Battle of Neighborhoods

Seattle, Washington

Problem Statement

- Select the best neighborhood to live, to buy a house or to build a restaurant etc in Seattle, Washington .
- Understanding the similarities and differences between the neighborhoods using Unsupervised K-Mean Clustering Algorithm

Objective

- Collecting Neighborhood’s top trending venues using Foursquare API(Beautiful Soup, http request)
- Forming neighborhood clusters based on venue categories using unsupervised k-mean clustering algorithm(sklearn)
- Identifying and understanding the similarities and differences between two chosen neighborhoods to retrieve more insights and to conclude with ease which neighborhood wins over other.

Python packages and Dependencies:

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Matplotlib – Python Plotting Module
- Sklearn – Python machine learning Library
- Folium – Map rendering Library

Work flow

- Web Scraping and Data Wrangling
- Top Trending Places Extraction and Clustering
- Decision Making based on the clustered neighborhoods, Population Distribution, School Ratings, Median House Price Analysis

Web Scraping and Data Wrangling

Beautiful Soup

Collecting Neighborhood/Postal code

http://seattlearea.com/zip-codes/

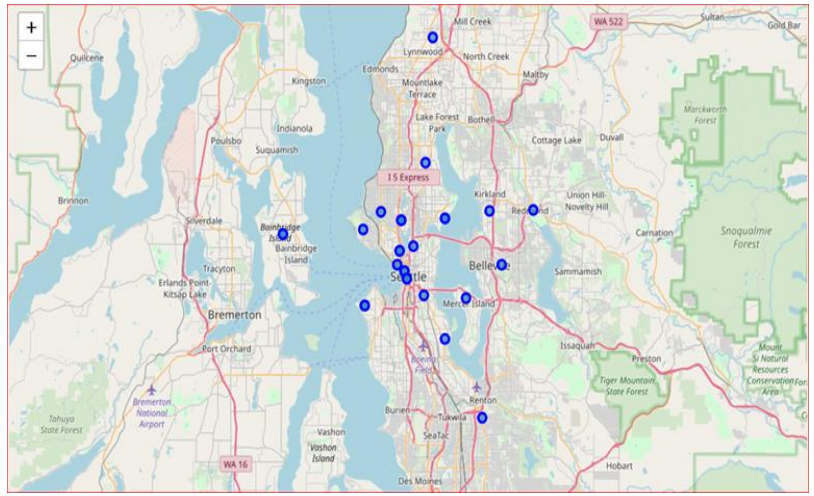
Google Maps API

To Collect Geographical Data

	PostalCode	Neighborhood	Latitude	Longitude
0	98003	Federal Way	47.316504	-122.322397
1	98005	Bellevue	47.615044	-122.171758
2	98033	Kirkland	47.668830	-122.192387
3	98037	Lynnwood	47.841952	-122.288181
4	98040	Mercer Island	47.582423	-122.233123

	PostalCode	Neighborhood
0	98003	Federal Way
1	98005	Bellevue
2	98033	Kirkland
3	98037	Lynnwood
4	98040	Mercer Island

Folium Visualization for Seattle Neighborhood



Venues Extraction using Four Square API and Clustering

Four Square API Calls to Collect Neighborhood Venue Category and LAT/LNG

	name	categories	lat	lng
0	Il Corvo	Italian Restaurant	47.602522	-122.331952
1	Biscuit B*tch	Breakfast Spot	47.603237	-122.332010
2	Columbia Tower Club	Social Club	47.604507	-122.330484
3	Juicy Cafe	Café	47.604329	-122.330958
4	Tat's Delicatessen	Sandwich Place	47.601901	-122.332423

One Hot Encoding to Convert Labels into Numbers

	Neighborhood	Yoga Studio	ATM	Accessories Store	Advertising Agency	Airport
0	Alki Beach	0.00	0.0	0.0	0.025	0.0
1	Bainbridge Island	0.00	0.0	0.0	0.000	0.0
2	Ballard	0.01	0.0	0.0	0.000	0.0
3	Bellevue	0.00	0.0	0.0	0.000	0.0
4	Belltown	0.00	0.0	0.0	0.000	0.0

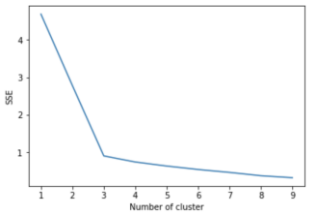
Venues Grouped by Neighborhood 229 Unique Venues

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Alki Beach	Park	Ice Cream Shop	Coffee Shop	Thai Restaurant
1	Bainbridge Island	Tree	Women's Store	Fabric Shop	Forest
2	Ballard	Coffee Shop	Brewery	Mexican Restaurant	Sandwich Place
3	Bellevue	Indian Restaurant	Coffee Shop	Other Repair Shop	Spa
4	Belltown	Bar	Coffee Shop	Gym	Sushi Restaurant

K-Means Clustering

Silhouette Score

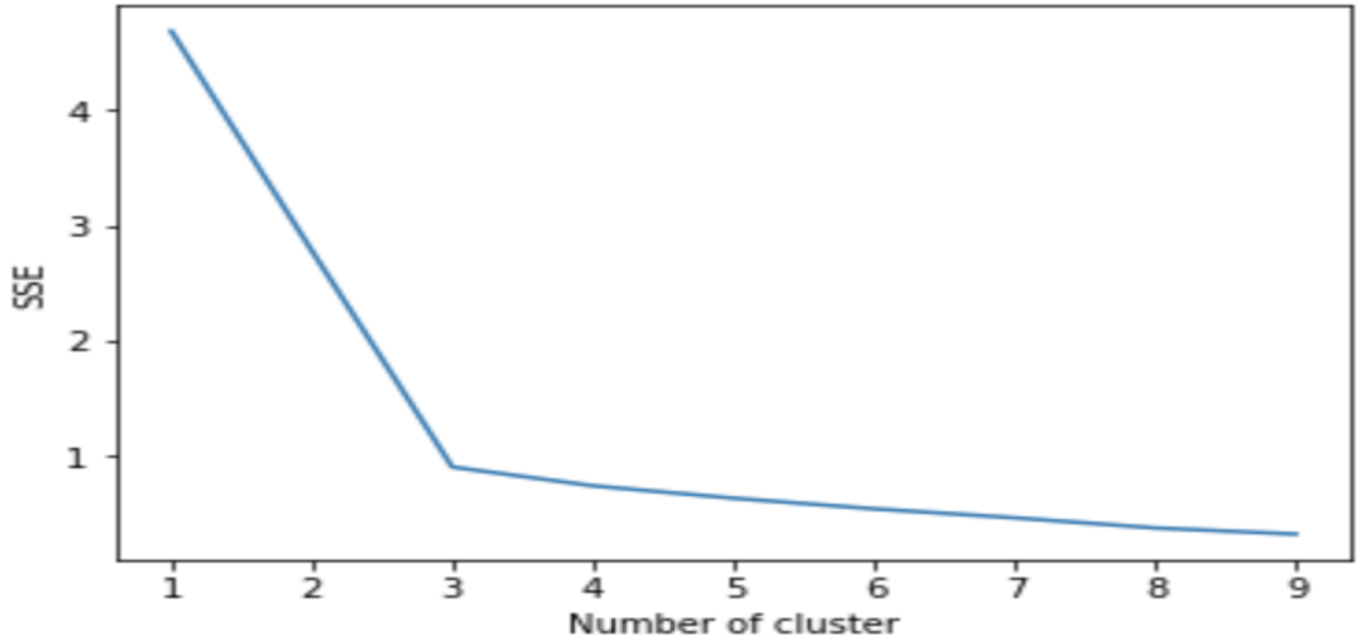
Elbow Method



- Cluster 1
- Cluster 2
- Cluster 3

Elbow Criterion Method

The idea behind **elbow method** is to run k-means clustering on a given dataset for a range of values of k and for each value of k, calculate sum of squared errors (SSE). After that, plot a line graph of the SSE for each value of k.



sklearn.metrics.silhouette_score

The Silhouette Coefficient is calculated using the

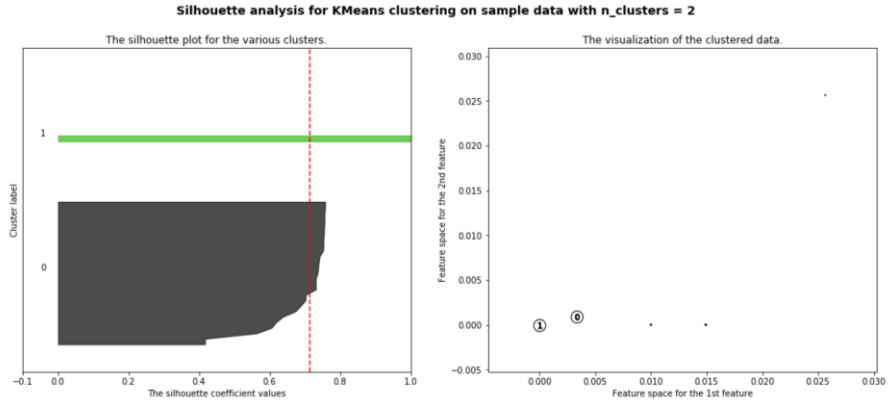
mean intra-cluster distance (a) and
the mean nearest-cluster distance (b) for each sample.

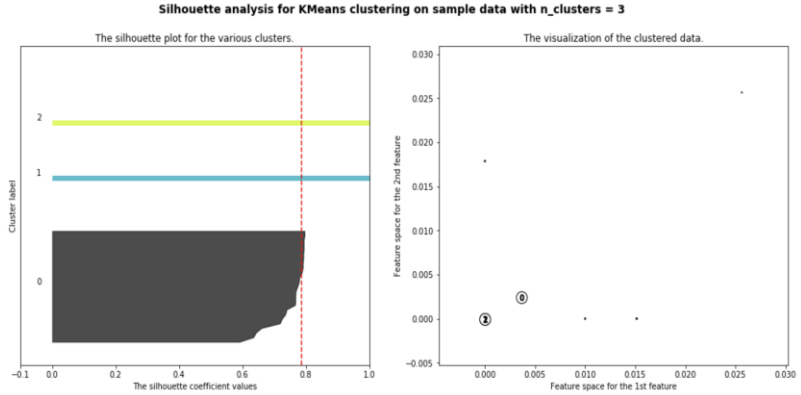
The Silhouette Coefficient for a sample is (b - a) / max(a, b).

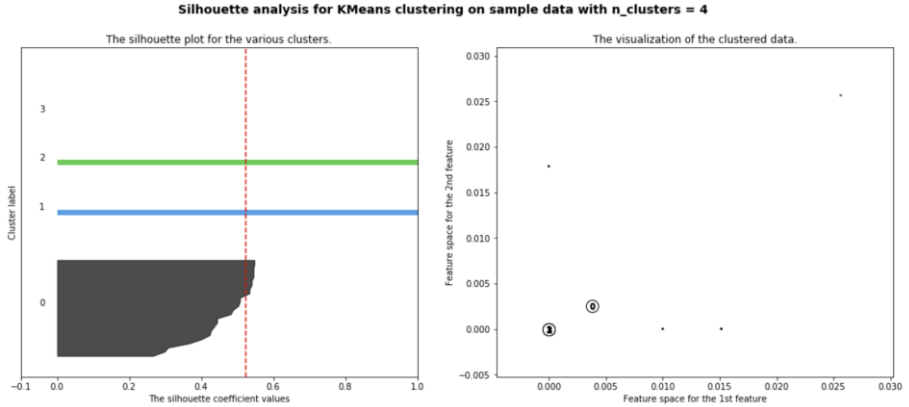
The best value is 1 and the worst value is -1. Values near 0 indicate overlapping clusters. Negative values generally indicate that a sample has been assigned to the wrong cluster, as a different cluster is more similar.

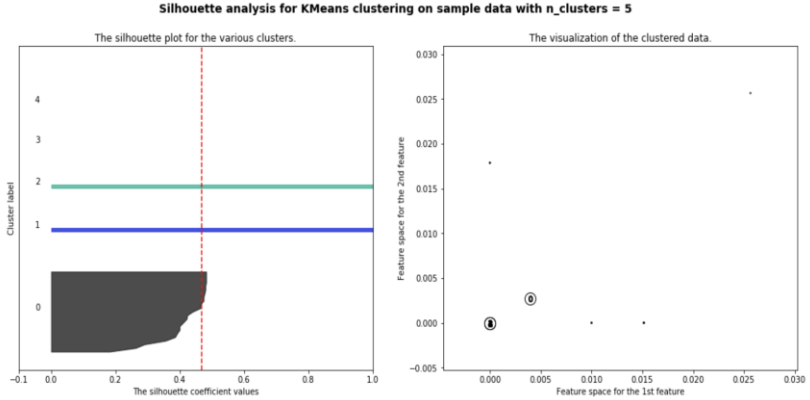
Siluhouette Score and Cluster Visualizations

```
For 2  Clusters  the average silhouette_score is : 0.6878385550826527
For 3  Clusters  the average silhouette_score is : 0.785035745080499
For 4  Clusters  the average silhouette_score is : 0.523387518795663
For 5  Clusters  the average silhouette_score is : 0.4692436167127194
```

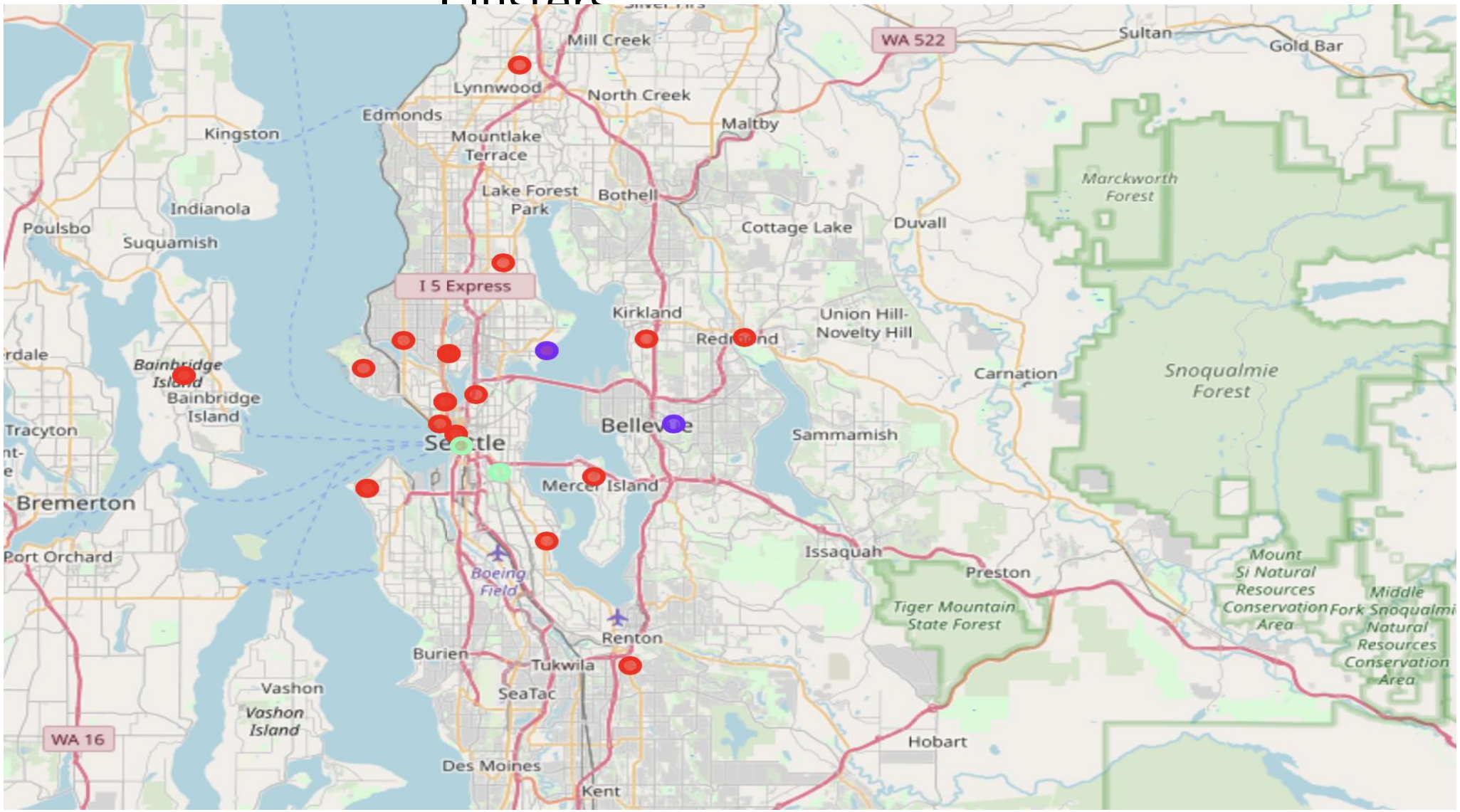




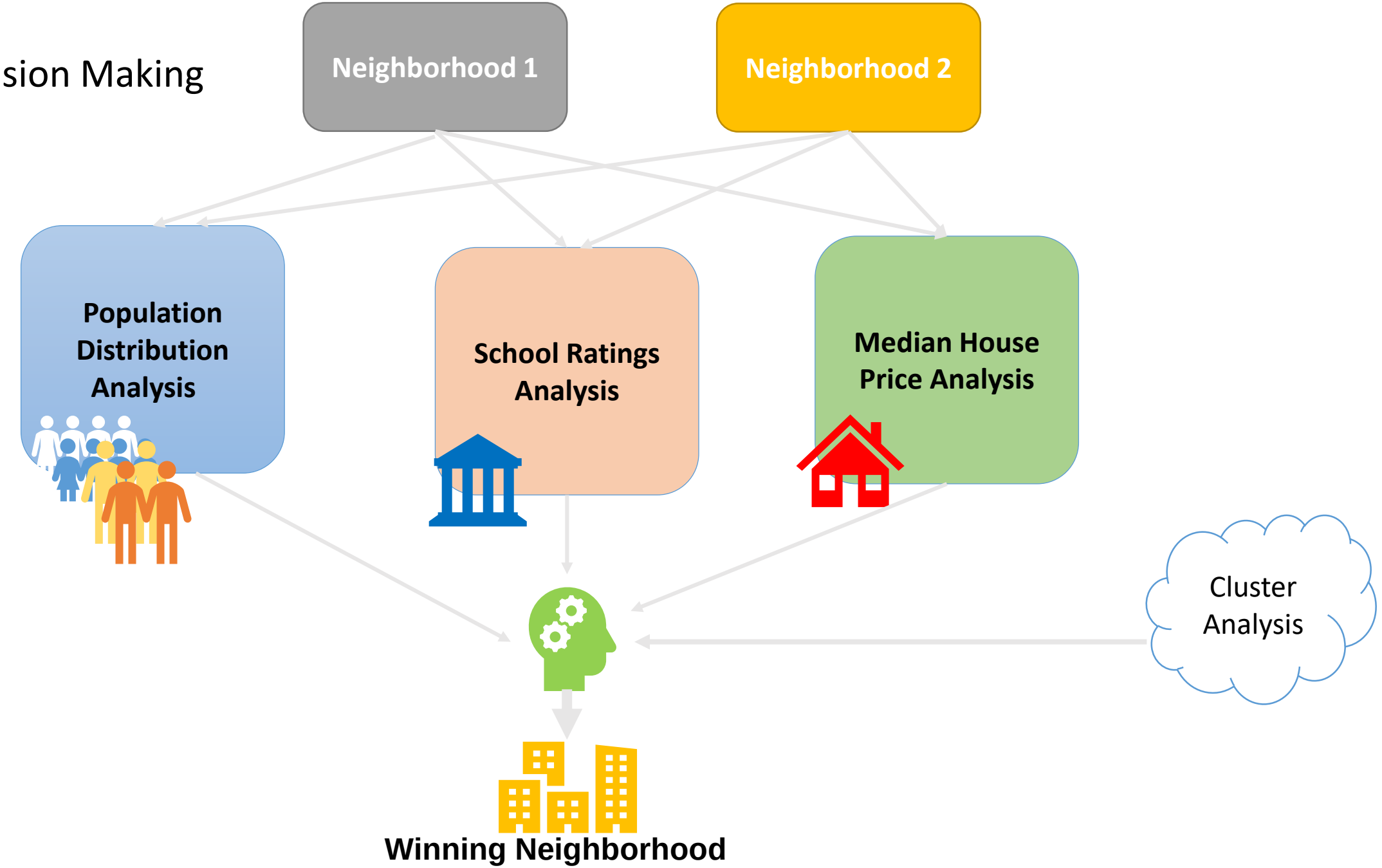




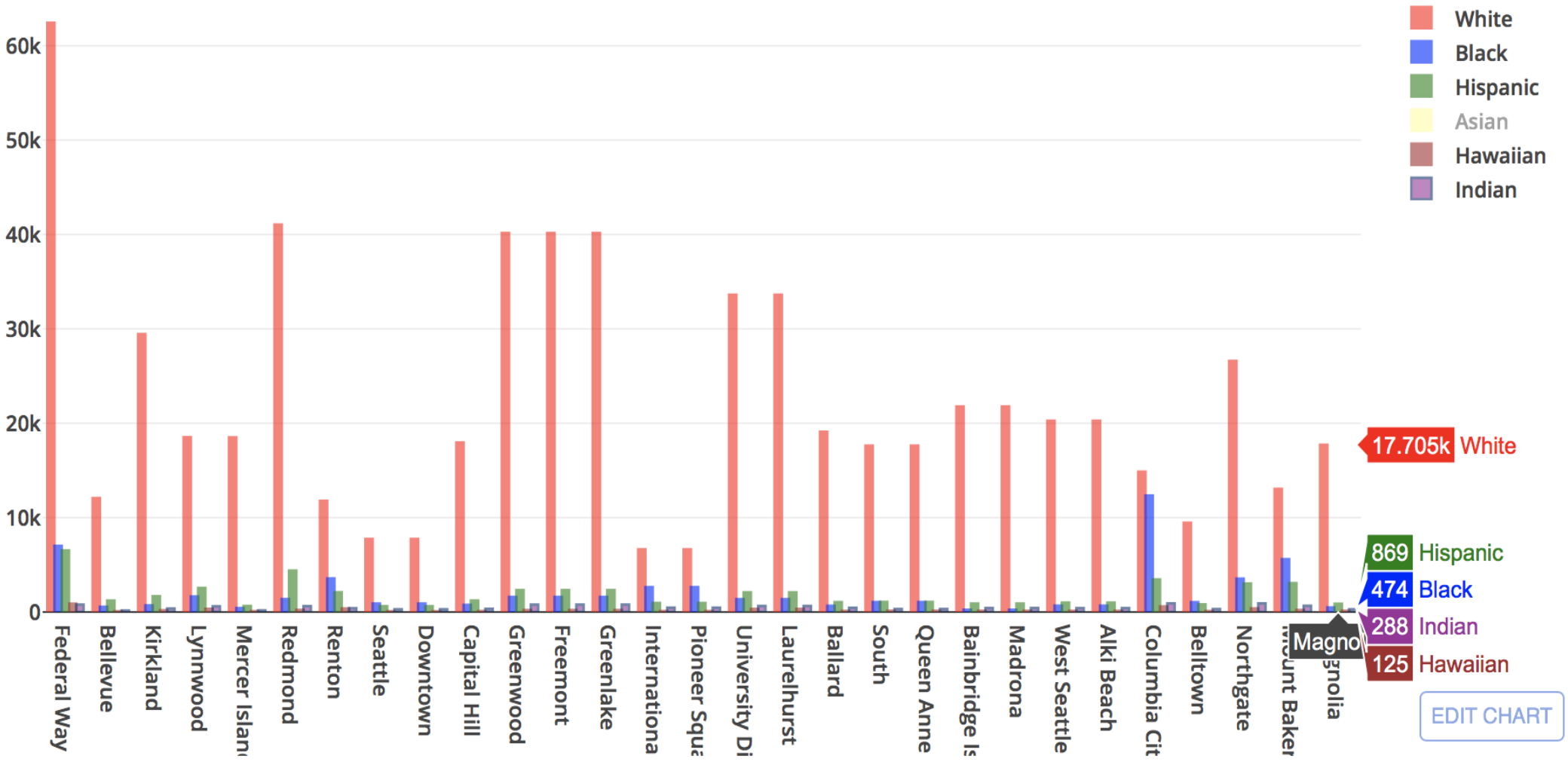
Neighborhood Clusters



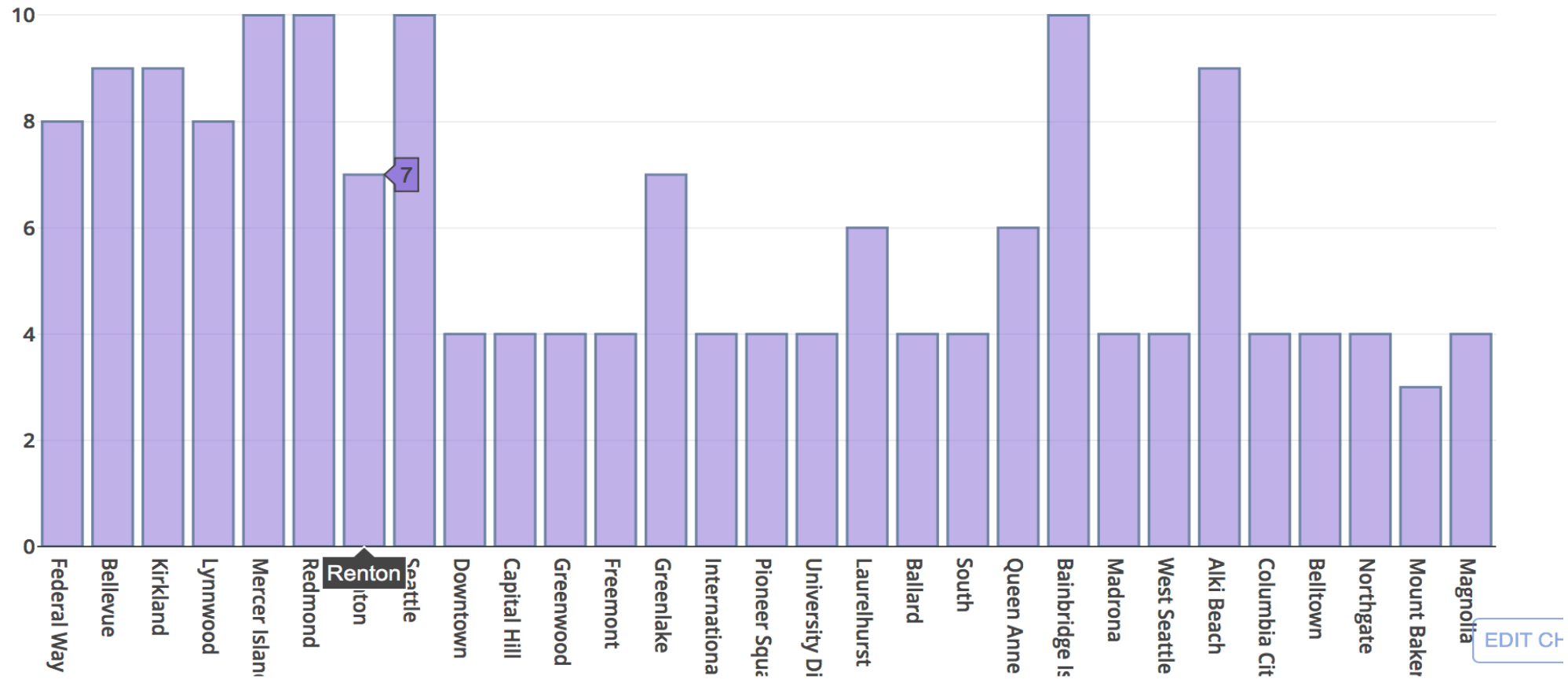
Decision Making



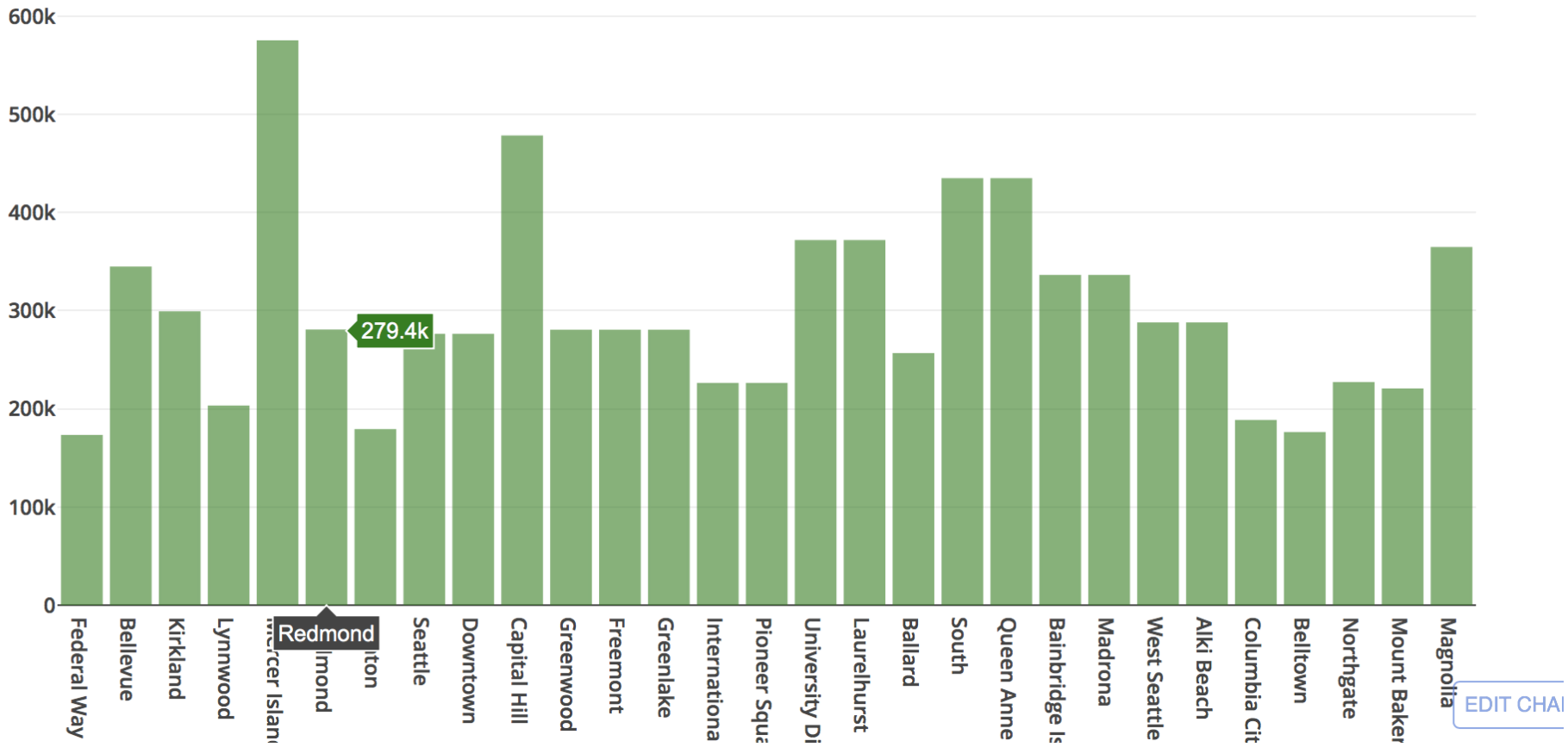
POPULATION DISTRIBUTION



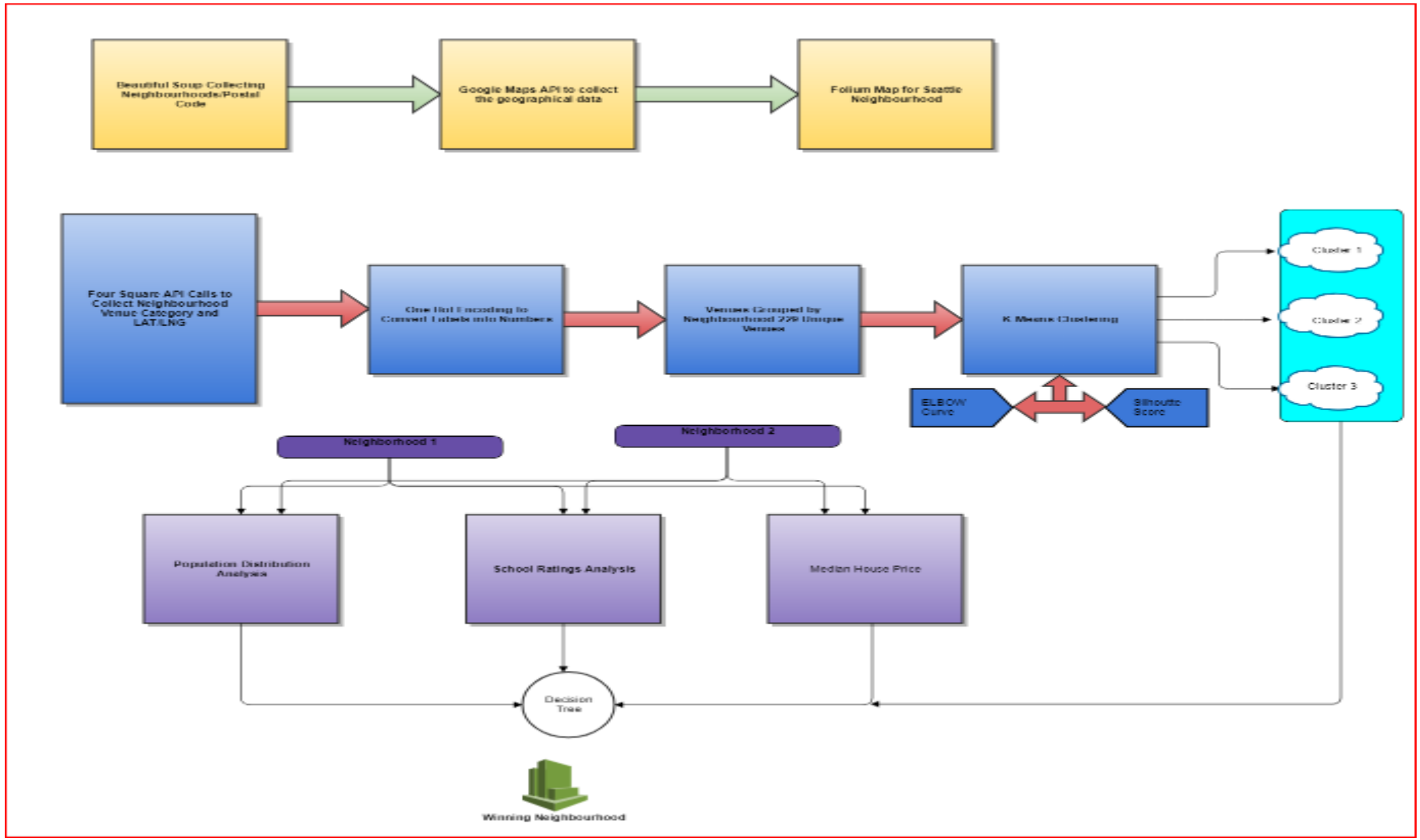
AVERAGE SCHOOL RATINGS



MEDIAN HOUSE PRICE



Complete Workflow



Comparison between Neighborhoods

Bellevue and Kirkland

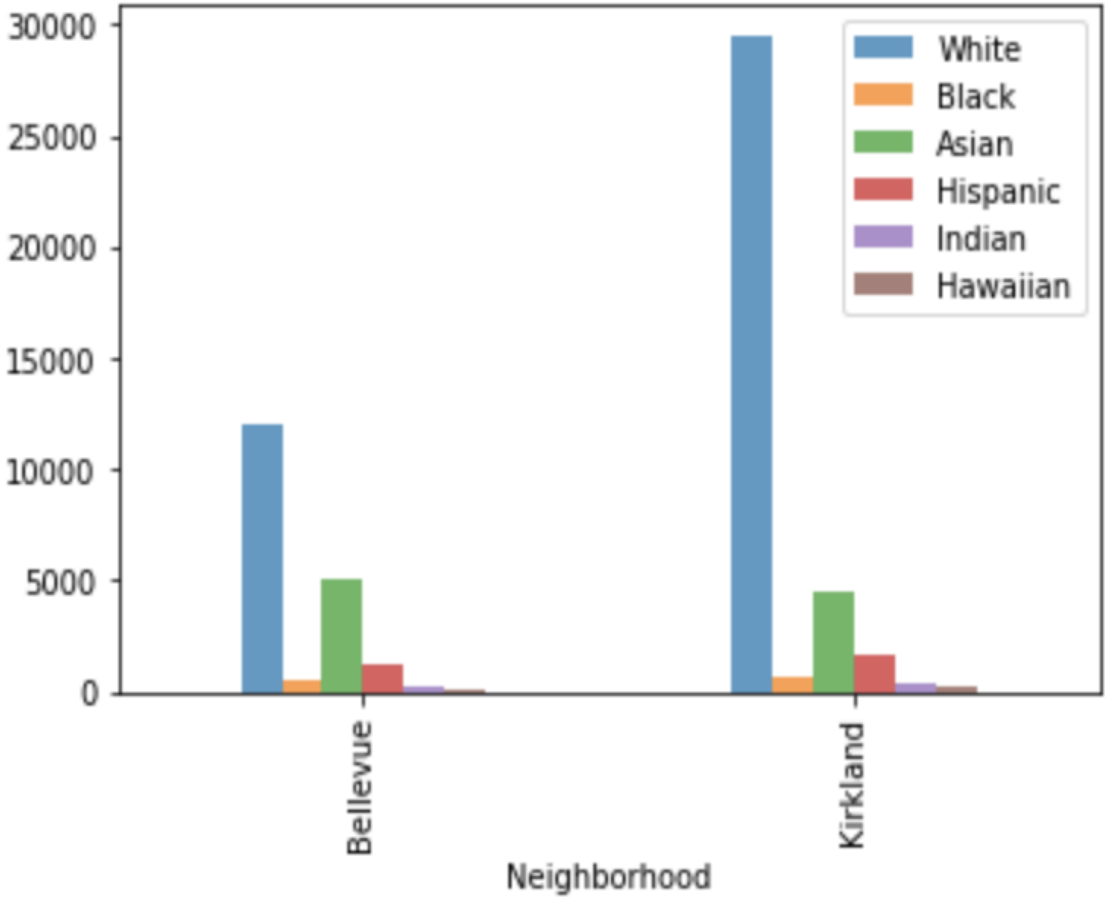
Now lets compare 2 neighborhoods to choose one that best matches our requirements as given below

1. More Indian Population
2. Higher School Rating
3. Reasonable Housing Price in the Range of 300k to 500k
4. Comfortable Neighborhoods

Neighborhood Venues

Neighborhood	Bellevue	Kirkland
PostalCode	98005	98033
Latitude	47.615	47.6688
Longitude	-122.172	-122.192
Cluster Labels	2	0
1st Most Common Venue	Indian Restaurant	Café
2nd Most Common Venue	Vietnamese Restaurant	Baseball Field
3rd Most Common Venue	Bagel Shop	Music Venue
4th Most Common Venue	Automotive Shop	Sandwich Place
5th Most Common Venue	Mexican Restaurant	Asian Restaurant
6th Most Common Venue	Coffee Shop	Bus Stop
7th Most Common Venue	Other Repair Shop	Grocery Store
8th Most Common Venue	Fish & Chips Shop	Food & Drink Shop
9th Most Common Venue	Fish Market	Brewery
10th Most Common Venue	Filipino Restaurant	Forest

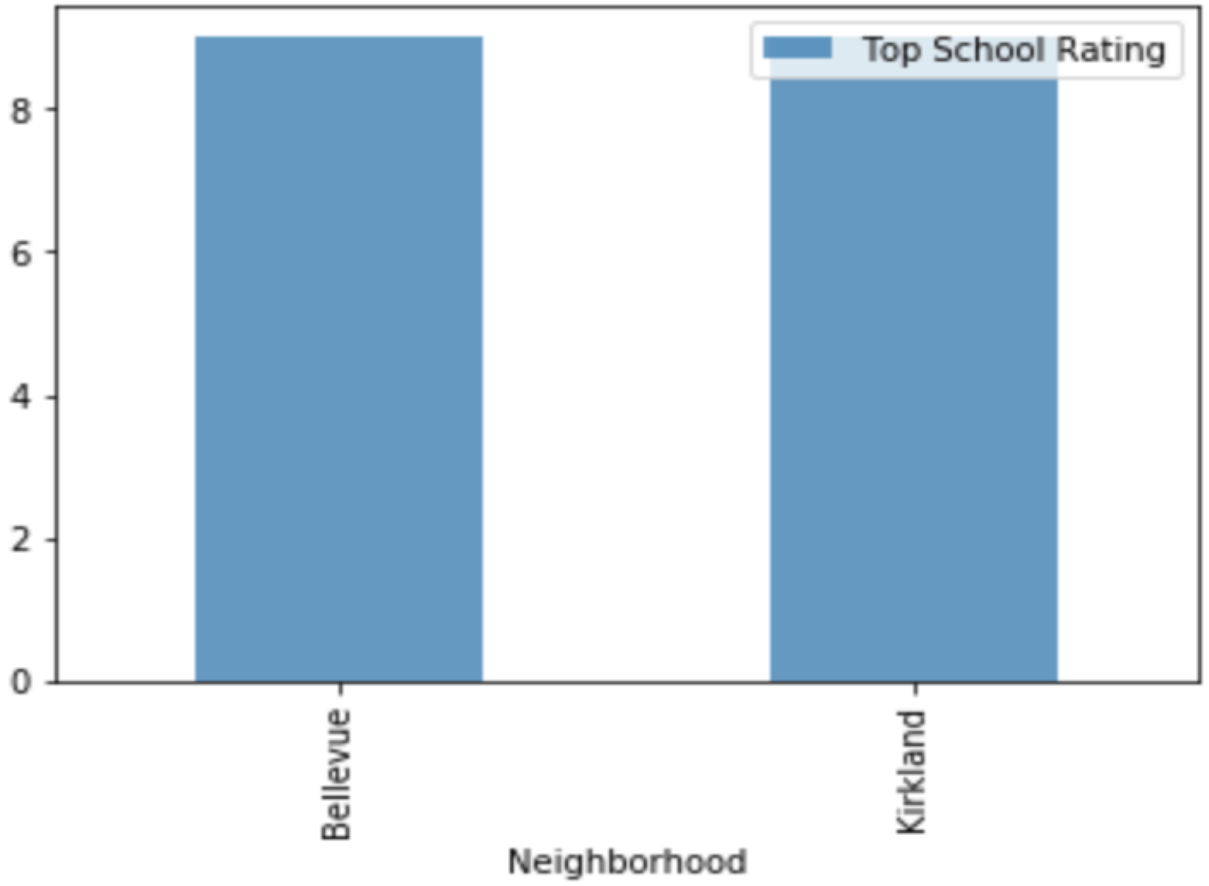
Population distribution




```
Population_Comparison['Indian']
```

```
Neighborhood
Bellevue     157
Kirkland     366
```

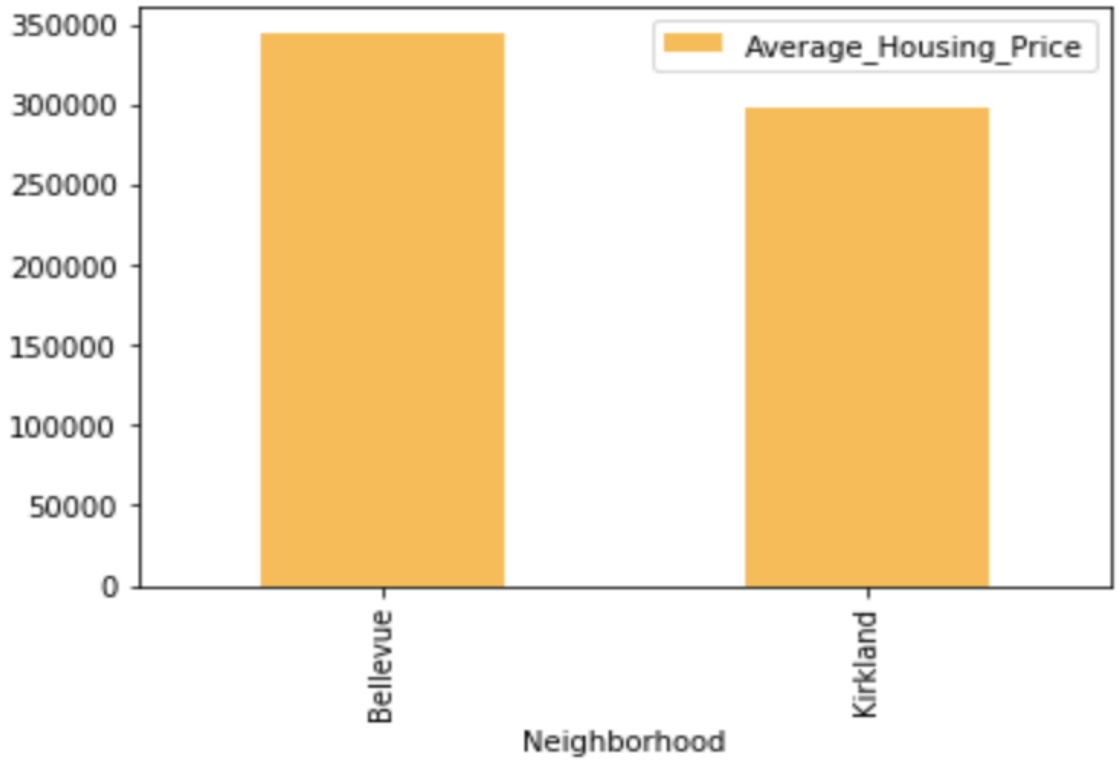
School ratings




```
School_rating_comparison
```

Neighborhood	Top School Rating
Bellevue	9
Kirkland	9

Average housing price



Neighborhood	Average_Housing_Price
Bellevue	343500.0
Kirkland	297900.0

conclusion

- This Analysis concludes that compared to Bellevue ,
  - Kirkland has the higher number of Indian population
  - Good school rating of 9
  - Reasonable average housing price of approximately 300k
  - Top 10 common venues shows Kirkland has gotten a good neighborhood with Cafe, Restaurants, Sandwich Place, Baseball Field etc.

  Hence Kirkland wins over Bellevue!

Thank you