**The battle of Chicago neighborhoods**

**April 1st, 2019**

1. **Introduction:** 
   1. **Background:** There’s plethora of reasons why people love coffee shops.Few of them being; a coffee shop can be considered a third place away from home and work, it is considered as a creative place to get one’s creative juices flowing and most important would be the caffeine that makes one love cafes so much. But, starting a coffee shop can be an exciting and profitable venture for an entrepreneur with a passion and love for coffee and customer service.Chicago is one of the most popular and busiest Us cities. A business man decides it would be a good business move to open a new café in Chicago.
   2. **Problem:** Before opening a café, it would be obvious to analyze the possible factors that could aid or harm the business. In the scope of this project my goal is to analyze the best neighborhood in Chicago to open a new café. The possible factors to be considered are: Which is the liveliest neighborhood amongst all neighborhoods? How many cafés are there in each neighborhood?
   3. **Interest:** The stakeholders involved in opening a new café would be interested in the analysis of the neighborhoods.

**2. Data Acquisition and Cleaning:**

**2.1. Data Sources:** Many data set relevant tothis project can be found on sites like Kaggle or data.gov. I used web scraping to get the relevant information from Wikipedia page (url: <https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago>). The latitudes and longitudes corresponding to each neighborhood was calculated using OpenStreetMap API and geopy. geocoders nominatim() function.

**2.2.**  **Data Cleaning:** The data scraped from the web was converted into a data frame. There were some rows with null values which were dropped. The data frame consisted of 4 attributes namely: Neighborhood and Community area. The data frame also consisted of 126 values.

**2.3. Feature Selection:** As only relevant attributes were in the data frame this step was skipped.

**3. Exploratory data analysis:**

**3.1. Exploring the most popular neighborhood:** The loop one of Chicago’sneighborhood is considered the central business district and downtown area of the city. Considering the popularity of this neighborhood it would be ideal to analyze the number of total venues and the number of total café’s in this neighborhood.

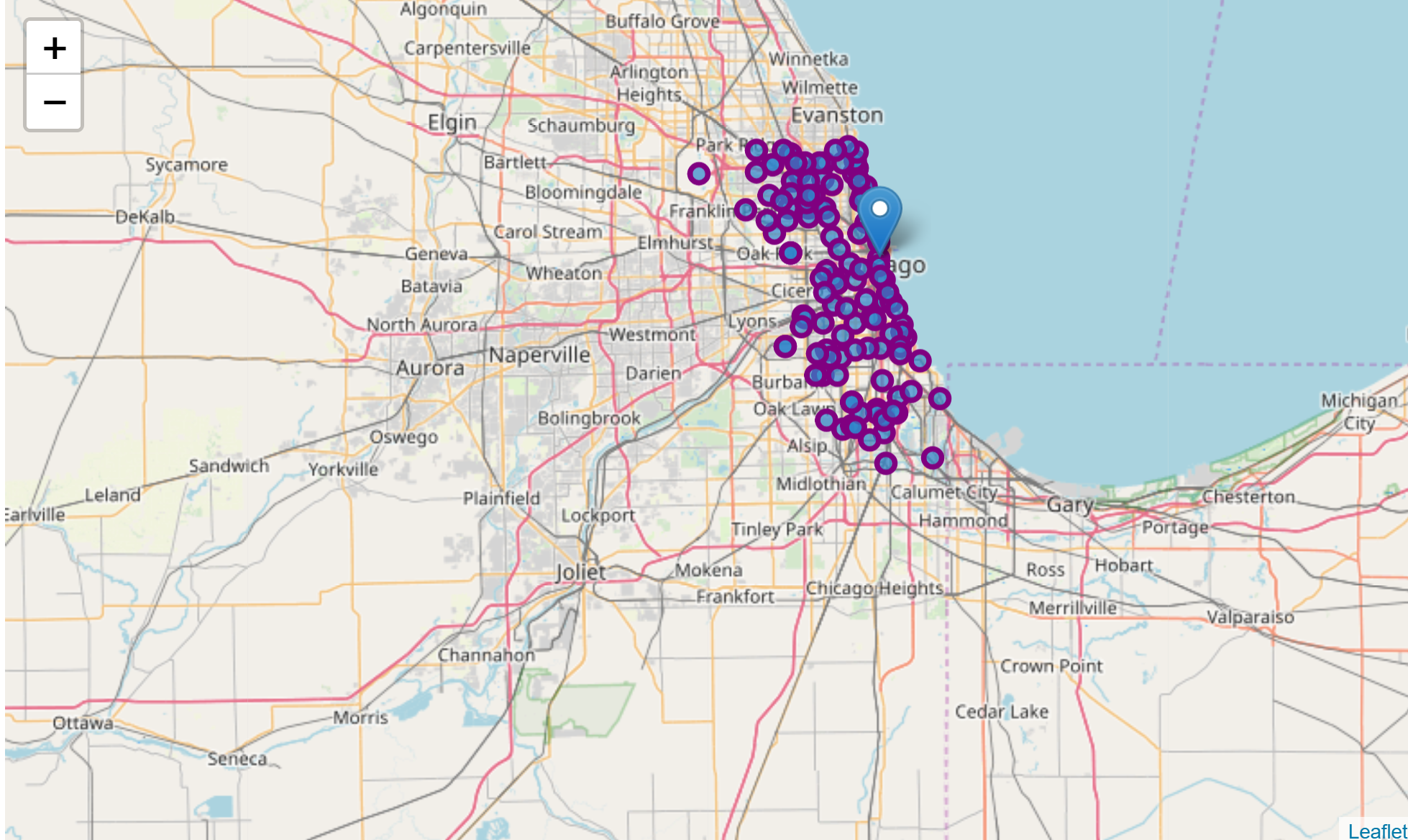


Fig 1: Map indicating the location of The loop (with marker) and other neighborhoods.

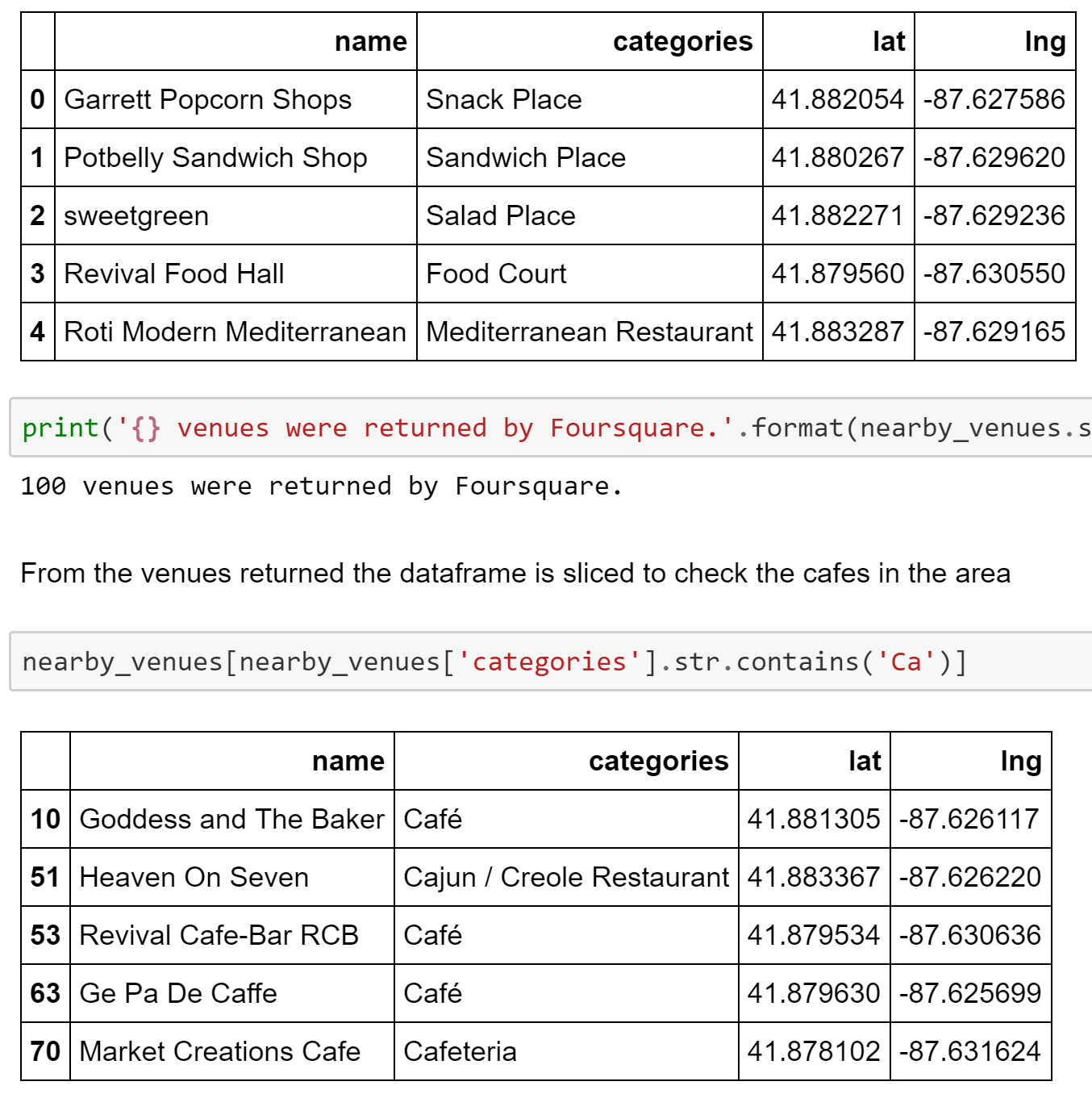


Fig 2: Show the total venues and total café in The Loop.

From the above figure it would be safe to assume that there are a lot of venues in the loop neighborhood. There seem to be a decent number of cafes in this area (4 to be precise). Therefore, it would be plausible to open a new café in this area. But, for better insight other neighborhoods were analyzed as well.

**3.2. Analyzing all neighborhoods:** In order to be sure that a correct location is chosen all other neighborhood should be analyzed as well. Top venues in each neighborhood were found. A new data frame containing relevant information was returned. The data frame contains venue attributes and neighborhood attributes. This data frame contains 2761 values and 7 attributes. The count of cafes in each neighborhood was calculated and stored in a new data frame. This new data frame was used to plot a histogram.

A screenshot of a cell phone

Description automatically generated

Fig 3: Histogram of café count per neighborhood vs frequency.

The histogram can be analyzed for further insight. On a positive side there seem to be no cafes in a few neighborhoods. But it seems that most neighborhoods seem to have already many cafés. The names of these neighborhoods with many cafés (<=3 | >=4) was retrieved to keep them in mind, as it wouldn't be wise to open our new cafe in such café-rich neighborhoods.

A new data frame was created consisting of neighborhood and occurrence of each venue category for that neighborhood. Finally, this data frame was used to get the top 5 venues in each neighborhood.

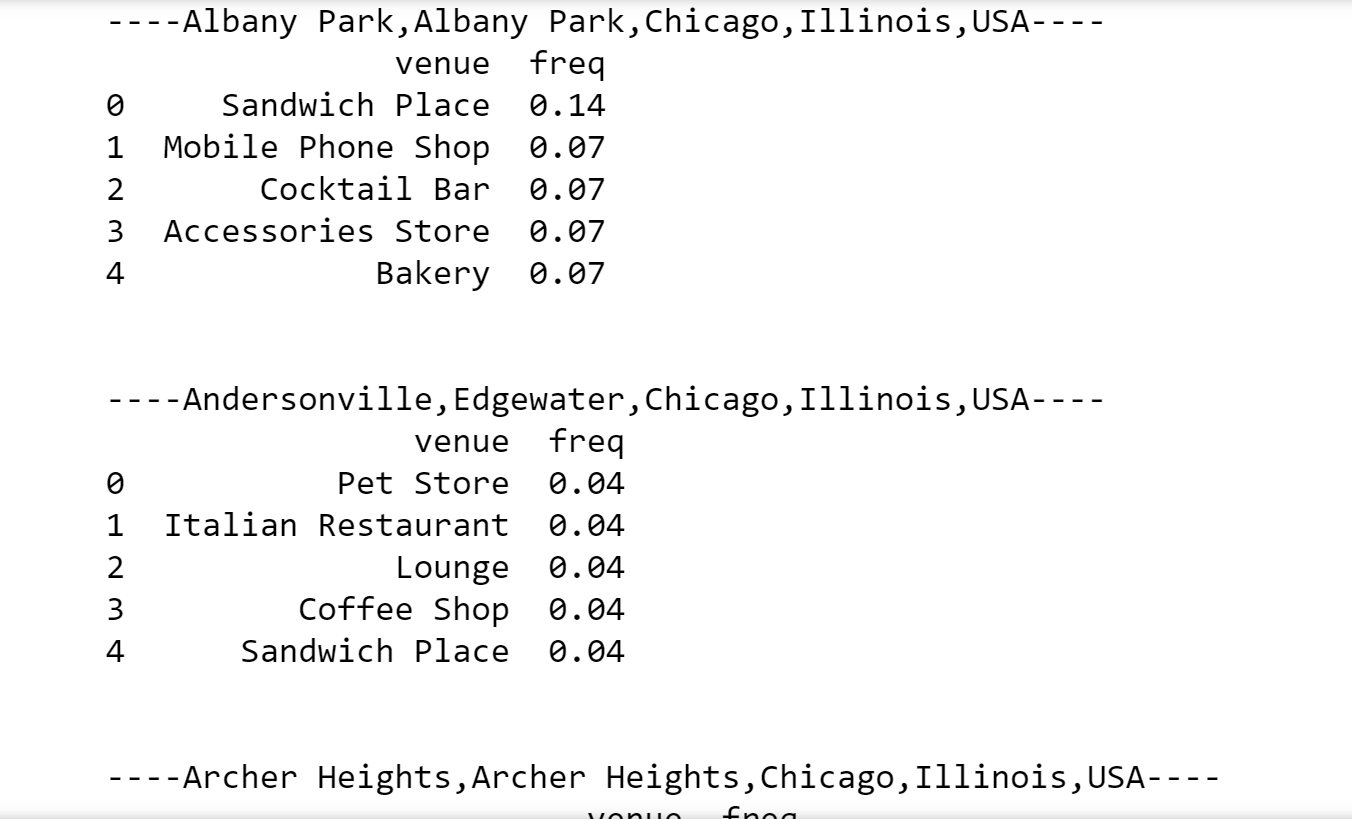


Fig 4: A snippet of 5 venues in each neighborhood.

This information could be used to analyze whether to open our new café in a neighborhood or not. If a neighborhood contains café in top 5 venue categories then it would not be a wise business move to open our café in these neighborhoods.

**4. Clustering:** Finally, the neighborhoods were segregated into three clusters and the clusters were analyzed for their similarities.

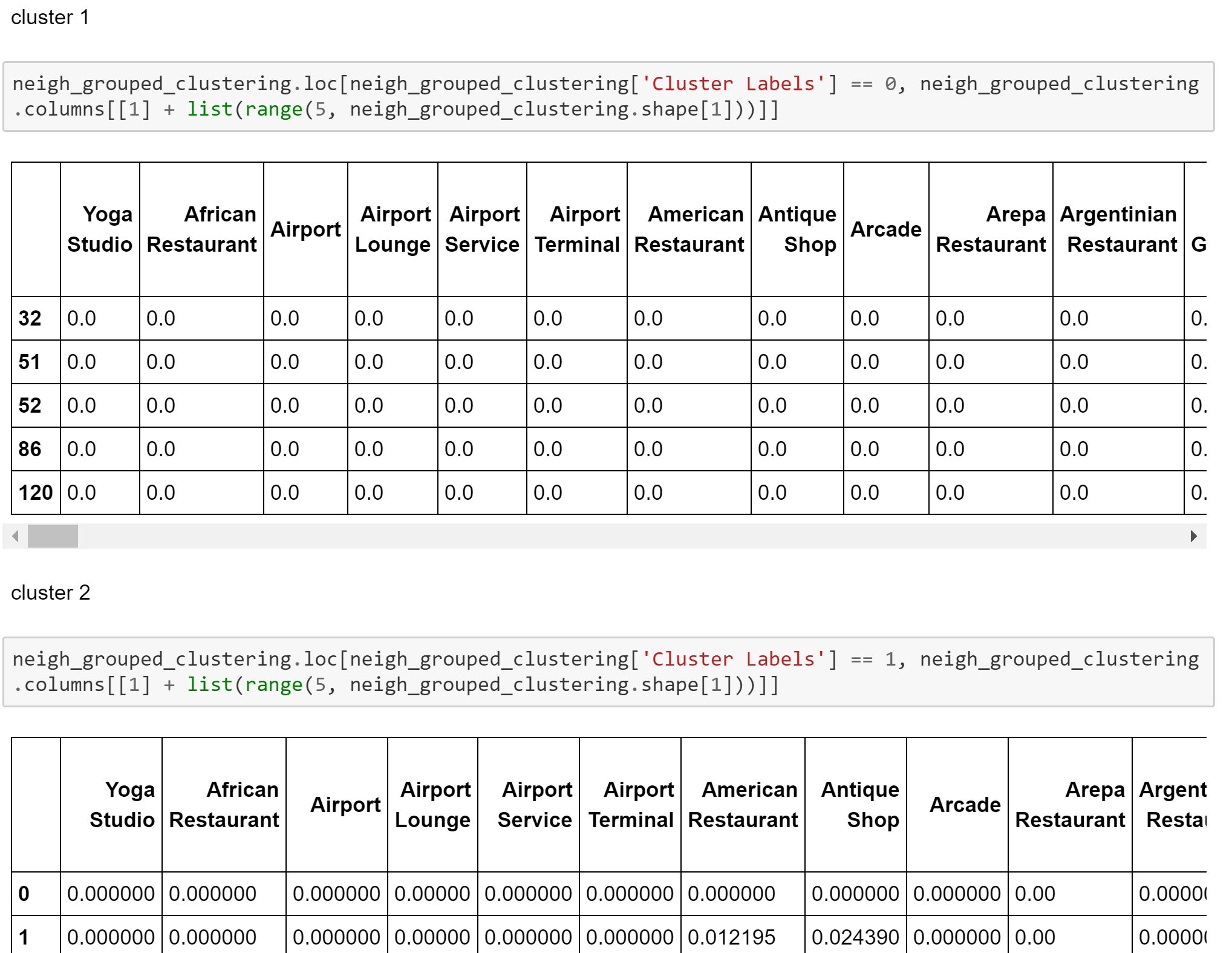


Fig 6: Snippet of Cluster 1

**5. Conclusion:** We have a nice overview of Chicago's neighborhoods along with their venue characteristics, number of coffee shops. This analysis should aid investors in finding the optimal place to open a new café. The analysis gives a list of lively neighborhoods with many other Restaurant venues around, but not too many cafés (>5). In my opinion The Loop is the ideal location to open a new café since its located well in the city and has not too many cafés (4).

**6. Future Scope:** This project can be modified and used to analyze other venue types as well. For example; if a business person wants to know whether to open an Asian restaurant or not. This project can also be expanded to analyze other popular cities and their corresponding venues.