battle_of_neighborhoods_part1

August 1, 2019

0.1 Capstone Project - The Battle of Neighborhoods (Week 1)

For this week, you will required to submit the following: Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

This submission will eventually become your Introduction/Business Problem section in your final report. So I recommend that you push the report (having your Introduction/Business Problem section only for now) to your Github repository and submit a link to it.

0.2 1.Description of the problem

We will explored New York City and London and segmented and clustered their neighborhoods. Both cities are very diverse and are very similar. Both cities are a densely populated boroughs that's among the world's major commercial, financial and cultural centers. . We will to compare the neighborhoods of the two cities and determine how similar or dissimilar they are. We will define that people like to do more in the cities, which places are often visited. Knowing this information we can think of how to use this. For exemple, open a new restaurant or supermarket, entertainment center or gift shop. As we can see in the next task that although there are Mexican restaurants in London, but they are not popular, entertainment is centrally located and almost none in areas farther from the center. We may also use this information for advertising purposes, etc

.

0.3 2. Description of Data.

This project will rely on public data from Wikipedia and Foursquare.

London is the capital of and largest city in England and the United Kingdom. It is administered by the City of London and 32 London boroughs.

We will get information about the areas of London https://en.wikipedia.org/wiki/List_of_areas_of_London

London dataset:

	neighborhoods	Dial code	borough	posttown	postcode	Latitude	Longitude
0	Aldwych	020	Westminster	LONDON	WC2	51.51651	-0.11968
1	Bayswater	020	Westminster	LONDON	W2	51.51494	-0.18048
2	Bedford Park	020	Ealing	LONDON	W4	51.48944	-0.26194
3	Bloomsbury	020	Camden	LONDON	WC1	51.52450	-0.12273
4	Charing Cross	020	Westminster	LONDON	WC2	51.51651	-0.11968

I will use dataset https://geo.nyu.edu/catalog/nyu_2451_34572 for information about boroughs of NYC

NYC dataset:

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

For the second week, the final deliverables of the project will be: Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.

This submission will eventually become your Data section in your final report. So I recommend that you push the report (having your Data section) to your Github repository and submit a link to it.

A link to your Notebook on your Github repository, showing your code. (15 marks)

A full report consisting of all of the following components (15 marks):

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

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

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.

Results section where you discuss the results.

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

Conclusion section where you conclude the report.

```
In [ ]: Conclusion
        In this project all the techniques learned in previous courses were applied. The main g
```

3. Your choice of a presentation or blogpost. (10 marks)