# Battle of Neighborhoods – Where should construction of next Apartment complex happen in Austin, TX?

Reshma Kotwani

April 2021

## 1 Introduction

### 1.1 Background

Austin is capital of Texas in United States of America and is one of the fastest growing cities. According to Austin City Government, the City of Austin has crossed the threshold of becoming a Majority-Minority city, which means that no demographic group exists as most of the City's population. As per rankings by U.S. News and World Report, 2020, Austin is in 3rd spot for best city to live/retire. It has expanded even more in the past one year. Global pandemic has in fact accelerated its growth rather than slowing down as it was observed that more and more people and businesses have moved their base to Austin. Many people in tech industry prefer to call Austin as their home and have moved from expensive Bay Area and its enormous prices, be it in terms of houses, rents or in general a very high cost of living to relatively cheaper Austin which has a lot of offer in every aspect.

Austin really is unique: fascinating and quirky, creative, and outdoorsy, hugely musical – and, given its setting, bizarrely anomalous. Being a fast-growing city with diverse ethnicities, Austin is a great place for entrepreneurs to start and grow their businesses. The city is also well known for its outstanding food and great live music venues.

### 1.2 Business Problem

Construction in Austin is not slowing down. There are lot of big projects that are in pipeline and the opportunities are endless. With massive influx of young people in this city, there is an ever-growing demand to have more Apartment communities in great neighborhoods which have lot of amenities in and around. This project aims to help Builders & Contractors with suggestions on which Neighborhoods would be optimum for the next construction that would appeal to the youth.

## 2 Data Acquisition and Cleaning

## 2.1 Data Sources

Austin neighborhood data was fetched from the city's official open data portal. It had been provided by the Housing and Planning Department. This data includes Neighborhood name, geometric information and several other fields. Then Google's Geocoding API was used to get latitude and longitude for all the neighborhoods that were fetched earlier. Four square API was used to get top 100 venues that were in neighborhoods within a radius of 2000 meters. Finally, demographic information was collected from austintexas.gov (Table II) which tells about total population in different neighborhoods in Austin and also the percent distribution among different age groups. Data preparation will be discussed in next section in detail.

## 2.2 Data Cleaning

Data downloaded from official open portal had 103 rows. Below is a snapshot of data.

```
# Load Austin Neighborhood data. Source: https://data.austintexas.gov/Building-and-Development/Neighborhoods/a7ap-j2yt
df = pd.read_csv("C:/Users/reshma.v.kotwani/Downloads/Neighborhoods.csv")
df.head()
```

| | the_geom | FID | TARGET_FID | NEIGHNAME | SqMiles | Shape_Leng | Shape__Area | Shape__Length |
|---|---|---|---|---|---|---|---|---|
| 0 | MULTIPOLYGON (((-97.792307359674 30.4567073495... | 3 | 3 | ANDERSON MILL | 8.669086 | 154458.205390 | 2.416802e+08 | 154458.205390 |
| 1 | MULTIPOLYGON (((-97.670762852964 30.3085399639... | 95 | 95 | WINDSOR PARK | 2.383074 | 40527.378654 | 6.643628e+07 | 40527.378654 |
| 2 | MULTIPOLYGON (((-97.753526659646 30.2387648363... | 19 | 19 | DAWSON | 0.495535 | 17697.924998 | 1.381473e+07 | 17697.924998 |
| 3 | MULTIPOLYGON (((-97.738154269236 30.3027463827... | 91 | 91 | WEST UNIVERSITY | 0.738442 | 24981.044476 | 2.058658e+07 | 24981.044476 |
| 4 | MULTIPOLYGON (((-97.682624533084 30.2858668623... | 49 | 49 | MLK | 1.545283 | 33470.415430 | 4.308001e+07 | 33470.415430 |

For every neighborhood, location details were found using Google's Geocoding API and latitudes and longitudes were added to Neighborhood names. Data was cleaned and irrelevant columns were dropped.

```
df.head()
```

| | NEIGHNAME | latitude | longitude |
|---|---|---|---|
| 0 | ANDERSON MILL, Austin, TX, USA | 30.455835 | -97.807096 |
| 1 | WINDSOR PARK, Austin, TX, USA | 30.313549 | -97.691095 |
| 2 | DAWSON, Austin, TX, USA | 30.232926 | -97.761418 |
| 3 | WEST UNIVERSITY, Austin, TX, USA | 30.285220 | -97.733893 |
| 4 | MLK, Austin, TX, USA | 30.284035 | -97.694001 |

For every neighborhood, top 100 venues in 2000-meter radius venue data were found using Foursquare API. Below is the snapshot of all neighborhoods and their venue information.

```
austin_venues.head()
```

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | ANDERSON MILL, Austin, TX, USA | 30.455835 | -97.807096 | Freda's Seafood Grille | 30.464196 | -97.803776 | Seafood Restaurant |
| 1 | ANDERSON MILL, Austin, TX, USA | 30.455835 | -97.807096 | Millrun Park | 30.451548 | -97.802975 | Park |
| 2 | ANDERSON MILL, Austin, TX, USA | 30.455835 | -97.807096 | CC Zing | 30.460979 | -97.816818 | Smoothie Shop |
| 3 | ANDERSON MILL, Austin, TX, USA | 30.455835 | -97.807096 | Thundercloud Subs | 30.461629 | -97.795651 | Sandwich Place |
| 4 | ANDERSON MILL, Austin, TX, USA | 30.455835 | -97.807096 | Almarah Mediterranean Cuisine | 30.461054 | -97.817103 | Mediterranean Restaurant |

It was observed that there were too many venue categories and many of the venue categories fell under same category.

```
print('There are {} unique categories of venues.'.format(len(austin_venues['Venue Category'].unique())))
```

```
There are 268 unique categories of venues.
```

```
pd.set_option("display.max_rows", None, "display.max_columns", None)

print(austin_venues['Venue Category'].value_counts())
```

```
Mexican Restaurant                           137
Coffee Shop                                  119
Park                                          72
Pizza Place                                   67
Taco Place                                    63
Food Truck                                    62
Sandwich Place                                59
Burger Joint                                  57
Hotel                                         47
American Restaurant                           45
Ice Cream Shop                                45
Brewery                                       43
Vietnamese Restaurant                         41
Fast Food Restaurant                          41
BBQ Joint                                     41
Convenience Store                             40
Grocery Store                                 40
Bar                                           38
Gym                                           34
Café                                          32
```

On observing data and the counts, 4 main categories were determined.

1. Food & drinks which includes all different types of restaurants like Mexican, American etc. along with Pizza, Burger places. All desserts, beverages (alcoholic + non-alcoholic) were clubbed under this category. This category had the maximum counts.
2. Recreation category included venues like Parks, Spas, Golf courses, Arcade, etc.
3. Shopping category included Supermarkets, malls, grocery stores, convenience stores, pharmacy, etc.
4. Gym was the final category that was determined after observing the data.

All venue categories were corrected to either of the 4 categories listed above and data was cleaned.

| | | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Category |
|---|---|---|---|---|---|---|
| Gym | 8 | ANDERSON MILL, Austin, TX, USA | 30.455835 | -97.807096 | Punch Austin Kettlebell Gym | Gym |
| | 81 | DAWSON, Austin, TX, USA | 30.232926 | -97.761418 | HEAT Bootcamp & Personal Training | Gym |
| | 183 | MLK-183, Austin, TX, USA | 30.279398 | -97.673922 | YMCA | Gym |
| | 269 | SOUTHEAST, Austin, TX, USA | 30.347189 | -97.714563 | North Austin Rock Gym (NARG) | Gym |
| | 283 | SOUTHEAST, Austin, TX, USA | 30.347189 | -97.714563 | Beyond Fit | Gym |
| | 357 | JESTER, Austin, TX, USA | 30.387517 | -97.799267 | Anytime Fitness | Gym |
| | 418 | WESTGATE, Austin, TX, USA | 30.227962 | -97.800881 | LA Fitness | Gym |
| | 441 | GALINDO, Austin, TX, USA | 30.235578 | -97.768515 | HEAT Bootcamp & Personal Training | Gym |
| | 451 | GALINDO, Austin, TX, USA | 30.235578 | -97.768515 | MOD Fitness | Gym |
| | 464 | BRODIE LANE, Austin, TX, USA | 30.144467 | -97.852284 | Villages Of Shady Hollow Pool | Gym |
| | 482 | HYDE PARK, Austin, TX, USA | 30.304412 | -97.730448 | Hyde Park Gym | Gym |

```
df_final['Venue Category'].value_counts()
```

```
Food Drinks    1429
Recreation       75
Gym              73
Shopping         42
Name: Venue Category, dtype: int64
```

This data was further prepped using one hot encoding technique for Machine Learning algorithm which will convert categorical Venue Categories into numbers. Snapshot of final data prepped is shown below which was used for K-means clustering algorithm. This will be discussed in Methodology section.

```
austin_grouped = austin_onehot.groupby('Neighborhood', as_index=False).agg({'Food Drinks': 'sum', 'Gym': 'sum', 'Recreation': 's
austin_grouped.head()
```

| | Neighborhood | Food Drinks | Gym | Recreation | Shopping |
|---|---|---|---|---|---|
| 0 | ALLANDALE, Austin, TX, USA | 22 | 1 | 1 | 0 |
| 1 | ANDERSON MILL, Austin, TX, USA | 14 | 1 | 2 | 1 |
| 2 | BARTON HILLS, Austin, TX, USA | 18 | 1 | 0 | 0 |
| 3 | BERGSTROM, Austin, TX, USA | 13 | 0 | 0 | 0 |
| 4 | BLUFF SPRINGS, Austin, TX, USA | 16 | 2 | 0 | 1 |

Demographic data was downloaded, and irrelevant columns and rows were omitted from file. Columns were labelled and a snapshot is shown below.

| | Neighborhood | Total Population | Age 0-4 | Age 5-9 | Age 10-14 | Age 15-17 | Age 18-19 | Age 20-24 | Age 25-34 | Age 35-44 | Age 45-54 | Age 55-59 | Age 60-64 | Age 65-74 | Age 75 8 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | ALLENDALE | 6643 | 0.063224 | 0.054945 | 0.043504 | 0.024086 | 0.010537 | 0.042451 | 0.156104 | 0.171609 | 0.149631 | 0.075267 | 0.056601 | 0.067891 | 0.05449 |
| 1 | ANDERSON MILL | 28473 | 0.062410 | 0.073192 | 0.077020 | 0.047940 | 0.022056 | 0.046500 | 0.116637 | 0.157693 | 0.176799 | 0.064658 | 0.055175 | 0.060935 | 0.02813 |
| 2 | AVERY RANCH--LAKELINE | 14785 | 0.111532 | 0.090971 | 0.066824 | 0.030977 | 0.011701 | 0.045384 | 0.202164 | 0.216638 | 0.120528 | 0.038147 | 0.026581 | 0.026987 | 0.00892 |
| 3 | BARTON CREEK MALL | 5147 | 0.052263 | 0.070915 | 0.083932 | 0.056926 | 0.019623 | 0.044492 | 0.101418 | 0.137167 | 0.171751 | 0.080435 | 0.059452 | 0.047601 | 0.03574 |
| 4 | BARTON HILLS | 8022 | 0.030666 | 0.034156 | 0.029294 | 0.018325 | 0.010222 | 0.107205 | 0.291573 | 0.147469 | 0.114934 | 0.062329 | 0.048990 | 0.048118 | 0.03178 |

Several columns were clubbed together to form age groups and reduce the number of columns. Below is the snapshot of data.

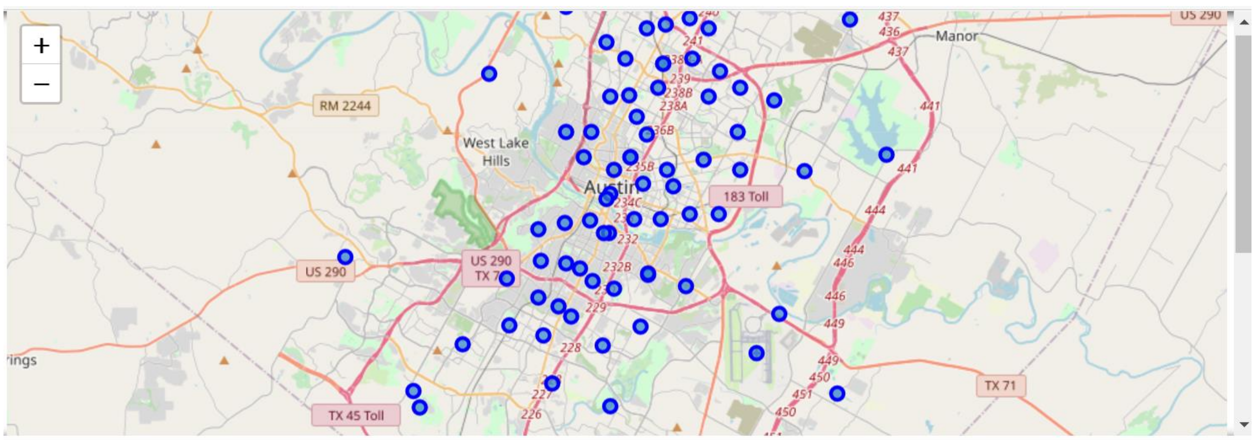
```
demo.head()
```

| | Neighborhood | Total Population | Age < 18 | Age 18-24 | Age 25-34 | Age 35-44 | Age 45-54 | Age 55-64 | Age 65 Plus |
|---|---|---|---|---|---|---|---|---|---|
| 0 | ALLENDALE | 6643 | 0.185759 | 0.052988 | 0.156104 | 0.171609 | 0.149631 | 0.131868 | 0.152040 |
| 1 | ANDERSON MILL | 28473 | 0.260563 | 0.068556 | 0.116637 | 0.157693 | 0.176799 | 0.119833 | 0.099919 |
| 2 | AVERY RANCH--LAKELINE | 14785 | 0.300304 | 0.057085 | 0.202164 | 0.216638 | 0.120528 | 0.064728 | 0.038553 |
| 3 | BARTON CREEK MALL | 5147 | 0.264037 | 0.064115 | 0.101418 | 0.137167 | 0.171751 | 0.139887 | 0.121624 |
| 4 | BARTON HILLS | 8022 | 0.112441 | 0.117427 | 0.291573 | 0.147469 | 0.114934 | 0.111319 | 0.104837 |

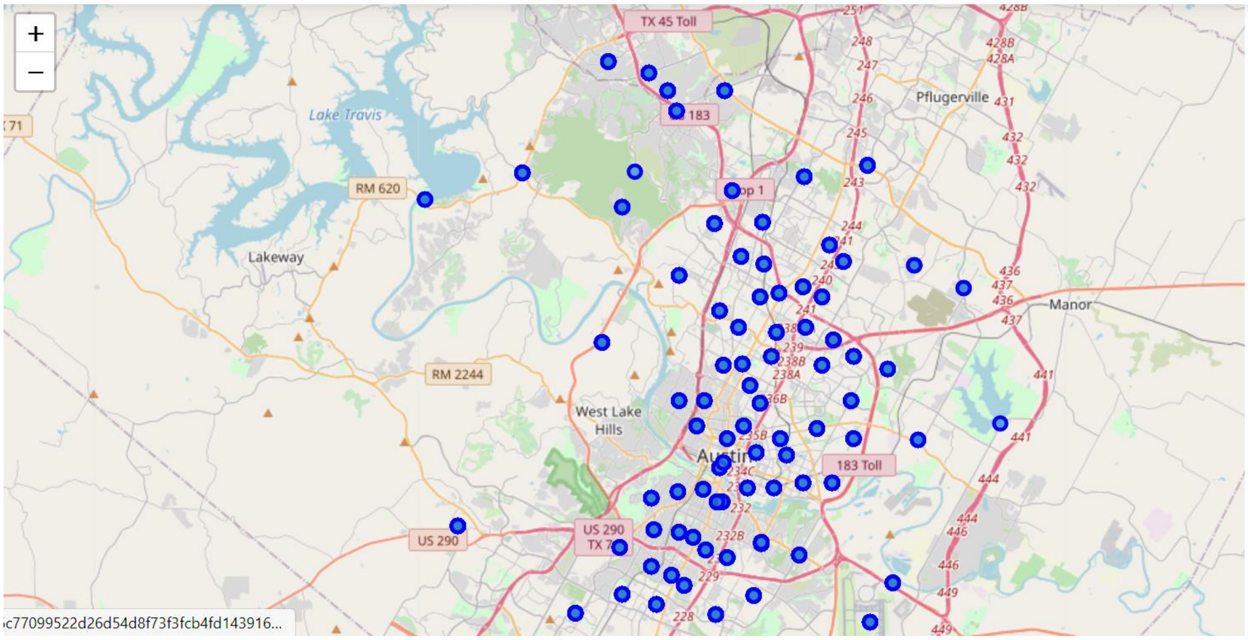
Further analysis and data cleaning was done on data to combine clustering results with demographic information will be discussed in Methodology section.

# 3 Methodology

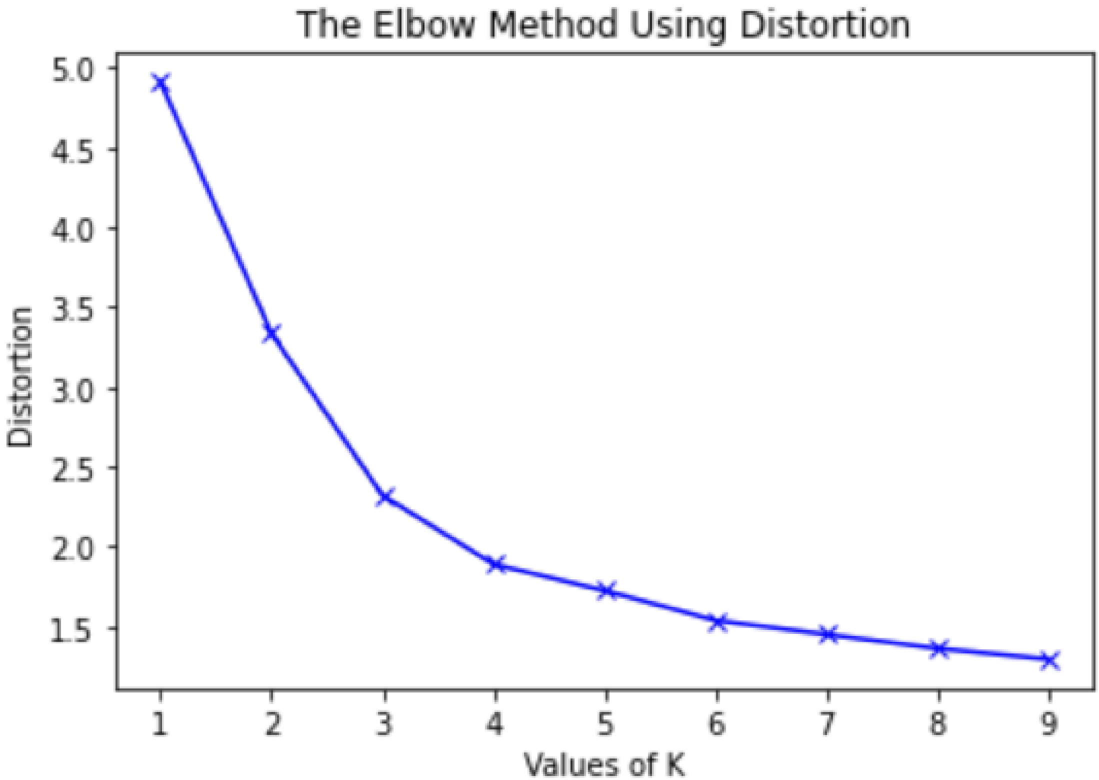
Data was collected from various sources on web and several APIs like Geocoding and Foursquare was used to prep it for further analysis as was seen in earlier section. Below is the snapshot of map of neighborhoods in Austin.

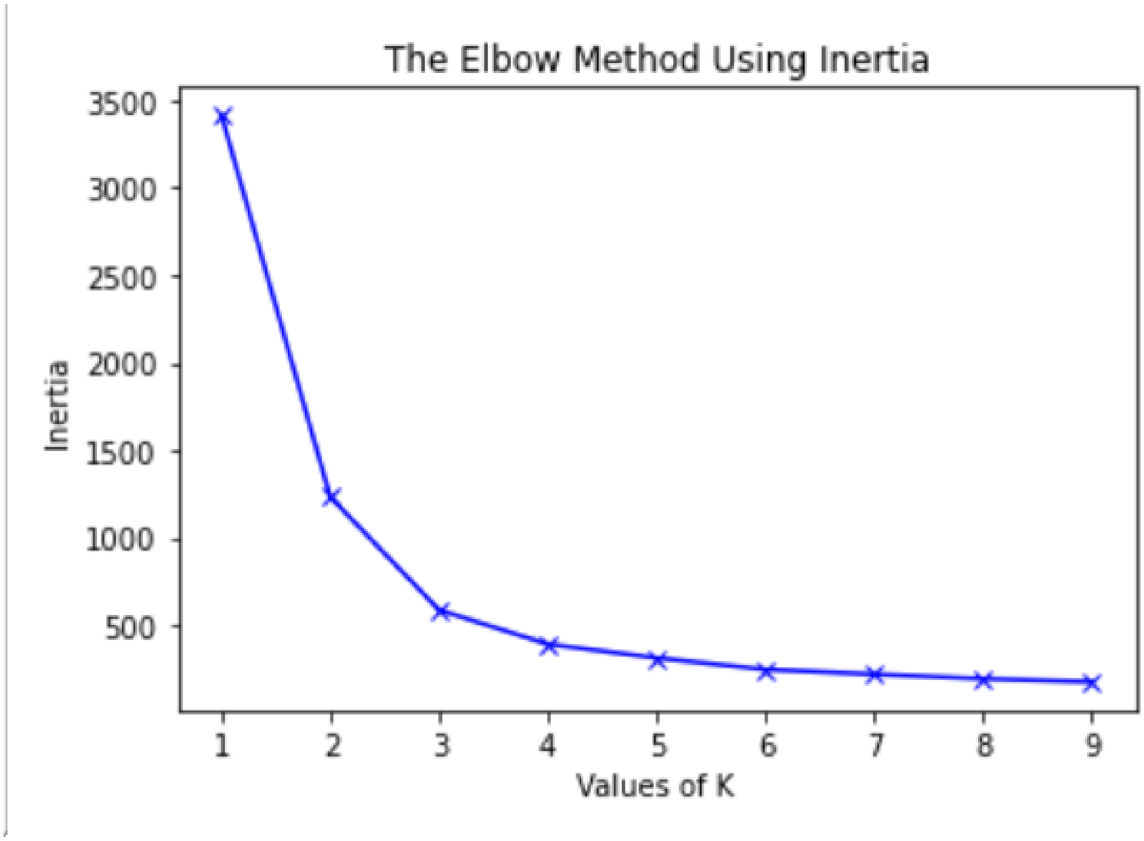



Below is the snapshot of 100 venues returned by Foursquare API in 2000-meter radius of neighborhoods.



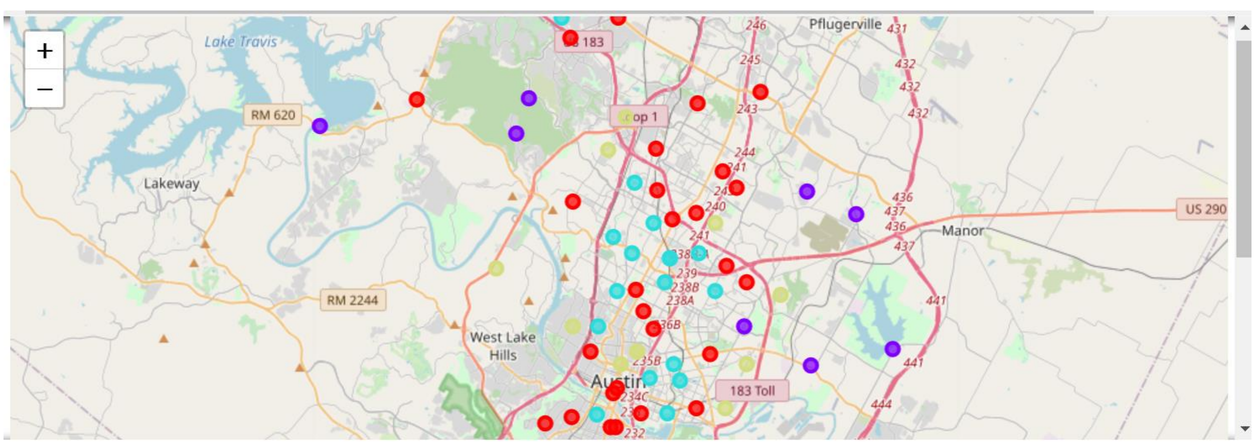

Data was to be analyzed of this venue information and intention was to find similar characteristics within it and ultimately and group them. This became ideal Clustering program. It fell under unsupervised learning as similarities were not known and we had to consider all four categories of Venue to form a cluster of neighborhoods. The most common clustering algorithm was used – K means. Using venue information, neighborhoods were grouped into clusters. Best K was found using elbow method using both Distortion and Inertia to be sure about the value of K which would be used to run algorithm on the prepped data.


The Elbow Method Using Distortion
Distortion
Values of K
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1
2
3
4
5
6
7
8
9

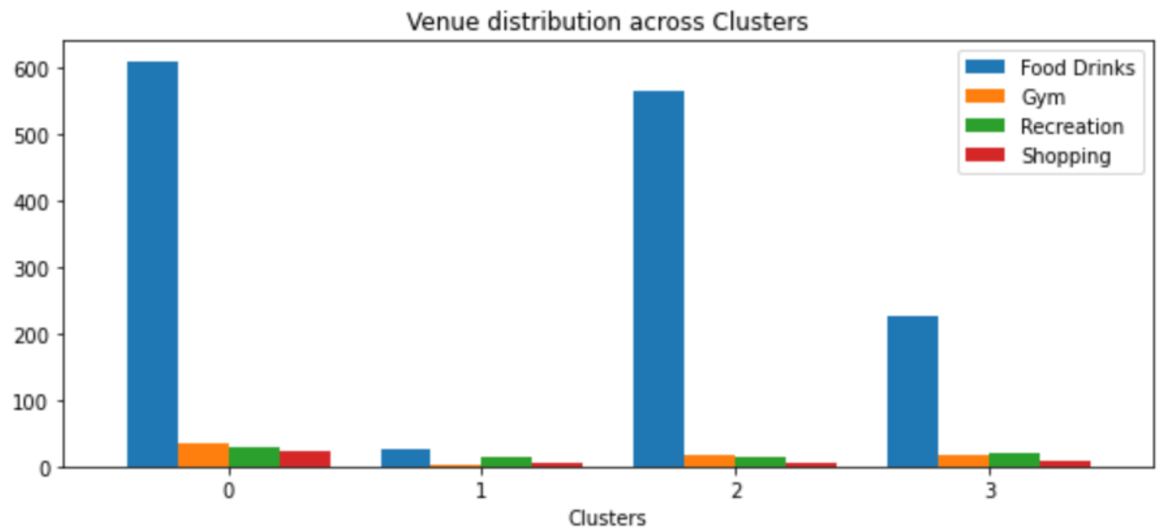



K value was observed at elbow points to check at what point was distortion/inertia decreasing in linear fashion. Value of K as 4 is considered as optimal to run clustering algorithm from the observations of elbow method.

Below is snapshot of 4 clusters that were formed.




Analysis was done to find how the Venues were distributed among clusters and a bar chart was plotted.




As, it can be observed from figure above, Cluster 0 had the maximum count of Venues considering all 4 categories in comparison to remaining 3 clusters. We will be checking further with demographic information to determine whether Cluster 0 would have the answer to our problem.

# 4 Results

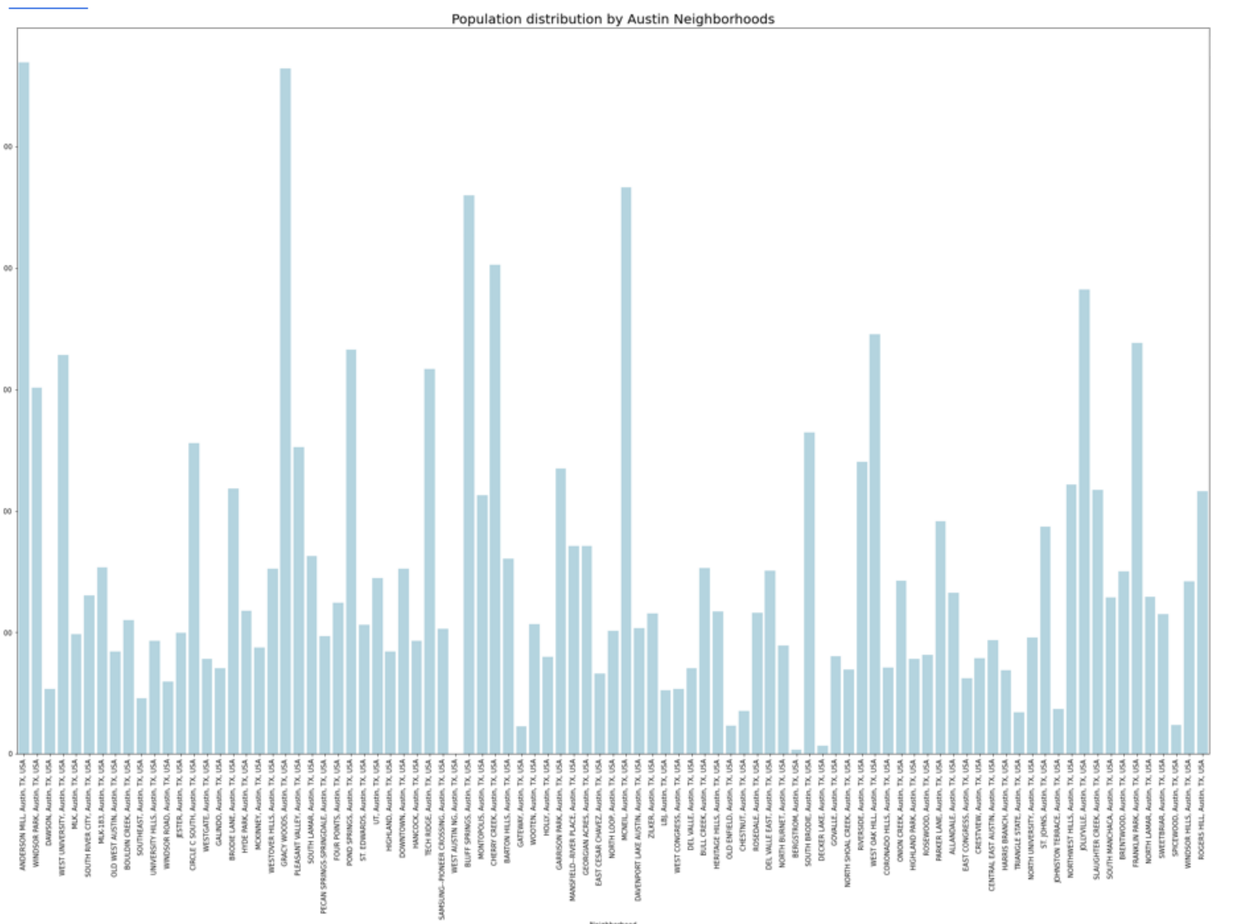
Demographic data of Austin was collected and prepped as explained in section 2 of this report. It had information of neighborhoods, total population and % of population among different age groups for all the neighborhoods. This demographic data was joined with data of Clusters to further the analysis.

There were some anomalies found in neighborhood names in both tables which was not giving the correct count on joining tables. Neighborhood names were corrected in demographic table, and tables were joined to get cluster information clubbed with demographic information.

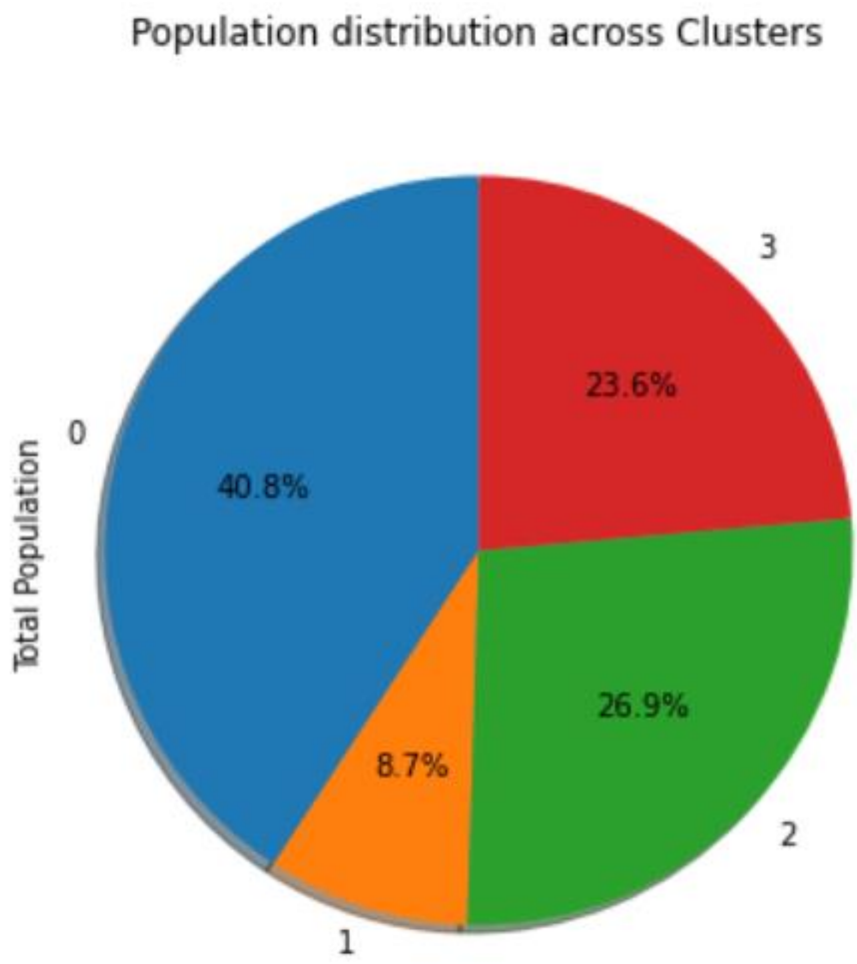
```
df_austin
```

| | Clusters | Neighborhood | latitude | longitude | Food Drinks | Gym | Recreation | Shopping | Total Population | Age < 18 | Age 18-24 | Age 25-34 | Age 35-44 | Age 45-54 | Ag |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | ANDERSON MILL, Austin, TX, USA | 30.455835 | -97.807096 | 14 | 1 | 2 | 1 | 28473.0 | 0.260563 | 0.068556 | 0.116637 | 0.157693 | 0.176799 | 0.11 |
| **1** | 2 | WINDSOR PARK, Austin, TX, USA | 30.313549 | -97.691095 | 20 | 0 | 1 | 1 | 15086.0 | 0.253016 | 0.118255 | 0.204958 | 0.147090 | 0.120178 | 0.07 |
| **2** | 2 | DAWSON, Austin, TX, USA | 30.232926 | -97.761418 | 22 | 1 | 0 | 0 | 2670.0 | 0.149813 | 0.144569 | 0.259551 | 0.159925 | 0.119476 | 0.07 |
| **3** | 3 | WEST UNIVERSITY, Austin, TX, USA | 30.285220 | -97.733893 | 12 | 0 | 0 | 0 | 16408.0 | 0.011519 | 0.872928 | 0.071672 | 0.017735 | 0.012433 | 0.00 |
| **4** | 0 | MLK, Austin, TX, USA | 30.284035 | -97.694001 | 18 | 0 | 3 | 1 | 4917.0 | 0.293065 | 0.110230 | 0.177954 | 0.127720 | 0.119178 | 0.07 |

Population distribution across neighborhood was analyzed. Below is the snapshot of the visualization.

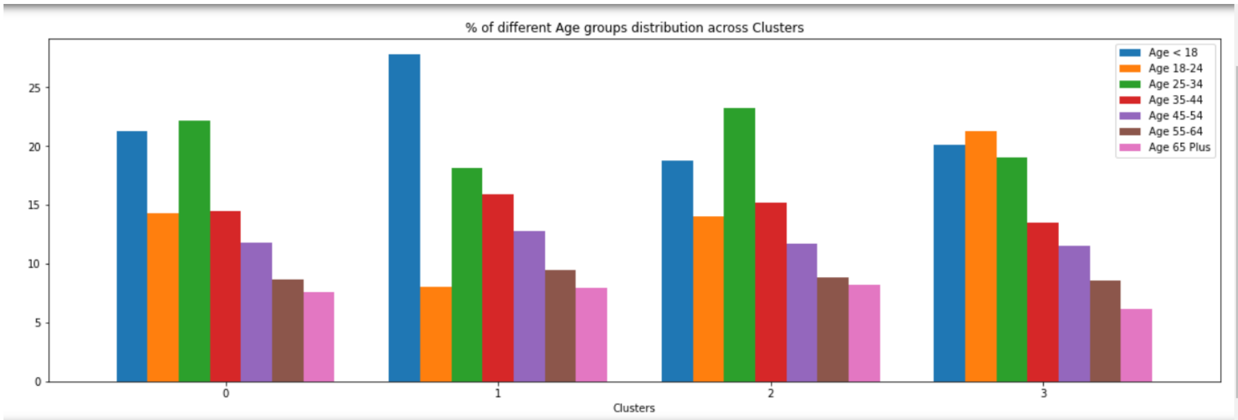


This was difficult to read across all neighborhoods so population distribution across Clusters were determined to get an idea on which Clusters had the maximum population density. Pie chart is ideal in such scenario.



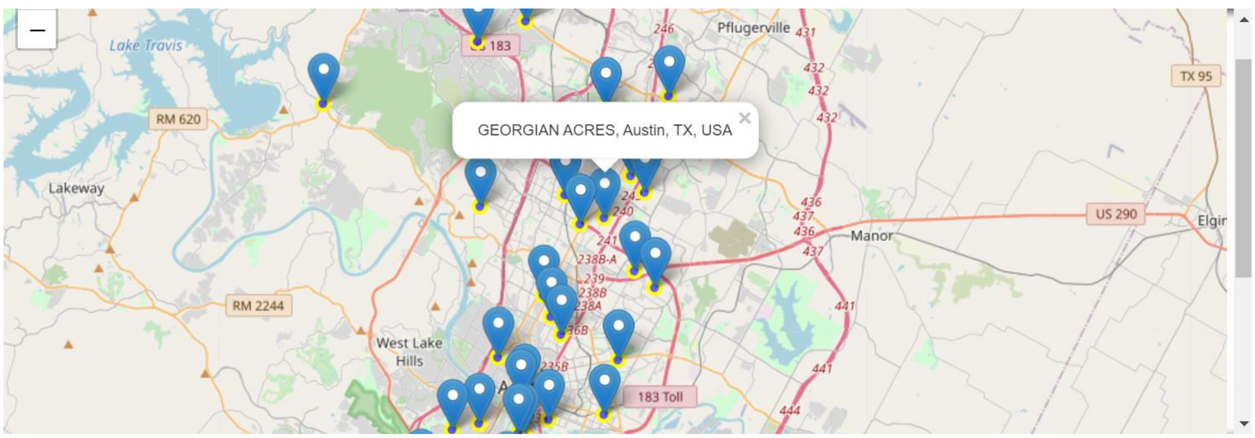

As it was observed, Cluster 0 had the highest total population.

Further analysis was done to understand how different age groups were distributed among these clusters and what percentage individual group had in those clusters. Below is the snapshot of the visualization that was created.

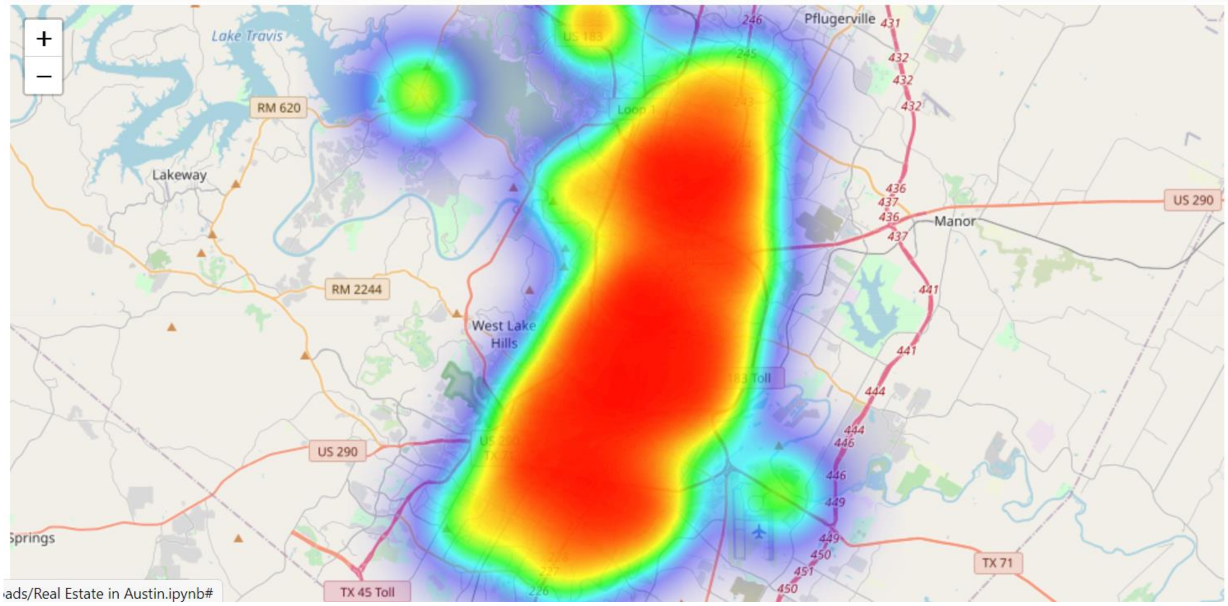



As it was observed, cluster 0 had more uniform distribution among different age groups. More importantly, % of population in age groups <18, 18-24, 25-34 is good, which categorizes under youth, and is the base of the project.

Further analysis was done on Cluster 0 and below are snapshots of Neighborhoods in Cluster 0.




Below is heat map of venue across neighborhoods in Cluster 0.




# 5 Discussion

So where should Builders & Contractors build the next Apartment complexes? Clustering analysis revealed that Cluster 0 had the maximum number of venues across all categories. Demographic information supports this which was inferred as part of further analysis done using demographic and cluster information. It was observed that Cluster 0 had the highest total population and the percent distribution among the youth population also was significant. With all this visualization, it was inferred that Cluster 0 is what would be recommended to Builders & Contractors for setting up new Apartment complexes in the neighborhoods in it as the current trend supports it and it has all the things that could appeal to youth population.

Some of the neighborhoods in Cluster 0 than can be picked are Downtown, Cherry Creek, Georgian Acres, East Congress. Although, all neighborhoods in the cluster are promising.

# 6 Conclusion

In this project, neighborhood data was collected and combined with location and venue data to get characteristics on which neighborhoods can be grouped in clusters. Demographic information containing population distribution across neighborhood was added to analysis, all to support and answer the business problem - where should construction of next Apartment complex should happen in Austin, TX? It was inferred that several neighborhoods like Downtown, Georgian Acres to name a few are promising locations where construction could happen as it will appeal the young crowd for the various amenities they offer and ease of accessibility of restaurants, parks, gyms, etc. Current population trend supports this inference and this project can be used to predict the likely and preferred neighborhoods.

This study can be extended, and several factors can be added to analysis like safety of neighborhoods, proximity to offices, public transport and many more which will help in coming up with better recommendations. This will help Builders & Contractors meet the needs of their future residents and attract more of them. It will also benefit residents as they would get desirable and comfortable accommodation suiting their requirements.