## **Introduction: Business Problem**

The objective of this project is to determine a location for a new business within the municipality of Denver, Colorado USA. This business will be a Wine Bar. It will feature a tasting menu as well as a variety of interesting wines from around the world. Wines will be available by the bottle, glass, or in flights. Craft beer has been popular in Colorado for a long time, but wine is also popular. Many of the wine bars are outside of Denver in the surrounding suburbs.

Over the past ten years or so there have been many, trendy neighborhoods that have been attracting more affluent millennials and others who enjoy living in a bustling urban area close to parks and quality restaurants.

The stakeholders want to capitalize on these trends and place a wine bar in one of these hot, trending neighborhoods. They would also consider a more established neighborhood with the correct demographics that is lacking a similar business.

## Data

The data to be considered for this project will come from the following sources:

- A list of 25 distinct Denver Neighborhoods found in an article called The 25 Best Neighborhoods in Denver
- An article in the popular local magazine, 5280, called The 25 Best Neighborhoods in Denver
- Data downloaded utilizing the Foursquare API including: most popular venues and locations of wine bars
- Neighborhood Latitudes and Longitudes will be acquired using the Nominatim package from geopy.geocoders.
- Google Maps was used to fill in missing or incorrect data as discovered in the initial exploratory analysis.

The data will be used to determine which of these neighborhoods already have wine bars, and which one would be favorable for the stakeholder's project. Popular venue lists from Foursqaure will be used gain an understanding on what these neighborhoods are like and what

5280's data comes from the Denver Police Department web page, American Community Survey's 2013-2017 neighborhood data, and School Performance Framework. Their data was determined by using the following criteria quoted below from the article:

"Our ranking uses four variables: home prices, crime data, school rankings, and an X factor score that accounts for things that can't be easily quantified, such as access to open space, nearby public transportation, and the prevalence of restaurants and shops. Each category is weighted: 30 percent for year-over-year percentage increase in home values; 25 percent for safety; 15 percent for neighborhood school ratings; and 30 percent for the X factor."

This report will include maps of all of the Denver neighborhoods overlaid with the wine bar locations. It will use the venue data to find a location near other sit-down restaurants but not too close to any of the existing wine bars.

Kmeans clustering based on the top ten most popular venues for each neighborhood will be used to narrow down neighborhood candidates.

The criteria for selecting a neighborhood for a new wine bar will include:

- Trending neighborhood mandatory (all of them meet this criteria)
- Near a wine shop
- Not in a neighborhood with many fast food restaurants or discount stores mandatory
- In a cluster that contains a neighborhood another wine bar mandatory (clusters TBD in the Analysis section)
- In a neighborhood with other bars or breweries
- Doesn't already have a wine bar mandatory
- Near a park

Here is the URL for the neighborhood list and data:

https://www.5280.com/neighborhoods/

The documentation for the Foursquare API is here:

https://developer.foursquare.com/docs/places-api/

I used Python's Beautiful Soup package to scrape the list of neighborhoods off of the 5280 article. I noticed it was contained in the table so I looped around to populate lists to create a data frame. I included some of their demographic data to get a little bit better picture of the neighborhoods. I ended up only using the list of names for the scope of this project. The end result was this table:

Out[408]:	Out	[40	8]	:
-----------	-----	-----	----	---

	Neighborhoods	Rank	AvgSalePrice2019	CrimeRank	XFactorScore
0	South Park Hill	1	\$ 804,250	8	8.5
1	Washington Park	2	\$ 1,119,585	3	9
2	Congress Park	3	\$ 680,522	15	9
3	West Highland	4	\$ 661,257	25	8.5
4	Cherry Creek	5	\$ 1,165,333	52	8.5
5	Speer	6	\$ 505,815	44	8
6	Wellshire	7	\$ 812,084	1	7
7	Highland	8	\$ 700,576	51	9.5
8	Hilltop	9	\$ 983,055	4	8
9	University Hills	10	\$ 596,061	46	6.5
10	Berkeley	11	\$ 651,844	22	9
11	Union Station	12	\$ 823,351	72	9.5
10	Indian Creek	13	¢ 335 564	2	4.5

After I had the scrape of the table including the list of the neighborhoods I needed to acquire the longitude and latitude for mapping purposes. To look the up using the Nominatim package from geopy.geocoders I had to append the addresses with Denver, CO. I also understand that Denver neighborhoods are evolving and changing constantly so I anticipated that this search would not return all neighborhoods.

#### Append Denver, CO to each neighborhood in order to acquire the latitude and longitudes

```
In [409]: addresses = []
          for i in top_25['Neighborhoods']:
             address = i + ', Denver, CO
             addresses.append(address)
          len(addresses)
Out[409]: 25
```

#### Acquire the neighborhood coordinates using the Novinatim API - create a list of all neighboorhoods that were not found

```
In [410]: Latitudes = []
          Longitudes = []
          found = []
          not_found = []
          for address in addresses:
                  geolocator = Nominatim(user_agent="denver explorer")
                  location = geolocator.geocode(address)
                  latitude = location.latitude
                  longitude = location.longitude
                 Latitudes.append(latitude)
                 Longitudes.append(longitude)
                 found.append(address)
              except AttributeError:
                 not_found.append(address)
          not found
Out[410]: ['Cory-Merrill, Denver, CO']
```

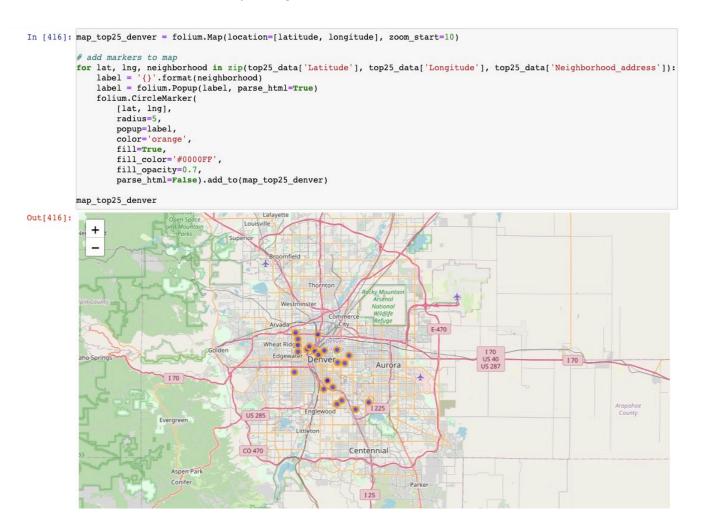
This resulted in only one neighborhood I needed to grab coordinates for on Google maps. Once that was added I used the pandas value counts to explore the latitudes and longitudes to make sure that Nominatim didn't duplicate anything.

Denver neighborhoods are changing rapidly. Verify that the Novinatim API identified unique neighborhoods and did not lump any together

```
In [413]: top25_data['Longitude'].value_counts().head()
Out[413]: -104.971034
           -104.931128
           -104.977986
           -105.043968
           -87.899960
          Name: Longitude, dtype: int64
In [414]: top25_data['Latitude'].value_counts().head()
Out[414]: 39.702081
           39.775231
           39.686780
          39.754658
           39.752426
          The Novinatim API has the same coordinates for Washington Park and Washington Park West.
          The correct coordinates will be updated with data from Google Maps
In [415]: top25_data.replace({'Latitude': {1: 39.697937 , 13: 39.7005822}})
          top25_data.replace({'Longitude': {1: -104.975186 , 13: -104.9980279}})
```

I found that Washington Park and Washington Park West have the same coordinates according to Nominatim, so I updated using Goole maps.

I then was able to create this map using Folium:



As you can see they are not all concentrated in the same area. This is why using the Foursquare and KMeans to cluster like neighborhoods is so critical to this project.

# Methodology

For this project we are going too need to determine what kind of venues are popular in each neighborhood. Acquiring a Foursquare developer account is mandatory for this project. The advantages of using the Foursquare API is that it has generous API call limits for the free version. Once the Foursquare account is set up and credentials are established for Client ID and Client Secret, then the venue data can be acquired. Here is a clip of the process using the GetNearbyVenues function created in the Coursera Capstone lab:

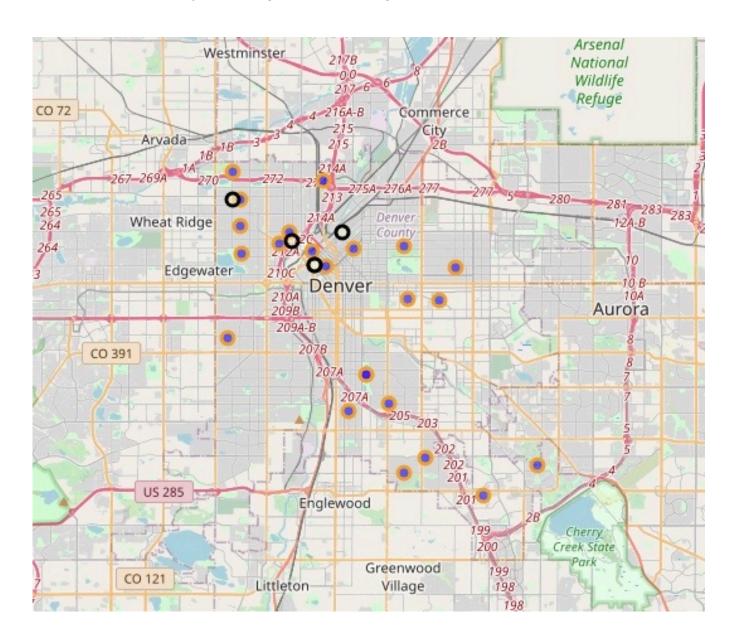
Once the venue data has been acquired and processed into a data frame, then a separate data frame can be created with only Wine Bars. The initial data frame had two wine bars with the exact same coordinates listed in two different neighborhoods. A Google maps search revealed which was the correct one and the duplicate was removed:

Part 3 - Find the wine bars in these neighborhoods

Berkeley, Denver, CO

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
287	Speer, Denver, CO	39.756877	-105.018439	The Truffle Table	39.758129	-105.011643	Wine Bar
412	Highland, Denver, CO	39.761583	-105.012500	The Truffle Table	39.758129	-105.011643	Wine Bar
591	Berkeley, Denver, CO	39.775231	-105.039261	BookBar	39.775213	-105.043888	Wine Bar
736	Union Station, Denver, CO	39.753630	-105.000748	Cru Wine Bar	39.747963	-104.998908	Wine Bar
907	Five Points, Denver, CO	39.754658	-104.977986	Mile High Winery	39.761417	-104.983637	Wine Bar
1237	Central Business District, Denver, CO	39,747378	-104.992737	Cru Wine Bar	39.747963	-104.998908	Wine Bar
There	e is a mistake on Foursquare. The	e Truffle Table, listed t	wice, is actually in the H	lighland neighb	orhood, not Sp	eer. Removing th	nat row.
wine	e is a mistake on Foursquare. The _bars = wine_bars.drop(ind bars	100 100 100 100 100 100 100 100 100 100	wice, is actually in the H	lighland neighb	orhood, not Sp	eer. Removing th	nat row.
wine	_bars = wine_bars.drop(ind	100 100 100 100 100 100 100 100 100 100	wice, is actually in the H	lighland neighb	orhood, not Sp	eer. Removing th	nat row.
wine	_bars = wine_bars.drop(ind _bars	ex=287)	wice, is actually in the h	lighland neighb		eer. Removing th	
wine	_bars = wine_bars.drop(ind _bars Neighborhood	ex=287)				-	Venue Category
wine	_bars = wine_bars.drop(ind _bars   Neighborhood   Highland, Denver, CO	ex=287)  Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category Wine Bar
wine wine	_bars = wine_bars.drop(ind _bars  Neighborhood  Highland, Denver, CO  Berkeley, Denver, CO	ex=287)  Neighborhood Latitude 39.761583	Neighborhood Longitude -105.012500	Venue The Truffle Table	Venue Latitude 39.758129	Venue Longitude -105.011643	Venue Category Wine Bar Wine Bar
wine wine 412	_bars = wine_bars.drop(ind_bars  Neighborhood  Highland, Denver, CO  Berkeley, Denver, CO  Union Station, Denver, CO	ex=287)  Neighborhood Latitude  39.761583 39.775231	Neighborhood Longitude -105.012500 -105.039261	Venue The Truffle Table BookBar	Venue Latitude 39.758129 39.775213	Venue Longitude -105.011643 -105.043888	

The Neighborhoods will be mapped with an overlay of the existing wine bars to get a better idea as to where they sit among the various neighborhoods as shown here:



The neighborhood and venue data will then be processed so that we can perform Kmeans clustering to group the neighborhoods into clusters based on their most popular venues. This was done using the one hot encoding method to reshape the data to get a count of each venue in each neighborhood as shown here:

```
In [426]: # one hot encoding
  denver_onehot = pd.get_dummies(top25_denver_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
  denver_onehot['Neighborhood'] = top25_denver_venues['Neighborhood']

# move neighborhood column to the first column
  col_name="Neighborhood"
  first_col = denver_onehot.pop(col_name)
  denver_onehot.insert(0, col_name, first_col)
  denver_onehot.head()
Out[426]:
```

	Neighborhood	ATM	Accessories Store	American Restaurant	Antique Shop	Aquarium	Arcade	Argentinian Restaurant		Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports		BBQ Joint
0	South Park Hill, Denver, CO	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	South Park Hill, Denver, CO	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	South Park Hill, Denver, CO	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	South Park Hill, Denver, CO	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	South Park Hill, Denver,	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Next, a new data frame is created to include a column each for the top ten venue types in each neighborhood.

#### Part 6 - Preprocess and perform kmeans clustering on the selected cluster from part 5

This is the cluster containing the neighborhoods with all of the existing wine bars

	Neighborhood	ATM	Accessories Store	American Restaurant	Antique Shop	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Automotive Shop
0	Barnum West, Denver, CO	0.000000	0.0	0.125000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0
1	Berkeley, Denver, CO	0.000000	0.0	0.034884	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0
2	Central Business District, Denver, CO	0.000000	0.0	0.070000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.010000	0.000000	0.0
4	Congress Park, Denver, CO	0.011494	0.0	0.022989	0.0	0.0	0.0	0.0	0.0	0.0	0.011494	0.011494	0.000000	0.0
5	Cory-Merrill, Denver, CO	0.000000	0.0	0.058824	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.029412	0.058824	0.0

Here is the processing for the Kmeans including methodology for the selection of the number of clusters and generating the cluster matrix:

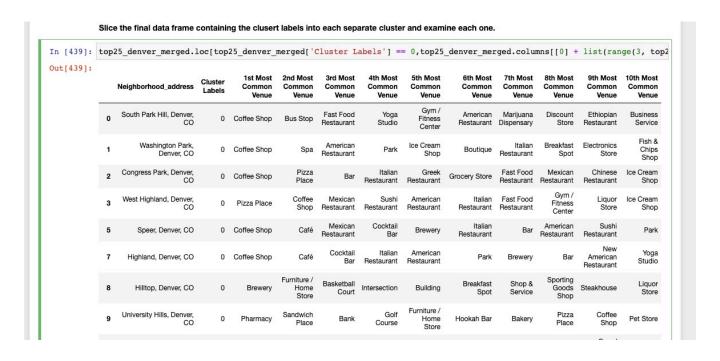
```
Create the dataset for the Kmeans. Loop through 25 times to create the elbow graph to select the optimum number of clusters
In [433]: X = denver_grouped.drop('Neighborhood', 1)
In [434]: wcss = []
           for i in range(1,25):
               kmeans = KMeans(n_clusters=i, init ='k-means++', max_iter=600, n_init=25,random_state=0)
               kmeans.fit(X)
               wcss.append(kmeans.inertia_)
In [435]: import matplotlib.pyplot as plt
           plt.plot(range(1,25),wcss)
           plt.title('The Elbow Method Graph')
           plt.xlabel('Number of clusters')
           plt.ylabel('WCSS')
           plt.show()
                             The Elbow Method Graph
             1.75
             1.25
           SS 1.00
             0.75
             0.50
             0.25
              0.00
                                  10
                                          15
           There is a bend in the elbow graph just past 5 so we will use 6 clusters.
In [436]: kmeans = KMeans(n_clusters=6, init = k-means++', max_iter=300, n_init=6,random_state=0)
           y_kmeans = kmeans.fit_predict(X)
           kmeans.labels [0:10]
Out[436]: array([0, 0, 0, 3, 0, 0, 5, 0, 0, 0], dtype=int32)
```

Next I created a new data frame with the cluster numbers inserted:

Insert the cluster labels into the combined neighborhood and top ten venue data frame. neighborhoods venues sorted.insert(0, 'Cluster Labels', kmeans.labels ) top25 denver merged = top25 data top25 denver merged = top25 denver merged.join(neighborhoods venues sorted.set index('Neighborhood'), on='Neighborhood' top25 denver merged.head() # check the columns! Out[494]: 9th Mo Commo South Park Hill, Denver, 39.746650 -104.922043 American Restaurant Marijuana Discount Ethiopia Bus Stop 39.702081 -104.971034 Boutique Congress Park, Denver, CO Fast Food 39.733720 -104.948367 Pizza Place Restaurant Gym / West Highland, Denver, 39.764466 -105.039271 Mexican Sushi American Italian Fast Food Restaurant Center Cherry Creek, Denver, 39.663610 -104.877444 Zoo Event Donut Electronics Ethiopian Gym Exhib

The next step will be to select all of the clusters with existing wine bars and combine them into one data frame. If the cluster contains at least 15 neighborhoods then we will re-cluster the data using Kmeans to narrow down the choices.

Finally, we will revisit the selection criteria highlighted in the Data section and make a final selection. Cluster one contains all of the neighborhoods that already have wine bars. There are 20 neighborhoods in this cluster.



I created a new data frame and then repeated the above steps to get a better feel for how these neighborhoods are grouped. There were three different clusters. One was by itself and already had a wine bar. There are two that contain neighborhoods with wine bars. These were examined for other necessary features. Here are the final two considered:

	Neighborhood_address	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Washington Park, Denver, CO	2	Coffee Shop	Spa	American Restaurant	Park	Ice Cream Shop	Boutique	Italian Restaurant	Breakfast Spot	Electronics Store	Fish & Chips Shop
2	Congress Park, Denver, CO	2	Coffee Shop	Pizza Place	Bar	Italian Restaurant	Greek Restaurant	Grocery Store	Fast Food Restaurant	Mexican Restaurant	Chinese Restaurant	Ice Cream Shop
7	Highland, Denver, CO	2	Coffee Shop	Café	Cocktail Bar	Italian Restaurant	American Restaurant	Park	Brewery	Bar	New American Restaurant	Yoga Studio
9	University Hills, Denver, CO	2	Pharmacy	Sandwich Place	Bank	Golf Course	Furniture / Home Store	Hookah Bar	Bakery	Pizza Place	Coffee Shop	Pet Store
10	Berkeley, Denver, CO	2	Coffee Shop	Brewery	Pizza Place	Park	Mexican Restaurant	Breakfast Spot	American Restaurant	Italian Restaurant	Gym / Fitness Center	Sushi Restaurant
11	Union Station, Denver, CO	2	American Restaurant	Cocktail Bar	Mexican Restaurant	Hotel	Pizza Place	Italian Restaurant	New American Restaurant	Restaurant	Bar	Lounge
13	Washington Park West, Denver, CO	2	Coffee Shop	Spa	American Restaurant	Park	Ice Cream Shop	Boutique	Italian Restaurant	Breakfast Spot	Electronics Store	Fish & Chips Shop
17	Southmoor Park, Denver, CO	2	Mexican Restaurant	Pizza Place	Coffee Shop	Sandwich Place	Grocery Store	Mediterranean Restaurant	Japanese Restaurant	Bar	Mobile Phone Shop	Deli / Bodega
18	Regis, Denver, CO	2	Brewery	Bar	Sporting Goods Shop	Bakery	Park	Outdoor Supply Store	Discount Store	Sandwich Place	Dog Run	Café

In [484]: cluster1\_new.loc[cluster1\_new['Cluster Labels'] == 2,cluster1\_new.columns[[0] + list(range(3, cluster1\_new.shape[1]))]]

10th Most Common Venue	9th Most Common Venue	8th Most Common Venue	7th Most Common Venue	6th Most Common Venue	5th Most Common Venue	4th Most Common Venue	3rd Most Common Venue	2nd Most Common Venue	1st Most Common Venue	Cluster Labels	Neighborhood_address	
Park	Sushi Restaurant	American Restaurant	Bar	Italian Restaurant	Brewery	Cocktail Bar	Mexican Restaurant	Café	Coffee Shop	4	Speer, Denver, CO	5
Cocktail Bar	Sandwich Place	Restaurant	Lounge	Steakhouse	Italian Restaurant	Theater	Coffee Shop	American Restaurant	Hotel	4	Central Business District, Denver, CO	21
Yoga Studio	Coffee Shop	Wine Shop	Art Gallery	Tennis Court	Garden Center	Business Service	Brewery	Park	Bagel Shop	4	Sloan Lake, Denver, CO	23

# **Results and Discussion**

### Narrowing down the potential neighborhoods

This new Wine Bar should be located in a neighborhood in the following criteria as mentioned in the data section:

- Trending neighborhood mandatory (all of them meet this criteria)
- Near a wine shop
- In a cluster that contains a neighborhood another wine bar mandatory
- Not in a neighborhood with many fast food restaurants or discount stores mandatory
- In a neighborhood with other bars or breweries
- Doesn't already have a wine bar mandatory
- Near a park

Out[484]:

We can already eliminate the following clusters:

- Cluster 1 Only one neighborhood which already has a wine bar
- Cluster 2 No wine bars in the cluster
- Cluster 4 Only one neighborhood and has many fast food restaurants and a discount store

Next we'll build a chart listing each neighborhood left over and how many criteria they have. Neighborhoods with wine bars already will not be considered.

Lets only consider those with all four of the mandatory criteria. Considering the above criteria the following neighborhoods should be considered. Here is the final table showing each neighborhood's score:

Neighborhood	Criteria
Washington Park	5
University Hills	4
Wash Park West	5
Southmooor Park	4
Speer	6
Sloan Lake	7

While all of these neighborhoods would be an excellent location based on their similarity with other neighborhoods with wine bars, only one had all seven criteria. The stakeholders were fairly particular on checking all of the boxes for this location. The reasoning behind their criteria is as follows.

- 1. They wanted a walkable neighborhood. Neighborhoods with parks are generally going to attract more foot traffic.
- 2. They wanted to attract people willing to sit down for a meal rather than wanting to grab a quick bite at a fast food restaurant.
- 3. They wanted to be located near other brewery or pub type restaurants to give customers a wider selection in alcohol based dining.
- 4. And, finally, they wanted to have wine shops near by so that they could partner with them for marketing purposes.

## Conclusion

# Sloan Lake is the winning neighborhood because it is the only one that has all seven of the criteria.

The stakeholders are satisfied with this selection. This neighborhood has several new developments that are a combination of business on the first floor and residential space on the upper floors. They will be selecting a space in one of these locations. They will make a

selection based on the availability of space for an outdoor patio and with good nearby parking.

Here is a snapshot of the map showing the location of this neighborhood. The black and yellow circles are the existing wine bars. The cluster color for this neighborhood cluster is red.

