Coursera Capstone Project

EXPLORING COFFEE SHOP LOCATIONS IN SOUTH AFRICA

INTRODUCTION

South Africa has upon recently adopted a strict coffee culture with alongside it's booming financial districts. Amazingly so Starbucks and any new to market coffee companies have cropped up in all places. Our client is highly interested in startup up a coffee franchise that will compete locally in this market which is not yet saturated.

BACKGROUND

South Africa is located at the southern most part of the African continent with a mixture of people and seasons, as much as it is of course close to the equator, the country still has some snow and cool temperatures through winter. This calls for the hot beverage industry to shine in these cool winter months.

Companies like Starbucks have recently entered into the South African Coffee market and failed dismally in a short duration. Starbucks will not be opening any new stores as reported by Business Insider https://www.businessinsider.co.za/taste-holdings-owners-of-starbucks-franchise-in-south-africa-halt-expansion-and-no-longer-opening-more-starbucks-stores-in-south-africa-anytime-soon-2018-11

Interestingly so, the demand for local coffee shops are still on the rise and our Client would like to take advantage of that.

PROBLEM STATEMENT

How can we use existing location-based data to deepen our understanding of coffee shops and their locations within South African Financial Hubs?

THE CUSTOMER (TARGET AUDIENCE)

This will compromise all investors or coffee entrepreneurs who desire to use location based quantitative data in addition to their data to determine the successful location of a coffee shop or explore existing locations for purchase.

SELECTION CRITERIA

We will review the province of Gauteng, South Africa and the 3 Most popular financial hubs are:

1. Pretoria
2. Sandton
3. Johannesburg

DATA FLOW

This project would use Foursquare API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

Foursquare

Using central location identifications of Foursquare API features of near-by places of the neighborhoods would be mined and then searched for the Key term "coffee". Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 1000 due to the distances between areas in South Africa.

DATA EXPLORATION

We will review the different coffee shops in the set locations and explore tips, reviews and later cluster the shops according to location.

CONCLUSION

It will be deemed that this data can serve as insight into coffee shop location for the rest of South Africa and more specifically financial hubs.

LIBRARIES:

- ☑ Pandas: For creating and manipulating data frames
- ☑ Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- ☑ Scikit Learn: For importing k-means clustering

- ☑ JSON: Library to handle JSON files
- ☑ Geopy: To retrieve Location Data
- ☑ Requests: Library to handle http requests
- ☑ Matplotlib: Python Plotting Module