Applied Data Science

week 2

Capstone Project — IBM Data Science

THE BATTLE OF NEIGHBORHOODS

An analysis of neighborhoods in Paris using data science methodologies

Abstract

Currently, we are sitting on a mountain of data, information that must be classified and ordered so that it can be useful, in this case, we will use as an example the deployment of a business model, where information and data can determine the success or failure of it. We will try to determine the best location for a restaurant in the city of Versailles, Paris, France.

1 Introduction

1.1 Background Information

Just like all over the world, Asian food is taking an important place in people' s food culture, and especially France is no exception.

This project aims to estimate the best location to open a business of this type in the city of Versailles, near Paris.

Before launching and starting a business, it is important to know if the model will be a good opportunity. For this, this report will try to collect data on the location of other restaurants, possible competition and the best location.

In this project, we are going to use Foursquare data to explore and compare different neighborhoods. Applying machine learning algorithms to find the most promising district in Paris.

This data could be used to determine a subsequent business plan.

1.2 Problem statement:

The problem or uncertainty is where exactly to place the restaurant.

As the goal of this is to create a business plan in the end, we need to make sure data from API are correct. We also need to check that customer could be interested in this specific business.

In order to do so, a survey in Paris and Versailles will be done in addition to data gathering. I'll go in the cities and check at different hours if restaurants are working, if streets are full and so on, and what king of restaurant works well. This survey will allow to validate the data analysis done here.

2 Data

2.1 Description of the Data

This notebook is highly inspirated by the template given in the course. I will keep the idea of clustering the city by area and then plot heatmap to find better area.

I will change some data:

- ✓ Country/City: France, Paris.
- ✓ Goal: Open a restaurant/little shop for workers in weekday and maybe Saturday

I will use the following API:

- ✓ Foursquare API: to find restaurant/venues
- ✓ Google API: reverse geolocalisation.

2.2 How the data will be used to solve the problem

Let' s create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is aprox. 1.5km killometers centered around Versailles city center.

Let' s first find the latitude & longitude of Versailles city center, using specific, well known address and Google Maps geocoding API.

We' ll consider the Prefecture to be the city center, as a lot of companies are around.