BATTLE OF
THE NEIGHBORHOODS

South Africa

Table of Contents
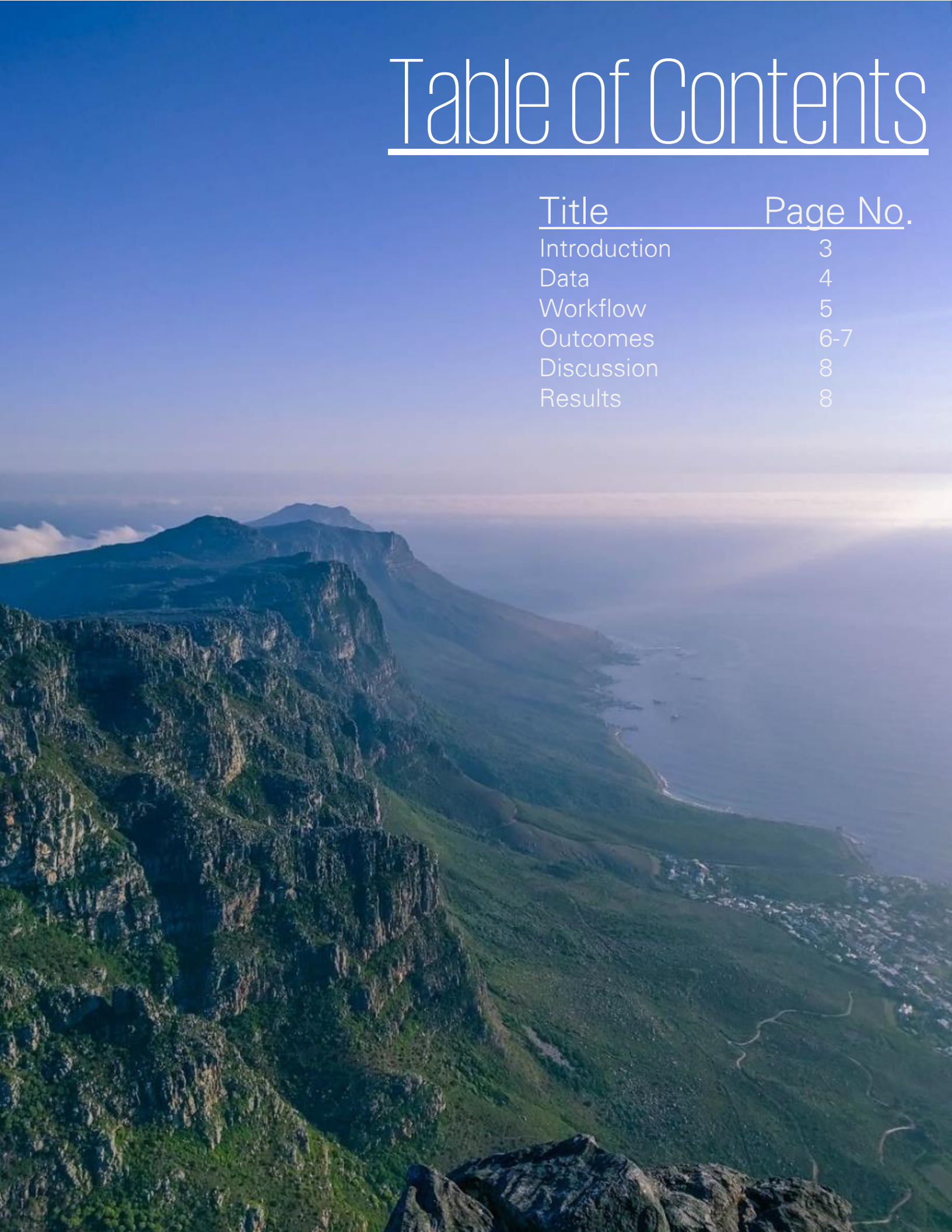
Introduction

When thinking about relocating to a new city or country for work purposes or to start a new life, people tend to research areas before moving. This research includes population rate, average house price, school ratings, crime rates, weather conditions, recreational facilities etc.

Based on the above, a search engine algorithm would be an efficient tool to use that will allow users to enter cities and get the neighborhood name that best suits their lifestyle or living conditions. This could be done using an algorithm that will perform an extensive analysis on the similarities and dissimilarities between neighborhoods in the two cities of the user's search criteria, and determine which neighborhoods best suits their lifestyle.

For this project, I will be developing a recommendation system using the following cities in South Africa as my search criteria:

VS

Data

The data used for this project will be acquired from http://www.sapostalcodes.info/. The datasets consists of the postal codes and suburb names of each city. In addition, the Foursquare API search feature will be used to collect neighborhood venue information as well as the longitude and latitude details of each suburb.

Details about local venues and locality will provide insight into the qualities of a neighborhood. In addition to Foursquare, various python packages will be used to create maps and machine learning models to gather further insights and provide efficient recommendations and results. This packages include:

- ✓ Pandas - Library for Data Analysis
- ✓ NumPy – Library to handle data in a vectorized manner
- ✓ JSON – Library to handle JSON files
- ✓ Geopy – To retrieve Location Data
- ✓ Requests – Library to handle http requests
- ✓ Matplotlib – Python Plotting Module
- ✓ Sklearn – Python machine learning Library
- ✓ Folium – Map rendering Library

Workflow

- ✓ HTTP requests would be made to this Foursquare API server using postal codes of Cape Town Suburbs and Johannesburg Suburbs to pull out the latitude and longitude which will be used for creation of the map as well data analysis.

- ✓ Foursquare API search feature would be enabled to collect the nearby places of the suburbs. Due to http request limitations, the number of places per suburb parameter would be set to 100 and the radius parameter would be set to 700.

- ✓ Folium- Python visualization library would be used to visualize the suburbs cluster distribution of Johannesburg and Cape Town over an interactive leaflet map.

- ✓ Extensive comparative analysis of two suburbs world be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

- ✓ Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighborhoods. These clusters from each of those two chosen suburbs would be analyzed individually collectively and comparatively to derive the conclusions.

Outcomes – Cape Town, South Africa

The K-means method was used to cluster the suburbs of Cape Town into 5 clusters.
The details of the clusters are as follows:

- Cluster 1
    - 17 Suburbs
    - Common Venus include Coffee Shops, Restuarants, Nightclubs and Hotels
- Cluster 2
    - 22 Suburbs
    - Common Venues include Bookstores, Coffee Shops, Convenience Stores and Restaurants
- Cluster 3
    - 29 Suburbs
    - Common Venues include Hotels, Pubs, Accessory Stores and Coffee Shops
- Cluster 4
    - 10 Suburbs
    - Common Venues include Fast Food Restaurants, Winery, Beaches and Cafes
- Cluster 5
    - 2 Suburbs
    - No common venues found
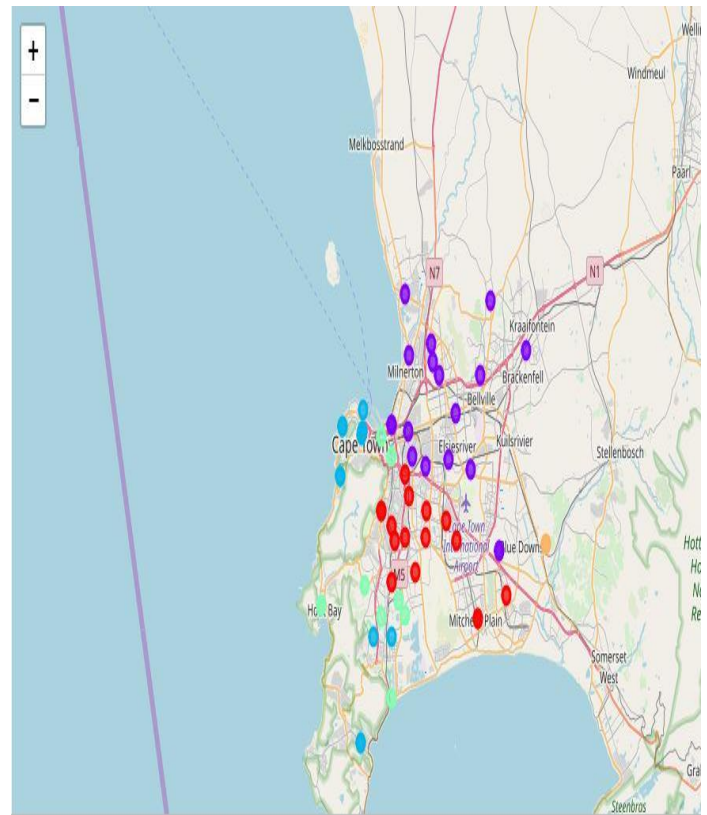
Outcomes – Johannesburg, South Africa

The K-means method was used to cluster the suburbs of Cape Town into 5 clusters. The details of the clusters are as follows:

- Cluster 1
  - 30 Suburbs
  - Common Venues include Hotels, Spas, Parks and Coffee Shops
- Cluster 2
  - 24 Suburbs
  - Common Venue is Hotels
- Cluster 3
  - 28 Suburbs
  - Common Venues include Hotels, Turkish Restaurants and Gyms
- Cluster 4
  - 21 Suburbs
  - Common Venues include Hotels, Greek Restaurants, and Food and Drink Shops
- Cluster 5
  - 16 Suburbs
  - Common Venues include Hotels, Shopping Malls and Steakhouse
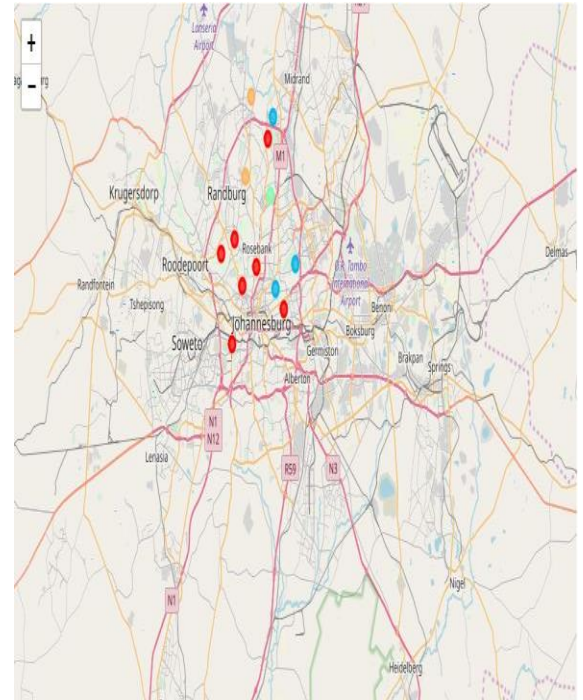
Discussion

- Cape Town has 83 suburbs with 675 venues. In addition, the geographical coordinate of Cape Town, South Africa is -33.928992, 18.417396. The best suburb to stay in is Durbanville with the following venues:

    - 1st Most Common Venue   Café
    - 2nd Most Common Venue   Gym
    - 3rd Most Common Venue   Thai Restaurant
    - 4th Most Common Venue   Pub
    - 5th Most Common Venue   Breakfast Spot
    - 6th Most Common Venue   Coffee Shop
    - 7th Most Common Venue   Fast Food Restaurant
    - 8th Most Common Venue   Steakhouse
    - 9th Most Common Venue   Grocery Store
    - 10th Most Common Venue   American Restaurant

Johannesburg has 119 suburbs with 198 venues. In addition, the geographical coordinate of Johannesburg, South Africa is -26.205, 28.049722. The best suburb to stay in is Bergvlei with the following venues:

- 1st Most Common Venue   Hotel
- 2nd Most Common Venue   Turkish Restaurant
- 3rd Most Common Venue   Gym
- 4th Most Common Venue   Greek Restaurant
- 5th Most Common Venue   Golf Course
- 6th Most Common Venue   French Restaurant
- 7th Most Common Venue   Food & Drink Shop
- 8th Most Common Venue   Exhibit
- 9th Most Common Venue   Deli / Bodega
- 10th Most Common Venue   Convenience Store

Many of the neighborhoods are homogenous and are very similar to each other. Both Johanessburg and Cape Town consist of suburb clusters that contain majority of the suburbs.

Conclusion

Cape Town had a significant more number of suburbs and venues than Johannesburg therefore it would be the better option to relocate to Cape Town, specifically Durbanville as the most efficient choice. Cape Town offers a variety in choices for restaurants, gyms, grocery stores, and extracurricular activities for individuals and families.