**Jainabou Barry Danfa**

**SI 507 Project Proposal**

**Data Sources**

1. Zillow API
   1. The User will input the zipcode and using an API request from Zillow(<https://www.zillow.com/howto/api/GetSearchResults.htm>) and return the Zestimate data for homes in that zipcode
2. Yelp API
   1. The user will input the zipcode and using the API request from Yelp(<https://www.yelp.com/developers/documentation/v3/business_search>) and return the restaurants in that zipcode + the types of restaurants + attributes
3. US household Income Stats
   1. The user will input the zipcode and using the Kaggle csv file (<https://www.kaggle.com/goldenoakresearch/us-household-income-stats-geo-locations#kaggle_income.csv>) I will populate a database that will store this data

**Total Challenge Score: 4** (Zillow API) + **4**(Yelp API) + **2**(US household income stats) = **10**

**Data Presentation:**

This program will be interactive and allow the user to have 4 different outputs:

1. Map the data (Housing estimates and restaurants in area + median income for area) on plotly with the returned data with the user having the
2. Ability to get specific information on specific homes and or restaurants by entering *more information* in the command prompt which will open the URL in their browser for that specific website 30% of income on housing in that area)
3. User can request to see a stacked bar chart of the different types of restaurants in that zipcode and will group the restaurants based on food type (middle eastern, Italian, Chinese, ect. )

**Data Collection:**

1. **Create database for US household income stats** 
   1. Income.db

|  |
| --- |
| **Zipcodes** |
| Id  *(Primary Key)* |
| Zipcode  *(integer)* |
| State Id  *(integer)* |
| Lat  *(integer)* |
| Lon  *(integer)* |

|  |
| --- |
| **States** |
| Id  *(primary key)* |
| Name  *(Text not null)* |
| Abbr  *(Text)* |

|  |
| --- |
| **Rental Prices** |
| Id  *(Primary Key)* |
| Zipcode\_id  *(integer)* |
| 2017\_avg  *(integer)* |
| 2017\_std  *(integer)* |
| 2018\_avg  *(integer)* |
| 2018\_std  *(integer)* |

|  |
| --- |
| **Housing Prices** |
| Id  *(Primary Key)* |
| Zipcode\_id  *(integer)* |
| 2017\_avg  *(integer)* |
| 2017\_std  *(integer)* |
| 2018\_avg  *(integer)* |
| 2018\_std  *(integer)* |

|  |
| --- |
| **Income Levels** |
| Id  *(Primary Key)* |
| Zipcode\_id  *(integer)* |
| Mean\_income  *(integer)* |
| Median\_income  *(integer)* |
| Std\_income  *(integer)* |

Questions:

1. How to get Id to properly increment on database

1. **Create API request for Yelp [**populate\_yelp\_table()**][**yelp\_api()**]**
   1. **Inputs**
      1. Base: 'https://api.yelp.com/v3/businesses/search'
      2. HTTP Auth: headers={'Authorization': 'bearer %s'%YELP\_KEY}
      3. Parameters: term=’restaurants’
      4. *Zipcode search*
         1. Params: Lat and Lon [from SQL query on zipcodes table]
      5. *Full Address search*
         1. Params:location [from user input]
   2. **Returns** (json file)
      1. 20 restaurants in location with:
         1. Business name
         2. Business type
         3. Business lat and lon
         4. Business price($-$$$$)
         5. Ratings(1-5)
         6. URL to site

|  |
| --- |
| **YelpResults** |
| Id  *(Primary Key)* |
| Zipcode\_id  *(integer)* |
| Business Name  *(integer)* |
| lat  *(integer)* |
| Lon  *(integer)* |
| Business\_price  *(text)* |
| Ratings  *(integer)* |
| Category  *(text)* |
| Site\_url  *(text)* |

1. **Create API request for Zillow[**populate\_zillow\_table()**][**Zillow\_api()**]**
   1. **Inputs**
      1. Base: ‘http://www.zillow.com/webservice/GetSearchResults.htm?’
      2. Key: zws-id=<ZWSID>
      3. Parameters: address[user input],citystatezip[user input],rentzestimate[‘true’]
   2. **Returns** (html file)
      1. URL to home details page
      2. Zestimate house value ($)
      3. Zestimate valuation high and low ($)
      4. Percentile value
      5. Rent Zestimate ($)
      6. Rent Zestimate valuation high and low ($)

|  |
| --- |
| **ZillowResults** |
| Id  *(Primary Key)* |
| Zipcode\_id  *(integer)* |
| Zestimate\_home\_value  *(integer)* |
| Zestimate\_rent\_value  *(integer)* |
| lat  *(integer)* |
| Lon  *(integer)* |
| Zestimate\_home\_high\_val  *(integer)* |
| Zestimate\_home\_low\_val  *(integer)* |
| Zestimate\_rent\_high\_val  *(integer)* |
| Zestimate\_rent\_low\_val  *(integer)* |
| Site\_url  *(text)* |

1. **Create Queries [**zipcode\_query()**]**
   1. **Zipcode + Home** 
      1. Return: zipcode, state, mean income+std, 2017 mean housing, 2017 std housing
   2. **Zipcode + Rent** 
      1. Return: zipcode, state, mean income+std, 2017 mean rent, 2017 std rent, lat(zip), lon(zip)
   3. **Zipcode + Yelp**
      1. Return: zipcode, state, name, business price, rating,category, lat(zip), lon(zip), lat(business), lon(business)
   4. **Zillow Address + Home**
   5. **Zillow Address + Rent**
   6. **Zillow Address + Yelp**
2. **Process Queries**
   1. **Zipcode +Yelp [**process\_query\_yelp()**]**
      1. [name],[price],[rating],[category],[lat\_buisness],[lon\_buisness], [lat\_zip], [lon\_zip]
   2. **Zipcode + Rent**
   3. **Zipcode + Yelp**
   4. **Zillow Address + Home**
   5. **Zillow Address + Rent**
   6. **Zillow Address + Yelp**
3. **Plotly Plots**
   1. **Zipcode + Yelp [**yelp\_plotly()**]**
      1. Shows all restaurants in the area
   2. **Home Box Plot [**homeprices\_plotly()**]**
      1. Show the mean median and mode for home values in zipcode
      2. Takes 2017 and 2018 mean +- standard deviation for each value
   3. **Rent Box Plot[**homeprices\_plotly()**]**
      1. Show the mean median and mode for rent values in zipcode
      2. Takes 2017 and 2018 mean +- standard deviation for each value
      3. Just change query from home to rent to get proper output