Battle of the Cape Town Neighbourhoods.
Coursera Capstone Project

Introduction

We note the growing popularity of social media applications, notably FourSquare (FSq). FSq is notable in that they provide location intelligence to people.

The cape Town CBD is expensive for lcocals to stay in and commute to. Based on data of places that Capetonians frequent, as evidenced by Foursquare data,

We would like to help determine the various segments of areas and their most liked venue categories to determine if it will be a good fit as requried.

Business Problem

If a business is launched in the wrong area, relative to the business type, we could expect that staff would find it difficult to travel to work.

Knowing the different clusters / groups of people and their likes could help a business determine where next to either launch a little coffee shop, or decide where to locate their new offices.

Description of the Data

We obtained South African post codes information via an internet search on http://www.sapostalcodes.info/ .
This data had no geo co-ordinates so we utilised a locational geocorodicates tool to generate the corodinates. The resultant data was uploaded to the Jupyter Workbook.

Also, we will use the following FSq endpoints to collect data on:
    Venues and Categories

This data will give us information related venues most visited and the categories within which they fall.

How will the data be used

The data will be used in various phases, first to determine frequency of categories liked in the various neighbourhoods, followed an unsupervised phase wherein we cluster the neighbourhoods and their preferences.

Methodology
The methodology is elaborated on in the model in the workbook publishhed to Github.
The main steps in our methodology are described below:

1. Install relevant python packages
2. Load Cape Town postal cordinates
3. Build Foursquare API call. We obtained data in a 5km radius as the data returned was little.
4. Pass out the FoursquareJSon data.
5. Create a function to loop through all the SA cities and obtain their nearest venues.
6. Hotcode the results of category frequencies per Venue
7. Perform KMeans Cluster analysis - we provided four clusters as responses were low incertain areas.
 8. We plotted the results using folium.

Results

We derived four clusers . The clusters were representative of the various neighbourhoods in CApe Town.
There was a clear distinction between areas close to the city, the suburban areas wherein we see people who work close to the city reside.
Lastly the poorer areas further away from the city wherein there were alow response rates.

Cluster 1 - City and Suburban Areas
    Areas in and around the City appear to be where holiday makers go to hotels and a clear love for Coffee!

Cluster 2 - Working Class
    The working class areas of Cape town frequently checkin at Fast Food shops and Airports.
    We note that the these areas of Cape Town are around the airport as well.

Cluster 3 - City Outskirts
    These areas showed a low participation rate. They are far away from the city center.

Cluster 4 - Suburban Areas.
    The Suburban Areas are just outside of the city. These areas offer more affordable housing generally.
    We see the Capetonians in these areasfrequenting Grocery stores and Fast Food joints often!

Discussion

The results determined by the KMeans Clustering could be improved upon if there were a greater response rate across the western cape.

It can be seen that FourSquare is used more frequenty and wider closer to the City.

It was interesting to note that there are two areas to the North and South of the City where locals frequent.
This group of people frequent shopping malls to do grocery shopping and go to fast food places.
Typically these would consist of people in middle income brackets.

The group of CLuster 1, the City and Suburb areas, show that this group of people consist primarily of tourists as evidenced by the number of hotels likes.
There is also wide spread popularity of a love of coffee shops coming through in the results sets.

Conclusion
If a company wants to open up a coffee shop, then there is huge popularity in and around the City areas. They would easily have a large market opportunity within whihc to attract customers. However, they would need to be mindful of excess competition.

If a company were to open a new office location, then it makes sense to start an office in ares outside the city as this is where lots of locals reside so as to avoid the inner city high prices and traffic into the city center.

Links

Notebook on Watson:
https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/21134960-7411-4511-8643-6ca37a98a9dc/view?access_to-ken=d393400ca7ef24a4f7787240bd261f148601a089a17780ea5a72aea1ef8fee1b

https://github.com/czr1803/GitHub_Main/blob/master/Applied%20Data%20Science%20Capstone.ipynb

Analysing Cape Town's Neighbourhoods

Coursera Capstone Project Report

Our intent for this project was to analyse the various neighbourhoods in Cape Town using FourSquare's locational data.

Once we obtained a Foursquare developer account we were able to use their endpoints to provide locational data that we would use for further analysis.
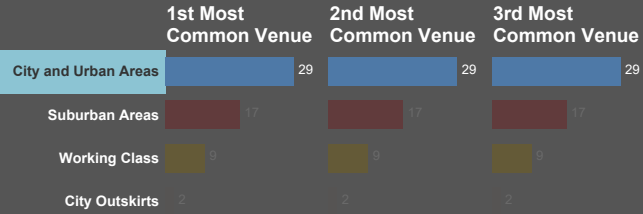
We also obtained South Africn Postal Codes from the internet and used Tableau to determine their postal coordinates. We fed this data into a pythin constructed macro so that we could obtain information for all the suburbs in Cape Town.

Having done that we determined the Categories of venues most frequented by Foursquare users in the various Cape Town suburbs. Thereafter we hotcoded these venues and their categories so that our little machine could learn easier!
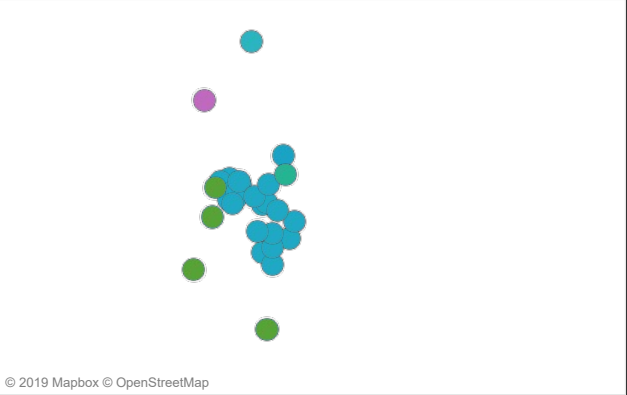
We found tha certain suburbs, especially those in East, and away from the City center had a low participation rate and thus needed to widen our sample to a 5km radius.

Once the sample size was sufficient, we performed a K-Means Cluster analysis in order to segment the various neighbourhoods. This dashboard helps us visualise the results.
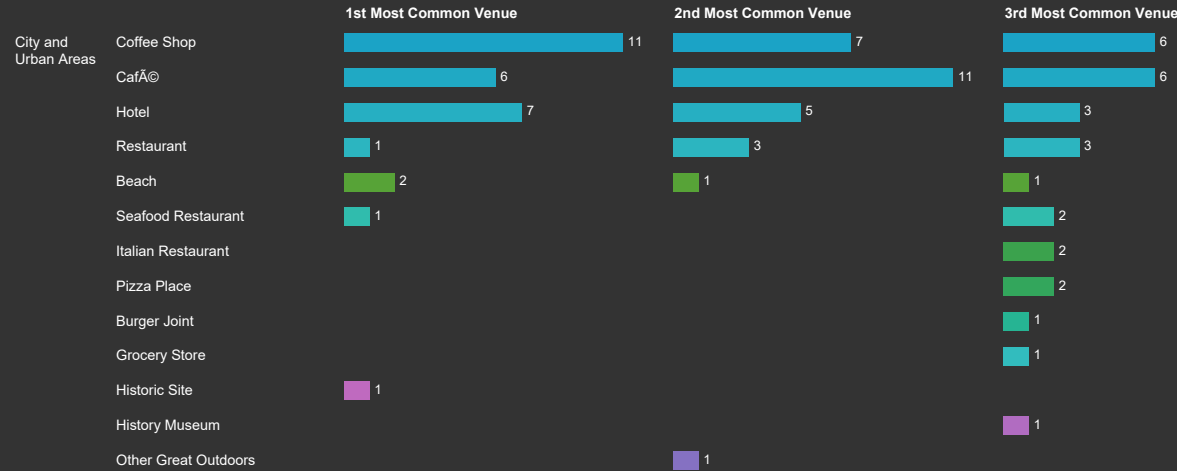
Select Most Venue to Analyse

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
City and Urban Areas	29	29	29
Suburban Areas	17	17	17
Working Class	9	9	9
City Outskirts	2	2	2

Map of Most Common Venues

© 2019 Mapbox © OpenStreetMap

● City and Urban Areas

Most Common Venues by Cluster

City and Urban Areas

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Coffee Shop	11	7	6
Café	6	11	6
Hotel	7	5	3
Restaurant	1	3	3
Beach	2	1	1
Seafood Restaurant	1		2
Italian Restaurant			2
Pizza Place			2
Burger Joint			1
Grocery Store			1
Historic Site	1		
History Museum			1
Other Great Outdoors		1	

Analysis of Clusters

City and Urban Areas

Areas in and around the City appear to be where holiday makers go to hotels and a clear love for Coffee!