Opening a Hungarian restaurant in New York

•••

Coursera Capstone project

Goal is to recommend a suitable place in NYC

The project's main goal is to recommend an area in NYC where an owner could open up a new Hungarian restaurant or venue to eat.

This is achieved by:

- Setting up methods / assumptions based on which to do the analysis
- Gathering data and respective data sources to be used
- Analyze data to find suitable areas
- Recommend one or more areas for the owner

Only NYC is considered (not greater metropolitan) and the analysis is simplified - there are more than 26'000 restaurants!

Goal is to recommend a suitable place in NYC

The project's main goal is to recommend an area in NYC where an owner could open up a new Hungarian restaurant or venue to eat.

This is achieved by:

- Setting up methods / assumptions based on which to do the analysis
- Gathering data and respective data sources to be used
- Analyze data to find suitable areas
- Recommend one or more areas for the owner

Only NYC is considered (not greater metropolitan) and the analysis is simplified - there are more than 26'000 restaurants!

Taking two approaches (always have a plan B :)

Method A - Finding areas with already operating similar restaurants

- Gather New York Borough and neighborhood data
- For each 'hood find Hungarian and related (Czech, Slovak, Polish, Romanian) restaurants
- Explore the result set and find high and low concentration of restaurants

If Method A does not yield a meaningful result, find neighborhoods with the highest density of restaurants in Manhattan.

- Only Manhattan areas are considered
- Take the 100 top venue in each 'hood
- Do a one-hot encoding
- Given the number of venues (and licensing on Foursquare API) small number of clusters are used - group them into 3 clusters by K-Means
- Analyze the clusters
- Recommend an area or neighborhood

Method A - finding Central European restaurants

```python
In [12]: #so there are 306 new york neighborhoods, need to look up Hungarian and other central / eastern european restaurants
         # reading Hungarian, Romanian, Czech, Slovakian and Polish restaurants in one go rather than having 5 calls to the API.
         #(While there is an Eastern European category in Foursquare it is not the one we're looking for)

         column_names=['Borough', 'Neighborhood', 'Latitude','Longitutde', 'Hun', 'CEE']
         restaurants = pd.DataFrame(columns = column_names)

         cee = ["Hungarian Restaurant","Czech Restaurant","Slovak Restaurant", "Polish Restaurant", "Romanian Restaurant"]

         #iterating through all neighborhoods, getting venues and adding 2 columns, Hun count and CE count (Hun + all others)

         for row in new_york_data.values.tolist():
             Borough, Neighborhood, Latitude, Longitude=row
             venues = get_venues(Latitude,Longitude)
             hun_restaurants =venues[venues['Category']=='Hungarian Restaurant']
             cee_restaurants=venues[venues['Category'].isin(cee)]

             print('Hungarian Resturants in '+Neighborhood+', '+Borough+':'+str(len(hun_restaurants)))
             print('CEE Resturants in '+Neighborhood+', '+Borough+':'+str(len(cee_restaurants)))
             #adding a new row to our restaurants frame
             restaurants = restaurants.append({'Borough': Borough,
                                   'Neighborhood': Neighborhood,
                                   'Latitude': Latitude,
                                   'Longitutde' : Longitude,
                                   'Hun' : len(hun_restaurants),
                                   'CEE' : len(cee_restaurants)
                                                    }, ignore_index=True)


         Hungarian Resturants in Wakefield, Bronx:0
         CEE Resturants in Wakefield, Bronx:0
         Hungarian Resturants in Co-op City, Bronx:0
         CEE Resturants in Co-op City, Bronx:0
         Hungarian Resturants in Eastchester, Bronx:0
         CEE Resturants in Eastchester, Bronx:0
         Hungarian Resturants in Fieldston, Bronx:0
```

```
CEE Resturants in Bronxdale, Bronx:0
Hungarian Resturants in Allerton, Bronx:0
CEE Resturants in Allerton, Bronx:0
Hungarian Resturants in Kingsbridge Heights, Bronx:0
CEE Resturants in Kingsbridge Heights, Bronx:0
Hungarian Resturants in Erasmus, Brooklyn:0
CEE Resturants in Erasmus, Brooklyn:0
Hungarian Resturants in Hudson Yards, Manhattan:0
CEE Resturants in Hudson Yards, Manhattan:0
Hungarian Resturants in Hammels, Queens:0
CEE Resturants in Hammels, Queens:0
Hungarian Resturants in Bayswater, Queens:0
CEE Resturants in Bayswater, Queens:0
Hungarian Resturants in Queensbridge, Queens:0
CEE Resturants in Queensbridge, Queens:0
Hungarian Resturants in Fox Hills, Staten Island:0
CEE Resturants in Fox Hills, Staten Island:0
```

```python
In [13]: restaurants.shape

Out[13]: (306, 6)

In [14]: restaurants.head()

Out[14]:
```

	Borough	Neighborhood	Latitude	Longitutde	Hun	CEE
0	Bronx	Wakefield	40.894705	-73.847201	0	0
1	Bronx	Co-op City	40.874294	-73.829939	0	0
2	Bronx	Eastchester	40.887556	-73.827806	0	0
3	Bronx	Fieldston	40.895437	-73.905643	0	0
4	Bronx	Riverdale	40.890834	-73.912585	0	0

Seems like Central Europeans prefer Jersey and Chicago

Method A did return a meaningful result, but does not help to locate a neighborhood.

- There are no dedicated Hungarian restaurants in NYC (there are though in Jersey)
- There are only 9 places dedicated to Central European cuisine and these are scattered in 7 neighborhoods in 4 boroughs.
- 3 in one neighborhood is not significant enough

'Hoods with restaurants of interest:

- Greenpoint / Brooklyn : 3
- Arrochar / Staten Island : 1
- Blissville / Queens : 1
- Lenox Hill / Manhattan : 1
- Ridgewood / Queens : 1
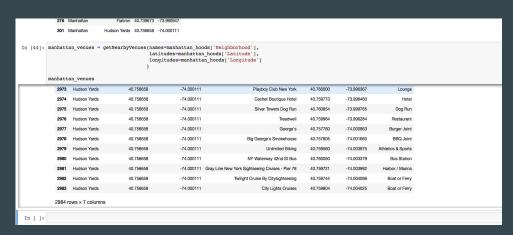- Roosevelt Island / Manhattan : 1
- Steinway / Queens : 1

Finding Manhattan areas with lots of restaurants

This was achieved in a similar way as discussed during the course:

- Filtering our source list for Manhattan neighborhoods only
- For each 'hood read the top 100 venues
- Group the venues by categories
- Run a one hot encoding and K-Means clustering (taking K=3 for simplicity)
- Visualize the results
- Analyze the clusters to pinpoint areas
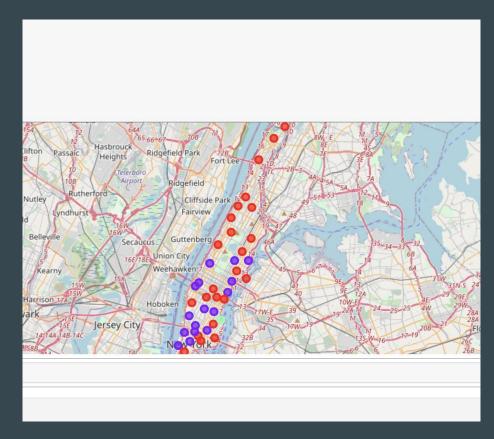
Finding Manhattan areas with lots of restaurants

- 2984 venues
- 321 unique categories -> this diversity can affect our results as well

Finding Manhattan areas with lots of restaurants

- 3 clusters, one can be excluded
- There are minor / subtle differences between the 2 interesting clusters
  - More upscale and concentrated more (not only) on lower Manhattan
  - Less upscale with more cafes, bakeries, delis and more scattered in high-end residential areas
- We can recommend areas based on the type of place the owner wants to open

Recommendation based on what kind of restaurant to open

If the owner wants to open a more upscale niche restaurant with more refined attributes:

- Upper East Side
- Greenwich Village
- Soho
- West Village

If the owner wants to open a lower profile restaurant or cafe, deli, bakery:

- Marble Hill
- Washington Heights
- East Village