

IBM Capstone Report

BATTLE OF THE SOUTH AFRICA CITIES

Contents

1. Introduction	2
1. Background.....	2
2. Business Problem	2
2. Data Sources	2
1. Web Articles	2
2. Geolocator	2
3. Foursquare.....	3
4. Census data	3
5. Tools and Modules	3
3. Methodology	4
1. Extract the location suburb data per main place and city	4
2. GEO Code the Subplaces.....	5
3. Explore the data	5
4. Results.....	8
5. Discussion	9
6. Conclusion	9

1. Introduction

1. Background

South Africa is renowned for its richness in raw minerals, friendly people and wealth of opportunities. An offshore investment group wants leverage this and sees an opportunity to invest into an upmarket restaurant in the South African Market.

2. Business Problem

The need is understanding what at the options for investing into a restaurant business in 3 of the key metros within South Africa being Cape Town, Johannesburg or Durban. These cities where chosen due to them biggest in South Africa for locals as well has being seen from overseas as top travel Destinations.

[Quality of life in Durban, Cape Town and Johannesburg](#)

[Best Holiday Destinations Durban vs Cape Town](#)

Further to this, is to understand of the type of restaurants and the location of such within the cities.

2. Data Sources

1. Web Articles

Title	Comment
Source	Internet
Extract method	Web scrape
Use	Review previous analysis to understand key business and tourism cities, disposable income etc.

2. Geolocator

Title	Comment
Source	GitHub
Extract method	Web scrape
Use	geocode the Suburbs

3. Foursquare

Title	Comment
Source	www.foursquare.com
Extract method	Web scrape. Extract venue info for suburbs in city
Use	this data will be used to understand the different restaurants available in a specific area

4. Census data

Title	Comment
Source	https://census2011.adrianfrith.com/place/798013
Extract method	Web scrape. Extract suburbs within city
Use	Extract the suburbs for under the main place

5. Tools and Modules

Title	Comment
Jupyter notebook	https://jupyter.org/
Pandas	https://pandas.pydata.org/
Numpy	https://numpy.org/
IBM cloud services	https://cloud.ibm.com/login
Geo	https://pypi.org/project/geopy/
BeautifulSoup	https://pypi.org/project/beautifulsoup4/
Folium	https://github.com/python-visualization/folium

3. Methodology

The following process was followed from extracting and scraping the data, to geocoding the extracted data. And then building maps and graph explore the data.

1. Extract the location suburb data per main place and city
2. GEO Code the Subplaces
3. Explore the data, Visualize graphically and maps
4. Use FourSquare API to extract the Restaurant Types
5. Prep the data for Clustering Model
6. Use Elbow Method to determine Optimum k
7. Cluster the Subplaces based on Restaurants

1. Extract the location suburb data per main place and city

I reached on the web regarding the top cities in South Africa for both Tourism and Business Hubs.

From the following articles, it was established that the Study will focus on Sandton in Johannesburg. Durban City and Cape Town City. They were seen as the best from a disposable income perspective and from tourism perspective.

[Quality of life in Durban, Cape Town and Johannesburg](#)

[Best Holiday Destinations Durban vs Cape Town](#)

[Most Expensive Cities in South Africa](#)

I further to this extracted the subplaces (suburbs) under these Main places areas by scraping a South Africa Census 2011 source. <https://census2011.adrianfrith.com>

Below is a sample of the table extracted for Cape Town Main Place:

	Name	Population	Area (km²)	Main Place
0	Airlie	312	0.26	Cape Town
1	Alphen	754	2.32	Cape Town
2	Bakoven	2209	1.35	Cape Town
3	Bantry Bay	820	0.38	Cape Town
4	Barbarossa	927	0.57	Cape Town

Below is a sample of the table extracted for Durban Main Place:

	Name	Population	Area (km ²)	Main Place
0	Athlone	2851	1.38	Durban
1	Athlone Park	3818	4.35	Durban
2	Avoca	7568	2.53	Durban
3	Bayhead	365	12.07	Durban
4	Beachwood Mangroves	201	2.37	Durban

Below is a sample of the table extracted for Sandton Main Place:

	Name	Population	Area (km ²)	Main Place
0	Atholl	4153	2.24	Sandton
1	Barlow Park	234	0.51	Sandton
2	Beverley	3828	2.70	Sandton
3	Bramley North	553	0.30	Sandton
4	Bramley Park	1161	0.54	Sandton

2. GEO Code the Subplaces

A function using GEOPY Plugin was then built to extract each Latitude and Longitude for each sub place per Main place table. The data extracted was then formatted to be used within the Folium Library.

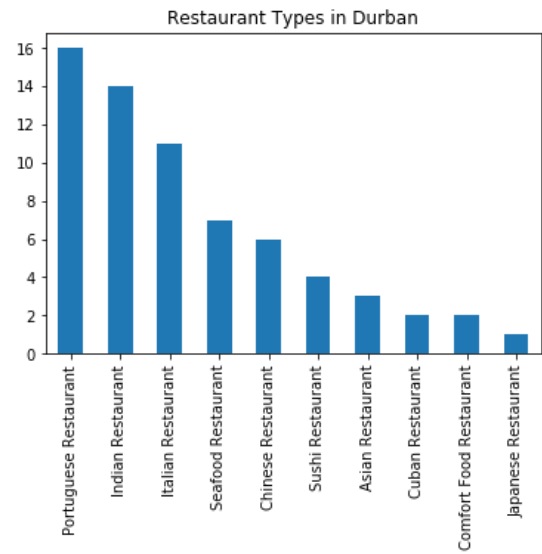
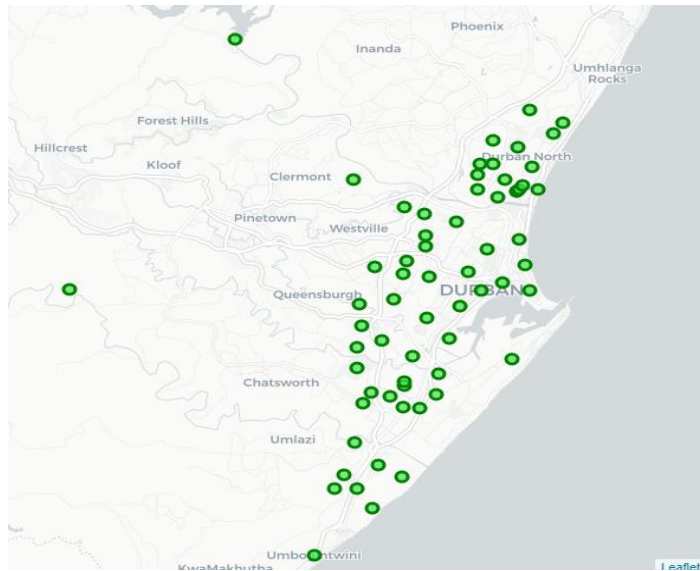
3. Explore the data

Below are the visuals per City using Folium. Each point is a subplace within the Main place.

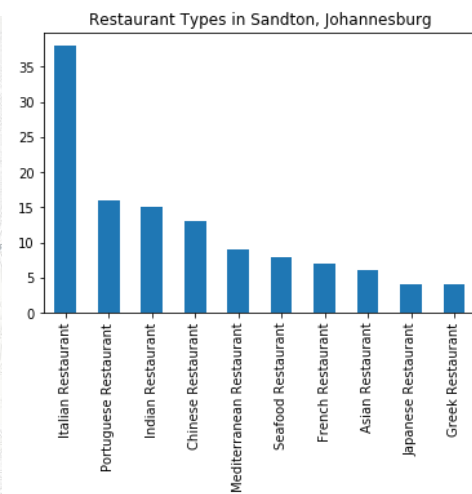
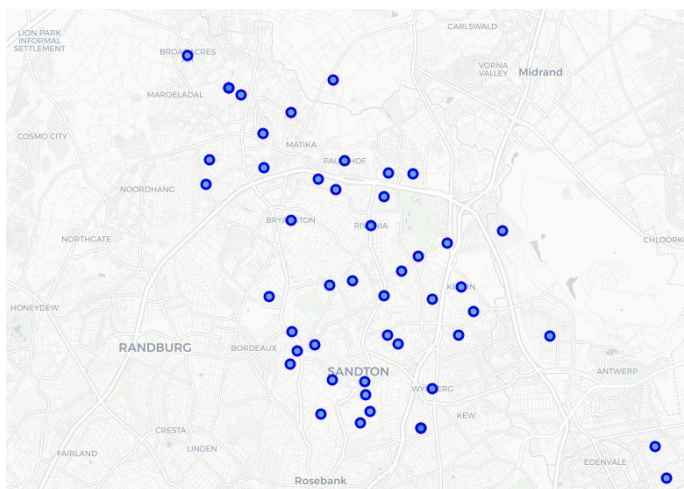
Cape Town Map:



Durban Map:

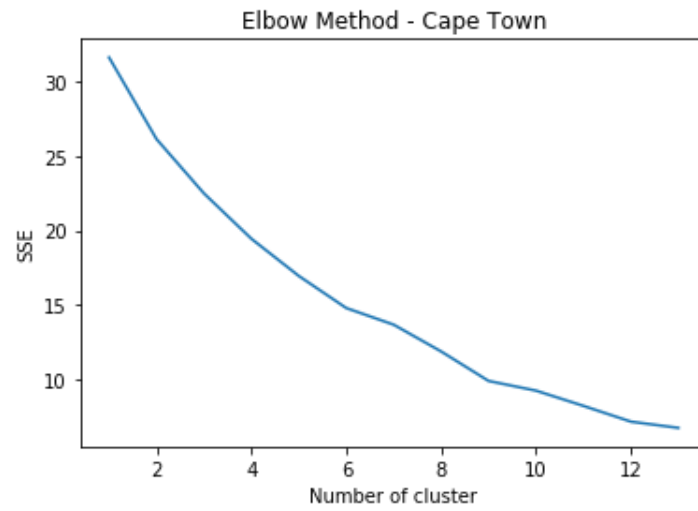


Sandton Map:



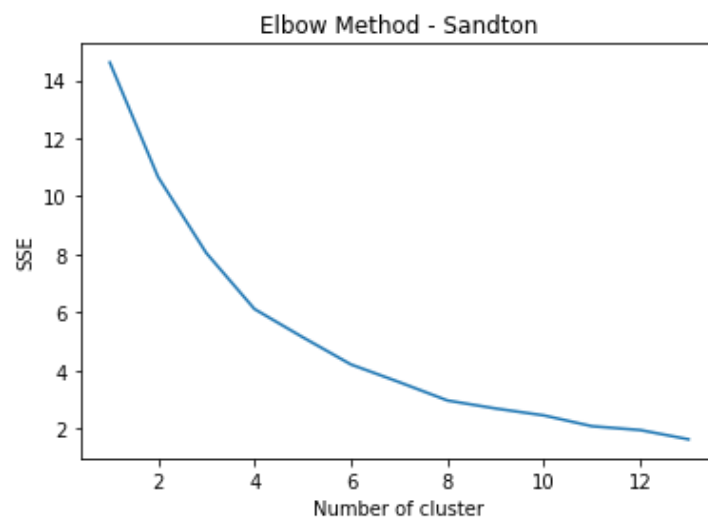
Elbow Joint results

Cape Town.



Durban

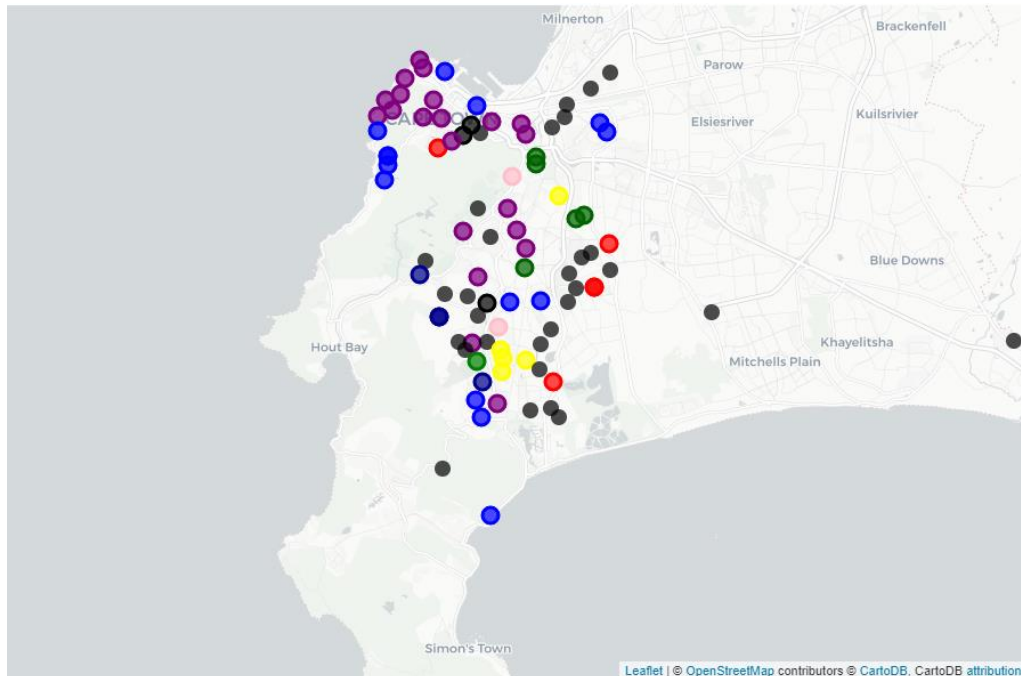
Sandton



4. Results

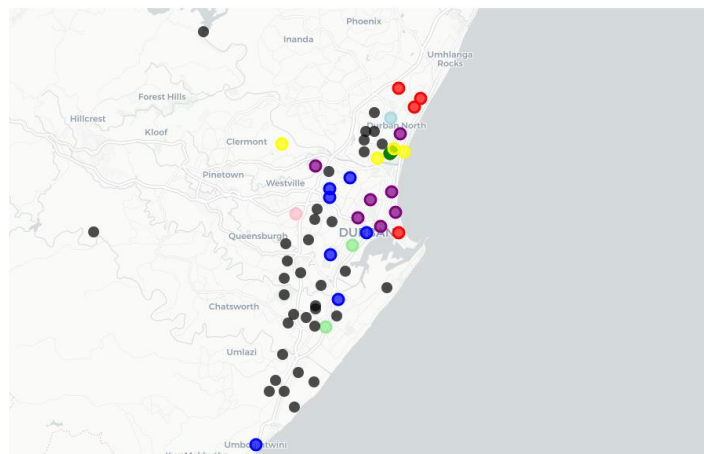
1. Cape Town

The common theme observed in the key restaurants around Cape Town was that Italian restaurants topped the list while Spicy and Asian Type food in general have a fair presence in the top 10 including Indian, Chinese and Mexican.



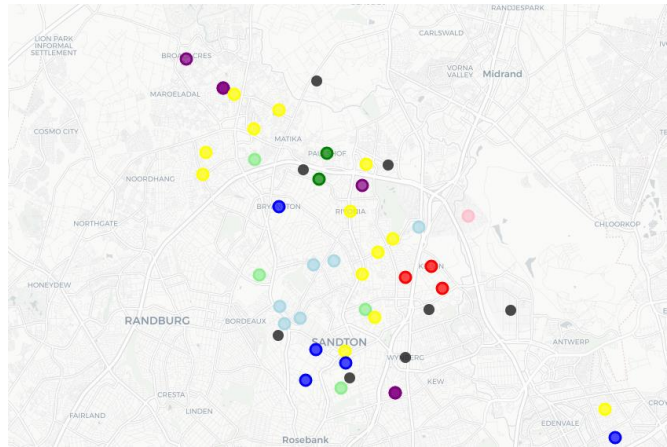
2. Durban

The common theme observed in the key restaurants around Durban was that Indian, Portuguese also very prominent restaurants, including Italian which all where in the top 5 types also following the wholesome, spicy and Asian food theme.



3. Sandton

The common theme observed in the key restaurants around Sandton was that again Italian, Indian and Chinese is very prominent. Restaurants types also following the Spicy and Asian food theme.



5. Discussion

Due to limited time and data sources, recommendation for further investigations to include household income data and population age stats to further fine tune the outputs to consider age appropriate target segment and more granular affordability.

Further recommendations to expand into Pretoria area which has a high disposable income and many universities, so to capture a younger market.

6. Conclusion

Based on the findings, the final recommendations to the investors are that the 3 key cities being Durban, Cape Town and Johannesburg are a good investment opportunity. The market considers both tourist and business hubs, so as to maximise both day time and night time occupancy rates. The recommended choice of exclusive restaurant would be of a Thai Theme. This is due to the high preference of food of Asian preference, observed by high demand of Indian, Asian, Chinese themed restaurants. The minimum occurrence of Thai restaurants in these areas presents an ideal investment opportunity.