

# 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

```
In [12]: #so there are 106 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', 'Longitude', '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 CEE 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 Restaurants in '+Neighborhood+', '+Borough*':'+str(len(hun_restaurants)))
    print('CEE Restaurants in '+Neighborhood+', '+Borough*':'+str(len(cee_restaurants)))
    #adding a new row to our restaurants frame
    restaurants = restaurants.append({'Borough': Borough,
                                     'Neighborhood': Neighborhood,
                                     'Latitude': Latitude,
                                     'Longitude': Longitude,
                                     'Hun': len(hun_restaurants),
                                     'CEE': len(cee_restaurants)
                                     }, ignore_index=True)
```

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

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

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

```
Out[13]: (306, 6)
```

```
In [14]: restaurants.head()
```

```
Out[14]:
```

	Borough	Neighborhood	Latitude	Longitude	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

```
276 Manhattan Flatiron 40.739673 -73.990947
301 Manhattan Hudson Yards 40.756658 -74.000111
```

```
In [44]: manhattan_venues = getNearbyVenues(names=manhattan_hoods['Neighborhood'],
                                             latitudes=manhattan_hoods['Latitude'],
                                             longitudes=manhattan_hoods['Longitude'])
```

manhattan_venues						
2973	Hudson Yards	40.756658	-74.000111	Playboy Club New York	40.760000	-73.996367 Lounge
2974	Hudson Yards	40.756658	-74.000111	Cachet Boutique Hotel	40.759773	-73.996460 Hotel
2975	Hudson Yards	40.756658	-74.000111	Silver Towers Dog Run	40.760854	-73.999765 Dog Run
2976	Hudson Yards	40.756658	-74.000111	Treadwell	40.759964	-73.996284 Restaurant
2977	Hudson Yards	40.756658	-74.000111	George's	40.757760	-74.000963 Burger Joint
2978	Hudson Yards	40.756658	-74.000111	Big George's Smokehouse	40.757805	-74.001660 BBQ Joint
2979	Hudson Yards	40.756658	-74.000111	Unlimited Biking	40.759560	-74.003975 Athletics & Sports
2980	Hudson Yards	40.756658	-74.000111	NY Waterway 42nd St Bus	40.760050	-74.003379 Bus Station
2981	Hudson Yards	40.756658	-74.000111	Gray Line New York Sightseeing Cruises - Pier 78	40.759721	-74.003962 Harbor / Marina
2982	Hudson Yards	40.756658	-74.000111	Twilight Cruise By Citysightseeing	40.759744	-74.004096 Boat or Ferry
2983	Hudson Yards	40.756658	-74.000111	City Lights Cruises	40.759804	-74.004025 Boat or Ferry

2984 rows x 7 columns

```
In [ ]:
```

```
In [46]: # one_hot_encoding
manhattan_onehot = pd.get_dummies(manhattan_venues[['Venue Category']], prefix="", prefix_sep="")

# 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()
```

```
Out[46]:
```

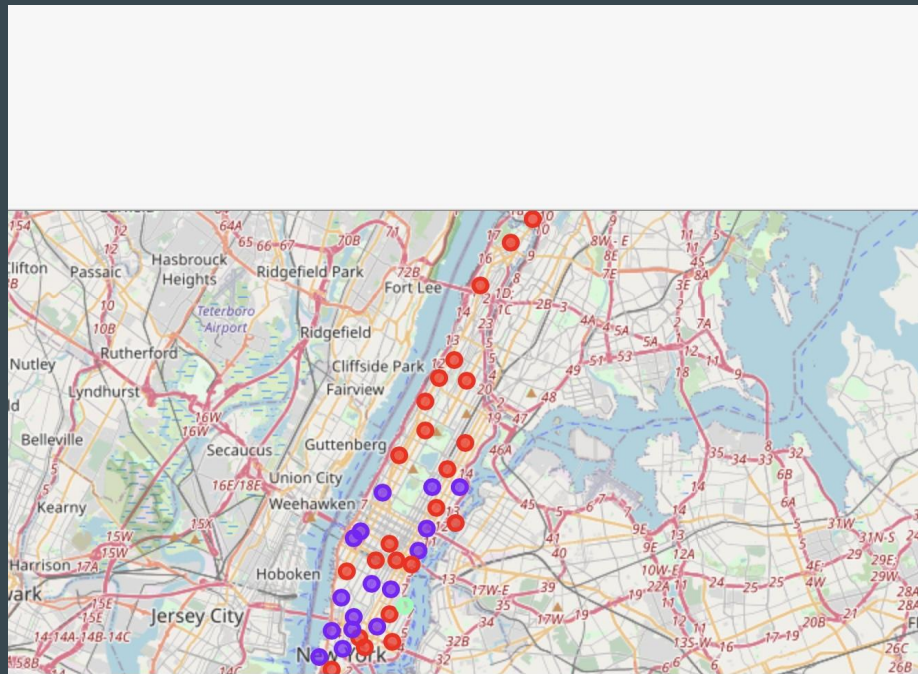
	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Araps Restaurant	Argentinian Restaurant	...	Video Store	Vietnamese Restaurant	Volleyball Court	Waterfront	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yo
0	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows x 322 columns



# 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