1. Introduction:

1.1. 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 cafe in Chicago.

1.2. Problem: Before opening a cafe, 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?

1.3. Interest: The stakeholders involved in opening a new cafe would be interested in the analysis of the neighborhoods.

2. Data Acquisition and Cleaning:

2.1. Data Sources: Many data set relevant to this 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.