**Data Description:**

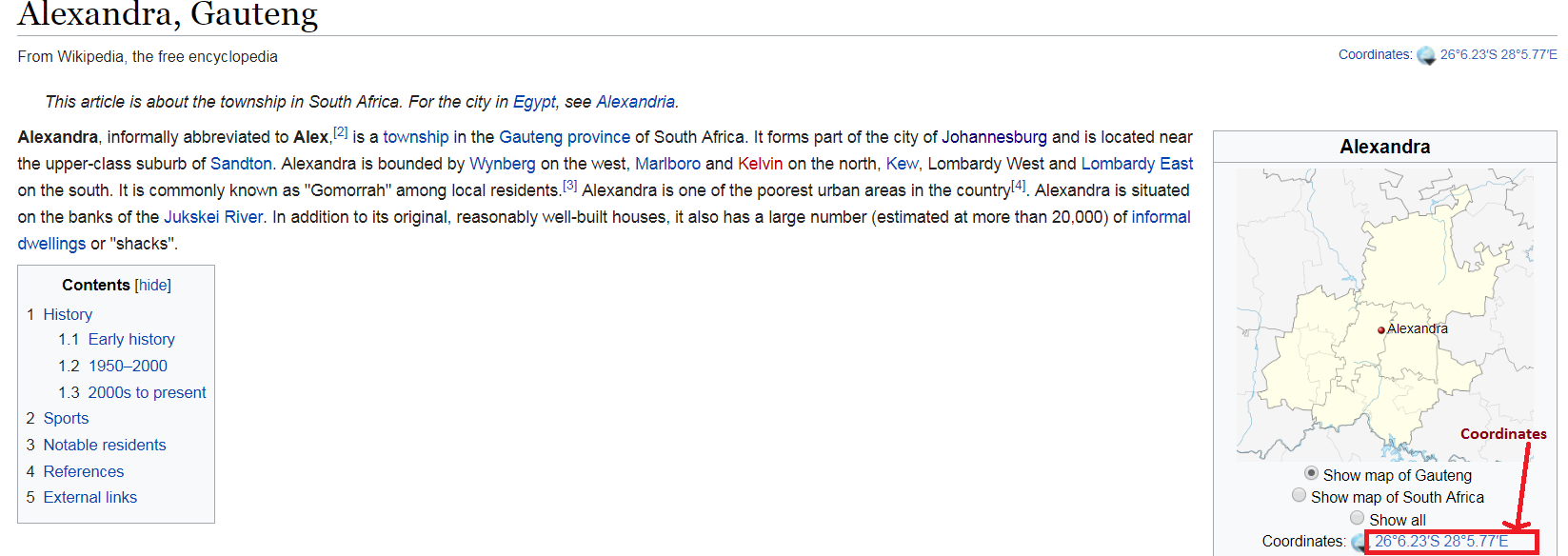
The dataset for this project will built from the following sources:

1. City of Johannesburg Municipality Wikipedia Page: <https://en.wikipedia.org/wiki/City_of_Johannesburg_Metropolitan_Municipality>)
2. Crime Stats SA: <https://www.crimestatssa.com/index.php>
3. Foursquare location data: [https://foursquare.com](https://foursquare.com/)

**City of Johannesburg Municipality Wikipedia Page:**

This page contains list of suburbs that are under the City of Johannesburg Municipality as well as the links to individual Wikipedia pages for each suburb, which then has coordinates the suburb. The City of Johannesburg Municipality is divided into 7 regions (Region A – Region G) and each suburb belongs to a particular region.

For Example below is the screenshot of the Alexandra suburb Wikipedia page:



The name of the suburb will be used to identify the suburb and the coordinates will be used in the Foursquare API and also for visualizing the results of the project on a map of Johannesburg.

**Crime Stats SA:**

Crime stats is a website that displays the latest and historic South African crime statistics in an easy-to-understand format. The website has crime data for areas all over South Africa. Total crime data is a sum of 39 different crime categories for each suburb.

For example below is the screenshot, from the Crime Stats SA website, showing crime data for Sandton, Johannesburg.



Figure 1Total Crime Sandton in 2018.

The crime data will be used to determine the crime rate per suburb.

**Foursquare Location Data:**

Foursquare is a technology company that built a massive dataset of accurate location data.

Foursquare location data will be used to find number of available schools and healthcare facilities within a determined radius for each suburb. For schools we will be concentrating on primary schools, elementary schools, high schools and colleges, but for healthcare facilities the project will consider all healthcare establishments within a suburb.