Introduction: Business Problem

In this project I will find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening a **Mexican Food** in **Newcastle**, UK.

Since there are lots of restaurants in Newcastle, we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Italian restaurants in vicinity**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Based on definition of our problem, factors that will influence our decision are:

* number of existing restaurants in the neighborhood (any type of restaurant)
* number of and distance to Italian restaurants in the neighborhood, if any
* distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

* centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
* number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
* coordinate of Newcastle center will be obtained using **Google Maps API geocoding** of well-known Newcastle location.

Neighborhood Candidates

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is approx. 12x12 kilometers centered around Newcastle city center.

Let's first find the latitude & longitude of Newcastle city center, using specific, well known address and Google Maps geocoding API.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Exploring All Venues in Newcastle

In the code below, API calls are done for each neighborhood in the first DataFrame. These JSON files are appended to a list, including the neighborhood of each venue. Just like the code above, these JSON files are converted and appended to a single DataFrame and formatted in such a way that the neighborhood, category and geographical coordinates of each venue are shown.

	Neighborhood	Venue	Venue Category	Venue Latitude	Venue Longitude
0	Malvern, Rouge	Wendy's	Fast Food Restaurant	43.807448	-79.199056
1	Malvern, Rouge	Interprovincial Group	Print Shop	43.805630	-79.200378
2	Rouge Hill, Port Union, Highland Creek	Royal Canadian Legion	Bar	43.782533	-79.163085
3	Rouge Hill, Port Union, Highland Creek	Affordable Toronto Movers	Moving Target	43.787919	-79.162977
4	Guildwood, Morningside, West Hill	RBC Royal Bank	Bank	43.766790	-79.191151

Methodology

Clustering the Neighborhoods in Newcastle

Because of the dummy variables of the venue categories, I am able to apply the K-Means Clustering model to the dataset. For this dataset, I have used 4 clusters. The clusters are shown in the map below. Note: Some neighborhoods had "NaN" als value in Cluster Labels because these neighborhoods don't have any nearby venues. These are grouped as cluster 5 in order to convert the values as int but aren't used in the analysis.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1	2	3	4
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	1	Fast Food Restaurant	Print Shop		
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	1	Bar	Moving Target		
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	1	Bank	Breakfast Spot	Electronics Store	Intersection
3	M1G	Scarborough	Woburn	43.770992	-79.216917	1	Coffee Shop	Indian Restaurant	Korean Restaurant	
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	1	Athletics & Sports	Bakery	Bank	Caribbean Restaurant

Results
Analyzing Clusters

In order to analyze these four clusters, I have created a function that returns (1) the number of neighborhoods in each cluster, and (2) the top 10 most common venues in each cluster. This data gives us an idea of how these clusters differ from each other.

It turns out that the first cluster contains neighborhoods that include lots of venues where you are able to have a lunch or dinner. The second cluster contains neighborhoods that mainly include a park and some coffee shops, cafes and restaurants. The third cluster contains five neighborhoods that have a wide variation of nearby venues, including a yoga studio and a theater. The fourth cluster contains a wide variation of restaurants, including some vegetarian restaurants.

	Venue	Count
0	Coffee Shop	22
1	Café	19
2	Restaurant	17
3	Italian Restaurant	16
4	Sandwich Place	14
5	Pizza Place	12
6	Sushi Restaurant	10
7	Grocery Store	10
8	Park	9
9	Diner	9

	Venue	Count
0	Park	22
1	Coffee Shop	16
2	Pizza Place	15
3	Bank	14
4	Sandwich Place	11
5	Fast Food Restaurant	10
6	Grocery Store	9
7	Pharmacy	9
8	Liquor Store	9
9	Intersection	8

	Venue	Count
0	Café	6
1	Restaurant	6
2	Coffee Shop	6
3	Vegetarian / Vegan Restaurant	5
4	Seafood Restaurant	5
5	Art Gallery	5
6	Japanese Restaurant	4
7	Hotel	4
8	Gastropub	4
9	American Restaurant	4

Conclusion

As we can see the outputs of the different clustering give us the best option of places around the city center of Newcastle.