# Texas Derby – Austin, Dallas, or Houston?

**Yuhao Sun**

**IBM Data Science Professional Certificate Capstone Project**




# Business Problem

Texas has been consistently ranked number one for the inflow of migrants from other states across America due to employment opportunities and relatively low living costs.

Most of the people inflow will settle down in one of the three cities (Austin / Dallas / Houston).

***Which city among the three is the most convenient in terms of living facilities?***

# Business Problem

- This comparison can help **People moving from one city to the other** to filter for areas similar (or even different, if you are up for something new) to what you are used to.
- This comparison can help **People moving from other states to Texas** to filter for neighborhoods similar or different from their original came from, and introduce them into local fun and popular places
- **Companies expanding within one of the cities** might want to look for a similar type of neighborhoods, as they are targeting a specific user group. The comparison can offer the first indication.
- **Companies expanding from one city to the other** can use their experience from the original city and look for a fitting (e.g., similar) neighborhood in the second one.

# Data Collection

Wikipedia. *List of cities in Texas Counties*

SimpleMaps.com *US Zip Codes and Location*

Foursquare API. *Top venues and their categories for each city*

# Methodology

- data processing and cleaning
- exploration data analysis (EDA)
- data scraping using Foursquare API. top 10 venue categories for each city's vector profile for clustering
- k-means clustering algorithm to identify patterns in the data, k=8
- data/clustering visualization with Folium

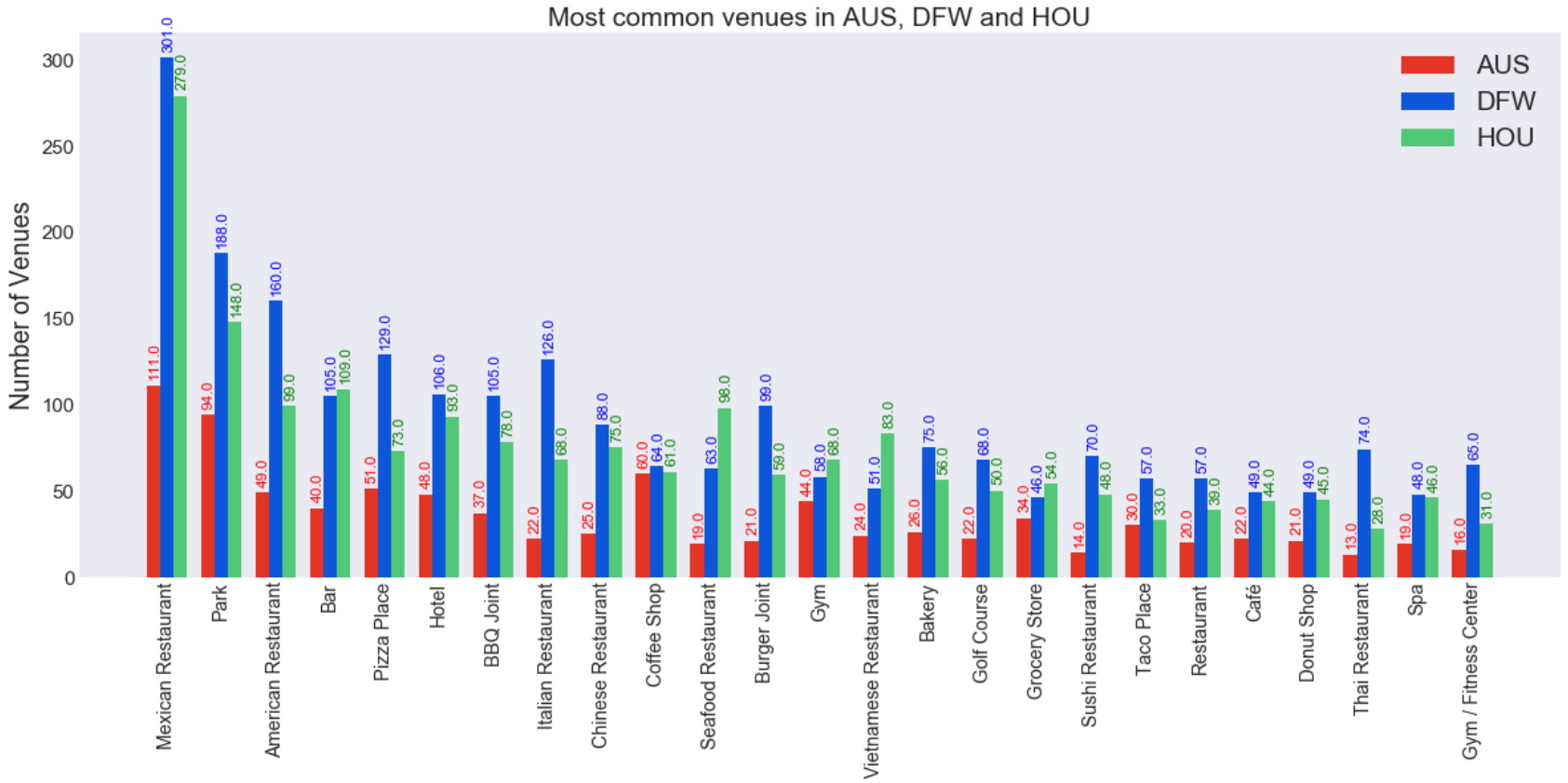
# Results/Discussion

EDA

For a better comparison, we should look at those numbers when **normalized with the number of inhabitants of all three cities**.

Normalized data shows that, in general, **AUS is the most convenient city** amongst the three. It almost topped every category

**Mexican restaurants** are still the most popular venue for all three cities




```
print(f'Shape of AUS ZIP-code only dataframe: {AUS_zip.shape}')
print(f'Shape of DFW ZIP-code only dataframe: {DFW_zip.shape}')
print(f'Shape of HOU ZIP-code only dataframe: {HOU_zip.shape}')
print(f'Shape of filtered data frame including all ZIP-codes for AUS,DFW and HOU: {AUS_DFW_HOU_zip.shape}')
```

```
Shape of AUS ZIP-code only dataframe: (81, 9)
Shape of DFW ZIP-code only dataframe: (201, 9)
Shape of HOU ZIP-code only dataframe: (191, 9)
Shape of filtered data frame including all ZIP-codes for AUS,DFW and HOU: (473, 9)
```

# Results/Discussion

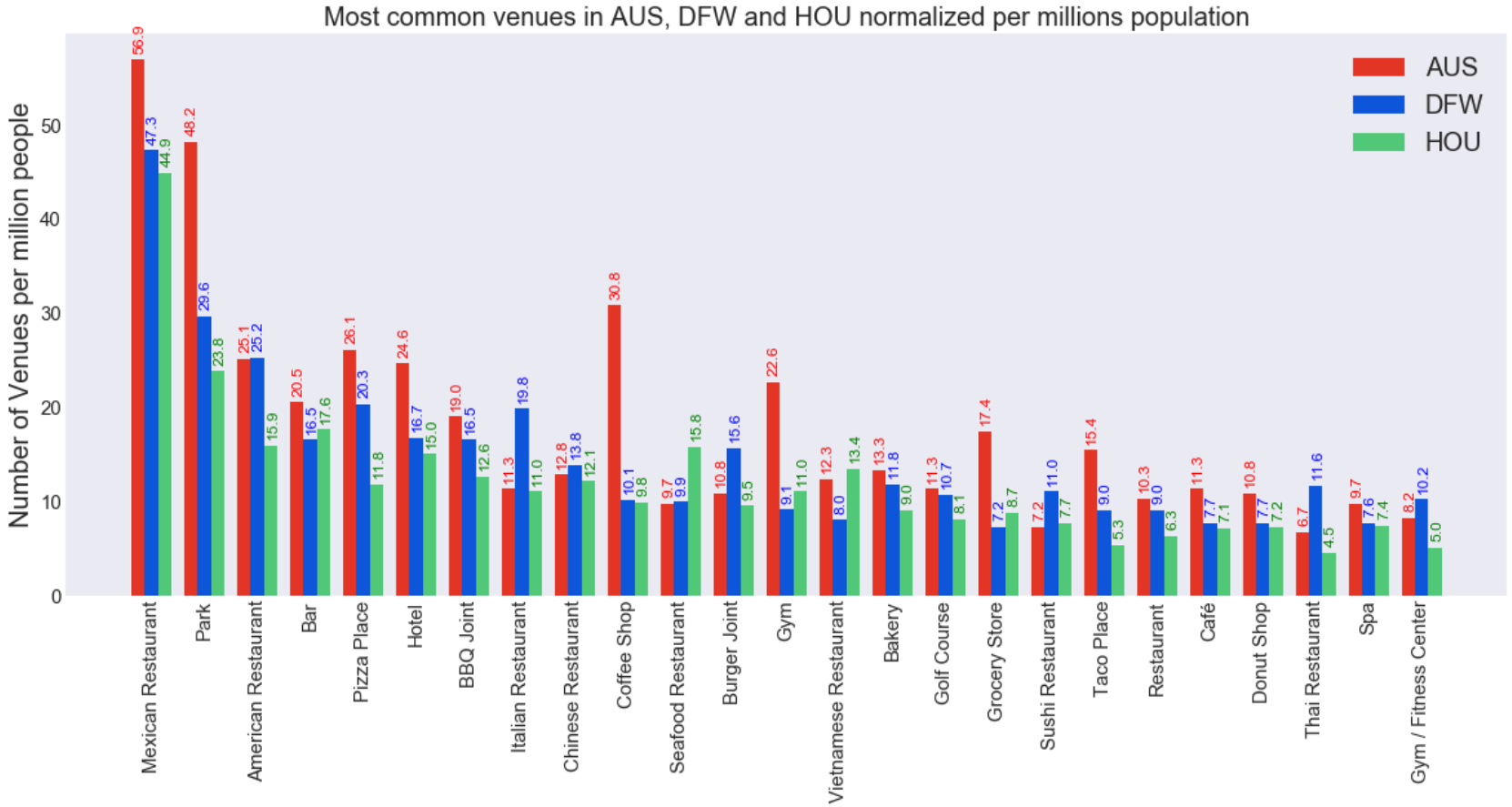
EDA

Mexican restaurants are most popular in all cities.

All 3 cities have roughly the same number of coffee shops.

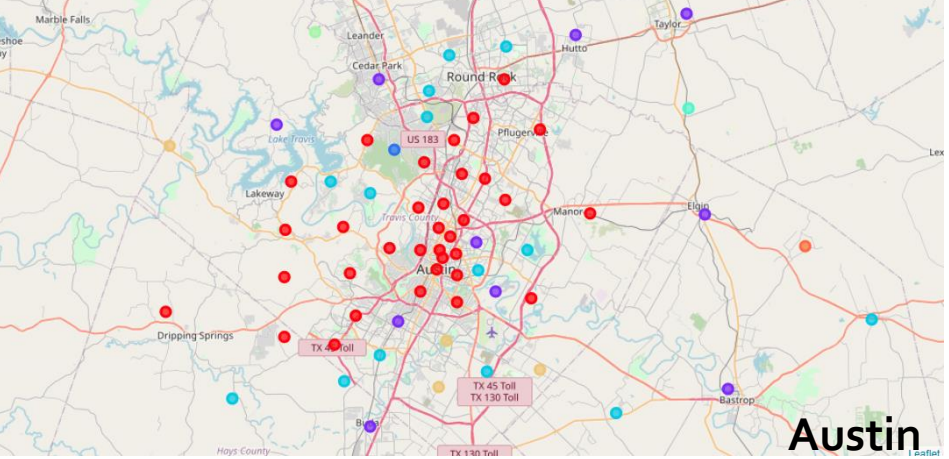
DFW and HOU are about the same size, while AUS is less than half the size of the two mega-cities.

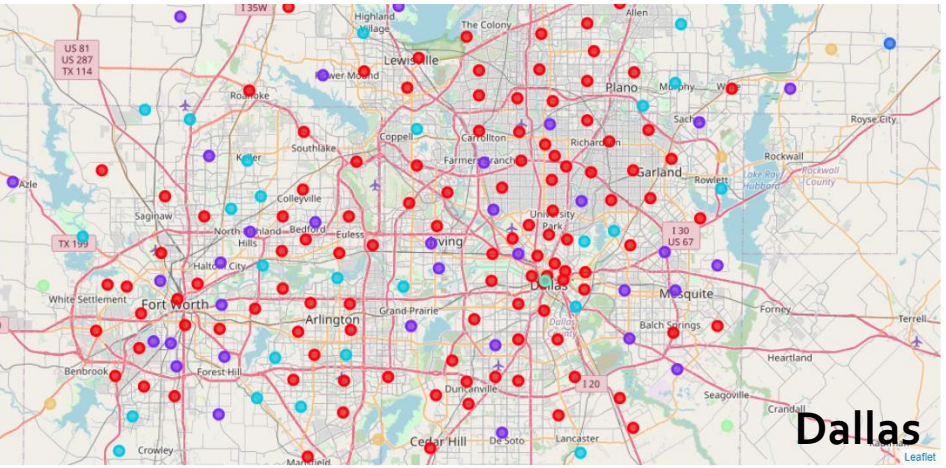
in terms of the absolute number of venues returned, AUS is the least, while DFW is slightly better than HOU overall

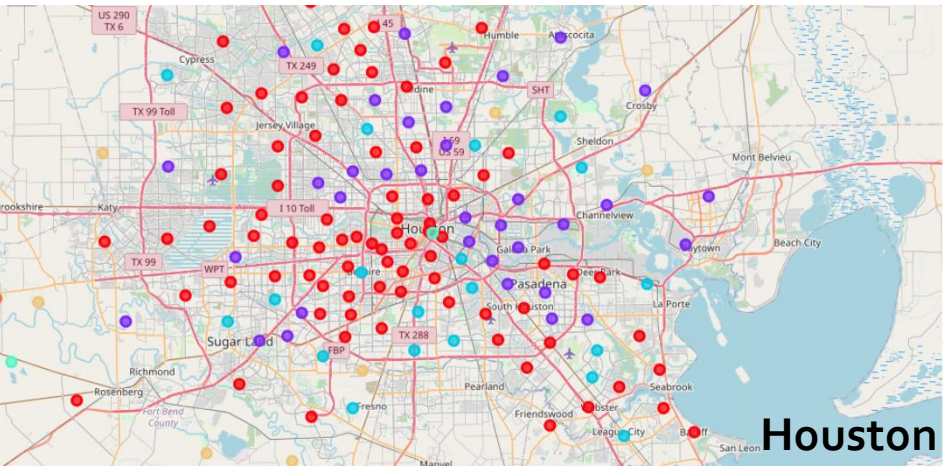



```
print(f'Population of AUS: {AUS_DFW_HOU_SUM["population"][0]}')
print(f'Population of DFW: {AUS_DFW_HOU_SUM["population"][1]}')
print(f'Population of HOU: {AUS_DFW_HOU_SUM["population"][2]}')
```

```
Population of AUS: 1950712
Population of DFW: 6357043
Population of HOU: 6209033
```




# Results/Discussion

Clusters

As can be seen, certain clusters are more common, while some are unique or rare. This is expected, as some zipcode areas share a common cause, e.g., housing. These neighborhoods fall into the same cluster as they share a similar venue structure. They are close to the city center.

At the same time, zipcode areas located at the center of all three cities share similar trends too. These are more venue crowded neighborhoods.

# Conclusion

**Austin** is the most convenient city among the three in terms of the number of most common venues in the data sets, normalized by the cities population.

For the two major metropolises**, DFW is slightly better than HOU**.

Texas people love **Mexican food**, as the number of Mexican restaurants is the most in all three cities.

In general, the three cities have **very similar zip code areas/clustering**; one should expect **roughly the same level** of living standard in any of the cities.