Optimizing an Advertising Campaign by Location Intelligence.

1. Introduction

1.1 Background

According to the National Restaurant Association, restaurants are a driving force in New York's economy. They provide jobs and build careers for thousands of people, and play a vital role in local communities throughout the state.

- In 2018 there were 50,153 Eating and drinking place locations in New York.
- The Estimated sales in New York's restaurants were \$51.6 billion.
- In 2019 there were 865,800 restaurant and foodservice jobs in New York , representing 9% of employment in the state.
- Every dollar spent in the tableservice segment contributes $1.76 to the state economy.

Half of all American adults have worked in the restaurant industry, and millions have sucessful professions in this vibrant industry.

On the other hand, small businesses can't hire advertising agencies for campaigns involving a huge amount of money, so if any of those brands wants to set up an advertising campaign they should:

- Define their advertising goals.
- Pick what they want to promote.
- Identify their target audience.
- **Determine where to find their audience**.
- Decide the campaign timing.
- Set an advertising budget.
- Select outlets to advertise in.
- Create the advertising message and graphics.

1.2 Problem

A new and small wine brand would want to be widely known in NYC in a shortest period of time. In order to achive this, they need to set up an optimal advertising campaign minimizing the cost and maximizing its profits.

Following the steps above, they want to promote their young red wine and their fruity white one. For this they seek areas (neighborhoods) localized whith high density of restaurants and high afluent of customers

2. Data

TLC Trip Record Data

The yellow and green taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. The data used in these datasets were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP).

With this information we will work to find those neighborhoods where there is high density of drop-off during the time ranges that we will stablish for lunch and dinner.

- https://nyc-tlc.s3.amazonaws.com/trip+data/yellow_tripdata_2019-12.csv
- https://nyc-tlc.s3.amazonaws.com/trip+data/green_tripdata_2019-12.csv

We took trip data from december 2019 trying to look for a short period representative enough about the activity in restaurants in NYC.

Taxi Zones

Official zones that correspond with pick-up and drop-off locations of taxi trips.

- https://data.cityofnewyork.us/api/geospatial/d3c5-ddgc?method=export&format=GeoJSON

Centroids of Neighborhoods

In geometry, the centroid of a plane is the intersection of all straight lines that divide it into two parts of equal moment about the line.

We calculated those centroids by QGIS, an Open Source Geographic Information System.

Foursquare API

It will provide venues information (restricted to restaurant) for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City

** Please note in this analysis the situation due to Covid-19 has not been considered**

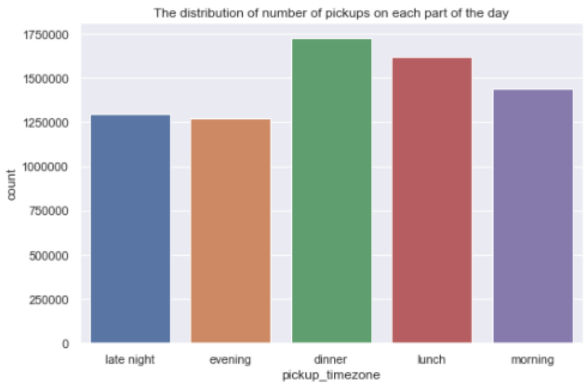
3. Methodology

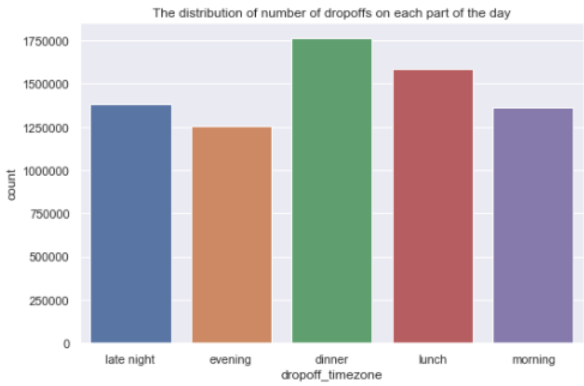
We started working with taxi trips dataset, we had a look at the **shape, data type**, and the **first 5 rows** of the data, which gave us an overview of the data.

We needed to stack yellow and green taxi data, so we had to make some changes like rename columns.

We observed the columns **pickup_datetime** and **dropoff_datetime** are stored as object which must be converted to datetime.

As part of our análisis we looked at the distribution of pickup and drop off, firstly from days of the week and after from hours of the day.

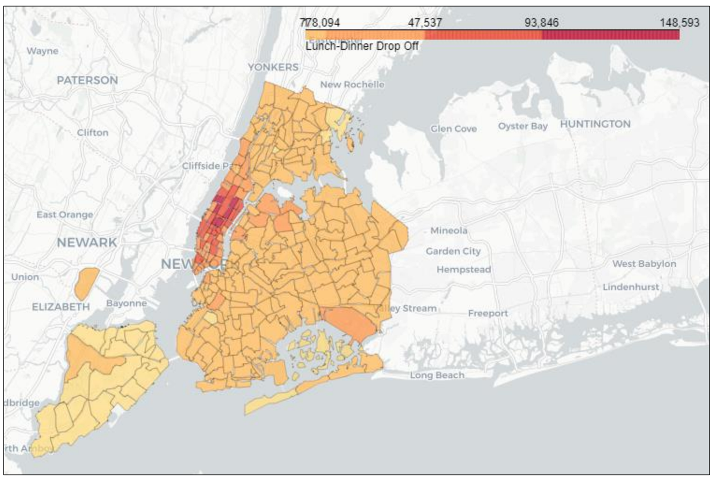




For this hourly análisis, in order to bring the data closer to the final goal of our analysis, we divide the times into 5 time zones: morning (5 hrs to 11 hrs), lunch (11 hrs to 15 hrs), evening (15 hrs to 18 hrs), dinner (18 hrs to 22 hrs) and late night (22 hrs to 5 hrs).

Next step was select those trips that happened at lunch or dinner time and convert categorical variables into dummies. This is because we'll need just one record per drop off location (neighborhood), and we'll aggregate all the information by adding it for each feature and record.

This map shows the density of drop off during lunch or dinner time. The darker the higher the density.



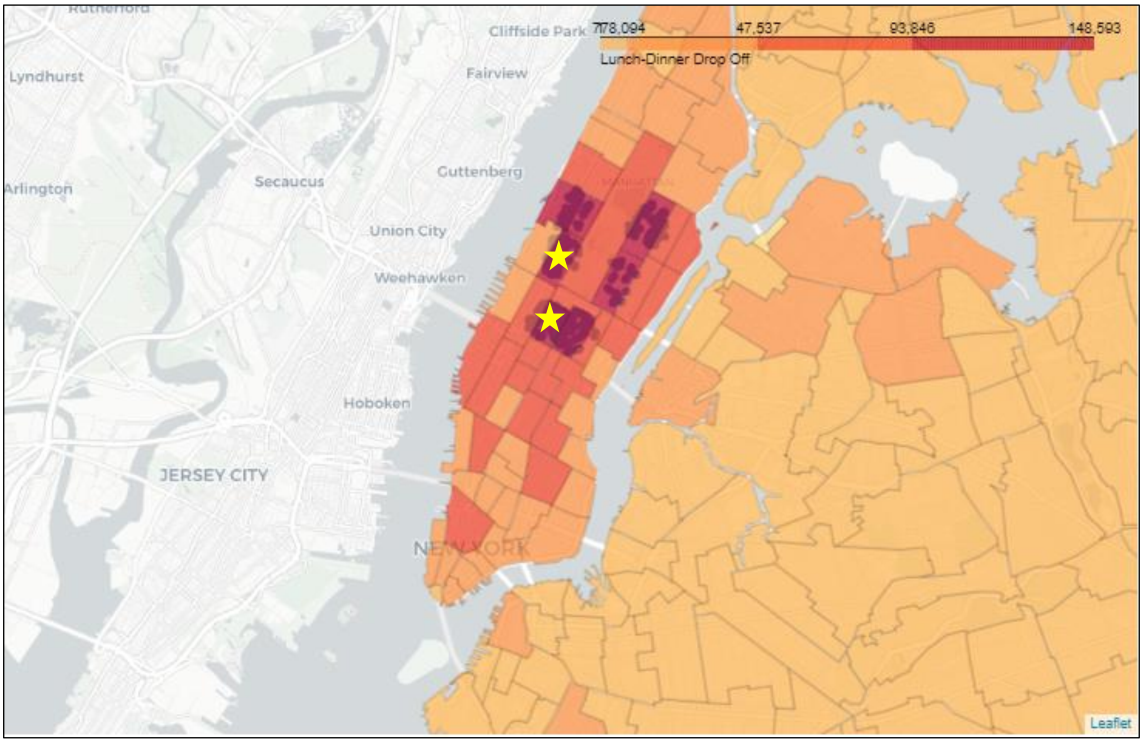
Finally used Foursquare API to look for restaurants within a radius of 500 meters from the centroids of our top 6 neighborhoods.

- Lincoln Square East
- Midtown Center
- Times Sq/Theatre District
- Upper East Side North
- Upper East Side South
- Upper West Side South

4. Results

With the steps above our goal was to find the neighborhood with the highest density of visits at lunch or dinner time, through taxi trips (drop off data), and which also has the highest density of restaurants.

Acording the trips and the restaurants density, we obtain:



Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Lincoln Square East	81	81	81	81	81	81
Midtown Center	59	59	59	59	59	59
Times Sq/Theatre District	100	100	100	100	100	100
Upper East Side North	93	93	93	93	93	93
Upper East Side South	30	30	30	30	30	30
Upper West Side South	100	100	100	100	100	100

5. Discussion

At this point, probably the results are not surprising, Times Sq/Theatre District could be an obvious location, especially having used the data of taxi trips from December.

The second option, with the same number of restaurants is Upper West Side South, which is considered one of Manhattan's cultural and intellectual hubs.

If we want to improve this study, we could take a longer period of time, maybe a whole year and calculate the average of taxi trips values, in order to avoid any possible temporary dependence.

On the other hand, we could as well repeat the Foursquare search several times in order to obtain a higher number of restaurants, being the results more representative.

6. Conclusion

In this study we have reached a first approach, with a couple of location where our theoretical wine company could deploy their advertising campaign.

The results make sense but, as we said earlier, there would be a couple of improvements that could make them more robust.