Clustering and Comparing the Neighborhoods of Fairfax County in Virginia

Farhad Omar

April 5, 2021

1. Introduction

Fairfax County is a suburb of Washington D.C. in northern Virginia. It is the most populous county in the state and has some of the most expensive housing markets in the DC-Maryland-Virginia (DMV) metropolitan region. Despite the global impact of the coronavirus pandemic, the real estate market in Fairfax County remains strong and competitive. A relatively strong economy mainly drives this competitiveness in the DMV area. Other factors that influence the real estate market are lack of inventory, low mortgage rates, and trim down payment options. However, the real estate market boom may not be uniform across different neighborhoods in the county; therefore, there is an opportunity to use data analysis to compare various localities and generate insights for potential investments.

1.1. Business Problem

Insights and data often assist successful investments. This project will utilize a machine learning technique to derive these insights and provide a mechanism for potential homebuyers or investors to make an informed decision. In particular, the business problem that we are trying to answer is: What neighborhood offers the best opportunity for purchasing a home or an investment property in Fairfax County? To solve this problem, we will cluster various communities in Fairfax County, using local venues and amenities like coffee shops, parks, etc., and a smoothed and seasonally adjusted value of typical homes across the region. We will then compare these clusters' results with the forecasted increase in home prices in each neighborhood to derive valuable insight for purchasing a home or an investment property.

2. Data Acquisition

In this project, we need three types of data to answer our business question. The data that we need include names of neighborhoods, venues and amenities, average home prices, and the forecasted increase in housing prices. The neighborhood dataset consists of the names of various localities in Fairfax County, VA. Using each neighborhood's name, we can obtain their complete address, latitude, and longitude information from the GeoCoder library in Python. We can use the Foursquare API to get relevant information about different venues and amenities in each neighborhood; and housing data from Zillow. To clean, analyze, and process the data, we will utilize a pandas dataframe.

We will load the neighborhood list in Fairfax County from a CSV file into a pandas dataframe. The CSV file was downloaded from https://www.fairfaxcounty.gov/demographics/interactive-map-communities-places-and-towns. Fairfax County offers an interactive map for selecting communities and towns within its borders.

The Foursquare API offers access to the largest sources of locational data, such as venues and amenities. It is a location technology platform. Zillow is the leading real estate and rental marketplace that serves the entire lifecycle of owning a home, such as buying, selling, renting, financing, and remodeling. Zillow makes housing data available for download via the following link: https://www.zillow.com/research/data/. We will utilize the Zillow Home Value Index (ZHVI) and Zillow Home Value Forecast (ZHVF) for this project, where ZHVI is a smoothed,

seasonally adjusted measure of typical home values and market conditions for a given region. Zillow publishes ZHVI for different price tiers, but for this project, we are using the mid-tier dataset, which reflects home values in the 35th to 65th percentile range. ZHVF is the one-year forecast of ZHVI. Both ZHVI and ZHVF files are downloaded as CSV files and will be uploaded into pandas dataframes for post-processing.

2.1. Neighborhood Data

To accomplish this project's objective and analyze investment opportunities in Fairfax County, VA, we need the names of neighborhoods and their geographical coordination. The names of various areas were downloaded from the Fairfax County government website [1]. The dataset was provided in a CSV file. The CSV file contains additional information such as land area, population, total housing units, etc.; we are only interested in the name of neighborhoods in this dataset. After processing and cleaning the dataset, the corresponding geo-coordinates such as latitude and longitude values as well as zip codes for each district were obtained by utilizing the *geolocator* function of *geopy* library in Python. The cleaned dataset was stored in a *pandas* dataframe. Figure 1 shows a snippet of the neighborhood data.

	Neighborhoods	Full_Address	Latitude	Longitude	Zipcode
0	Burke Centre	Burke Centre, Fairfax County, Virginia, 22015-...	38.790992	-77.300519	22015
1	Fair Oaks	Fair Oaks, Fairfax County, Virginia, 22035, Un...	38.863427	-77.359230	22035
2	Crosspointe	Crosspointe, Laurel Hill, Fairfax County, Virg...	38.724002	-77.265078	22079
3	Wakefield	Wakefield Forest, Fairfax County, Virginia, 22...	38.835395	-77.239581	22003
4	McNair	McNair, Fairfax County, Virginia, 20171-3025, ...	38.952295	-77.414295	20171

Figure 1. Various neighborhoods their addresses, geo-coordinate information, and zip codes in Fairfax County, VA

Using the data in Figure 1, we can create a map of Fairfax County's border and various neighborhoods using the Python Folium library. The communities are plotted using latitude and longitude values, and the border is plotted using data from a *geojson* file, providing the geo-coordination of the county's borders [2]. Figure 2 shows the map of Fairfax County and its neighborhoods in our dataset.

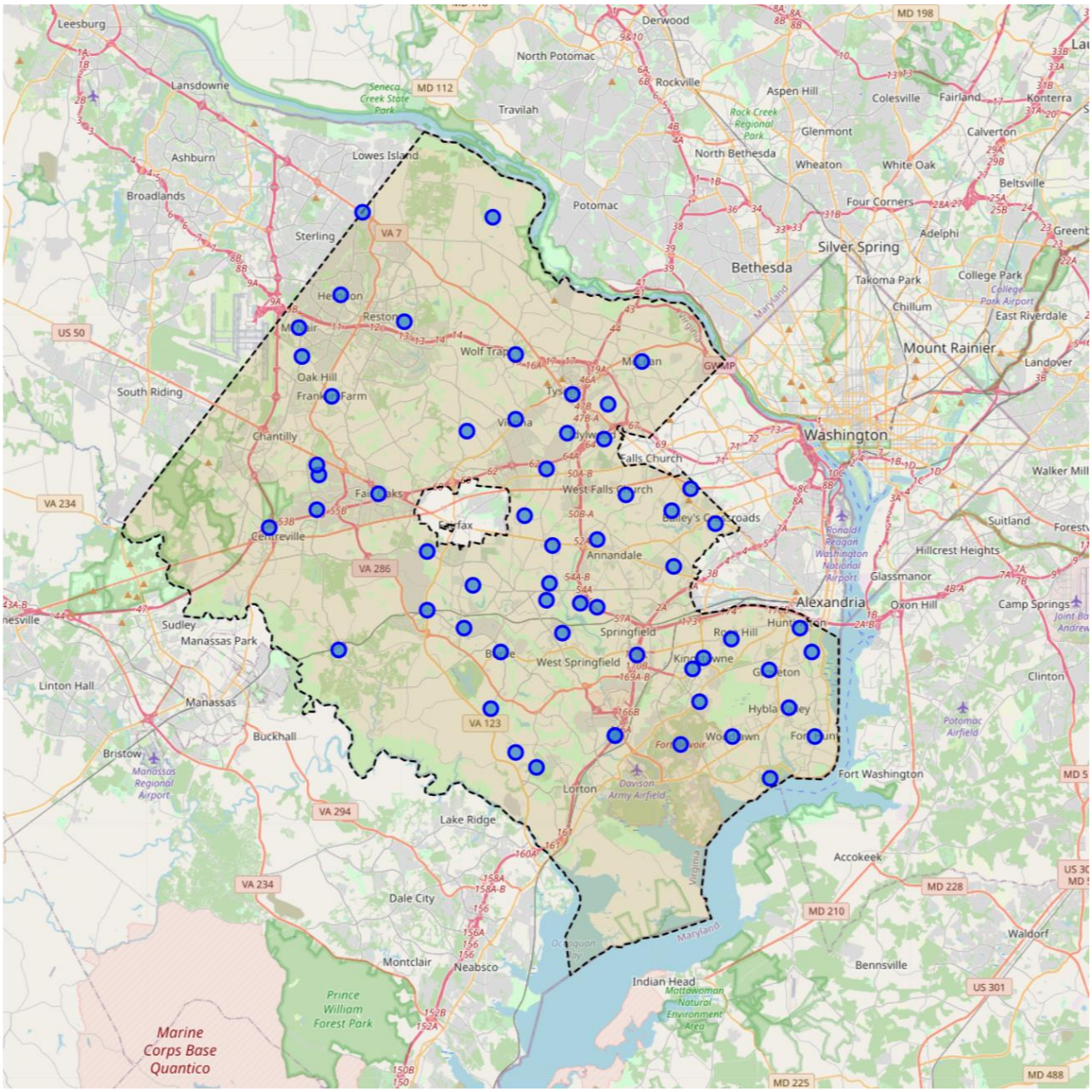

Figure 2. Map of Fairfax County, VA, and its neighborhoods in our dataset

The next task is to obtain housing price data for each neighborhood in the dataset. The housing price information and the forecasted percentage change in the housing prices were downloaded from Zillow [3]. The datasets were provided in CSV files. Both CSV files were read into *pandas* dataframes.

The CSV file for forecasted percentage change in the housing prices contains more than thirty thousand entries, covering various states and counties in the U.S. Figure 3 shows a slice of the forecasted percentage change in the housing prices.

	Region	RegionName	StateName	CountyName	CityName	ForecastedDate	ForecastYoYPctChange
0	Zip	00501	NY	Suffolk County	Holtsville	2022-02-28	5.9
1	Zip	00602	AL	Lee County	Auburn	2022-02-28	4.5
2	Zip	00606	MO	Newton County	Neosho	2022-02-28	6.8
3	Zip	00693	PA	Westmoreland County	Greensburg	2022-02-28	8.2
4	Zip	00705	UT	Kane County	Aibonito	2022-02-28	3.2

Figure 3. Initial dataset for forecasted percentage change in the price of houses in the U.S.

As shown in Figure 3, each locality is designated by its zip code (RegionName). We will utilize these zip codes to extract the forecasted percentage change in housing prices for each neighborhood in Fairfax County, VA. Figure 4 shows various neighborhoods and the corresponding forecasted percentage change in the housing prices in Fairfax County, VA.

	Neighborhoods	Full_Address	Latitude	Longitude	Zipcode	Region	StateName	CountyName	CityName	ForecastedDate	ForecastYoYPctChange
0	Burke Centre	Burke Centre, Fairfax County, Virginia, 22015-...	38.790992	-77.300519	22015	Zip	VA	Fairfax County	Burke	2022-02-28	5.8
1	Crosspointe	Crosspointe, Laurel Hill, Fairfax County, Virg...	38.724002	-77.265078	22079	Zip	VA	Fairfax County	Lorton	2022-02-28	5.8
2	Wakefield	Wakefield Forest, Fairfax County, Virginia, 22...	38.835395	-77.239581	22003	Zip	VA	Fairfax County	Annandale	2022-02-28	5.9
3	McNair	McNair, Fairfax County, Virginia, 20171-3025, ...	38.952295	-77.414295	20171	Zip	VA	Fairfax County	Herndon	2022-02-28	5.7
4	South Run	South Run, Fairfax County, Virginia, 22039, Un...	38.747437	-77.281867	22039	Zip	VA	Fairfax County	Lorton	2022-02-28	5.8

Figure 4. Various neighborhoods and their corresponding forecasted percentage change in the housing prices for Fairfax County

Thus far, we only have the name of neighborhoods, geo-coordinates, and the forecasted percentage change in the housing prices. However, we also need an estimate of the current housing prices in each neighborhood.

The CSV file for the housing prices contains a historical estimated record of housing prices from 1996 to 2021 for various states and counties across the U.S. Using this dataset, we will extract the housing prices for Fairfax County. Figure 5 shows a plot of housing prices from 1996 to 2021 in our dataset.

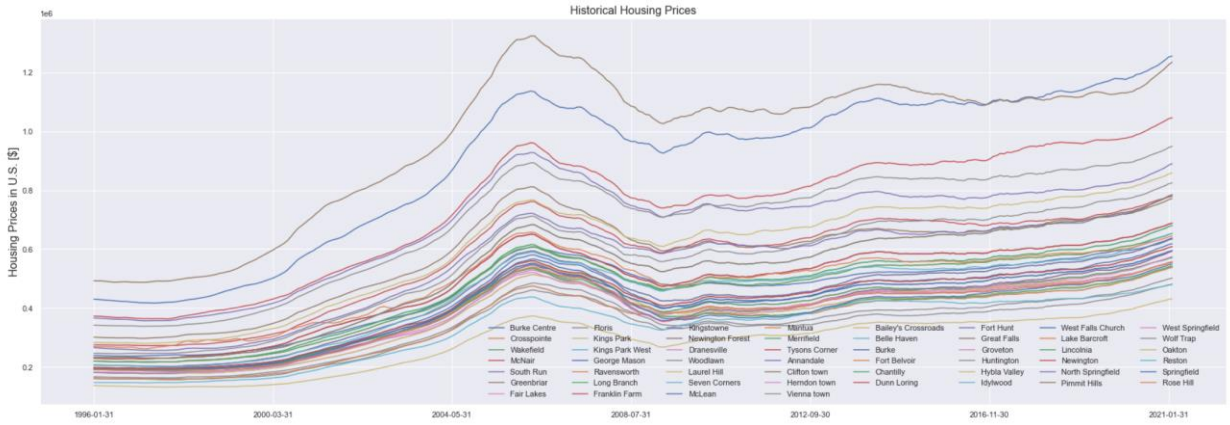

Figure 5. Historical housing prices by Zillow for Fairfax County in the dataset

The housing prices have an overall positive slope; however, notice the price dip due to the recession in 2008 and the recovery period. We will choose the most recent housing price estimate of 2021 to represent an average cost of a typical home in various neighborhoods in Fairfax County for our analysis purposes. Figure 6 shows the complete dataframe for Fairfax County, including a list of neighborhoods, geo-coordinates, zip codes, addresses, forecasted percentage change in the housing prices, and the housing prices. The housing prices are in U.S. dollars.

	Neighborhoods	Full_Address	Latitude	Longitude	Zipcode	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price
0	Burke Centre	Burke Centre, Fairfax County, Virginia, 22015-...	38.790992	-77.300519	22015	VA	Burke	Fairfax County	2022-02-28	5.8	595250.0
1	Crosspointe	Crosspointe, Laurel Hill, Fairfax County, Virg...	38.724002	-77.265078	22079	VA	Lorton	Fairfax County	2022-02-28	5.8	555413.0
2	Wakefield	Wakefield Forest, Fairfax County, Virginia, 22...	38.835395	-77.239581	22003	VA	Annandale	Fairfax County	2022-02-28	5.9	637630.0
3	McNair	McNair, Fairfax County, Virginia, 20171-3025, ...	38.952295	-77.414295	20171	VA	Herndon	Fairfax County	2022-02-28	5.7	688402.0
4	South Run	South Run, Fairfax County, Virginia, 22039, Un...	38.747437	-77.281867	22039	VA	Lorton	Fairfax County	2022-02-28	5.8	890128.0

Figure 6. A complete datafram to represent Fairfax County, including the housing price information

3. Venues Data – Foursquare API

Thus far, we have generated a clean dataframe that contains a list of neighborhoods in Fairfax County, VA. The dataframe includes neighborhood names, complete addresses, geo-coordinates, zip codes, the forecasted percentage change in housing prices, and most recent housing prices based on Zillow's data. Our next task is to use each neighborhood's latitudes and longitudes and the Foursquare API to obtain a list of categories of different venues within a certain radius of each neighborhood center. To retrieve data from Foursquare using their API, we need to create a search query URL. A research query URL can be generated in the following code snippet:

```
LIMIT = 100
radius = 2000
#url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={}&v={},{}&radius={}&limit={}'.format(
# CLIENT_ID, CLIENT_SECRET, latitude, longitude,ACCESS_TOKEN, VERSION, search_query, radius, LIMIT)
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
url
```

,

where CLIENT_ID and CLIENT_SECRET are developers credentials, VERSION is the Foursquare API version (20210404), neighborhood's latitude and longitude of the desired location, the radius is the distance in meters between specified location and venues, and LIMIT is used to restrict the number of venues returned by the API. Figure 7 shows a snippet of the venues returned for one neighborhood in Fairfax County by the Foursquare API in JSON format.

```
results = requests.get(url).json()
results
```


```
{'meta': {'code': 200, 'requestId': '606a0b9c25fa6e064bf8dbea'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'},
    {'name': '$-$$$$', 'key': 'price'}]},
  'headerLocation': 'Tysons Corner',
  'headerFullLocation': 'Tysons Corner',
  'headerLocationGranularity': 'city',
  'totalResults': 175,
  'suggestedBounds': {'ne': {'lat': 38.93450901800002,
    'lng': -77.20269462550311},
   'sw': {'lat': 38.89850898199999, 'lng': -77.24887707835362}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4f626ea7e4b0ea77cba053b8',
       'name': 'CAVA',
       'location': {'address': '8048 Tysons Corner Ctr',
        'lat': 38.9171941,
        'lng': -77.2236294,
        'labeledLatLngs': [{'label': 'display',
          'lat': 38.9171941,
          'lng': -77.2236294}],
        'distance': 201,
        'postalCode': '22102',
        'cc': 'US',
        'city': 'McLean',
        'state': 'VA',
        'country': 'United States',
        'formattedAddress': ['8048 Tysons Corner Ctr',
         'McLean, VA 22102',
         'United States']},
       'categories': [{'id': '4bf58dd8d48988d1c0941735',
         'name': 'Mediterranean Restaurant',
         'pluralName': 'Mediterranean Restaurants',
         'shortName': 'Mediterranean',
         'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/mediterranean_',
          'suffix': '.png'},
         'primary': True}],
       'photos': {'count': 0, 'groups': []}},
      'referralId': 'e-0-4f626ea7e4b0ea77cba053b8-0'},
     {'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4ae3a5e6f964a520a99721e3',
       'name': 'Nordstrom',
       'location': {'address': '8075 Tysons Corner Ctr',
        'lat': 38.9165482,
        'lng': -77.2220314,
        'labeledLatLngs': [{'label': 'display',
          'lat': 38.9165482,
          'lng': -77.2220314}],
        'distance': 325,
        'postalCode': '22102',
        'cc': 'US',
        'city': 'McLean',
        'state': 'VA',
        'country': 'United States',
        'formattedAddress': ['8075 Tysons Corner Ctr',
```


Figure 7. Snippet of data returned by the Foursquare API for a neighborhood in Fairfax County

After processing the Foursquare data into *pandas* dataframe, we can calculate how many venues were returned for each neighborhood as shown in Figure 8. The Foursquare API produced 287 unique categories.

Neighborhoods	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Annandale	85	85	85	85	85	85
Bailey's Crossroads	100	100	100	100	100	100
Belle Haven	92	92	92	92	92	92
Burke	48	48	48	48	48	48
Burke Centre	57	57	57	57	57	57
Chantilly	50	50	50	50	50	50
Clifton town	17	17	17	17	17	17
Crosspointe	33	33	33	33	33	33
Dranesville	47	47	47	47	47	47
Dunn Loring	82	82	82	82	82	82
Fair Lakes	56	56	56	56	56	56
Floris	64	64	64	64	64	64
Fort Belvoir	45	45	45	45	45	45
Fort Hunt	23	23	23	23	23	23
Franklin Farm	33	33	33	33	33	33
George Mason	47	47	47	47	47	47
Great Falls	23	23	23	23	23	23
Greenbriar	84	84	84	84	84	84
Groveton	88	88	88	88	88	88
Herndon town	100	100	100	100	100	100
Huntington	100	100	100	100	100	100
Hybla Valley	78	78	78	78	78	78
Idylwood	81	81	81	81	81	81
Kings Park	30	30	30	30	30	30
Kings Park West	37	37	37	37	37	37
Kingstowne	100	100	100	100	100	100
Lake Barcroft	75	75	75	75	75	75
Laurel Hill	36	36	36	36	36	36
Lincolnia	68	68	68	68	68	68
Long Branch	29	29	29	29	29	29
Mantua	83	83	83	83	83	83
McLean	46	46	46	46	46	46
McNair	97	97	97	97	97	97
Merrifield	100	100	100	100	100	100
Newington	56	56	56	56	56	56
Newington Forest	56	56	56	56	56	56
North Springfield	24	24	24	24	24	24
Oakton	36	36	36	36	36	36
Pimmit Hills	93	93	93	93	93	93
Ravensworth	29	29	29	29	29	29
Reston	100	100	100	100	100	100
Rose Hill	42	42	42	42	42	42
Seven Corners	100	100	100	100	100	100
South Run	17	17	17	17	17	17
Springfield	100	100	100	100	100	100
Tysons Corner	100	100	100	100	100	100
Vienna town	93	93	93	93	93	93
Wakefield	13	13	13	13	13	13
West Falls Church	64	64	64	64	64	64
West Springfield	49	49	49	49	49	49
Wolf Trap	34	34	34	34	34	34
Woodlawn	54	54	54	54	54	54

Figure 8. Venue categories for each neighborhood in Fairfax County returned by Foursquare API

To understand various Foursquare data categories, we can count the frequency of occurrences of venues in each neighborhood and return the ten most common sites. Figure 9 shows the ten most common venues for a few neighborhoods.

	Neighborhoods	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
0	Annandale	Korean Restaurant	Asian Restaurant	Pizza Place	Bakery	Park	Pharmacy	Dessert Shop	Bank	Vietnamese Restaurant	Bubble Tea Shop
1	Bailey's Crossroads	Mexican Restaurant	Department Store	Furniture / Home Store	Latin American Restaurant	American Restaurant	Ethiopian Restaurant	Coffee Shop	Taco Place	Bakery	Thai Restaurant
2	Belle Haven	Pizza Place	Hotel	Pharmacy	American Restaurant	Chinese Restaurant	Donut Shop	Convenience Store	Bank	Bakery	Grocery Store
3	Burke	Pizza Place	Mexican Restaurant	Asian Restaurant	Gym / Fitness Center	Fast Food Restaurant	ATM	Park	Discount Store	Pet Store	Pharmacy
4	Burke Centre	Coffee Shop	Pizza Place	Park	Sandwich Place	Fast Food Restaurant	Big Box Store	Mexican Restaurant	Bank	Gym	Spa

Figure 9. Ten most common venues in different Fairfax neighborhoods

4. Clustering Neighborhoods

In this section, we will cluster various neighborhoods based on their similarity of characteristics or features. Clustering is a mechanism for finding similar items in a dataset. For this project, we will use the *k-means* clustering algorithm of the *Scikit-learn* library in Python. To achieve the objectives of this project, we will perform the following clustering:

1. Cluster neighborhoods based on venue categories;
2. Cluster neighborhoods based on housing prices;
3. Cluster neighborhoods based on the forecasted percentage change in housing prices.

To perform clustering using k-means, we need to remove categorical descriptions from the dataset. The k-means clustering algorithm requires numerical values representing features. We apply one-hot encoding to venue categories, housing prices, and forecasted percentage change in housing prices.

4.1. Clustering Based on Venue Category

After applying one-hot encoding, we group rows by neighborhood and count each category's frequency of occurrence. The result of this operation is shown in Figure 10. For brevity, we only offer a section of the resulting dataframe. As shown in Figure 10, there are four American Restaurants in Belle Haven, for example.

	Neighborhoods	ATM	Accessories Store	Advertising Agency	Afghan Restaurant	American Restaurant	Antique Shop	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	Automotive Shop	BBQ Joint	Bagel Shop	Bakery	Bank	Bar
0	Annandale	0	0	0	0	1	0	0	0	0	6	0	0	1	0	0	4	2	0
1	Bailey's Crossroads	0	0	0	1	3	0	0	0	1	1	0	0	0	0	1	2	2	0
2	Belle Haven	0	0	0	0	4	1	0	0	0	1	0	0	0	1	0	3	3	0
3	Burke	1	0	0	0	1	0	0	0	0	2	0	0	0	0	0	1	0	0
4	Burke Centre	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	2	0

Figure 10. Grouping rows by neighborhood and count the frequency of occurrence of each category

The result of applying a k-means clustering algorithm based on venue category is shown in Figure 11.

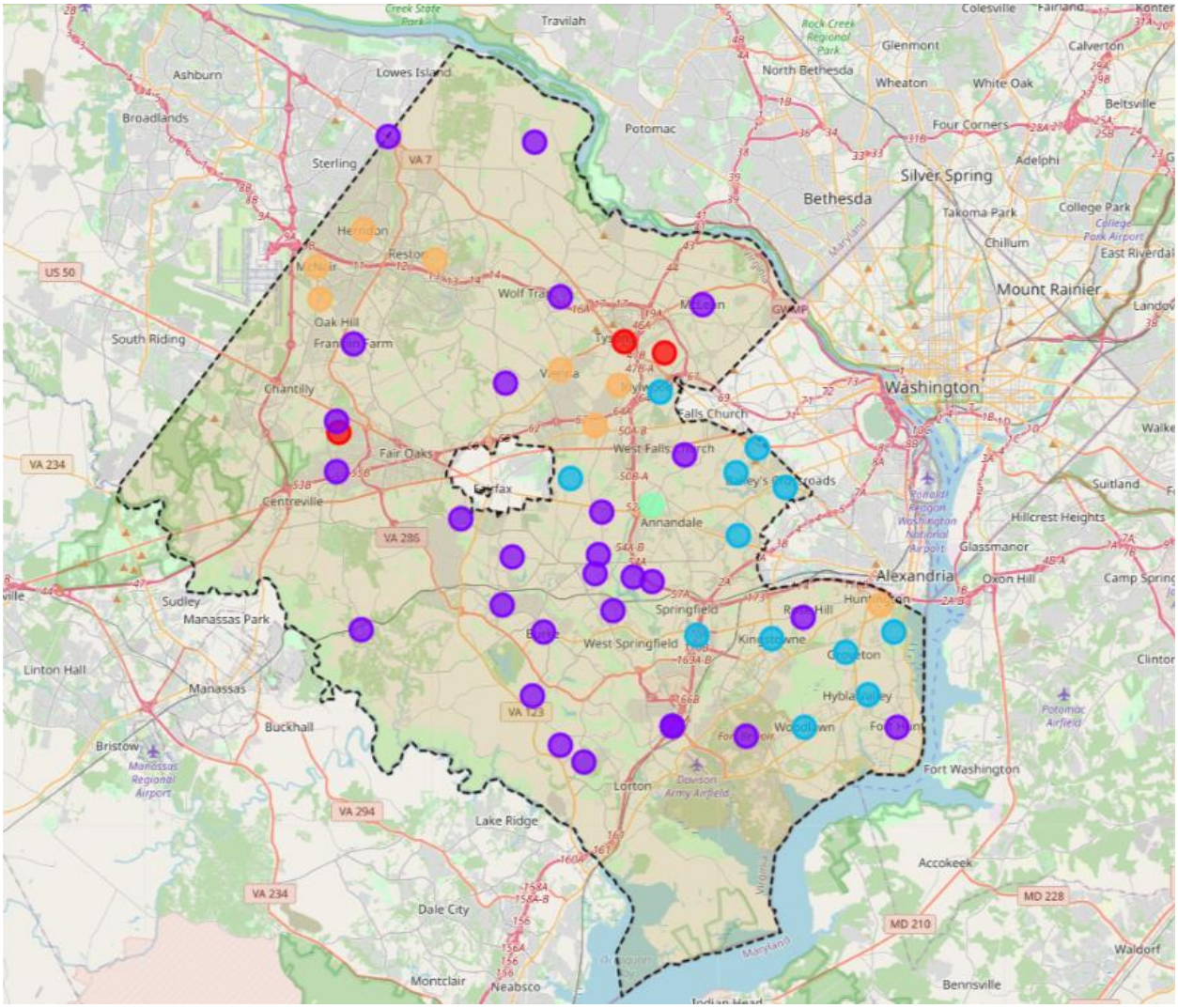

Figure 11. Clustering neighborhoods based on venue category

4.2. Clustering Based on House Price

To perform k-means clustering based on housing prices, we extracted a subsection of the dataframe as shown in Figure 12.

	Neighborhoods	Latitude	Longitude	House_Price	ForecastYoYPctChange
0	Burke Centre	38.790992	-77.300519	595250.0	5.8
1	Crosspointe	38.724002	-77.265078	555413.0	5.8
2	Wakefield	38.835395	-77.239581	637630.0	5.9
3	McNair	38.952295	-77.414295	688402.0	5.7
4	South Run	38.747437	-77.281867	890128.0	5.8

Figure 12. A snippet of dataframe for k-means clustering based on housing prices

As before, we applied one-hot encoding to our dataframe and used the k-means clustering algorithm. Notice that we did not group our data based on frequency occurrences because each neighborhood has one value for the price, representing the cost of a typical home in that neighborhood. The result of applying the k-means clustering algorithm based on venue category is shown in Figure 13.

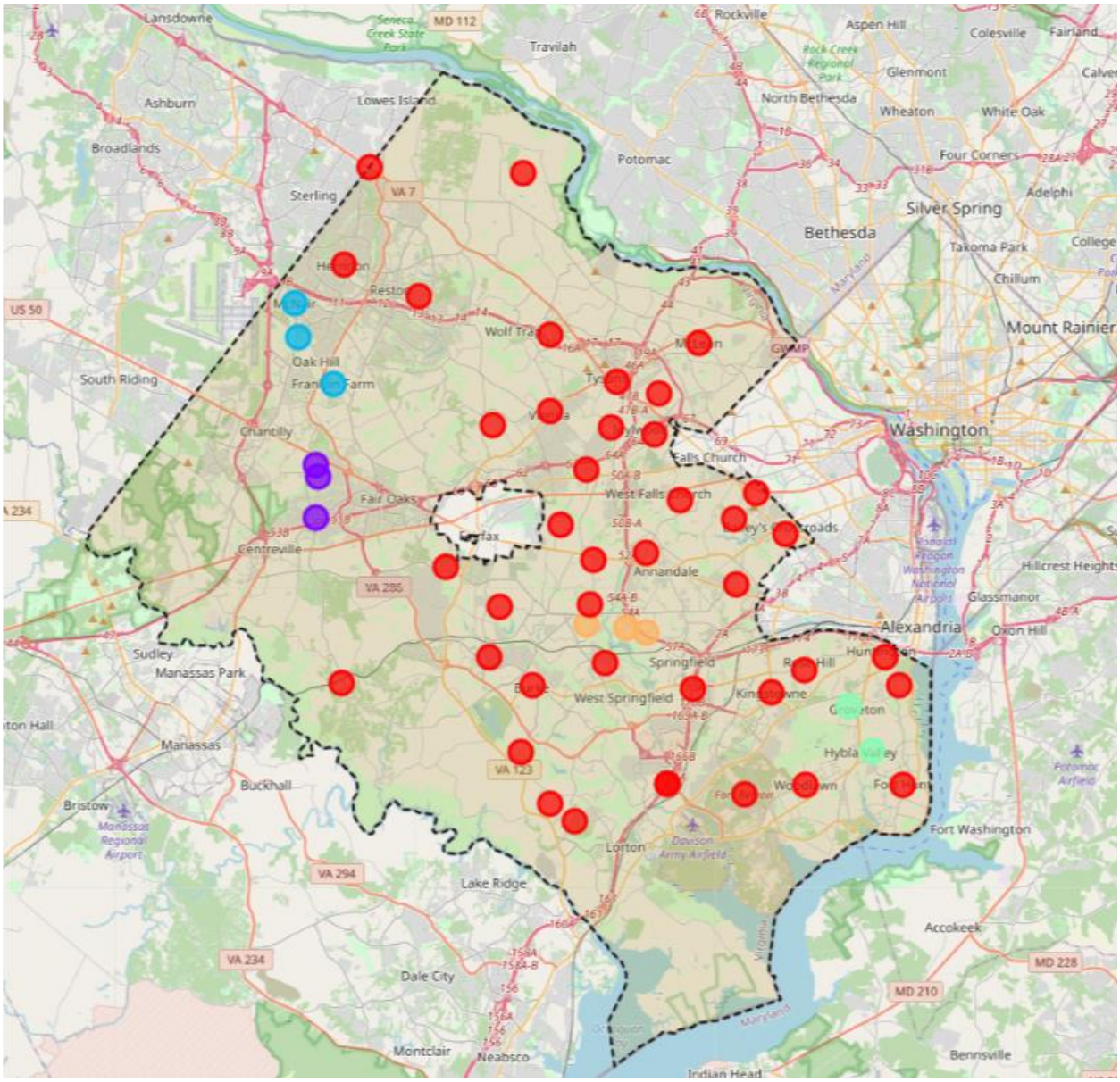

Figure 13. Clustering neighborhoods based on housing prices

4.3. Clustering Based on the Forecasted Percentage Change in Housing Prices

To perform k-means clustering based on the forecasted percentage change in housing prices, we extracted a subsection of the dataframe as shown in Figure 14.

	Neighborhoods	Latitude	Longitude	House_Price	ForecastYoYPctChange
0	Burke Centre	38.790992	-77.300519	595250.0	5.8
1	Crosspointe	38.724002	-77.265078	555413.0	5.8
2	Wakefield	38.835395	-77.239581	637630.0	5.9
3	McNair	38.952295	-77.414295	688402.0	5.7
4	South Run	38.747437	-77.281867	890128.0	5.8

Figure 14. A snippet of dataframe for k-means clustering based on the forecasted percentage change in housing prices

As before, we applied one-hot encoding to our dataframe and used the k-means clustering algorithm. Notice that we did not group our data based on frequency occurrences because each neighborhood has one value for forecasted percentage increase in housing prices. The result of applying the k-means clustering algorithm based on venue category is shown in Figure 15.

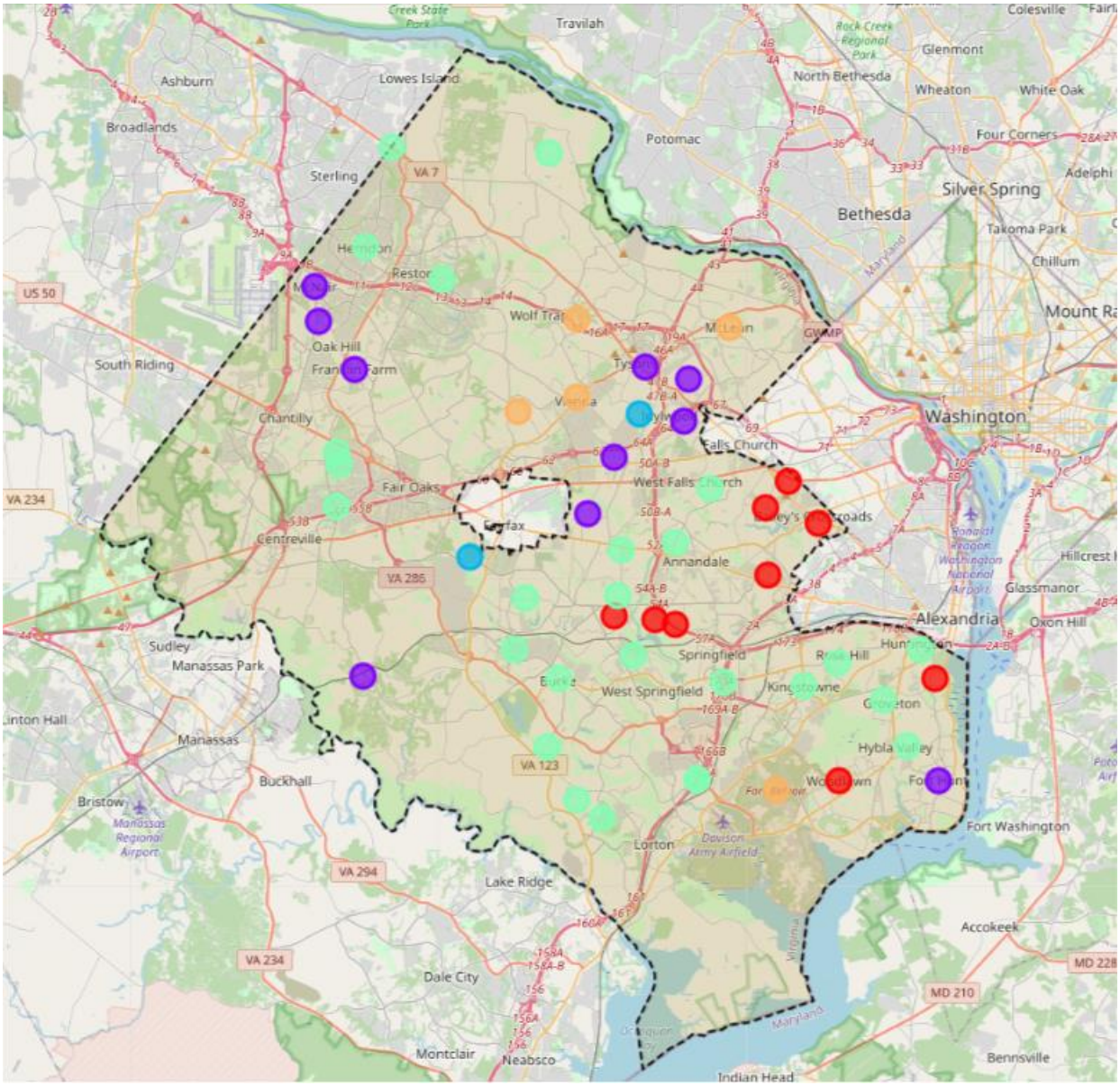

Figure 15. Clustering neighborhoods based on the forecasted percentage change in housing prices

5. Analyzing the Results for Each Clustering Approach

This section will report the results of applying the k-means clustering algorithm based on venue category, housing prices, and forecasted percentage change in housing price features. Our object is to analyze these clustering approaches for investment opportunities in different neighborhoods.

5.1. Cluster 1 – Venue Category

In this group, only three neighborhoods with similar venues are clustered together. The range of housing prices is between $57000 to $780000. The forecasted percentage change in housing prices has relatively identical magnitudes. Still, the results in Figure 16 suggest that we should purchase a house in the Greenbriar neighborhood if we want to maximize our return on investment.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
5	Greenbriar	VA	Fairfax	Fairfax County	2022-02-28	5.8	571477.0	0
23	Tysons Corner	VA	McLean	Fairfax County	2022-02-28	5.7	779114.0	0
45	Pimmit Hills	VA	Falls Church	Fairfax County	2022-02-28	5.6	783622.0	0

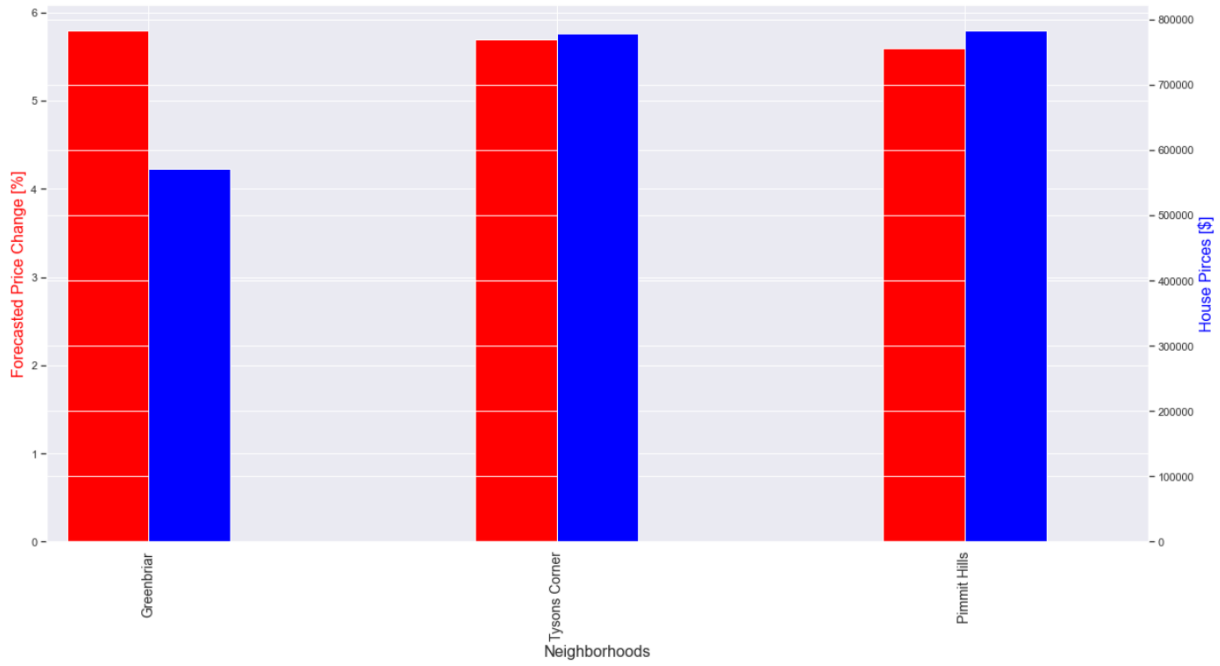

Figure 16. Cluster 1-Venue Category

5.2. Cluster 2 – Venue Category

In this group, twenty-eight neighborhoods with similar venues are clustered together. The range of housing prices is between $550000 to $1.2 million, and the range forecasted percentage change in housing prices is 5.2 % to 6.0 %. The results in Figure 17 Figure 16suggest that most neighborhoods provide us an opportunity to gain on our investments. The best locations for purchasing an investment property are Chantilly, Rose Hill, West Springfield, and Laurel Hill neighborhoods. However, there are several neighborhoods that we should avoid investing. According to the data in Figure 17, we should avoid investing in neighborhoods such as McLean, Wolf Trap, Oakton, Great Falls, and South Run.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
10	George Mason	VA	Fairfax	Fairfax County	2022-02-28	5.2	634644.0	1
20	McLean	VA	McLean	Fairfax County	2022-02-28	5.3	1255540.0	1
47	Wolf Trap	VA	Vienna	Fairfax County	2022-02-28	5.4	949536.0	1
31	Fort Belvoir	VA	Fort Belvoir	Fairfax County	2022-02-28	5.4	644641.0	1
48	Oakton	VA	Vienna	Fairfax County	2022-02-28	5.5	858957.0	1
25	Clifton town	VA	Clifton	Fairfax County	2022-02-28	5.6	771612.0	1
34	Fort Hunt	VA	Fort Hunt	Fairfax County	2022-02-28	5.7	784641.0	1
13	Franklin Farm	VA	Herndon	Fairfax County	2022-02-28	5.7	688402.0	1
0	Burke Centre	VA	Burke	Fairfax County	2022-02-28	5.8	595250.0	1
46	West Springfield	VA	Springfield	Fairfax County	2022-02-28	5.8	574793.0	1
35	Great Falls	VA	Great Falls	Fairfax County	2022-02-28	5.8	1234994.0	1
32	Chantilly	VA	Fairfax	Fairfax County	2022-02-28	5.8	571477.0	1
30	Burke	VA	Burke	Fairfax County	2022-02-28	5.8	595250.0	1
18	Laurel Hill	VA	Lorton	Fairfax County	2022-02-28	5.8	555413.0	1
6	Fair Lakes	VA	Fairfax	Fairfax County	2022-02-28	5.8	571477.0	1
4	South Run	VA	Lorton	Fairfax County	2022-02-28	5.8	890128.0	1
1	Crosspointe	VA	Lorton	Fairfax County	2022-02-28	5.8	555413.0	1
16	Dranesville	VA	Herndon	Fairfax County	2022-02-28	5.8	557876.0	1
15	Newington Forest	VA	Springfield	Fairfax County	2022-02-28	5.9	608921.0	1
12	Long Branch	VA	Fairfax	Fairfax County	2022-02-28	5.9	680395.0	1
9	Kings Park West	VA	Fairfax	Fairfax County	2022-02-28	5.9	680395.0	1
40	West Falls Church	VA	Falls Church	Fairfax County	2022-02-28	5.9	618723.0	1
43	Newington	VA	Springfield	Fairfax County	2022-02-28	5.9	608921.0	1
2	Wakefield	VA	Annandale	Fairfax County	2022-02-28	5.9	637630.0	1
51	Rose Hill	VA	Rose Hill	Fairfax County	2022-02-28	5.9	555972.0	1
11	Ravensworth	VA	Springfield	Fairfax County	2022-02-28	6.0	593580.0	1
8	Kings Park	VA	Springfield	Fairfax County	2022-02-28	6.0	593580.0	1
44	North Springfield	VA	Springfield	Fairfax County	2022-02-28	6.0	593580.0	1

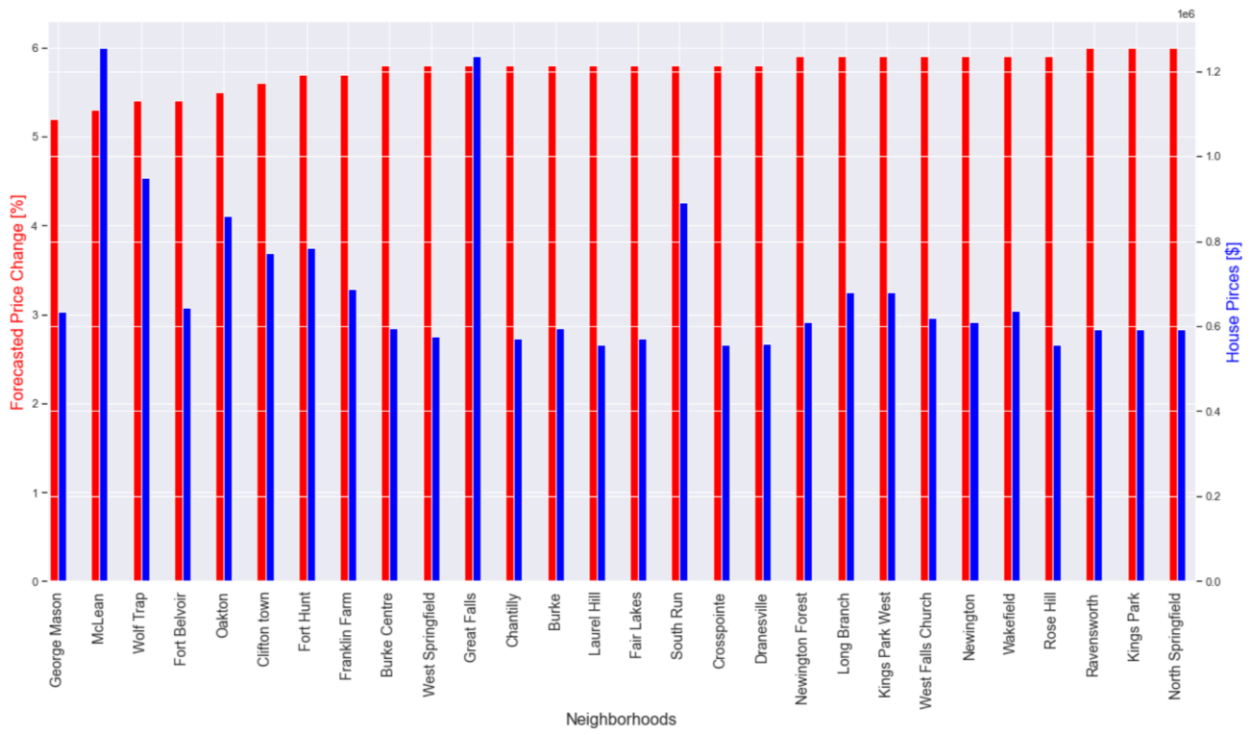

Figure 17. Cluster2 - Venue Category

5.3. Cluster 3 – Venue Category

In this group, twelve neighborhoods with similar venues are clustered together. The range of housing prices is between $430000 to $780000, and the range of forecasted percentage change in housing prices is 5.6 % to 6.1 %. The results in Figure 18 Figure 16suggest that the best place to invest in is Bailey's Crossroads. Even though the forecasted percentage change in the price of homes has similar values in most neighborhoods, the initial investment in Bailey's Crossroads is less than all other locations.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
21	Mantua	VA	Fairfax	Fairfax County	2022-02-28	5.6	653093.0	2
39	Idylwood	VA	Falls Church	Fairfax County	2022-02-28	5.6	783622.0	2
36	Groveton	VA	Hybla Valley	Fairfax County	2022-02-28	5.8	543755.0	2
38	Hybla Valley	VA	Hybla Valley	Fairfax County	2022-02-28	5.8	543755.0	2
50	Springfield	VA	Springfield	Fairfax County	2022-02-28	5.8	549804.0	2
14	Kingstowne	VA	Rose Hill	Fairfax County	2022-02-28	5.9	555972.0	2
42	Lincolnia	VA	Lincolnia	Fairfax County	2022-02-28	6.0	541869.0	2
17	Woodlawn	VA	Mount Vernon	Fairfax County	2022-02-28	6.1	501693.0	2
19	Seven Corners	VA	Falls Church	Fairfax County	2022-02-28	6.1	538331.0	2
28	Bailey's Crossroads	VA	Falls Church	Fairfax County	2022-02-28	6.1	431954.0	2
29	Belle Haven	VA	Belle Haven	Fairfax County	2022-02-28	6.1	653496.0	2
41	Lake Barcroft	VA	Falls Church	Fairfax County	2022-02-28	6.1	538331.0	2

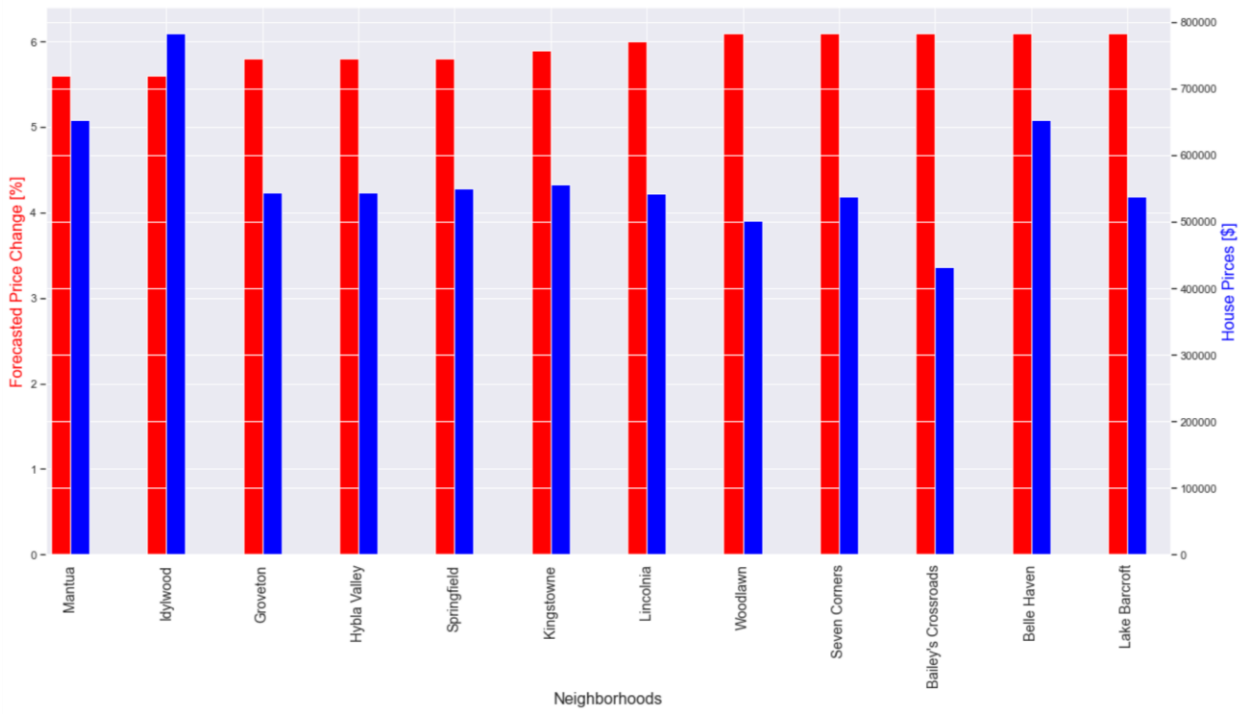

Figure 18. Cluster3 - Venue Category

5.4. Cluster 4 – Venue Category

Cluster 4 only has one neighborhood with a 5.9 % potential increase in investment.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
24	Annandale	VA	Annandale	Fairfax County	2022-02-28	5.9	637630.0	3

5.5. Cluster 5 – Venue Category

In this group, eight neighborhoods with similar venues are clustered together. The range of housing prices is between $480000 to just over 1 million dollars, and the range of forecasted percentage change in housing prices is 5.1 % to 5.8 %. The results in Figure 19 Figure 16suggest that the best places to invest are Huntington and Reston neighborhoods.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
33	Dunn Loring	VA	Dunn Loring	Fairfax County	2022-02-28	5.1	1046272.0	4
27	Vienna town	VA	Vienna	Fairfax County	2022-02-28	5.4	824915.0	4
22	Merrifield	VA	Fairfax	Fairfax County	2022-02-28	5.6	653093.0	4
3	McNair	VA	Herndon	Fairfax County	2022-02-28	5.7	688402.0	4
7	Floris	VA	Herndon	Fairfax County	2022-02-28	5.7	688402.0	4
26	Herndon town	VA	Herndon	Fairfax County	2022-02-28	5.8	557876.0	4
37	Huntington	VA	Huntington	Fairfax County	2022-02-28	5.8	481086.0	4
49	Reston	VA	Reston	Fairfax County	2022-02-28	5.8	479682.0	4

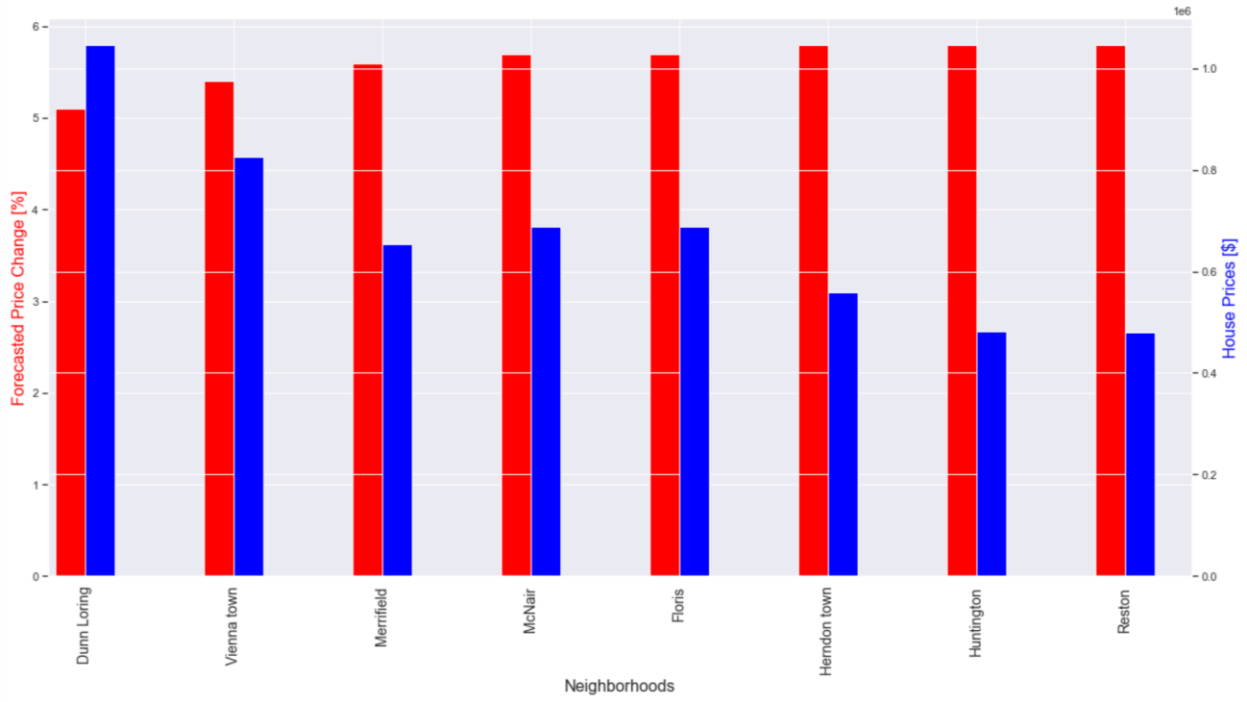

Figure 19. Cluster 5 - Venue Category

5.6. Cluster 1 – House Price

In this group, forty-one neighborhoods with similar housing prices were clustered together. The range of housing prices is between $430000 to $1.3 million, and the range of the forecasted percentage change in housing prices is between 5.1 % to 6.1 %. The results in Figure 20 Figure 16suggest that the best places to invest are Bailey's Crossroads, and Woodlawn followed by Reston and Huntington neighborhoods.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
33	Dunn Loring	VA	Dunn Loring	Fairfax County	2022-02-28	5.1	1046272.0	0
10	George Mason	VA	Fairfax	Fairfax County	2022-02-28	5.2	634644.0	0
20	McLean	VA	McLean	Fairfax County	2022-02-28	5.3	1255540.0	0
27	Vienna town	VA	Vienna	Fairfax County	2022-02-28	5.4	824915.0	0
47	Wolf Trap	VA	Vienna	Fairfax County	2022-02-28	5.4	949536.0	0
31	Fort Belvoir	VA	Fort Belvoir	Fairfax County	2022-02-28	5.4	644641.0	0
48	Oakton	VA	Vienna	Fairfax County	2022-02-28	5.5	858957.0	0
45	Pimmit Hills	VA	Falls Church	Fairfax County	2022-02-28	5.6	783622.0	0
39	Idylwood	VA	Falls Church	Fairfax County	2022-02-28	5.6	783622.0	0
22	Merrifield	VA	Fairfax	Fairfax County	2022-02-28	5.6	653093.0	0
25	Clifton town	VA	Clifton	Fairfax County	2022-02-28	5.6	771612.0	0
21	Mantua	VA	Fairfax	Fairfax County	2022-02-28	5.6	653093.0	0
23	Tysons Corner	VA	McLean	Fairfax County	2022-02-28	5.7	779114.0	0
34	Fort Hunt	VA	Fort Hunt	Fairfax County	2022-02-28	5.7	784641.0	0
30	Burke	VA	Burke	Fairfax County	2022-02-28	5.8	595250.0	0
35	Great Falls	VA	Great Falls	Fairfax County	2022-02-28	5.8	1234994.0	0
46	West Springfield	VA	Springfield	Fairfax County	2022-02-28	5.8	574793.0	0
49	Reston	VA	Reston	Fairfax County	2022-02-28	5.8	479682.0	0
37	Huntington	VA	Huntington	Fairfax County	2022-02-28	5.8	481086.0	0
50	Springfield	VA	Springfield	Fairfax County	2022-02-28	5.8	549804.0	0
0	Burke Centre	VA	Burke	Fairfax County	2022-02-28	5.8	595250.0	0
1	Crosspointe	VA	Lorton	Fairfax County	2022-02-28	5.8	555413.0	0
4	South Run	VA	Lorton	Fairfax County	2022-02-28	5.8	890128.0	0
18	Laurel Hill	VA	Lorton	Fairfax County	2022-02-28	5.8	555413.0	0
16	Dranesville	VA	Herndon	Fairfax County	2022-02-28	5.8	557876.0	0
26	Herndon town	VA	Herndon	Fairfax County	2022-02-28	5.8	557876.0	0
40	West Falls Church	VA	Falls Church	Fairfax County	2022-02-28	5.9	618723.0	0
2	Wakefield	VA	Annandale	Fairfax County	2022-02-28	5.9	637630.0	0
9	Kings Park West	VA	Fairfax	Fairfax County	2022-02-28	5.9	680395.0	0
12	Long Branch	VA	Fairfax	Fairfax County	2022-02-28	5.9	680395.0	0
43	Newington	VA	Springfield	Fairfax County	2022-02-28	5.9	608921.0	0
51	Rose Hill	VA	Rose Hill	Fairfax County	2022-02-28	5.9	555972.0	0
14	Kingstowne	VA	Rose Hill	Fairfax County	2022-02-28	5.9	555972.0	0
15	Newington Forest	VA	Springfield	Fairfax County	2022-02-28	5.9	608921.0	0
24	Annandale	VA	Annandale	Fairfax County	2022-02-28	5.9	637630.0	0
42	Lincolnia	VA	Lincolnia	Fairfax County	2022-02-28	6.0	541869.0	0
17	Woodlawn	VA	Mount Vernon	Fairfax County	2022-02-28	6.1	501693.0	0
19	Seven Corners	VA	Falls Church	Fairfax County	2022-02-28	6.1	538331.0	0
29	Belle Haven	VA	Belle Haven	Fairfax County	2022-02-28	6.1	653496.0	0
41	Lake Barcroft	VA	Falls Church	Fairfax County	2022-02-28	6.1	538331.0	0
28	Bailey's Crossroads	VA	Falls Church	Fairfax County	2022-02-28	6.1	431954.0	0

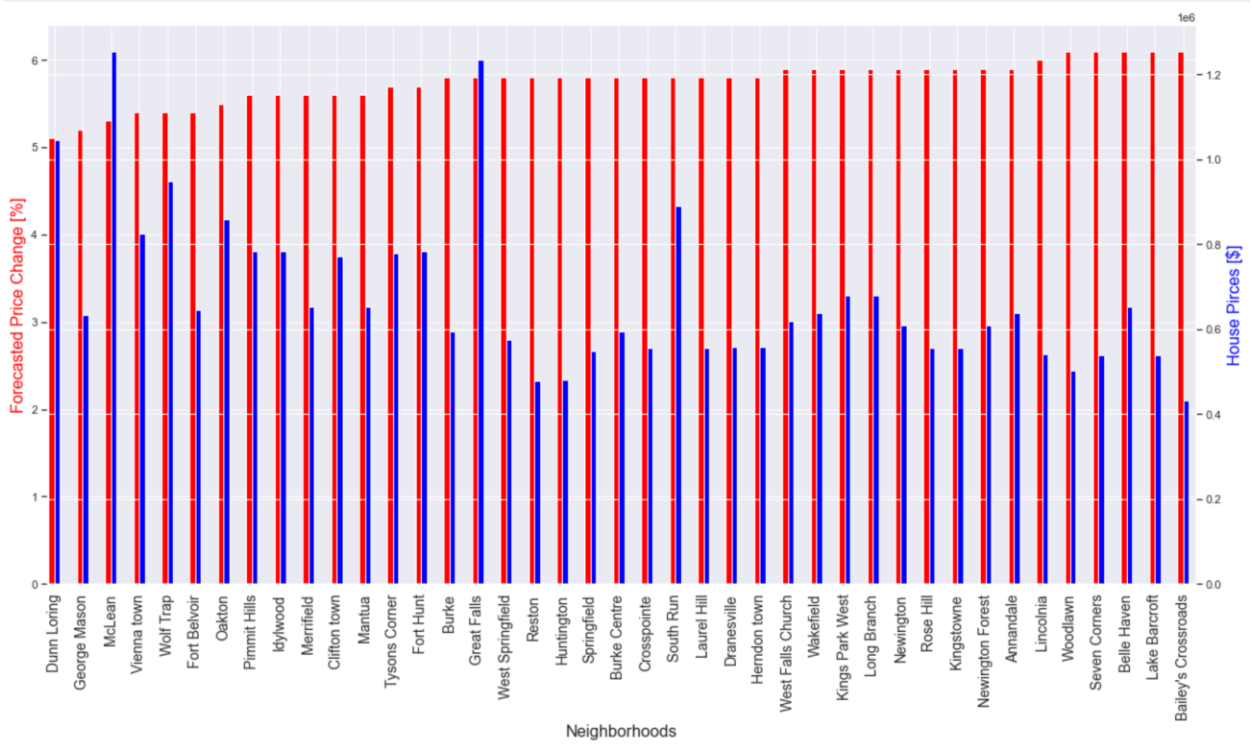

Figure 20. Cluster 1 - House Price

5.7. Cluster 2 – House Price

In this group, three neighborhoods with similar house prices were clustered together. All three communities could offer a great investment opportunity because the cost of initial investment is lower. The results in Figure 21 Figure 16confirm this assertion.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
5	Greenbriar	VA	Fairfax	Fairfax County	2022-02-28	5.8	571477.0	1
6	Fair Lakes	VA	Fairfax	Fairfax County	2022-02-28	5.8	571477.0	1
32	Chantilly	VA	Fairfax	Fairfax County	2022-02-28	5.8	571477.0	1

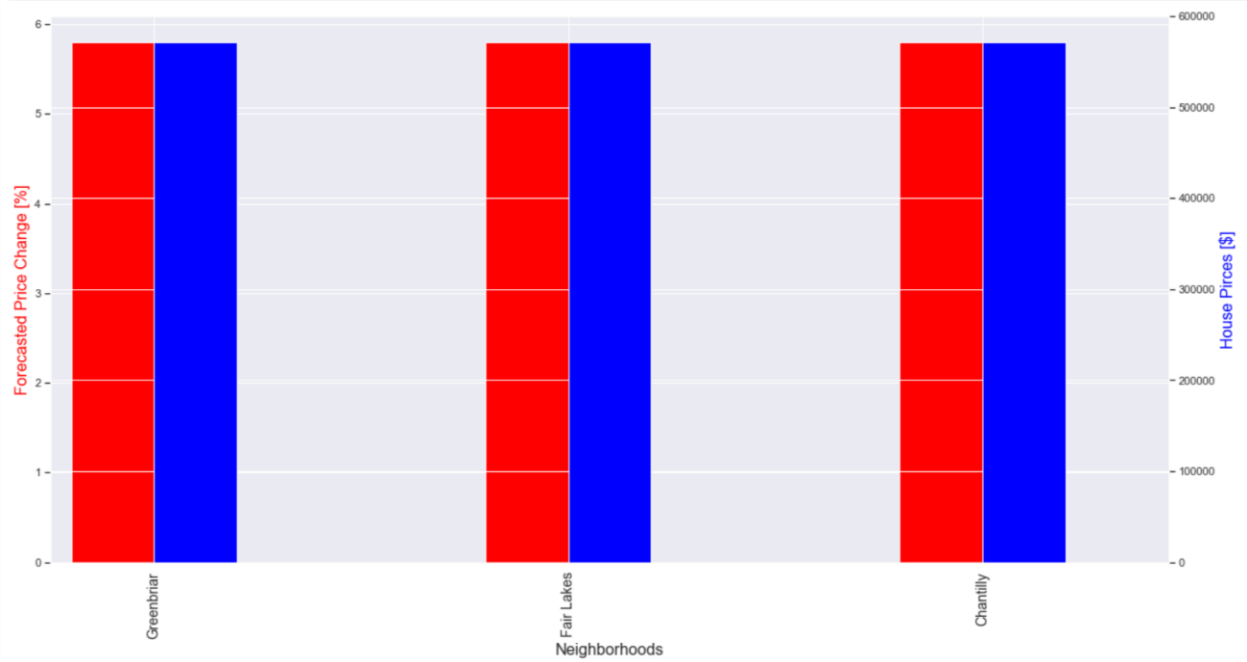

Figure 21. Cluster 2 - House Price

5.8. Cluster 3 – House Price

In this group, three neighborhoods with similar house prices were clustered together. All three areas could offer a good investment opportunity because the cost of initial investment is lower. The results in Figure 22 Figure 16confirm this assertion.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
3	McNair	VA	Herndon	Fairfax County	2022-02-28	5.7	688402.0	2
7	Floris	VA	Herndon	Fairfax County	2022-02-28	5.7	688402.0	2
13	Franklin Farm	VA	Herndon	Fairfax County	2022-02-28	5.7	688402.0	2

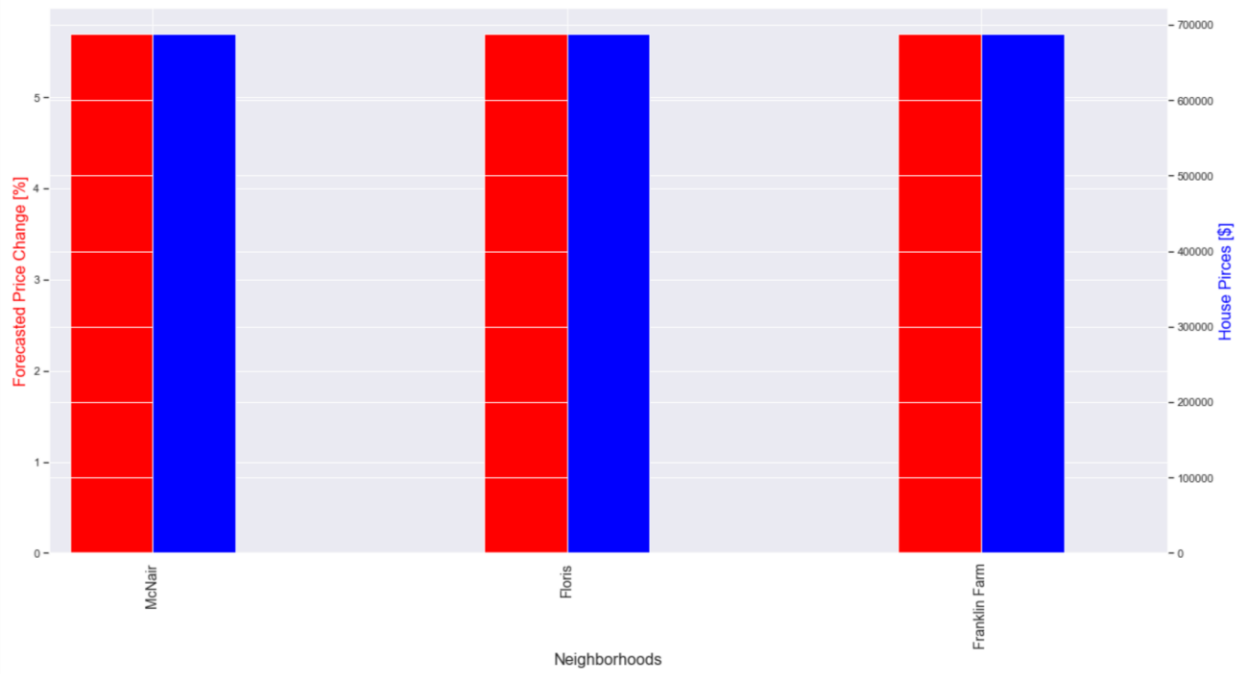

Figure 22. Cluster 3 - House Price

5.9. Cluster 4 – House Price

In this group, two neighborhoods with similar house prices were clustered together. Both communities offer a great investment opportunity because the cost of initial investment is lower.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
36	Groveton	VA	Hybla Valley	Fairfax County	2022-02-28	5.8	543755.0	3
38	Hybla Valley	VA	Hybla Valley	Fairfax County	2022-02-28	5.8	543755.0	3

5.10. Cluster 5 – House Price

In this group, three neighborhoods with similar house prices were clustered together. All three communities could offer a good investment opportunity because the initial investment cost is lower, and the forecasted percent increase in homes' values is higher.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
8	Kings Park	VA	Springfield	Fairfax County	2022-02-28	6.0	593580.0	4
11	Ravensworth	VA	Springfield	Fairfax County	2022-02-28	6.0	593580.0	4
44	North Springfield	VA	Springfield	Fairfax County	2022-02-28	6.0	593580.0	4

5.11. Cluster 1 – Forecasted Change in Price

In this group, nine neighborhoods with the similar forecasted percentage change in housing prices were clustered together. Again, Bailey's Crossroads offers the best investment opportunity, followed by Woodlawn. The results in Figure 23 Figure 16confirm this assertion.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
8	Kings Park	VA	Springfield	Fairfax County	2022-02-28	6.0	593580.0	0
11	Ravensworth	VA	Springfield	Fairfax County	2022-02-28	6.0	593580.0	0
42	Lincolnia	VA	Lincolnia	Fairfax County	2022-02-28	6.0	541869.0	0
44	North Springfield	VA	Springfield	Fairfax County	2022-02-28	6.0	593580.0	0
17	Woodlawn	VA	Mount Vernon	Fairfax County	2022-02-28	6.1	501693.0	0
19	Seven Corners	VA	Falls Church	Fairfax County	2022-02-28	6.1	538331.0	0
28	Bailey's Crossroads	VA	Falls Church	Fairfax County	2022-02-28	6.1	431954.0	0
29	Belle Haven	VA	Belle Haven	Fairfax County	2022-02-28	6.1	653496.0	0
41	Lake Barcroft	VA	Falls Church	Fairfax County	2022-02-28	6.1	538331.0	0

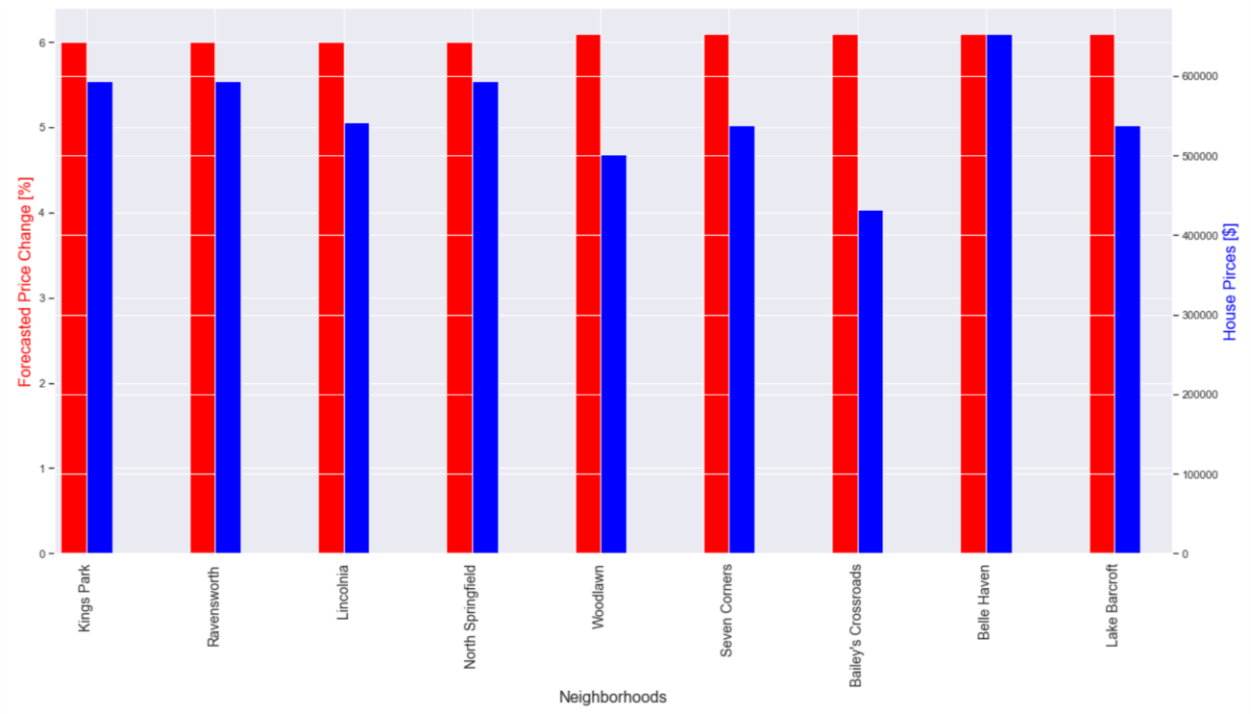

Figure 23. Cluster 1 - Forecasted Change in Price

5.12. Cluster 2 – Forecasted Change in Price

In this group, ten neighborhoods with the similar forecasted percentage change in housing prices were clustered together. None of the locations offer a good investment opportunity because the initial cost is high. The results in Figure 24 Figure 16confirm this assertion.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
21	Mantua	VA	Fairfax	Fairfax County	2022-02-28	5.6	653093.0	1
22	Merrifield	VA	Fairfax	Fairfax County	2022-02-28	5.6	653093.0	1
25	Clifton town	VA	Clifton	Fairfax County	2022-02-28	5.6	771612.0	1
39	Idylwood	VA	Falls Church	Fairfax County	2022-02-28	5.6	783622.0	1
45	Pimmit Hills	VA	Falls Church	Fairfax County	2022-02-28	5.6	783622.0	1
3	McNair	VA	Herndon	Fairfax County	2022-02-28	5.7	688402.0	1
7	Floris	VA	Herndon	Fairfax County	2022-02-28	5.7	688402.0	1
13	Franklin Farm	VA	Herndon	Fairfax County	2022-02-28	5.7	688402.0	1
23	Tysons Corner	VA	McLean	Fairfax County	2022-02-28	5.7	779114.0	1
34	Fort Hunt	VA	Fort Hunt	Fairfax County	2022-02-28	5.7	784641.0	1

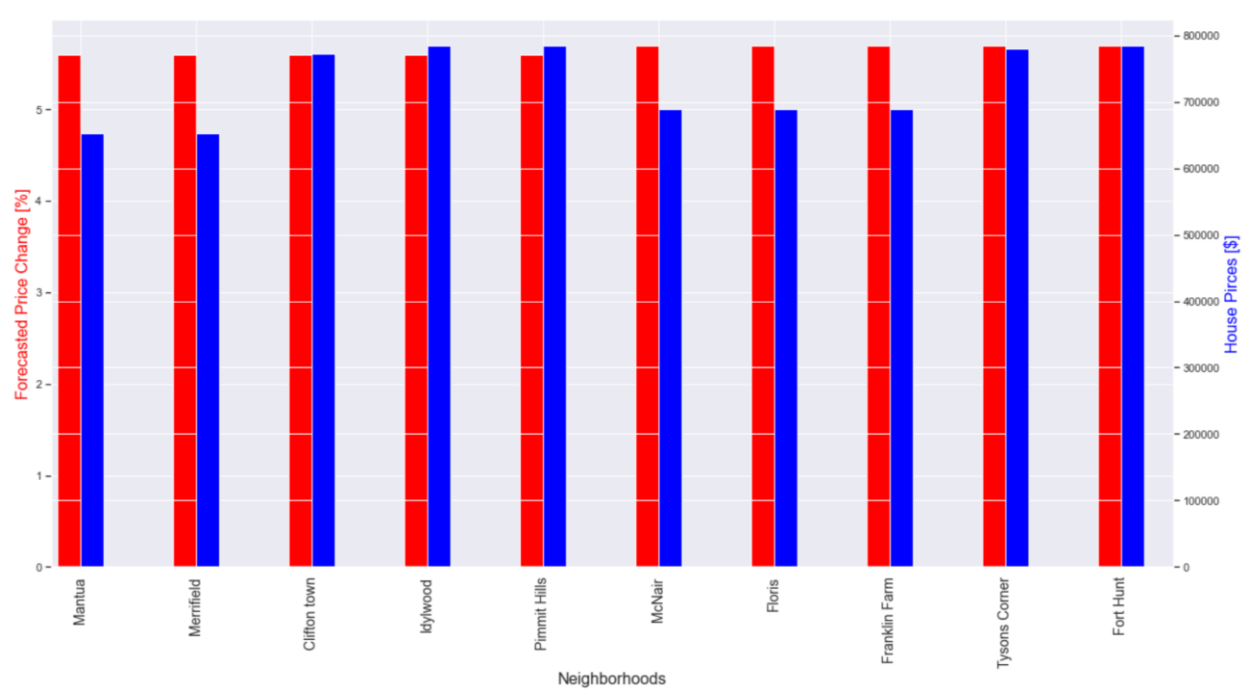

Figure 24. Cluster 2 - Forecasted Change in Price

5.13. Cluster 3 – Forecasted Change in Price

In this group, only two neighborhoods with the similar forecasted percentage change in housing prices were clustered together. However, the initial cost of investment in each neighborhood is significantly different. Based on the investment cost, the George Mason location offers the best opportunity for making a profit. The results in Figure 25 Figure 16confirm this assertion.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
33	Dunn Loring	VA	Dunn Loring	Fairfax County	2022-02-28	5.1	1046272.0	2
10	George Mason	VA	Fairfax	Fairfax County	2022-02-28	5.2	634644.0	2

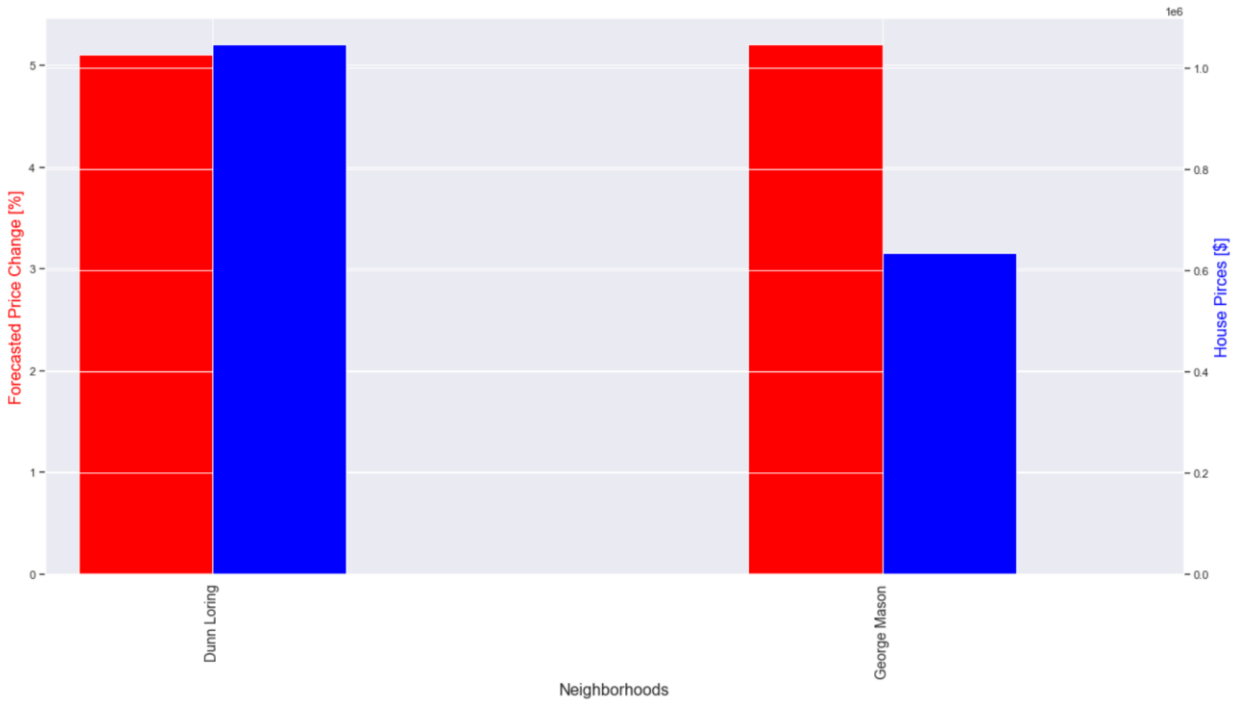

Figure 25. Cluster 3 - Forecasted Change in Price

5.14. Cluster 4 – Forecasted Change in Price

In this group, twenty-six neighborhoods with the similar forecasted percentage change in housing prices were clustered together. Most neighborhoods offer a good investment opportunity due to lower initial costs and a more significant increase in the forecasted percentage change of housing prices. The best location for investment is Reston, followed by the Huntington neighborhood, and the worst areas for investment are Great Falls and South Run. The results in Figure 26 Figure 16confirm this assertion.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
0	Burke Centre	VA	Burke	Fairfax County	2022-02-28	5.8	595250.0	3
35	Great Falls	VA	Great Falls	Fairfax County	2022-02-28	5.8	1234994.0	3
32	Chantilly	VA	Fairfax	Fairfax County	2022-02-28	5.8	571477.0	3
30	Burke	VA	Burke	Fairfax County	2022-02-28	5.8	595250.0	3
26	Herndon town	VA	Herndon	Fairfax County	2022-02-28	5.8	557876.0	3
50	Springfield	VA	Springfield	Fairfax County	2022-02-28	5.8	549804.0	3
18	Laurel Hill	VA	Lorton	Fairfax County	2022-02-28	5.8	555413.0	3
16	Dranesville	VA	Herndon	Fairfax County	2022-02-28	5.8	557876.0	3
36	Groveton	VA	Hybla Valley	Fairfax County	2022-02-28	5.8	543755.0	3
38	Hybla Valley	VA	Hybla Valley	Fairfax County	2022-02-28	5.8	543755.0	3
46	West Springfield	VA	Springfield	Fairfax County	2022-02-28	5.8	574793.0	3
6	Fair Lakes	VA	Fairfax	Fairfax County	2022-02-28	5.8	571477.0	3
5	Greenbriar	VA	Fairfax	Fairfax County	2022-02-28	5.8	571477.0	3
4	South Run	VA	Lorton	Fairfax County	2022-02-28	5.8	890128.0	3
49	Reston	VA	Reston	Fairfax County	2022-02-28	5.8	479682.0	3
1	Crosspointe	VA	Lorton	Fairfax County	2022-02-28	5.8	555413.0	3
37	Huntington	VA	Huntington	Fairfax County	2022-02-28	5.8	481086.0	3
43	Newington	VA	Springfield	Fairfax County	2022-02-28	5.9	608921.0	3
40	West Falls Church	VA	Falls Church	Fairfax County	2022-02-28	5.9	618723.0	3
24	Annandale	VA	Annandale	Fairfax County	2022-02-28	5.9	637630.0	3
14	Kingstowne	VA	Rose Hill	Fairfax County	2022-02-28	5.9	555972.0	3
12	Long Branch	VA	Fairfax	Fairfax County	2022-02-28	5.9	680395.0	3
9	Kings Park West	VA	Fairfax	Fairfax County	2022-02-28	5.9	680395.0	3
2	Wakefield	VA	Annandale	Fairfax County	2022-02-28	5.9	637630.0	3
15	Newington Forest	VA	Springfield	Fairfax County	2022-02-28	5.9	608921.0	3
51	Rose Hill	VA	Rose Hill	Fairfax County	2022-02-28	5.9	555972.0	3

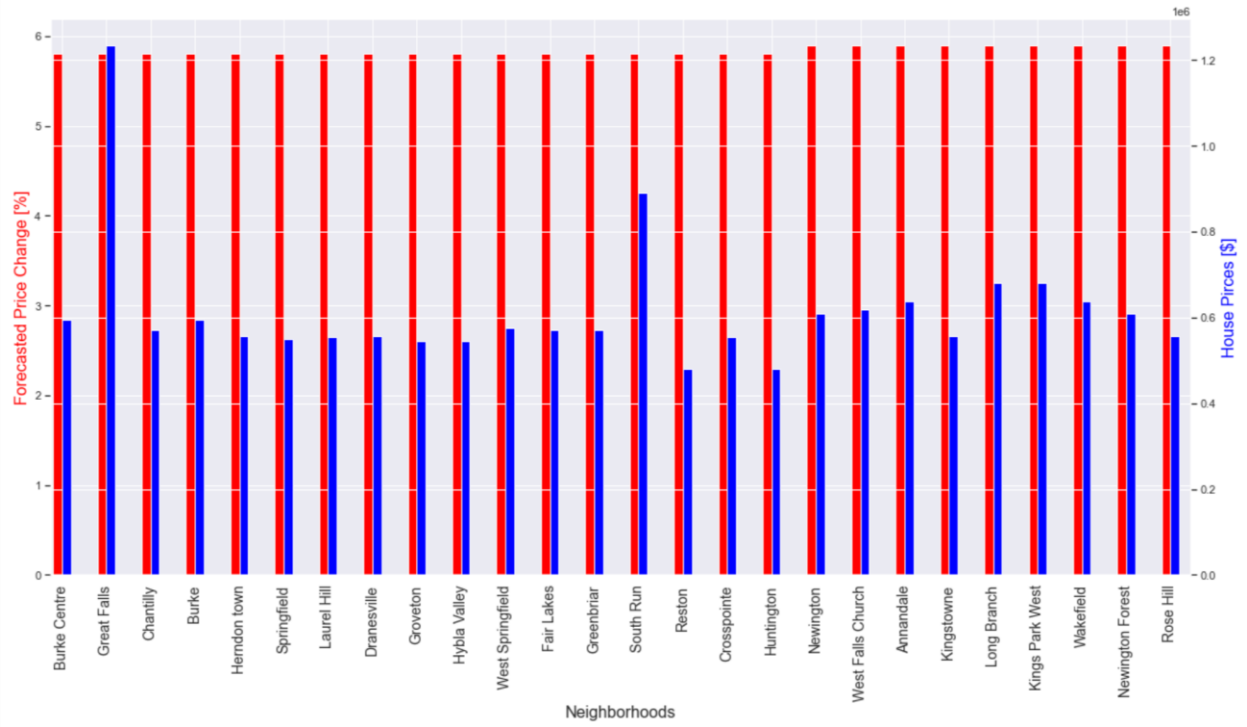

Figure 26. Cluster 4 - Forecasted Change in Price

5.15. Cluster 5 – Forecasted Change in Price

In this group, five neighborhoods with the similar forecasted percentage change in housing prices were clustered together. The best investment location is Fort Belvoir followed by Vienna town due to lower initial costs and a more significant increase in the forecasted percentage change of housing prices. The results in Figure 27 Figure 16confirm this assertion.

	Neighborhoods	StateName	CityName	CountyName	ForecastedDate	ForecastYoYPctChange	House_Price	Cluster Labels
20	McLean	VA	McLean	Fairfax County	2022-02-28	5.3	1255540.0	4
27	Vienna town	VA	Vienna	Fairfax County	2022-02-28	5.4	824915.0	4
31	Fort Belvoir	VA	Fort Belvoir	Fairfax County	2022-02-28	5.4	644641.0	4
47	Wolf Trap	VA	Vienna	Fairfax County	2022-02-28	5.4	949536.0	4
48	Oakton	VA	Vienna	Fairfax County	2022-02-28	5.5	858957.0	4

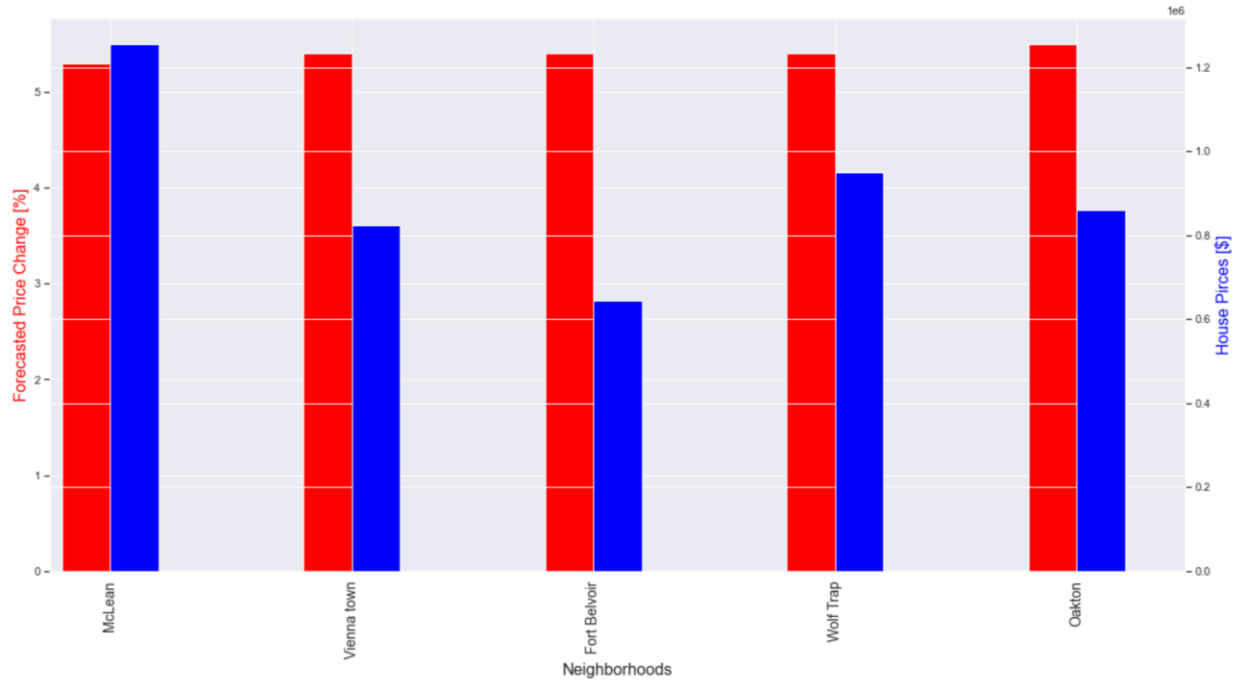

Figure 27. Cluster 5 - Forecasted Change in Price

6. Conclusion

Fairfax County is a suburb of Washington D.C. in northern Virginia. It is the most populous county in the state and has some of the most expensive housing markets in the DC-Maryland-Virginia (DMV) metropolitan region. Despite the global impact of the coronavirus pandemic, the real estate market in Fairfax County remains strong and competitive. A relatively strong economy mainly drives this competitiveness in the DMV area. Other factors that influence the real estate market are lack of inventory, low mortgage rates, and trim down payment options. However, the real estate market boom may not be uniform across different neighborhoods in the county. We employed three different clustering approaches based on venue category, housing prices, and forecasted percentage change in the housing prices to compare various localities and generate insights for potential investments. Using data analysis and machine learning techniques, mainly k-means clustering, we identified the best neighborhoods for investment. We also identified locations that are the worst neighborhoods for investment purposes. This analysis provides an insight into different areas and allows us to invest in our desired neighborhoods. Based on the number of times a community was ranked the best or worst places for investment in all clustering approaches, the information in Table 1 was compiled.

Table 1. The Best and Worst Neighborhoods for Investment

Investment Ranking	Neighborhood
1st Best	Bailey's Crossroads Huntington Reston
2nd Best	Chantilly Greenbriar Woodlawn
1st Worst	Great Falls South Run

References:

[1] https://www.fairfaxcounty.gov/demographics/interactive-map-communities-places-and-towns

[2] https://catalog.data.gov/bg/harvest/5e0e7936-9905-4d14-be00-72ea8f67d01f

[3] https://www.zillow.com/research/data/