

Battle of Neighborhoods: Chicago, Illinois

Context and Background:

When people move to a new place, they explore and try to fetch as much information as possible. It can be the neighborhood, locality, businesses, price of apartments or single family homes and many more factors including neighborhood analysis like school ratings, crime rates, weather conditions, recreational facilities etc. It would be beneficial and nice to have a report or analysis which could make a family's move as easy as it could get.

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 analyzes 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:

Chicago neighborhood data can be web scraped from wiki pages as well as already available files over the internet. Here I have already scraped and put the details in to a csv file.

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 or drink. 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.

Workflow:

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 700.

Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of Seattle 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 like:

- Pandas
- NumPy
- 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:

The workflow of the project starts with the web scraping and data wrangling. Using the Beautiful Soup library, the postal code and the neighborhood is processed to derive the latitude and longitude of the Chicago neighborhood. With the folium Map, the latitude and longitude of the Chicago neighborhood provides the choropleth visualization.

FourSquare API and K-means clustering methods are used to retrieve the top trend venues of the Chicago neighborhood.

Elbow criterion method is used here to the optimum number of cluster present in the dataset.

Silhouette Coefficient analysis is also used to find the number of clusters.

In this project, decision of buying, rental or setting up business is recommended based on the clustered neighborhoods, Population Distribution analysis, School Ratings, Median House Price Analysis

Result:

With the help of above methodologies, the project can easily help the user to decide which neighborhood is better to stay based on the factors like population, localities, schools rating of the particular neighborhood and availability of top trend venues.

Discussion:

This project is beneficial in many terms, it will save the users time and money. Many a times when people move to the new place, they worry a lot before settling down. This analysis recommends the better places in a very less time.

Conclusion:

This Analysis concludes that the two places of Chicago Armour Square , Parkview both has great amenities and locality, but out of these two Parview has better prospects for buying houses or choose for rental houses. Parkview has the higher number of Indian population ,good school rating of 7 and a reasonable avg housing price of around 330k ,also top 10 common venues shows Parkview has got a good neighborhood with Gas station, Italian and American Restaurants, Train Station, Clothing Store , Parks, Donut Shops and many more. Hence Parview wins over Armour Square!

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