Opening a new mall in the city of Strasbourg

1.Introduction

1.1 Background

People like to do several activities during the weekends or during the holidays like shopping, eating at restaurants with friends. Several shopping malls particulary in the city of Strasbourg allow them to realize their dreams by offering them a wide range of activities as eating in a restaurant. For an investor who wants to meet the customer's needs, opening a mall in a place where customer traffic is important is important. That will determine whether he is going to have a return on investment or go bankrupt. Therefore, doing web scraping in order to leverage data by cleaning, exploring and clustering the neighborhoods is crucial.

1.2 Business Problem

Data that might contribute to determine the best place to open a mall should include the list of neighborhoods in Strasbourg, Latitude and Longitude coordinates of these neighborhoods and venue data particularly the venue related to shopping malls. This project aims to determine where will a stakeholder open a mall in order to raise his turnover and make a perk

1.3 Targeted Audience

This project particularly concerns those who want to become shareholders or who want to open a shopping mall in the best places on the neighborhoods in Strasbourg

2. Data acquisition and wrangling

2.1 Data sources

https://fr.wikipedia.org/wiki/Liste_des_quartiers_de_Strasbourg is a wikipedia page that have all the neighborhoods in Strasbourg. There are 15 neighborhoods. We will scrape the Wikipedia page and wrangle the data, clean it, and then read it into a *pandas* dataframe. Many data science skills are required to lead this project as :

•Using web scraping allows us to extract the data with the help of libraries such as BeautifulSoup4 and pandas. Then we will clean the data

•Importing Nominatim library from Python Geocoder package in order to transform addresses of the neighborhoods into latitude and longitude coordinates. This will be useful for visualize the data

•Using Foursquare API in order to get the venue data of those neighborhoods. Foursquare provides many categories of the venue data but we are interested in Shopping mall category.

•Doing Machine Learning by using K-means clustering

•Visualize the data by using the folium library