Where to live in Vancouver ?

1. Business Problem
2. Data aquisition
3. Methodology
4. Results and Discussion
5. Conclusion

1. Business Problem

- Job offer in Vancouver, BC, Canada
- Need a location to move to fullfilling following criteria
  - **Workplace nearby** and **airport not too far** away
  - **Good** public transportation
  - Lots of **restaurants** and **coffee places**
  - Close to **parks** and/or **sports facilities**
  - **Education** opportunities for children

2. Data Aquisition

- **Neighborhoods** of Vancouver, BC scraped with **requests** and **Beautiful Soup**
  (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_V)
- **Location Data** for Neighboorhoods from **geocoder.arcgis** API
- **Venue Data** from **Foursquare API** with focus on categories that fit the relevant criteria
- **Maps** created with **folium** package

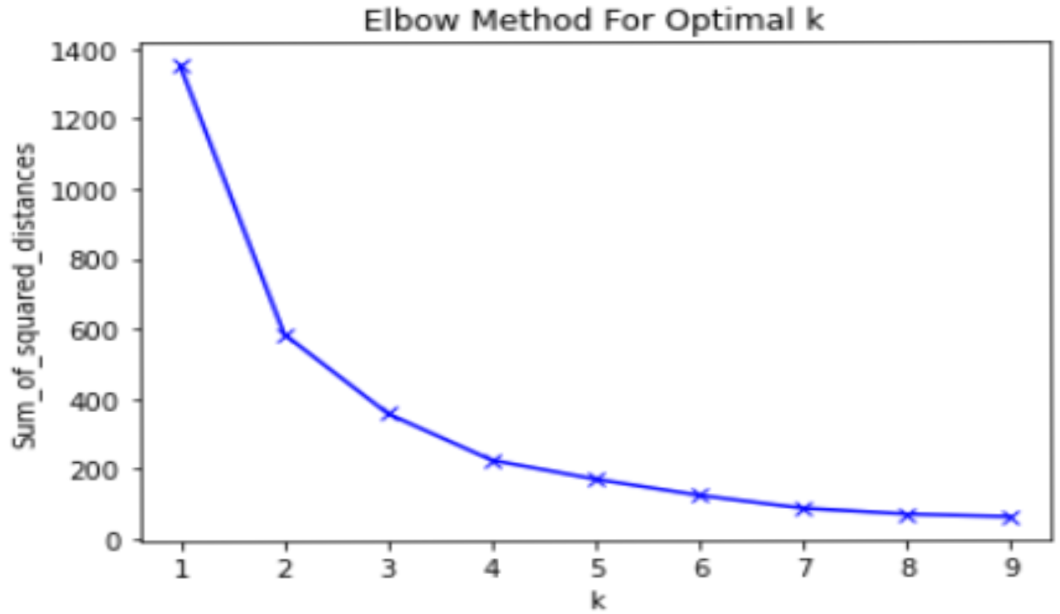
Methodology

- For Clustering the data is preprocessed to a DataFrame with shape 44x187
- Venue Categories are One Hot Encoded to make the clustering work

	Neighborhood	Accessories Store	Airport Terminal	American Restaurant	Amphitheater	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	...	Trail	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Ware
0	Bentall Centre	0.0	0.014085	0.014085	0.0	0.014085	0.0	0.0	0.0	0.014085	...	0.0	0.000000	0.000000	0.0
1	Central Kitsilano, Greektown	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.020408	0.020408	0.0
2	East Central	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.5	0.000000	0.000000	0.0
3	East Fairview, South Cambie	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.050000	0.0
4	East Mount Pleasant	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.058824	0.0
5	Inner East	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0
6	Killarney	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0
7	NE Downtown, Gastown, Harbour Centre, Internat...	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0
8	NW Arbutus Ridge, NE Dunbar-Southlands	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.000000	0.0

Methodology

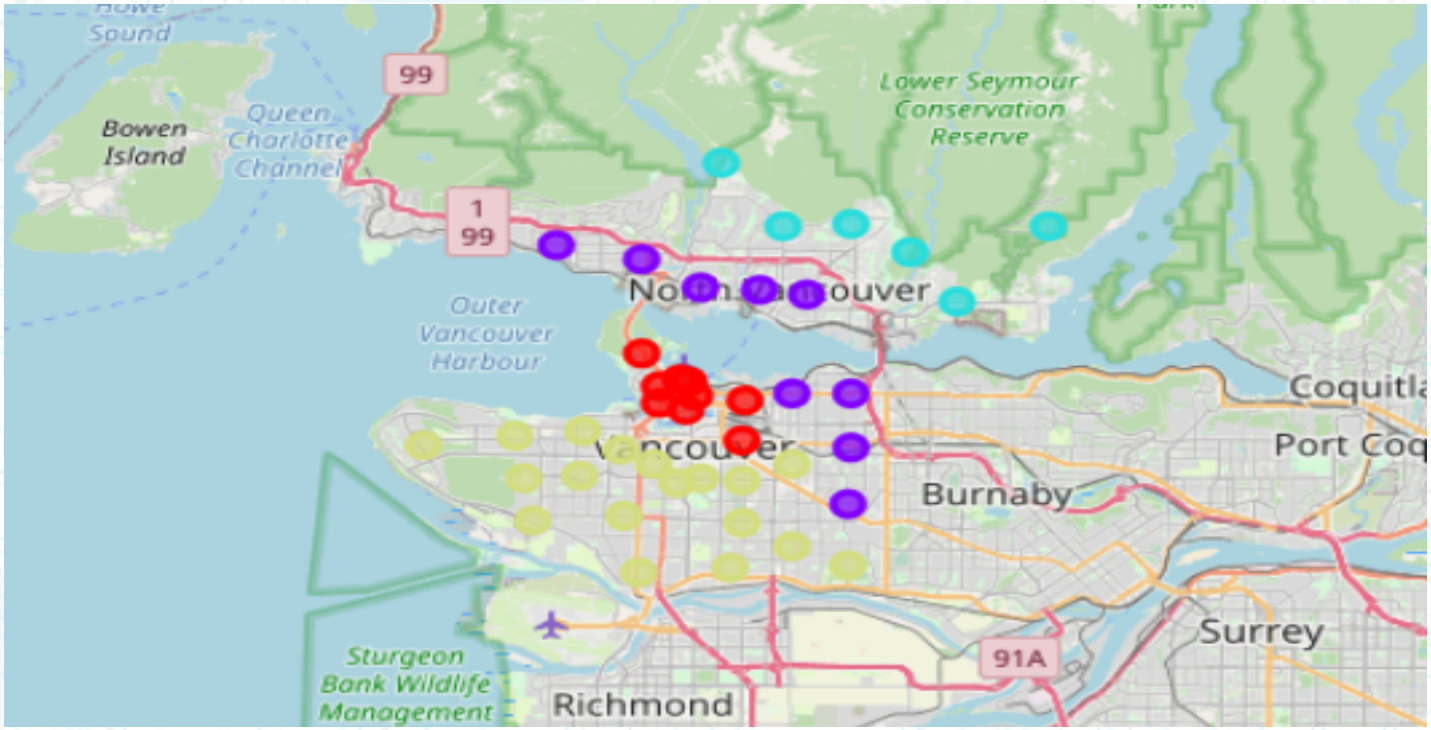
- For Clustering the **K-Means Algorithm** is selected
- The K-Means Algorithm needs a predefined **number of clusters**, which is defined using the so called **Elbow Method**



- **Selected Number of Clusters = 4**

Methodology

- The Neighborhoods and their cluster labels are visualized on the Map of Vancouver



- Like you can see on the map the location of the neighborhoods seems to influence the clustering to further describe the clusters the top 10 features per cluster are determined

Results and Discussion

Cluster 0:

- moderate distance to airport
- short distance to work
- lots of hotels
- lots of restaurants and cafés
- a few parks

Cluster 1:

- far distance to airport
- moderate distance to work
- sports facilities and parks
- coffee shops and bakeries

Cluster 2:

- far distance to airport
- far distance to work
- lots of trails and parks, close to nature
- a few coffee shops

Cluster 3:

- close distance to airport
- close to moderate distance to work
- lots of restaurants and coffee shops
- good public transport
- many parks

- **Result:** Neighborhoods in **Cluster 3** fullfill the relevant criteria the most and should be used for further analysis

- **Diskussion:** Especially the **housing/rental prices** and **availability** have **not been considered** during the analytics within this notebook and are definitely relevant for the decision making process. Also some **crime stats** and soft factors like **noise, air quality** etc. could be interesting for deciding where to move to.

Conclusion

Final decision on **optimum housing location** will be made by stakeholders based on specific characteristics of neighborhoods and locations in **recommended zone cluster 3**, taking into consideration **additional factors** like **levels of noise** / **proximity to major roads**, **housing availability**, **prices**, **social dynamics** of every neighborhood etc.