Segmenting and Clustering the Southern Suburbs and Northern Suburbs of Cape Town (S.A)

HOW SIMILAR ARE THE SOUTHERN SURBURBS TO THE NORTHEN SURBURBS OF CAPE TOWN?

Introduction

BACKGROUND

Cape Town is a City in the Western Cape Province of South Africa. The city is generally divided remarkably different from the rest of the cities in this country. Unlike in Europe, Neighborhoods are determined by street codes rather than postal codes. For example in the Southern Suburbs, Observatory street code 7925 is close to Kenilworth street code 7708. Looking at their postal code Kenilworth 7745 while Observatory is 7935. We can further compare a suburb in the South Peninsula Muizenberg with Postal Code 7950. If we use postal codes, this would suggest that Muizenberg is near Observatory. This is incorrect as Kenilworth is closer to Muizenberg.

We are thus going to use this fact to answer a simply question, which is mainly of interest to Employment Agencies and Prospective employees.
Employment Agencies have a tough time convincing prospective employees that the Northern suburbs are comparably similar to Southern Suburbs.
This is a very important question in Cape Town because in the south is where most of the beaches are found, as well as popular restaurants and hence attractive to residents.

PROBLEM STATEMENT

We are going to use suburb data that includes postal codes, street codes, longitude and latitude and some metrics to describe “neighborhoods”. We are then going to use foursquare API to see if there are comparatively popular places in the South as there are in the North.

INTERESTS GROUP

Employment Agencies will be interested to know how similar southern suburbs are to the northern suburbs. Prospective employees will also be interested in this result in making future decision with regards to change of employment

Data Acquisition and Cleaning

DATA ANALYSIS

To collect the data, I web scrapped a Wikipedia page that shows the suburbs of Cape Town, Postal Codes and Street Codes. The page can be found here.

In addition, data concerning longitudes and latitudes of the central points of the suburbs was obtained from the municipality of Plumstead. Using this data, I was able to allocate longitudes and latitudes of the central areas in the common street codes.

DATA CLEANING

Data was downloaded and scraped form the Wikipedia page. The data included missing street codes with some missing postal codes.
I proceeded by dropping all rows that had missing street code values. The remaining data was concatenated into a single data frame for further processing.
The final data frame had 126 from 132 street codes and 4 columns, which describe these areas.
The longitude and Latitude data was then added to the clean data frame.

FEATURE SELECTION

After examining the features, I discarded postal codes from the analysis.
I then mathematically calculated the correlation between street code and suburb.
Street code is highly correlated to suburb (Pearson Coefficient = 0, 85)

1.1

Exploratory Data Analysis

To visually classify the data, I used folium to show the distribution of neighborhoods in Cape Town. This allows us to see that the northern suburbs are on the other side of the Mother City as initially mentioned.
I then explored the neighborhoods and placing interest in the areas of interest in both south and north.
I then explored the categories of the areas of interest, using the central points of both Southern and Northern Suburbs.
I then plotted them on a map using foursquare API

Fig 1. Distribution of Neighborhoods in Cape Town

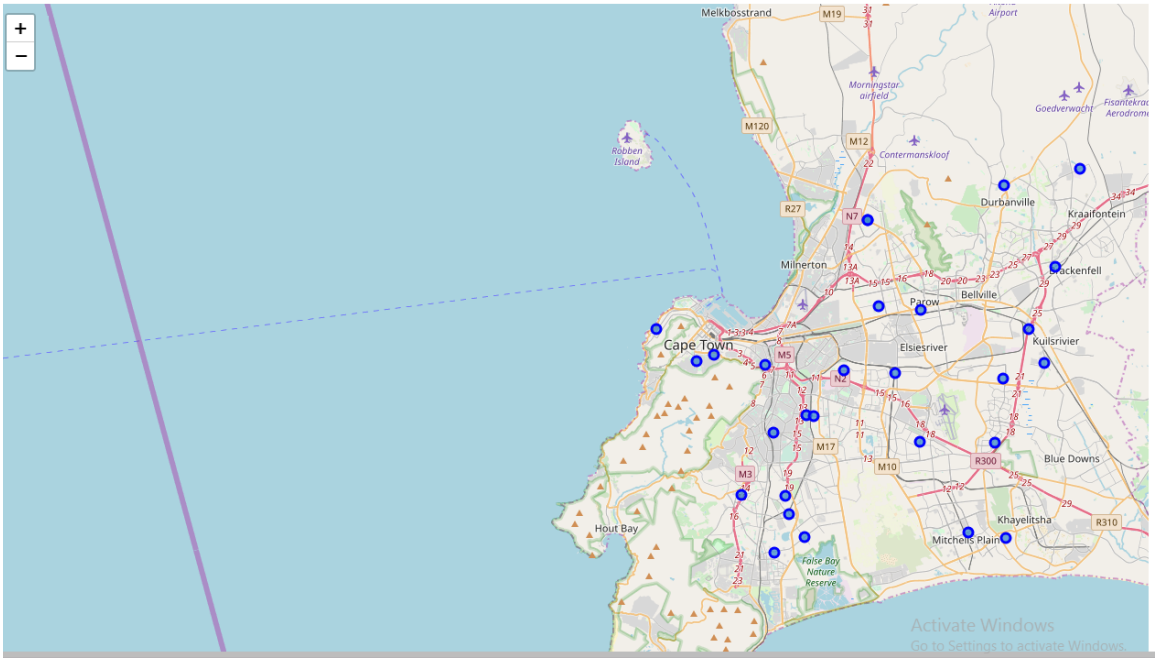


Fig 2 Exploration of Areas of interest in the Northern Suburbs

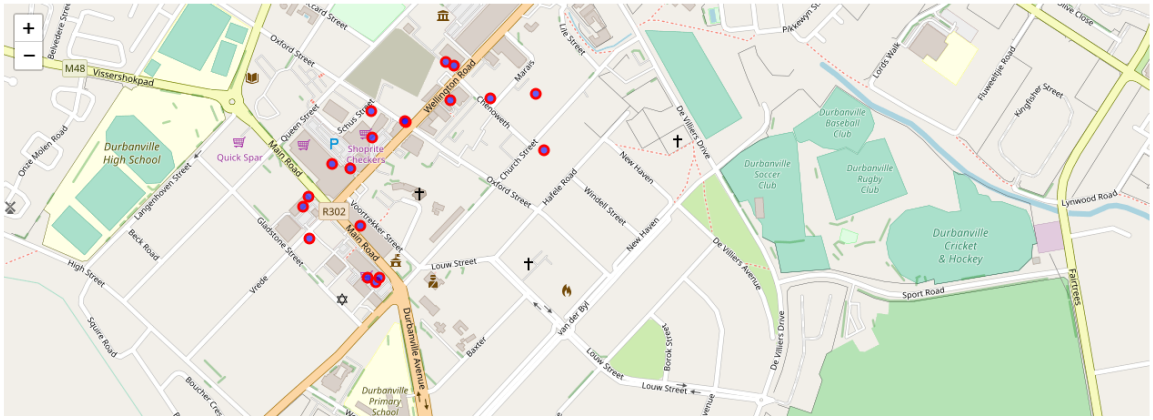
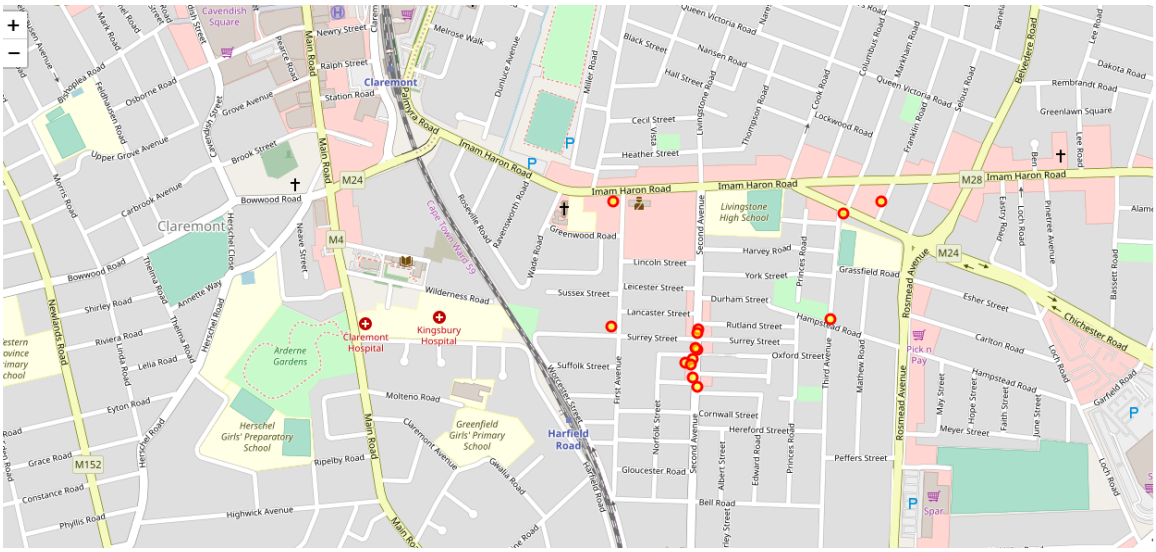


Fig 3 Exploration of Areas of interest in the Southern Suburbs



Clustering the neighborhoods

After clustering the neighborhoods is was observed that the Northern Suburbs have more variety in terms of categories of places than the Southern suburbs.
The southern suburbs have more concentrated places in vicinities.

Conclusion

In this study I analyzed that there are comparatively more amenities in the North than the South.
The scarcity of these venues is more prevalent in the North than the south.
Therefore, though the south and north are similar there are more attractive places in the north than the south.