

# Report

February 22, 2019

## 1 INTRODUCTION: Gourmet Bussines

A chef has contacted me, since he needs to do a study of the businesses that have been successful in the culinary field he develops.

With respect to his experience, he has been based on Gourmet businesses in certain areas of Mexico City where they have been well received by the population.

His intention is to open a gourmet-type business in a potential place for his success and where the concept has not yet arrived.

The proposal that he makes to me is the following:

### 1.1 Customer's proposal

Dear Minerva: I am in contact with you to formally make my intention to require your artificial intelligence / machine learning service, for which I will detail below the requirements of the project you will develop.

Approach

As part of a professional growth, I intend to open a gourmet store, which will have as main selling axes:

Butchery: National and imported cuts of beef (prime quality) will be sold, wagyu meat (imported from Japan), aged meat (inside the store by us in aging chambers), pork cuts, piglets, exotic meats such as wild boar, crocodile. Fish and seafood: Daily selection attached to sustainable fishing, fish and shellfish national and imported. Spices, fruits, vegetables and other items related to gourmet preparations. Workshops, courses.

Location

As part of the support I require from you, I need to determine the ideal location of the store, which could be in the Cuahutemoc or Benito Juarez delegation, for which I need to consider the following points in order to make a decision:

Purchase incidence: Take into account the purchasing trends of customers, according to nearby businesses. Opening hours and closing of the store. Market segmentation.

Taking these points into account, I ask for your expertise to be able to create a model and suitable with the needs that I have.

Without further ado, I remain attentive to your response.

## 2 DATA

We will rely on Foursquare to have the necessary information of the places, as well as their popularity and the business line that is of interest.

In order to locate each of the colonies in the determined areas of Mexico City, we will access the INEGI portal, which is the National Institute of Statistics and Geography of Mexico, through a web query, we can download a csv file with the geographic coordinates of all the neighborhoods of the City and to be able to work on these data.

In the question of the procedure, we can highlight the places of Gourmet turnaround related to the interest of my client and associate the types of businesses and sites that exist around it, in order to associate a set of behavior related to the gourmet sector and thus be able to extrapolate this set to a new town in the city

## 2.1 csv File filter and Downloaded from <https://www.inegi.org.mx/app/geo2/cng/>

Import dependencies that we will need

```
In [3]: import pandas as pd
import numpy as np

In [8]: pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

#!conda install -c conda-forge geopy --yes # uncomment this line if you haven't complete
from geopy.geocoders import Nominatim # convert an address into latitude and longitude v

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

#!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't c
import folium # map rendering library

print('Libraries imported.')
```

Libraries imported.

Preview of the Data obtained

```
In [9]: geo=pd.read_csv('borough_data.csv')
geo.columns = [c.replace(' ', '_') for c in geo.columns]
geo.rename(index=str, columns={'Nombre_de_municipio': 'Borough', 'Nombre_geografico': 'Nei
geo.head()
```

```

Out[9]:
      Neighborhood  Nombre_conocido  Clase Termino_generico \
0      10 de Abril                NaN  Localidad          Colonia
1      10 de Mayo                NaN  Localidad          Colonia
2     12 de Diciembre                NaN  Localidad          Colonia
3      15 de Agosto                NaN  Localidad          Colonia
4     16 de Septiembre                NaN  Localidad          Colonia

      Clave_de_entidad  Nombre_de_entidad  Clave_de_municipio \
0                9  Ciudad de Mexico                16
1                9  Ciudad de Mexico                17
2                9  Ciudad de Mexico                 7
3                9  Ciudad de Mexico                 5
4                9  Ciudad de Mexico                16

      Borough  Clave_de_localidad  Nombre_de_localidad  Latitud \
0    Miguel Hidalgo                1    Miguel Hidalgo 19ř26'55'' N
1  Venustiano Carranza                1  Venustiano Carranza 19ř26'01'' N
2      Iztapalapa                1      Iztapalapa 19ř19'45'' N
3  Gustavo A. Madero                1  Gustavo A. Madero 19ř29'05'' N
4    Miguel Hidalgo                1    Miguel Hidalgo 19ř24'07'' N

      Longitud  Latitude  Longitude
0  99ř12'30'' W  19.448611  99.208333
1  99ř07'05'' W  19.433611  99.118056
2  99ř06'02'' W  19.329167  99.100556
3  99ř06'23'' W  19.484722  99.106389
4  99ř12'25'' W  19.401944  99.206944

```

### 2.1.1 Filter the DataFrame with the data of Borough,Neighborhood,Latitude and Longitude

```

In [10]: df = pd.DataFrame()
          df['Borough']=geo.Borough
          df['Neighborhood']=geo.Neighborhood
          df['Latitude']=geo.Latitude
          ###multiply longitud by -1 in order to use in Foursquare
          df['Longitude']=(geo.Longitude)*-1
          df.head()

```

```

Out[10]:
      Borough  Neighborhood  Latitude  Longitude
0    Miguel Hidalgo      10 de Abril  19.448611 -99.208333
1  Venustiano Carranza      10 de Mayo  19.433611 -99.118056
2      Iztapalapa     12 de Diciembre  19.329167 -99.100556
3  Gustavo A. Madero     15 de Agosto  19.484722 -99.106389
4    Miguel Hidalgo     16 de Septiembre  19.401944 -99.206944

```

Use geopy library to get the latitude and longitude values of Mexico City.

```

In [11]: address = 'Mexico City, MX'

```

```

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Mexico City are {}, {}'.format(latitude, longitude))

```

The geograpical coordinate of Mexico City are 19.4326009, -99.1333416.

### Create a map of Mexico City with neighborhoods superimposed on top.

```

In [25]: # create map of Mexico using latitude and longitude values
map_mx = folium.Map(location=[latitude, longitude], zoom_start=11.48)

# add markers to map
for lat, lng, borough, Neighborhood in zip(df['Latitude'], df['Longitude'], df['Borough'], df['Neighborhood']):
    label = '{} {}'.format(Neighborhood, borough)
    popup = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=popup,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_mx)

map_mx

```

Out[25]: <folium.folium.Map at 0x7f18f5e4e6d8>

## 3 Methodology

The main component of analysis in this study is by means of the Foursquare search tool.

Through this tool we will obtain the necessary information about the most visited venues in Mexico City.

### 3.0.2 Dimension of the Data.

The 16 Delegations or Boroughs that make up Mexico City have a number of neighborhoods, the list of these neighborhoods and their coordinates are obtained through the INEGI database mentioned in the previous section.

Therefore, when downloading said database, we obtain that in Mexico City there are 1232 neighborhoods.

Due to the immense number of neighborhoods, and therefore the queries of the neighborhood venues that will be on Foursquare exceeds the allowed number of daily consultations and customer requirements, We will analyze the main Boroughs that are Miguel Hidalgo and Cuauhtemoc to base our study

These Boroughs are the most representative of the restaurant sector of the city.

For this reason, my client wants this to be the area in which the behavior of the target businesses (Safood, Gourmet, Aged meat) is analyzed with respect to the most visited places in this area.

### **3.1 Clustering**

The process will consist in obtaining first the venues with the characteristics that we are looking for in your category.

once obtained these data, they will be contrasted with the most visited places in these Boroughs.

To carry out this analysis, a cluster will be made regarding the most visited venues in the Boroughs Miguel Hidalgo and Cuautemoc.

In this step we will compare how the most visited places are geographically related to the target venues in our project.

Once we have said analysis, we will obtain the information of the venues of interest and the most visited ones but for the Borough Benito Juarez, which is the Borough where the client wishes to open his new business.

### **3.2 Target Borough**

Within the Borough target that is Benito Juárez, it was found which is the neighborhood with the greatest potential due to its proximity to the top venues previously analyzed. Once the target neighborhood has been selected, the venues of that neighborhood are obtained and filtered with respect to which already exist that its category is similar to that which the client seeks to open.

In this case, you get what category of venue is needed in that neighborhood and what is the distribution of the other venues for that geographical area

## **4 Results**

In this analysis we were able to observe which are the largest venues in the city of Mexico that belong to the Cuautemoc Boroughs and Miguel Hidalgo, as well as where the areas of greatest concentration of people that can be potential clients are located.

The areas with the highest concentration of successful venues showed us a behavior with respect to the venues of the target category to which we want to describe.

It is common for these to be near the most visited venues or venues top, however there is some relaxation in distance with respect to these clusters.

The economic sector of the people who visit the venues of the two main Boroughs of this study is greater than what is intended to be opened in the Benito Juarez Borough.

The Narvarte Poniente neighborhood represents a potential area for a place specializing in aged meat and gourmet products, due to the closeness it has with well-visited venues and the fact that there is no such business in the neighborhood.

## **5 Discussion**

### **5.0.1 Access to Foursquare**

One of the issues to improve the study is to have greater access to queries in Foursquare to have a larger database and to make predictions with more formidable data.

The large number of neighborhoods that Mexico City contains makes it more convenient to do this analysis with queries on Foursquare that require premium queries.

### **5.0.2 Venues categories**

Specialized gourmet places in Mexico City sometimes do not carry a good categorization in the field of venue category that offers foursquare. It is useful in future studies access to data that are not specifically of this tool to give more support to the real category that the venues have.

## **6 Conclusion**

From the study we can conclude that the Benito Juarez Borough is a potential location for the establishment of gourmet venues due to the good geographic location that exists with the most frequented venues of the clientele that prefers these types of places for consumption.

It is considered that the location within the neighborhood Narvarte Poniente is better because the category of venues in that area still does not exploit the specialized gourmet sites within the area, although if there are only two places of Seafood and Steakhouse but no precisely meet the specialized gourmet category sought by our client