

As you have already learned in this bootcamp, as a Data Scientist you need to be able to display a very important set of skills. You need to be able to write code that can be used to clean, organise, analyse and visualise data. These coding and analytical skills need to be combined with the ability to communicate your findings to stakeholders effectively.

In this task, you will be given the opportunity to showcase all the skills which you have been developing throughout this bootcamp. You will be tasked with analysing a dataset of your choice. Furthermore, you will create a report that contains several data visualisations that will clearly communicate your findings. Finally, you will present your report for assessment.

You are free to choose any dataset of your choice to analyse. Here are some sources of datasets:

- [United Nations datasets](#)
- [South African Open Data](#)
- [World Bank Data](#)

Your report should include the following:

- A brief description of the dataset you have chosen
- A description of how your dataset was cleaned
- A description of how you managed missing data
- A detailed description of your findings accompanied by appropriate visualisations
- A wide selection of appropriate data visualisations. Besides Matplotlib, we encourage you to explore and use libraries other than with more interesting visualisations

A template for the report (**EDA_doc.odt**) accompanies this task.

Ensure that your report is professional! Remember, such a report can be added in your professional portfolio.

Capstone Project

- Select a dataset. **Note:** Once you've selected a dataset please make sure to check it with a reviewer before starting.
- Create the report as described in this task.
- When you submit your report for review, also book a call with a reviewer so that you can present your report and findings to them.