The Battle of Neighborhoods | Business Proposal

Introduction:

As we all know that U.S.A is a large country with many states. And many people like to visit these state. But there are many cities to explore in each state. So knowing about the capital of each state will brief us about each state. This project will segment these capitals on the basis of similarity of the venue found around them which will help the explorer/visitor to know about how similar these capitals are with each other and also will be able to find out the most common venue in each cluster.

Problem Which Tried to Solve:

1) Determine the cluster label for each capital.
2) Visualize the cluster label in the map
3) Determine most common venue in each cluster.

Data:

For the data we use all the state capital of USA from Wikipedia.

Foursquare API:

This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. 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 500.

Clustering Approach:

To compare the similarities of two cities, we decided to explore venues, segment them, and group them into clusters to find similar capitals. To be able to do that, we need to cluster data this is a form of unsupervised machine learning: k-means clustering algorithm.

Libraries Which are Used to Develope the Project:

Pandas: For creating and maniulating dataframes.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.