

Battle of Neighborhoods

Dallas, Texas



Project Description:

Whenever people move to any other place, they explore the place and try to fetch as much information as possible about it. It can be the neighborhood, locality, market, price of the place and many more factors including neighborhood analysis. This is can be termed as request for a search algorithm which usually returns the requested features such as population rate, school ratings, crime rates, weather conditions, recreational facilities etc.

It would be beneficial and nice to have an application which could make easy by considering a comparative analysis between the neighborhood with provided factors.

This project helps the end user or the stakeholder to achieve the results which will not only recommend but also saves a lot of time in manual search. This will indeed save the time and money of the user.

This project can be used by the user at the time of rental apartment or buy house in a locality based on the distribution of various facilities available around the neighborhood. As an example, this project would compare 2 randomly picked neighborhoods and analyses the top 10 most common venues in each of those two neighborhoods based on the number of visits by people in each of those places. Also, this project uses K-mean clustering unsupervised machine learning algorithm to cluster the venues based on the place category such as restaurants, park, coffee shop, gym, clubs etc. This would give a better understanding of the similarities and dissimilarities between the two chosen neighborhoods to retrieve more insights and to conclude with ease which neighborhood wins over other.

Data Sets and APIs:

- Foursquare API: This API has a database of more than 105 million places. This project would use Four-square API as its prime data gathering source. Many organizations are using to geo-tag their photos with detailed info about a destination, while also serving up contextually relevant locations for those who are searching for a place to eat, drink or explore. This API provides the ability to perform

location search, location sharing and details about a business. Foursquare users can also use photos, tips and reviews in many productive ways to add value to the results

In [80]:

Out[80]:

	PostalCode	Neighborhood	Latitude	Longitude	Cluster Labels
0	75001	Addison	32.960047	-96.838522	2
1	75006	Carrollton	32.962141	-96.898585	2
2	75019	Coppell	32.963417	-96.984664	2
3	75040	Garland	32.928902	-96.619782	2
4	75048	Sachse	32.968973	-96.580671	2

• Work Flow:

HTTP requests would be made to this Foursquare API server using zip codes of the Chicago city neighborhoods to pull the location information (Latitude and Longitude). Foursquare API search feature would be enabled to collect the nearby places of the neighborhoods. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 900.

In [81]:

Out[81]:

	name	categories	lat	lng
0	The Joule	Hotel	32.780558	-96.798247
1	Pioneer Plaza	Plaza	32.776616	-96.801539
2	Weekend	Coffee Shop	32.780309	-96.798191
3	AT&T Store	Mobile Phone Shop	32.779447	-96.799028
4	Spice in the City	Indian Restaurant	32.780014	-96.797829

- Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of Dallas city over an interactive leaflet map. Extensive comparative analysis of two randomly picked neighborhoods would be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

- Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighborhoods. These clusters from each of those two chosen neighborhoods would be analyzed individually collectively and comparatively to derive the conclusions.

Python packages and Dependencies:

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Matplotlib – Python Plotting Module
- Sklearn – Python machine learning Library
- Folium – Map rendering Library

Methodology

Folium makes it easy to visualize data that's been manipulated in Python on an interactive leaflet map. It enables both the binding of data to a map for choropleth visualizations as well as passing rich vector/raster/HTML visualizations as markers on the map

. The workflow of the project starts with web scraping and data wrangling. Using the HTML library , the postal code and the neighborhood is processed to derive the latitude and longitude of the Dallas neighborhood. With the Folium Map the Latitude and Longitude of the Dallas neighborhood provides the Choropleth visualization

. FourSquare API and K means Clustering methods are used to retrieve the top trend venues of the Dallas Neighborhood

Elbow Criterion method is used here the optimum number of cluster present in the dataset. Silhouette Coefficient Analysis is also used to find the number of Cluster

In this project , neighborhood and the populaton of different ethnicity has been considered to give an analysis of the two random neighborhoods of Dallas, Texas

Result:

This Analysis concludes that the two places of Carallton,Garland both has great amenities and locality, but out of these two Garland has the higher number of Mexican population around 80k ,also top 10 top venues shows Garland has the most common restaurants are Mexican restaurants.Carallton has got a good neighborhood with Korean restaurant, American Restaurant, Coffee shop, Pizza place and many more. Hence Both oCarallton and Garland has thier own advantages !!

This study helps user to compare two neighborhood and recommend options with facts.

In []: