Finding the optimal coastal city for investment between Cape Town and Durban (South Africa)

By: Sihle Mkaza

Project For: Applied Data Science Capstone course (IBM Data Science Professional Certificate)

Table of Contents

1. Introduction

1.1 Background

Cape Town and Durban are two of the most popular cities in South Africa and both are coastal cities[1] [2]. Both cities are business centers, with Cape Town having many tourist attractions such as Table Mountain and Robben Island and with Durban having tourist attractions such as the Durban Botanic Gardens. Many international companies choose Cape Town as an initial entry into the African market and Durban is the first choice for holiday destinations. Setting up a viable business in one of the cities would be a good investment for any entrepreneur or investor.

1.2 Problem

Seeing that both Cape Town and Durban share many similar features such as being costal cities and business hubs, it is hard for investors and/or entrepreneurs to choose a city to start a business in when having to pick between the two. The aim of this report and project is to provide an analysis of both cities and determine which of the cities would be the most optimal one to invest in. This will be accomplished by providing information on what types of businesses (or business categories) are the most popular in each of the cities and suggesting on what business category to invest in and in which city to invest in.

The information produced from this report would be most useful to investors and/or entrepreneurs that have identified Cape Town and Durban as potential business locations and are not sure what kind of business to start yet and in which city.

2. Data

2.1 Data Acquisition

The names of the neighbourhoods in each city is needed for an analyses of businesses categories in each city. No single website with names of neighbourhoods for each city and the associated postal codes exist, so two csv files were manually created with data obtained from multiple websites. The first csv file contains Cape Town neighbourhoods and their associated postal codes [3] and the second csv file contains Durban neighbourhoods and their associated postal codes [4].

The Foursquare API will be used to obtain popular venues (businesses/places) for each of the neighbourhoods in each city using the 'venues/explore' foursquare endpoint [5].

The latitude and longitude of each neighbourhood is needed in order to get popular venues using the foursquare 'venues/explore' endpoint. The GeoPy Nominatim library [6] was used to obtain the latitude and longitude of each of the neighbourhoods in Cape Town and Durban. The retrieved latitude and longitude values were merged with the dataframe containing neighbourhood names.

2.2 Data Cleansing

On some occasions the geopy nominatim module fails to get the latitude and longitude values of a neighbourhood and results in the neighbourhood dataframe containing NaN values where the latitude and longitude values are meant to be. All neighbourhoods with NaN values in the latitude and longitude column were discarded as it would not be possible to get popular venues for them using the foursquare API.

2.3 Feature Selection

Even though each of the neighbourhoods have postal codes, some neighbourhoods (two or more) share the same postal code, for example Esplanade, Essenwood and Glenmore in Durban share the postal code 4001. Postal code is there not very representative of each neighbourhood as two

neighbourhoods that share the same postal code can very in the trends of popular businesses in each neighbourhood [7]. Postal code was therefore discarded from the data.

	Neighbourhood	City	Latitude	Longitude
0	Bakoven	Cape Town	-33.960000	18.382778
1	Bantry Bay	Cape Town	-33.928151	18.378970
2	Camps Bay	Cape Town	-33.954774	18.381852
3	Clifton	Cape Town	-33.935285	18.379070
4	Fresnaye	Cape Town	-33.925194	18.387743

Figure 1: first five rows of Cape Town data after feature selection

	Neighbourhood	City	Latitude	Longitude
0	Addington	Durban	-29.868127	31.043306
1	Austerville	Durban	-29.945278	30.980833
2	Avondale Road	Durban	-29.844753	31.009084
3	Bayhead	Durban	-29.891389	30.991389
4	Beach	Durban	-29.861825	31.009909

Figure 2: first 5 rows of Durban data after feature selection

Methodology

Results

Discussion

Conclusion

References

[1] https://www.southafricavisa.com/cities-to-visit-in-south-africa/

[2] https://www.touropia.com/best-cities-to-visit-in-south-africa/

[3] https://github.com/sihlemkaza/Coursera_Capstone/blob/main/CAPETOWN_CITY.csv

[4] https://github.com/sihlemkaza/Coursera_Capstone/blob/main/DURBAN_CITY.csv

[5] https://developer.foursquare.com/docs/api-reference/venues/explore/

[6] https://geopy.readthedocs.io/en/stable/#nominatim

[7] https://towardsdatascience.com/stop-using-zip-codes-for-geospatial-analysis-ceacb6e80c38