**Task – L02T18**

**Exploratory Data Analysis on the South African Hospitals Survey 2011-2012 Dataset**

Introduction

The dataset is a summary of a survey conducted between 2011-2012 of the South African hospitals and clinics in various provinces. Some of the results of the survey include the hospital or clinic performance, facilities, patient safety, cleanliness, waiting times, availability of medicines and supplies, doctors, nursing staff and other staff at the facilities

The dataset was extracted from the followingwebsite [South African Hospitals Survey 2011-2012 - Datasets - OpenUp Data Portal](https://data.openup.org.za/dataset/south-african-hospitals-survey-2011-2012-ft5b-smjr)

Data Cleaning

In order to understand the dataset I studied the dataset by displaying the first five rows of the dataset, displayed the total number of columns and rows of the dataset and also checked for missing values.

In the dataset there were a lot of missing values in some of the columns. I therefore performed the following procedure as part of data cleaning:

* I used the drop function to delete the columns that had missing values

Another method which I used as part of data cleaning was the following:

* For loop to iterate through the location column in order to access the provinces. Then I used the replace method to replace the values that were not relevant in the location column.

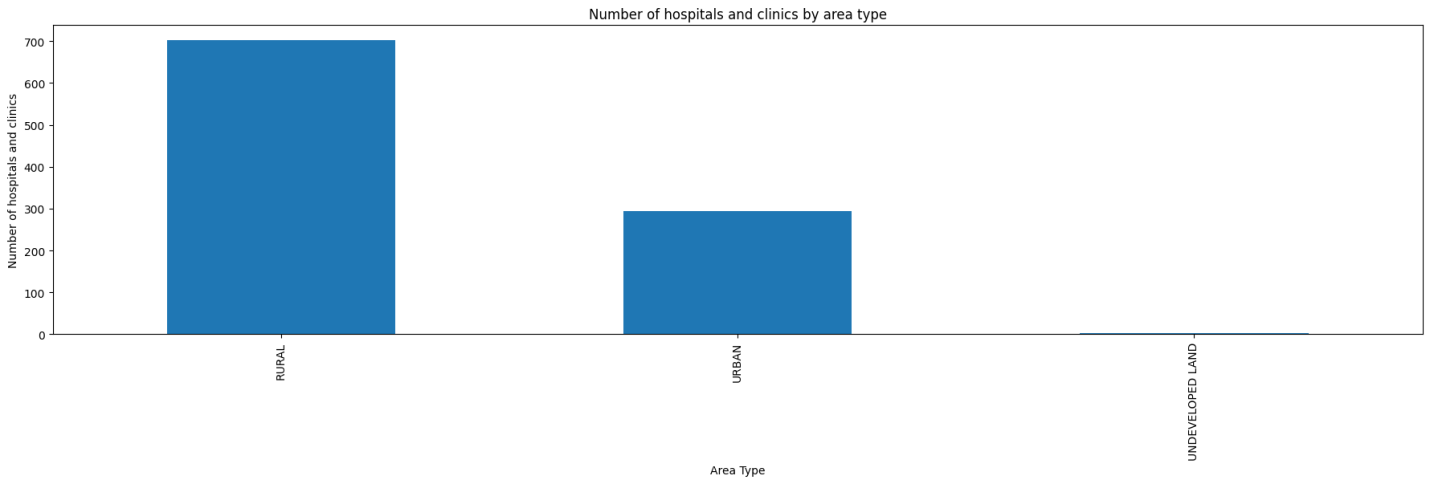
Data Stories and Visualisations

1. Sorting in ascending order by column 'overall\_performance'



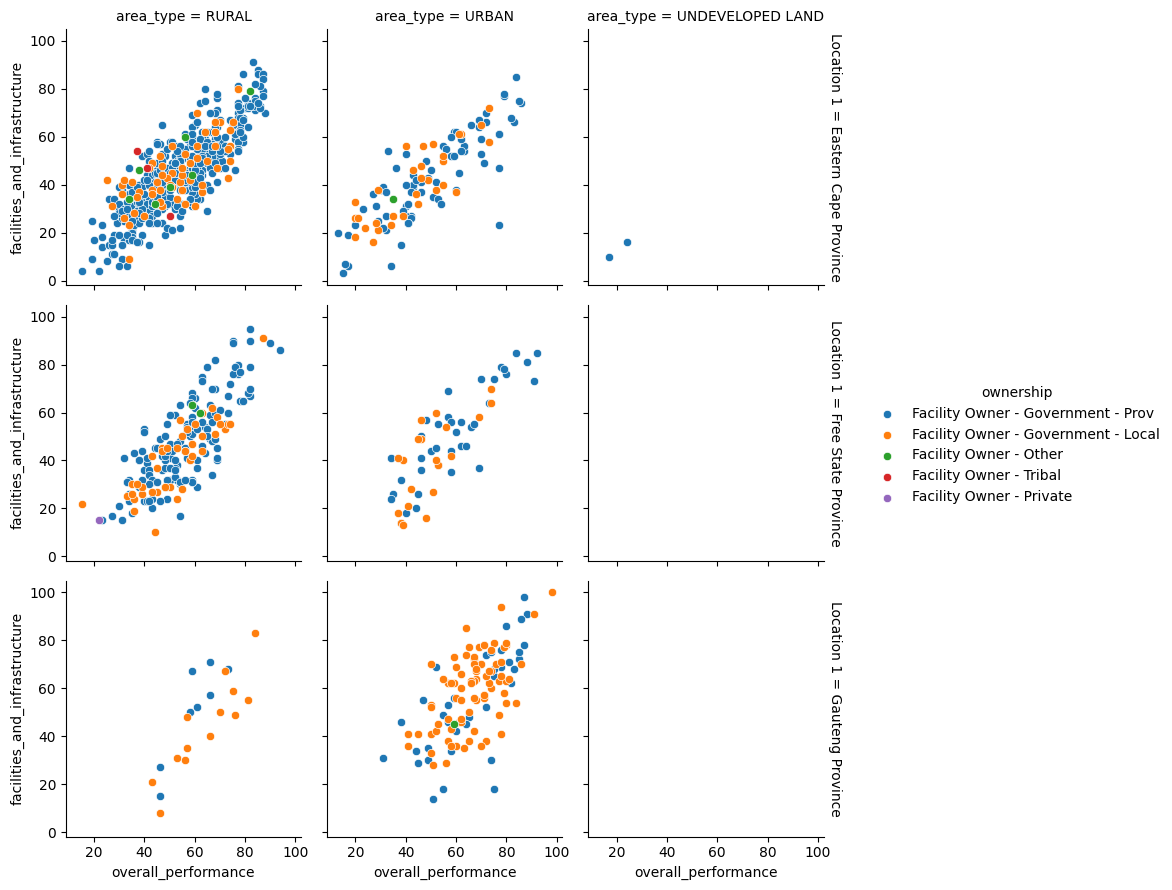
I sorted the dataset by overall performance in order to extract the least performing hospitals and clinics according to the survey dataset. According to the results, only clinics are appearing in the least performing hospitals and clinics. Eastern Cape Province has the most number of clinics that did not perform well.

1. Plotted a bar graph which shows the number of hospitals and clinics by area type



The graph shows that there are more rural hospitals and clinics that took part in the survey than urban and undeveloped land

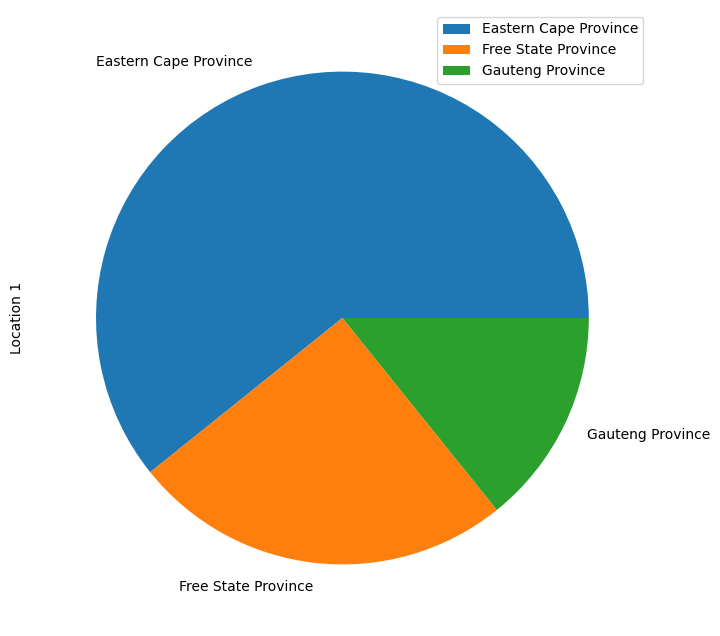
1. Plotted a multi-plot grid



The multi-plot grid shows that the higher the percentage of availability of facilities and infrastructure the higher the overall performance.

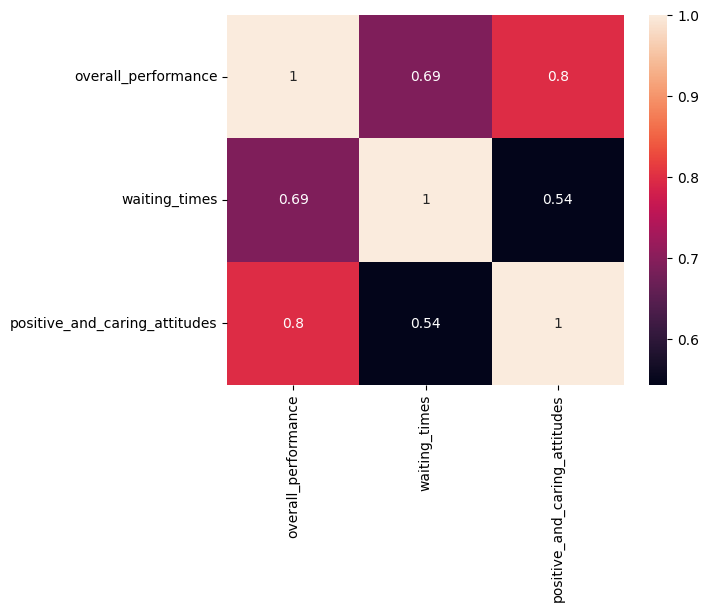
In addition there were more hospitals and clinics from the Eastern Cape Province than from Free State and Gauteng Province that participated in the survey.

1. Plotted a pie chart



The pie chart shows that Eastern Cape Province had the most hospitals and clinics in the survey, followed by Free State Province and then Gauteng Province.

1. Plotted a heat map to see the relationship between overall performance, waiting times and positive and caring attitudes.



There is a large correlation between waiting times, positive and caring attitude, and overall performance since the correlations are all above 0.5.

We can see that hospitals and clinics that scored high in the attribute of positive and caring ranked high in overall performance.

THIS REPORT WAS WRITTEN BY: MITCHEL MUNAKANDAFA