

# **Luxury Condominium Development in the D.C. Metro Area**

## **IBM Data Science Professional Certificate**

### **Capstone Project**

#### **Introduction**

Many young professionals are moving to the Washington D.C. area to pursue a range of employment options in the federal workforce, government contracting, non-profit sector, and private industry. The D.C. Metro Area has many unique communities. Housing in this region is at a premium as increasing job opportunities attract many professionals to the region.

Throughout the D.C. Metro areas there many cities and suburbs which provide its residents with rich amenities and cultural experiences. This data science project investigates different cities, suburbs, and neighborhoods in the D.C. Metro area to determine the best location to build a new development of luxury condominium units. The ideal customer for these units would be early to mid-career professional singles and couples interested in easy access to amenities including restaurants, bars, health clubs, and shopping. The target price of the planned condominium units would range from 500,000- 750,000 depending upon unit specifications (1-3 bedrooms; 2-4 bathrooms). To investigate this question, the researcher collected data on the demographics, property values, income, and popular venues in different neighborhoods in the D.C., Maryland, Virginia (DMV) area. After some initial investigations and analysis, twenty areas were selected to represent a diverse collection of cities, suburbs, and neighborhoods throughout the Metro Area. The ideal location of the new development would have a relatively low median age, high median family income, and moderately high property values. The selected area would also have access to numerous, highly rated amenities for young professionals.

#### **Data**

Demographic data for this research study was collected from the U.S. Census available at [data.gov](https://data.gov). Median household income and property values were collected from <https://datausa.io/>. Datausa.io collects data from numerous government organizations including the Department of Housing and Urban Development, Bureau of Labor Statistics and the Bureau of Economic Analysis. This information was used to assess cities, suburbs and neighborhoods in the D.C. Metro area that may be attractive locations for the new housing complex. From initial observations and data analysis, twenty cities/neighborhoods were selected as potential sites for the condo development.

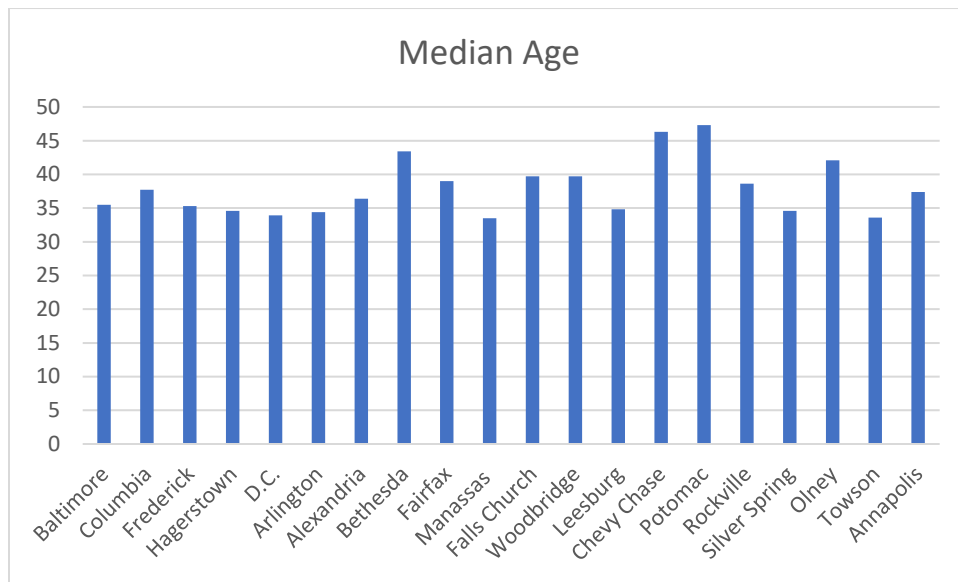
<i>State</i>	<i>City</i>	<i>Med Age</i>	<i>Med Income</i>	<i>Med Prop Value</i>	<i>Population</i>	<i>Latitude</i>	<i>Longitude</i>
<i>MD</i>	Baltimore	35.5	51000	167800	619493	39.2904	-76.6122
<i>MD</i>	Columbia	37.7	103707	367600	99615	39.2037	-76.861
<i>MD</i>	Frederick	35.3	67828	255100	71408	39.4143	-77.4105
<i>MD</i>	Hagerstown	34.6	40761	148300	40306	39.6418	-77.72
<i>DC</i>	D.C.	33.9	85203	617900	633427	38.9072	-77.0369
<i>VA</i>	Arlington	34.4	112138	634300	234965	38.8816	-77.091
<i>VA</i>	Alexandria	36.4	93370	537900	144301	38.8048	-77.0469
<i>MD</i>	Bethesda	43.4	154559	877300	63374	38.9847	-77.0947
<i>VA</i>	Fairfax	39	106870	501900	22549	38.8462	-77.3064
<i>VA</i>	Manassas	33.5	77551	307000	39300	38.7509	-77.4753
<i>VA</i>	Falls Church	39.7	114795	742000	12751	38.8823	-77.1711
<i>VA</i>	Woodbridge	39.7	102844	333600	54275	38.6582	-77.2497
<i>VA</i>	Leesburg	34.8	105844	394000	54215	39.1157	-77.5636
<i>MD</i>	Chevy Chase	46.3	173333	897000	9381	38.9693	-77.0789
<i>MD</i>	Potomac	47.3	187568	893600	44865	39.0182	-77.2086
<i>MD</i>	Rockville	38.6	100436	498200	68401	39.084	-77.1528
<i>MD</i>	Silver Spring	34.6	76608	477800	71452	38.9907	-77.0261
<i>MD</i>	Olney	42.1	136786	491300	33844	39.1532	-77.0669
<i>MD</i>	Towson	33.6	82062	345500	55197	39.4015	-76.6019
<i>MD</i>	Annapolis	37.4	81143	395900	39321	38.9784	-76.4922

In order to determine more information about the selected cities and neighborhoods, Foursquare API is used to collect information on the types of venues and their frequencies and the top attractions in each neighborhood. The ideal location for the neighborhood would have access to amenities including restaurants, bars, and music venues. This information from Foursquare is instrumental in selecting the perfect location to attract young, working professionals to purchase the new condominium units. Data from Foursquare is analyzed to determine communities with high frequencies of restaurants, bars, and coffee shops, along with high necessities and services like pharmacies and grocery stores. Neighborhoods and cities will be analyzed using K-means clustering to determine similar neighborhoods. From there, each neighborhood will be further investigated to determine top venues in each vicinity. This will help determine the best location for the target demographic.

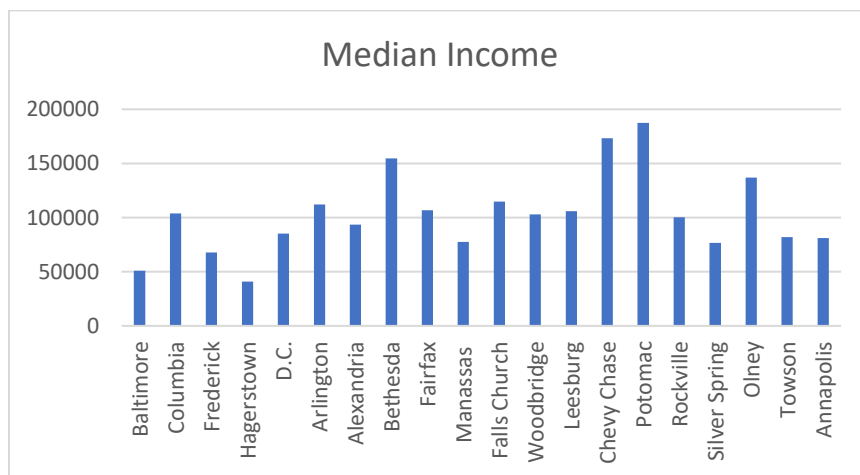
## Methodology

There are many neighborhoods and cities within the D.C. Metro area. In order to select the location luxury condominium unit, an initial set of criteria were developed. The location must be in the city of D.C. or a surrounding suburb. Areas should be within an hour and half drive of D.C. The initial investigation examined the median age, population, median income, and median property values of each location. This helped identify areas that meet the target demographic for the housing development. This data was collected from the United States Census and datausa.io. Twenty cities, neighborhoods, and suburbs were identified. Graph 1 highlights the median age of each location. Graph 2 shows median income.

Graph 1: Median Age of Each Neighborhood in the Study

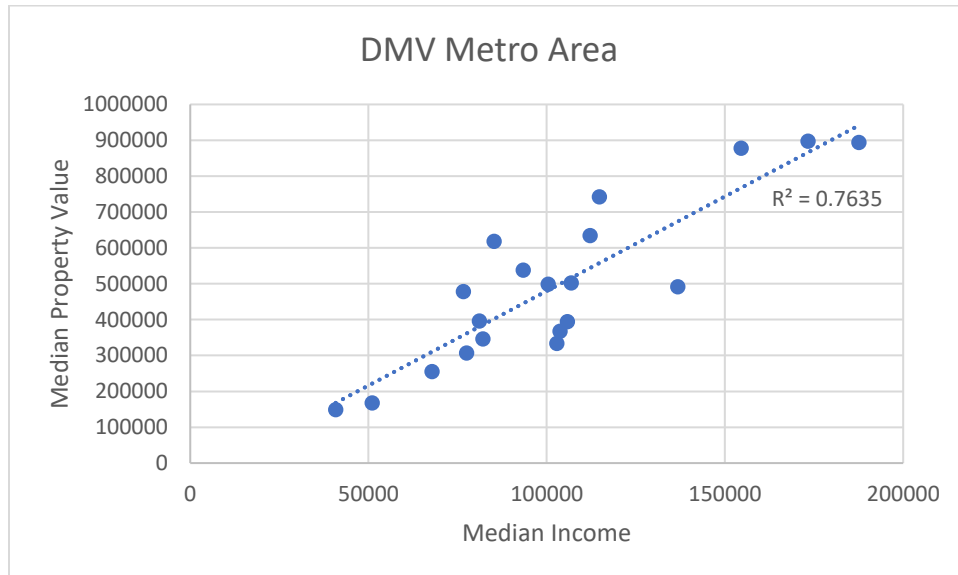


Graph 2: Median Income of Each Neighborhood in the Study



An investigation of property values relative to median income reveals a positive correlation and strong goodness of fit (R-squared value of 0.76). This is logical given that consumers with higher incomes will purchase homes in more desirable areas.

Graph 3: Income Versus Property Values



Through this initial investigation, the researcher identified that target neighborhoods should have a low median age and average property values of higher than \$500,000 but lower than \$750,000. While Bethesda, Chevy Chase, and Potomac have high median incomes, the median age is higher than similar areas. After this initial analyses Arlington, Columbia, Alexandria, Fairfax, Falls Church, Leesburg, Rockville emerge as top contenders for the new development.

In order to determine the best locations, further data was collected on the top venues in the area using the Foursquare API. The frequency of these venues was determined as well as the top-rated venues in each location. Based on this data, the cities and neighborhoods were clustered using K-means clustering. This helped divide the locations into similar neighborhoods. While there was a strong amount of similarities, about half of the locations were removed from consideration after this analysis. Several different numbers of clusters were investigated with 10 being the optimal level. From this division, one cluster was selected as having the highest concentration of bars, restaurants, music venues, and amenities (grocery stores, pharmacies, etc). This list was compared to the initial selections to narrow down the finalists to Rockville, Falls Church, and Alexandria. Each of these cities were further investigated and based on top venues and recommendations from Foursquare, Falls Church, Virginia was the final selection.

## Results

After initial examination of the data set, several neighborhoods, cities, or suburbs emerged as potential sites for the new housing development. Foursquare API was used to gain additional insights into the types of venues and amenities available. Foursquare returned information on 983 venues in the cities and neighbors investigated. A sampling of this information is below.

	City	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Baltimore	39.2904	-76.6122	Larry Flynt's Hustler Club	39.289615	-76.609833	Strip Club
1	Baltimore	39.2904	-76.6122	Under Armour Performance Center	39.289126	-76.613973	Clothing Store
2	Baltimore	39.2904	-76.6122	Chez Hugo	39.289108	-76.611811	French Restaurant
3	Baltimore	39.2904	-76.6122	The Quintessential Gentleman	39.288056	-76.612107	Salon / Barbershop
4	Baltimore	39.2904	-76.6122	Halal Food Cart	39.291095	-76.613960	Food Truck
5	Baltimore	39.2904	-76.6122	Kimpton Hotel Monaco Baltimore Inner Harbor	39.289793	-76.615647	Hotel
6	Baltimore	39.2904	-76.6122	Hotel RL Baltimore Inner Harbor	39.288828	-76.611903	Hotel
7	Baltimore	39.2904	-76.6122	Ida B's Table	39.292654	-76.610420	Southern / Soul Food Restaurant
8	Baltimore	39.2904	-76.6122	Baltimore Farmers' Market & Bazaar	39.293297	-76.610490	Farmers Market
9	Baltimore	39.2904	-76.6122	Chesapeake Shakespeare Company Downtown Theater	39.288970	-76.612150	Theater

Analysis of these venues determined these frequencies of venue types in each location. The example for Alexandria is below. This information allowed the researcher to examine the types of amenities that are in each neighborhood. This helps to narrow down potential locations.

----Alexandria---- venue freq

0 American Restaurant 0.05

1 Spa 0.04

2 Boutique 0.03

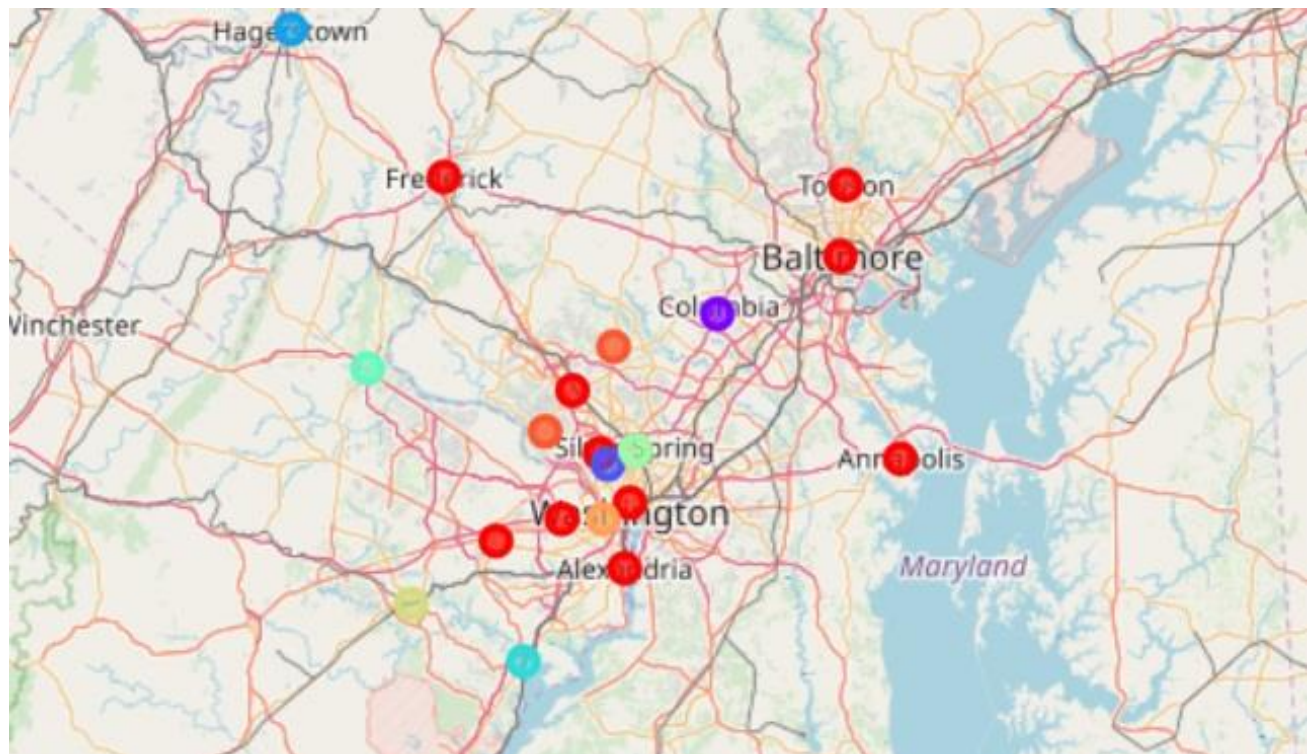
3 Coffee Shop 0.03

4 Clothing Store 0.03

Further analysis showed the 10 most common venues in each neighborhood/city.

	City	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
0	Alexandria	American Restaurant	Spa	Italian Restaurant	Clothing Store	Coffee Shop	French Restaurant	Boutique	Cajun / Creole Restaurant	Café
1	Annapolis	Bar	Sandwich Place	Coffee Shop	Pub	Historic Site	Gift Shop	Restaurant	Mexican Restaurant	Ice Cream Shop
2	Arlington	Park	Café	Convenience Store	Korean Restaurant	Spa	Liquor Store	South American Restaurant	Gas Station	Mediterranean Restaurant
3	Baltimore	Coffee Shop	American Restaurant	Food Truck	Hotel	Clothing Store	Gym / Fitness Center	Lounge	Pizza Place	Café
4	Bethesda	Coffee Shop	Deli / Bodega	Spa	Juice Bar	American Restaurant	Seafood Restaurant	Bakery	Sandwich Place	French Restaurant

This venue information was added to the initial dataframe and used to generate a K-means clusters for the twenty locations. These locations were clustered based on the venue types to reveal similar cities and neighborhoods. The map generated is below.



## Clusters and Top Venues

### Cluster 0

Name	Most Common	Second	Third	Fourth	Fifth
Baltimore	Coffee Shop	Amer. Restaurant	Food Truck	Hotel	Clothing Store
Frederick	Amer. Restaurant	Park	Coffee Shop	Art Gallery	Bar
D.C.	Hotel	Amer. Restaurant	Hotel Bar	Coffee Shop	Salad Place
Alexandria	Amer. Restaurant	Spa	Italian Restaurant	Clothing Store	Coffee Shop
Bethesda	Coffee Shop	Deli	Spa	Juice Bar	Amer. Restaurant
Fairfax	Pizza	Yoga Studio	Smoothie Bar	Restaurant	Record Shop
Falls Church	Bar	Italian Restaurant	Grocery Store	Chinese Restaurant	Coffee Shop
Rockville	Bakery	Ice Cream Shop	Coffee Shop	Salon	Bubble tea
Towson	Cosmetic Store	Amer. Restaurant	Pizza	Café	Jewelry Store
Annapolis	Bar	Sandwich Shop	Coffee	Pub	Historic Site

### Cluster 1

Name	Most Common	Second	Third	Fourth	Fifth
Columbia	Business services	Photography Studio	Performing Arts Center	Yoga Studio	Fast Food

### Cluster 2

Name	Most Common	Second	Third	Fourth	Fifth
Chevy Chase	Sandwich shop	Bank	Electronic Store	Shoe Store	Shoe Repair

### Cluster 3

Name	Most Common	Second	Third	Fourth	Fifth
Hagerstown	Theater	Amer Restaurant	Bus Station	Pizza	Convenience Store

#### Cluster 4

Name	Most Common	Second	Third	Fourth	Fifth
Woodbridge	Pizza	Sandwich shop	Ice Cream	Thrift store	Dance studio

#### Cluster 5

Name	Most Common	Second	Third	Fourth	Fifth
Leesburg	Brewery	Amer. Restaurant	Chinese Restaurant	Sandwich Shop	Coffee

#### Cluster 6

Name	Most Common	Second	Third	Fourth	Fifth
Silver Spring	Asian Restaurant	Coffee	Brewery	Hotel	Greek Restaurant

#### Cluster 7

Name	Most Common	Second	Third	Fourth	Fifth
Manassas	Coffee Shop	Amer. Restaurant	Brewery	Mexican Restaurant	Bar

#### Cluster 8

Name	Most Common	Second	Third	Fourth	Fifth
Arlington	Park	Coffee Shop	Convenience store	Korean Restaurant	Spa

#### Cluster 9

Name	Most Common	Second	Third	Fourth	Fifth
Potomac	Bar	Grocery Store	Drug Store	Health Services	Restaurant
Olney	Ice Cream	Sandwich Shop	Italian Restaurant	Mexican Restaurant	Pizza

The results of the clusters reveal that cluster “0” was most appropriate for the new housing development. From this selection, Rockville, Alexandria, and Falls Church were the finalists. Each of these locations were more closely examined using data from the Foursquare API. Falls church had the highest frequency of bars, restaurants, grocery stores, and other popular amenities among young professionals.



----Falls Church----

	venue	freq
0	Bar	0.06
1	Sushi Restaurant	0.04
2	Chinese Restaurant	0.04
3	Italian Restaurant	0.04
4	Grocery Store	0.04

#### Top Venues in Fall Church

	name	categories	lat	lng
0	Dogwood Tavern	Bar	38.882998	-77.171894
1	Rare Bird Coffee Roasters	Coffee Shop	38.883206	-77.172715
2	Ireland's Four Provinces	Bar	38.882527	-77.171579
3	Northside Social	Bistro	38.884074	-77.171954
4	Maneki Neko	Sushi Restaurant	38.883342	-77.172961

## Discussion

This investigation examined different cities, suburbs, and neighborhoods in the D.C. Metro Area. After an initial analysis, twenty locations were selected as potential sites for a luxury condominium development geared toward young professionals. Selection required obtaining median incomes, property values, and median ages of the neighborhoods using data from the U.S. Census Bureau. Additional data collected from Foursquare identified top venues and frequencies of available amenities. Special attention was given to bars, restaurants, and grocery stores. The analysis of a K-means clustering of the neighborhoods revealed that 9 of the cities or neighborhoods met these criteria. However, when this data was combined with initial research, three finalists were established: Rockville, Falls Church, and Alexandria. After analyzing many of the top venues, Falls Church was selected as the best location for the new development.

## Conclusions

This project examines location data and information from the Foursquare API to learn about different neighborhoods and cities. The results of this investigation revealed that many of the neighborhoods, cities, and suburbs in the D.C. metro area have similar characteristics and venues that appeal to young, working professionals. However, given the proximity to D.C., available amenities, and high property value, Falls Church is the best location for a new housing development.