SACAIR Hackathon – 4 December 2023

Getting Started

Table of Contents

[1. Introduction 3](#_Toc151477099)

[2. Generate Problem Statement 3](#_Toc151477100)

[3. Design 3](#_Toc151477101)

[4. Development / Modelling 3](#_Toc151477102)

[5. Presentation 4](#_Toc151477103)

[6. Assessment Criteria 4](#_Toc151477104)

[7. Environment and Submissions 4](#_Toc151477105)

[7.1 Starting resources 4](#_Toc151477106)

[7.2 Environment 5](#_Toc151477107)

[7.3 Submissions 6](#_Toc151477108)

[8. Dataset Details 6](#_Toc151477109)

[8.1 Eskom Data 6](#_Toc151477110)

[8.2 Weather Data 6](#_Toc151477111)

# Introduction

A hackathon, also known as a codefest, is a social coding event that brings computer programmers and other interested people together to improve upon or build a new software program.

In this Hackathon, we will be reviewing data from Eskom, finding insights, designing a process flow and building a model to predict something interesting.

You will be assessed based on the novelty, creativity and usefulness of the problem you are try to solve, the model built, and how the insights are presented.

As this is only a day-long event, you will not be judged on the accuracy of the data model (they can take weeks/months to develop properly). You will be assessed on insightfulness shown, the creativity of the problem you’re trying to solve, the use of good data science practice and principles, making clever, logical assumptions, and presenting your findings in a clear, professional manner that tells the data story.

The following sections will guide you as to what is expected for this Hackathon.

# Generate Problem Statement

Explore the dataset. Decide on which problem you are trying to solve. What is a question that would be most useful to answer? Think about the business, engineering, social and/or environmental benefits in solving this problem. Be Creative!

Suggested problems could include:

* 1. How much diesel does Eskom use?
  2. When would be the optimal times to perform planned maintenance on the generation fleet?

Feel free to find additional datasets to answer the question you’re trying to solve.

# Design

Given the problem you are trying to solve, think about where that data might come from, and how it could be ingested into your platform, processed, stored and served. Think about and discuss how this could be implemented in practice.

Give a basic, high-level process flow from source data to front-end reporting.

# Development / Modelling

Identify the dependent and independent variables in what you are trying to represent and find correlations between them. Identify the most appropriate model to solve the problem and build it. What are the assumptions included in your modelling and/or decision-making?

Of course, you will not be able to build an end-to-end solution. Credit is awarded for good data science practice, making of good, logical assumptions, identifying and using the correct data to address the problem statement.

# Presentation

You will be required to present a 3-minute elevator pitch with slide deck. This should tell the data story. Talk us through the model, the assumptions and the solution. Credit is awarded for the content of the presentation and talk, not necessarily how pretty the slides look.

# Assessment Criteria

You will be assessed on the following criteria (from most to least important):

1. Completeness of vision.
2. Creativity – Is the problem you are trying to solve unique and insightful?
3. Good data science practice.
4. Presentation – data storytelling.
5. Accuracy of the model.
6. Collaboration.

# Environment and Submissions

## Environment

This Python assumes that you will be working in Python and using Notebooks for coding. The suggested environment for this project is Google Colab.

Navigate to: <https://colab.google/> and sign in with your Google account, and select **Open Colab**

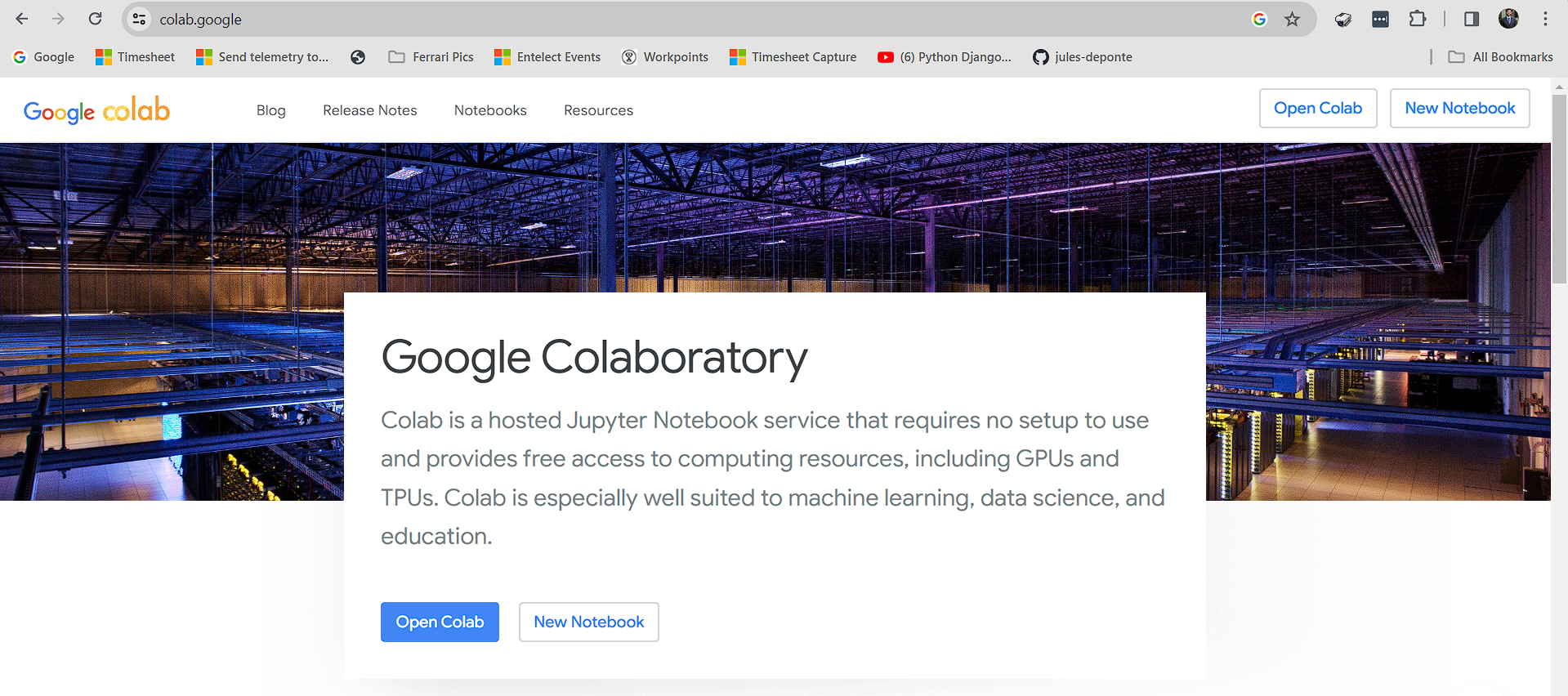


Figure 1 – Google Colab

Select **GitHub**, and enter the following URL:

<https://github.com/jules-deponte-entelect/SACAIR-Hackathon-2023-Getting-Started.git>

**NOTE**: Make sure to select the repo called: **SACAIR-Hackathon-2023-Getting-Started**

Select the notebook called: **SACAIR Hackathon Getting Started.ipynb**

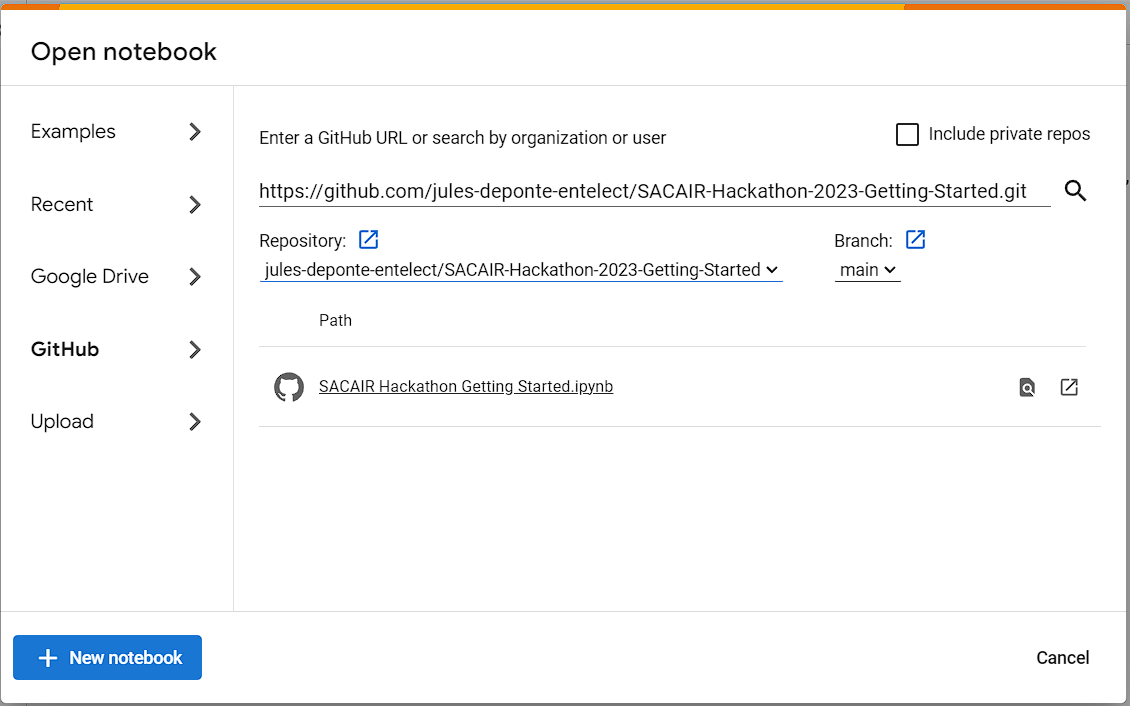


Figure 2 – Open Colab and select GitHub. Choose the SACAIR Hackathon Getting Started.ipynb notebook

Run the first cell, which will clone the repo to the environment. Wait until the cell finishes running. Once it is done, you will find a folder called **SACAIR-Hackathon-2023** in your Google Drive account. This contains the data and starter code for your project.

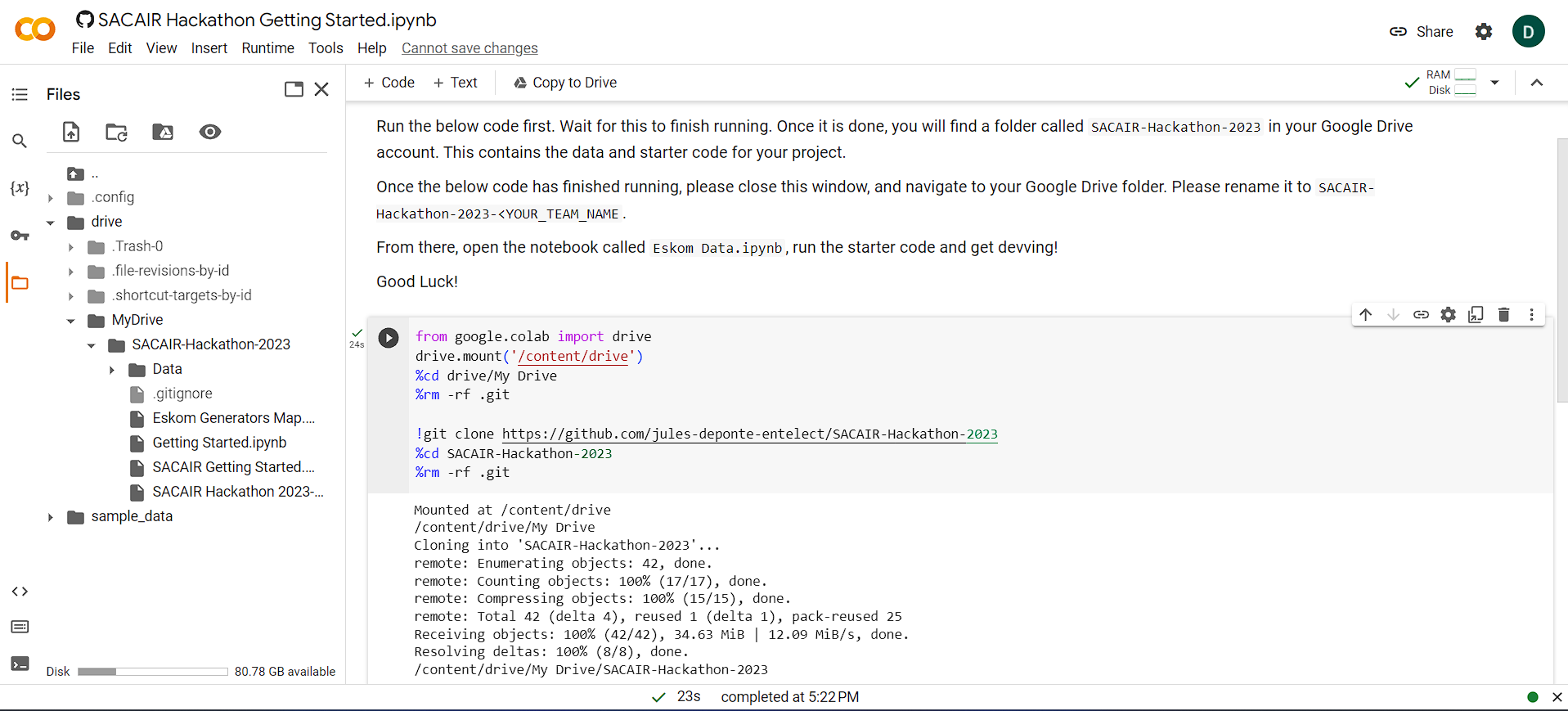


Figure 3 – Run code from the SACAIR Hackathon Getting Started.ipynb notebook

Once the above code has finished running, please close this window, and navigate to your Google Drive folder. Please rename it to **SACAIR-Hackathon-2023-<YOUR\_TEAM\_NAME>**.

From there, open the notebook called **Eskom Data.ipynb**, run the starter code and get devving!

A screenshot of a computer

Description automatically generated

Figure 4 – SACAIR Hackathon 2023 folder.

This starter code will mount your Google Drive to the notebook, and allow you to use the data stored in the Google Drive.

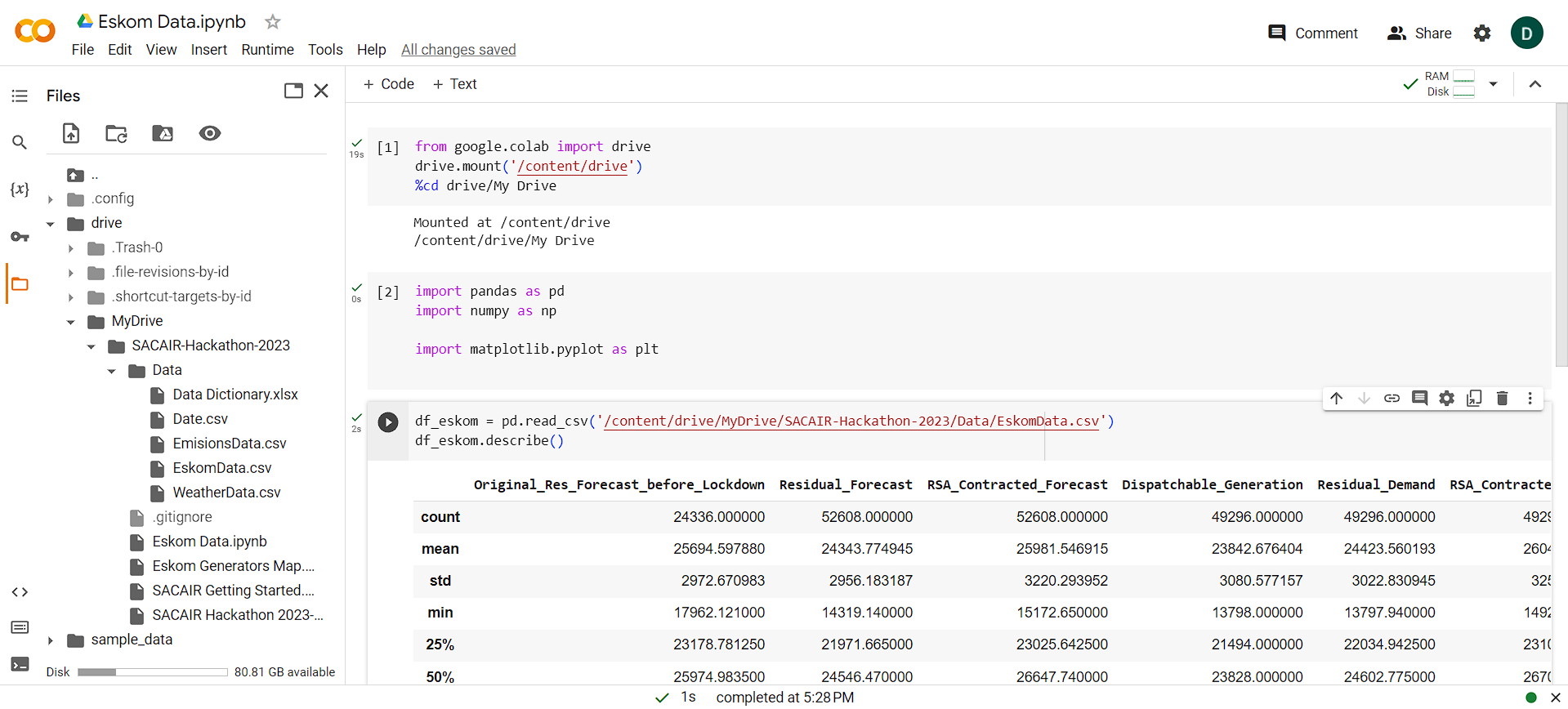


Figure 5 – Starter code from Eskom Data.ipynb

## Submissions

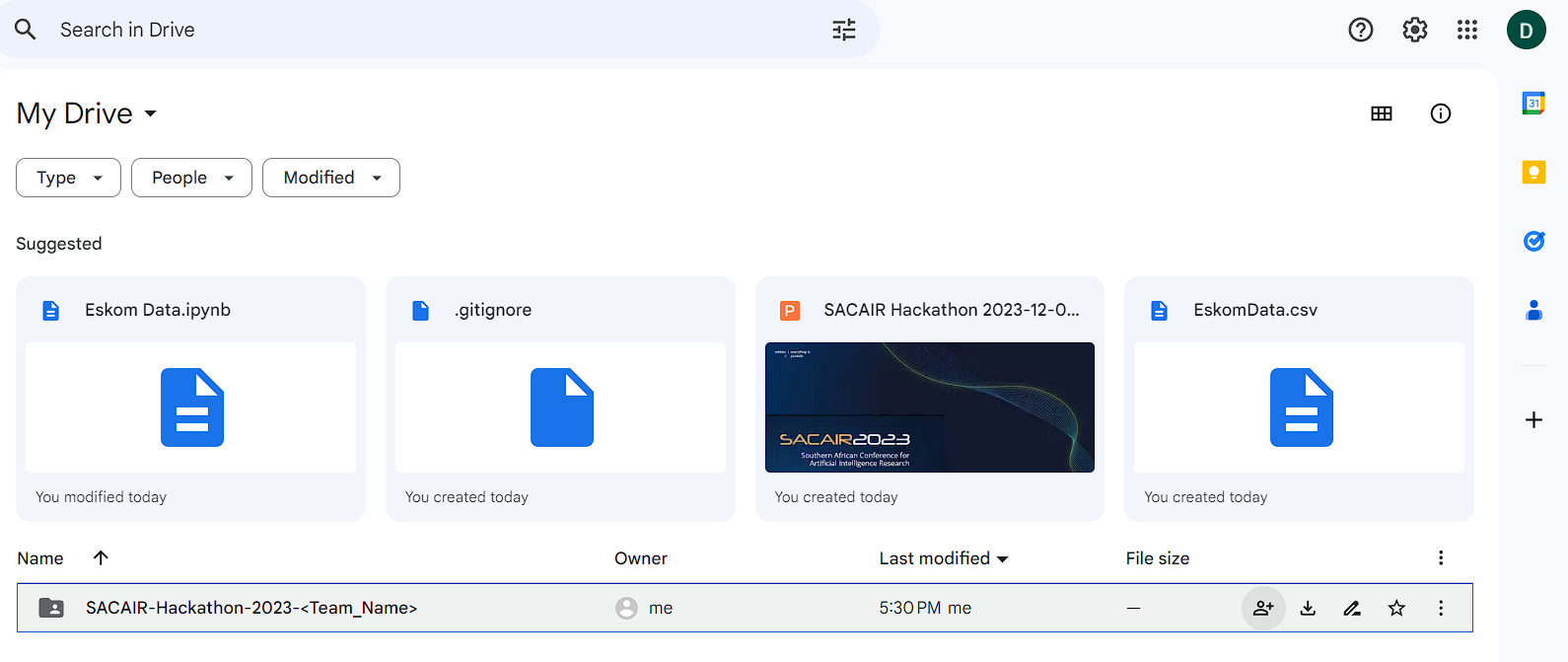
When you’re done with your project, ensure that all of your files, your notebooks, presentation, etc. are stored in the **SACAIR-Hackathon-2023-<TEAN\_NAME>** folder in your Google Drive.

Share the folder with:

[jules.deponte@entelect.co.za](mailto:jules.deponte@entelect.co.za)

[danielle.winter@entelect.co.za](mailto:danielle.winter@entelect.co.za)

Share with **Editor** access. This will count as your submission. **If the folder name does not contain your team’s name, your submission will not count.**



# Dataset Details

The following datasets are provided. Feel free to find additional data to support your project.

## Eskom Data

1. Data is supplied at an hourly granularity.
2. Data about Eskom power demand, energy generation and capacity (MW), averaged over an hour interval.
3. Energy generation by generation/supply type (international imports, coal, gas, open cycle gas turbine, hydro-electric)
4. Installed generating capacity.
5. The rated energy generation capacity by generation type.
6. The various types of outages that contribute to load shedding (these outages are described in the data dictionary)
7. Energy consumption by power plants used to generate power (referred to as Synchronous Condenser Operation. A technical understanding of this is not needed; a basic description is given in the data dictionary).
8. Tons of Carbon emissions per month. This is the lowest level of granularity of data available.

## Weather Data

1. Data is supplied at a daily granularity.
2. Data is supplied for the following locations:
   1. Johannesburg
   2. Pretoria
   3. Cape Town
   4. Durban
   5. Port Elizabeth
   6. Bloemfontein
   7. Upington
   8. Koekenaap
3. Temperatures, min, max, average (deg C)
4. “Feels like” temperature min, max, average.
5. Solar energy ()

Wind speed (km/h) and direction (deg) (North at 0 deg)