

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, author 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.

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 beautifulsoup 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

<https://api.foursquare.com/v2/venues/explore> 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 be combined into a 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.

References

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

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