Business Problem

Moving into a big city like Seattle can be overwhelming and time consuming. Recent graduates from college tend to have a limited budget and time to find the right apartment to move in--a person can spend days researching about specific neighborhoods and decide about whether renting an apartment or house within a specific neighborhood, is a good option.

This project takes this requirement as its main idea to help users reduce wasted time trapped in endless search engines. More specifically, how a college student with a limited budget can access real time information of a neighborhood within Seattle.

This project will focus on using average rent and housing prices and demographic info from rentcafe.com to determine what kind of neighborhood is available for a recent graduate versus a person who is more established financially. The foursquare API will assist in showing the most common venues of the neighborhoods.

The project will finally use K-means clustering as the machine learning algorithm to cluster the venues based on the foursquare place categories. This would give a better understanding of the kind of venues present in each of the neighborhoods and compare them with the demographic info.

Data

Foursquare API:

This project will use Foursquare API as its prime data gathering source as it has a database of million places, especially their places API which provides the ability to perform location search, location sharing and details about local restaurants.

https://foursquare.com/developers/verify

Rentcafe:

Rentcafe is another API that uses geo location to map median prices of rentals in different communities. Like Foursquare, data is fed by the users who rent in each of the neighborhoods, providing accurate and real time data.

https://www.rentcafe.com/

Seattle Neighborhood Postal Codes:

Using the Beautifulsoup library, I will scrape the codes for each of the neighborhoods analyzed in this project.

Methodology

Used BeautifulSoup library to collect postalcodes from Seattle.

```
#Collecting Neighborhoods and their Postalcodes by web scraping from Seattle webpage
url='http://seattlearea.com/zip-codes/'
source = requests.get(url).text

soup = BeautifulSoup(source, 'lxml')#Beautiful Soup to Parse the url page
divi = soup.find('div',attrs={'class': 'entry-content'})
```

```
l=[]
for i in divi:
    l.append(i)
```

```
neighborhood=[]
zipcode=[]
for b in l[11:]:

    try:
            k=b.text.split()
            if len(k)==5:
                zipcode.append(k[1])
                neighborhood.append(k[3]+' '+k[4])

            else:
                zipcode.append(k[1])
                neighborhood.append(k[3])

    except:
           pass
```

5]:

	PostalCode	Neighborhood
0	98003	Federal Way
1	98005	Bellevue
2	98033	Kirkland
3	98037	Lynnwood
4	98040	Mercer Island
5	98052	Redmond
6	98055	Renton

Obtaining the venue info from Foursquare and getting them into a json file:

```
sea_latitude=47.6062
sea_longitude=-122.3321
url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={
```

```
#Getting the results into a json file
results = requests.get(url).json()
results
```


```
{'meta': {'code': 200, 'requestId': '5c0dbeab4c1f671924b19e5f'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': '$-$$$$', 'key': 'price'},
    {'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Seattle Central Business District',
  'headerFullLocation': 'Seattle Central Business District, Seattle',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 87,
  'suggestedBounds': {'ne': {'lat': 47.6107000045, 'lng': -122.32543810034119},
   'sw': {'lat': 47.6016999955, 'lng': -122.3387618996588}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended'.
```


Preprocessing the data using the one hot encoder to get it ready to create the K-cluster algorithm.

```
# one hot encoding
sea_onehot = pd.get_dummies(sea_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
sea_onehot['Neighborhood'] = sea_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [sea_onehot.columns[-1]] + list(sea_onehot.columns[:-1])
sea_onehot = sea_onehot[fixed_columns]

sea_onehot.head(10)
```

	Neighborhood	ATM	Accessories Store	Airport	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Workshop	BBQ Joint	Bagel Shop	Bakery	Bank	Ba
0	Federal Way	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	(
1	Federal Way	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	(
2	Federal Way	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	(
3	Federal Way	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	(

K-means Clustering. For this project, I decided that using 5 clusters should be enough for the analysis and amount of neighborhoods anlaized. The amount of venues was also limited the 10 most common.

```
[30]: # import k-means from clustering stage
from sklearn.cluster import KMeans

# set number of clusters
kclusters = 5

sea_grouped_clustering = sea_grouped.drop('Neighborhood', 1)


# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(sea_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

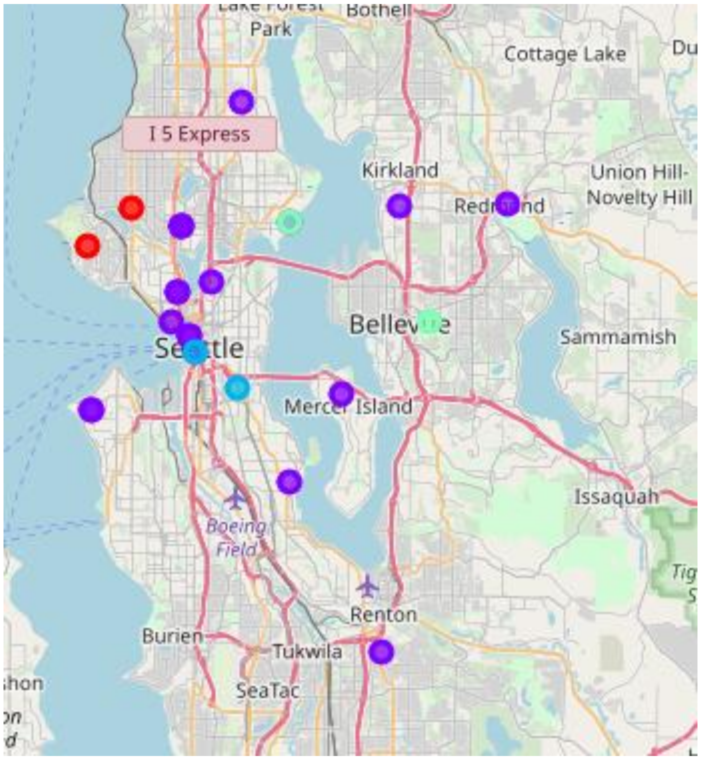
```
[30]: array([0, 3, 1, 4, 1, 1, 1, 1, 1, 1], dtype=int32)
```

```
[31]: sea_merged = coor_sea

# add clustering labels
sea_merged['Cluster Labels'] = kmeans.labels_

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
sea_merged = sea_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

sea_merged
```

[31]:

	Unnamed: 0	Neighborhood	PostalCode	Latitude	Longitude	Rent in US Dollars	Cluster Labels	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
0	0	Federal Way	98003	47.316504	-122.322397	1349	0	Coffee Shop	Mexican Restaurant	Gym / Fitness Center	Grocery Store	Yoga Studio	Eastern European Restaurant	Fish Market

A map of each of the clusters created is displayed the visualized how each of the neighborhoods was classified according to the venues.



Clusters created for the neighborhoods based on the venues.

```
[35]: sea_merged.loc[sea_merged['Cluster Labels'] == 0, sea_merged.columns[[1] + list(range(5, sea_merged.shap
```

[35]:

	Neighborhood	Rent in US Dollars	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Federal Way	1349	0	Coffee Shop	Mexican Restaurant	Gym / Fitness Center	Grocery Store
17	Ballard	2039	0	Brewery	Coffee Shop	Food Truck	Vietnamese Restaurant
28	Magnolia	3200	0	Park	Bus Stop	Trail	Athletics & Sports

```
[36]: sea_merged.loc[sea_merged['Cluster Labels'] == 1, sea_merged.columns[[1] + list(range(5, sea_merged.shap
```

[36]:

	Neighborhood	Rent in US Dollars	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
2	Kirkland	1849	1	Café	Baseball Field	Asian Restaurant	Bus Stop
4	Mercer Island	2168	1	Coffee Shop	Sandwich Place	Pizza Place	Thai Restaurant
5	Redmond	1937	1	Furniture / Home Store	Bakery	Thai Restaurant	Accessories Store
6	Renton	1633	1	Coffee Shop	Pizza Place	Bar	Yoga Studio
7	Seattle	2134	1	Coffee Shop	New American Restaurant	American Restaurant	Hotel

Discussion and Results

Loading demographic data on housing prices as well as ethnical distribution within each neighborhood, we can enrich the analysis provided by the k-means algorithm. Looking at two opposite neighborhoods in terms of rent prices, we can start analyzing the type of neighborhood a person could live and what venues he/she will have access to.

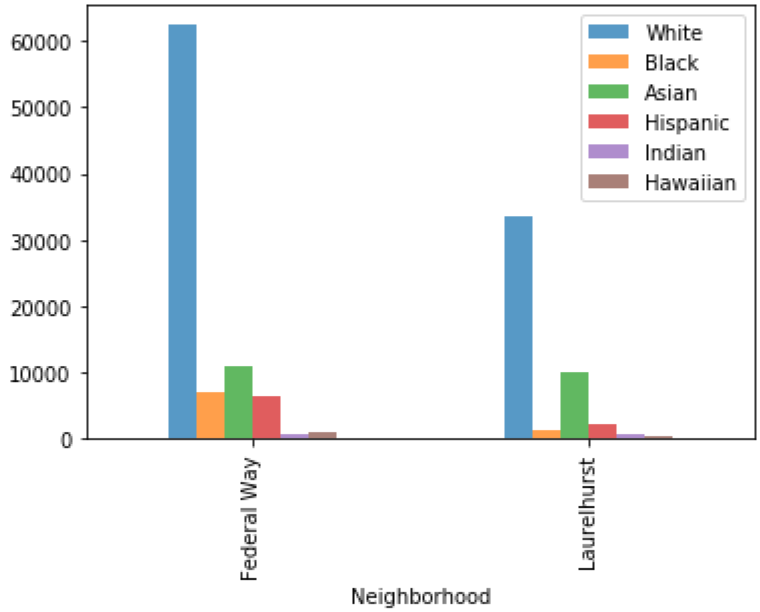
In this case, we will compare Federal way, with an average rent price of $1349 per month and an average housing price of $172000 against Laurelhurst, with an average rent price of $3200 and $370000 of average housing price.

Looking at the venues for each of the neighborhoods, the only clear difference is the proximity to the beach of Laurelhurst, which could be the biggest factor contributing to the high housing prices and hence rent. Within the same cluster as Lauelhurst was Bellevue, that had the "Health and Beauty Service" as the most common venue. These type of venues demonstrate that the type of people living in the area belong to a higher socioeconomic status, where more luxury based venues are frequented.

[81]:

Neighborhood	Federal Way	Laurelhurst
Unnamed: 0	0	16
PostalCode	98003	98105
Latitude	47.3165	47.6614
Longitude	-122.322	-122.268
Rent in US Dollars	1349	3200
Cluster Labels	0	3
1st Most Common Venue	Coffee Shop	Beach
2nd Most Common Venue	Mexican Restaurant	Yoga Studio
3rd Most Common Venue	Gym / Fitness Center	Food Truck
4th Most Common Venue	Grocery Store	Flower Shop
5th Most Common Venue	Yoga Studio	Fish Market
6th Most Common Venue	Eastern European Restaurant	Filipino Restaurant
7th Most Common Venue	Fish Market	Fast Food Restaurant
8th Most Common Venue	Filipino Restaurant	Farmers Market
9th Most Common Venue	Fast Food Restaurant	Falafel Restaurant
10th Most Common Venue	Farmers Market	Event Space

If we look at socio demographic information, we can see a clear difference between the diversity of population living in each of the neighborhoods. While Laurelhurst tends to be predominantly white, Federal Way does have more representation of minority groups like Hispanic, Black and Asian. Something does worth recalling is the high number of White people living in the neighborhood, which shows that the area is not dominated by a specific sociodemographic group. This could potentially be the case of a group of white people under employed or recently graduated from college who cannot afford paying three times the rent of another neighborhood.



Conclusion

Based on the comparison of the two neighborhoods Laurelhurst and Federal Way, I recommend a recent graduate or low-income worker to rent in Federal Way. Laurelhurst and other neighborhoods in a similar cluster rent an apartment in the area due to its proximity to the beach and venues that could be considered luxurious and extra. For surviving, the venues located in Federal Way and other neighborhoods within that cluster serve different basic necessities ranging from meals, to groceries to access to public transportation.