**Introduction**

This dataset contains information about the use of various forms of renewable energy in the 9 South African provinces.

Along with the type of technology the dataset also has capacity values for each renewable energy system.

**DATA CLEANING**

- Import the numpy and pandas packages to read-in the csv file and analyse the information in it.

- Import seaborn and matplotlib to create visualizations for different aspects of the dataset.

- Use the .describe() method to return a detailed numeric and overall analysis of the dataset to find unique and missing values, as well as the mean value for columns with numeric data.

- Drop columns with redundant data (Owner, Programme, Technology, Website and Company Name).

- After loading the dataset I noticed that the location 1 column had co-ordinates attached to the end of each Province name, this was probably the exact location where the renewable energy system could be found, but since the province name was what I needed I created 2 functions to edit the data in this column:

- The first function converts every value in the location 1 column to a list and then return everything before the co-ordinates (the province name).

- The other function converts all the now list values in the column to string values.

- Create the new\_data object to save the dataset with the edited column values.

MISSING DATA

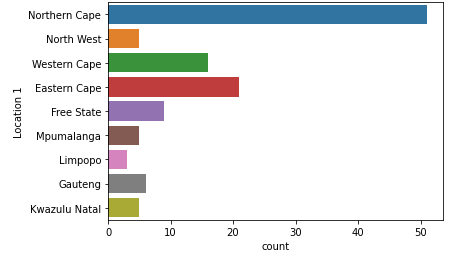
After using the .describe() method mentioned above I noticed that the Company name column had over 30 rows with None as the value, since there is another column called project name with no missing values that could be used to find the company name I decided to drop the company name column.

**DATA STORIES AND VISUALIZATIONS**

**Bar graph**

-With the seaborn module I displayed a bar graph showing which provinces used renewable energy resources.

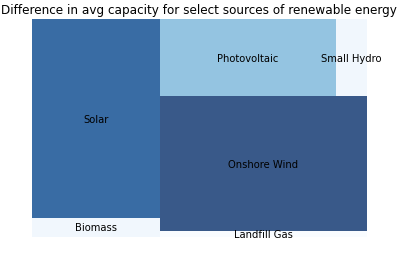
-The visual showed that the Northern Cape Province used the most renewable energy resources at 51.



**Treemap**

-With the squarify module I displayed a treemap showing the difference in the average capacity value for each form of renewable energy.

-It can be seen that out of the six forms of renewable energy in this dataset, the forms of renewable energy with the highest average capacity are onshore wind, solar and photovoltaic energy.

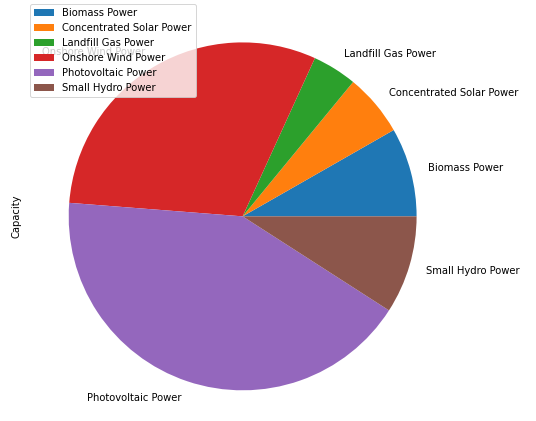


**Pie chart**

-Displayed a pie chart showing how often each form of renewable energy was used in the country.

-From this chart we learned that onshore wind and photovoltaic power are the 2 most used forms of renewable energy, with landfill gas and solar power being the

2 least used forms of renewable energy in the dataset.



**Conclusion**

Based on the information gathered from the visuals created it is clear that onshore wind power has the highest average capacity, however since the technology for this resource can only be used in coastal regions I would suggest the use of solar power for people livng inland. Solar power has the second highest average capacity and can be used anywhere with direct sunlight.

**THIS REPORT WAS WRITTEN BY : Darren Vandeyar**

**Reference:**

[**https://data.openup.org.za/**](https://data.openup.org.za/)