THE BATTLE OF NEIGHBORHOODS

The Case of Washington DC

DECEMBER 21, 2019

Contents

Introduction: Background

Washington D.C is the political capital city of the United States. It is one of the most popular tourist destinations in the world and annually receives more than 20 million visitors. As of July, 2018 the U.S. Census Bureau estimated more than 702,444 people live in the city.

Washington D.C has a diversified growing economy. Restaurant industry plays important role in the state's economy. According to [National Restaurant Association Report, in 2018 the industry's estimated total sale was $4.4 billion with 2,447 eating and drinking locations in the District.

Due the facts Washington DC population is rising, number of workforces entering the city by day is expanding and among the visitors of the city most of them are internals might be the recipes for opening new American restaurant. Due to these and other facts stakeholders might be interested in opening American Restaurant in Washington DC.

Business Plan

In this study three major tasks were conducted:

1. Analysis of food venues distribution in Washington D.C

2. Analysis of restaurants distribution in Washington D.C and

3. Finding optimal locations for opening a new American restaurant.

In the first and second studies food venues and restaurants were analyzed lightly. While in the third study optimum locations for opening new American restaurants were identified. The criterion implemented to get optimum locations were:

   i.   Not populated with many restaurants

  ii.   Close to the center of Washington D.C and

 iii.   No American restaurant exists in a certain distance range

Some of the algorithms and methods used in this work were adopted from *'Capstone Project - The Battle of the Neighborhoods: Applied Data Science Capstone by IBM/Coursera'*. In this study stakeholders who have intention to open restaurant specifically American restaurant in Washington D.C were targeted.

Data

Data used for this project were obtained from different sources. Mainly we used for sources:

1. US Government's open data Geospatial: Data of the United States containing coordinates and shapes of states was obtained from an open data source portal of the US government. The complete geometry of Washington DC is required in order to generate neighborhood centers throughout Washington DC. We extracted the coordinates of the entire Washington DC geometry from the US shapely object file.

2. Open Data DC: Data of 39 Neighborhood Clusters currently used by the District of Columbia for different purposes were retrieved. The geometry of the clusters was used to visualize food venues/restaurants distribution in Washington DC using folium map.

3. Candidate neighborhoods data for Washington D.C was produced algorithmically.

4. Foursquare Food venues including type and geographical locations were retrieved by calling Foursquare API.

5. Mapbox: Approximate addresses of each neighborhoods were obtained by calling Mapbox reverse geocoding API.

Neighborhood Candidates

Washington D.C. has 129 neighborhoods and 39 Clusters whose geographic coordinates can be easily retrieved from the city's government Open Data DC dataset portal. However, these neighborhoods and clusters don't lend themselves for the analysis we intended to do. To serve our purpose neighborhoods having equal size will be generated algorithmically throughout Washington D.C. Each neighborhood generated will be circular shape and equally sized.

Four Square

Now we have all the information needed about the neighborhoods. The next step will be obtaining food venues located in each neighborhood. However, to analyze the results returned by Foursquare API it is important to understand how food venues are categorized hierarchically in foursquare database.

In general venues in four square datasets are categorized hierarchically. To mention some of the Venue Categories on the highest level of the hierarchy: "Arts & Entertainment", "College & University", "Event", and "Food". However, in this work our focus will be only on Food Venues

which has category ID of '4d4b7105d754a06374d81259'. In the hierarchy of food venue there are 91 categories (Asian, African, American ...). Among the 91 categories 21 of them have specific categories and placed on the next level of the hierarchy. For example, Asia restaurant was classified on the top of the food venue hierarchy with the next hierarchy contains China, Burmese, Cambodian and other 13 restaurants. Furthermore, on the hierarchy of China restaurant there are 31 categories such as Anhui, Beijing and 29 more restaurants each having unique category ID.

When we call Foursquare food venues API, each venues result fall exclusively in either of the hierarchical categories explained above (for instance either Asian, or Chinese, or Beijing restaurant category not both).

Methodology

The methodology used here is more or less adopted from "Capstone Project - The Battle of the Neighborhoods: Applied Data Science Capstone by IBM/Coursera". In this project our main focus will be identifying potential locations(neighborhoods) in Washington DC for opening American restaurant. The selected locations are mainly characterized by their proximity (2000-meter range of radius) to the center of Washington D.C. Moreover, there shouldn't be any American restaurant within 250 meters of radius. And finally, no more than two restaurants should exist within 400-meter radius range from their respective center.

In this work four major tasks were conducted. The first task was collecting food venues data throughout Washington DC. During this stage food venues data was generated by calling foursquare API and saved into dataframe.

While in the second stage the dataframe obtained from the first step was manipulated and restaurant dataframe was created. Both food venue and restaurant dataframes were grouped categorically and the results were visualized using bar plot. The most common restaurant venue (American restaurant) will be considered for the next stage analysis.

During the third step the restaurant selected out in the previous stage (i.e. American restaurant) will be analyzed. In this analysis area distribution of this restaurant will be visualized using heat map. Again, density of restaurants over Washington D.C will be visualized using heat map. Furthermore, neighborhoods throughout Washington DC will be analyzed in terms of how far they are from the closest American restaurant and number of restaurants available in each of them.

In the fourth and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without American restaurants in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Results and Discussions

The first work of this project was analyzing the distribution of food venues category in Washington D.C and plot the result using bar chart. While in the second work
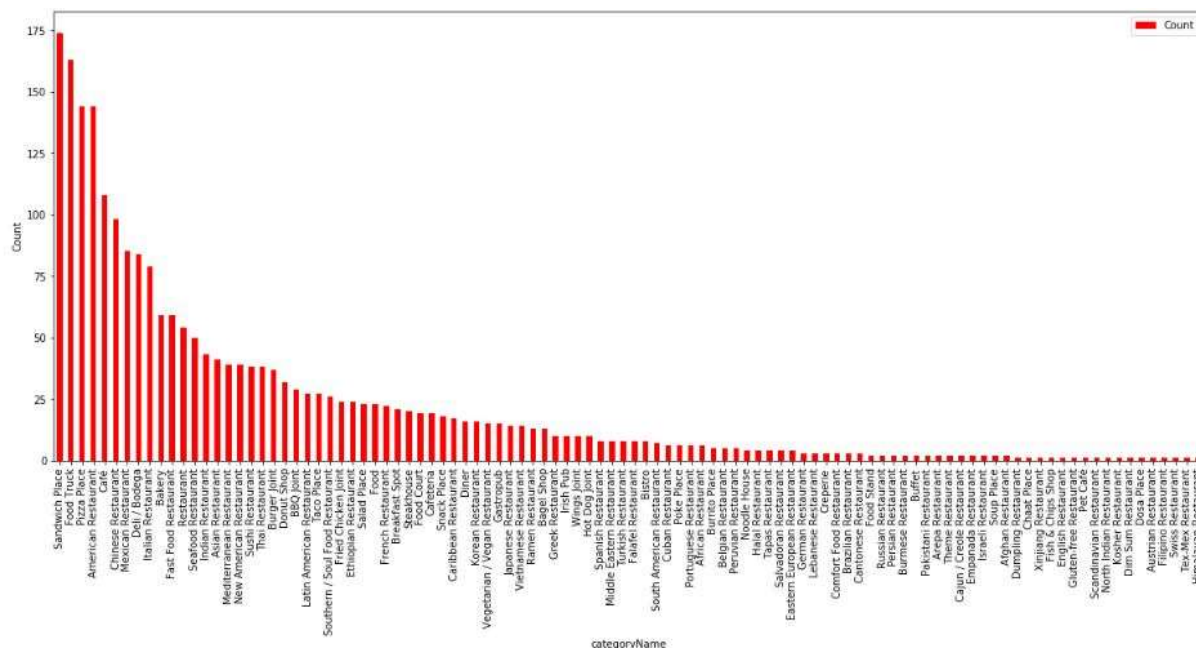

Figure 1 Bar plot of food venues category distribution in Washington D.c

As it can be seen from the bar plot Sandwich Place is the number one most populated food venue in Washington DC and followed by food truck, American restaurant and Pizza place. The diversity of Washington DC is quite reflected in its restaurants. The restaurants variety ranges from South America to Asia and from Africa to Europe spanning all over the entire world. Among the exotic restaurants Chinese restaurant is the most common and followed by Mexican, Italian and Indian

restaurants. By far American restaurants are the most populated restaurants in DC. Italian restaurants are the leading in number among the European restaurants. Apparently Ethiopian restaurant are the only establishment from African continent.

While in the second part of analysis restaurants dataset was extracted out from food venues dataset. The result set was grouped categorically and was plotted into bar chart. The result figure is shown below
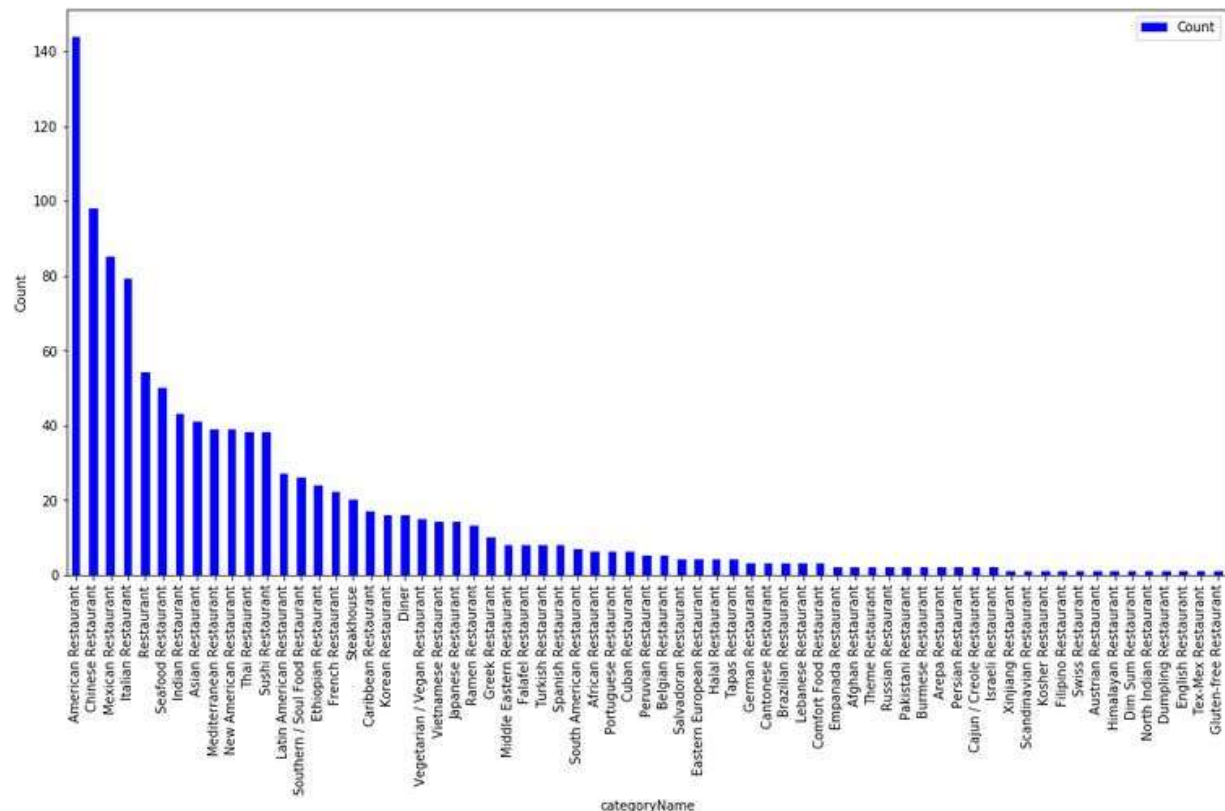

Figure 2 Restaurant venues category distribution in Washington D.C plotted using bar chart

Among the restaurants American restaurant is the number one most populated restaurant in Washington DC and followed by Chinese restaurant. Whereas Mexican and Italian restaurants trailed third and fourth. However, the number of restaurant venues of each category might go up as subcategories of some restaurants were considered independently. For instance, New American Restaurant is considered independently as shown above in the map, although it is under the category of American restaurant. Later when we do area/location analysis of American restaurants we will merge both together subtly.

In the third stage of analysis folium visualizing tool will be used. Restaurant number distribution was visualized using folium heat map plot. American restaurant density distribution was visualized using folium heat map too. Regions of low distribution of restaurants were identified visually. The resulting heat map plots are shown below in figure 3&4.
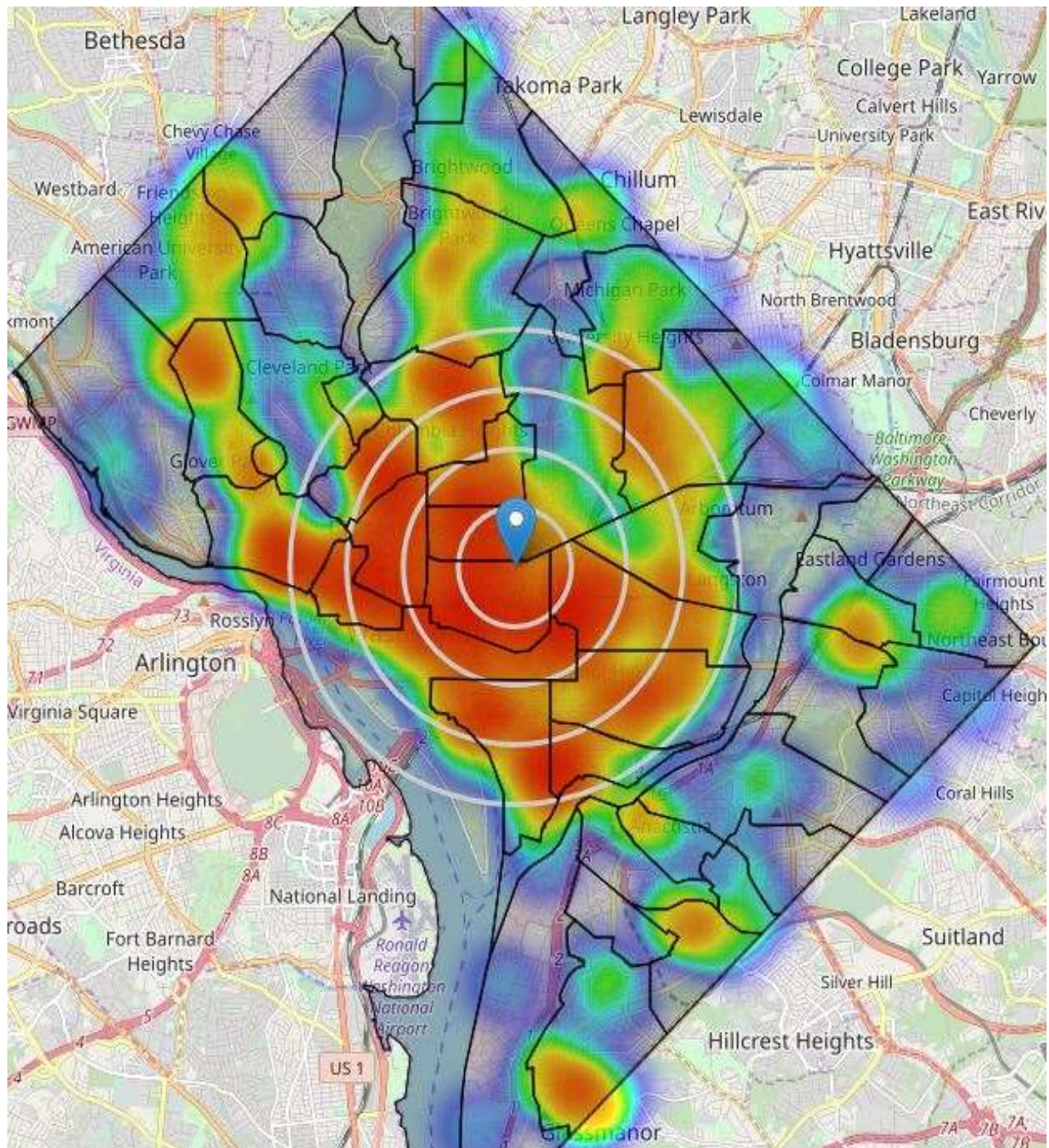

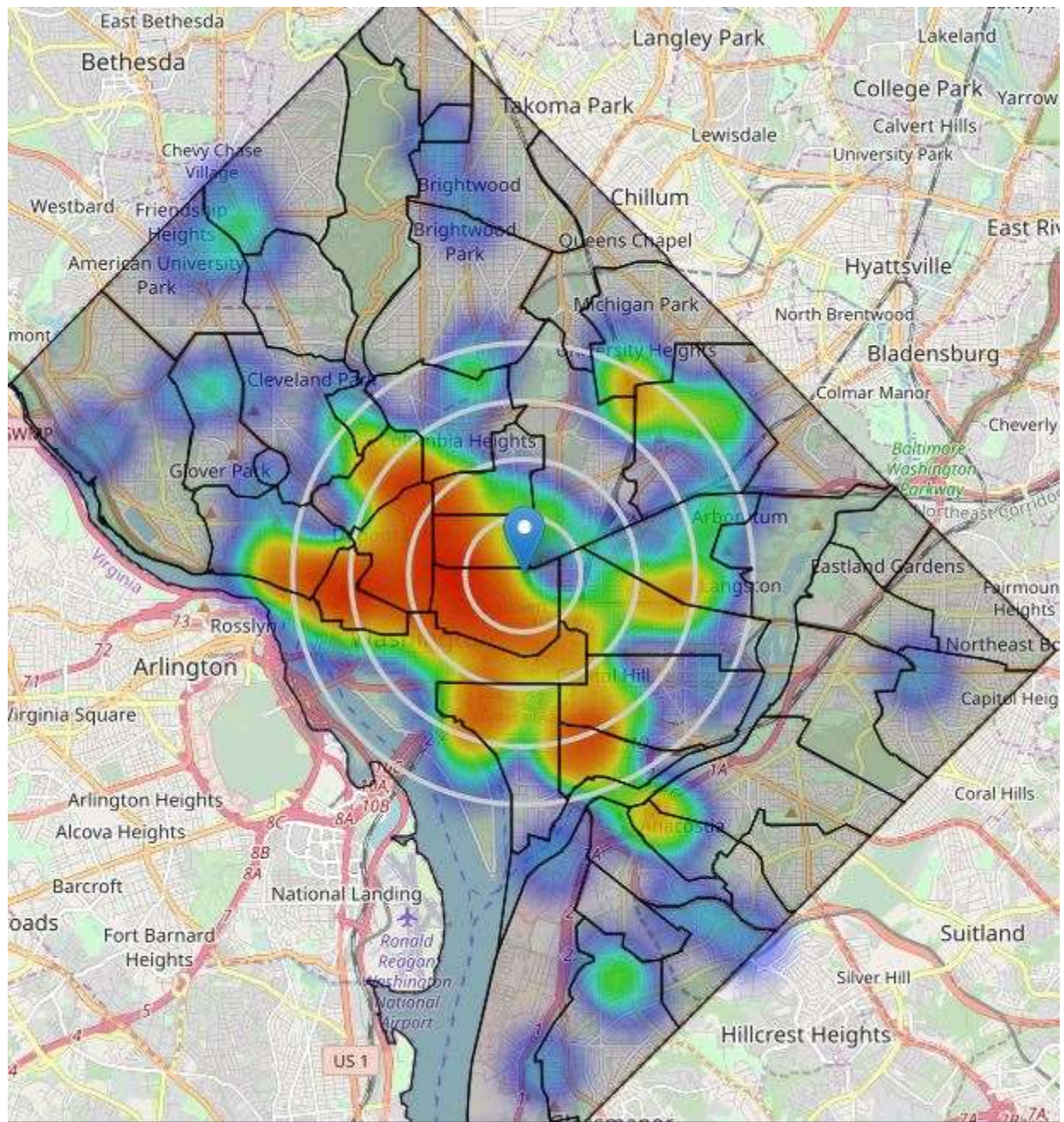*Figure 3 Heat map plot of restaurant distribution in Washington D.C*

Figure 4 Heat map of American restaurant distribution in Washington D.C

As shown above in the heat map (figure 3) most of the restaurants are distributed within 4000 meters of the centroid. Areas of highest density of restaurants are West, North West and South West of Washington DC. In fact, most of the restaurants are found in Cluster 1,2 3, 4, 5, 6, 7 and 8. The neighborhoods in each cluster can be seen in the above interactive folium map.

American restaurant represents less than 16.5 percent of the total restaurants. It is evident from the map above (figure 4) the density of American restaurants is low. From restaurant heat map low concentration of restaurants observed along east, north east and south east areas. Most of restaurant activities are limited within 4000 meters distance range from the center. Most of tourist attracting sites including White House, Museums, and U.S. Capitol are located close to the center of Washington D.C.

In the last stage of analysis, for detecting optimum locations for opening American restaurant we limited our area search in 2000 meters of radius from the city center. Neighborhood centers having 100 meters radius were generated in the selected region. They were tested for criterion of "not more than two restaurants in 250 meters" and "no American restaurant in 400 meters radius". The locations satisfying both conditions were filtered out and further clustered to create zones of interest. All the locations obtained were grouped together in 10 clusters. Approximate addresses for centers of each cluster were obtained using Mapbox reverse geocoding technique. The real neighborhoods and cluster of these locations were identified.
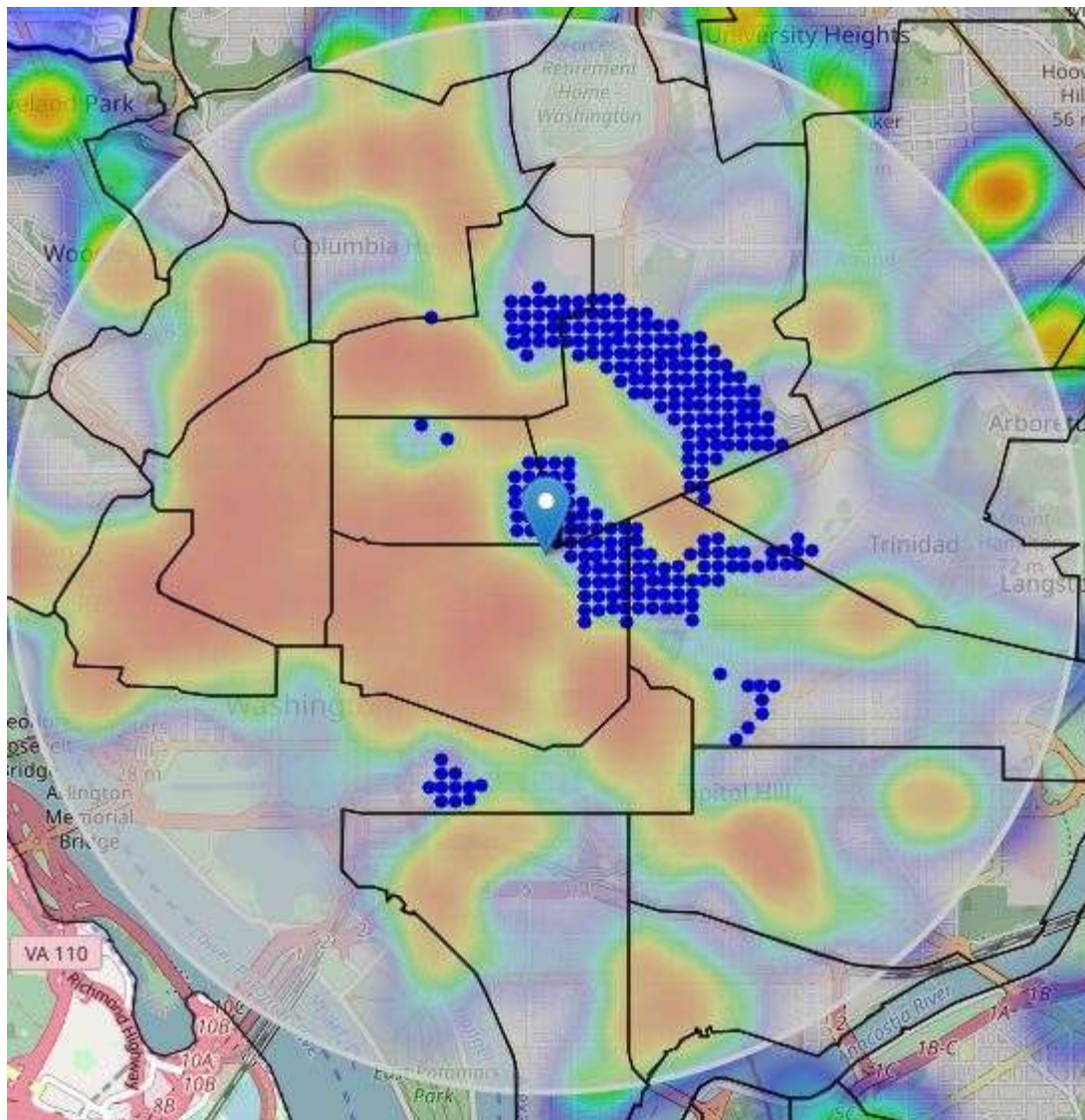
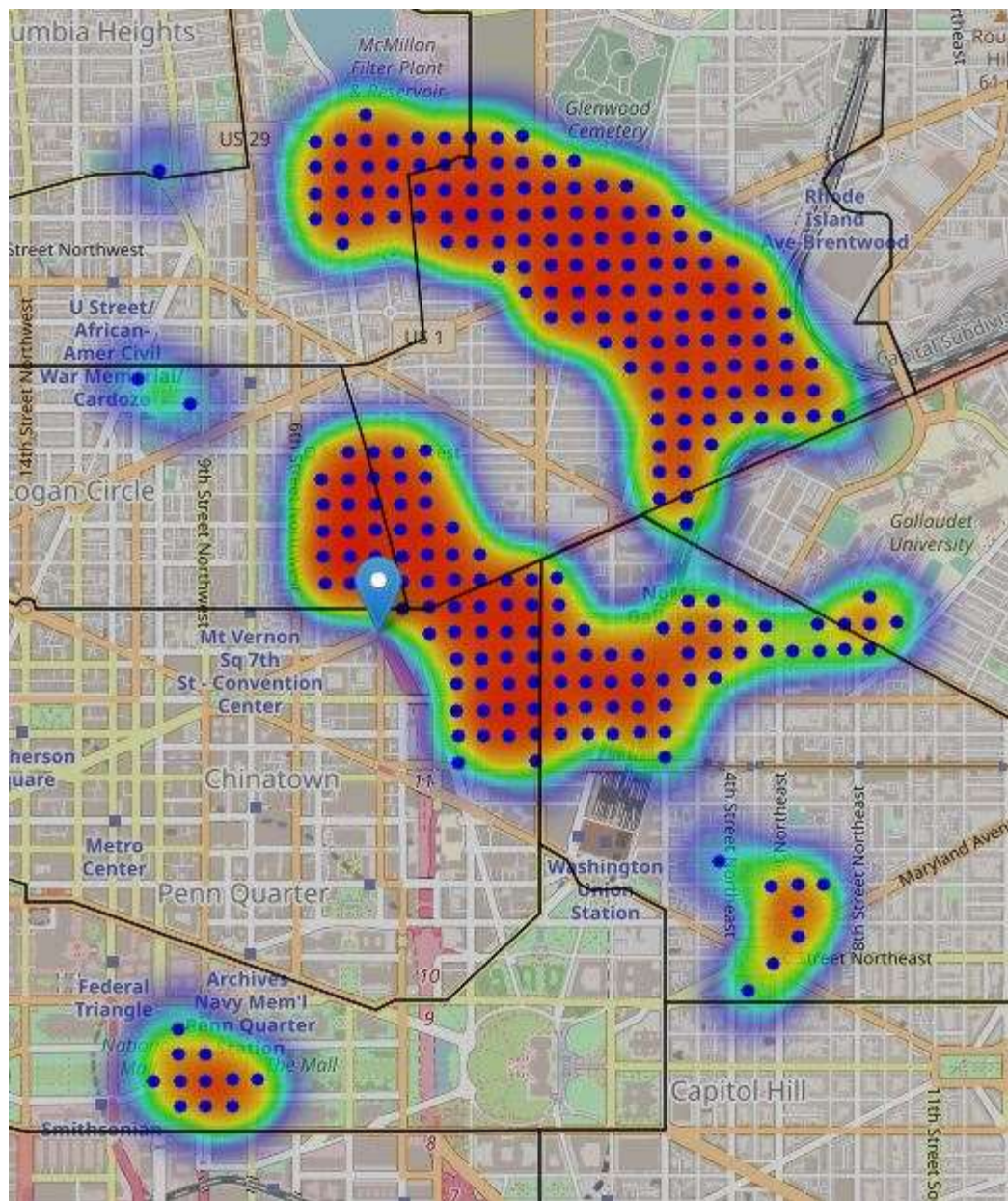*Figure 5 Location fulfilling test criterion plotted along heat map of restaurants distribution*

Figure 6 Heat map of potential locations for opening new American restaurant
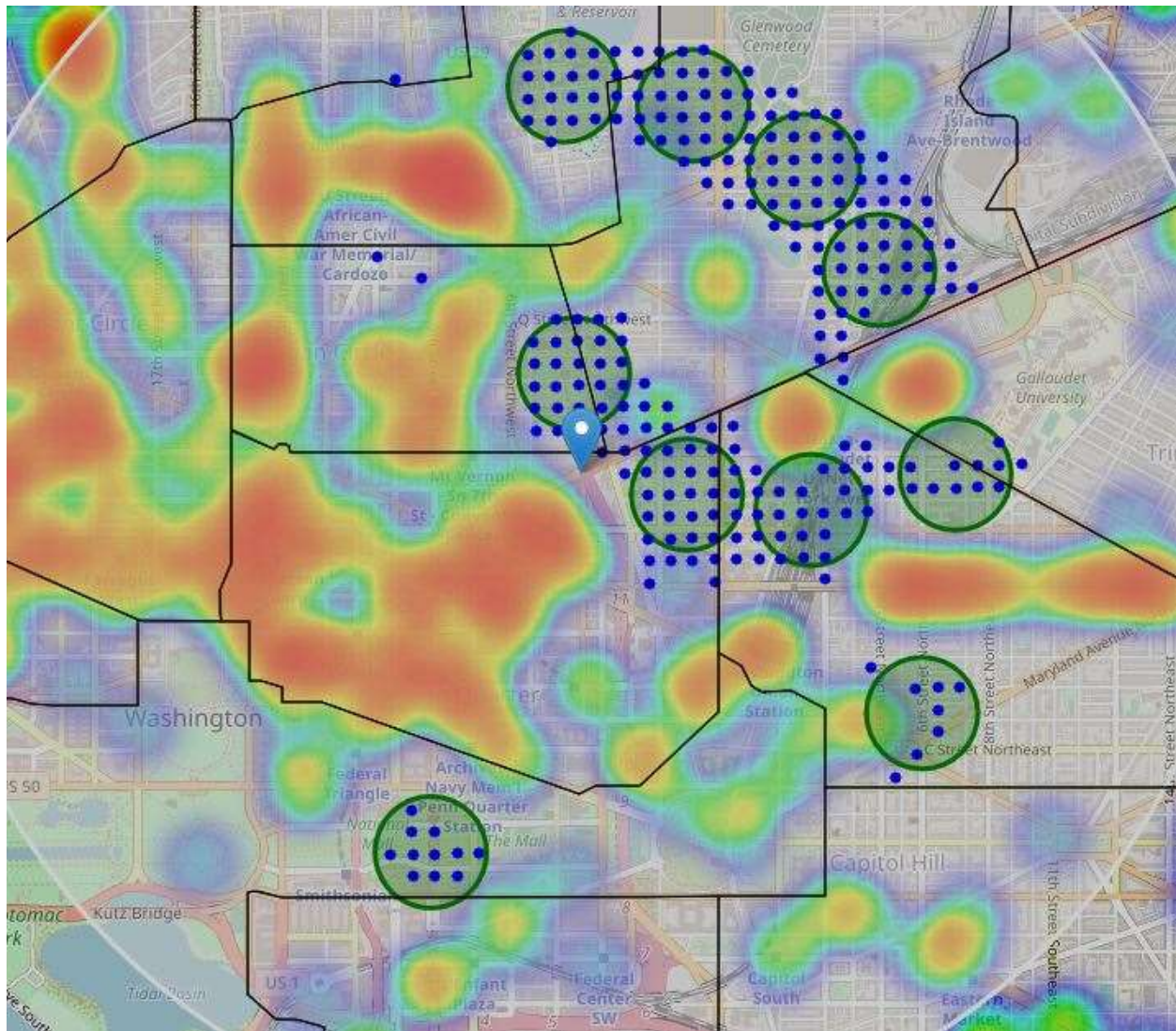
Figure 7 The result locations were grouped into 10 clusters and visualized using heat map

Conclusions

The main purpose of this study was to detect optimum locations for opening American restaurant in Washington DC. The 10 narrowed locations obtained are located in Clusters 3,7,8,21,23 and 25 whose natural neighborhoods can be easily identified from the interactive map generated above.

In connection with it some basic exploratory analysis were done on food venues located in Washington DC. The study reveals some restaurants have got only one or couple of venues. Where as many locations of the city food venues were hardly to find specially as we go further from center. Most of restaurants are located in a certain radius from the center where tourist attraction sites and governmental offices are located. Separately this study shows that sandwich places, food

trucks and pizza places are the three most popular food venues in Washington DC. Most of them as the rest of food venues are populated near the center of the city.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, parking garages availability, prices, crime rate, social and economic dynamics of every neighborhood etc.