Keith Jones Capstone Project: Data Section

This project will utilize the following sources of data:

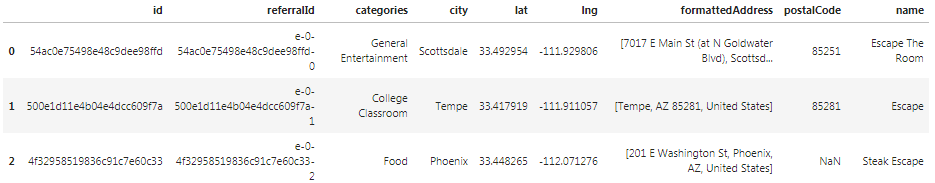
Foursquare API

The Foursquare database will be used to provide information of 3 primary types:

1. Search data about the various escape rooms within a given metropolitan area
2. Venue data about different escape rooms within that metropolitan area
3. Explore data about each zip code in the metropolitan area

Escape Room Search Data: FourSquare’s search functionality will be used to gather listings of venues in a given metropolitan area that contain the word ‘escape’ in some form in their listing. It will use query strings of the form: https://api.foursquare.com/v2/venues/explore?client\_id=ase8sf8a9823alf&client\_secret=23pouf2893u24p&v=20181120&ll= 33.6050991,-112.4052495 &radius=100000&limit=100&query=escape

Once obtained and converted into dataframe form (and unnecessary columns are removed), data will be of the form:

 This data will need to be cleaned up in a few ways, most notably in that it includes venue information for venues not in the escape room industry. More about this process will be included in the Methodology section.

Escape Room Rating Data: Because many escape rooms do not yet have ratings on Foursquare, ‘likes’ will suffice for purposes of determining venue successfulness. This information is gained through queries of the following type: https://api.foursquare.com/v2/venues/189732975983?client\_id=ase8sf8a9823alf&client\_secret=23pouf2893u24p&v=20181120

Once obtained, the query results will be picked through to obtain the number of ‘likes’ if any (using the ‘response’🡪’venue’🡪’likes’🡪’count’ dictionary key progression) and the number of ‘dislikes’ if any (using the ‘response’🡪’venue’🡪’dislike’🡪’count dictionary key progression).

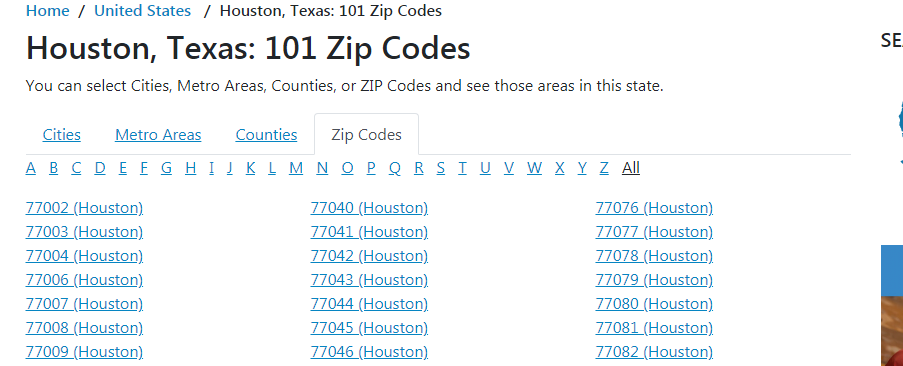
Zip Code Data: The analysis will require venue exploration information from the various zip codes in the chosen metropolitan area. Each zip code will require a query of the following type: https://api.foursquare.com/v2/venues/explore?client\_id=ase8sf8a9823alf&client\_secret=23pouf2893u24p&v=20181120&ll=33.4764,-112.2980&radius=500&limit=100

Once obtained and converted into dataframe form (and unnecessary columns are removed), data will be of the form:



www.BestPlaces.net

Data from BestPlaces.net will be used to obtain lists of zip codes for appropriate metropolitan areas/cities. The principal section of the applicable websites that contains the needed data looks like this:

The Beautiful Soup package will be used to scrape the zip codes from the data to provide us with the needed zip codes for venue profiling.

Splitwise Blog Posting on “Free US Population Density And Unemployment Rate By Zip Code”

Available at <https://blog.splitwise.com/2014/01/06/free-us-population-density-and-unemployment-rate-by-zip-code/> (accessed 12/12/18). This blog hosts free access to csv and Excel files containing population and population density information for every zip code in the United States (see example below). This information will help to indicate which zip codes in a given metropolitan area should be removed from consideration due to low population density, as well as provide helpful context regarding population densities of the remaining zip codes in the final analysis of each metropolitan area.

|  |  |  |  |
| --- | --- | --- | --- |
| Zip/ZCTA | 2010 Population | Land-Sq-Mi | Density Per Sq Mile |
| 80622 | 268 | 13.179 | 20.33538 |
| 80623 | 1028 | 0.666 | 1543.544 |
| 80624 | 946 | 31.361 | 30.16485 |
| 80631 | 48603 | 102.846 | 472.5804 |