

## Business Problem/Objectives:

Metro Phoenix is a metropolitan area, centered on the city of Phoenix, that includes much of the central part of the state of Arizona. It is one of the fastest growing major metropolitan areas, gaining nearly 600,000 residents from 2010 to 2017. As of Census Bureau's 2017 population estimates, Metro ranks 11th largest Metropolitan Area in the nation with 4,737,270 residents. Metro Phoenix has historically been the center of the state's economy, ranks 5th in the nation in economic growth. The economic and population growth in the area draw attentions in the nation and from the world.

An investment group is seeking opportunities to invest on short-term real estate rental business in Phoenix metro area in Arizona. One key factor to a successful short term rental business is the highly convenient surrounding businesses. The investment group would like to have a general idea on phoenix metro area as how the business located in major suburb cities downtown area, are different suburb cities have same rental business potential based the business distributions, how is the business diversity and amenity and etc. Based on the request, we came across the following objectives:

1. the business diversities and amenities analysis in the metro phoenix area in order to help with their business decision.
2. the similarity of the suburb cites based on business distribution to avoid the repeated investments.

## Data:

The analysis will be performed on each individual suburb cities business data. In order to collect the business data, first of all, we extracted each city latitude and longitude from <https://www.latlong.net/>. Then based on the geographic location of each city, we use foursquare API to pull each city's business(ventures) data from <https://foursquare.com/>. The JSON file of each city's ventures data will be downloaded, transformed and appended into a single python dataframe, then the type, popularity of business and similarity of rental market based on each city's business type will be analyzed.

First we selected metro area suburb cities and download the latitude and longitude of each city to make a dataframe.

After knowing city's geographical data, we loop through each city and use foursquare API to download each city venues information into json file, parse each file and append the venue name, venue type, latitude and longitude into one data frame.

# Methodology

The methodology is the a guiding strategy for solving the problem, it provides a framework that what methods and processes that are used to obtain the answers and solve the problem. We have the following stages that composite the methodology of the project:

1. Collect inspection data
2. Explore and understand data
3. Cleanse and prepare data
4. Modeling

## 1. Collect data

We first select major suburb cities in metro phoenix area and collect the geographical latitude and longitude from <https://www.latlong.net/>, then made a function to in order to parse the json file saved from foursquare. We assigned all foursquare API parameters values into variables and the loop through the cities and send API request to foursquare to download each city's json file, parsed file and put the data into dataframe.

## 2. Explore and understand the busniess data

To understand what type of the business in different suburb city downtown area, what are the unique business types are available in each city, we calculate the count of each type of business of each city, total unique business type of each city:

## 3. Prepare and Analyze data

We group rows by suburb cities and

by taking the mean of the frequency of occurence of each category, we also check the correlation to understand how different type of business correlate to each other:

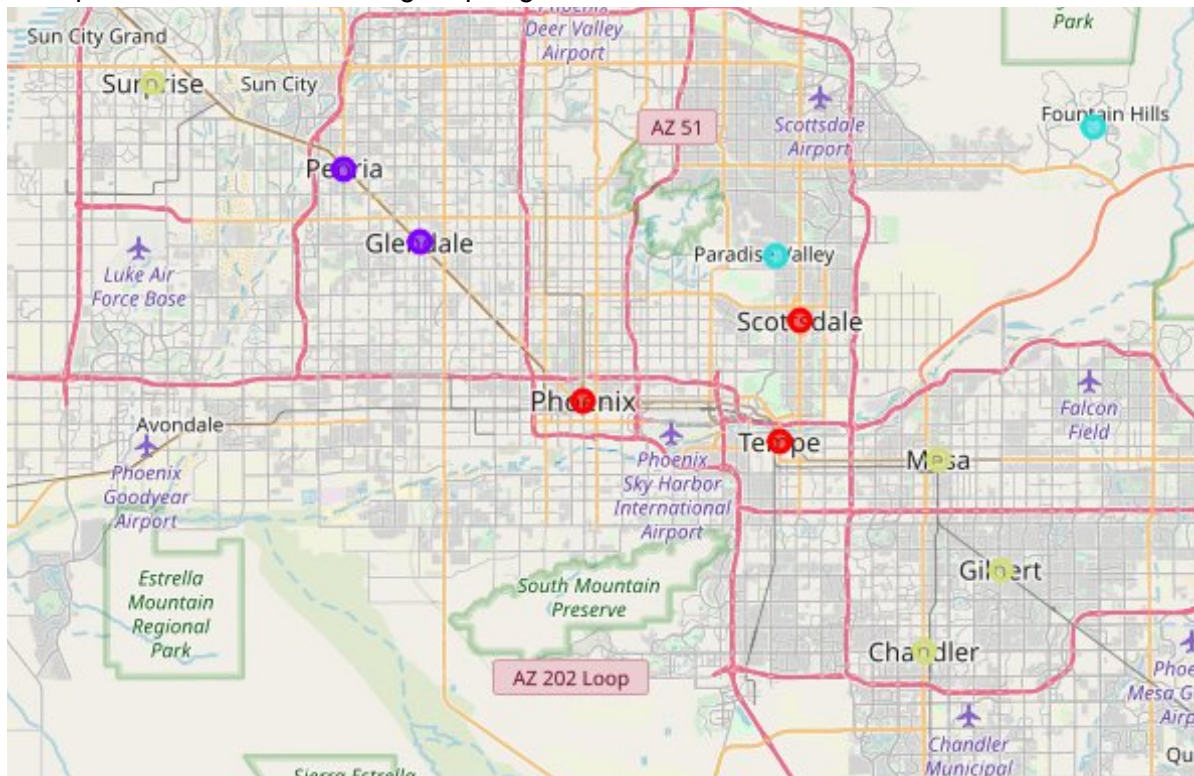
Also we analyze what are the common business in each city:

## 4. Modeling

Finally we use the K mean cluster to discover among all cities, what cities share the similar investment environment based on the business type.

## Result

A map of the K mean clustering map is generated for the client to better understand the result



## Discussion

The project is purely based on the business data downloaded from foursquare. The result may not reflect the short term rental business potentials objectively. In order to get more accurate result to help with the business decision, other factors like the average rents, population, average housing price and etc. of each city will need to be considered.

# Conclusion

From our analysis, we see: all cities have similar diversity of business and amount of the business. Among all cities, Mesa, Gilbert and Chandler geographically located close with each other and also with very similar type of businesses. City of Surprise though located far from these 3 cities, because of the nature of the business, it's still considered the similar investment environment. City of Paradise Valley and City of Fountain Hill are located apart, from the type of business, they both own trails and golf courses that are different than other cities, these 2 cities have similar investment potential. Center cities City of Phoenix, City of Tempe and City of Scottsdale are similar type of investment environment; City of Peoria and City of Glendale are similar due to the similarity of businesses.

Again, other factors like the average income, population, average rent, and etc. of each city should be combined with this analysis conclusion in order to achieve more accurate result, therefore to achieve wiser business decision.