COMPARING BIG CITIES

ESSENTIAL INFORMATION ABOUT CITIES IS VALUABLE FOR INVESTORS AND INTERESTED PEOPLE

- In order to open new venues we need to have a sense of the most popular ones.
- If one venue is successful in one city, it might be successful in another similar city.
- It is necessary to open a new venue in a place where it is not crowded

DATA ACQUISITION AND CLEANING

- I had some really hard problems trying to get the data from Neighborhoods.
- For California neighborhoods:

  https://www.geonames.org/postal-codes/US/CA/california.html

- For Toronto neighborhoods:

  https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- For Manhattan neighborhoods:

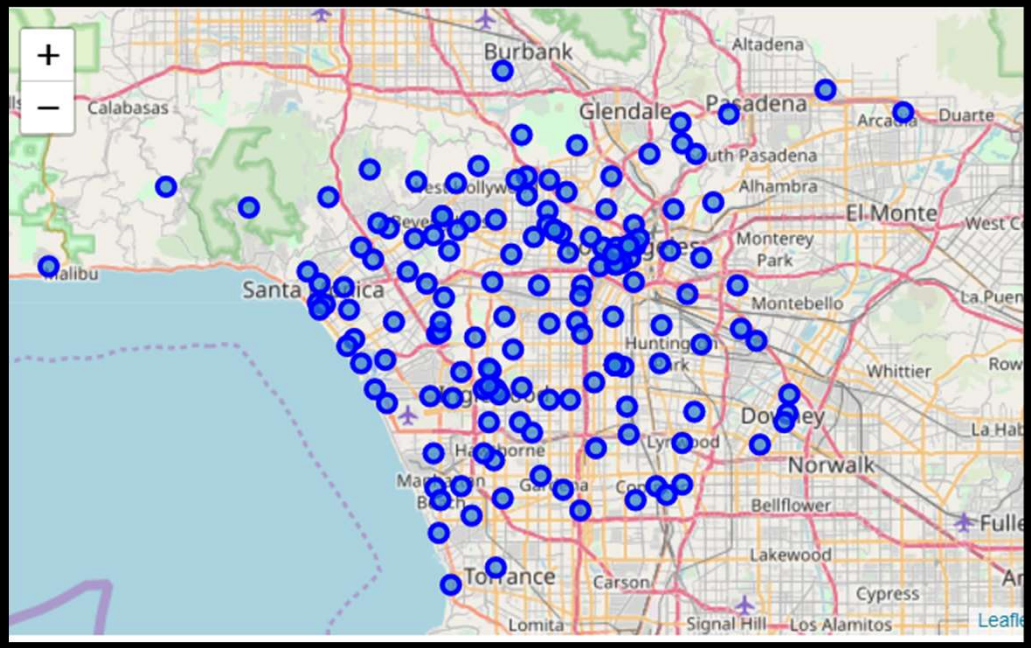
  https://cocl.us/new_york_dataset

- To get the coordinates I used the Arcgis method

USIGN THE ARCGIS METHOD

```
import geocoder

Code_1=df_clean['Code']


for postcode in Code_1:
    g= geocoder.arcgis(postcode)#maybe we need to change equal to the format below
    current_g= g.json
    neighborhood_lat= current_g['lat']
    neighborhood_lon= current_g['lng']

    df1 = df1.append({'Latitude':neighborhood_lat,
                                'Longitude': neighborhood_lon}, ignore_index=True)
```

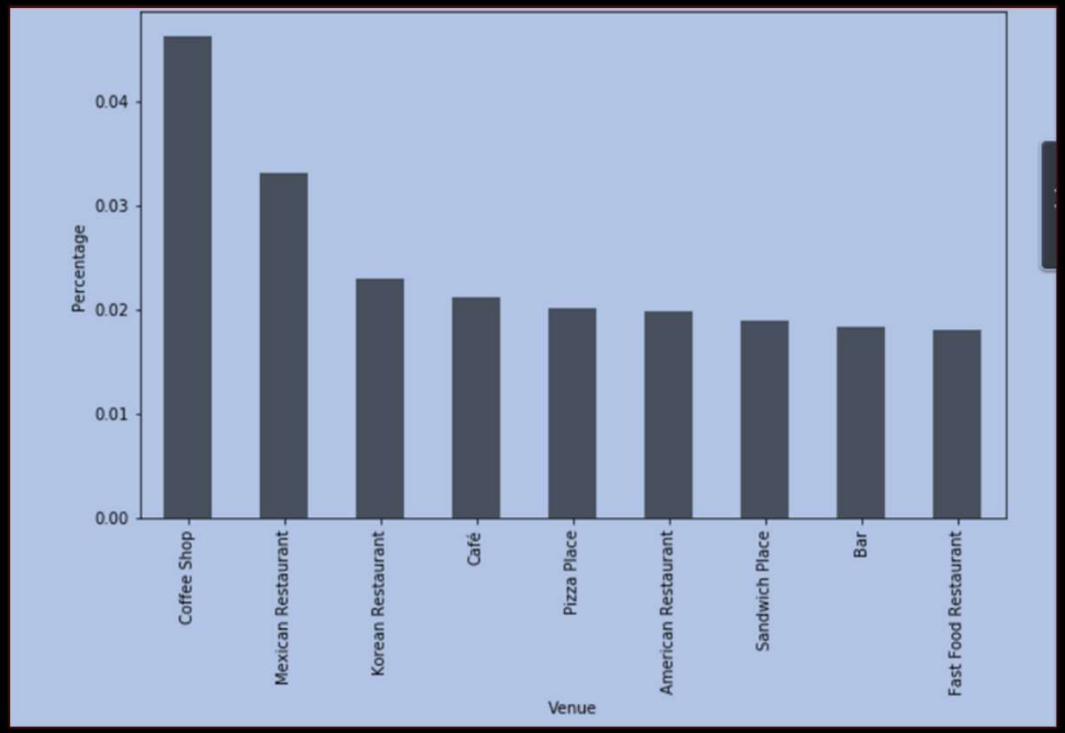
CALIFORNIA


Burbank
Altadena
Calabasas
Glendale
Pasadena
Arcadia
Duarte
Alhambra
El Monte
Monterey Park
Montebello
Whittier
Norwalk
Bellflower
Lakewood
Torrance
Carson
Cypress
Lomita
Signal Hill
Los Alamitos

```
# one hot encoding
California_onehot = pd.get_dummies(California_venues[['Venue Category']], prefix="", pr

# add neighborhood column back to dataframe
California_onehot['Neighbourhood'] = California_venues['Neighbourhood']

# move neighborhood column to the first column
fixed_columns = [California_onehot.columns[-1]] + list(California_onehot.columns[:-1])
California_onehot = California_onehot[fixed_columns]

California_onehot.head()
```

	Neighbourhood	ATM	Accessories Store	Acupuncturist	Afghan Restaurant	African Restaurant	Airport	Airport Terminal	Am Resta
0	Beverly Hills	0	0	0	0	0	0	0	
1	Beverly Hills	0	0	0	0	0	0	0	
2	Beverly Hills	0	0	0	0	0	0	0	
3	Beverly Hills	0	0	0	0	0	0	0	
4	Los Angeles	0	0	0	0	0	0	0	

Using one hot encoding for California venues



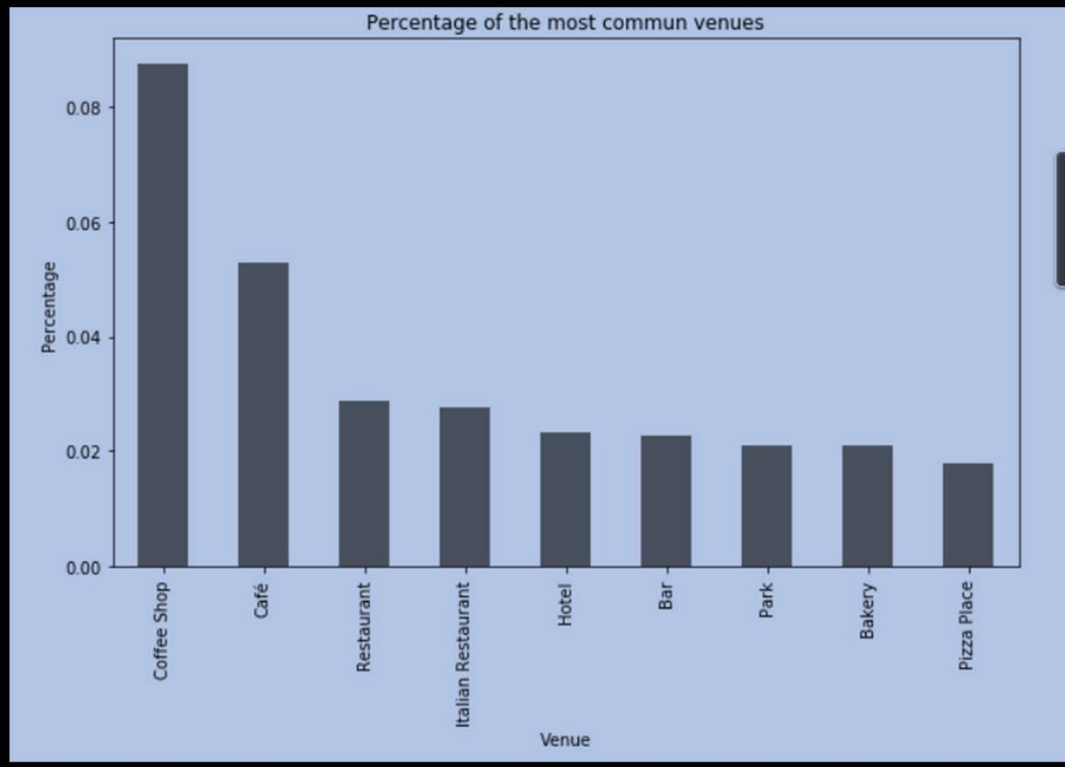
Most popular venues in California

TORONTO

```
# one hot encoding
Toronto_onehot = pd.get_dummies(Toronto_venues[['Venue Category']], prefix="", prefix_s

# add neighborhood column back to dataframe
Toronto_onehot['Neighbourhood'] = Toronto_venues['Neighbourhood']

# move neighborhood column to the first column
fixed_columns = [Toronto_onehot.columns[-1]] + list(Toronto_onehot.columns[:-1])
Toronto_onehot = Toronto_onehot[fixed_columns]

Toronto_onehot.head()
```

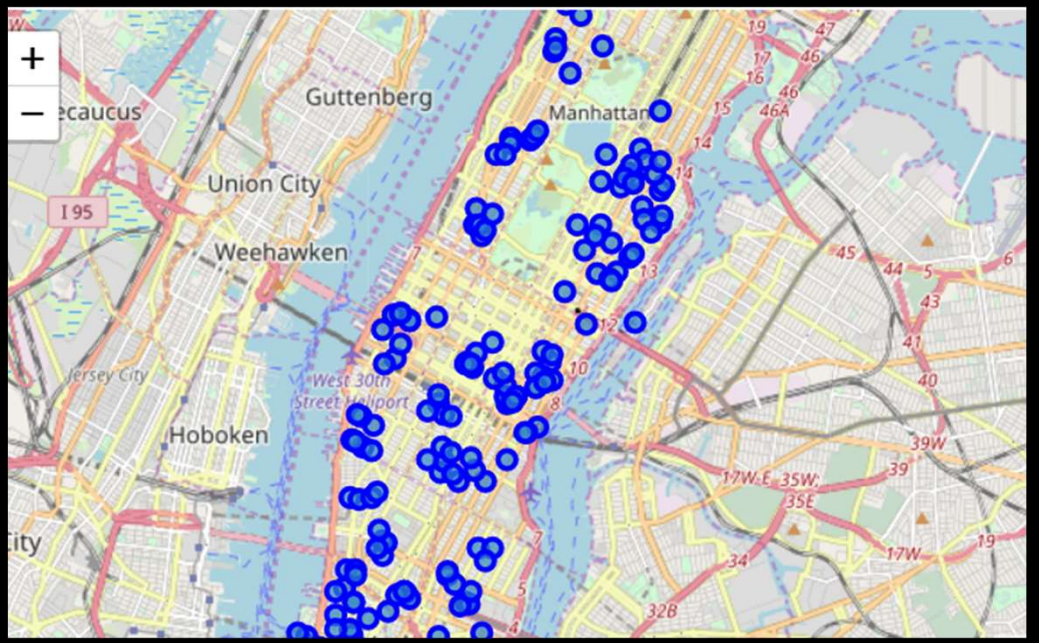
	Neighbourhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant
0	Harbourfront,Regent Park	0	0	0	0	0	0	0	0
1	Harbourfront,Regent Park	0	0	0	0	0	0	0	0
2	Harbourfront,Regent Park	0	0	0	0	0	0	0	0
3	Harbourfront,Regent Park	0	0	0	0	0	0	0	0
4	Harbourfront,Regent Park	0	0	0	0	0	0	0	0

Using one hot encoding for Toronto venues



Most popular venues in Toronto

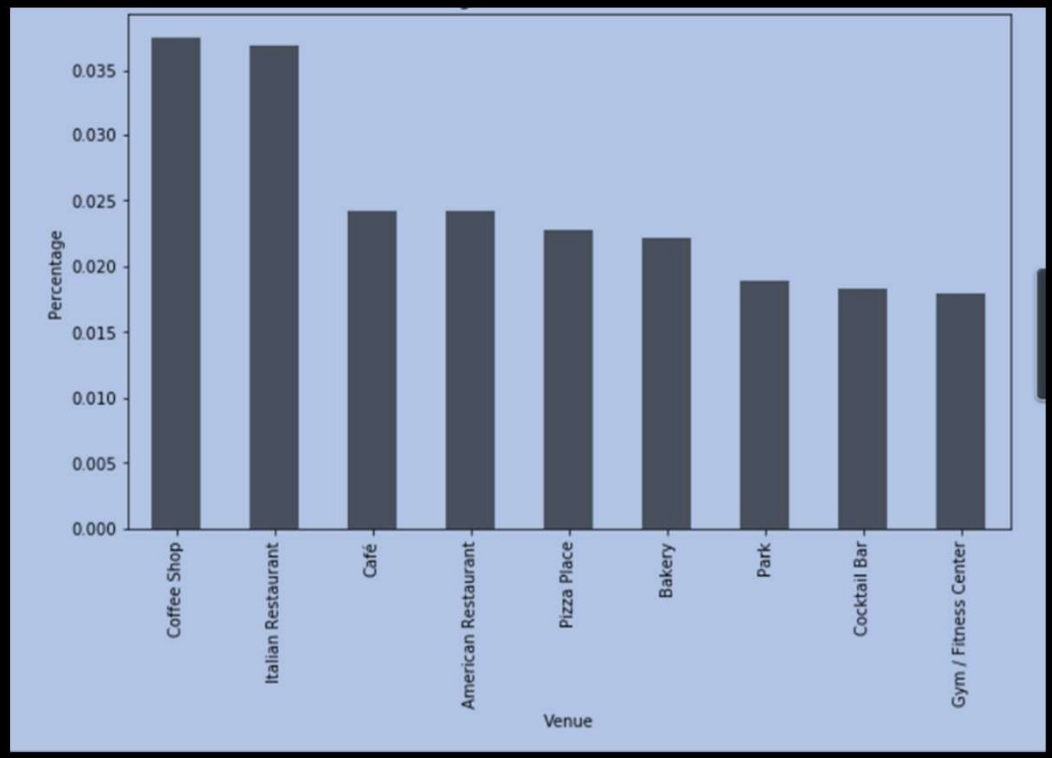
MANHATTAN


Guttenberg
Manhattan
Union City
Weehawken
Hoboken
West 30th Street Heliport
Jersey City
I 95

```
# one hot encoding
manhattan_onehot = pd.get_dummies(manhattan_venues[['Venue Category']], prefix="", pre

# add neighborhood column back to dataframe
manhattan_onehot['Neighborhood'] = manhattan_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [manhattan_onehot.columns[-1]] + list(manhattan_onehot.columns[:-1])
manhattan_onehot = manhattan_onehot[fixed_columns]

manhattan_onehot.head()
```

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	A Resta
0	Marble Hill	0	0	0	0	0	0	0	
1	Marble Hill	0	0	0	0	0	0	0	
2	Marble Hill	0	0	0	0	0	0	0	
3	Marble Hill	0	0	0	0	0	0	0	
4	Marble Hill	0	0	0	0	0	0	0	

Using one hot encoding for Manhattan venues

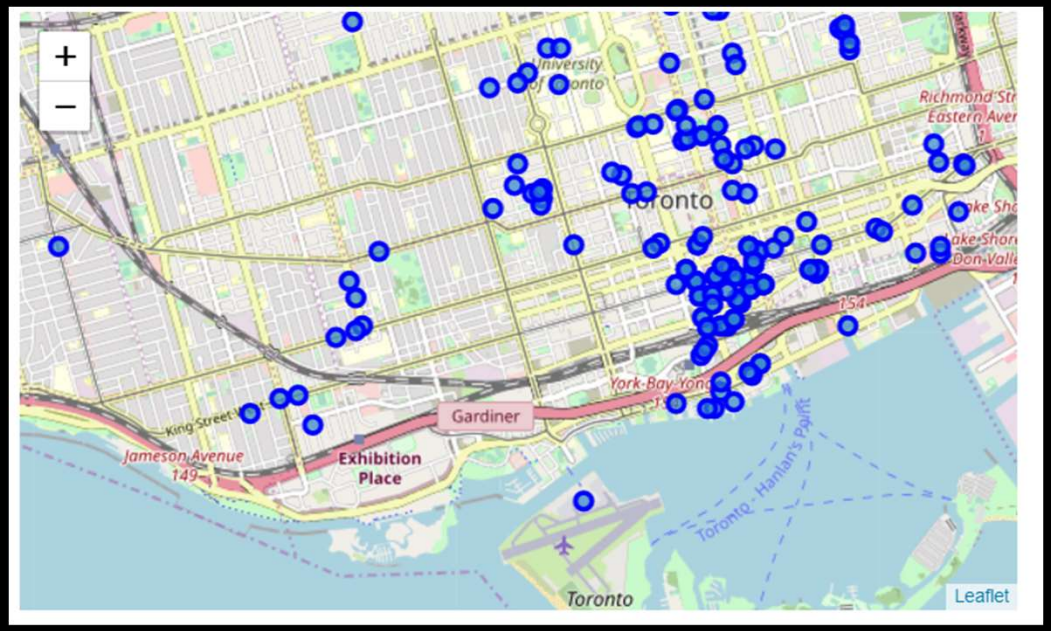


Most popular venues in Manhattan

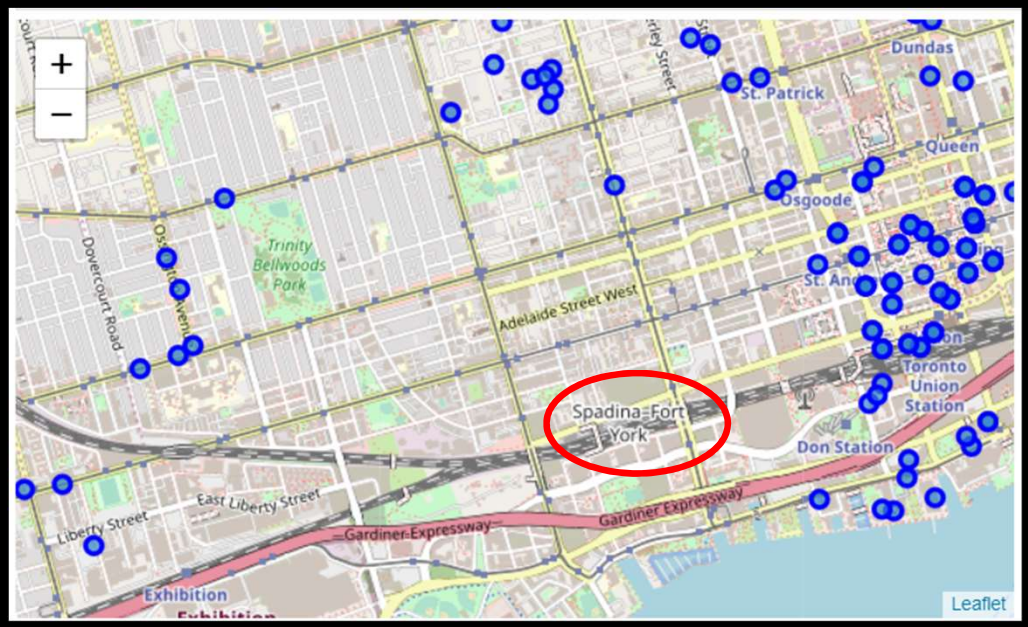
MOST POPULAR VENUE

Coffee Shops

Only Coffee Shops in Toronto



Downtown is crowded with coffee shops

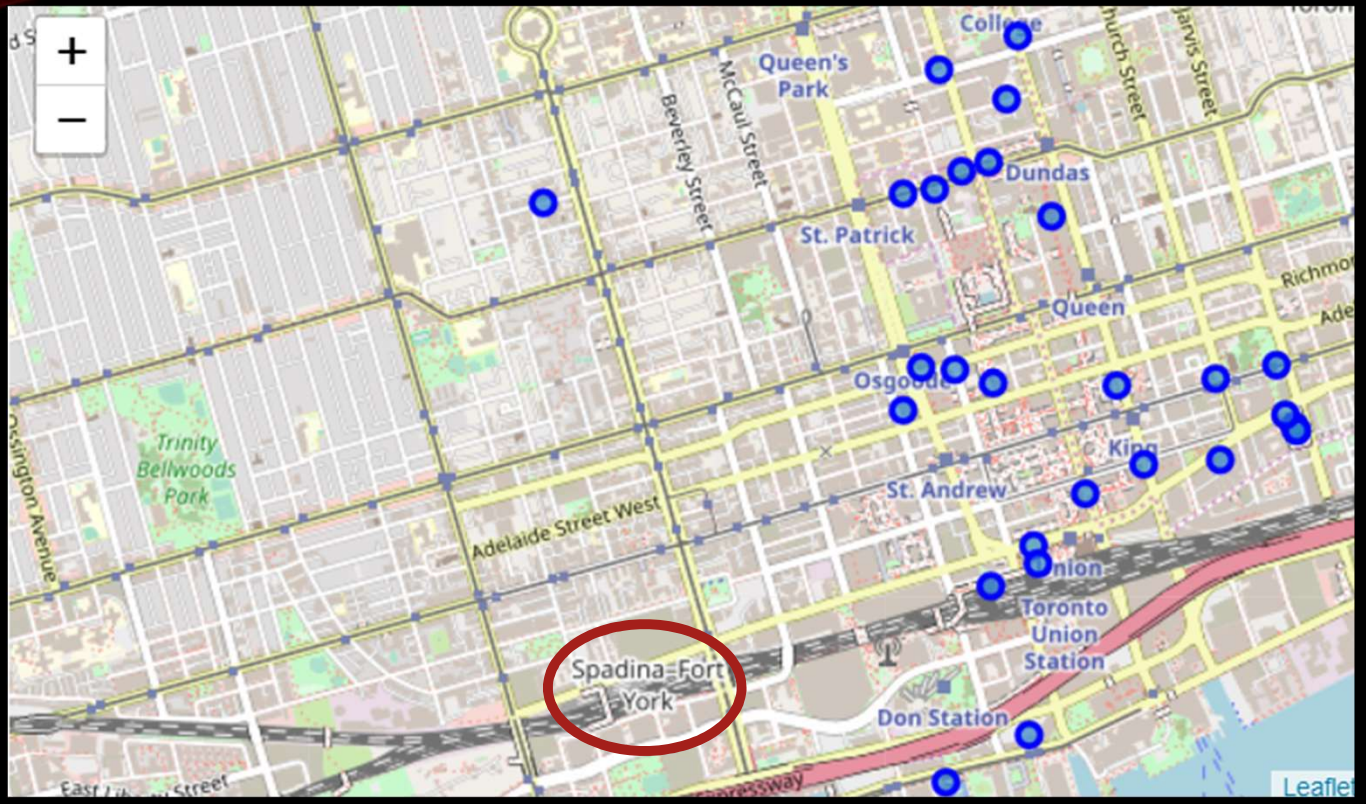


Near Spadina Fort York would be a good place to open a coffee shop

PIZZA PLACE

It might be a good idea to open a Pizza place in Toronto

Pizza Places in Toronto



NEAR SPADINA FORT YORK IS AGAIN A CONVENIENT PLACE TO OPEN A NEW VENUE

CONCLUSION

- The model gave some useful data from the cities and showed a map with their locations
- We built a relatively simple model that will help us to choose the most profitable venue.
- It also suggests some places where it would be adequate to open a venue.
- There is plenty of room for improvent, specially in the complexity of algorithms, but I used what they taught us stricly.