

A. Introduction

A.1. Description & Discussion of the Background

Seattle is a seaport city on the west coast of the United States, and is the largest city in Washington State. With 2 big tech companies (Amazon and Microsoft) headquartered in the area, together with a continuously growing number of tech companies increasing their footprint in the city, Seattle is becoming one of the fastest growing tech hubs in the United States ^[1]^[2]. With its lower cost of living than the San Francisco Bay Area ^[3] and its low to no income tax ^[4], it is becoming a more and more attractive city for software developers to further pursue their career.

One of the first things to consider when moving to a new city is to find a comfortable environment to settle in. When searching for a place to live, several important factors to consider include safety, convenience, affordability, etc. For example, a neighborhood with low crime rates, good schools, plenty of recreational venues (e.g. parks), and convenient access to public transport is generally considered a desirable district to live in.

However, the desirable qualities of a neighborhood can be subjective. For instance, young professionals may prefer a more vibrant and diverse neighborhood with lots of entertainment venues, whereas those with families may prefer a quiet neighborhood with a good school district. Hence, maps and information charts showing the distributions of various types of venues and crime data can be helpful for determining the best suitable neighborhood to settle in based on individual criteria.

This information will be very useful for those who are new to the area looking to settle down, as well as for those who are considering making a big move to Seattle.

A.2. Data Description

The following data have been collected and will be used for further analysis:

- A list of neighborhoods in Seattle obtained from Wikipedia ^[5]
 - This data will be used for identifying the individual neighborhood in Seattle in order to further explore the details of each neighborhood. A map of Seattle superimposed with the different neighborhoods together with their corresponding

venue category and density information will be useful for identifying desirable districts to live in.

- The latest crime data from January 2020 to March 2020 published by the Seattle Police Department ^[6]
 - This data will be used to create a choropleth map of crime rates in Seattle, showing the intensity and categories of crimes in each neighborhood. This information will be useful for determining the safety level of the various neighborhoods in Seattle.
- List of venue information in each Seattle neighborhood obtained by using Foursquare API ^[7]
 - This data will be used to find the top 10 most common venues for each neighborhood to classify neighborhoods with distinct characteristics, which will be useful for highlighting the particular neighborhoods that are best suited according to personal preferences.

B. Methodology

B.1. Exploratory Data Analysis for Seattle Neighborhood Venues

In order to obtain the list of neighborhoods in Seattle for further clustering, the data was first scraped from Wikipedia using the “Beautiful Soup” Python package. As the resulting data contained other additional information for each neighborhood (e.g. “Locator Map”, “Street Map”, “Street Image”, “Notes”, etc) which were not of interest for this project, only the “Neighborhood” and “District” data was extracted for further processing.

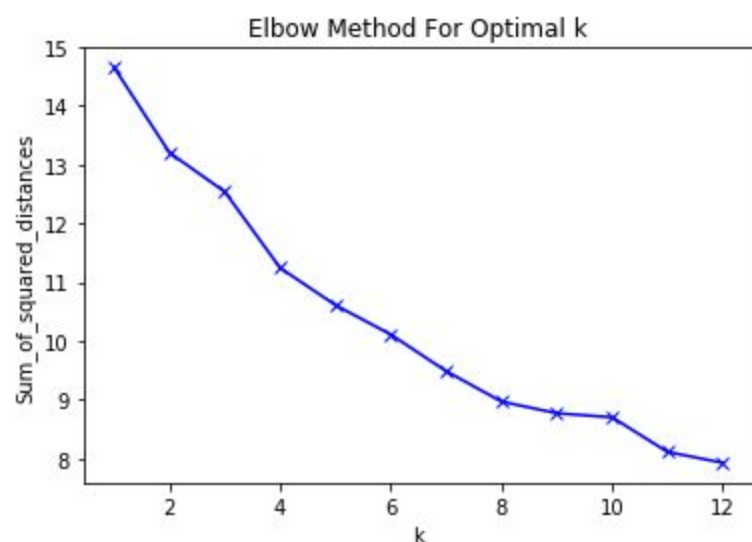
The neighborhoods and district data was then cleaned and formatted to be ready for later handling. For instance, the in-text citation in the neighborhood’s name was removed, overlapping and duplicated neighborhoods were consolidated, etc. In order to retrieve the various venue information for each neighborhood, its corresponding latitude and longitude data needs to be acquired. Using the “Nominatim” geolocator, the geographical coordinates of each neighborhood was obtained and merged with the neighborhoods and district dataset.

	Neighborhood	District	Latitude	Longitude
0	Broadview	North Seattle	47.722320	-122.360407
1	Bitter Lake	North Seattle	47.726236	-122.348764
2	North Beach	North Seattle	47.696210	-122.392362
3	Crown Hill	North Seattle	47.694715	-122.371459
4	Greenwood	North Seattle	47.690981	-122.354877

Given the list of neighborhoods with its corresponding location information, the Foursquare API was used to retrieve 100 nearby venues for each neighborhood that are within a radius of 500 meters. Each neighborhood was then analyzed to find the top 10 most common venues.

	Neighborhood	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	Adams	Bus Station	Dog Run	Gym Pool	Soccer Field	Park	Food	Flower Shop	Fish Market	Fish & Chips Shop	Food & Drink Shop
1	Alki Point	Beach	Lighthouse	Baseball Field	Baseball Stadium	Frozen Yogurt Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Field	Financial or Legal Service
2	Arbor Heights	Spa	Professional & Other Places	Zoo Exhibit	Fair	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Field	Financial or Legal Service	Fish & Chips Shop
3	Belletown	Hotel	Bakery	Bar	Seafood Restaurant	New American Restaurant	Sushi Restaurant	Pizza Place	Deli / Bodega	Movie Theater	Marijuana Dispensary
4	Bitter Lake	Fast Food Restaurant	Gym / Fitness Center	Food	Seafood Restaurant	Sandwich Place	Pet Store	Noodle House	Mobile Phone Shop	Hotpot Restaurant	Hotel

In order to use k-means clustering to segment neighborhoods with similar characteristics, the elbow method was used to identify the optimal k to use for k-means. Based on the plotted graph, a k of 10 was chosen.



B.2. Exploratory Data Analysis for Seattle Crimes

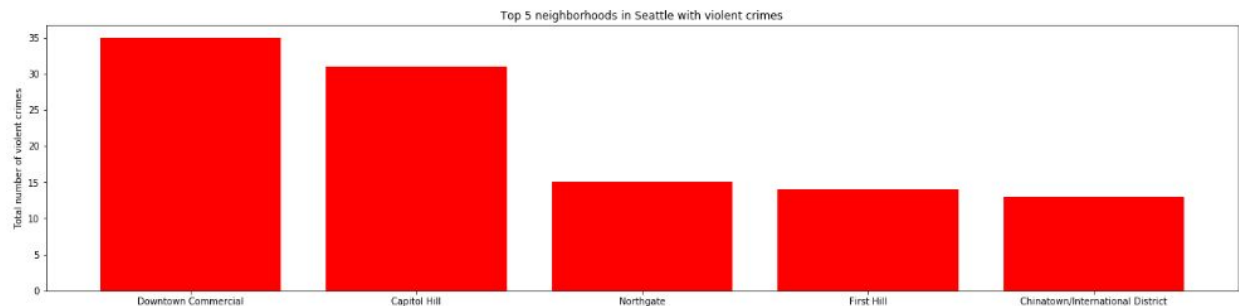
To obtain the latest crime data in Seattle, the 2020 crime dataset published by the Seattle Police Department was downloaded. Only the relevant columns of data (i.e. “Neighborhood”, “Category”, “Offense”) were extracted and the resulting dataset was further cleaned to drop “unknown” neighborhoods and properly capitalize and format neighborhood names.

	Neighborhood	Category	Offense
0	Bitterlake	Violent Crime	Rape
1	University	Violent Crime	Aggravated Assault
2	University	Violent Crime	Aggravated Assault
3	University	Violent Crime	Aggravated Assault
4	Capitol Hill	Violent Crime	Aggravated Assault

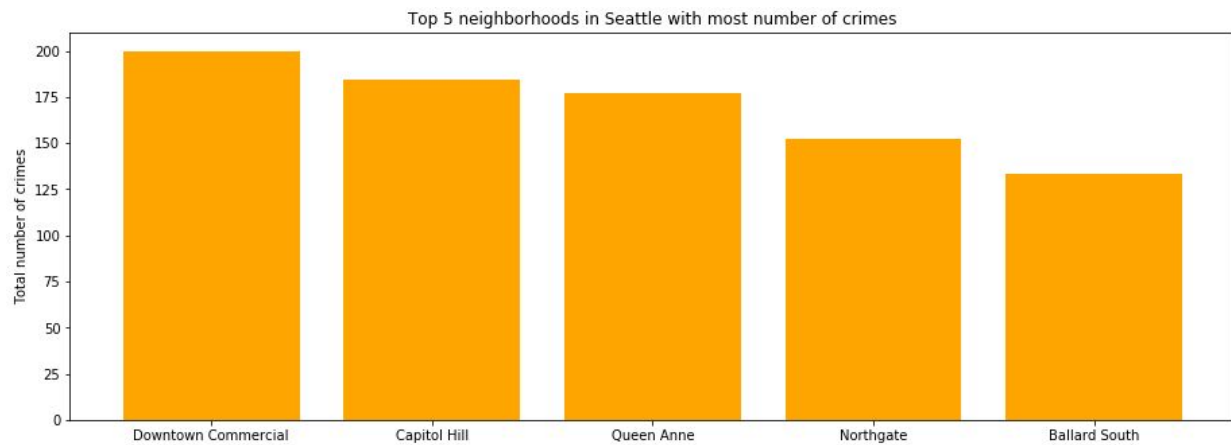
To explore the safety of different neighborhoods, the list of crime incidents were grouped by their category into “Violent Crime” and “Property Crime”.

	Neighborhood	Category	Offense
0	Alaska Junction	Property Crime	120
1	Alaska Junction	Violent Crime	8
2	Alki	Property Crime	32
3	Alki	Violent Crime	1
4	Ballard North	Property Crime	132

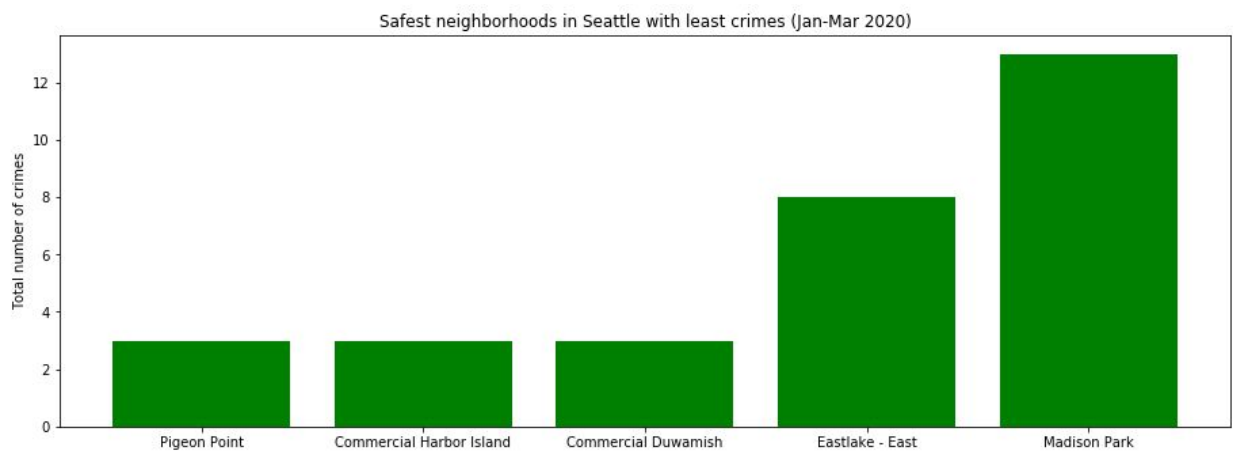
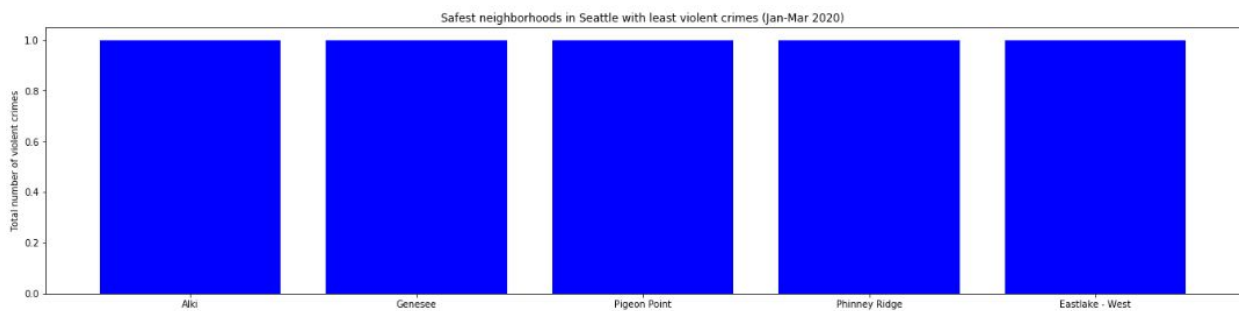
The top 5 neighborhoods with the most violent crimes were identified and plotted as follows.



The top 5 neighborhoods with the most total number of crimes (in all categories) were also identified and plotted as follows.

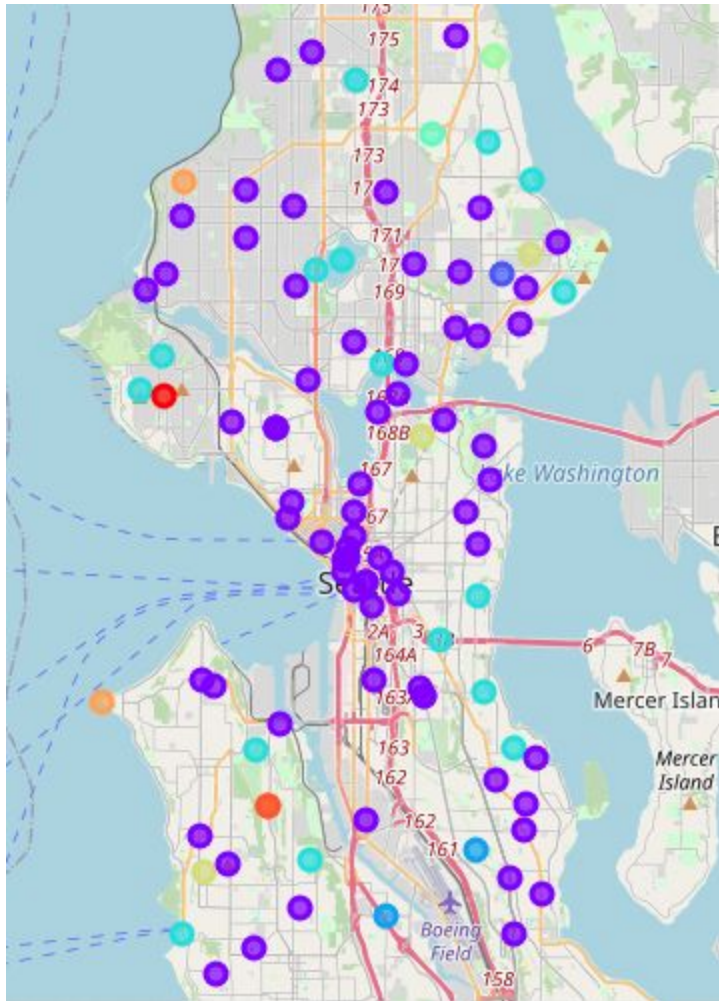


In contrast, the safest 5 neighborhoods with the least violent crimes and least total number of crimes were also identified and plotted as follows.

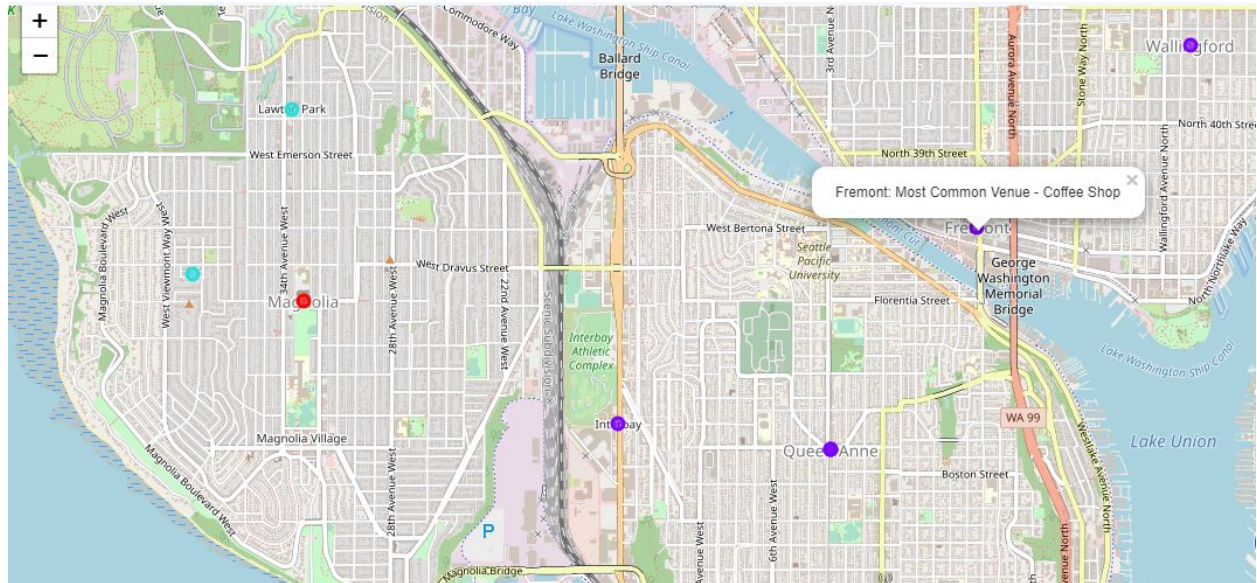


C. Results

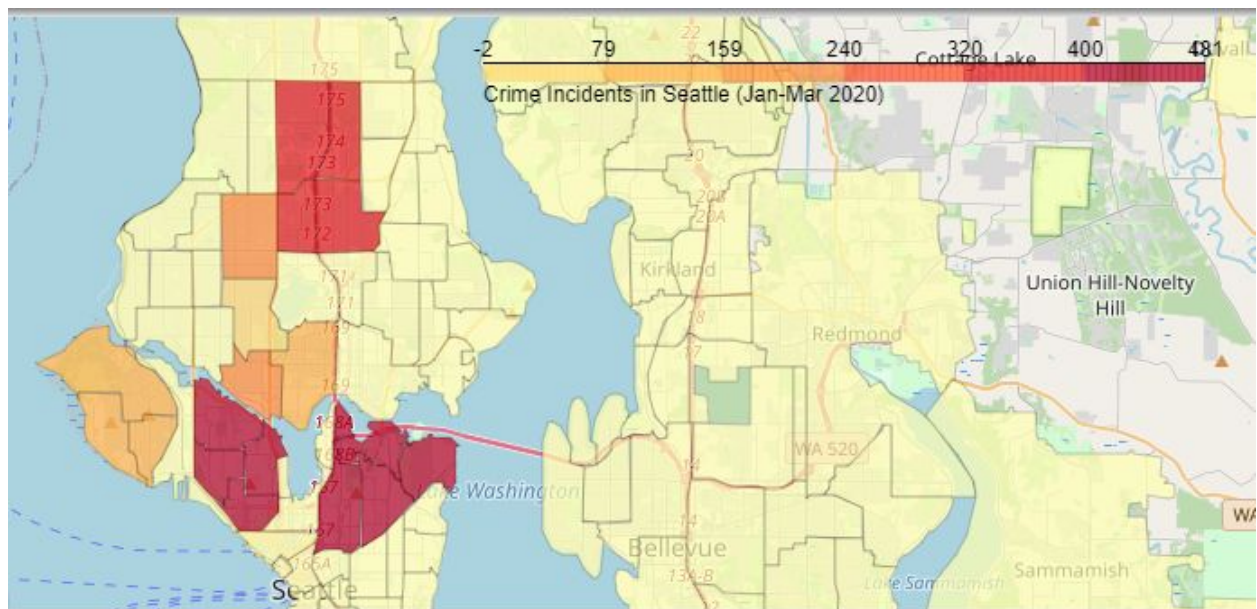
Using k-means clustering, the different neighborhoods in Seattle were segmented into 10 clusters with similar characteristics. The resulting clusters can be visualized with the plotted folium map below, with markers of same color representing the same cluster.



To further explore the details of each neighborhood, the map can also be zoomed in and contains information of each neighborhood with corresponding popup labels indicating the most common venue in that area. This is useful for highlighting the characteristics of the neighborhood which will help decide whether the area is desirable according to personal preferences.



In order to illustrate the distribution of crimes across various neighborhoods, a choropleth map was generated as follows, where neighborhoods with higher color intensity indicate greater levels of crimes. This is helpful for highlighting which neighborhoods are particularly dangerous relative to other neighborhoods, and indicating corresponding neighborhoods that are adjacent to the dangerous zones.



To specifically highlight the districts with most violent crimes (as violent crime incidents are particularly alarming), another choropleth map was generated with solely violent crime data.

someone who is interested in particular types of crime. For example, women are likely to be more concerned with incidents of rape than men.

- Housing price is also a common criteria for deciding where to live. Hence it would be very helpful to obtain the average housing price information for each neighborhood and display a corresponding choropleth map with the data.

E. Conclusion

The desirable qualities of the best neighborhoods to live in can be subjective and based on personal preferences. Some common criteria are safety and types of venues in the nearby area.

This report gathered the latest 2020 crime dataset published by the Seattle Police Department to identify the safety level of the different neighborhoods in Seattle, and used Foursquare API to explore the top venues in each neighborhood to segment and cluster similar neighborhoods.

Resulting maps were generated illustrating the distribution of crime levels in each neighborhood, as well as clusters of neighborhoods with similar characteristics.

Given additional time, further enhancements can be made to better analyze each neighborhood in more detail, such as classifying neighborhoods according to specific interests, breaking down the crime incidents into more granular categories, and obtaining the average housing price information for each neighborhood.

These results would be useful for people who are looking to settle down in Seattle, as well as for those who are considering making a big move to the area.

References

[1] *Seattle, The Fastest Growing Tech Hub*, viewed April 11, 2020
<<https://www.mobiledeveloper.net/blog/seattle-fastest-growing-tech-hub>>

[2] *Report: Seattle Leads Nation as Fastest Growing Tech Hub*, viewed April 11 2020,
<<https://www.lightercapital.com/blog/report-seattle-leads-nation-as-fastest-growing-tech-hub/>>

[3] *2020 Cost of Living Calculator: Seattle, Washington vs San Francisco, California*, viewed April 11 2020, <<https://www.bestplaces.net/cost-of-living/seattle-wa/san-francisco-ca/115000>>

[4] *Washington state Supreme Court denies Seattle's bid for income tax on wealthy households*, The Seattle Times, viewed April 11 2020,

<<https://www.seattletimes.com/seattle-news/politics/washington-state-supreme-court-denies-seattles-bid-for-income-tax-on-wealthy-households/>>

[5] *List of neighborhoods in Seattle*, Wikipedia, viewed April 11 2020,
<https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle>

[6] *Crime Dashboard*, Seattle Police Department, viewed April 12 2020,
<<https://www.seattle.gov/police/information-and-data/public-data-sets>>

[7] Foursquare API for Developers, <<https://developer.foursquare.com/>>