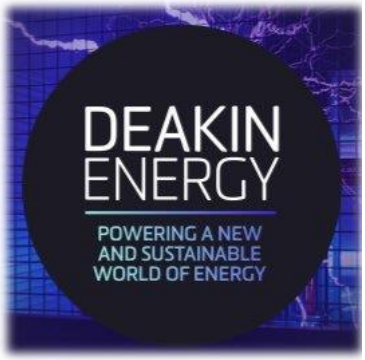
DATA 2 INTELLIGENCE CONSULTING: DEAKIN ENERGY



CLIENT REPORT

ELECTRICAL NETWORK RELATED DATA ANALYSIS

TRIBE LEADER
DR. THANH THI NGUYEN

ACADEMIC MENTORS
JEMAL ABAWAJY
SUNIL ARYAL

Table of Contents

1. Project Overview

Energy Safe Victoria is an independent government agency responsible for the safe generation, supply and use of electricity, gas, and pipelines. ESV works together with Deakin Energy to analyse their data, build models, predictions, and recommendations

The current project aims for Semester 1, 2021 are:

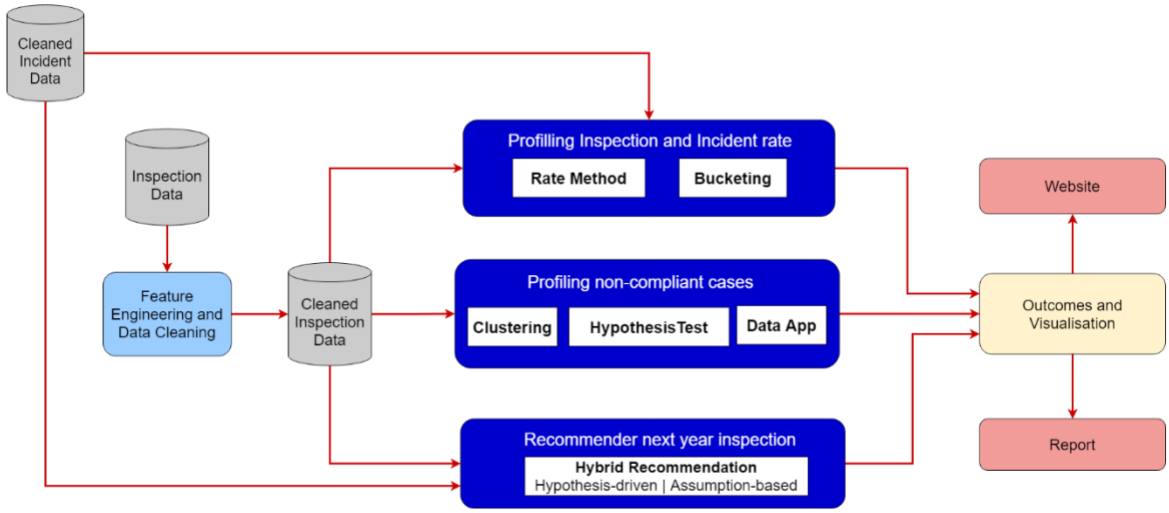
1. To figure out if there is any place that has high vegetation contact related incident rate but a low inspection rate or vice versa
2. To profile non-compliant cases and to check if there are any trends or patterns
3. To build a recommender for locations of next year inspections

2. Document Purpose

To report the outcomes of our experiments regarding addressing the project requirements.

3. Methods

Here is our product pipeline:



3.1. EDA

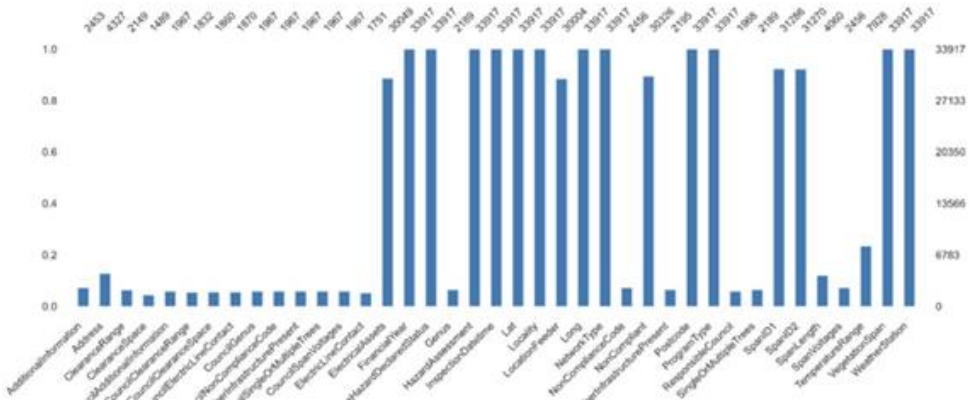
- Performed EDA using Python libraries, generated detailed report on the variable types, missing values, outliers and statistical measures.
- Understood the data dictionary/datatypes and values by meeting with the tribe leader
- Cleaned the unsupported variables and generated fresh reports

Dataset: span_inspections.csv

Overview of dataset:

Number of variables	39
Number of observations	33917
Missing cells	707850
Missing cells (%)	53.5%
Duplicate rows	69
Duplicate rows (%)	0.2%
Total size in memory	10.1 MiB

Data Instances:



Data Cleaning and Feature extraction

- Removing duplicates records (68 rows)
- Combining columns that serve similar purposes such as CouncilClearanceRange and ClearanceRange columns
- Filling out some missing values

Data Visualisations

- Visualise the data based on the location, weather station and vegetation span.
- This visualisation helps us to identify the inspection areas based on the weather station and financial year.
- We used Tableau to visualise data with filters (Weather station, Financial year and Vegetation span) and integrated it into the web application.

3.2 Identify places that have High Vegetation Inspection Rate and Low Incident Rate and vice versa

We pre-processed the Incident Dataset to find the number of vegetation related incidents then we find the relationship between Incidents and Inspections. The data for both datasets covers the period 2017 to 2021.

We applied two methods:

- *Bucketing Method*: We use the median to group Incidents Inspections by their total cases
- *Rate Analysis*: We applied mathematics calculation to determine the rate (total incidents/total inspections).

Based on both methods, we selected the top 10 postcodes that have High incidents low inspections and the top 10 postcodes that have Low incidents high inspections

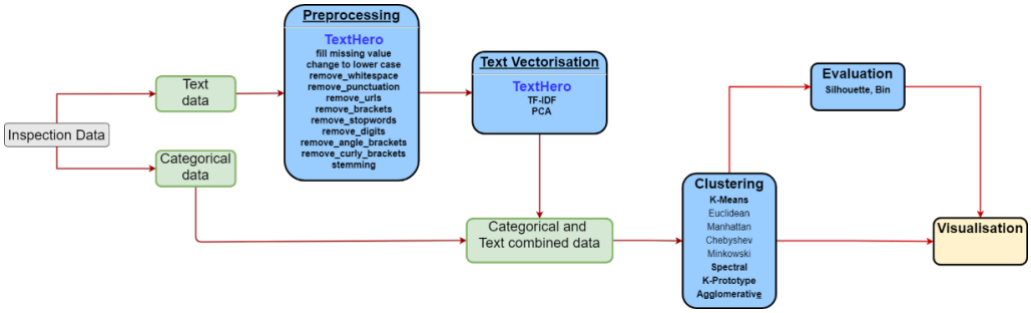
3.3 Profiling Non-Compliant Cases

For this task, we applied three methods:

- *Clustering and Evaluation*

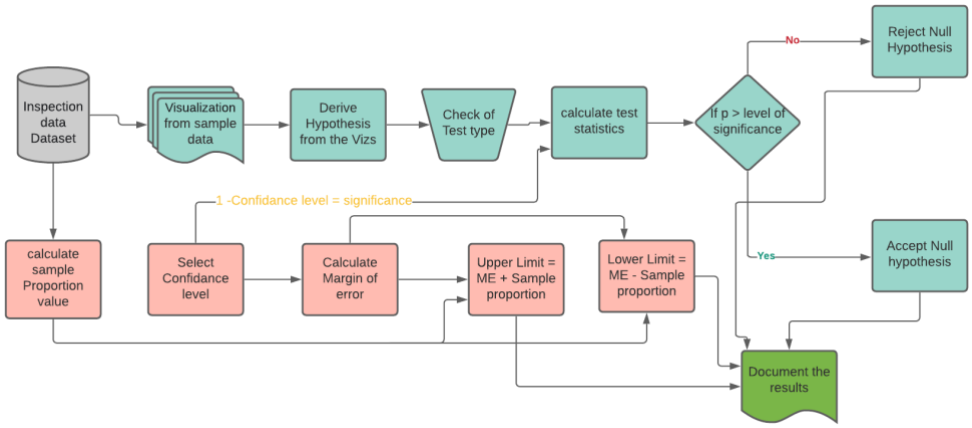
  After experimenting and evaluating different clustering techniques, the best number of clusters is 6. From this result, we profile the non-compliance cases.

  Here is the non-compliant pipeline:



- *Statistical hypothesis testing*

  Following is the hypothesis test model design

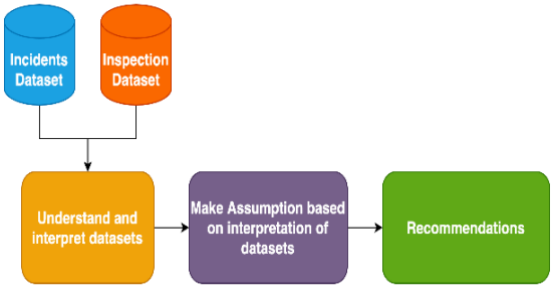


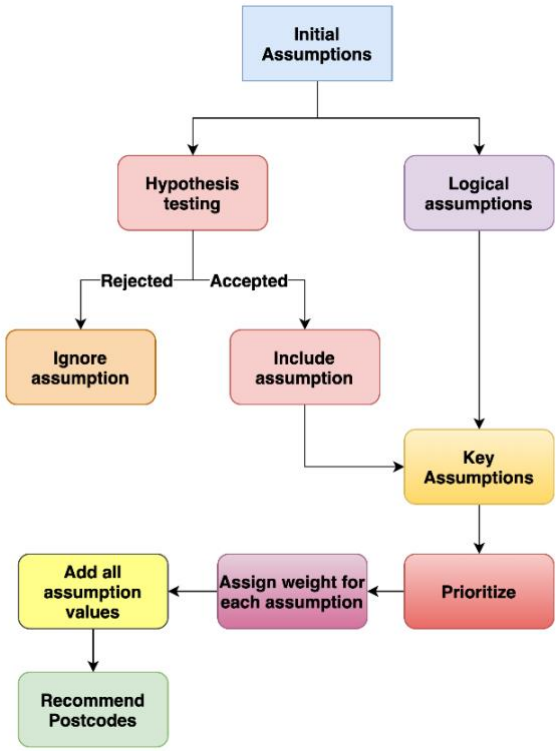
- *Data Analysis Tool*

  Development of data application which can be used as data analysis tool using filters.

3.4 Recommender locations of next year inspection.

The following pipeline demonstrates the analytic process to generate the Recommender





3.5 Web Application

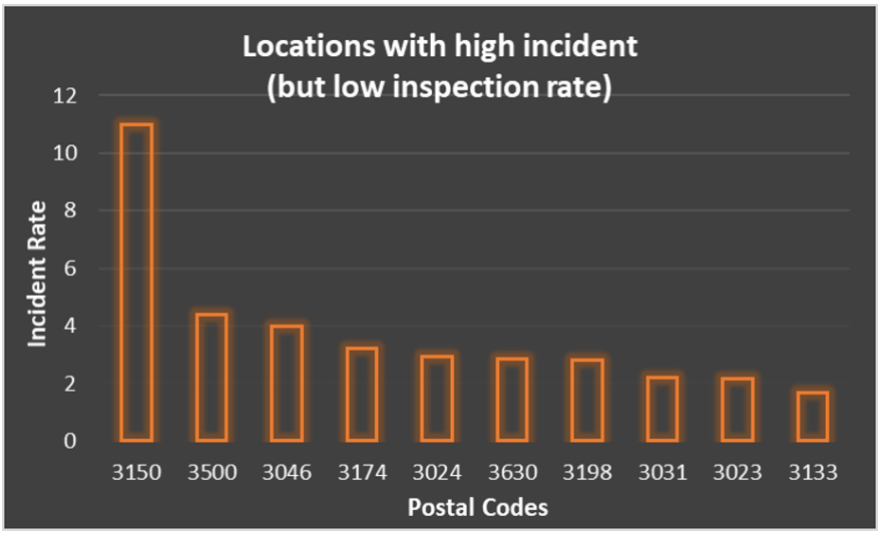
We improved the UI design, added a landing page, added content and web application testing

We also included interactive maps to display the Vegetation Incidents and Inspections and Recommender for next year inspection. We developed the maps using Plotly Chloropleth.

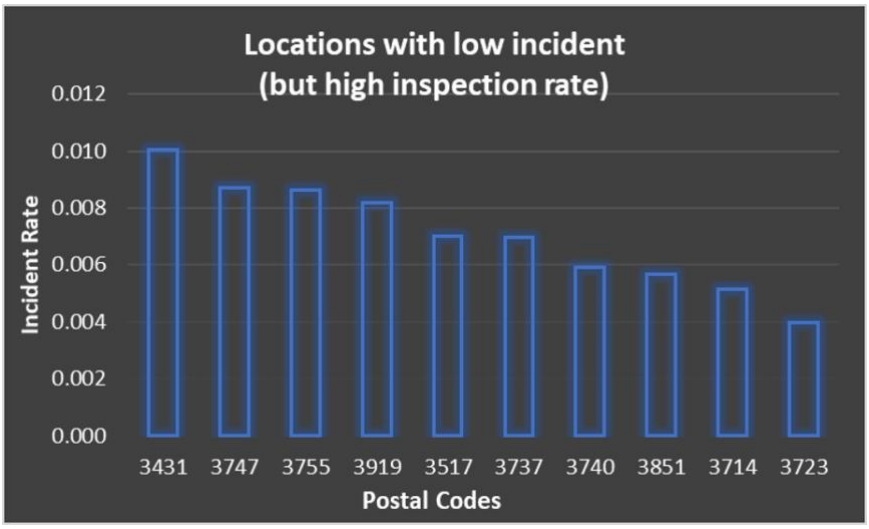
4. OUTCOMES

4.1 Vegetation related incidents and inspections

Top 10 postcodes with high vegetation related incidents



Top 10 postcodes with low Vegetation related incidents but high inspections



4.2 Profiling non-compliances result

Hypothesis test results for non-compliances cases

The below findings are given within their individual interval (Upper and Lower Limit) which has been studied under 95% level of confidence.

Each of the value in the table represents the proportion of non-compliant cases in each subcategory and main category. For instance, in the main category of [Network Type], there exist non-compliant cases ranging between 10.41% to 11.70% of the ANS sub-type. This prediction is ascertained under 95% of confidence level.

Network Type									
ANS		CP		JEN		PC		UE	
Upper Limit	Lower Limit	Upper Limit	Lower Limit	Upper Limit	Lower Limit	Upper Limit	Lower Limit	Upper Limit	Lower Limit
11.70%	10.41%	72.77%	65.00%	24.89%	21.00%	9.89%	8.80%	18.26%	16.51%

Financial Year							
2017 - 2018			2018- 2019		2019 - 2020		
Upper Limit		Lower Limit	Upper Limit	Lower Limit	Upper Limit		Lower Limit
7.05%		5.60%	15.52%	14.33%	24.74%		23.32%

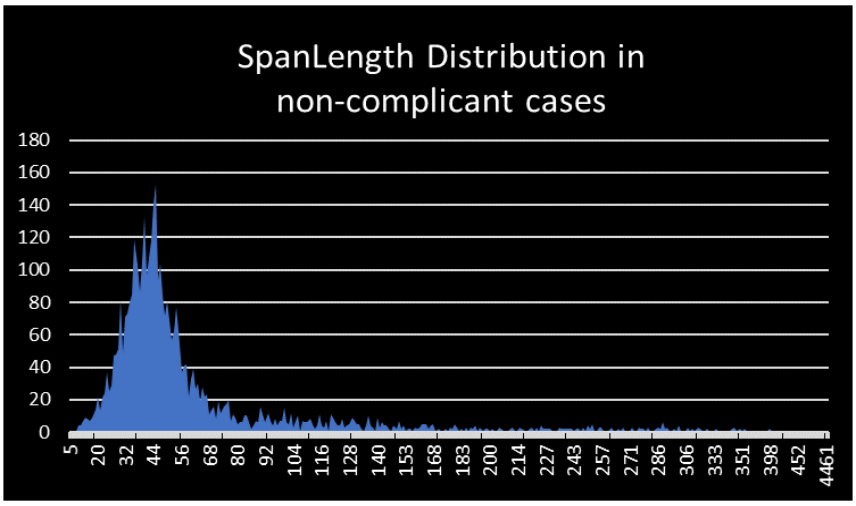
Inspection Type					
HBRA			LBRA		
Upper Limit		Lower Limit	Upper Limit		Lower Limit
6.14%		5.42%	24.75%		23.32%

Low Bush Fire Related Area					
Declared			Non Declared		
Upper Limit		Lower Limit	Upper Limit		Lower Limit
36.85%		34.53%	14.50%		12.90%

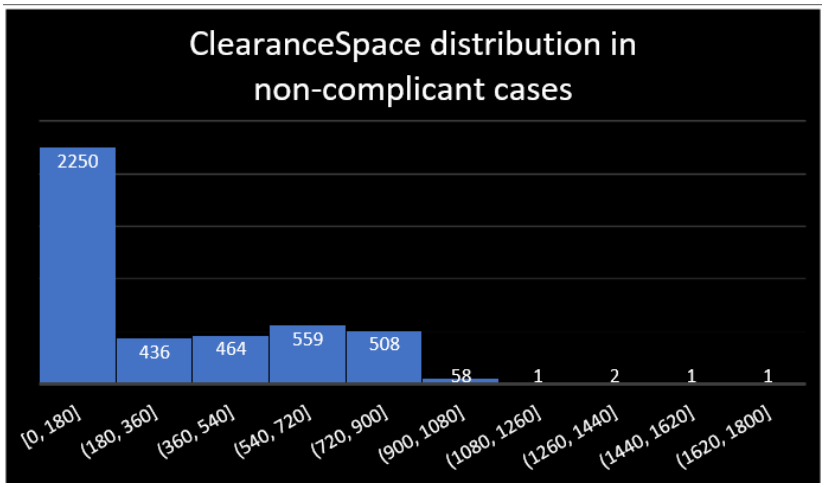
High Bush Fire Related Area					
Declared			Non Declared		
Upper Limit		Lower Limit	Upper Limit		Lower Limit
4.42.%		1.25%	6.18%		5.45%

Clustering results and analysis

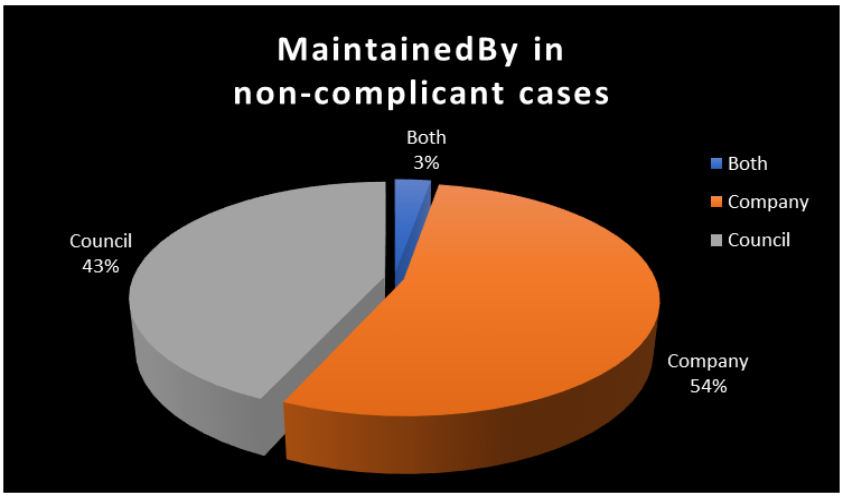
This section presents the findings of non-compliant patterns in relation to a collection of selective features including SpanLength, ClearanceSpace, MaintainedBy, ResponsibleCouncil, and SpanVoltages.



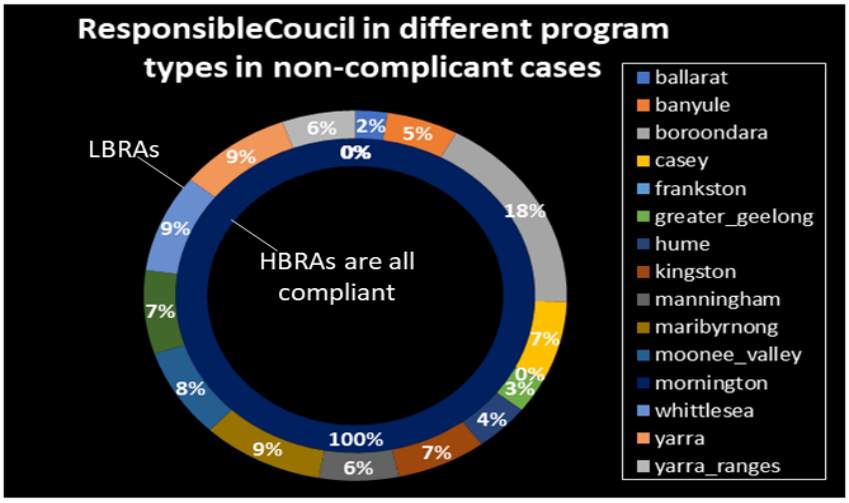
- This graph shows the span length distribution in non-compliant cases.
- The amount of non-compliant cases piled up in the areas where the spans were low.
- More and more non-compliant cases subsided as the span length expended.
- The peak of non-compliant cases extended between 30 to 50.



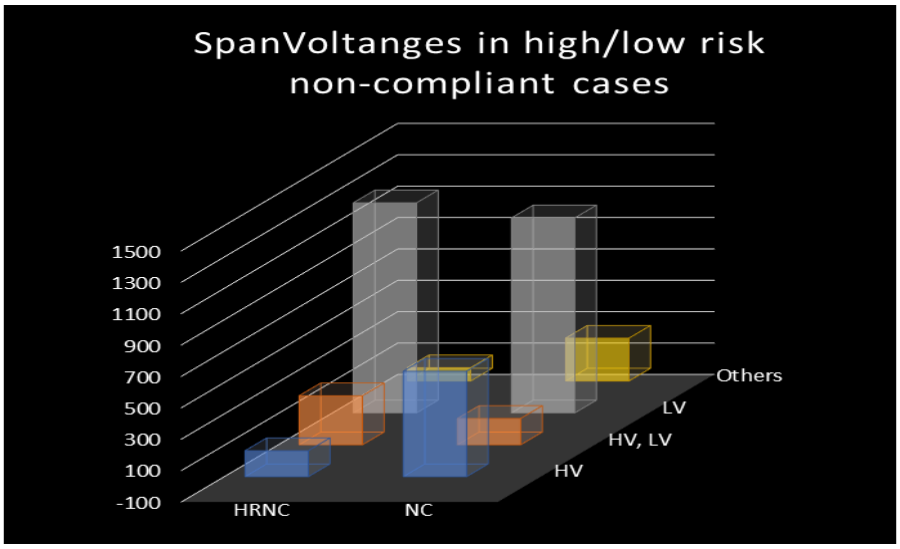
- This graph illustrates Clearance Space distribution in non-compliant cases.
- Evidentially the amount of non-compliant cases was highest at the range between 0-180.
- As the clearance space increased, the non-compliant cases began to drop and almost reached zero when the range was significantly higher i.e. from 1080-1800.



- This pie chart shows the proportional percentage of non-compliant cases, either maintained by the council, company, or both.
- Within an area which was maintained by both council and company, the percentage of the non-compliant case was significantly low i.e. 3%.
- The percentage became higher when the areas were maintained by the company or council individually.



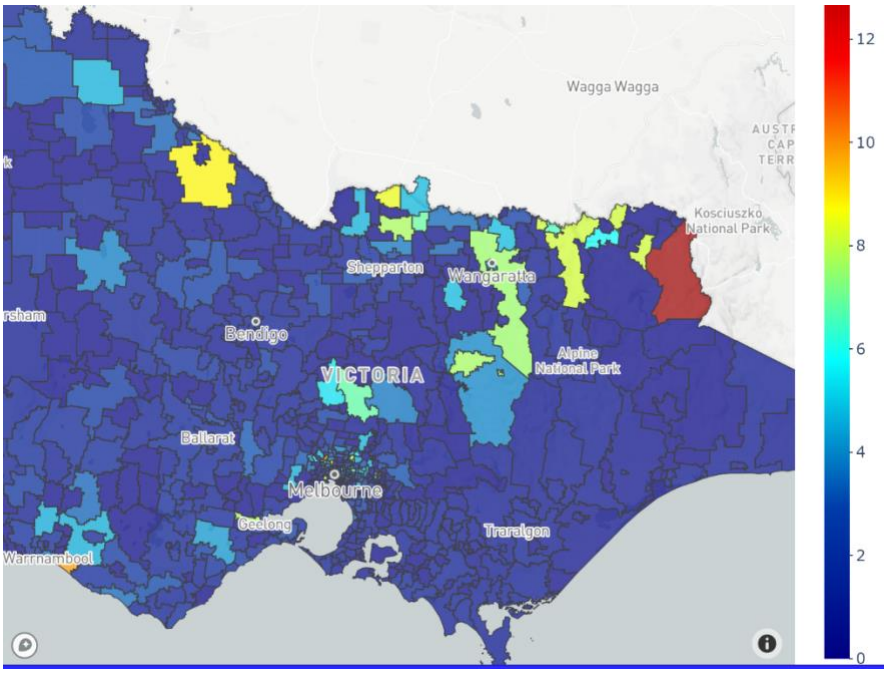
- This plot compares the percentage of non-compliant cases by responsible councils for two different programs i.e. LBRAs and HBRAs.
- The cases with HBRAs were 100% compliant by responsible councils.
- However, within the LBRAs program, 15 councils were found to be non-compliant.
- Boroondara was on the lead with up to 18% of non-compliant cases as compared to others.



- This chart demonstrates the amount of non-compliant cases within each type of span voltages which further classified into high risk (HRNC) and low risk (NC) scenario.
- As seen in the graph, low voltage [LV] locations shared the largest proportion of non-compliance, both in HRNC and NC cases.
- High voltage [HV] locations were responsible for the second-largest proportion in NC cases
- while locations that ran [HV, LV] similarly claimed the second rank in HRNC cases.
- The combination of 20 other types of locations (Others) did not seem to have any significant contribution to these non-compliant cases.

4.3 Recommended postcodes for next year inspections

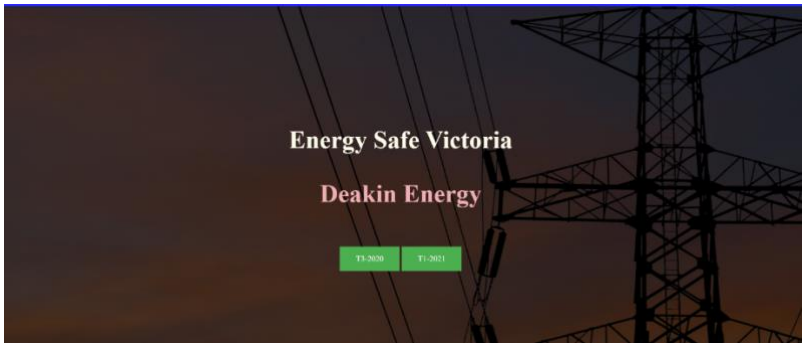
Each colour on the map indicates the score for that postcode. The higher the score, the more urgent the area needs inspections and the further up it is on the colour scale.



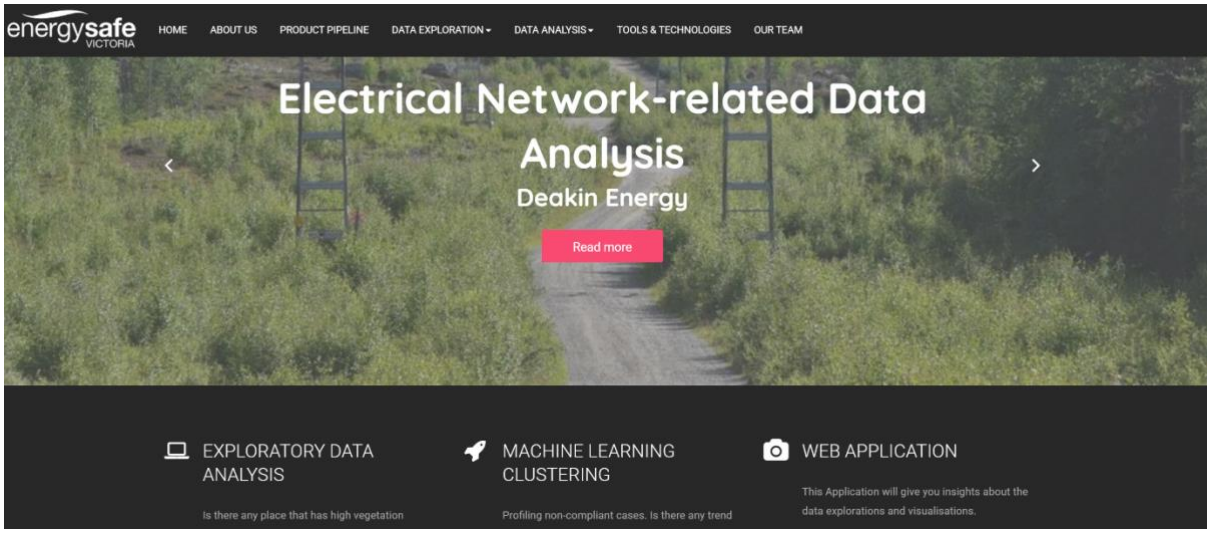
Postcode	Score
3765	14.52
3707	13.81
3043	9.93
3277	9.83
3044	9.19
3579	8.80
3641	8.50
3015	8.20
3691	8.17
3636	7.70

4.3 Web Application

We developed a landing page where users can select between the previous project (Incident website) or the current project (Inspection website)



The current website displays the methods and outcomes of experiments.



5. FURTHER STUDY

Further study is recommended to improve the recommendation for next year inspections, such as:

1. Evaluate the performance of recommendation.
2. Improve the recommendation of places for next year inspection
   - Identify most critical assumptions
   - Monitor current assumptions to understand if they are critical for the process
   - Experiment with various critical assumptions and priority of them.
3. Get more data and perform detailed analysis.