

Settling in New Orleans

1.Introduction

1.1 Background

New Orleans is a trending location in the US with a lot of famous venues, like ferry rides, cafes, seafood restaurants, and more. This attracts a lot of people into the area, for tourism and settlement. There are quite a few neighborhoods in New Orleans each of which have its own specialty, and venues.

1.2 Problem

For people shifting to completely new State/Country, it is often a problem finding a good place to settle down at. Making a wrong decision could lead to a lot of problems. At this time, it is useful, if the person had some knowledge about the nearby neighborhoods, venues, trendy places etc.. which suit their taste, thus helping them in their decision of finding the right neighborhood to settle down in.

In this project, we look into the neighborhoods of one of the most trending places in the US, New Orleans. For a person who's completely new to the place, it would be difficult to find out information about nearby facilities and venues, which would suit him the best, and this project aims at providing, not just for New Orleans , but for any place in the world, a descriptive outlook of each Neighborhood, which will help the person in settling down.

1.3 Interest

People wanting to move to a local neighborhood in a new State/ Country would really find it useful to have this data in hand for selecting the right neighborhood to settle in.

2. Data acquisition and cleaning

2.1 Data sources

For this project, I have used The neighborhood data of New Orleans from the following wikipedia page:

https://en.wikipedia.org/wiki/Neighborhoods_in_New_Orleans

This page consists of a table with the neighborhoods of New Orleans and their respective coordinates, and further in the notebook, we access the Nearby venues of a given coordinate using Foursquare API.

2.2 Data Cleaning

Since the data was html <td> tags I had to use a loop to access each of them and append into a list, there I converted them into strings and got rid of the <td> and </td> tags and trimmed off any excess space that might have crept in. Then I converted the final data into a dictionary with 3 keys as Neighborhood, Latitude and longitude and converted it into a dataframe. For using the coordinates in a map, I converted them from string to Float.

2.3 Feature Selection

Since there were only 3 columns involved and all of them were integral, there was no feature selection done.

3. Exploratory Data Analysis

3.1 Calculation of target variable

Since our main goal is clustering of neighborhoods, we used an unsupervised learning method where the cluster labels are the target variables.

3.2 Relation between columns

Since we use simple coordinates, there is no relation between columns other than the fact that the latitude and longitude columns are the coordinates of the particular neighborhood

3.3 Acquiring Nearby Venues Data

Using foursquare API I accessed the nearby venues for each neighborhood and appended the data to the original table.

4. Model Selection

4.1 DBSCAN

DBSCAN was my first choice for clustering of Data since it takes into consideration outliers, and is based on density of points, and it did perform well, But the model only gave 2 clusters as output and hence wasn't a good choice of model, for giving suggestions to a person who'd want to select from a variety of groups.

4.2 Agglomerative Hierarchical Clustering

This was also a choice for clustering, and with the choice of selecting more clusters, it was a better choice than DBSCAN, but the results were poor as the cluster distribution was not uniform , with 95% of the data in one cluster and very little in others to even consider.

4.3 K-Means Clustering

This was the final, and the best selection for a clustering model. Although the default ' kmeans++ 'init parameter didn't perform as well and I hence had to use the 'random' parameter and it performed well, for 7 clusters, with a decently uniform distribution among them.

5. Result

The result was a total of 7 clusters which divided the neighborhoods based on their most common venues, with one cluster having more café's and restaurants and others having more bars and nightclubs and so on.. Finally this division of clusters made it very easy to differentiate the group of neighborhoods based on their venues and hence allowing the person to select a neighborhood based on his/her liking and preferences.

6.Conclusion

In this study I scraped neighborhood data from a Wikipedia page, and found information about nearby venues using FQ API , cleaned the data and clustered it based on common venues to give suggestions for the person in selecting a neighborhood based on his/her preferences.

7. Future directions

New Orleans was a test location to test my project, and the dataset wasn't big enough to get good results, but with a bigger dataset of more neighborhoods, a different location with more venues, and maybe a better clustering algorithm, better and more uniform results can be obtained.