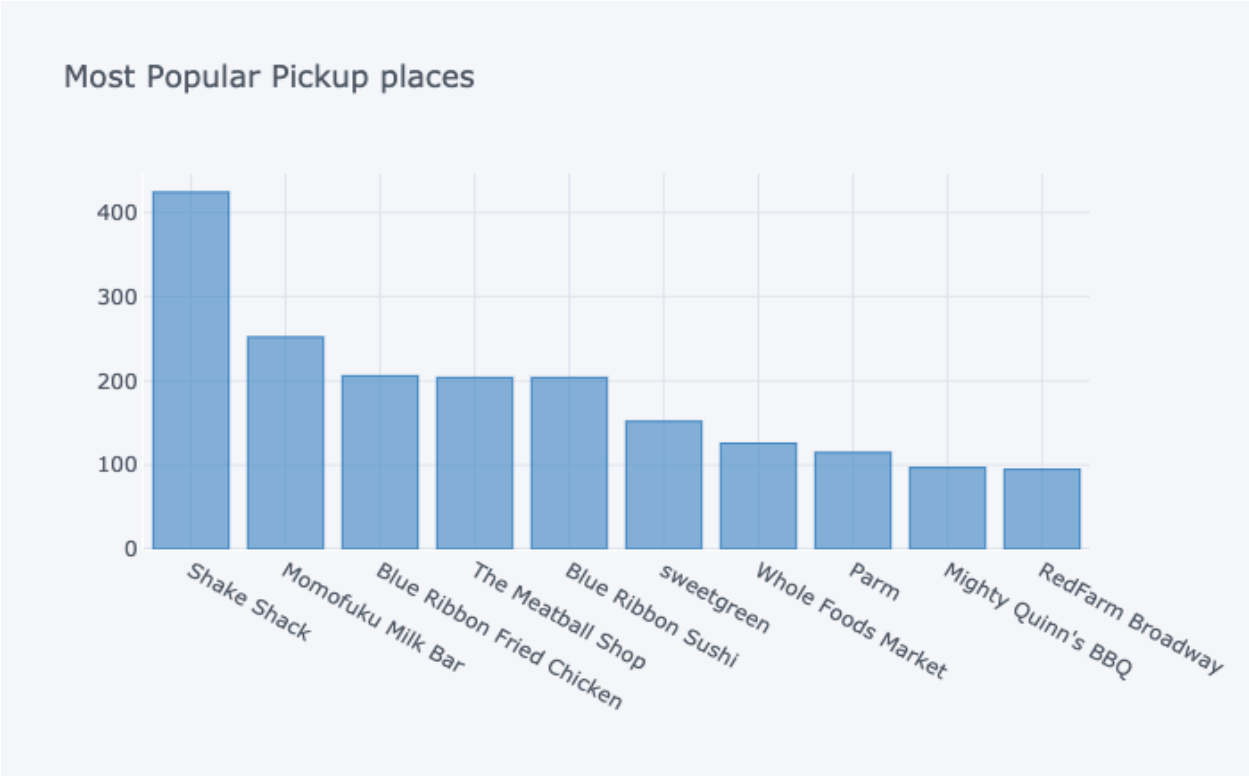


There are multiple orders which took more than 30 mins to place an order . These orders can be considered outliers and needs to be investigated to see if there is any performance issues with the app. This is critical since it reduces and churn and helps to increase user retention.

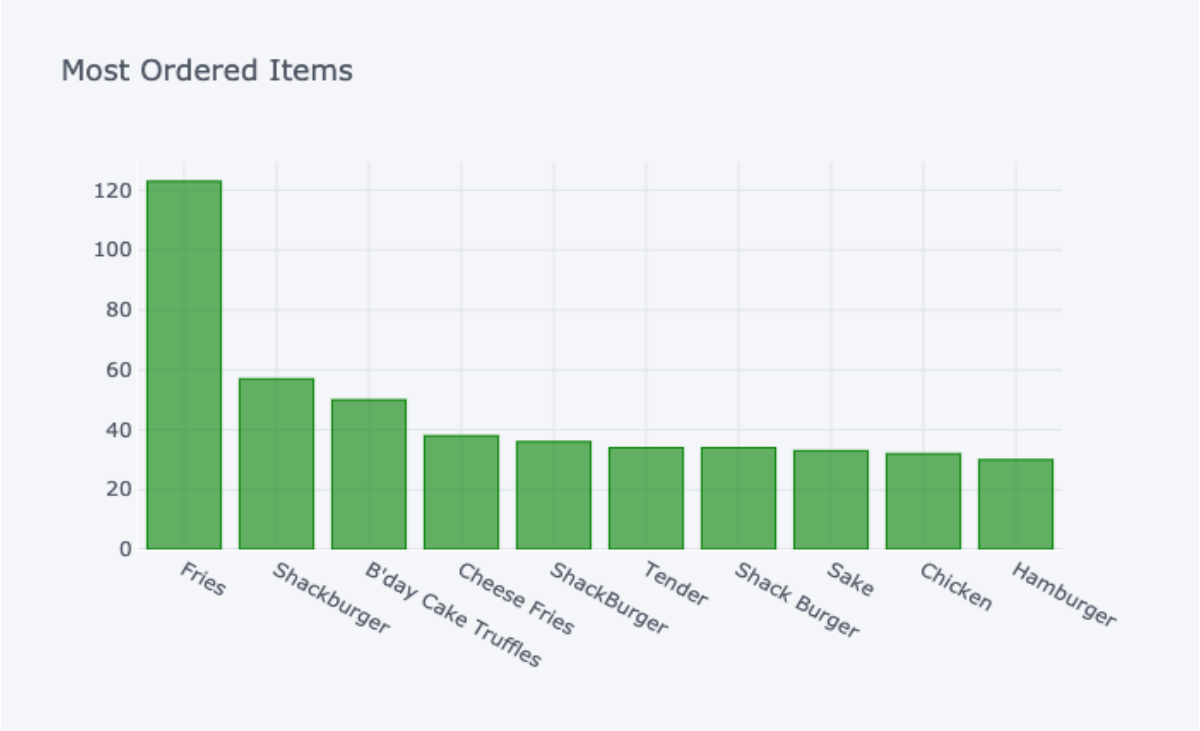
The other data integrity issue that needs attention is the columns when_the_Jumpman_arrived_at_pickup and when_the_Jumpman_left_pickup have exact same number of null values indicating some of the orders doesn't record the datetime attribute of the delivery and must be inspected further.

POPULARITY ANALYSIS THROUGH VISUALIZATION

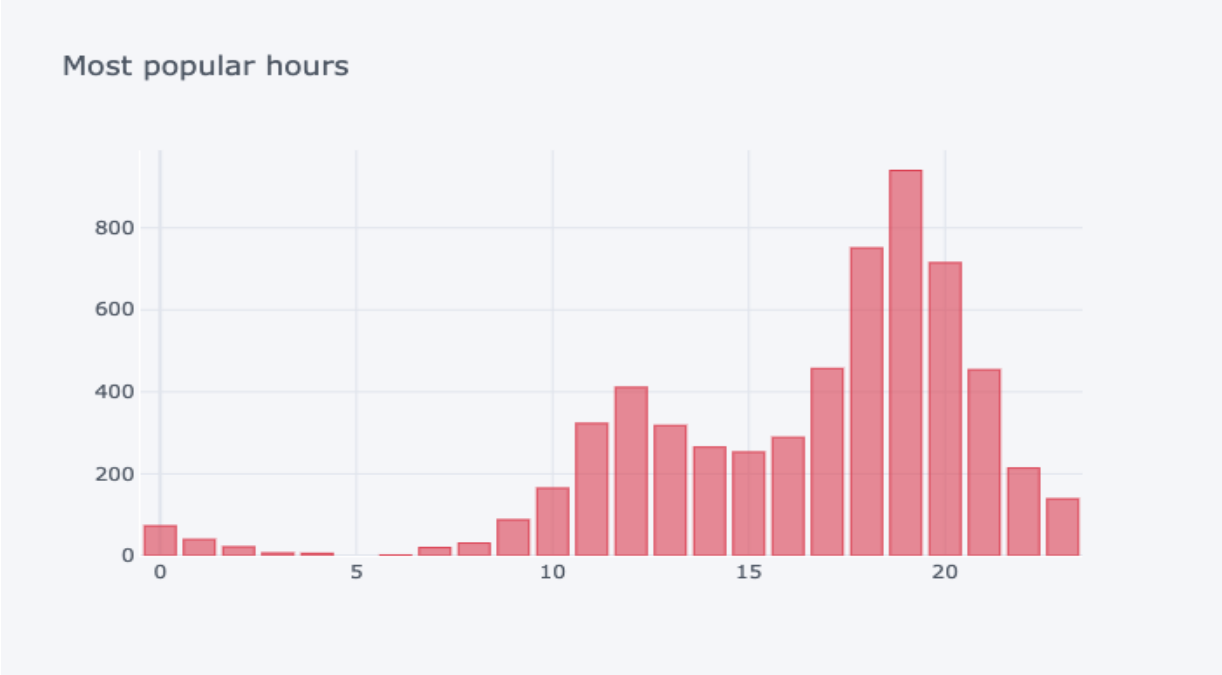
TOP PICKUP PLACES BASED ON ORDERS



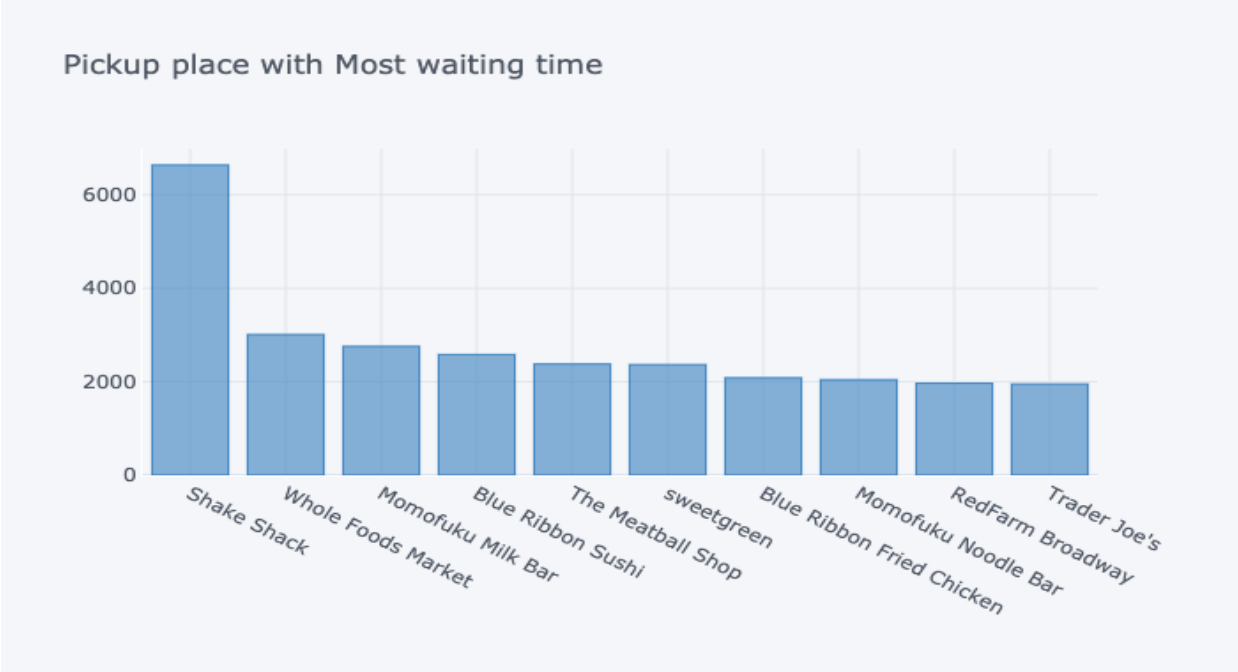
MOST ORDERED ITEMS



MOST POPULAR HOURS



PICKUP WITH MOST WAITING TIME



These 3 popularity metrics are essential for our analysis. It would improve efficiency when we pre-order in following cases

1. With more than 6000 minutes of overall pickup waiting time as well as being the most popular pickup place, Preordering from Shake Shack would increase

2. Also other most ordered foods like Fries during the busy hours will help us to reduce the waiting time for JumpMen and also alleviates order congestion.

From the graph, we can infer that the demand is high during Dinner time between 6 pm to 8 pm.

MOST COMMON MODE OF TRANSPORT

```
bicycle       4274
car           1215
walker         274
van             76
scooter         75
truck           48
motorcycle      21
```

HEATMAPS

Pickup regions

Dropoff regions

Based on the two graphs above, it is evident that the dropoff zones are much more spread out(Queens, Brooklyn, Upper-Manhattan) than the pickup zones. The graphics are irrespective of time, and focus on aggregated interest points.

CONCLUSION

The data preprocessing steps have reduced inconsistencies. By solving the identified data integrity issues, we can get more accurate data for our analysis . This in turn will be useful to make better decisions for retaining and adding new users and also reduce the overall delivery time in a busy city like New York.