Report: Battle of the Neighborhoods

Battle of the Cape Town Suburbs

Introduction

Moving to a new city is daunting and it helps to know as much information as possible about the city and its suburbs before settling on a place to live. Information such as crime rates, proximity to hospitals, school ratings, and nearby amenities are important factors to consider before choosing a place to live. This problem calls for an algorithm which will help the user explore suburbs in an area with regard to such factors, in order to make an easy comparative analysis. This project aims to develop an application which will perform the task of suburb comparison and recommend the best place to live under user specifications. It will assist in rapid decision making, saving the user time and money.

South Africa is a notoriously high-crime country, so immigrants seek areas that have low crime rates. The project compares two Cape Town suburbs in terms of their crime rate and proximity to nearest hospital, and recommend the user the most suitable suburb to live in. Furthermore, this project aims to use K-means clustering to cluster the suburbs based popular venues such as restaurants, parks, coffee shops, gyms, etc., helping to provide a better understanding of the similarities and dissimilarities between the two chosen suburbs and to retrieve more insights so as to easily choose the winning suburb.

Data and Methods

Primary data:

- Cape Town list of suburbs scraped from wikipedia webpage.
- Cape Town suburbs crime rate data downloaded from the South African Police Service webpage.
- Distance from suburb to nearest hospital sourced from Google maps.

Foursquare:

The Foursquare Places API provides location based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, and geo-tagging. This project uses Foursquare API to retrieve data on Cape Town suburbs in terms of their most popular venues. Requests will be sent to Foursquare API using the names of Cape Town suburbs to return most popular venues within a 500m radius.

K-means clustering:

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms, which clusters data based on their similarities. K-means clustering will be used to cluster suburbs based on their most popular venues.

Folium:

Folium is a powerful data visualization library in Python that was built primarily to help people visualize geospatial data. Folium will be used to create interactive leaflet maps of Cape Town and its suburbs.

These aforementioned processes require the following python packages:

- Pandas: for analysing datasets
- Numpy: for handling vector data
- JSON: for handling json files used in folium
- Geopy Nominatim: for getting location data
- matplotlib: for plotting in Python
- sklearn: for kmeans clustering
- folium: to generate maps

Results

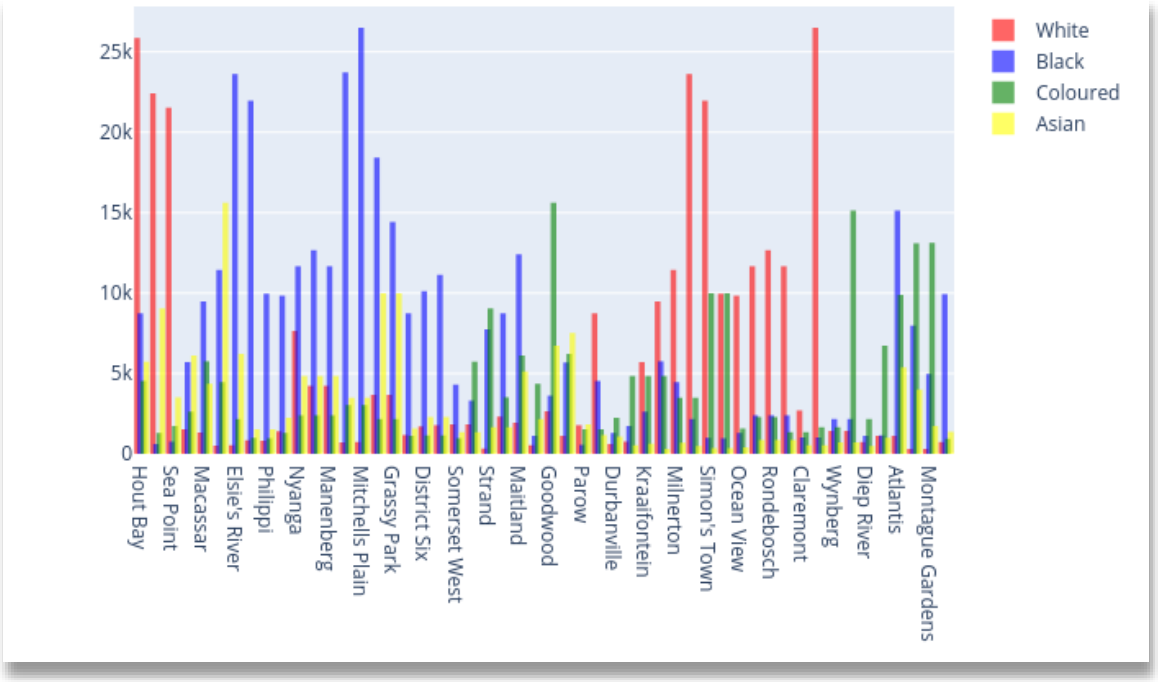

Figure 1: Bar plot of population demographics of Cape Town suburbs

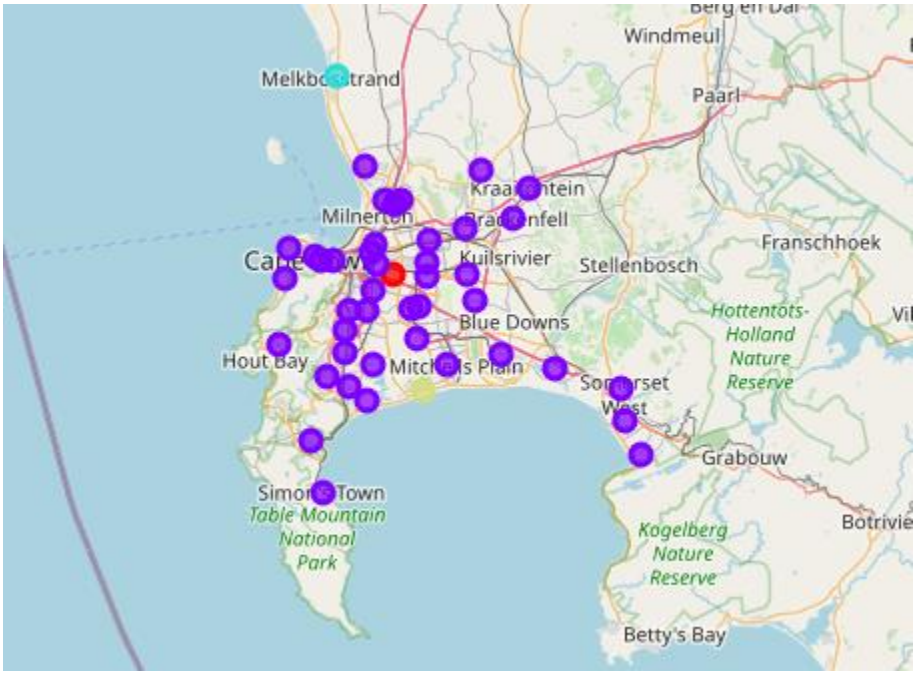

Figure 2: Map of Cape Town suburbs clustered using K-means clustering, generated with folium

Comparison of two user-selected suburbs

The user wants to compare Sea Point (left) and Muizenberg (right), two very popular seaside suburbs of Cape Town.

- Crime comparison

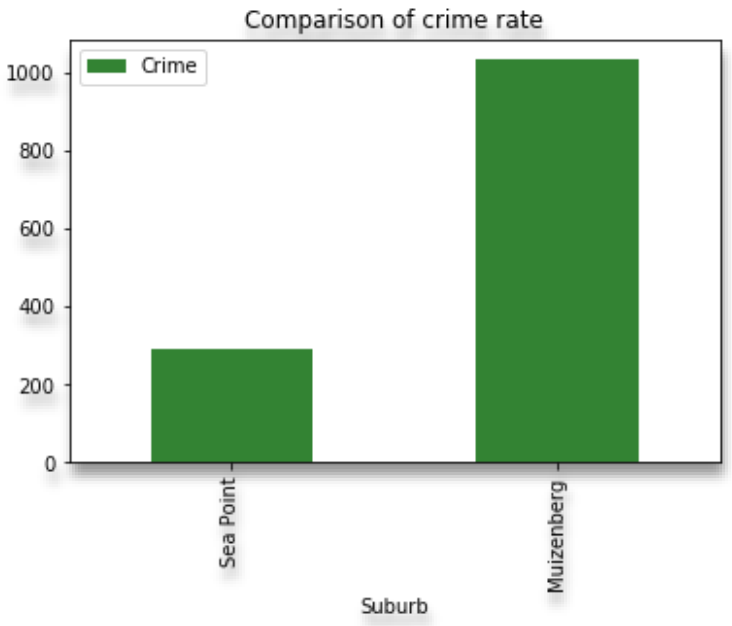

Figure 3: Bar plot comparing crime rate in Sea Point and Muizenberg

We can see that Muizenberg has a much higher crime rate, almost three times higher, than Sea Point.

- Demographic comparison

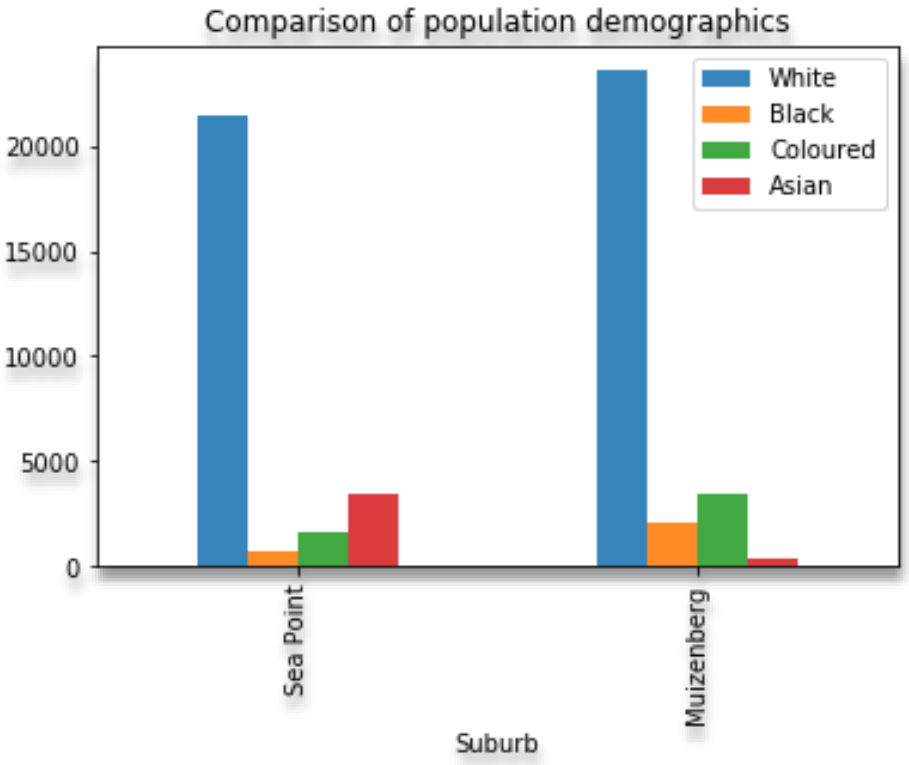

Figure 3: Bar plot comparing population demographics in Sea Point and Muizenberg

The user is immigrating from Japan and would therefore prefer a larger Asian population. Sea Point has a substantially higher Asian population, and, for that reason, Sea Point is more appealing to the user.

- Hospital comparison

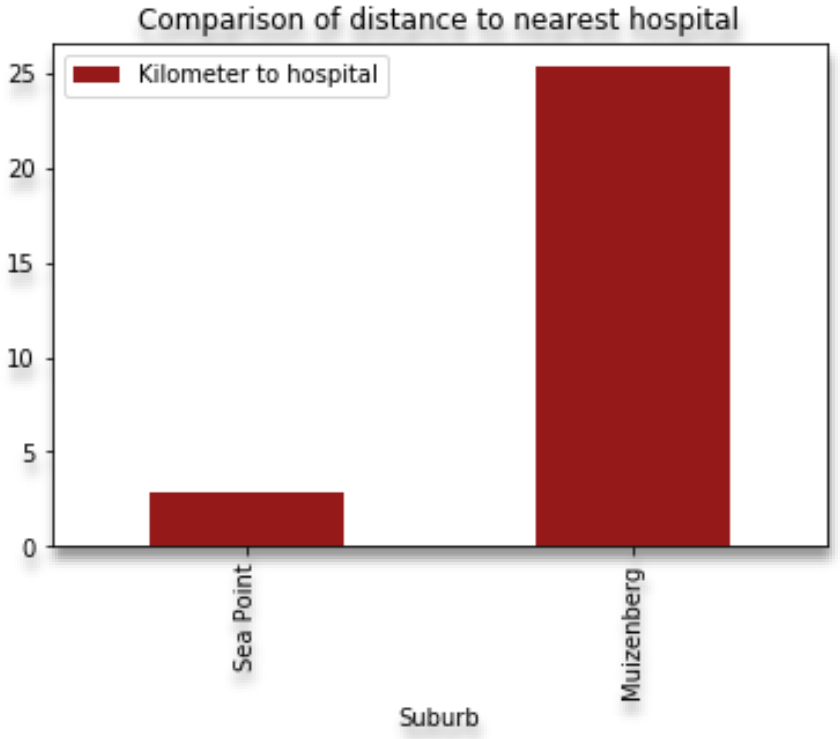

Figure 4: Bar plot comparing distance to nearest hospital from Sea Point and Muizenberg

The user suffers from renal issues so it is important the suburb is in close proximity to a hospital that offers dialysis.

Sea Point is much closer to a major hospital and is therefore a far safer place to live in case of emergencies.

- Comparison of popular venues

Table 1: Sea Point and Muizenberg's top 10 most popular venues

Suburb	Sea Point	Muizenberg
1st Most Common Venue	Hotel	Coffee Shop
2nd Most Common Venue	Pizza Place	Event Space
3rd Most Common Venue	Burger Joint	Flea Market
4th Most Common Venue	Coffee Shop	Fishing Store
5th Most Common Venue	Sushi Restaurant	Fish & Chips Shop
6th Most Common Venue	Italian Restaurant	Fast Food Restaurant
7th Most Common Venue	Restaurant	Factory
8th Most Common Venue	Grocery Store	Fabric Shop
9th Most Common Venue	Shopping Mall	Ethiopian Restaurant
10th Most Common Venue	Indian Restaurant	Convenience Store

Table 1 shows us that Sea Point is a more upmarket suburb by the likes of what it offers: hotels, restaurants and shopping malls, compared to Muizenberg's flea markets, fast food and fishing store and convenient stores, which are less upmarket.

Conclusion

The user in this example is a Japanese immigrant with kidney problems, looking for a secure and upmarket place in Cape Town to live. Sea Point has a substantially lower crime rate and a more Asian population, making him feel more comfortable. Muizenberg is too far from a hospital in case of emergency or need for dialysis, whereas Sea Point has a major hospital just 3km away. Sea Point is more upmarket and offers nice restaurants and malls, both amenities which the user enjoys.

Therefore, Sea Point is the best suburb for this user.