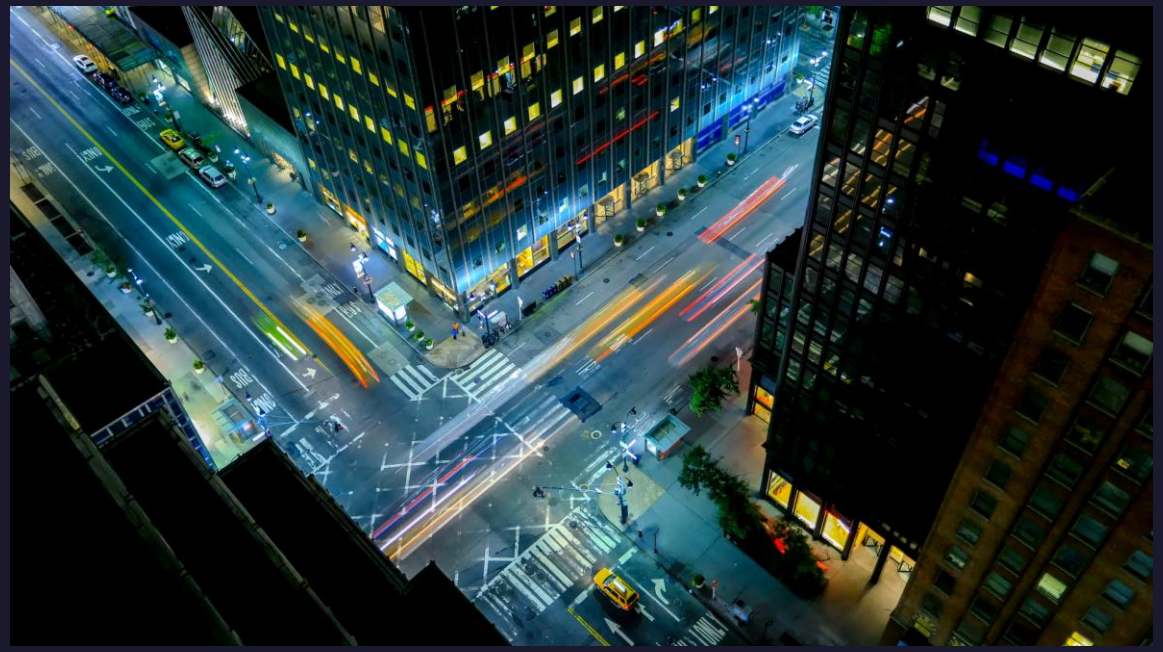
The Tale of Two Cities



Project by-

Shaswat Tripathi

Introduction-

- The Tale of two cities is a project designed to pick out the similarities in the neighborhoods of two cities.
- This program uses foursquare API to solve this problem.
- The same colored clusters are the neighborhoods which belong to the same categories.
- The project can help people looking for moving or relocating to a different place

Problem

- This program is designed to help people in the US or any other parts of the world to compare the similarities between two cities.
- Its target audience is the people who are trying to move from one city to another.
- The people who are afraid that the different city and the different neighborhood they may move in, might be totally different from their existing neighborhood.
- If people have been living somewhere for a while it becomes their comfort zone, and moving into an area which was like their previous one might make them more comfortable and ease the process of settling in.
- The program uses the resemblance of the venues located near a neighborhood in one city and compares it to another city and find the same distribution of the venues in the other city.

Data sources

- I have used the foursquare data to explore the input neighborhoods and classify them into similar clusters and the areas which are similar.
- The Same colored clusters are the neighborhoods which belong to the same categories.
- The Foursquare data also provided the latitude and the longitudes of the locations it returned.
- The Latitude and the longitude values are used to plot the locations on the graphs.

Feature and Data segmentation

- The data is Segmented into binary encoding, also known as one hot encoding to fit the data into the k means algorithm fast and will give much accurate results.

	American Restaurant	Bakery	Bank	Bar	Bridal Shop	Clothing Store	Coffee Shop	Cosmetics Shop	Deli / Bodega	Department Store	Diner	D
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	1	0	0	0
2	0	0	0	0	0	0	1	0	0	0	0	0
3	0	0	0	0	1	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0

- As you can see all the places are converted in form of a dataframe and then fit into the k means algorithm which will follow this step.
- The application finds the popular places located near a particular neighborhood. Distributes them into categories and helps find the similar categories in the place selected to move.

Clustering

- K means clustering is used for this feature to find the similar cluster areas.
- The k-means algorithm takes the similar types of places in a particular area and clusters them into similar groups.
- The K means Algorithm uses k points around which the data is clustered.
- Then the points are moves to minimize the inter-data point distance and maximize the inter-point distance.
- The K means algorithm is also called a slow learner.

Limitations-

- The logarithm is based on the Foursquare API so it can only access the locations thoroughly explored by the foursquare.
- If the place you entered returns less than at least less than 50 values by the Foursquare then the algorithm might not be able to categories places correctly.
- The algorithm might be more accurate if the categories of the foursquare api return more concise and not verbose categories
- The project is only available in the Jupyter notebook and you can run in your system.

Conclusions-

- The algorithm can predict any two matching neighborhoods and can help people find their new home fast.