IBM Data Science Professional Certificate

Capstone Project

The Battle of Coffee-Shops in Prague City

Antonio Miranda

31st January 2021

Contents

1. INTRODUCTION

1.1. Business Problem

An investor wants to open a new coffee shop in Prague but he isn't convinced yet about the location based only in the information they have acquired so far. He requested us to complement his study analysing the local based in few more criteria which were missed out in the initial report:

- A detailed mapping of existents coffee shops per region
- A demographic study of each region to help them to target the right population niche
- Add tourism potential into perspective. Prague receives about 8.2 million tourists per year
- Use companies size (# employees) to spot areas with high business movement

We will add to the data analysis above Machine Learning algorithm to mining for the gold spot and find the best location to open a coffee-shop in Prague based on these criteria.

1.2. Target Audience

This project is aimed towards investors or business owners who want to open a new coffee-shop in Prague city and also to current coffee shop owners who wants to improve their current business. The results and findings on this report provides crucial information to back-up the place location decision.

2. DATA

2.1. Data Acquisition

Data used in this work were obtained from available public databases and come in different formats.

- Prague City and administrative districts borders

  Shape file format which contains geographic coordinates, borders and area of each administrative district of Prague. Essential to visualize in maps using folium library. Shapefile data: link here. shapefile to geojson: link here

- Foursquare venues location data API endpoint

  Based on the geodata and geo coordinates, we will use foursquare venues API endpoint to obtain the list of coffee-shops in each administrative district of Prague city. Link here

- Prague city demographic distribution data

  Demographic of Prague city per administrative districts with 2019 census. This data set will give us a recent and in depth understand of population in each area based on sex and age. Data source as excel file at Czech Statistical Office: link here

- Prague city tourism distribution data

  Guests in collective tourist accommodation establishments by administrative district in 2019. Data source as excel file at Czech Statistical Office: link here

- Prague city companies size distribution per administrative district

  Companies size of Prague city by administrative districts 2019 census. Data source as excel file at Czech Statistical Office: link here

---------- to come…..

3. METHODOLOGY

4. MACHINE LEARNING

5. RESULTS AND DISCUSSIONS

6. CONCLUSION