
COURSERA CAPSTONE: NORTHERN VIRGINIA NEIGHBORHOODS



NORTHERN VIRGINIA SUBURBS ARE GROWING DOUBLE DIGITS AND PRESENT BUSINESS OPPORTUNITIES.

- Northern Virginia comprises cities and towns which are suburbs to Washington DC. Population growth since the 2010 Census was 12.4%, comprising two-thirds of Virginia's growth¹.
- Job growth in Northern Virginia is expected to grow by 45% from 2010 to 2030, with the fastest growth occurring in the Services Sector². This represents a shift in dependency on the government sector. In 2019, Amazon awarded a Northern Virginia location – Crystal City - the location for their East Coast headquarters which is expected to bring 25,000 jobs to the Northern Virginia area over the next 10 years³.
- Understanding the profile of Northern Virginia suburbs can help potential entrepreneurs pin-point business opportunities.
- Additionally, understanding popular venues within each neighborhood helps families identify fun activities.

Sources:

¹<https://news.virginia.edu/content/state-population-growth-slows-most-new-residents-northern-virginia>

²https://cra.gmu.edu/pdfs/research_reports/recent_reports/Future%20of%20NVA%20Economy.pdf

³<https://online.mason.wm.edu/blog/amazon-hq2-jobs-virginia>

DATA ACQUISITION & METHODOLOGY

DATA

- Used dataset from Open Data Soft (<https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/>), which contains US Cities, States, and their Longitude and Latitudes.
- Enhanced the dataset to include an identifier for Northern Virginia cities and towns
- Created a dataframe of Northern Virginia cities and towns

Methodology

- Use FourSquare to access the top 5 venues for each Northern Virginia city or town, using a radius of 33,000 meters (or 20 miles). Unlike large cities which are walkable and have broad-reaching transit systems, people who live in Northern Virginia drive to most destinations.
- Used k-means cluster analysis to segment Northern Virginia into 10 clusters. Since Northern Virginia encompasses an area of approximately 50 miles, used a higher count of clusters.

K-MEANS CLUSTERING – IDENTIFYING BUSINESS OPPORTUNITY BY LOOKING AT LOW VARIETY OF VENUES AND PRESENCE OF FAMILY ATTRACTIONS.

- A K-Means cluster analysis for 10 clusters produced the top 5 venues for each cluster. Venues were largely focused on food & drink type of services.
- A potential entrepreneur would be interested in identifying gaps in a market where demand is not met. We can look at the variety of venues represented in the cluster analysis as a potential measure of gaps which could present a business opportunity.
- Additionally, a potential entrepreneur would be interested in the available attractions which could generate additional business from outside the town or city.
- Areas which could be of interest to an entrepreneur are highlighted in blue – 5 clusters.

Cluster	# zipcodes	Characteristic	Variety of Venues?	Family Attractions?
0	4	Classic suburb	No	Yes
1	8	People who cook	No	No
2	59	Metropolitan	Yes	Yes
3	11	Brewery town(s)	Yes	Yes
4	20	Fast Food County	No	Yes
5	54	Monument Towns	Yes	Yes
6	4	Wine Country	No	Yes
7	14	Coffee & Carbs	No	No
8	2	International Foodi	No	Yes
9	1	Scenic Lookout	No	Yes



RECOMMENDED ADDITIONAL ANALYSIS TO FURTHER VALIDATE BUSINESS OPPORTUNITY IN HIGHLIGHTED CLUSTERS

- Incorporate population, demographic, socioeconomic, and tourist data into the dataset to gauge the potential size of the market. This analysis should not only provide the volume of potential customers but what they are willing to pay for food and/or drink.
- Incorporate data about potential street locations which would be desirable to locate a new business – where there are no food & drink venues, where there is dense traffic, and availability of real estate.



BACKUP – CLUSTERS AND TOP 5 VENUES


CLUSTER 0 – LOW VARIETY OF RESTAURANTS. OPPORTUNITY TO EXPLORE PARK, HISTORIC SITE AND ZOO.

```
In [27]: NVA_merged.loc[NVA_merged['Cluster Labels'] == 0, NVA_merged.columns[[1] + list(range(5, NVA_merged.shape[1]))]]
```

Out[27]:

	City	Timezone	NVA	PWC	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
34	Clifton Forge	-5	NVA	NaN	0	American Restaurant	Gas Station	Sandwich Place	Resort	Zoo
50	Fort Belvoir	-5	NVA	NaN	0	Park	American Restaurant	Historic Site	French Restaurant	Supermarket
56	Fairfax Station	-5	NVA	NaN	0	American Restaurant	Greek Restaurant	Supermarket	Movie Theater	Grocery Store
113	Occoquan	-5	NVA	PWC	0	Historic Site	American Restaurant	Supermarket	Pool	Farmers Market

CLUSTER 1 – FAMILY-ORIENTED SUBURB (DOMINANCE OF GROCERY STORE). LOW VARIETY OF RESTAURANTS (DOMINATED BY ITALIAN, FAST FOOD, AMERICAN).

 NVA_merged.loc[NVA_merged['Cluster Labels'] == 1, NVA_merged.columns[[1] + list(range(5, NVA_merged.shape[1]))]]

rt[28]:

	City	Timezone	NVA	PWC	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Mount Vernon	-5	NVA	NaN	1	Grocery Store	Italian Restaurant	American Restaurant	Fast Food Restaurant	Golf Course
65	Centreville	-5	NVA	NaN	1	Grocery Store	Italian Restaurant	Fast Food Restaurant	American Restaurant	Supermarket
72	Chantilly	-5	NVA	NaN	1	Grocery Store	Brewery	Supermarket	Wine Shop	American Restaurant
74	Centreville	-5	NVA	NaN	1	Grocery Store	Italian Restaurant	Fast Food Restaurant	American Restaurant	Supermarket
86	Clifton	-5	NVA	NaN	1	Grocery Store	Park	American Restaurant	Italian Restaurant	Supermarket
94	Chantilly	-5	NVA	NaN	1	Grocery Store	Brewery	Supermarket	Wine Shop	American Restaurant
100	Chantilly	-5	NVA	NaN	1	Grocery Store	Brewery	Supermarket	Wine Shop	American Restaurant
116	Centreville	-5	NVA	NaN	1	Grocery Store	Italian Restaurant	Fast Food Restaurant	American Restaurant	Supermarket

CLUSTER 2 – LARGEST CLUSTER (NOT ALL ROWS SHOWN). VARIETY OF VENUES.

```
NVA_merged.loc[NVA_merged['Cluster Labels'] == 2, NVA_merged.columns[[1] + list(range(5, NVA_merged.shape[1]))]]
```

29]:

	City	Timezone	NVA	PWC	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Fairfax	-5	NVA	NaN	2	Supermarket	Greek Restaurant	Coffee Shop	Grocery Store	American Restaurant
6	Merrifield	-5	NVA	NaN	2	Greek Restaurant	Wine Shop	Grocery Store	Supermarket	Taco Place
8	Burke	-5	NVA	NaN	2	Greek Restaurant	Grocery Store	Supermarket	Pizza Place	Movie Theater
11	Merrifield	-5	NVA	NaN	2	Greek Restaurant	Wine Shop	Grocery Store	Supermarket	Taco Place
12	Fairfax	-5	NVA	NaN	2	Supermarket	Greek Restaurant	Coffee Shop	Grocery Store	American Restaurant
18	Springfield	-5	NVA	NaN	2	Pizza Place	Taco Place	Supermarket	Greek Restaurant	Brewery
19	Fairfax	-5	NVA	NaN	2	Supermarket	Greek Restaurant	Coffee Shop	Grocery Store	American Restaurant
24	Herndon	-5	NVA	NaN	2	Brewery	Grocery Store	Taco Place	Greek Restaurant	Coffee Shop
25	Springfield	-5	NVA	NaN	2	Pizza Place	Taco Place	Supermarket	Greek Restaurant	Brewery
28	Falls Church	-5	NVA	NaN	2	Taco Place	American Restaurant	Pizza Place	Wine Shop	Coffee Shop
30	Herndon	-5	NVA	NaN	2	Brewery	Grocery Store	Taco Place	Greek Restaurant	Coffee Shop

CLUSTER 3 – CLASSIC SUBURB, DOMINATED BY BREWERIES. MODERATE VARIETY OF RESTAURANTS.

```
NVA_merged.loc[NVA_merged['Cluster Labels'] == 3, NVA_merged.columns[[1] + list(range(5, NVA_merged.shape[1]))]]
```

0]:

	City	Timezone	NVA	PWC	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
31	Dulles	-5	NVA	NaN	3	Brewery	American Restaurant	Supermarket	Café	Vineyard
61	Purcellville	-5	NVA	NaN	3	Brewery	Winery	American Restaurant	Pizza Place	Park
77	Dulles	-5	NVA	NaN	3	Brewery	American Restaurant	Supermarket	Café	Vineyard
80	Leesburg	-5	NVA	NaN	3	American Restaurant	Brewery	Sandwich Place	Wine Bar	Pizza Place
91	Dulles	-5	NVA	NaN	3	Brewery	American Restaurant	Supermarket	Café	Vineyard
131	Purcellville	-5	NVA	NaN	3	Brewery	Winery	American Restaurant	Pizza Place	Park
136	Leesburg	-5	NVA	NaN	3	American Restaurant	Brewery	Sandwich Place	Wine Bar	Pizza Place
137	Leesburg	-5	NVA	NaN	3	American Restaurant	Brewery	Sandwich Place	Wine Bar	Pizza Place
149	Dulles	-5	NVA	NaN	3	Brewery	American Restaurant	Supermarket	Café	Vineyard
174	Leesburg	-5	NVA	NaN	3	American Restaurant	Brewery	Sandwich Place	Wine Bar	Pizza Place
175	Dulles	-5	NVA	NaN	3	Brewery	American Restaurant	Supermarket	Café	Vineyard

CLUSTER 4 – THIRD LARGEST CLUSTER, DOMINATED BY FAST FOOD RESTAURANTS AND PARKS.

`NVA_merged.loc[NVA_merged['Cluster Labels'] == 4, NVA_merged.columns[[1] + list(range(5, NVA_merged.shape[1]))]]`

`t[32]:`

	City	Timezone	NVA	PWC	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
15	Woodbridge	-5	NVA	PWC	4	Fast Food Restaurant	Italian Restaurant	American Restaurant	Park	Grocery Store
17	Gainesville	-5	NVA	PWC	4	Fast Food Restaurant	Golf Course	Historic Site	Brewery	Grocery Store
20	Manassas	-5	NVA	PWC	4	Fast Food Restaurant	Grocery Store	Italian Restaurant	Park	American Restaurant
26	Haymarket	-5	NVA	PWC	4	Fast Food Restaurant	Historic Site	Brewery	Golf Course	Park
32	Manassas	-5	NVA	PWC	4	Fast Food Restaurant	Grocery Store	Italian Restaurant	Park	American Restaurant
37	Nokesville	-5	NVA	PWC	4	Brewery	Historic Site	Mexican Restaurant	Fast Food Restaurant	Convenience Store
45	Quantico	-5	NVA	PWC	4	Fast Food Restaurant	Italian Restaurant	Park	Greek Restaurant	Grocery Store
54	Bristow	-5	NVA	PWC	4	Brewery	Convenience Store	Fast Food Restaurant	Historic Site	Supermarket
60	Nokesville	-5	NVA	PWC	4	Brewery	Historic Site	Mexican Restaurant	Fast Food Restaurant	Convenience Store
79	Manassas	-5	NVA	PWC	4	Fast Food Restaurant	Grocery Store	Italian Restaurant	Park	American Restaurant
92	Manassas	-5	NVA	PWC	4	Fast Food Restaurant	Grocery Store	Italian Restaurant	Park	American Restaurant
126	Gainesville	-5	NVA	PWC	4	Fast Food Restaurant	Golf Course	Historic Site	Brewery	Grocery Store
129	Manassas	-5	NVA	PWC	4	Fast Food Restaurant	Grocery Store	Italian Restaurant	Park	American Restaurant
139	Haymarket	-5	NVA	PWC	4	Fast Food Restaurant	Historic Site	Brewery	Golf Course	Park
140	Woodbridge	-5	NVA	PWC	4	Fast Food Restaurant	Italian Restaurant	American Restaurant	Park	Grocery Store

CLUSTER 5 – SECOND LARGEST CLUSTER (NOT ALL SHOWN), DOMINATED BY MONUMENTS/LANDMARKS, PARKS. DIVERSE VENUES.

```
NVA_merged.loc[NVA_merged['Cluster Labels'] == 5, NVA_merged.columns[[1] + list(range(5, NVA_merged.shape[1]))]]
```

ut[33]:

	City	Timezone	NVA	PWC	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Alexandria	-5	NVA	NaN	5	Monument / Landmark	Park	Pizza Place	Coffee Shop	American Restaurant
3	Alexandria	-5	NVA	NaN	5	Monument / Landmark	Park	Pizza Place	Coffee Shop	American Restaurant
4	Alexandria	-5	NVA	NaN	5	Monument / Landmark	Park	Pizza Place	Coffee Shop	American Restaurant
7	Arlington	-5	NVA	NaN	5	Monument / Landmark	Coffee Shop	Park	Italian Restaurant	Plaza
9	Alexandria	-5	NVA	NaN	5	Monument / Landmark	Park	Pizza Place	Coffee Shop	American Restaurant
10	Alexandria	-5	NVA	NaN	5	Monument / Landmark	Park	Pizza Place	Coffee Shop	American Restaurant
13	Alexandria	-5	NVA	NaN	5	Monument / Landmark	Park	Pizza Place	Coffee Shop	American Restaurant
14	Arlington	-5	NVA	NaN	5	Monument / Landmark	Coffee Shop	Park	Italian Restaurant	Plaza
27	Arlington	-5	NVA	NaN	5	Monument / Landmark	Coffee Shop	Park	Italian Restaurant	Plaza
36	Arlington	-5	NVA	NaN	5	Monument / Landmark	Coffee Shop	Park	Italian Restaurant	Plaza
38	Alexandria	-5	NVA	NaN	5	Monument / Landmark	Park	Pizza Place	Coffee Shop	American Restaurant

CLUSTER 6 – WINE COUNTRY.

```
NVA_merged.loc[NVA_merged['Cluster Labels'] == 6, NVA_merged.columns[[1] + list(range(5, NVA_merged.shape[1]))]]
```

34]:

	City	Timezone	NVA	PWC	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
16	Middleburg	-5	NVA	NaN	6	Winery	Brewery	Vineyard	Coffee Shop	American Restaurant
88	Marshall	-5	NVA	NaN	6	Winery	Vineyard	American Restaurant	Coffee Shop	Golf Course
125	Marshall	-5	NVA	NaN	6	Winery	Vineyard	American Restaurant	Coffee Shop	Golf Course

CLUSTER 7 – SMALL TOWNS DOMINATED BY LOCAL COFFEE SHOP / BAKERY.

```
NVA_merged.loc[NVA_merged['Cluster Labels'] == 7, NVA_merged.columns[[1] + list(range(5, NVA_merged.shape[1]))]]
```

t[35]:

	City	Timezone	NVA	PWC	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Culpeper	-5	NVA	NaN	7	Mexican Restaurant	Mediterranean Restaurant	Seafood Restaurant	Café	Burger Joint
21	Fredericksburg	-5	NVA	NaN	7	Coffee Shop	Fast Food Restaurant	Sushi Restaurant	Brewery	Supermarket
22	Spotsylvania	-5	NVA	NaN	7	Fast Food Restaurant	Grocery Store	Farmers Market	Donut Shop	Market
23	Fredericksburg	-5	NVA	NaN	7	Coffee Shop	Fast Food Restaurant	Sushi Restaurant	Brewery	Supermarket
29	Fredericksburg	-5	NVA	NaN	7	Coffee Shop	Fast Food Restaurant	Sushi Restaurant	Brewery	Supermarket
35	Fredericksburg	-5	NVA	NaN	7	Coffee Shop	Fast Food Restaurant	Sushi Restaurant	Brewery	Supermarket
42	Fredericksburg	-5	NVA	NaN	7	Coffee Shop	Fast Food Restaurant	Sushi Restaurant	Brewery	Supermarket
44	Fredericksburg	-5	NVA	NaN	7	Coffee Shop	Fast Food Restaurant	Sushi Restaurant	Brewery	Supermarket
85	Warrenton	-5	NVA	NaN	7	Bakery	Brewery	Winery	Fast Food Restaurant	American Restaurant
89	Warrenton	-5	NVA	NaN	7	Bakery	Brewery	Winery	Fast Food Restaurant	American Restaurant
106	Fredericksburg	-5	NVA	NaN	7	Coffee Shop	Fast Food Restaurant	Sushi Restaurant	Brewery	Supermarket

CLUSTER 8 – POPULAR GREEK / JAPANESE RESTAURANT. OPPORTUNITY TO EXPLORE FRENCH RESTAURANT AND NATIONAL PARK.

```
NVA_merged.loc[NVA_merged['Cluster Labels'] == 8, NVA_merged.columns[[1] + list(range(5, NVA_merged.shape[1]))]]
```

it[37]:

	City	Timezone	NVA	PWC	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
108	Dumfries	-5	NVA	PWC	8	Greek Restaurant	Fast Food Restaurant	Japanese Restaurant	Bakery	French Restaurant
163	Triangle	-5	NVA	PWC	8	Japanese Restaurant	Greek Restaurant	Fast Food Restaurant	Hotel	National Park

CLUSTER 9 – CLUSTER OF ONE, UNIQUE TOWN, MUST SEE!

```
NVA_merged.loc[NVA_merged['Cluster Labels'] == 9, NVA_merged.columns[[1] + list(range(5, NVA_merged.shape[1]))]]
```

[38]:

	City	Timezone	NVA	PWC	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
47	Washington	-5	NVA	NaN	9	Scenic Lookout	Winery	Trail	Coffee Shop	National Park