Capstone Project - The Battle of Neighborhoods

(Week 1)

Report

Introduction

For many shoppers, visiting malls is a great way to relax and have fun during weekends and holidays. Shopping centers are like a single destination for all types of buyers. For the traders the central location and the large crowd in the shopping centers constitute an excellent distribution channel for marketing their products and services. As a result, there are many shopping centers in Lille. Of course, as with any business decision, opening a new shopping center requires serious thought and is much more complicated than it seems. In particular, the location of the mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

Business problem

The objective of this project is to analyze and select the best locations in the city of Lille, in France, to open a new shopping center. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: if a property developer is looking to open a new shopping center, where would you recommend that they open it?
This project is particularly useful for property developers and investors who wish to open or invest in new shopping centers.

Data

To solve the problem, we will need the following data:
• List of districts of Lille.
• Latitude and longitude coordinates of these districts. This is necessary to draw the map and also to obtain the data of the place.
• Site data, in particular data relating to shopping centers. We will use this data to cluster on neighborhoods.
Data sources and methods to extract it :

This web page (http://www.mapcrow.info/Lille-FR-suburbs).We will use web scraping techniques to extract data from the web page, using queries Python and beautifulsoup packages. Then we will get the geographic coordinates of the neighborhoods using the Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.
After that, we will use the Foursquare API to get the location data for these neighborhoods. Foursquare has one of the largest databases of more than 105 million locations and is used by more than 125,000 developers.
The Foursquare API will provide many categories of data on the sites, we are particularly interested in the Shopping Mall category to help us solve the proposed business problem. This is a project that will use many skills in data science, from web scraping, to work with API (Foursquare), to data cleaning, to data manipulation, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps followed in this project, the data analysis we did and the machine learning technique that was used.