APPLIED DATA SCIENCE CAPSTONE

Suitable New Store Locations in Paris for a Fashion Retailer

Quang-Thinh BUI

Faculty of Electrical Engineering and Computer Science
VŠB - Technical University of Ostrava

My Tho City, April 09, 2019

Outline

1. Description
2. The First Week Work
3. The Second Week Work
4. Result and Visualization
5. Discussion and Conclusions

Description

This capstone project is worth 70% of your total grade. The project will be completed over the course of 2 weeks. Week 1 submissions will be worth 30% whereas week 2 submissions will be worth 40% of your total grade.

1. For first week, you will required to submit the following:
   - A description of the problem and a discussion of the background. (15 marks)
   - A description of the data and how it will be used to solve the problem. (15 marks)
2. For the second week, the final deliverables of the project will be:
   - A link to your Notebook on your Github repository, showing your code. (15 marks)
   - A full report consisting of all of the following components (15 marks)
3. Your choice of a presentation or blogpost. (10 marks)

The First Week Work

jupyter Applied Data Science Capstone (Week 1) Last Checkpoint: 16 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Markdown

2. The first week

The description of the problem and a discussion of the background

The problem

In a hypothetical situation, I play a role as a data scientist to support a fashion retailer to open new stores in busy traffic areas in Paris, France

I have been given the interesting task of helping them make data-driven decisions on the new locations that are most suitable for their new stores in Paris. This will be a main element of their decision-making process, the other being on the ground qualitative analysis of districts once when the report for this data is reviewed and studied.

This fashion retailer is positioned in the upper end of the fast fashion market. Therefore, they only focus to 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.

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

The discussion for the background

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 places often, so opening new stores in these locations are 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 for either further research to be conducted, advising agents of the chosen district, or on the ground searching for specific sites by the company's personnel.

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.

The data will provide better answers and better solutions to their task at hand.

More in

`https://github.com/mphuongngo/Applied-Data-Science-Capstone`

The Second Week Work

jupyter Applied Data Science Capstone (Week 2) (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Markdown

3. The second week

3.1. Data Research and Preparation

Import the Required Libraries

In [1]:

```
# library to handle data in a vectorized manner
import numpy as np

# library to handle JSON files
import json

# library to handle data in dataframe
import pandas as pd

# library to convert an address into latitude and longitude values
#!conda install -c conda-forge geopy --yes
from geopy.geocoders import Nominatim

# library to handle requests
import requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors
from bs4 import BeautifulSoup

# library to map rendering library
#!conda install -c conda-forge folium=0.5.0 --yes
import folium

print("Libraries imported successfully!")
```

More in

https://github.com/mphuongngo/Applied-Data-Science-Capstone

Result and Visualization

Frequency distribution for the top 3 venue categories including clothing for each neighborhood

The final 3 prospective neighborhoods for new fashion retail store locations are where 4 criteria, Restaurants, Cafés, Wine Bars, and Clothing Stores, are met: 3eme, 4eme, and 6eme Arrondissement.

Discussion

- I guess it is not a surprise that these districts are all very centrally located in the circular arrangement of Paris's arrondissements. Locations fitting the criteria for popular venues would normally be in central locations in many cities of the world.

- For this visualization it is clear that on a practical level, with no data to base decisions on, the circle of the 20 districts is very large, and researching and then visiting them all would be a daunting and time consuming task. We have narrowed the search area down significantly from 20 potential districts to 3 that should suit the client's retail business.

- More in
  `https://github.com/mphuongngo/Applied-Data-Science-Capstone`

Conclusions

- There are many ways this analysis could have been performed based on different methodologies and perhaps different data sources. I chose the method I selected as it was a straight forward way to narrow down the options, not complicating what is actually simple in many ways – meeting the criteria for the surrounding venues, and in my case, domain knowledge I have on the subject. I originally intended to use the clustering algorithms to cluster the data, but as it progressed it became obvious that this only complicated the task at hand.

- The analysis and results are not an end point, but rather a starting point that will guide the next part of the process to find specific store locations. The next part will involve domain knowledge of the industry, and perhaps, of the city itself. But the data analysis and resulting recommendations have greatly narrowed down the best district options based on the data and what we can infer from it.

THANKS FOR YOUR ATTENTION