

# The Battle of the Neighbourhoods.

Johannesburg, South Africa



VS

London, United Kingdom



# Is London the most ideal city to immigrate to?

Khrystyne Taylor

## Introduction.

We mostly believe, at one time or another, that ‘the grass is always greener on the other side’ and this is probably true for most people wanting to move outside of their home country with the belief that they would be improving their living conditions. Due to globalisation and increased connectivity because of technology advancements, people have become increasingly more interested in pursuing opportunities outside of their birth countries. This is particularly true for South Africans. Approximately 4.3 million South Africans have left their homes during the 20<sup>th</sup> and 21<sup>st</sup> century. The main reasons for South Africans looking for greener pastures is because of poor economic conditions, concerns about safety due to the high crime rate and better job opportunities abroad.

As per an article written by [FinGlobal](#):

*"As it turns out, psychology is very much at play when moving to a new country. Psychology tells us that as humans, we have an inbuilt need to belong. It's not enough for us just to belong to this huge clan of "humans", we actually want to have individualised belonging, which means we want to identify with specific groups of people in a population of humans. This comes into play quite evidently with emigration."*

We can, thus, assume that many South Africans, when deciding where to immigrate, would choose a location that would feel familiar. Given this, the analysis will focus on where South Africans most commonly live outside of South Africa and how similar this location is to their city in order to decide whether or not they would feel a sense of belonging and how much of a change they would have to adapt to.

## **Business Problem:**

The aim of our analysis is to help those South African that are looking to leave the country, in order to improve their living conditions, make a decision about which city would be best suited for them. The purpose and goal would be to uncover neighbourhoods in a given city that are most similar to the South African way of life. The findings from the analysis will help stakeholders make an informed decision based on their own preferences according to what the city has to offer, which includes, but is not limited to, entertainment, indoor and outdoor activities, different cuisines and stores.

# Data Description:

We first need to determine which countries are the most common for South Africans to move to. According to an article written by [BUSINESSTECH](#) the most common destinations for South African to move to are:

- .New Zealand

- .The United Kingdom

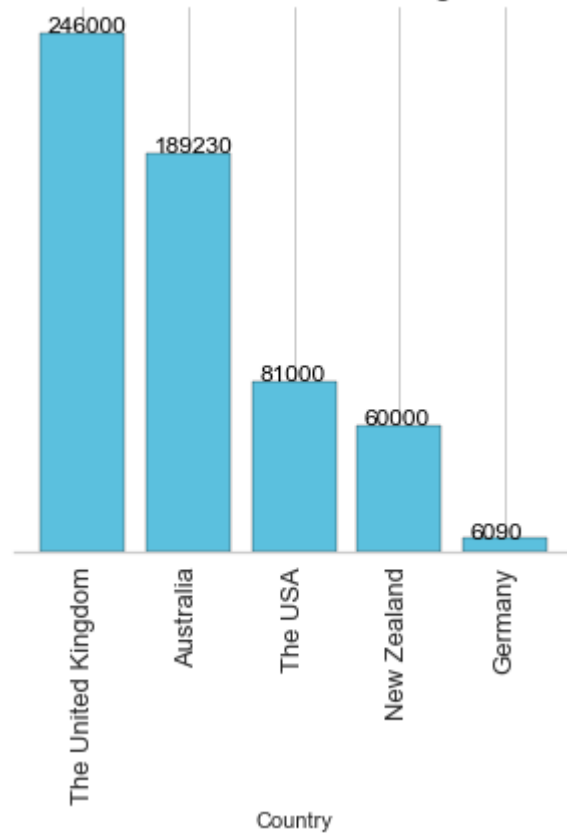
- .Australia

- .The USA

- .Germany

.It can be seen on the next slide the count of South African living in each of these countries:

Number of South Africans living Abroad



We can see from the above graph that the UK has the most South African immigrants, moreover, according to [BBC UK](#), the most popular areas in the UK for South Africans to reside are Wimbledon N, Putney E, Putney W, Golders Green, Wimbledon S, Richmond N, Highgate, Acton, Fulham, and Kensington. Therefore, we will collect and use data for London for our analysis. Additionally, because we are making a comparison of two locations, we will collect data for a South African city too. Moreover, to aid our analysis we will need geographical location data, that is, the longitude and latitude of each area in the city. The postal codes for each location will help us get started in this regard.

**- London data collection:**

The best possible source to gather London location data is from [https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London) as this page has geolocation data about all neighbourhoods in London that includes:

- .Location(area)
- .London Borough
- .Post town
- .Postcode district

We will gather this data by scraping this wikipedia page with the use of the Python library Pandas.

Unfortunately, the wikipedia page does not include information about the geographical locations we need for London. In order to solve this, we will use the pgeocode library to pull the necessary information from the web using each locations postal code.

**Pgeocode:**

pgeocode is a Python library for high performance off-line querying of GPS coordinates, region name and municipality name from postal codes. Distances between postal codes as well as general distance queries are also supported. The used GeoNames database includes postal codes for 83 countries.

We have used this library to obtain latitude and longitude coordinates based on each locations postal code. The resulting pandas dataset includes the coordinates for each neighbourhood(borough) in London.

**- Data collection of a South African City:**

According to Stats SA, the province of Gauteng has the largest population, with approximately 15,5 million people living here. We will, therefore, gather data for Johannesburg, which is the capital city of Gauteng.

This city data will be obtained from [www.blaauwberg.net](http://www.blaauwberg.net) where there is a publicly available database containing South African Postcodes.

- This website includes postal information for the cities that is spread across a total of 26 pages. Each page will be scraped to obtain the data needed. After scraping we will have a data frame consisting of the following variables:

- Place Name(neighbourhood)
- Street Code(Post Code)
- PO Box Code
- City
- Province



Unfortunately, the pgeocode library did not work well to obtain the geolocation data for the areas in Johannesburg. Due to this not working in our favour, we will use the arcGIS API instead.

**ArcGIS API:**

ArcGIS API for Python is a Python library for working with maps and geospatial data, powered by web GIS.

We will specifically use arcGIS API location functionality to retrieve the longitude and latitude coordinates based on each locations postal code. The resulting pandas dataset includes the coordinates for each neighbourhood in Johannesburg.

**Foursquare API data collection:**

The Fourquare Apl provides location based data for a specific are of interest. Bt querying this API we can obtain information on an areas food outlets, malls, parks, stadiums or any venue of interest that is location specific.

Retrieving this data based on areas in London and Johannesburg is essential in order to answer our initial research questions , and will, also, allow stakeholders to make informed decisions based on venues in each location.

Once we have retrieved postal and geospacial data on areas in both Johannesburg and London, we will connect to the Foursquare API with the use of each initial dataset and the location data therein in order to gather the venue data we require for our final analysis. For both Johannesburg and London, we will request the API to retrieve information on all venues within a 500 meter radius of the longitude and latitude coordinates of each area.

.

Once the venue data for each city has been retrieved from the API call, the resulting data obtained will be as follows:

Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bexley, Greenwich	51.4869	0.1075	Co-op Food	51.487650	0.113490	Grocery Store
Bexley, Greenwich	51.4869	0.1075	Bostal Gardens	51.486670	0.110462	Playground
Bexley, Greenwich	51.4869	0.1075	Meghna Tandoori	51.485709	0.101681	Indian Restaurant
City	51.5085	-0.1257	National Gallery	51.508876	-0.128478	Art Museum
City	51.5085	-0.1257	Gordon's Wine Bar	51.507911	-0.123293	Wine Bar

Where each variable has the following information:

- .Neighbourhood (areas in the city of choice)
- .Neighbourhood Latitude (the latitude coordinates for each area)
- .Neighbourhood Longitude (the longitude coordinated for each area)
- .Venue (a specific venue name within a specific area)
- .Venue Latitude (the latitude coordinates for a specific venue)
- .Venue Longitude (the longitude coordinates for a specific venue)
- .Venue Category (the specific category in which the venue belongs, e.g., restaurant, bar, grocery store, etc.)

**Once all data is obtained we will then build our models based on a cluster analysis using the K-means method in order to analysis venue data for each city and present results for our stakeholders.**