

IBM Applied Data Science Capstone

Analyzing Neighborhoods in Chicago, IL

Israel Tshikoji

I. Introduction

1. Background

Chicago is the most populous city in the US State of Illinois, and the third most populous in the United States. With an estimated population of about 3 million people, it is the most populous city in the US Midwest.

Chicago is an international hub for finance, commerce, industry, technology, etc.

Taking into consideration its big population size and its financial relevance, there is a definite opportunity to provide enough facilities (such as school, parks, restaurants, retail stores, etc.) to accommodate the needs of its residents and visitors.

Providing business amenities to a big population as the one under consideration is a business opportunity that any good investor would like to explore further. Although there are many types of businesses that can be studied and tested, this project will zoom in a retail business idea.

2. Business Problem

While it may be reasonably assumed that opening a retail store in the Chicago City area is a sound business idea, this project will try to locate which neighborhoods would offer a greater profitability for such an endeavor.

This project is carried out in order to advise a group of retail investors about the potential benefits of investing in the retail business in the Chicago city area.

II. Data

The list of Chicago neighborhoods will be scraped from the following website:

https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago

The raw data off the webpage will be transformed into a pandas dataframe.

The Geopy library will be used to get the latitude and longitude values of desired neighborhoods.

The Foursquare API will then be used to explore neighborhoods and segment them according to the most prominent business activity.

Analyzing the clustered neighborhoods, useful insights will be drawn to help recommend suitable locations for retail businesses.