**Determining the optimal neighborhood of where to locate a fast-food stall.**

**Daniel Ortiz**

**April 10, 2021**

1. **Introduction**
   1. **Background**

Manhattan is an island located at the mouth of the Hudson River in northern New York Harbor and is also one of five districts that are part of New York City, United States. In this place there are many kinds of venues where the people can do any types of things, therefore, the people pile up in certain places where there may be many kinds of locations that there are not now. Exist a distribution of locations that sell all types of food, as restaurants and food trucks where sell fast food.

* 1. **Problem**

Many people need to know the best way in they can locate their own business, depending on kind of business location there will be different factors that could affect to the performance of business sales as the amount of people that usually goes through there. Certain locations and venues make certain areas more crowded and that could be a good factor, however, certain locations could become a problem for the location that one person want to put there, as venues that sell the same kind of things.

* 1. **Interest**

This could be useful for many people, but specifically it will be focused in the correct location of a Food Truck which will sell fast food. There are many people that want to undertake with a new fast-food business and face the same kind of problem and that is: they do not know where locate it. All the time, the people face an incredible number of doubts to assume their business. Is for this, that all that people could be interested in this kind of projects.

1. **Data acquisition and cleaning**
   1. **Data sources**

This data was able on the IBM cloud and was used for a previous lab, and it is used also here for reasons of simplicity, but for obtain this, the IBM operators could have used web scrapping techniques with the library *beautifulsoup*. The type of file that

* 1. **Data cleaning**

Due the fact all that data was in JSON file it is necessary convert it to a dataframe using methods of the library of *pandas.* After that, for the sake of simplicity, we only take the data from Manhattan to explore a smaller number of data and not extend more than appropriate.

As after that we only have the names of the neighborhoods and data of latitude and longitude, we use the foursquare API to obtain the venues around that places.

* 1. **Feature selection**

As I said before, there are places that could affect positively are that venues that attract a great amount of people, as this: dance studio, dessert shop, cupcake shop, cosmetic shops, community center, circus, candy stores, chocolate shops, clothing store, comedy clubs, banks, auditoriums, gyms, gardens, theaters, pedestrian plazas, etc.

The places that could affect negatively are that venues that sell the same things that I am going to sell, or the same kind of things that I am going to sell, as this: other food trucks, food stands, hot dog joints, BBQ joints, burrito places, empanada restaurants, pizza places, sandwich places, taco places, kebab restaurants, etc.

Both the places that positively affect and those that negatively affect were selected in the analysis to make an estimated

1. **Exploratory Data Analysis**

We begin fetching the number of venues that there are on each neighborhood to make us an idea of how crowded could be each one of the neighborhoods.

Tabla

Descripción generada automáticamente

Table 1: first five places of the data and its respective number of venues around it

After that, analysis of how many unique categories exist in total, and there are 324 unique categories in this specific dataset. And analysis of each neighborhood grouping by neighborhood the amount points for venue. In this way, we can obtain the distribution frequency of each one venue on the neighborhood.

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

Table 2: taking the mean of the frequency of occurrence of each category.

After that, select the columns of the data that represents positive points which consist in the points that offer the venues that contribute good features for a good place to locate a food truck, what is the objective of this analysis, for example, venues that could attract many people. And also, select the columns of the data that represents negative points which consist in the points that offer the venues that contribute bad features for a good place to locate a food truck, for example, fast foods outlets or fast-food restaurants. Due the fact that this type of venues could greatly affect, I multiply by two the points that that has.

#### The total points are the subtraction between the positive points and negative points to obtain an estimation of the probability of success of a food truck in the area.

#### At the end we obtain the neighborhoods where is likely be successful in the sells of a Food Truck.

#### Tabla Descripción generada automáticamente

#### Table 3: five most convenient places to locate a Food Truck

1. **Conclusions**

#### The extent to which the variables affect may be different from those estimated, I mean, all those that negatively affected as fast-food stalls were only multiplied by a single constant (2) and the places that positively affect were left that way. The right thing to do would be to analyze the popularity of the places and depending on how crowded they are, make a better estimate of the extent to which it affects. But still, the estimate that was obtained can be of great help to many people.