

Final Report

How dissimilar is Murray hill and Hudson Yards?

Business problem

How dissimilar is Murray Hill and Hudson Yards?

This project has the main aim to find the dissimilarities from both neighborhood, the main idea regarding venues, like restaurants, cafe, actual trends. After this question is answered, the stakeholders will have a general idea about how both cities are different, for future decision making.

Methodology

The data were collected from two different websites, and all the results were analyzed by using jupyter notebook, since the aim of this project is to find dissimilarities of two cities we use the machine learning algorithm clustering, as well as some libraries from the python programming language were used.

Data

The data that will be used in this project, obtain the API foursquare in order to see restaurants, coffees and which neighbourhood has more traffic, and information about the neighbourhood our target neighbourhood are Murray Hills and Hudson Yards both are in New York city, it can be found in the following web address https://geo.nyu.edu/catalog/nyu_2451_34572 which is free..

Features :

In order to get and analyze the dissimilarities between both neighborhoods, it was used the API foursquare, this provided the information about different venues in both locations.
<https://foursquare.com/developers/apps>

About libraries:

In order to work with data obtained from different websites, the data were inserted in dataframes in the language of python, the library more used for this was the pandas, and for visualize the data were used the folium library.

Machine Learning Technique:

The technique used to solve this problem was clustering, which is an unsupervised machine technique, is useful when it comes to analyze data and group the data in sets that show the dissimilarities and similarities within the set. The name of this algorithm is kmeans

Results

The result were extracted from after analyze a certain amount of data, of differents neighborhoods in New York .It shows the id of the cluster and the city name,after run the algorithm several times.

Cluster id	City
1	Murray Hill
3	Hudson Yard

Discussion

Well used the algorithm, it was found that between both cities were different venues types and amount, in proportion the numbers were a little far from each other.

Conclusion:

Both cities differ because the location, amount avenues, Muray Hill and Hudson Yards, there were not similar at all. It the plan is to found similarities this study shows that both city are not the similar.

Biography:

Foursquare

<https://foursquare.com/developers/apps>

Geographical information about NY

https://geo.nyu.edu/catalog/nyu_2451_34572