EXPLORE || DIGITAL SKILLS

Visualisation for Data Science
Predict

Predict Summary

In this predict, we will use PowerBI to connect to new data sources related to Eskom's power generation, change the data source locations, and create new visuals to derive the insight required to appropriately answer several MCQs.

Theme	Objective	Outcome
1 Data Sourcing	Experiment with different PowerBI data connectors	• Changed the source of the PowerBI data files to the path on your local machine • Sourced additional Eskom Excel data files
2 Data Formatting	Clean, transform and format sourced data	• Filled the missing values of the Eskom Excel files with the approximate feature averages • Created feature relations

3 Data Visualisation	Fix 'broken' dashboard and create new visuals	• Added various visuals including:
		  ○ Slicers,
		  ○ Cards, and
		  ○ Line charts

Agenda

01 **Problem Introduction**

02 Connecting to the Data

03 Sourcing Additional Data and Setting Relationships

04 Creating Visuals

EXPLORE || DIGITAL SKILLS

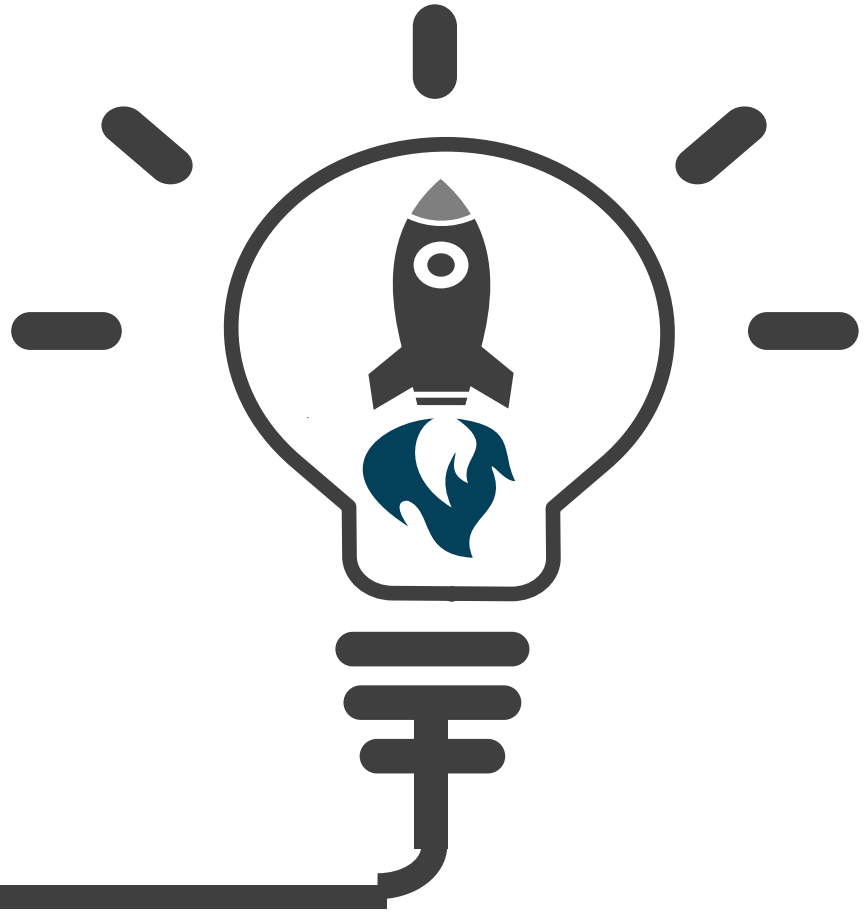
Problem Introduction: The National Energy Crisis

In this predict, we will use Eskom data to build an informative dashboard using Microsoft Power BI.
The dashboard, with its underlying reports, will be used to derive insight into the current South African energy crisis and answer multiple-choice questions.

You will be supplied with a 'broken' dashboard and the underlying data.

Your task is to:
1) Correctly source the data
2) Fix the relationships, and
3) Create the appropriate visuals

The reports you have created will be used to answer the multiple choice questions at the end of this predict.



EXPLORE | DIGITAL SKILLS

Agenda

01 Problem Introduction

02 **Connecting to the Data**

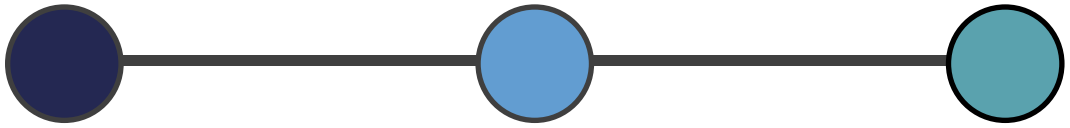
03 Sourcing Additional Data and Setting Relationships

04 Creating Visuals

EXPLORE DIGITAL SKILLS

Connecting to the Data

In this predict, we will use XLSX and CSV files to source the data.



Step 1

Download and unzip the folder with the XLSX and CSV files:

Eskom-Data-Power-BI-Predict.zip

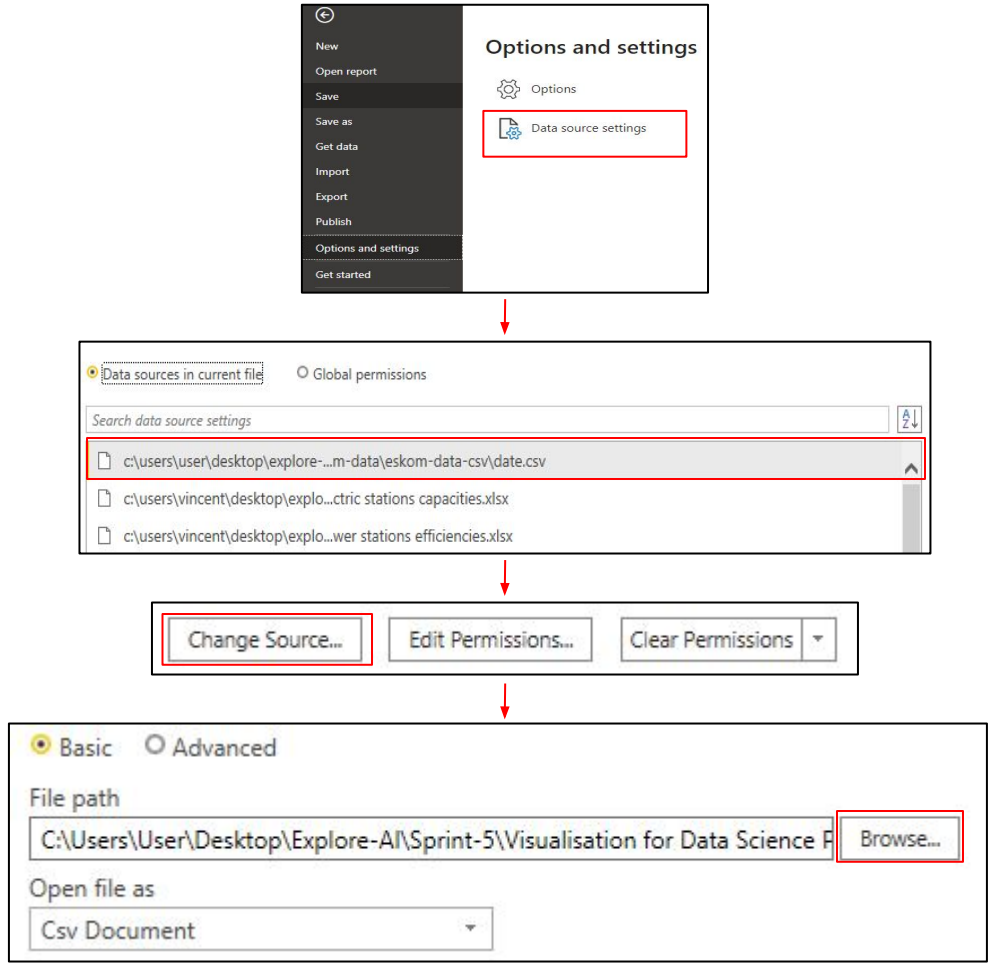
Step 2

Download the Power BI file containing the broken 'dashboard':

Eskom-Visualisation4DS-student-version.pbix

Step 3

Change the data source directory for each file in the *Eskom-Data-Power-BI-Predict* folder.



EXPLORE DIGITAL SKILLS

Agenda

01 Problem Introduction

02 Connecting to the Data

03 **Sourcing Additional Data and Setting Relationships**

04 Creating Visuals

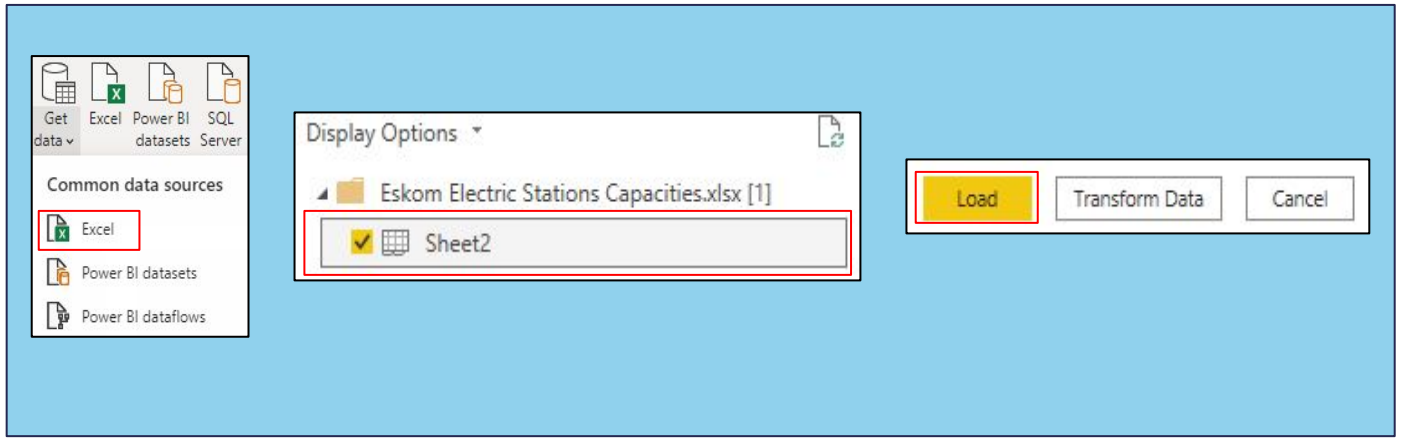
Sourcing Additional data and Setting Relationships

When opening the 'broken' dashboard, you will note that the '*Capacity and Efficiency*' visuals are not working. This is because the datasets for these visuals and their corresponding relationships are not set up. To properly source the data and set up the relationships, we will be following the steps below:

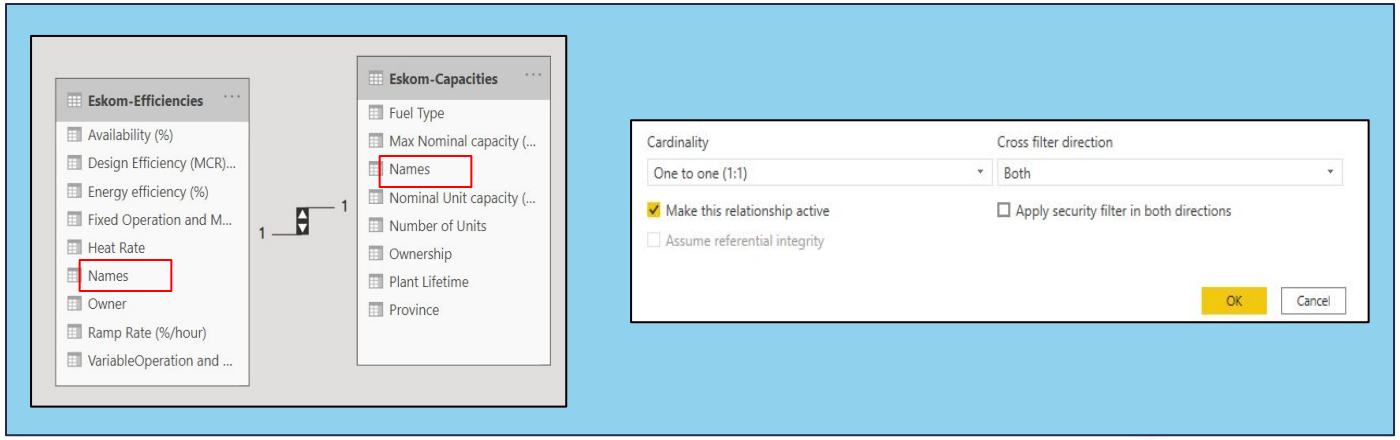
Steps for connecting to the Eskom Capacities and Efficiencies Datasets:

1. Use the Power BI '*Get Data*' Excel connector and connect to the *Eskom Electric Stations Capacities.xlsx* dataset
2. Select '*Sheet2*'
3. Click on *'Load'*
4. Repeat steps 1-3 for the *Eskom power stations efficiencies.xlsx* dataset
5. Rename the *Efficiencies* dataset to '*Eskom-Efficiencies*' and the *Capacities* dataset to '*Eskom-Capacities*'
6. Create a 1-1 relationship between the *'Names'* column of the *Eskom-Capacities* and *Eskom-Efficiencies* datasets

Source Data



Create Relationship



EXPLORE DIGITAL SKILLS

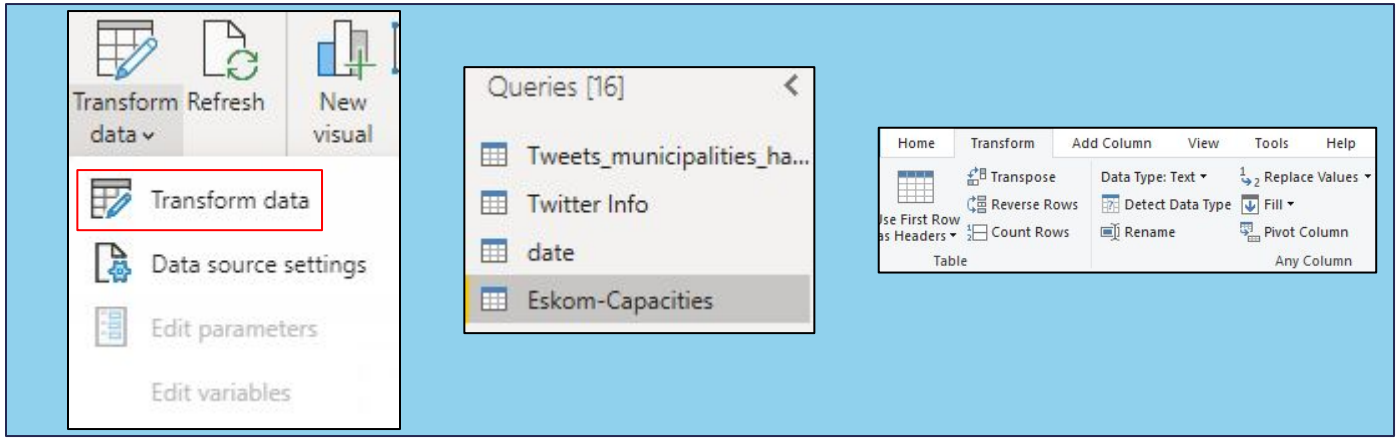
Sourcing Additional data and Setting Relationships

With the Eskom Efficiencies and Capacities datasets sourced and the relationships created, the next steps are concerned with cleaning the data to get it in a workable format. To do this, we are going to replace the empty and null value columns with the column average.

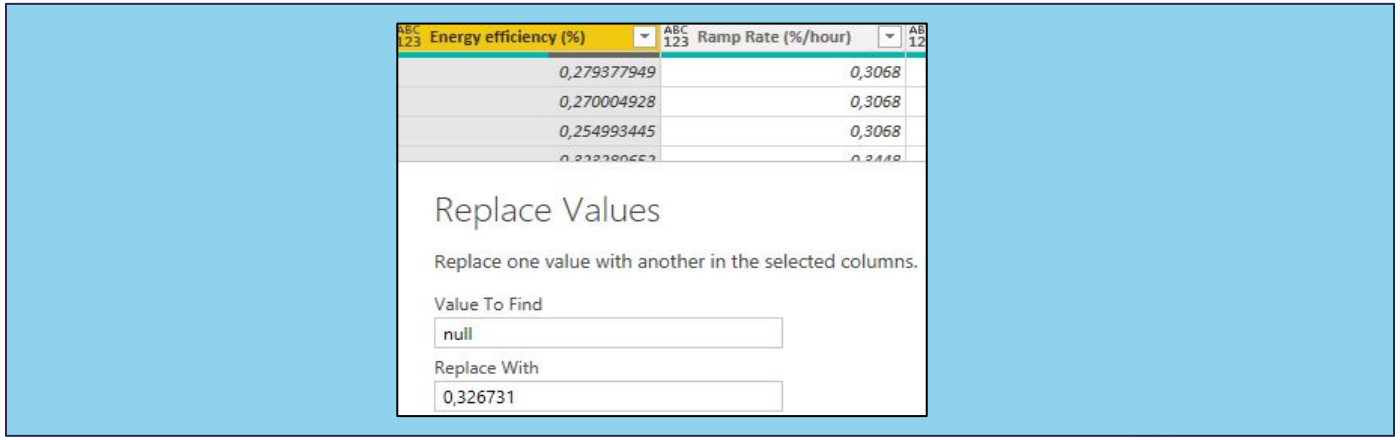
Steps for cleaning the Capacities Dataset

1. Access the Power Query Editor via the '*Transform Data*' button and navigate to the Eskom-Capacities dataset (you will notice a lot of columns with 'null' entries)
2. Remove 'Steam Capacity' and 'Column10'
3. For the 'max nominal capacity (MW)' column - fill the missing value with 1352
4. Select the 'Plant Lifetime' column and replace 'null' values with 50
5. For the 'Nominal Unit capacity (MW)' column - fill the missing values with 510
6. For the 'Number of Units column' - fill the missing values with 5

Access Datasets and Data Functions



Replace 'null' values



EXPLORE DIGITAL SKILLS

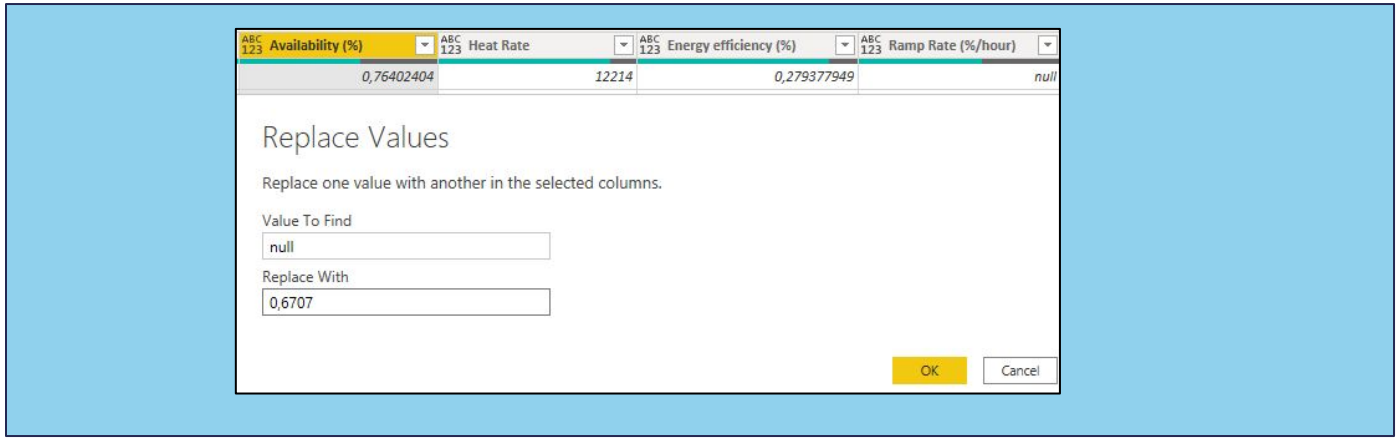
Sourcing Additional data and Setting Relationships

With the Eskom Efficiencies and Capacities datasets now sourced and the relationships created, the next steps are concerned with cleaning the data to get it in a workable format. To do this, we are going to replace the empty or null value columns with the column average

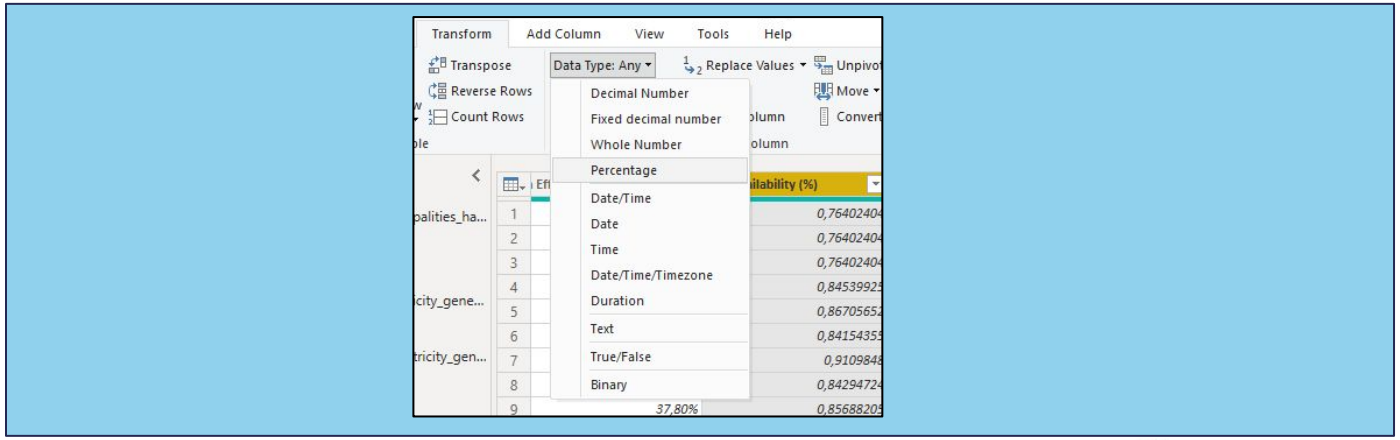
Steps for cleaning the Efficiencies Dataset

1. Remove 'Column10'
2. For 'Fixed Operation and Maintenance Cost' - fill the missing values with 188
3. For 'Variable Operation and Maintenance Cost' - fill the missing values with 45
4. For 'Ramp Rate per hour' - fill the missing values with .3068
5. For 'Energy Efficiency' - fill the missing values with .326731
6. For 'heat rate' - fill the missing values with 10589.22
7. For 'availability' - fill the missing values with .6707 (approx. average)
8. For 'design efficiency' - fill the missing values with .350889
9. Set the data types for all columns i.e. %, whole number, etc.
10. Remove the bottom 67 rows and the top 1 row

Replace 'null' values



Set the Data Type



EXPLORE DIGITAL SKILLS

Agenda

01 Problem Introduction

02 Connecting to the Data

03 Sourcing Additional Data and Setting Relationships

04 **Creating Visuals**

Infrastructure Dashboard

The infrastructure dashboard gives us insight into attributes of each station. Using Power BI slicers and filters we can draw correlations between station attributes, and explore and analyse this dataset in-depth.

Using what you've learnt up to this point, you are tasked to restore the supplied 'broken' infrastructure dashboard. You may use the below standing as guidance in your quest to rebuild the dashboard

Slicers

- Add a location slicer
- Add a station slicer
- Add a station type slicer

Cards

- Add an average electricity generated in MW card
- Add a total number of stations card
- Add a summary of stations card. Hint: Count of station_name feature and the status feature might be helpful

Table

- Add the features 'station_name' and 'count of id' to the table
- Add the features 'nominal_capacity_MW' and 'province' to the table
- Add the features 'operator_name' and 'installed_capacity' to the table

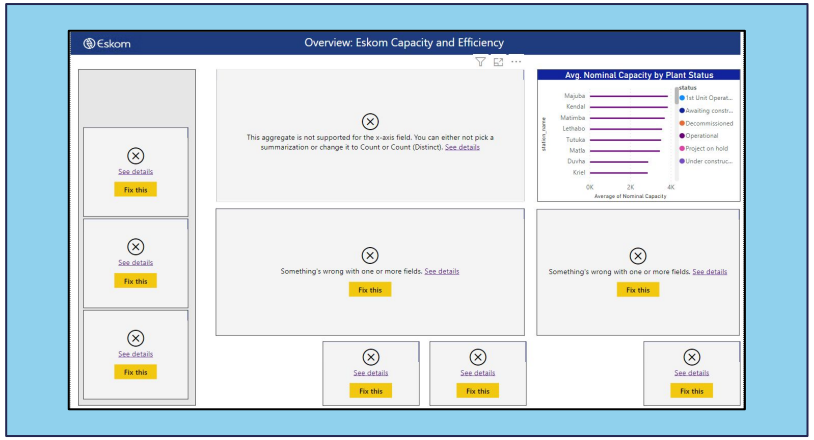
Visuals

- Add the feature 'Avg. of nominal_capacity_MW' to the column values field of the line and stacked column chart
- Add the feature 'commissioned_year_end ' to the line values field of the line and stacked column chart
- Add the feature 'commission_year_start' to the tooltips field of the line and stacked column chart

Capacity and Efficiency Dashboard

The Eskom capacity and efficiency dashboard is aimed at providing information on individual station performance.

Before data is sourced, cleaned and relationships are created



After data is sourced, cleaned and relationships are created

Guidelines for fixing the Eskom Capacity and Efficiency dashboard

Slicers

Add a 'number of units' slicer to the dashboard. Set the slicer type as between

Cards

Add the following cards to the report:

- Avg. of Ramp Rate (%/hr)
- Avg. of variable operations and maintenance costs

Visuals

- For the scatter chart, add the feature 'Max nominal capacity (MW)' to the size column
- For the clustered bar chart add the feature 'status' to the visualisations legend field

EXPLORE | DIGITAL SKILLS

Twitter Dashboard

The twitter dashboard summarises Eskom mentions (@) and tags (#) over a defined period

Municipality slicer

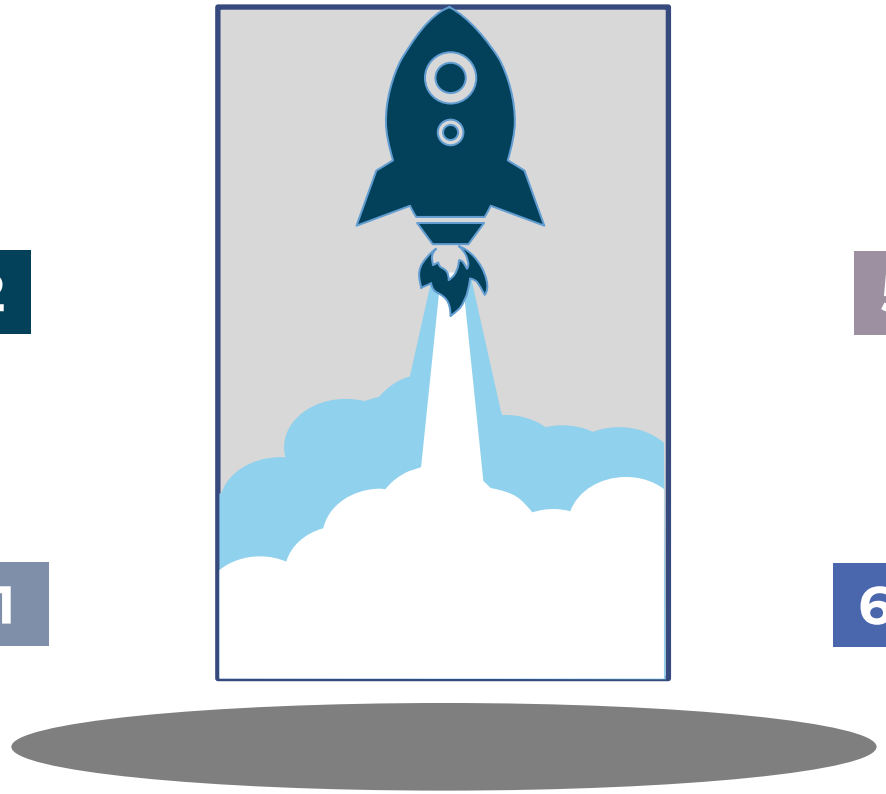
3

Add a municipality slicer of type ‘dropdown’ to the dashboard

Hashtags slicer

4

Add a hashtags slicer of type ‘dropdown’ to the dashboard



Date slicer

2

Add a date slicer of type ‘between’ to the dashboard

Multirow card

5

Add the feature ‘count of hashtags’ to the multi-row card visual

Import visuals

1

If you have not yet done so – click on the ‘Get more visuals’ button and download the WordCloud 2.0 visual and the Scroller visual

Addition of Tweets feature

6

To the table add the feature ‘Tweets’. Insert this feature in the table values column

EXPLORE | DIGITAL SKILLS

Predict-related FAQs

This page will be updated periodically with common predict-related questions which may arise during the Sprint. Consider consulting this space before asking your course facilitator a question.

EXPLORE | DIGITAL SKILLS