Coursera Capstone

IBM Applied Data Science Capstone

Opening a new Shopping Mall in Toulouse France

By: Nestor Carmona Moreno
January 2019

Introduction

For many shoppers, going to shopping malls is a great way to relax and enjoy themselves during weekends and holidays as they offer a wide variety of “activities” such as grocery shopping, dining, shopping, watching movies or any other activity. It is safe to say that shopping malls fit at least one person in the world. For retailers, the central location and the crowds at shopping malls makes it easier for them to sell their products and/or services to respond to the increasing demand.

Finally, opening a shopping mall allows property managers to have a consistent source of rental income. Nevertheless, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, its location is one of the key aspects as it can determine the success or failure of the investment.

Business problem

The objective of this capstone project is to analyse and select the best locations in Toulouse, France to open a new shopping mall using Data Science methodology and Machine Learning techniques like clustering. This project aims to provide solutions to answer a business problem: In the city of Toulouse, France, if a property manager is looking to open a new shopping mall, in which neighborhood would we recommend them to open?

Target audience of this project

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in Toulouse, France. This project is timely as the city currently offers a limited amount of shopping mall in its most centric neighborhoods.

Data

To solve the problem, we will need the following data:

- List of neighborhoods in Toulouse: chose city for this project.
- Latitude and longitude coordinates of those neighborhoods: required in order to plot the map and get the venue data.
- Shopping mall data: required to perform clustering on the neighborhoods.

Sources of the data:

- List of neighborhoods: We will use this Wikipedia page: https://fr.wikipedia.org/wiki/Catégorie:Quartier_de_Toulouse using web scraping techniques to extract the data with the help of Beautifulsoup packages and Python requests.
- Latitude and longitude coordinates: we will use Python GeoCoder
- Shopping mall data: we will use Foursquare API to get all the venue data and, in particular, the Shopping Mall category.

This is a project that showcases many data science skills, from web scraping to working with an API, Data Cleaning, Data Wrangling, Machine Learning and Data Visualization.

In the next section, we will present the Methodology where we will discuss the steps taken in order to answer the above business question, all the Data Analysis that we did and the Machine Learning techniques used.