

Capstone Project - Battle of the Neighborhoods (Week 1)

1. Introduction and Business Understanding

1.1 Problem Description

The business problem we are working to solve is: The United States 2021 Presidential Inauguration Day is rapidly approaching and many people may be (irresponsibly) traveling to Washington D.C. to observe. How can we support visitors of various backgrounds and tastes, by categorizing and visualizing the cast number of D.C. neighborhoods by primary restaurants?

1.2 Background Discussion

With a population of nearly 690,000 people in a land mass 5.6% the size of Rhode Island, Washington D.C. can expect some of the almost 330 million U.S. citizens, and international visitors, to converge on the capitol city on January 20th. The city is used to this kind of tourism and is well-known for its diverse culinary options; 2,233 to be exact!

A huge time-sink for many travelers is taking the time to find out what kind of food you want to have, where you can have it, and who does it the best for a reasonable price. With all the apps and opinions available to the average person nowadays, it quickly becomes an overwhelming amount of data to process. Plus, nobody wants to look like a tourist!

How can we use machine learning algorithms and the power of Foursquare's location data to make this process of finding a place to eat simpler? In this project, I will be using Foursquare and clustering algorithms to analyze and group the neighborhoods of Washington D.C. by their primary restaurant types in an attempt to answer this problem.

2. Data Requirements

To pursue this project, we will need the following open source data:

- A. Data that contains all the neighborhoods of Washington D.C. and their basic location information.
Source: This will be sourced from Wikipedia: https://commons.wikimedia.org/wiki/Category:Neighborhoods_in_Washington,_D.C.
- B. Utilization: I will use BeautifulSoup to scrape the neighborhoods of Washington D.C. from the Wikipedia webpage and obtain their prospective latitudes and longitudes using the geocoder from GeoPy. I will then process the data and toss out any invalid data entries.
- C. Data that contains all the restaurants located in each of the ~80 neighborhoods of Washington D.C.
Source: Foursquare API
Utilization: Using Foursquare, we will input the locational coordinates of each neighborhood to obtain all the venues for each area. Then, we will filter out the

restaurants for each area and drop the rest. Beyond this will be the performing of the machine learning analysis.