

Vacationing in CapeTown, an analysis of activities and accommodation rental prices

Paul Bolton June 16, 2020

1. Problem and Background discussion

Living in South Africa, one of the most visited Tourist spots is the wider Capetown area, as a potential visitor to this area one is inundated with a huge amount of options on what to do and where to stay so much so that it can be overwhelming on how to make the most advantages decision to get the most out of ones vacation to this beautiful area. In approaching this problem there are two questions that I want to focus on:

- What are my options on things that I can do?
- Where should I stay in relation to the activities that I want to perform and what is a reasonable price that I can spend on accommodation.

2. Data acquisition and cleaning

2.1 Data sources

From my understanding there is no available data that provides a list of Boroughs with their respective Neighbourhoods and coordinates, to get over this hurdle I scrapped a [property website](#) to obtain a list of Neighbourhoods that reside within this area, due to it being set up to cater for property sales as well as for the whole of the Western Cape area, I revised the scope in two ways:

- All neighbourhoods that resided within the larger Cape Town area were included under Cape Town.
- The list was revised where a person would be able to take a day trip to the specific area all other areas were removed from the list due to their proximity being too far from Cape Town.

Once a list of Neighbourhoods was obtained, I used Geopy to obtain the coordinates for each of the Neighbourhoods. With this list I then used it to obtain the following two additional data sources:

- As input into the **Forsquare API** to obtain the most common venues and activities for the given neighbourhoods within the wider Cape Town area.
- As input into the Air Bnb Website where I scrapped all options to stay with in each of these areas as well as their respective prices per night.

2.2 Data cleaning

As Mentioned earlier a lot of the information is not available in a tabular format and had to be scrapped from websites and enriched. Technologies utilised for these processes:

- BeautifulSoup to scrape websites
- Geopy to obtain the coordinates of the Neighbourhoods
- Foursquare API to obtain venues and activities with their respective coordinates.