# Team 2: A Cable Sealing End Conundrum

Adedoyin Adebayo
Joshua Omolegan
Esinam Dake
Victor Gachoki

# Table of Contents

# A Cable Sealing End Conundrum: The Problem

National Grid require intervention as their assets are at risk of failure or unnecessary replacement. In order to prevent a potential loss of £12 million over the next 5 years, our team was assigned to evaluate this problem.

# A Cable Sealing End Conundrum: Initial thoughts

First, we explored the effects of the issue in order to gather a greater understanding. National Grid had to restrict contacts hours therefore reducing productivity due to the risks when dealing with Cable Sealing Ends (CSEs).

The issue concerns the safety of workers, and produces a sustained cash loss. Our goal was to find a solution which did not result in the scrapping and replacement of all assets as this would pacify the issue rather than solve it.

# Missing and Incorrect Data

Over a third of the data in the year sampled column was incorrect. To counteract this, we chose to replace the incorrect values with the median.

With a sample size this small, the amount of missing data would have had a catastrophic influence on the results of our analysis. To prevent this, we deployed a number of techniques to limit the number of empty fields:

- We set the value of the missing damage flags to 0.
- For majority of the variables (i.e Hydrogen to Breakdown_voltage), we were able to use the average of the given data to fill in the blanks.
- We calculated age through the use of the 'Year_Installed' and the 'Year_Sampled'.

Summation of missing data

| | |
|---|---|
| Damage_flag | 14 |
| Hydrogen | 7 |
| Oxygen | 7 |
| Nitrogen | 7 |
| Carbon_dioxide | 7 |
| Carbon_monoxide | 7 |
| Methane | 1 |
| Ethylene | 1 |
| Ethane | 1 |
| Acetylene | 1 |
| Moisture | 1 |
| Breakdown_voltage | 3 |
| age | 34 |
| Year_Installed | 1 |

# Omitted Columns

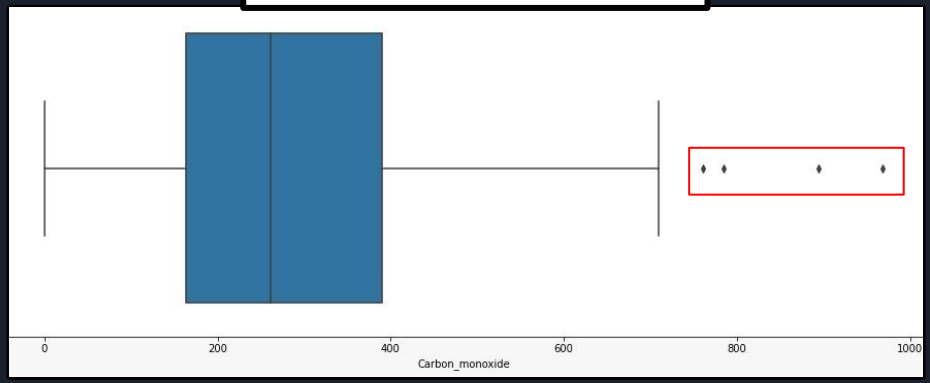
The data provided included many variables that we believed insignificant to the issue. On our first round of omittance, we removed:

- The year sampled and year installed columns as this information was already seen within the age column
- The ID column as this was just an indexing column and wasn't relevant for any of our predictions
- The asset, phase and type columns were categorical variables so we created new columns where each category in these columns was assigned a number. We then removed the original columns and did all our predictions using the new ones.

# Outliers

- We felt that due to the size of our data sample, it would be best to remove outliers completely as they would have a significant effect on our results.
- We chose to calculate whether or not a point was an outlier by checking if the data in any of its columns was more than 3 standard deviations away from the mean of that column.
- We ended up having 15 outliers, all of which were removed from the dataset

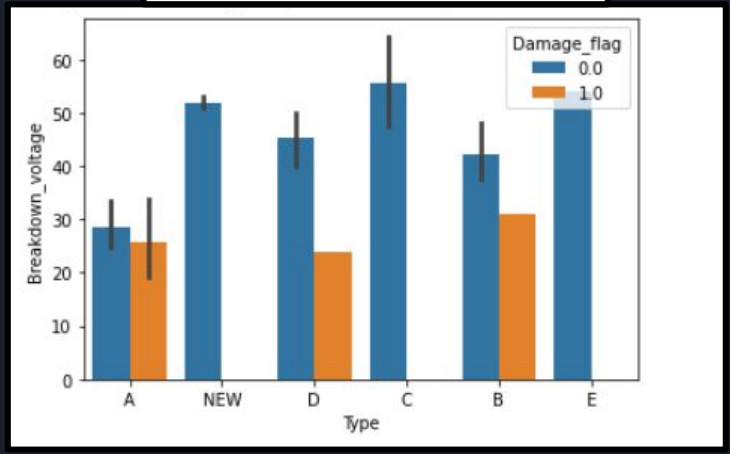
Identifying outliers



# The Effect Of Variables On Type

- Type A far more susceptible to damage
- Model C & E are seemingly the  best options however more data needs to be conducted to confirm.

The relationship between 'Type', 'Breakdown_voltage' and 'Damage_flag'



# The Assets

As we forecasted, there are many assets that do not require replacement as of now.

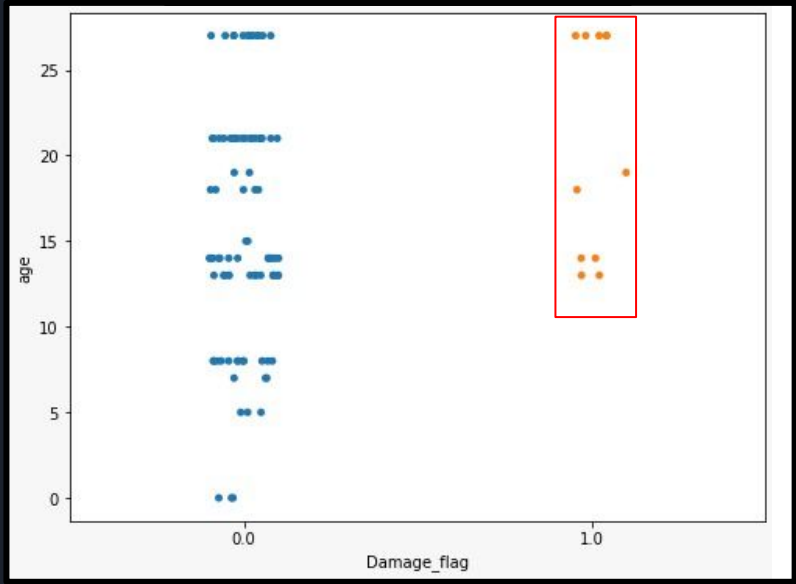
There are general rules that will benefit the preservation of assets:

- Any asset with a/an:
  - breakdown voltage of less than or equal to 25 should be replaced.
  - damage flag of 1 should be checked to see if it is repairable. If not, it should be replaced.
  - Any asset with a moisture level above 75 should be checked for damages.
  - Asset tag of “ALPHA” requires immediate replacement.

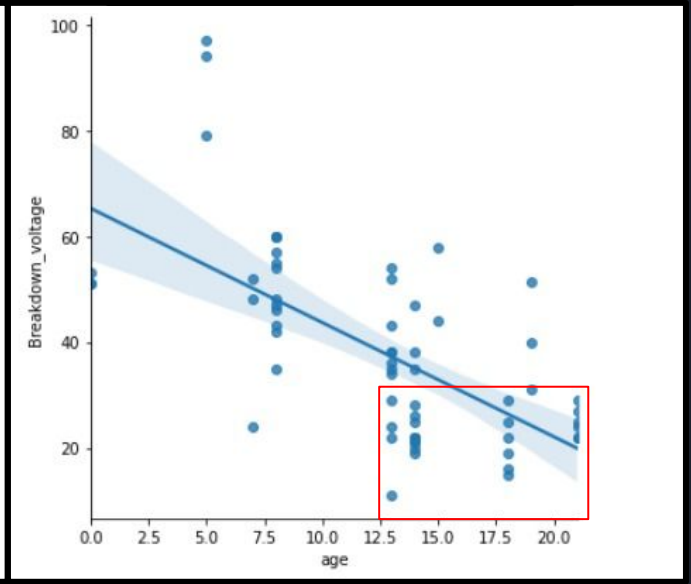
# The Assets

We found age to have  a strong correlation with both the damage flag and the breakdown voltage. We estimated ~12.5 years to be the age at which the assets enter the ‘critical zone’ in which a check up is essential to prevent a loss of the entire asset.

The relationship between damage flag and age

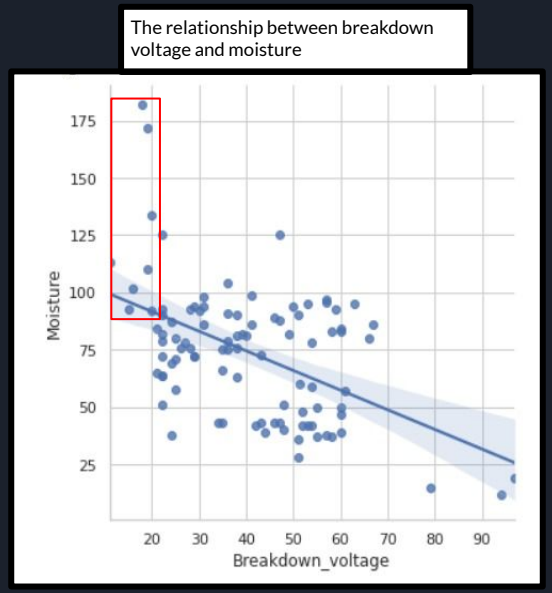



The relationship between age and breakdown voltage



# The Assets

- From our visualizations, we found moisture playing a substantial role in the breakdown voltage. The higher the moisture, the lower the breakdown voltage. Ideally the moisture should be kept below 100 on estimate to secure the breakdown voltage.



# Future data collection

National Grid should ensure that the engineers study further into what causes moisture to form and whether it is solvable  as this variable has the strongest correlation with the deterioration of the asset. With this, the assets are far more likely to increase in longevity.

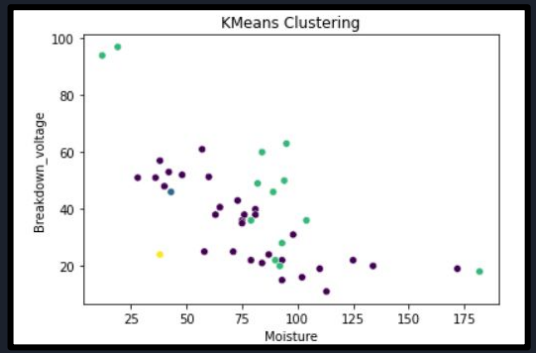
The data provided to the data scientist could be far better optimised with required second checks and the negation of irrelevant factors.

# Conclusive research

We initially used KMeans Assumption and DBSCAN, to create models to validate our assumptions.  We explored Kmeans to help us confirm our initial thoughts as we thought using a cluster model would make it easier to visualise and draw our conclusions.

We thought it would be best for data to be categorised as a false positive rather than a false negative as it is far more dangerous if the asset is accessed to be safe when it is not. Even with this bias, there are many assets that do not require replacement.

National Grid do not need to replace all of their assets, instead, more research should be done on the effects of ageing, moisture levels and the breakdown voltage.



# Q&A

We will now open the floor to questions

# Thank you for listening to our presentation