Coursera Capstone Week 4: Opening a Dunkin' Donuts Shop in Washington DC

By: George Liu

July 24, 2020

Introduction

Dunkin' Donuts, since its opening in 1950, has become a staple in American culture. As one of the most popular fast food chains in the country, it specializes in coffee and donuts. In Washington DC, a plethora of Dunkin' Donuts shops are scattered across the city, from end to end. It is no surprise that the demand for Dunkin' Donuts coffee and donuts is so prevalent in Washington DC; in turn, competition amongst shops is intense. Thus, for new business owners looking to take advantage of this American staple and open their own shop, location is crucial to the shop's success. There are countless other factors to opening a new business, but the widespread of Dunkin' Donuts in Washington DC makes location a vital one.

Business Problem

This exploration's aim is to closely analyze and pick out the most optimal areas in Washington DC to start a new Dunkin' Donuts shop, utilizing machine learning clustering algorithms as well as a data science methodology. Ultimately this will answer the business question: If a business owner in Washington DC, USA is looking to open a Dunkin' Donuts shop, where should they do so?

Data

To solve this business problem, we will several pieces of data:

- A list of neighborhoods in Washington DC, to define the scope if this exploration
- Geographical coordinates of those neighborhoods, to create a map and obtain venue data
- Venue data, to perform clustering

From this Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Washington,_D.C.) we can obtain the list of Washington DC neighborhoods by web scraping using the Python Requests and Beautiful Soup libraries. Next to obtain the longitude and latitude coordinates of each neighborhood, the Python Geocoder library will be used.

Finally, the Foursquare Locations API will be utilized to obtain venue data for those neighborhoods. The Foursquare Locations database has over 100 million locations and is used extensively, with over 100,000 developers. For this scenario, the focus will be on Dunkin' Donuts shop data from Foursquare. Once this data is obtained, wrangled and cleaned the K-means Machine Learning algorithm will be used to cluster each neighborhood. After the clustering, the map of Washington DC and its Dunkin' Donuts shops will be displayed along with its labelled clusters.