Coursera Capstone - Washington DC Apartments

Capstone Project for the IBM Data Science Professional Certificate.

1.0 Introduction:

Washington DC is known for being one of the most culturally diverse cities in the United States. With so many individuals coming from all around the globe for work purposes, many communities have found homes in the District over the years. A fantastic side effect of this cultural diversity is that you can find cuisine of almost any variety represented here.

The goal of this project is to identify apartments around DC that have the best food options available and what similar apartments can be grouped together.

1.1 Problem

If I am an individual moving to DC and I want to know how to which apartment building will give me the best dining options how would I do it?

To solve this problem we will need several things:

- Data of apartments in the DC area
- Data of different food venues around these apartments
- A way to analyze said data

1.2 Data

The data that will be used to address the above problem will come from 2 different sources:

1. The Basic Business Licenses data from the local DC government. The City of DC has large datasets that are available to the public. These data cover a variety of interesting topics and in our case we will use a set that contains information on businesses licensed

in the city. An apartment is a business and so we will be able to filter or just the DC area apartment buildings

2. Venue data form Foursquare API. We will find the top venues for each apartment by leveraging the Foursquare API. We can set parameters to gather only data on food venues to tailor our results to the problem.

2.0 Methodology:

In this section we will tackle all data import and processing as well as all technical analysis.

2.1 Libraries Utilized

The Python Libraries I used in this project are as follows:

- Pandas – Data import and transformation
- Numpy – Data Manipulation
- Geopy – Geographic data fetching
- Matplotlib – Data Visualization
- ScikitLearn – Algorithms
- Folium – Map Rendering

2.2 Data Import and load

The first thing I did was load in the raw csv file that I got from the public DC data website. I then did several manipulations on this to get a resulting data frame with only the variables I needed. Additionally, I filtered out any NA values as this would cause issues down the line when using the Foursquare API. Next I created a function that allowed me to systematically construct Foursquare API calls for every apartment in my data frame. I then retrieved the most common food venues within a 500 meter radius of every apartment and joined the results together.

The Foursquare API allows for a dynamic set of fields to be passed via the URL that is passed. For my purposes, my function essentially looped over every apartment entry and passed the latitude and longitude values for said apartment. The API then would search for venues surrounding that specific point and return them into a data frame.

2.3 Clustering

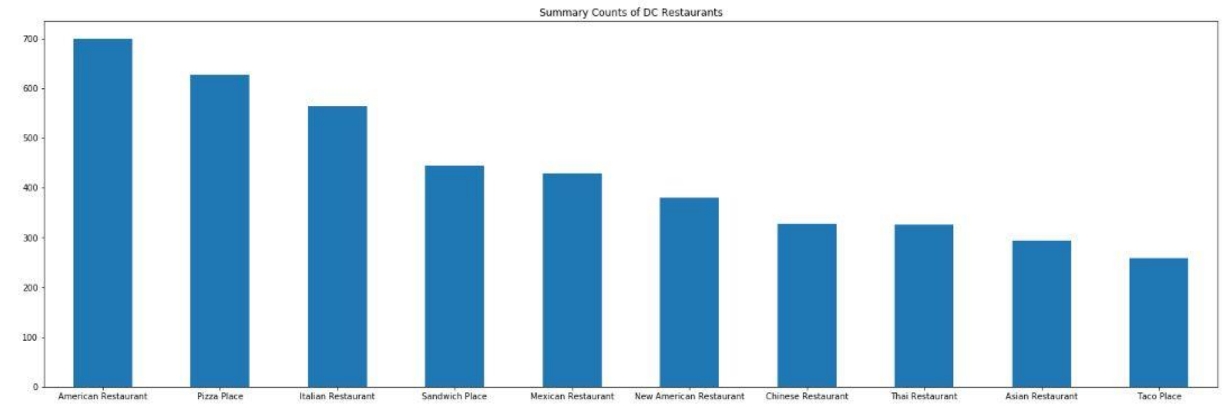
With my prepared data I conducted one hot encoding and then clustering via K Means Clustering. I conducted clustering using a k of 6 and my results were analyzed.

3.0 Results

In this results section I will hope to answer our original problem question of where in DC a food lover should live.

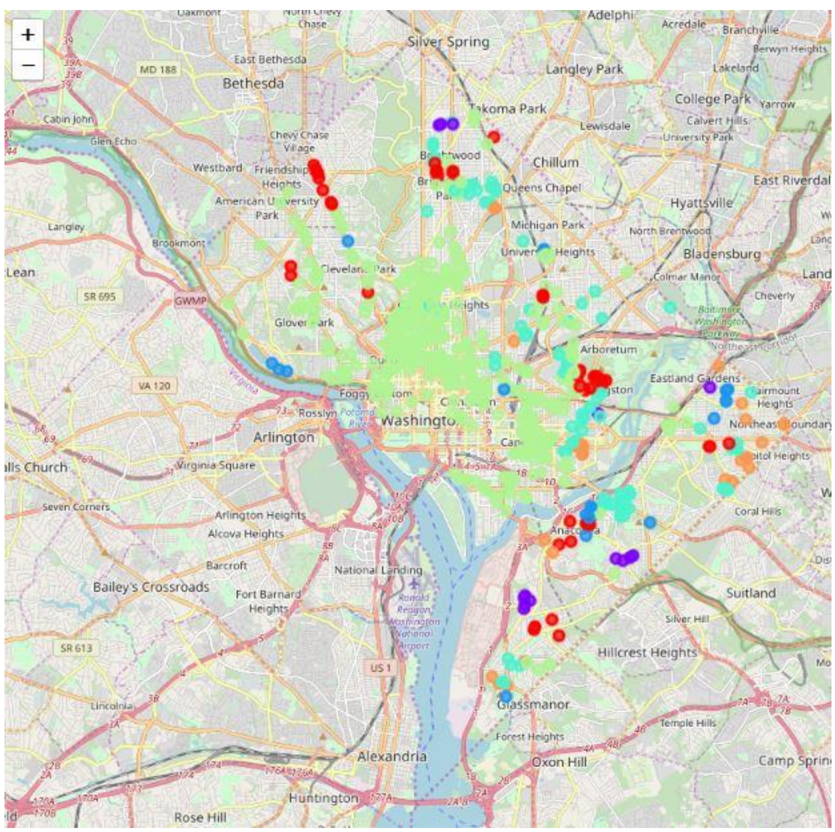
3.1 Venue Analysis

I plotted the overall counts of the Foursquare venues to get an overarching picture of what the restaurant scene was like for most apartments



The top results were that the 3 most common venues for DC apartments were American Restaurants, Pizza Places, and Italian Restaurants.

After this I mapped all of my clusters onto the DC area to visualize the breakdown.



The largest cluster was cluster 4. The majority of these apartments are located in central DC and my guess is that this resulted in these apartments having the broadest range of food venues. The algorithm likely created a "catch all" cluster that these apartments were dumped into. I take a closer look at some of the apartments within cluster 4 to confirm this assumption.

	Apartment	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2747	4827 RUGBY AVENUE, S BETHESDA MD 20814	Indian Restaurant	Pizza Place	American Restaurant	French Restaurant	Asian Restaurant	Food Truck	Sandwich Place	Spanish Restaurant	Mexican Restaurant	Filipino Restaurant
2766	PO BOX 4130 OAKTON VA 22124	Vietnamese Restaurant	Pizza Place	Italian Restaurant	Mexican Restaurant	Indian Restaurant	German Restaurant	Salad Place	Sandwich Place	Sushi Restaurant	Dumpling Restaurant
2769	P O BOX 11597 WASHINGTON DC 20008	American Restaurant	New American Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Mexican Restaurant	Sushi Restaurant	Belgian Restaurant	Pizza Place	Food Truck	Sandwich Place
2777	6600 SANY SPRING RD WASHINGTON DC 20012	Thai Restaurant	Ethiopian Restaurant	Southern / Soul Food Restaurant	African Restaurant	Asian Restaurant	Italian Restaurant	Sushi Restaurant	New American Restaurant	Middle Eastern Restaurant	Korean Restaurant
2786	3031 SUGAR LANE VIENNA VA 22181	American Restaurant	Thai Restaurant	Sushi Restaurant	Mexican Restaurant	Italian Restaurant	Indian Restaurant	Latin American Restaurant	Pizza Place	New American Restaurant	Ethiopian Restaurant
2789	730 24TH NW ST WASHINGTON DC 20037	American Restaurant	Italian Restaurant	Sushi Restaurant	Indian Restaurant	South American Restaurant	Falafel Restaurant	Mexican Restaurant	New American Restaurant	Chinese Restaurant	Restaurant
2816	717 IRVING ST NE WASHINGTON DC 20017	Food Court	American Restaurant	Chinese Restaurant	Latin American Restaurant	Food & Drink Shop	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant
2818	1350 CONNECTICUT AVE WASHINGTON DC 20036	Mediterranean Restaurant	Pizza Place	Thai Restaurant	Italian Restaurant	Sandwich Place	Filipino Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant
2821	901 TAYLOR STREET, N WASHINGTON DC 20017	Latin American Restaurant	Xinjiang Restaurant	Food & Drink Shop	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	Food
2836	2501 PORTER ST NW WASHINGTON DC 20008	Thai Restaurant	Italian Restaurant	Mediterranean Restaurant	Mexican Restaurant	Xinjiang Restaurant	Sandwich Place	American Restaurant	Asian Restaurant	Cafeteria	Chaat Place
2874	4938 HAMPDEN LN BETHESDA MD 20814	Italian Restaurant	Indian Restaurant	New American Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Salad Place	Sandwich Place	Mediterranean Restaurant	Taco Place
2904	1116 5TH ST NW WASHINGTON DC 20001	American Restaurant	New American Restaurant	Italian Restaurant	Southern / Soul Food Restaurant	Pizza Place	Korean Restaurant	Brazilian Restaurant	Vegetarian / Vegan Restaurant	Mexican Restaurant	Salad Place
2919	9615 Maury Rd FAIRFAX VA 22032	Restaurant	American Restaurant	New American Restaurant	Latin American Restaurant	African Restaurant	Thai Restaurant	Taco Place	Sushi Restaurant	Salad Place	Mexican Restaurant
2956	6172 DEER RIDGE TRL SPRINGFIELD VA 22150	Latin American Restaurant	Mexican Restaurant	Chinese Restaurant	South American Restaurant	Cuban Restaurant	Ethiopian Restaurant	Caribbean Restaurant	Sandwich Place	Mediterranean Restaurant	Fast Food Restaurant
2980	3460 14TH STREET NW WASHINGTON DC 20010	Pizza Place	Mexican Restaurant	Latin American Restaurant	Asian Restaurant	Vietnamese Restaurant	South American Restaurant	Caribbean Restaurant	Cantonese Restaurant	Ethiopian Restaurant	Cuban Restaurant

4.0 Discussion

The discussion section of this report will focus on what can be gained from the results, as well as what a real life example could look like.

4.1 Key Points

My results have some key takeaways.

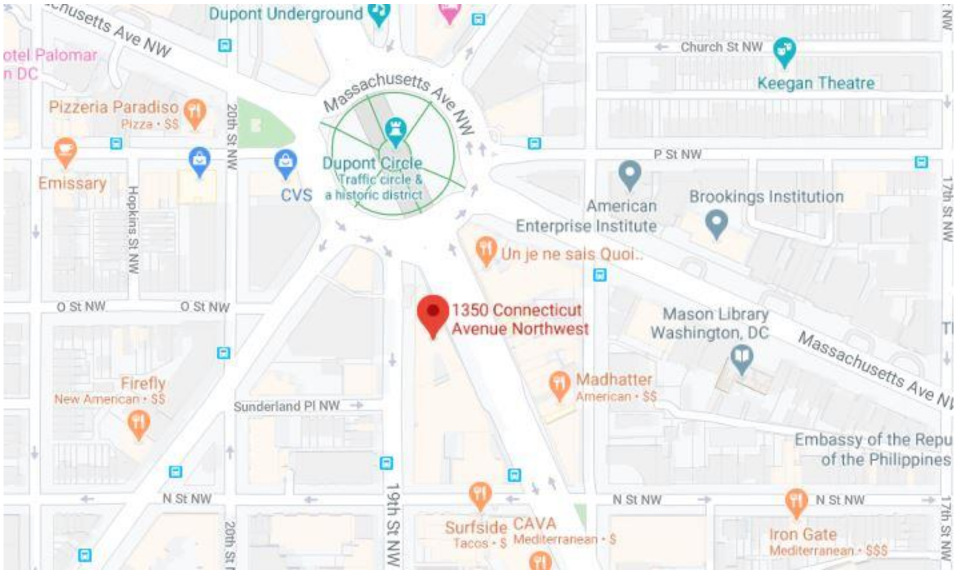
1. DC has a large volume of restaurants to choose from that encompass a variety of dining options. A summary of the most common venues for every apartment can be found below:
   - American Restaurants - apx. 700
   - Pizza Place - apx. 650
   - Italian Restaurant - apx. 600
2. Clustering provided us with good insights into what types of apartments could exist in terms of dining options. Someone looking for an apartment could use the following summary to help their decision:
   - Cluster 0 (West & SW DC) - Fast food and various Asian cuisines options
   - Cluster 1 (Central & West DC) - Chinese and Dumpling restaurants
   - Cluster 2 (Central & North DC) - Good variety with some preference towards Asian Cuisine
   - Cluster 3 (Central & West DC) - People who really like sandwich places and Chinese Food
   - Cluster 4 (Central DC) - Widest variety of options. Something for everyone
   - Cluster 5 (Central & South DC) - People who love American Restaurants and Sandwiches

4.2 Example

We can look at a real example to see how someone would actually use a project like this.

I have chosen a random Apartment out of Cluster 4 and have looked into the rent and available units.

The Apartment I chose is right in the middle of Dupont Circle and we can see via Google maps that it does in fact have the dining options that my research found. The company that manages the property can be found here and we can see there are listings available



4.2 Evaluating

I think the key pieces of this project to evaluate would be the use of K Means Clustering as my algorithm choice, as well as the reality of using only food venues to make a choice for the a arrangement.

Point 1:

While there are certainly other clustering methods available, I wanted to use K Means because it is fast and works just fine for working with geographic data. If I wanted to repeat this experiment with even more data, for instance, multiple cities worth, I could still see reasonable calculation speeds by sticking with K Means. In this sense it is quite scalable. The biggest downside I see with my use of the algorithm is the fact that I may not have gotten the most optimized number of clusters. Other algorithms such as DBSCAN arguably allow for greater assessment of spatial data. In this instance though I think my results along with the scalability of my project justify the use of K Means

Point 2:

In terms of this projects real world functionality I obviously recognize that food options are not the only thing to drive someone's housing decision. For the sake of this project acting as a proof of concept I think using restaurants as a variable for clustering is completely acceptable and also keeps things fun and relatable for everyone.

5.0 Conclusion

Overall this project accomplished its goal. The problem was meant to be rather light hearted while at the same time showcasing what machine learning methods are capable of from an analysis stand point. Who know, maybe a true foodie out there could actually love cuisine enough to base their entire living situation around it.

I hope you enjoyed this project and found it educational.

Best,

Kai Hannum