**Battle of the Neighborhoods in Saint Louis**

Kalpana Joshi

August 26th 2020

1. Introduction
   1. Background
   2. Problem
   3. Interested Parties
2. Data Acquisition
   1. Data Sources
   2. Data Transformation
3. **Introduction**

**1.a Background**

Saint Louis is a quiet city in Missouri, right in the heart of the United States of America. Its widely known for its very popular baseball team, the Cardinals, which attracts a lot of tourists during the game season. The city has several natural places to visit during summer time like lakes, adventure trails and parks. There are many facilities provided to the public free of cost like the Saint Louis Zoo, Art Museum, History Museum and the Science Centre. Along with these free avenues there are several malls, restaurants, eateries which provide the residents a plethora of activities to keep themselves busy during weekends or holidays. In this project we will be exploring the various Neighborhoods in the city of Saint Louis and the popular venues in each of these.

**1.b Problem**

The project aims to explore the venues in the different parts of Saint Louis and determine the most popular ones in a certain area. This analysis is aimed at finding out the best suited locations to start up a new business like a restaurant or a clothing store.

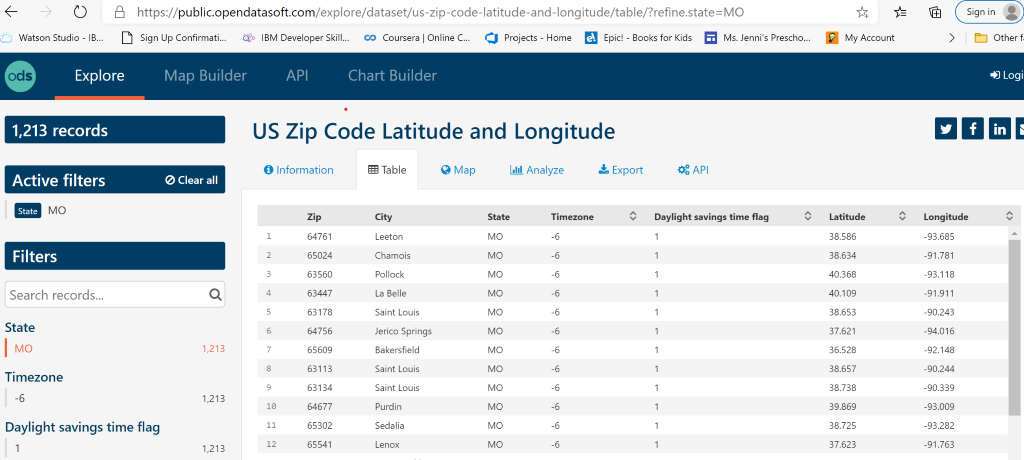
**1.c Interested Parties**

The result of this project will benefit many of the entrepreneurs in Saint Louis to determine the location that is best for them to set up shop. Since we will be assessing the popular venues in various locations, we will get an idea of the kind of customers and their purpose while visiting these locations. This will increase the chances of them visiting the stores of these entrepreneurs if they set up their business accordingly.

1. **Data Acquisition**

**2.a Data Sources**

The data for this project has been accessed from the website: <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/> , which has a detailed list of all the zip codes of the US cities along with other information such as City, State, Longitude, Latitude, Time Zone, Daylight savings time flag and Geopoint. This data can be filtered by State. Below is a snapshot of the data filtered for the State of Missouri where Saint Louis is located:



Above data can be exported in the csv format and then read into a pandas Dataframe.

We will also be utilizing the Foursquare API to determine the venues at each of these locations. Below is a screenshot of the information received from Foursquare for a specific location. The highlighted portions are the ones we will be focusing on during our analysis.



**2.b Data Cleaning**

The data extracted from the website below is in the csv format.

<https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/>

Fist we need to create a Dataframe with the data in the above file. This data is at the state level, so we need to filter it further by city (Saint Louis).

**2.c Feature Selection**:

We can further drop fields which will not be used during our analysis like ‘Time zone’, ‘Daylight Saving time flag’ and ‘Geopoint’. The final Dataframe had about 71 samples or information about 71 Postal Codes in the city of Saint Louis with their respective Latitudes and Longitudes.