Discover St. Louis (Business)  
Project 1 – RAJ SIKKA

August 20, 2018

# Overview

## Project Background and Description

**Group Members** – Rori Cooper, Rebecca Kalhorn, and Raj Sikka

**Background** – Our group initially thought about probing into ATM and banking data; however, after having difficulties finding public banking data we decided to try a different approach. It turns out Rebecca and Rori both reside in St. Louis City and collectively came up with the idea to go on a hunt for public data to explore what is unique about St. Louis neighborhoods. We eventually found out that identifying datasets was trickier than we thought, but that’s what made our project adventurous and fun. The sole purpose was a fact-finding mission using descriptive statics.

**Goals** - Our project is to describe each zip code in Saint Louis and provide an overview of their economies and demographics. We will compile and analyze multiple data sources and let the data show what is unique about each zip code.

**Questions** –

•What is unique about each zip code located within the Saint Louis city limits?

•How do the zip codes (neighborhoods) compare by a number of demographic variables using descriptive statistics?

•If you were moving to Saint Louis and were new to the area, which is the most attractive zip code to live and why?

**Description** - Use public data to show the distinctive neighborhood characteristics of St. Louis City.

## Project Scope

We delegated data-gathering duties amongst each teammate respectively:

Rori – Education

Rebecca – Census Demographics

Raj – Business

The objective was to gather the data related to the number of Businesses categories (see the list below) present in St Louis City. The analysis would provide a snap shot to an individual when he/she is checking out a particular area with the intend to live or visit.

* Food (Fast Food Joints, Restaurants)
* Stores (Department, Grocery, Gas)
* Financial Institutions (Banks, Loans etc.)
* Hospitals, Urgent Care Centers
* Universities/Colleges, Daycare Centers
* Hotels, Bed & Breakfast

## High-Level Requirements

|  |  |
| --- | --- |
|  |  |

## Business Data Sources

Yelp was the main source of data. Google was used to validate the response details received from Yelp.

## Methods

The method used to gather and process the business data is stated below.

1. Initial request sent via Postman to Yelp. This Postman tool had to be configured to carry out the request. It contained the designated API key details which I received from Yelp.
2. Captured the response which was in JSON format from Postman into notepad. Repeated this process and gathered the data for the business categories (see above).
3. Reviewed the JSON response and finalized the business categories.
4. Processed JSON file using a customized Python program and generated a CSV file for each business category present in the following zip codes.



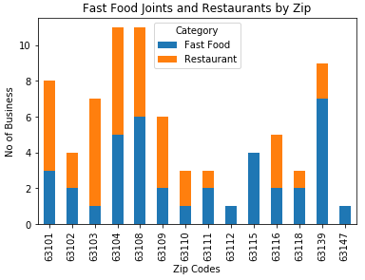
1. The generated CSV file contained the following information for each business.
   * Business Name
   * Address
   * City
   * Zip
   * State
   * Category
   * Latitude
   * Longitude

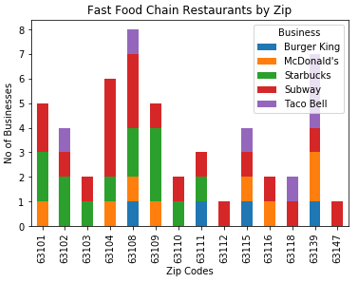
## Analysis.

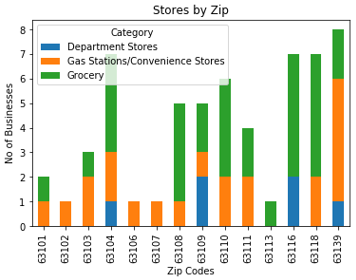
The Data analysis steps are stated below.

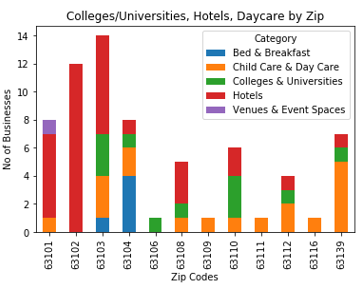
1. Converted each CSV file for each category into a Data Frame using Pandas.
2. Gathered the counts for a particular business category present in each Zip using Pandas.
3. Merged the Data Frames for each sub-categoriey to main categories and gathered counts for businesses present in zip.
4. Generated bar charts and stacked bar charts using Matplotlib to represent the information gathered.
5. Generated a consolidated CSV file by merging all the CSV files (generated in step a) using a customized Python program.
6. Displayed the results on a map using Excel 3D-Maps.

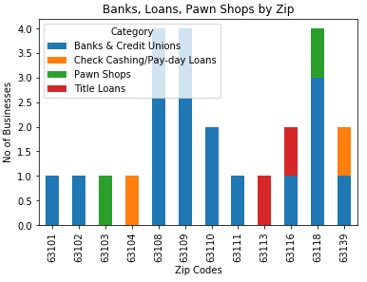
## Results

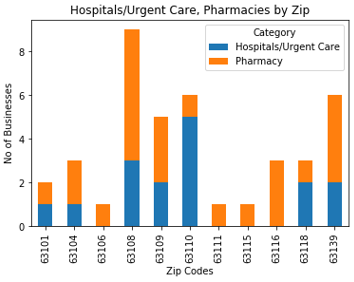










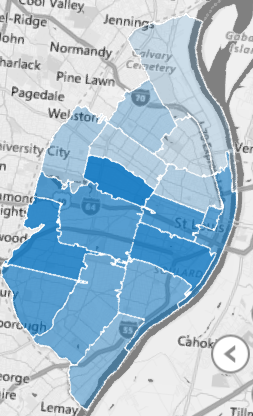


## Conclusions

1. Overall findings are:
   * 63108, 63139, and 63109 is the top zip codes as it contains all the businesses categories which an individual would want to see. Please note that there might be instances where there might be a Zip codes which have more business than the listed earlier.
   * 63101, 63102 and 63103 have 6, 12 and 7 hotels respectively. This because they are in the downtown vicinity which often has a lot of visitors.
   * It appears that north part of St louis city has fewer businesses than the south. The results state the fact. The lighter shade on the map below demonstrate the fact.

63105 – 3 63106 – 3 63107 – 1

63112 – 5 63113 – 2 63147 – 1



## Next Steps

I plan to do the following to improve the current solution by doing the following

* Plan to have a strategy when trying to pull the data using Google’s Geo locator for conducting Business search.
* Expanding the Data Visualizations by using maps.
* Providing a rating module which will assist in providing a rating for each zip code.

## Appendix

To view the map in Step 8 please refer to the following spreadsheet in Excel (version 2016) press the 3D-map button (present on the insert menu). Once the map is initiated hover over each shaded area to see the total number of businesses by Zip.



To see the businesses in a Zip by business categories. See the attached spreadsheet.

