**An English man in New York**

**Capstone Project: The Battle of Neighborhoods.**

This is the final assignment of the Applied Data Science Capstone Course by IBM on Coursera.

**Introduction**

*Introduction; where you discuss the business problem and who would be interested in this project.*

Meaning of “Englishman in New York” song by Sting

Englishman in New York is a song performed by English singer and songwriter Sting. The self-explanatory lyrics of the song talk about the life of an Englishman who has moved from his native England to live in New York City. The “Englishman” in the song is the famous English gay icon, writer and actor Quentin Crisp, who was the inspiration of the song. In 1986, Crisp relocated from London, England to Manhattan, New York City, and Sting wrote the song shortly after Crisp’s relocation.

In Sting’s 2007 book titled Lyrics, the singer said he felt so homesick the first time he moved to live in New York that he was forced to move from one English pub to the other just so he could meet with other Englishmen and enjoy the feeling of being home.

<https://www.songmeaningsandfacts.com/meaning-englishman-new-york-sting/>

So, let's go on a short journey to discover such places in New York City. Restaurants, Pubs, Fish and Chip shops, etc. This would be interesting if you were commissioned to write an article or tourist information brochure. But let's consider that a friend is looking to open their own such establishment, for which we need to do a small amount of analysis to get things started.

**Data**

*Data; where you describe the data that will be used to solve the problem and the source of the data.*

New York City data containing the neighborhoods and boroughs, longitudes and latitudes will be obtained from the data source: <https://cocl.us/new_york_dataset>

All data related to locations and quality of English restaurants will be obtained via the FourSquare API. The data in FourSquare under the category "English Restaurant" is actually quite limited. But there is also the category "Pub" and "Fish & Chips Shop", which are also terms that may uniquely describe an English restaurant. So, we will include all three categories when collecting data.

NOTE: The following developer resource on ForeSquare can be used to determine available venue categories.

<https://developer.foursquare.com/docs/build-with-foursquare/categories/>

**Methodology**

*Methodology section; which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.*

I will use this section to describe each relevant step of the processing (python code sections), AKA "exploratory data analysis".

Use a function to get initial New York data

This is where we read New York data from the data source: <https://cocl.us/new_york_dataset> and create a dataset containing the neighborhoods and boroughs, longitudes and latitudes.

Discover the total of number of different Neighborhoods in New York

Basically, the neighborhoods in New York are located within five boroughs. Bronx, Brooklyn, Manhattan, Queens and Staten Island. The following Wikipedia link helps to visualise this;

<https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City>

Perform initial data analysis

Here we graph the number of neighborhoods in each borough. Which has the most? Which has the least? Is this significant in our data analysis?

Further analysis to see how many English Restaurants there are in each neighborhood and borough.

This is where we try to determine where exactly in these New York boroughs/neghbourhoods our English restaurants reside. Mainly due to English terminology, we are using the categories 'English Restaurant', 'Pub', and 'Fish & Chips Shop' to determine all potential English Restaurants. This is also where we do the majority of your FourSquare API calls, which we save to a csv file in the next step to limit further calls.

Discover the locale of said English Restaurants

Now we graph the data we have collected. English restaurants per borough/neghbourhood (two graphs). This helps visualise popular places for such an establishment.

Discover the ratings of said English Restaurants

In this next part of our analysis, we collate the number of likes, ratings, and tips for each English restaurant.

Discover which English Restaurants have the greatest number of Likes, Ratings, Tips

The number of Likes, Ratings, and Tips is a good indication of how busy or popular a venue may be.

Discover the neighborhoods with the best average ratings of restaurants and discover the boroughs with the best average ratings of restaurants and visualize in a graph

These two parts of the analysis may be quite significant. People (both tourists and locals) will likely gravitate to areas where ratings are high.

Discover all the neighborhoods with average rating greater or equal to 7.5 (upper quartile) to visualize on a map

So now we take it to the next level and subsequently visualise on a map, but only those venues in the upper quartile for ratings.

## Results / Discussion / Observations

*Results section where you discuss the results, any observations you noted and any recommendations you can make based on the results.*

So, the total number of English restaurants is 34. As stated before, this includes venues from three FourSquare categories, "English restaurant", "Pub" and "Fish & Chips Shop". As potentially these could all be classed as English restaurants. I note that one venue in our dataset is called "Woodrow Fish Market". Which doesn't sound like a restaurant. So, we may need to refine our dataset further to validate each venue.

Of those 34, most appear to be in the borough of Manhattan. More specifically, in the neighbourhood of Murray Hill. Although Woodlawn in Queens appears to have quite a few also. Again, more analysis could be done here to determine precisely what type of English restaurants reside in which borough/neighbourhood. For example, one restaurant may serve a lot of steak dishes, another may provide a lot of choice in real ale and pies. And are such traits determined by locale?

Now imagine if the same restaurant transpired to have the most amount of Likes and tips, AND the best Ratings. Surely this would be the place to go as an English man in New York. Or perhaps a good business model to use when opening your own establishment? Well, there is one such establishment, the Cask Bar & Kitchen. And where is it? Murray Hill in Manhattan of course.

Interestingly though, when we look at neighborhoods with the best average ratings of restaurants, Murray Hill is not on the list. What does this mean? After all, this is where the most English restaurants reside. Perhaps all other English restaurants in this area have poor ratings? Or maybe there are a lot of new restaurants that quite simply have no ratings yet.

However, the borough of Manhattan does have the best average ratings of restaurants.

Finally, the map we generated shows neighborhoods with average rating greater or equal 7.5 (upper quartile). This seems to give a fairly widespread, so perhaps we need to narrow this down my increasing the average rating parameter? However, this has probably removed all the neighbourhoods that are not worth visiting.

## Conclusion

*Conclusion section where you conclude the report.*

The results of our analysis have given us a lot of useful information and insight. However, it has also raised a lot of questions. But this is good, as these are questions that we may not have previously thought of. And my friend does need to do make sure the research is thorough before investing any time and money. At the very least, it has given us some specific venues and neighbourhoods that are worth visiting while continuing said research.

The next step would be to refine our dataset / data analysis further. For instance, I did not use any machine learning, as I did not think it would be useful this time. So perhaps we could try a different analysis method and compare the results?

But we can also use what we have as the basis for more traditional research. For example, consider the following article.

The Five Boroughs of New York City <https://www.nycgo.com/neighborhoods-boroughs/about-nyc-five-boroughs/>

This describes the boroughs in great detail. Are there any parallels between this and our data analysis?

Just for fun, let's also consider the following film.

Gangs of New York: <https://en.wikipedia.org/wiki/Gangs_of_New_York>

"In the 1846 slum neighborhood of Five Points, Manhattan, two gangs, the Nativist Protestants, led by William "Bill the Butcher" Cutting, and a group of Irish Catholic immigrants, the "Dead Rabbits", led by "Priest" Vallon, engage in a battle to determine which faction will hold sway over the territory."

Does history have a part to play on the location of said venues? For example, there is no real difference between an English Pub and an Irish Pub. The food, drink and atmosphere are really quite similar. If a certain neighborhood has a certain amount of English or Irish history, this would be good for business.