Spent Coffee Grounds (SCG) production and analyses of coffee café market size in Cape town

Introduction & industrial problem

Sustainability and eco-friendly practises continue to be important topics in industry. This has further seen the Department of Science and Technology(DST) funding a public resource, named: Bioenergy Atlas through the South African Environmental Observation Network (SAEON). Such innovations are particularly of interest to the growing coffee community that has seen numerous “green” product developments.
The growing coffee market in South Africa has seen the country host it’s first coffee festival in Cape town at the Cape of good hope. This attracting numerous coffee houses and barristers. The coffee culture is certainly growing in popularity seeing the average coffee shop selling 300 cups of coffee per day. The major questions addressed by the project are: Is the industry, itself, growing by consumption and revenues seeing new coffee businesses opening shop. Secondly, how much waste is generated in the country, to become potential feedstock to emerging green technologies as supported by the DST, and where are the major regions producing the largest amount of SCG. These insights can inform 3 stakeholders being: the department of sustainable development, coffee enthusiasts looking to improve their waste management and entrepreneurs interested in innovative waste management solutions.

Data acquisition

The data of interest has 4 main attributes:

1. Amount of coffee consumed from ground coffee beans
2. The number of major coffee café’s and roasters
3. The city with a high number of coffee shops per area
4. Sales and market trends of the coffee industry.

The quantity of coffee beans consumed in the country can be obtained through global import and export figures given by the International coffee organization. A well sorted

data set of the trade of coffee can be obtained from the Index mundi in 60kg bags as a unit.
An elaborate market research report into the coffee market is available through Insight survey, breaking down the sales and performance of the market. The full report is only available upon purchase but key values can be obtained, followed by justifiable extrapolation in the data preparation. The Insight survey will give details of major coffee roasters and cafés in the country
The foursquare API will be used to obtain location data to determine the region where the most coffee cafés are located. Google data will provide the best central geolocation of each cluster.

Some data will be collected from different sources such as Stats-SA and consolidated into data frames for processing.

Data cleaning

Data was obtained from a range of sources to determine usability and relevance. The data collected was obtained from the city of Cape town database. This comprises of city planning data with city suburb address being the key feature required. The data was scrapped from the website using the beautiful soup method. Beautiful soup is ideal for specific data set and upon realizing the simplicity of the data the pandas read_html method proved effective for our purposes.

Geocoding was use to obtain the latitude and longitude of the suburbs. NaN values where left out as our aim is to get a wholistic center for the coffee shops and minor suburbs within do not have a significant impact on the result. The data was further cleaned to exclude areas on the outskirts of the city, centralizing the focus on the dataset.

Selecting ideal features and data

Only the suburb names and geographic coordinates were relevant to the research and as such, the rest of the features were left out. Coffee trade data from the world

coffee organization was clean and absolute in reflecting the amounts of 60kg bags shipped into the country with everything consumed.

Exploratory Data Analysis

The data frame scrapped from the website contained no much structure with clear value

	Suburb	Zip Code	Postal Code
0	NaN	8000.0	8001.0
1	Athlone	7760.0	7764.0
2	Athlone Industria 1	NaN	7764.0
3	Athlone Industria 2	NaN	7764.0
4	Bakoven	NaN	8005.0

Table 1. Data set for cleaning

This was converted into addresses using the postal codes and regenerated into absolute addresses. The location of our suburbs is pivotal to the purpose of the investigation. We looking to understand the layout of major suburbs. To view this data, it was subject to folium geo plotting and the following is the visualized map.

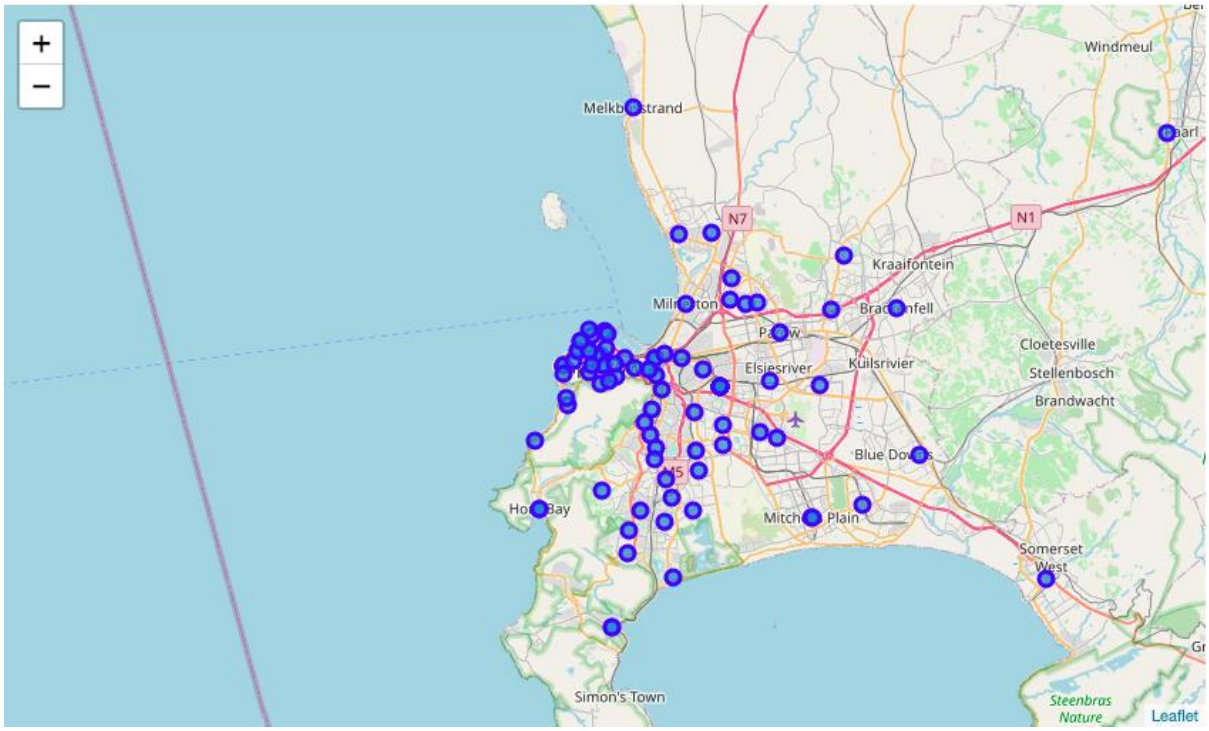


Figure 1. Geo location of suburbs in Cape town

The main city centre has the major suburbs as expected. Th visual clearly suggests solutions to our geographical problem. We Will further asses these locations through popularity of the coffee shops. The locations and the set up of the city make it ideal for our operational needs.

The Four Square API was used to obtain data on the most popular coffee shops in the city specifically. We know that a significant amount of coffee is consumed in the city centre and the big game players are all concentrated there. A closer look of the coffee shops is ideal in understanding the scatter matrix of the shops.

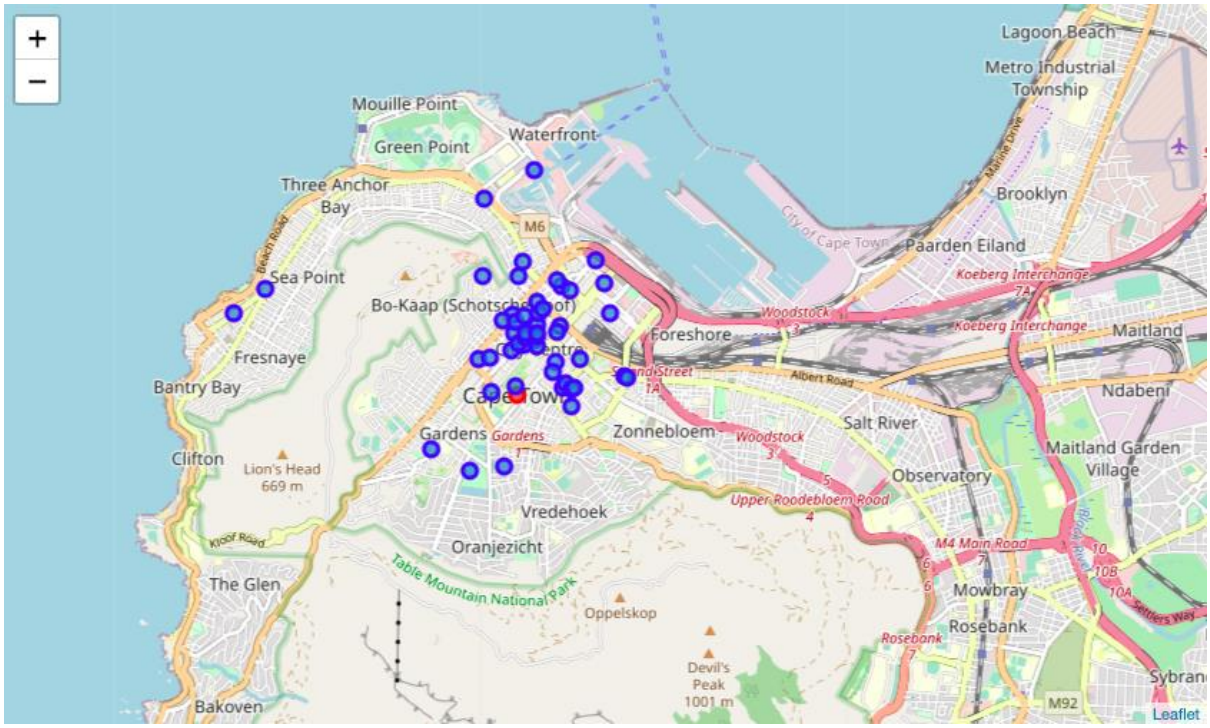


Figure 2 Major coffee shops in the City of Cape town

We have a clear indication of how our coffee shops are laid out in the city centre. What is now interesting to find out are the ideal clusters from which we will collect our Spent coffee grounds. No more than 4 points of collections are feasible.

The suburbs were further analysed to find our best points for the city. Kmeans is an ideal method for us to determine optimum number of collection points with the city centre being one of them.

Kmeans clustering

The ideal locations will have to be representative of the suburb layout and these coordinates were used in our the advanced python statistical tool, being kmeans.

After assigning cluster labels to each suburb coordinate, the following was obtained.

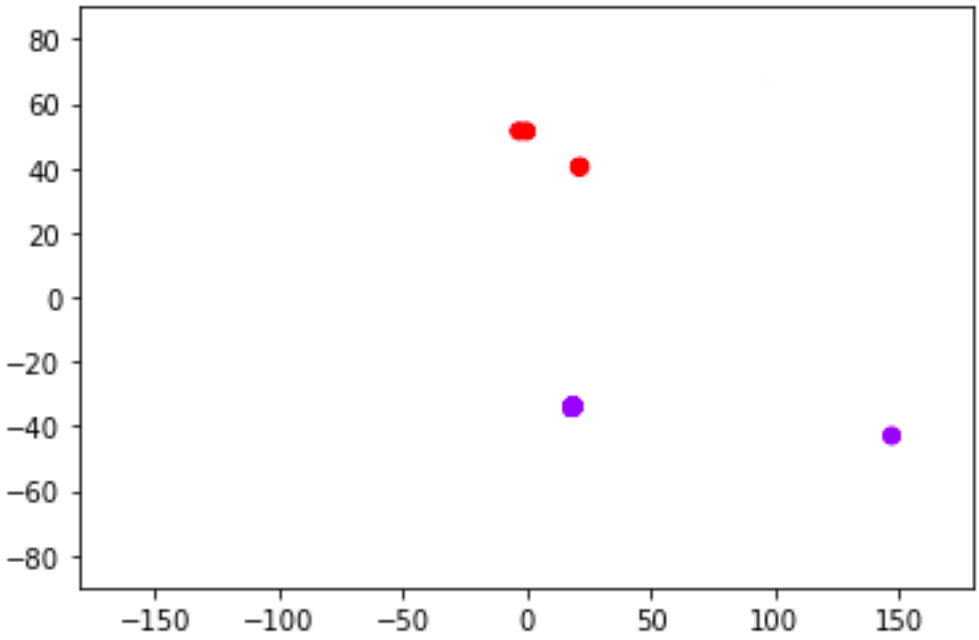
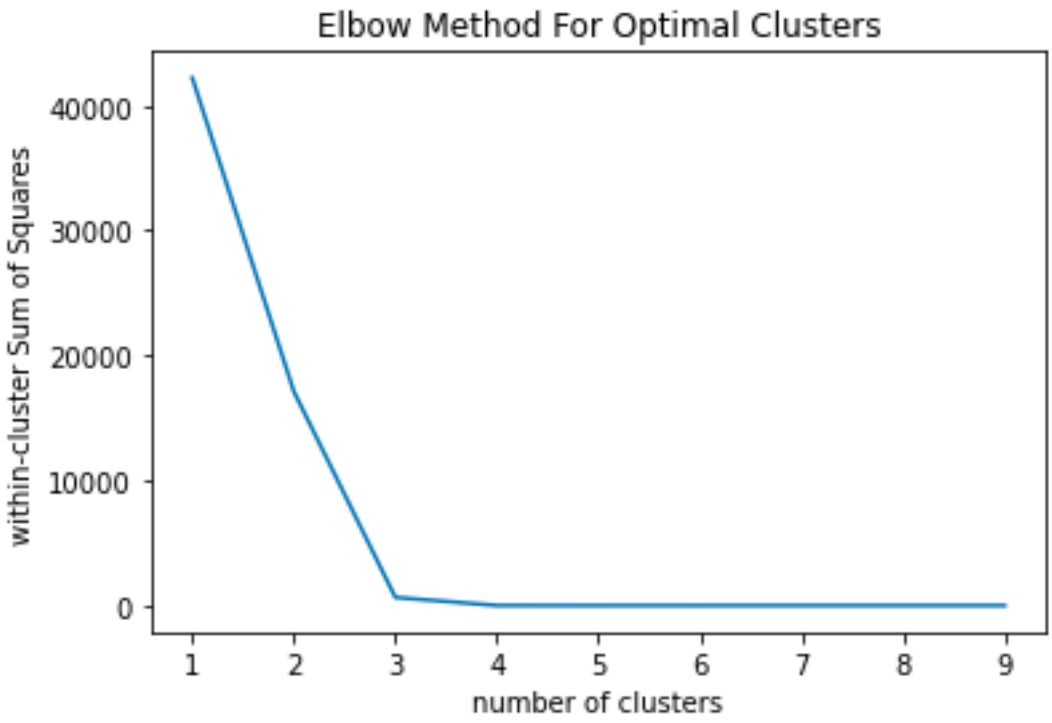


Figure 3 Cluster plot on Y. X axis

There is a clear outlier in the data set and we are looking to find centralized locations despite outliers. We employed the elbow method to determine the best number of clusters out of 10, to see what the optimum solution is.



Three clusters returned the lowest Within-Cluster sum of Squares.

Conclusion

We are required to obtain the best locations and the most operationally efficient positions for collection of Spent Ground Coffee and through the Kmeans analysis, we have identified 3 main collection points to be the most ideal cluster centre for the suburbs in cape town. The coffee shop output is clearly outlined by the coffee organization and specific values for Cape town are dynamic. The seasonality of the data calls for a further market analysis with a set season or time frame. Also time frame is a large factor, we do know that the city centre produces most of our waste of interest. The South African coffee market has increased at a compound annual growth rate of 13.3% from 556 000 60kg bags in 2014/15 to 809 000 60kg bags in 2017/18. This suggests it worthy to further the investigation into the waste produce as a feed source.

References

1. https://www.indexmundi.com/
2. Consumer behavioural intentions: https://repository.up.ac.za/bitstream/handle/2263/52904/VanDerMerwe_Behavioural_2016.pdf?sequence=1&isAllowed=y
3. 4. SOUTH AFRICAN COFFEE IMPORTERS AND ROASTERS (16 Pages)
   4.1 Bean There: Overview
   4.2 CATURRA Africa: Overview
   4.3 Coco Safar: Overview
   4.4 Deluxe Coffeeworks: Overview
   4.5 Importers Coffee Merchants: Overview
   4.6 Mastertons: Overview
   4.7 Origin: Overview
   4.8 Quaffee: Overview
   4.9 Rosetta Roastery: Overview
   4.10 Terbodore Coffee Roasters: Overview
   4.11 Tribe Coffee: Overview
   4.12 Truth Coffee Roasting: Overview
   4.13 South African Coffee Importers And Roasters: Other: https://insightsurvey.co.za/wp-content/uploads/2020/05/SA-Coffee-Industry-Landscape-Brochure-2020-1.pdf:
4. http://www.ico.org/trade_statistics.asp : trade statistics from International coffee trade
5. https://agribook.co.za/forestry-and-industrial-crops/coffee/