CAPSTONE PROJECT - THE BATTLE OF NEIGHBORHOODS

Created with IBM Watson Studio

SUITABLE NEW STORE LOCATIONS IN PARIS FOR A FASHION RETAILER

This notebook contains multiple parts:

Week 1

1. A description of the problem and a discussion of the background - Week 1
2. A description of the data and how it will be used to solve the problem - Week 1

Week 2

1. Methodology and Exploratory Data Analysis - Week 2
2. Inferences and Discussion - Week 2

The separate report goes into more description of the methodology.

INTRODUCTION AND DISCUSSION OF THE BUSINESS OBJECTIVE AND PROBLEM

Locations for New Fashion Stores in High Traffic Areas in Paris France

The Task At Hand

A digitally native vertical fashion retailer, with a substantial e-commerce footprint, has begun the rollout of brick and mortar stores as part of their Omni channel retail strategy. After rolling out stores in a few select cities by guessing where the best locations were to open, as part of their store expansion for Paris they have decided to be more informed and selective, and take the time to do some research.

I have given the exciting task of assisting them to make data-driven decisions on the new locations that are most suitable for their new stores in Paris. This will be a major part of their decision-making process, the other being on the ground qualitative analysis of districts once this data reviewed and studied.

The fashion brand not considered as high end, there positioned in the upper end of the fast fashion market. As such, they do not seek stores in the premium upmarket strips like Avenue Montaigne, but rather, in high traffic areas where consumers go for shopping, restaurants and entertainment. Foursquare data will be very helpful in making data-driven decisions about the best of those areas.

Criteria

Qualitative data from another retailer that they know, suggests that the best locations to open new fashion retail stores may not only be where other clothing is located. This data strongly suggests that the best places are in fact areas that are near ***French Restaurants, Cafés and Wine Bars***. Parisians are very social people that frequent these place often, so opening new stores in these locations is becoming popular.

The analysis and recommendations for new store locations will focus on general districts with these establishments, not on specific store addresses. Narrowing down the best district options derived from analysis allows either further research to be conduct, advising agents of the chosen district, or on the ground searching for specific sites by the company's personnel.

Why Data?

Without leveraging data to make decisions about new store locations, the company could spend countless hours walking around districts, consulting many real estate agents with their own district biases, and end up opening in yet another location that is not ideal.

Data will provide better answers and better solutions to their task.

Outcomes

The goal is to identify the best districts - *Arrondissements* - to open new stores as part of the company's plan. The results will be translate to management in a simple form that will convey the data-driven analysis for the best locations to open stores.

THE DATA SCIENCE WORKFLOW

Data Requirements

The main districts in Paris are divided into 20 *Arrondissements Municipaux* (administrative districts), shortened to *arrondissements*.

The data regarding the districts in Paris needs to research and a suitable useable source identified. If it founds but not in a useable form, data wrangling and cleaning will have to be performed.

The cleaned data will used alongside Foursquare data, which is readily available. Foursquare location data will be leveraged to explore or compare districts around Paris, identifying the high traffic areas where consumers go for shopping, dining and entertainment - the areas where the fashion brand are most interested in opening new stores.

The Data Science Workflow for Part 1 & 2 includes the following:

- Outline the initial data that is required:
  - District data for Paris including names, location data if available, and any other details required.

- Obtain the Data:
  - Research and find suitable sources for the district data for Paris.
  - Access and explore the data to determine if it can manipulated for our purposes.

- Initial Data Wrangling and Cleaning:
  - Clean the data and convert to a useable form as a data frame.

The Data Science Workflow for the next section:

- Data Analysis:
  - Foursquare location data will leveraged, to explore or compare districts around Paris.
  - Identifying the high traffic areas using Data Visualization and Statistical Analysis.
- Machine Learning:
  - Analysis and Visualization with Clustering.
  - Data Visualization using Choropleth Mapping.

DATA RESEARCH AND PREPARATION

Import the Paris District Data

Arrondissements Municipaux for Paris CSV (administrative districts)

Paris is divided into 20 Arrondissements Municipaux (or administrative districts), shortened to just arrondissements. They and normally referenced by the arrondissement number rather than a name.

Data for the arrondissements is necessary to select the most suitable of these areas for new stores.

Initially looking to get this data by scraping the relevant Wikipedia page (https://en.wikipedia.org/wiki/Arrondissements_of_Paris), fortunately, after much research, this data is available on the web and can be manipulated and cleansed to provide a meaningful dataset to use.

Data from Open|DATA France:
https://opendata.paris.fr/explore/dataset/arrondissements/table/?dataChart

Also available from Opendatasoft:
https://data.opendatasoft.com/explore/dataset/arrondissements%40parisdata/export/

DISCUSSION OF THE BUSINESS OBJECTIVE AND PROBLEM / THE DATA WORKFLOW

We now have located and imported the relevant data for the districts of Paris, and have constructed a data frame.

Our business objective, strategy and methods to achieve our goal. Which has laid out and a data workflow established.

Next up, we will leverage Foursquare location data to obtain data on high traffic areas - where consumers go for shopping, restaurants and entertainment - in all of the 20 districts. The Battle of Neighborhoods continues in the next section.