IBM Capstone Project in Applied Data Science: The Battle of the Neighborhoods
by
Caspah Lidiema
May 5, 2021

1. Introduction

1.1 Background

Durban is the largest City in South Africa after Johannesburg and Cape Town. It is located in the is located on the east coast of South Africa in KwaZulu-Natal province and touching the Indian ocean. Durban was one of the host cities of the 2010 FIFA World Cup. Therefore, hosting a FIFA world cup is not only prestigious in the world of sports but economically it promotes tourism, investment opportunities and culture change through tourism. This could include food culture and behavioral culture. Domareski et al (2019) opine that South Africa's GDP continued to grow, reaching an increase of 8,34% five years after the FIFA world cup mainly contributed by tourism. According to FIFA (2010), even though the tournament was held in winter, tourism in South Africa increased tremendously with most visited areas being Gauteng, KwaZulu-Natal, and Western Cape.

1.2 Problem Statement

We have noticed that FIFA world cup improves a nation's economy through tourism and business. Given that KwaZulu-Natal is one of the most visited provinces in South Africa, there is need for more business centers that caters for all cadres of tourists from all over the world. The author, therefore, wishes to establish a business within the Durban Suburbs

either in the hotel and restaurant sector. This project aims to establish which Suburb is the best and which type of restaurant to open in Durban using the following factors, density of restaurants and the type of restaurant. However, while some factors like distance from other suburbs and population density should be considered they are not within the scope of this study. This study will concentrate of restaurant concentration and type of restaurant currently available in Durban suburbs while keeping in mind the existence of other business in each surbub.

1.3 Interest

Any entrepreneur would be interested in getting to understand the concentration of the kind of business they wish to undertake or utilize the gaps in the market to come up with new products or improve on the existing ones. In this study the author wishes to understand the restaurant concentration in Durban suburbs and choose a suburb with the least concentration or and capitalize on it.

2. Data acquisition and cleaning

2.1 Data Sources

Data will mainly be sourced from Goggle search using the beautifilsoup python library and my making use of Wikipedia extract or scrap data from

https://en.wikipedia.org/wiki/Category:Suburbs_of_Durban. This data on suburbs will then be linked with respective coordinates for latitudes and longitudes. Lastly data for venues within Durban will be sourced from Foursquare API

and connect to the previous data to get the full

data set to use in our analysis.

2.2   Data Cleaning

Data downloaded or scraped from multiple sources will combined in into table first. Thereafter, checks on missing data points will be done and dropped. Once the data is clean the methodology process will begin first starting with Exploratory data analysis, and then clustering will be done. This will be done with a view of finding the best suburb and the best type of restaurant business to open.

3.   Methodology

This study will employ K-Mean clustering algorithm to establish the umber of clusters in Durban Suburbs to enable make informed decision on which in which cluster/suburb is appropriate to open which kind of restaurant.

3.1   Map of Durban Clusters
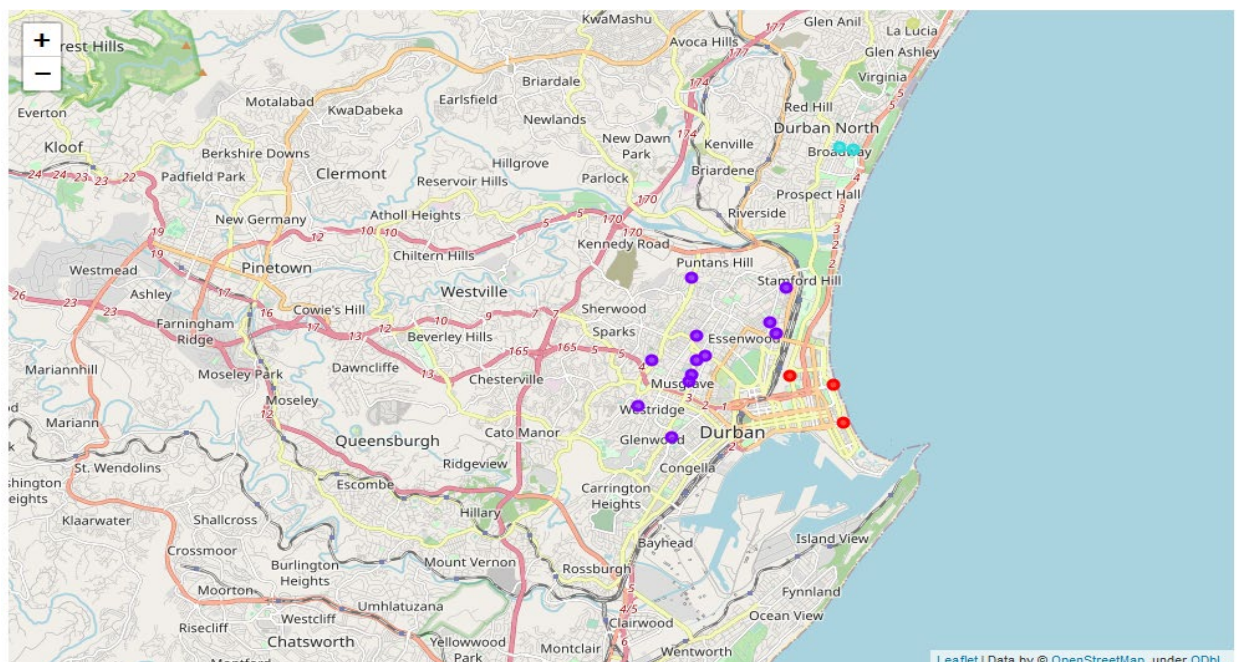
3.2   Results on Clustering

3.2.1   Cluster 1

Cluster 1 is the most concentrated cluster. This cluster contains businesses in Central, and Western part of Durban. The most common venues in this cluster are Coffee shops, Cafes, Grocery stores, Indian restaurant, and Portuguese Restaurant. Pizza Places and Fast-Food restaurants are also Common in this cluster.

	Surbubs	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Bellaire, Durban	1	Coffee Shop	Grocery Store	Café	Portuguese Restaurant	Steakhouse
13	Durban North	1	Café	Grocery Store	Coffee Shop	Indian Restaurant	Fast Food Restaurant
16	Gillitts	1	Café	Grocery Store	Coffee Shop	Fast Food Restaurant	Portuguese Restaurant
17	Glenwood, KwaZulu-Natal	1	Café	Coffee Shop	Seafood Restaurant	Indian Restaurant	Shopping Mall
23	La Lucia	1	Coffee Shop	Shopping Mall	Burger Joint	Indian Restaurant	Mexican Restaurant
28	Montclair, Durban	1	Coffee Shop	Indian Restaurant	Café	Restaurant	Burger Joint
30	Mount Vernon, Durban	1	Café	Indian Restaurant	Coffee Shop	Grocery Store	Restaurant
33	Ottawa, KwaZulu-Natal	1	Coffee Shop	Steakhouse	Grocery Store	Italian Restaurant	Indian Restaurant
39	Reservoir Hills	1	Indian Restaurant	Coffee Shop	Café	Burger Joint	Restaurant
44	South Beach, Durban	1	Hotel	Coffee Shop	Café	Restaurant	Pizza Place
45	Stamford Hill, Durban	1	Coffee Shop	Portuguese Restaurant	Shopping Mall	Steakhouse	Restaurant
54	Wiggins, Durban	1	Coffee Shop	Café	Hotel	Pizza Place	Portuguese Restaurant

3.2.2   Cluster 2

With only two suburbs this cluster has a mixture of Cafes, Burger Joints and Italian Restaurant as popular venues. This Cluster is located in the Western region of Durban. Other common business venues in this area are, Indian Restaurant and Convenience Stores.

	Surbubs	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	Botha's Hill	2	Café	Burger Joint	Italian Restaurant	Portuguese Restaurant	Convenience Store
12	Cowies Hill	2	Burger Joint	Café	Italian Restaurant	Convenience Store	Indian Restaurant

3.2.3 Cluster 3

Cluster 3 is a one-suburb cluster located in KwaMashu Township in North Durban. The cluster has a mixture of cafes, Italian restaurants, and liquor stores as common business venues.

3.2.4 Cluster 4

Cluster 4 is concentrated with Hotels and general Restaurant. This cluster located on the is the most Beach side of the Durban mainly in Umlazi, Sherwood and Musgrave.

4. Result and Discussion

After carrying out clustering using K-Means. We find 4 main clusters. The clusters are mainly located on the coastal side touching the beach, the central Durban Area, Western Durban and North Durban regions. From our analysis, cluster 1 is mainly located in central Durban, cluster 2 in the Western and Cluster 3 is located in the KwaMashu Township in North Durban. Cluster 4 is combines three major suburbs of Musgrave, Sherwood and Umlazi.

Cluster 1 is the most concentrated and this expected considering that central Durban is the most populated and hence a buzz if activities are expected. As a result, the K-means clustering algorithm used in this analysis picks most business concentration in Cluster 1 which include café Shops, Grocery stores, Coffee Shops, Fast Food restaurants and other specialized restaurants.

My analysis did not consider the population density, since Tourists mainly do not look at the population density but the area of attraction. The author's main objective is to open a restaurant in any suburb that is less concentrated worth providing value for money and services to Tourists. Therefore, the determining characteristic in this analysis is mainly the concentration of restaurants

and Hotels in the clusters identified by Durban Suburbs. For K-Means Clustering analysis, most of the suburbs in Durban have general restaurants, coffee shops and other specialized restaurants (Indian, Italian and Portuguese).

5. Conclusion

Clusters 2,3 and 4 would be ideal to open any restaurant. However, cluster 4 would not be ideal for a Hotel since Hotel business is the dominant and most popular ins all suburbs under cluster 4. Therefore, the following would be appropriate proposals: an Indian restaurant would be ideal in Cluster 4; Portuguese and Indian restaurant in Cluster 3 and Fast-food Restaurant would be ideal in Cluster 2 and cluster 3.

6. Future Research and Directions

This study was limited to clustering of businesses in Durban suburbs using K-Means Clustering. This study dd not take into consideration population density of Durban Suburbs. Therefore, future research and analysis can consider the element of population density and establish if the results would be different or similar. Equally the flow of tourists to each suburb and the income per capital in Durban Suburbs can be considered.

References

1. Domareski Ruiz, T. C., Chim Miki, A. F., & Dos Anjos, F. A. (2019). Competitiveness, economic legacy and tourism impacts: World Cup.

2. FIFA (2010). Study reveals tourism impact in South Africa. *Extracted from https://www.fifa.com/worldcup/news/study-reveals-tourism-impact-south-africa-1347377*