

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 use several pieces of data:

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

From 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 BeautifulSoup 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.

Methodology

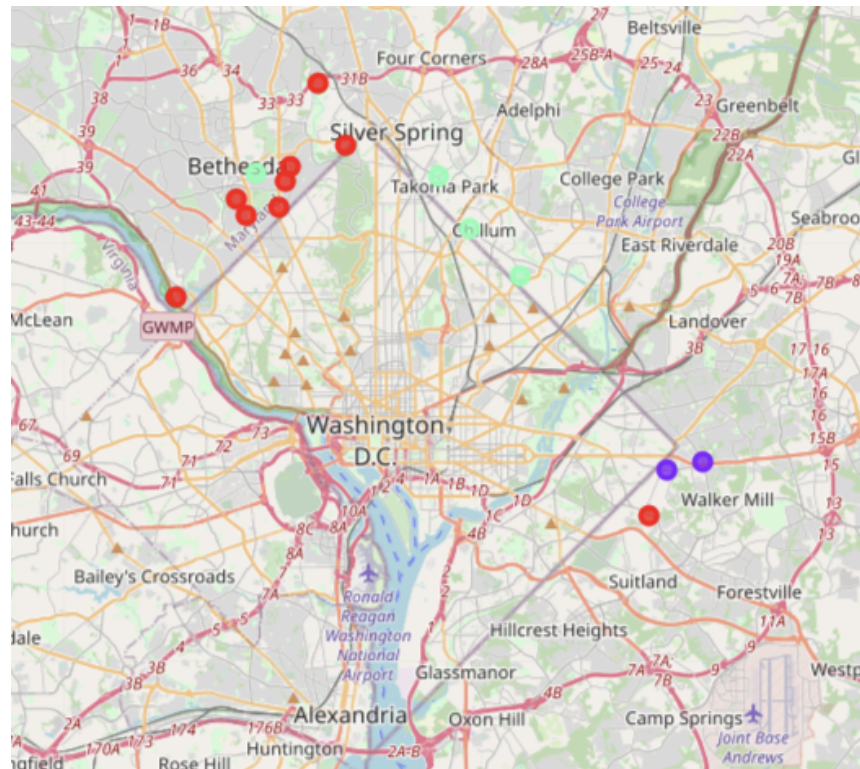
To begin, the list of neighborhoods in Washington DC must be obtained; it can be found at the link in the above section. Using the Python Requests and BeautifulSoup libraries we can scrape the list of neighborhoods. To be able to utilize the Foursquare Locations API, the neighborhoods must be along its geographical coordinates. Using the python Geocoder package, the corresponding coordinates can be obtained and merged into a data frame with the existing list. Utilizing the Folium package, a map of Washington DC and its neighborhoods superimposed can be visualized, which will ultimately aid in verifying the correctness of the processes.

Next using the Foursquare Locations API, the top 100 venues in a radius of 2km can be returned. Passing in my Foursquare Developer credentials, the venue data will be obtained as a JSON, from which the name, category and coordinates can be extracted. Next, we can analyze each neighborhood through grouping the rows for each neighborhood as well as taking the average of the frequency of occurrence for each venue category. After we can filter "Donut Shop" as the venue category for use.

Finally, clustering on our dataset can be done through the use of the k-means algorithm. For this exploration, 3 clusters will be formed based on their frequency of “Donut Shop”. These results will be useful in determining neighborhoods with varying concentrations of donut shops and will ultimately be useful for our initial question.

Results

The results of the k-means algorithm clustering are shown below.



- Cluster 0 (red): Neighborhoods with low concentrations of donut shops
- Cluster 1 (purple): Neighborhood with high concentrations of donut shops
- Cluster 3 (green): Neighborhoods with moderate concentrations of donut shops

Discussion

From the observations shown in the above section, interestingly enough, neighborhoods with the same cluster assignments are located quite close to each other. Perhaps, neighborhoods in specific areas overall do not prefer donuts as much as other neighborhoods. Another possible explanation could be a class or wealth difference among the different communities. Specifically,

near Silver Spring and Bethesda where a lot of neighborhoods were labeled with low concentrations of donut shops, there exists one neighborhood labeled in cluster 3. This suggests that even in low competition, that specific area only contains a moderate number of donut shops, likely meaning that area just does not prefer donuts as much. Taking a look at the bottom right portion of the map, near Walker Mill, there exists two neighborhoods with high concentrations of donut shops and one with a low concentration. This likely suggests that most donut customers in the sole Cluster 0 neighborhood prefer to go to nearby neighborhoods where there is a high concentration. Thus, to strike the perfect balance of avoiding strong competition and donut popularity, business owners are strongly suggested to look into the area near Takoma Park (Cluster 1) where there is moderate competition.

Limitations and Future Suggestions

In this exploration, the sole factor considered was frequency, so conclusions were not as straightforward. In a location, there exists many other factors like population and salary of residents. Therefore, in a future exploration, a larger data set with salaries and population can be merged all along with the original one. Unfortunately, not much of this data is available for free. Another limitation was the tier of the Foursquare API that was used in this exploration. The Sandbox Tier only allowed for 500 calls, whereas a premium one would come with many more.

Conclusion

For new business owners looking to open a donut shop such as Dunkin' Donuts in Washington DC, the best neighborhoods to do so are those labeled cluster 1 from our results section.

References

Category: Suburbs in Washington DC. *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Category:Suburbs_of_Washington,_D.C.