Coursera Capstone Project: IBM Data Science

Mwimbi Shindola

mwimbi.shindola@gmail.com

Lusaka, Zambia

1. **Introduction and Business Problem**

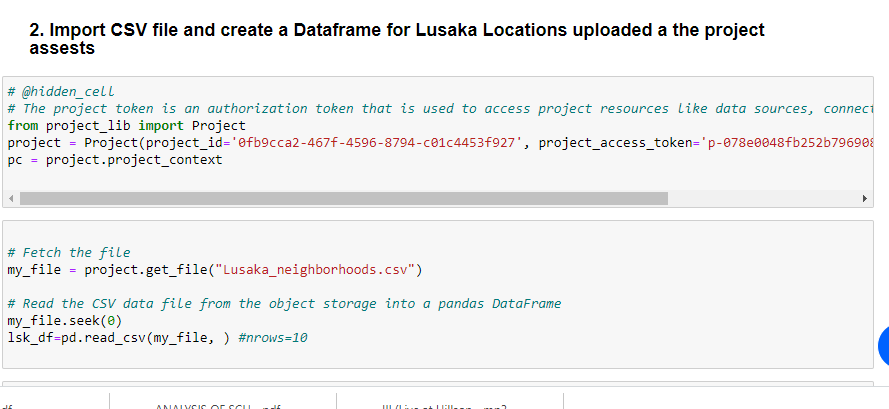
Lusaka is the capital city of Zambia and is the Largest city by population in the county. As of 2020 It has a population of about 2,774,000 people according to Macprotrends.net. Being the most populous city there is need to properly plan how institutions of learning are located. One of the challenges parents face with growing traffic is finding the nearest schools to take their children to. This project intends to identify areas where schools are located further away. Information from this research can be used by families with school going children when planning which areas to settle, the government when identifying which areas to build more schools or private investors who may intend to venture into private schools to bridge the gap in neighborhoods without schools.

1. **Data**

This section describes how data for the project was retrieved and processed. Multiple sources and methods where used to make the data usable for analysis. The following subsections sections describe the data collected and how it was processed at different levels.

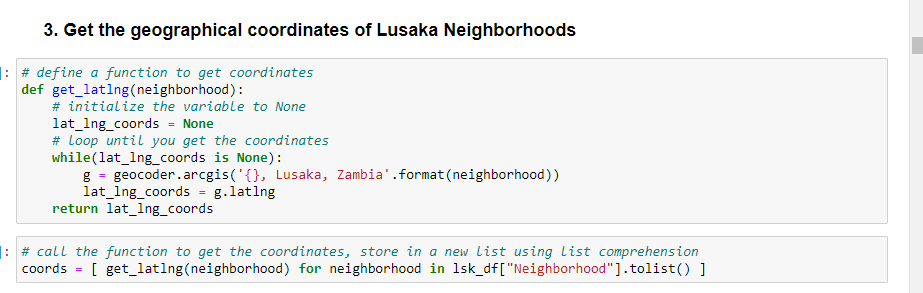
* 1. **Neighborhoods**

The data of the neighborhoods in Lusaka was collected from 2 different sources: Wikipedia page and macpro.info website then created a csv list of the neighborhoods was created and uploaded as a data asset on the Watson IBM project platform. This CSV file was then used to create a neighborhoods data frame in the project notebook.



* 1. **Geocoding**

The file contents from Lusaka\_nieghborhoods.csv are retrieved into a Pandas DataFrame. The latitude and longitude of the neighborhoods are retrieved using Google Maps Geocoding API. The geometric location values are then stored into the a dataframe which was later merged with the neighborhoods list dataframe.



* 1. **Venue Data**

From the location data obtained, the venue data is found out by passing in the required parameters to the FourSquare API, and creating another DataFrame to contain all the venue details along with the respective neighborhoods.

Code

