# **Finding the Best Neighborhoods to Build Coffee Shops in Washington, DC**

## **Introduction**

Data Analysis can help determine the best location to build a coffee shop in Washington, DC. Great locations for coffee shops are locations where there is a shortage of coffee shops compared to the number of coffee drinkers in the area, and where the location can be accessible to the most people.

The Foursquare API is a great data source for information needed to find locations that best match the above criteria. The Foursquare API can find coffee shop locations in DC and provide the number of check-ins at the coffee shop. This data will help us learn which areas have a high ratio of coffee shop customers to the number of coffee shops.

Additionally, we can combine the Foursquare data with neighborhood data from Washington, DC. DC's data includes clusters of neighborhoods, which can serve a reference points. We can also incorporate Metro stop data from DC. The Metro is DC's public railway system, and it serves a significant number of customers. We can use this data to determine the degree to which the number of Metro stops in a given area correlates with the number of coffee shops in the same area. This will be a helpful data point relevant to our analysis because Metro stops increase the accessibility of a coffee shop.

## **Data and Methodology**

To accomplish the above analysis, we'll take the following methodology:

1. Acquire and clean DC neighborhood geo data
2. Acquire and clean DC Metro stops geo data
3. Generate the number of Metro stops per area unit of each neighborhood
4. Query the Foursquare API for coffee shops in each neighborhood
5. Locate the neighborhoods that were in the vicinity of the coffee shops
6. Determine which neighborhoods had the highest number of coffee shops per area unit
7. Assess the correlation between Metro stops and coffee shops in each neighborhood

Due to some constraints using a free Foursquare API developer account, we’ll make some modifications to our approach and methodology along the way, while noting how our analysis could be improved with better access to the Foursquare API data. The following still demonstrates, however, the validity and usefulness of our methodology and approach and how it can be tailored to recommend the best locations for building a coffee shop in DC.

## **Results**

A picture containing arrow

Description automatically generated

**Figure 1**

After compiling and cleaning the data and querying the Foursquare API, we are able to calculate the ratio of coffee shops to area unit in the select DC neighborhoods, as seen in Figure 1 above.

Map

Description automatically generated

**Figure 2**

The five neighborhoods with the lowest number of coffee shops per area unit are marked in Figure 2.

Graphical user interface

Description automatically generated

**Figure 3**

Last, we measured the Pearson correlation between the number of coffee shops per area unit and the number of Metro stops per area unit. There is a slight inverse correlation between the two variables, meaning there are slightly fewer coffee shops in areas that have greater Metro access.

There is an opportunity here: locations closer to Metro stops increase the number of potential customers both by virtue of their proximity to a stop and because a coffee shop closer to a Metro stop would have an advantage over other coffee shops, which are typically farther away from Metro stops.

## **Conclusion**

Above demonstrates my method and approach for determining the best location for building a coffee shop in DC. Because of the limitations of the free developer account, I cannot offer a recommendation here. Despite lacking full access to the developer API, the method and approach are sound and can offer insights into determining the proper location. To recap the method and approach I proposed and demonstrated here, I:

1. Acquired DC neighborhood geo data
2. Acquired DC metro stops geo data
3. Generated the number of metro stops per area unit of each neighborhood
4. Queried the Foursquare API for coffee shops in each neighborhood
5. Located which neighborhoods the coffee shops were in the vicinity of
6. Determined which neighborhoods had the highest number of coffee shops per area unit
7. Assessed the correlation between metro stops and coffee shops

If time permitted, I would enhance my analysis in order to provide a stronger recommendation by:

1. Calculating the number of checkins at all coffee shops in each neighborhood
2. Calculating the number of checkins divided by the number of coffee shops divided by area unit
3. The neighborhoods with the highest values in Step 9, would be the best locations
4. We can provide a final assessment of those locations by analyzing the locations based on whether there is a strong correlation between the number of Metro stops in an area and factoring in the strengths and weaknesses of our data