Capstone – The Battle of Neighborhoods

Food Venues Analysis: The case of Washington DC

Introduction

Washington DC Background

- ✓ The capital city of the United States
- ✓ In 2018 US census bureau estimated more than 702,000 people live in the city
- ✓ Annually receives more than 20 million visitors. Most of them are internally

Restaurants Background

- ✓ In 2018 the restaurant industry's estimated total sale was $4.4 billion with 2,447 eating and drinking locations in the District

Prospects

- ✓ Visitors number is increasing each year
- ✓ Number of workforce entering the city by day is expanding
- ✓ City's expanding economy each year and high consumer's outlook push restaurant industry to expand

Business plan

In this study three major tasks were conducted

1. Exploratory statistical analysis of food venues in Washington DC
2. Exploratory statistical analysis of restaurant distribution in Washington DC
3. Finding optimal locations for opening American restaurant in the city

Criterion for choosing optimal location

I. The area is not populated with many restaurants

II. Close to the center of Washington D.C

III. No American restaurant exists in a certain distance range

Data sources

US Government's open data portal

- ✓ Geospatial data of US and its states was retrieved from it

Washington DC Government's open data portal

- ✓ The city's neighborhood's are arranged in Clusters. The clusters boundary and its geometry was retrieved from this portal

***Foursquare* location data and intelligence API**

- ✓ Food venues dataset in the city was obtained from it

***Mapbox* reverse geocoding API**

- ✓ Reverse geocoding (finding area addresses) was done using this API

Candidate neighborhoods generated algorithmically

- ✓ Grids were generated throughout Washington DC

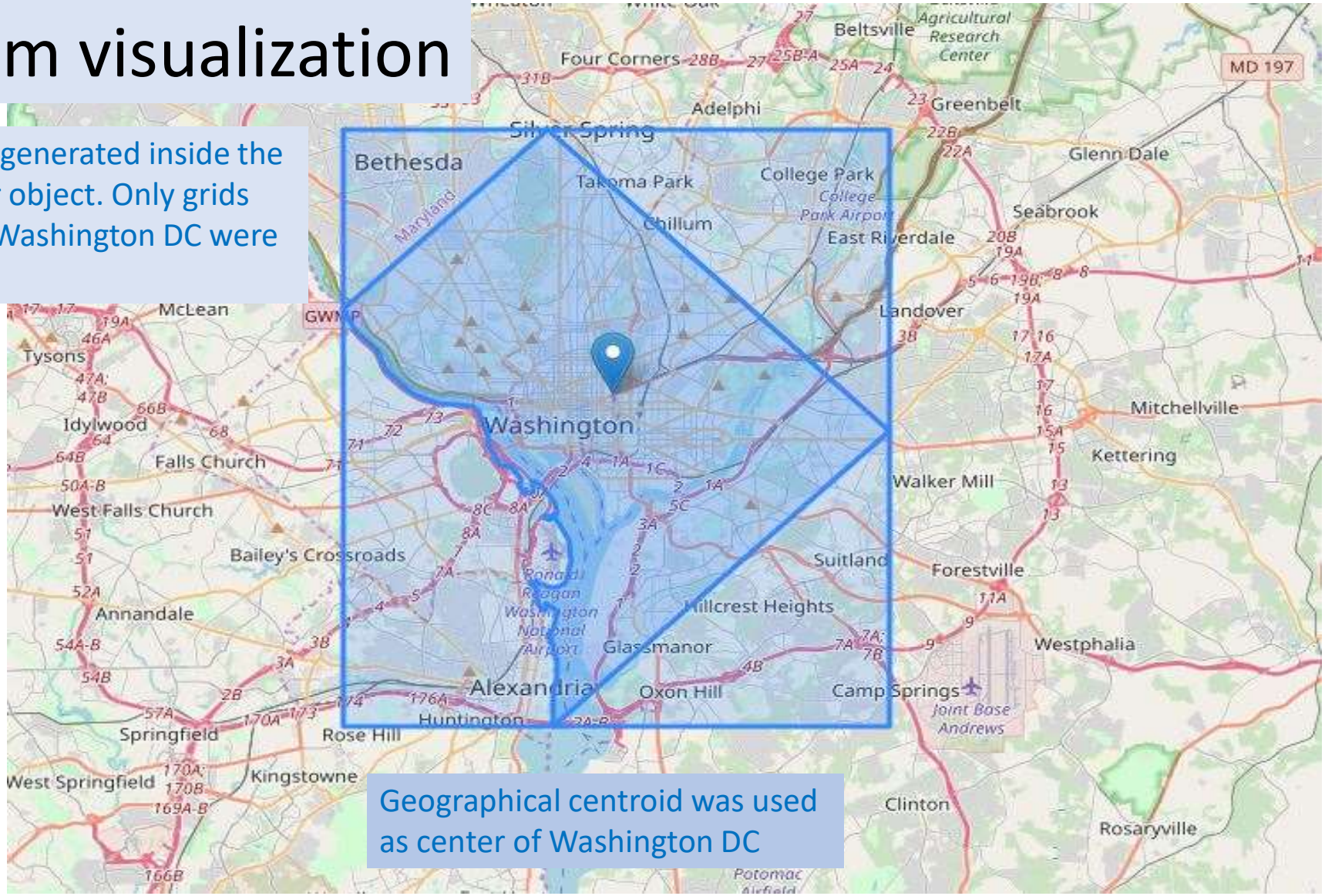
Data collection steps

1. ***Candidate neighborhoods generation***
   - ✓ Generated throughout the city
   - ✓ Circular shape of radius of 300 meters
   - ✓ Square grid objects were used to generate
   - ✓ The grids were first produced inside a rectangular object that completely contained Washington DC map
   - ✓ All the grids fall inside Washington DC were selected for further analysis

2. ***Mapbox reverse geocoding***
   - ✓ During this step addresses for each center of neighborhoods were obtained
   - ✓ Mapbox reverse geocoding was used to convert latitude/longitude coordinate systems to physical addresses.
   - ✓ In order to access datasets from Mapbox, latitude/longitude information and Mapbox key are required

Folium visualization

Grids were generated inside the rectangular object. Only grids lied inside Washington DC were selected

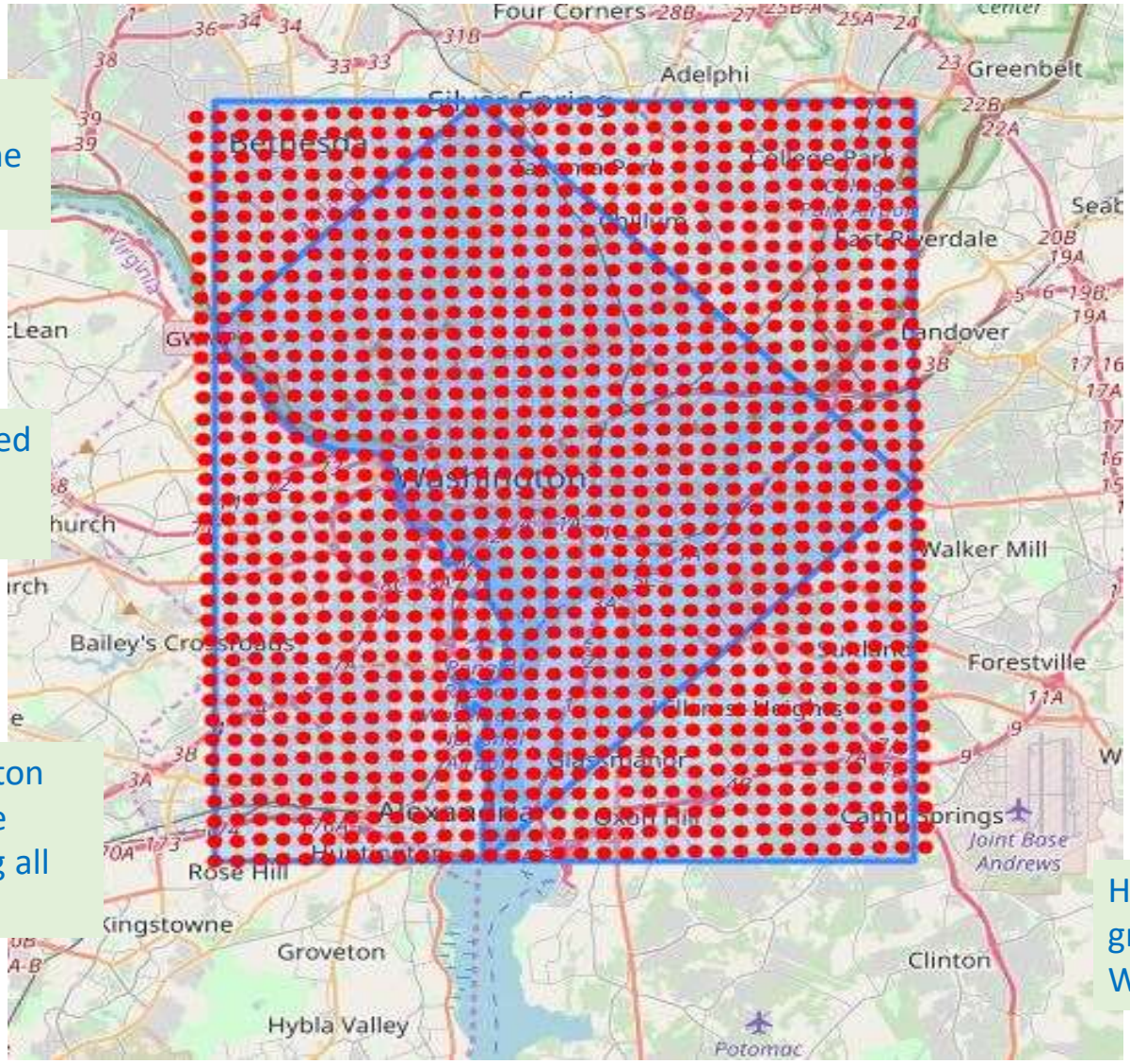


Geographical centroid was used as center of Washington DC

Red dot are grids generated throughout the rectangular object

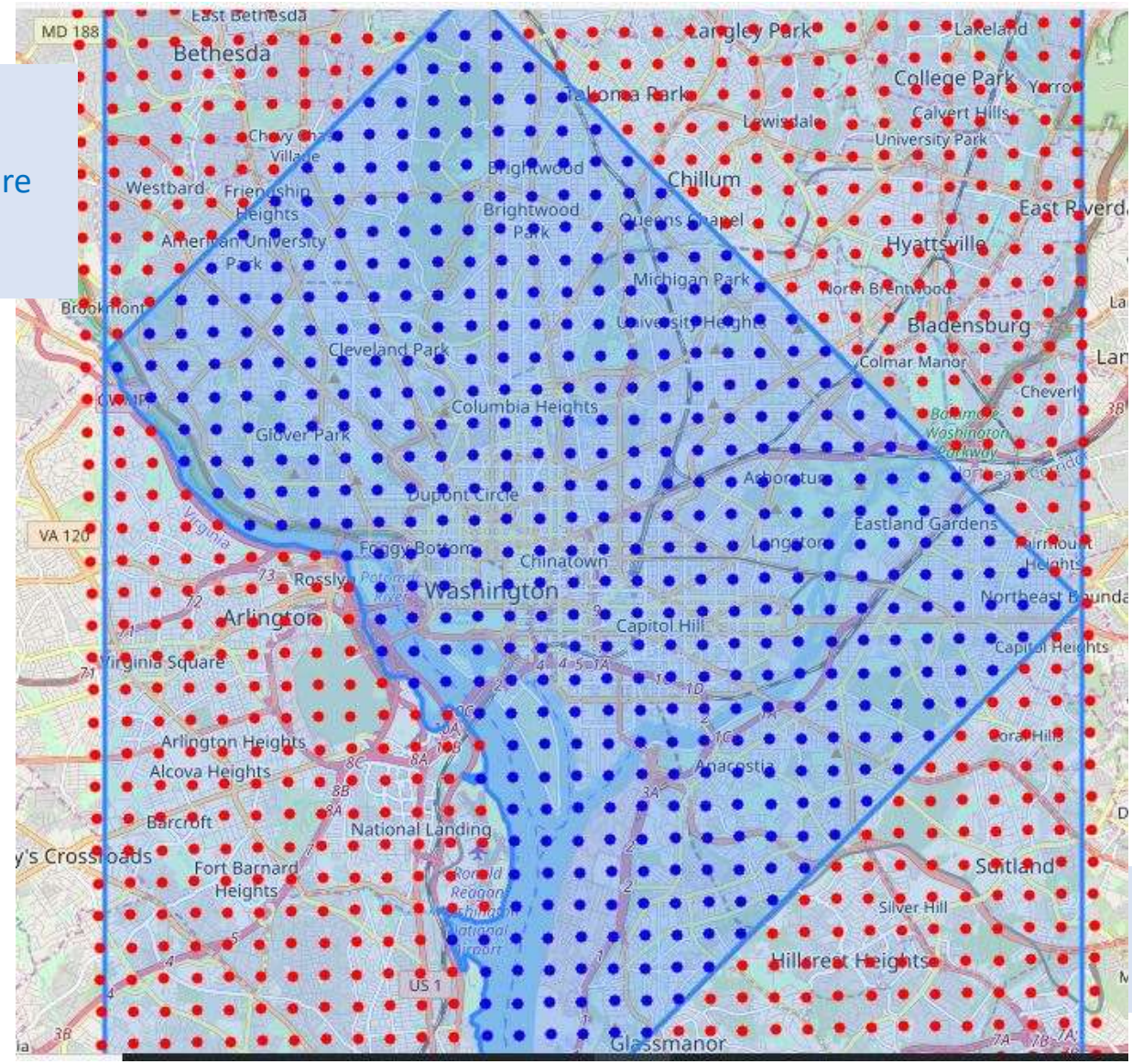
1216 grids were generated inside the rectangular object

The geometry of Washington DC was extracted from the geometry of US containing all the states geometry



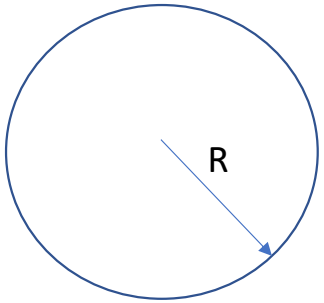
However we only need the grids contained inside Washington DC map

The blue dots are centers of candidate neighborhoods we are interested in further analysis



Among the 1216 grids only 496 are contained inside Washington DC map

3. Foursquare: Food venues located in Washington DC were collected by calling API from Foursquare.com



To obtain restaurants throughout Washington DC Foursquare API will be called for all 1217 candidate neighborhoods generated

Required information for retrieving food venues dataset in a certain neighborhood are:

- ✓ Radius of the candidate neighborhood
- ✓ Latitude/Longitude of the center
- ✓ Food venue ID
- ✓ Client secrete and client id obtained from Foursquare

The dataset obtained will be sorted out according to their type and stored into a dataframe object and saved

Food
4d4b7105d754a06374d81259

- Afghan Restaurant
  503288ae91d4c4b30a586d67
- African Restaurant
  4bf58dd8d48988d1c8941735
  - Ethiopian Restaurant
    4bf58dd8d48988d10a941735
- American Restaurant
  4bf58dd8d48988d14e941735
  - New American Restaurant
    4bf58dd8d48988d157941735
- Asian Restaurant
  4bf58dd8d48988d142941735

Images were extracted from Foursquare

Sample snippets showing how dataset are arranged in Foursquare database

- Asian Restaurant
  4bf58dd8d48988d142941735
  - Burmese Restaurant
    56aa371be4b08b9a8d573568
  - Cambodian Restaurant
    52e81612bcbc57f1066b7a03
  - Chinese Restaurant
    4bf58dd8d48988d145941735
    - Anhui Restaurant
      52af3a5e3cf9994f4e043bea
      Supported countries: CN, HK, MO, MY, TW, SG
    - Beijing Restaurant
      52af3a723cf9994f4e043bec
      Supported countries: CN, HK, MO, MY, TW, SG, JP

Methodology

1. ***Exploratory data analysis***
   - Food venues dataframe will be manipulated and restaurant dataframe will be produced
   - Both food venues and restaurant dataframes will be grouped according to their category.
   - The results will be visualized using matplotlib bar charts

2. ***American restaurant data analysis***
   - During this stage of analysis folium visualizing tool will be used
   - Restaurant number distribution will be visualized using folium heat map plot
   - American restaurant density distribution will be visualized using folium heat map
   - Regions of low distribution of restaurants will be identified visually

cont'd …

Methodology

… cont'd

3. Finding optimal locations for opening new American restaurant

- New neighborhoods having radius of 100 meters will be generated within 2000 meters of radius from the center
- Addresses of each neighborhood generated will be obtained using reverse geocoding method by calling Mapbox API.

- Number of restaurants lied within 250 meters of radius from the center of each neighborhood generated will be counted and stored in a dataframe
- Distance of the closest American restaurant from center of each neighborhood will be calculated and stored in a dataframe

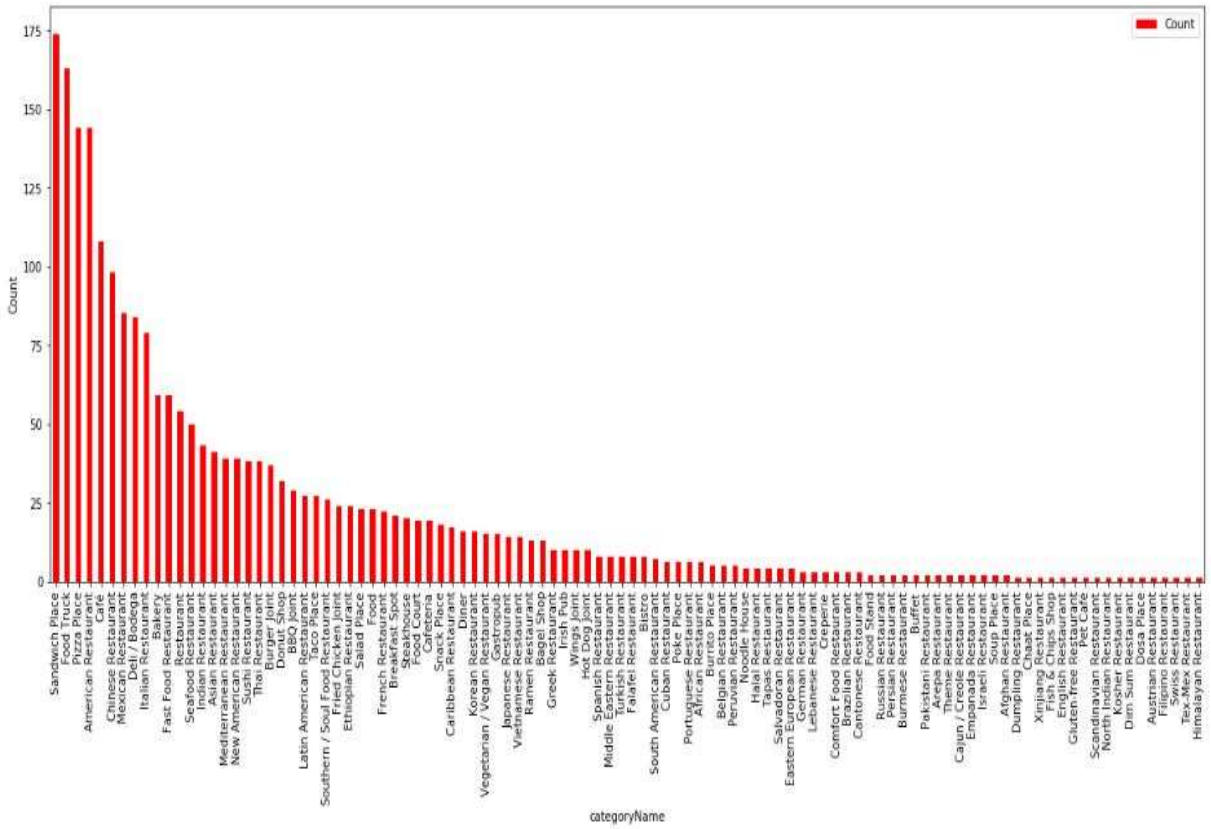
Cont'd …

Methodology

- Neighborhoods having less than 2 restaurants in 250 meters of radius from their respective center will be kept
- The locations selected above were further subjected to a test condition. Those locations where the closest American restaurant is not in 400 meters radius will be identified as potential locations

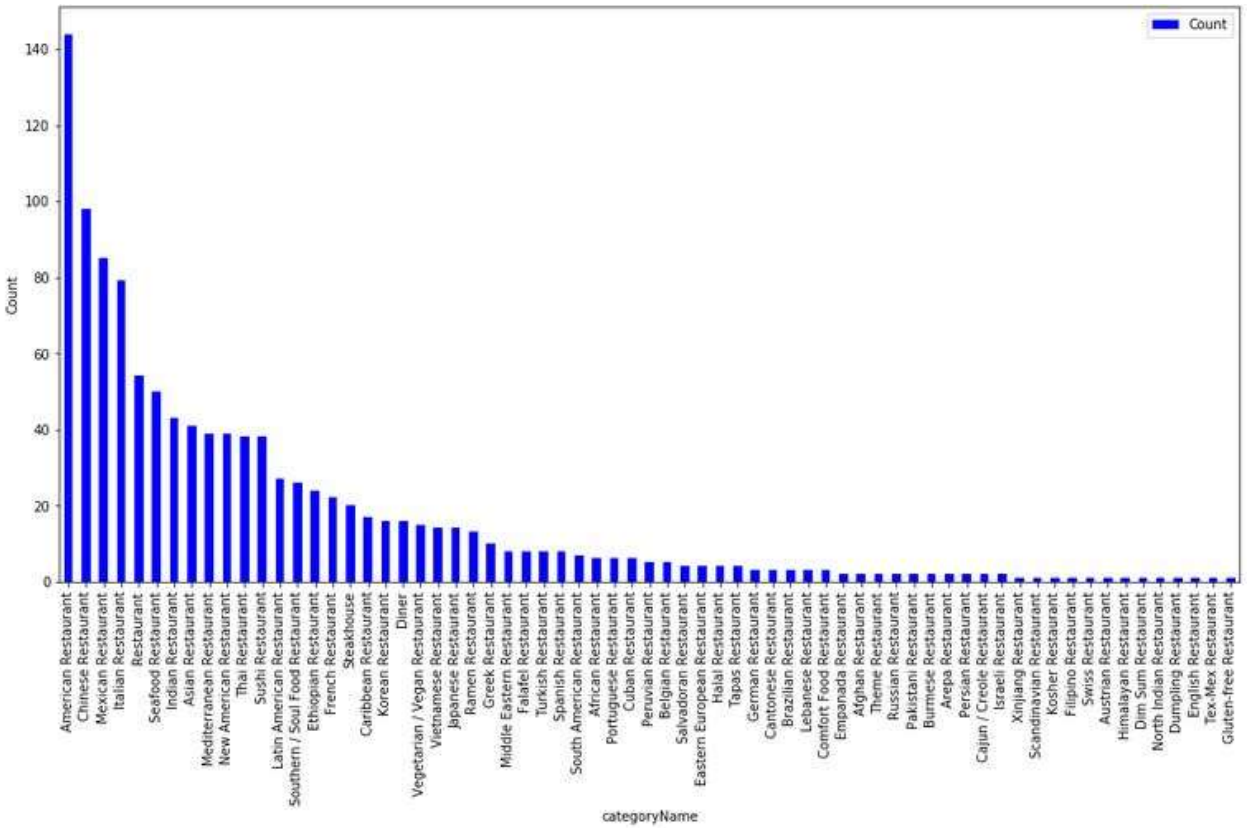
- These potential locations will be clustered using k-means algorithm and their centers will be identified
- Addresses for these cluster centers will be obtained using reverse geocoding technique by using Mapbox API
- Distance will be calculated from White House to each cluster's centers

Results and Discussions



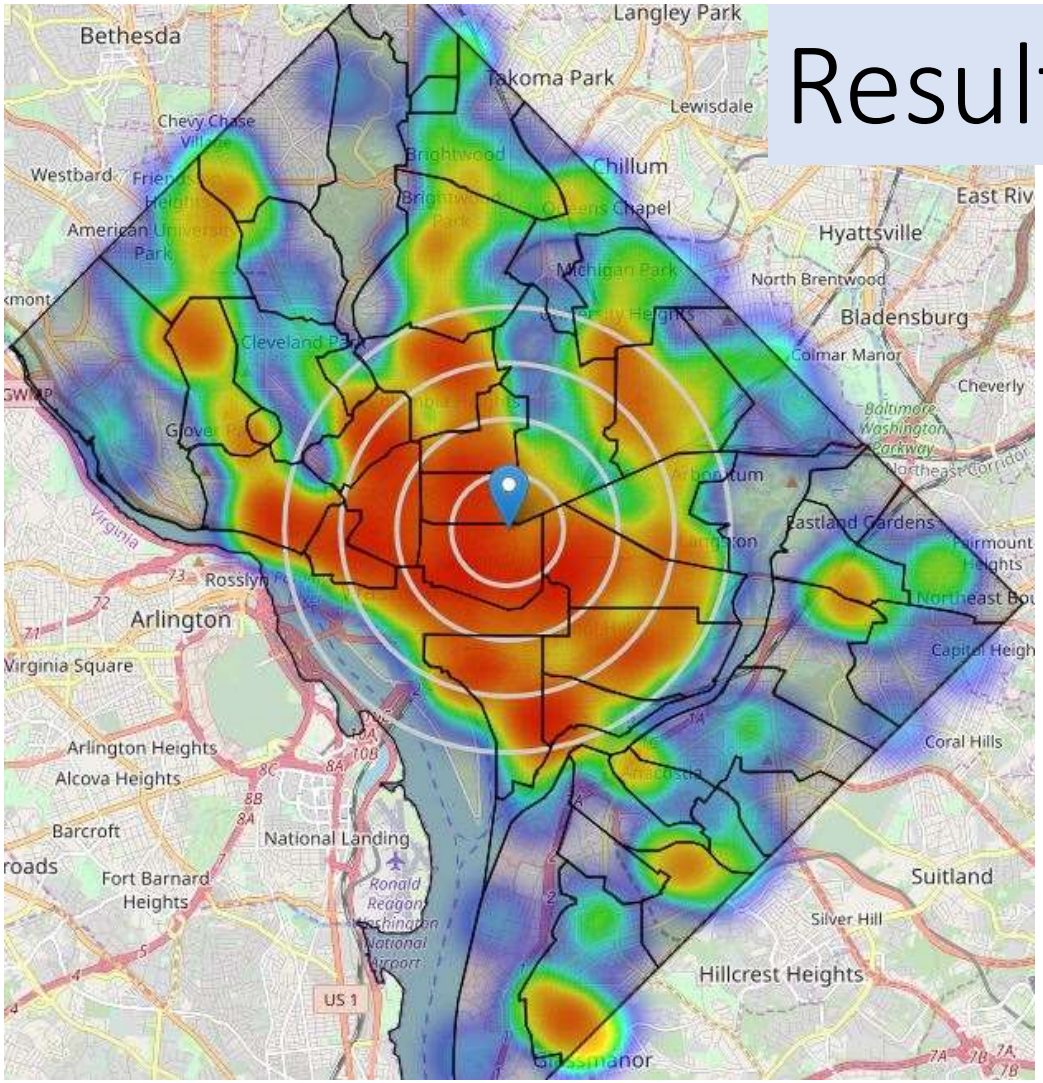
- Sandwich Place is the number one most populated food venue in Washington DC and followed by food truck, American restaurant and Pizza place.
- The restaurants variety ranges from South America to Asia and from Africa to Europe spanning all over the entire world.
- Among the exotic restaurants Chinese restaurant is the most common and followed by Mexican, Italian and Indian restaurants.
- By far American restaurants are the most populated restaurants in DC

Results and Discussions



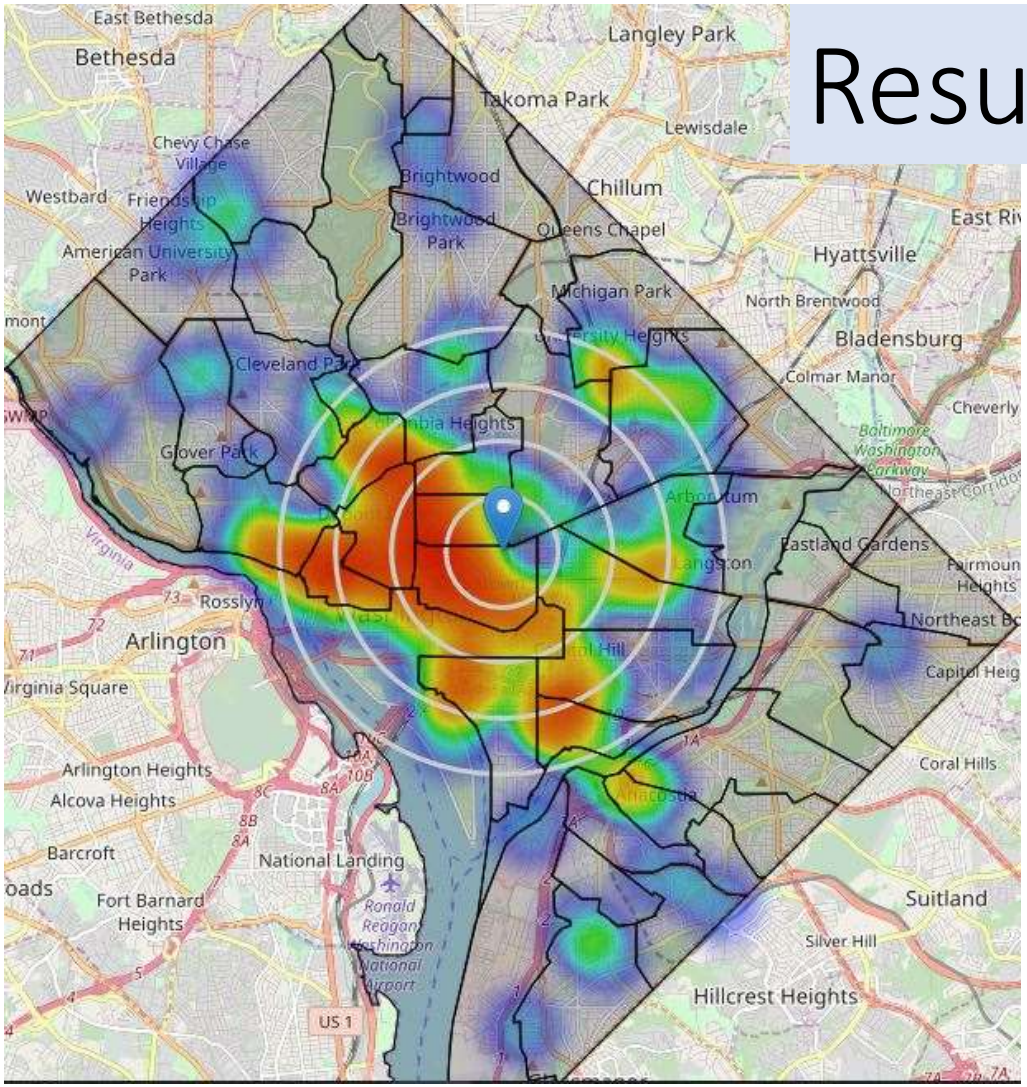
- Food venues were filtered out to fit the conventional definition of restaurant criteria and the result was categorized plotted as shown in the bar plot
- Among the restaurants American restaurant is the number one most populated restaurant in Washington DC and followed by Chinese restaurant
- Mexican and Italian restaurants trailed third and fourth
- A lot of restaurants have single or couple of venues in Washington DC area

Results and Discussions

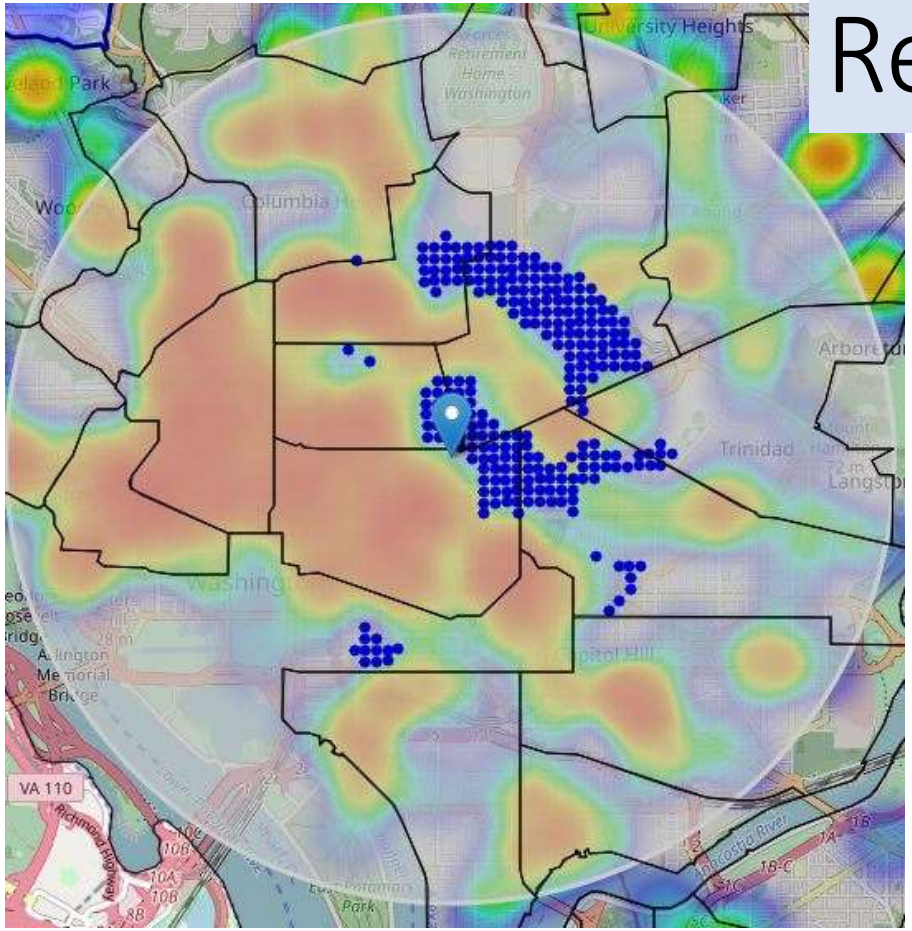


- Restaurants distribution was plotted as folium heat map as shown
- The heat map shows most of the restaurants are distributed with in 4000 meters of radius range from the centroid
- Areas of highest density of restaurants are West, North West and South West of Washington DC
- In fact most of the restaurants are found in Cluster 1,2 3, 4, 5, 6, 7 and 8
- The actual folium multi layered heat map shows restaurants distribution in the actual neighborhoods

Results and Discussions



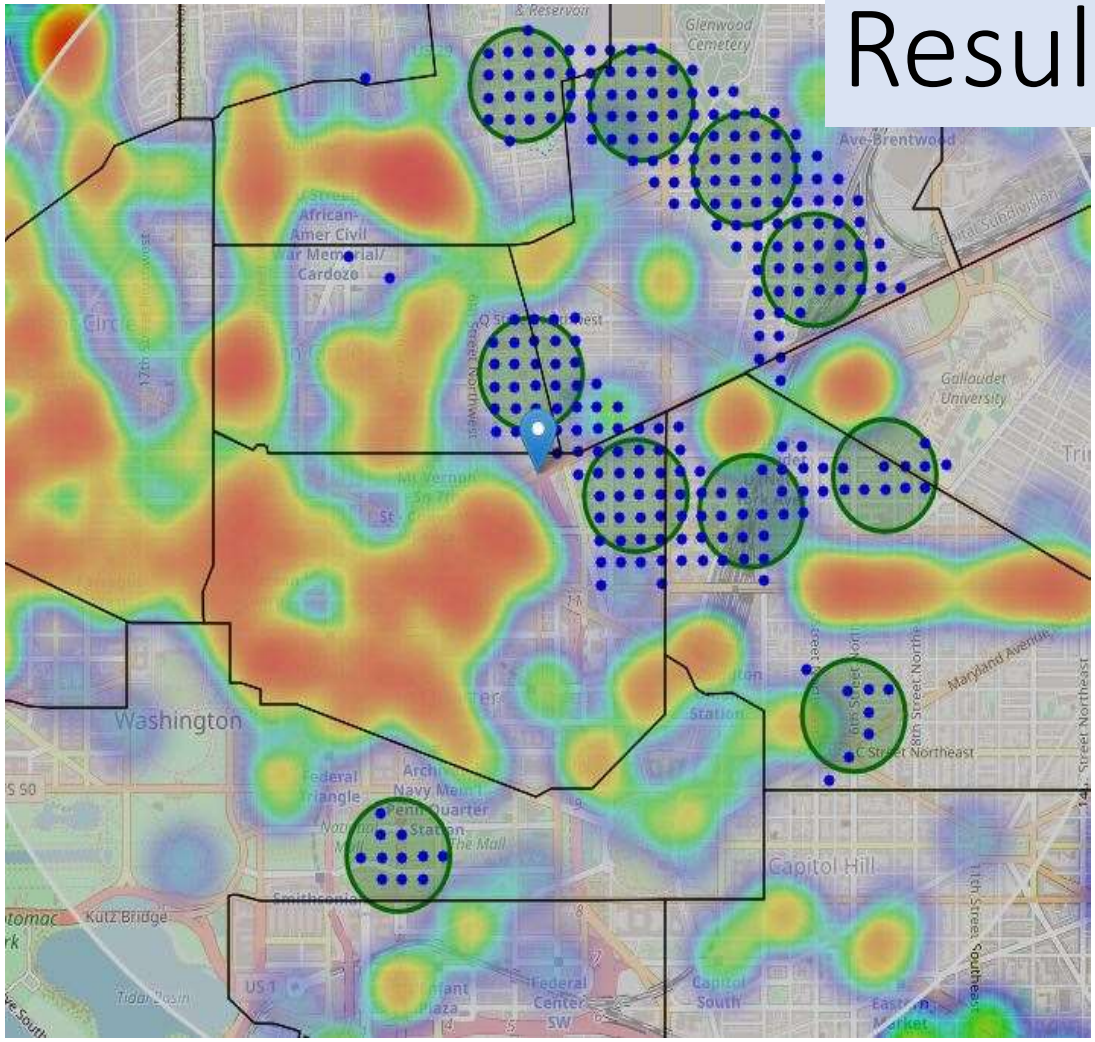
- American restaurants density distribution was plotted on folium heat map as shown
- American restaurant represents less than 16.5 percent of the total restaurants. It is evident from the map above the density of American restaurants is low
- From restaurant heat map low concentration of restaurants observed along east, north east and south east areas
- Most of restaurant activities are limited within 4000 meters distance range from the center



Results and Discussions

- The map shows potential areas for opening new American restaurant
- These areas are located in 2000 meters of radius range from the center of Washington DC
- Each of those locations has no more than two restaurants in radius of 250m
- No American restaurant is to be found in 400 meters of radius from the center

Results and Discussions



- Further narrowing the location detected 10 clusters were created based on k-means algorithm
- The folium map shows 10 clustered regions where the centers were found from k-mean algorithm
- The addresses of each clusters was also obtained by calling Mapbox API

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
- Parallelly 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
- many locations of the city food venues were hardly to find specially as we go further from center

- This study shows 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
- Additional factors considered for deciding the final location are: 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