

Introduction and Problem Statement

The title of the project is '*Choosing a suburb to stay in Johannesburg*'. Johannesburg is the largest city in South Africa and one of the 50 largest urban areas in the world. It is the provincial capital and largest city of Gauteng, which is the wealthiest province in South Africa, this makes Johannesburg attractive for immigrants, investors, and people hoping for a better life. In 1950, the population of Johannesburg was 910,550 and as of 2019, Johannesburg has a population of about 5,635,127. Immigration accounts for a significant 17% of SA's population growth. For a newcomer to the big city, it can be a daunting task to determine which part of Johannesburg to settle in depending on a person's preferences.

Deciding which suburb to move into in Johannesburg, should be trouble-free. People look for different things to determine a suitable place to stay depending on their needs, priorities and likings. Some of the factors people consider to determine neighborhoods to stay in includes crime rate, schools, healthcare availability, jobs, cost of living, etc. A platform, either mobile application or website, that allows people to screen suburbs in Johannesburg on the mentioned factors would make a huge difference and save people time, energy and money that goes into choosing the most favorable place to settle into.

At the present moment choosing the right suburb to settle into involves a lot of work and stress on individuals. This includes searching on the information internet, making calls and visiting different suburbs physically. A website, <http://www.teleport.org> provides information about the whole city of Johannesburg for people willing to relocate to the city but the site does not analyze individual suburbs within Johannesburg.

In response to this problem my project proposes to come up with a solution. My solution uses machine learning, clustering in particular, to analyze profiles of suburbs in Johannesburg and categorizes them. The project considers three factors which are crime rate, school availability and healthcare availability in each suburb. The suburbs are then categorized into three clusters, from best to worst based on crime rate, school availability and healthcare availability.