**Beyond the Charge: Decoding Patterns in Electric Vehicle Battery**

**-*Blog Writing on EDA***

**Introduction to EDA**

Exploratory Data Analysis (EDA) is a critical initial phase in the data analysis process. It allows analysts to better understand the data at hand - its properties, structure, and the stories hidden within it. Before diving into sophisticated statistical modeling or machine learning algorithms, EDA offers the opportunity to get up close with our data in an unbiased manner and without strict assumptions.

Through EDA techniques like calculating summary statistics, visualizing distributions, and probing relationships, we commence an open conversation with the data. Key insights start emerging about inherent trends, patterns and anomalies that can inform and optimize our future analysis. Running through exploratory exercises also allows us to check assumptions, spot issues with quality or completeness, and uncover nuances particular to that dataset.

Used Panda Packages for creating my synthetic dataset which EV Battery Health as the foundation. Together, we forged a robust sample of 1000 synthetic cases, carefully crafted with 9 features apiece. This fusion sparked a comprehensive exploration - transforming bland data into a vibrant landscape brimming with analytic potential. The synthesized cases of poor cell construction, thermal abuse, faulty assemblies, short circuits, temperature fluxes and more became the springboard to examine the interplay between maintenance, battery type, charging behaviors and environments. Segmenting by battery chemistry revealed insights into tailoring safer systems for both lithium-ion and lead-acid. This pivotal dataset was our guiding force to drive discoveries and make recommendations targeted to the nuances of each battery type’s performance.

**Feature Labels:**

1. **Continuous Features:**

1. Temperatures:

- Numeric indicating if temperature out of safe range

- Operation in extreme temps degrades battery over time

- Temperature monitoring is critical for safety

**II. Discrete Features:**

1. Poor cell design:

- Categorical feature indicating if there were flaws in how the cell was designed

- Could analyze if certain manufacturers have higher rates of flaws

- Flaws likely increase chance of blast, so important predictor

2. Poor battery assembly design:

- Categorical variable checking if the issue was due to battery assembly

- Could group by assembly manufacturers to check for patterns

- Important predictor since assembly issues can directly cause Chance of blasts

3. short circuits:

- Binary variable indicating if a short circuit triggered it

- Short circuits very frequently lead to blasts

- Important factor to capture

**III. Categorical Features:**

1. External abuse:

- Categorical variable capturing if there was external abuse like thermal, mechanical or electrical

- Would want to analyze which types of abuse are most common

- External issues explain some blasts so useful predictor

2. Overcharge or over-discharge:

- Categorical variable documenting if improper charging/discharging occurred

- Overcharging/discharging damages cells and very unsafe

- Key predictor of explosions

3. battery\_maintenance:

- Could be categorical variable indicating if proper maintenance occurred

- Lack of maintenance increases wear and unsafe conditions

- May require some feature engineering to quantify

4. battery\_health:

- Categorical variable classifying battery health

- Directly relevant to blast risk

- Key target variable to predict

5. Battery type:

- Categorical feature indicating either lead-acid or lithium-ion

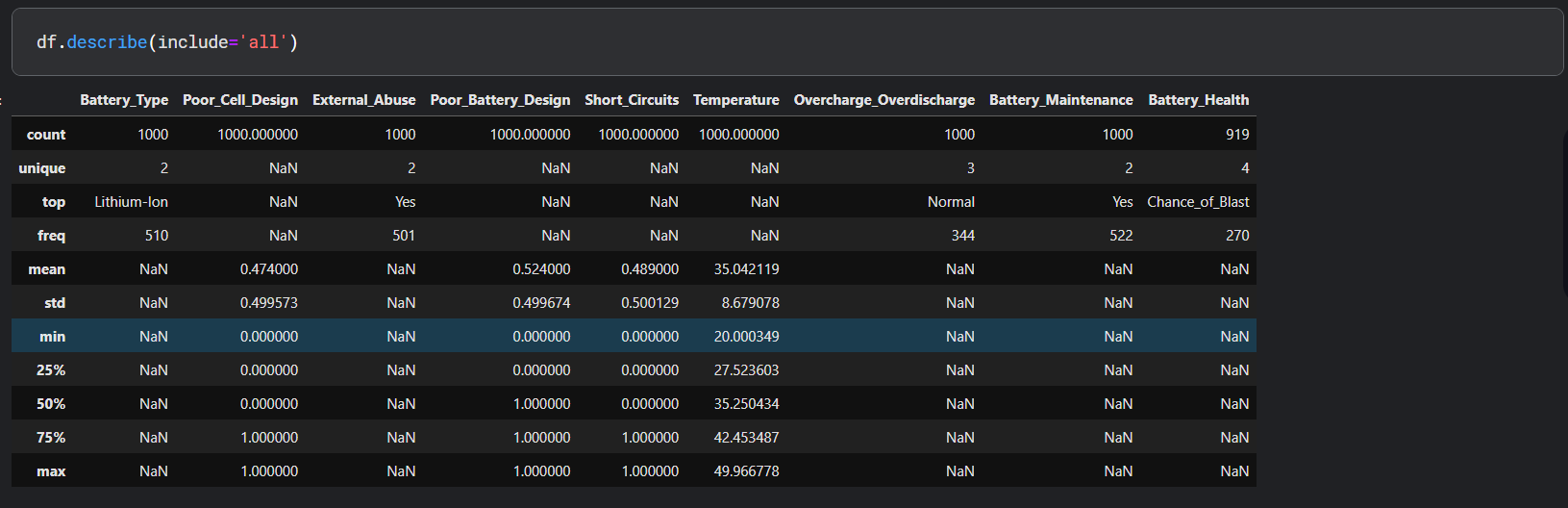
- Useful to analyze if certain battery types have higher failure rates

- Lithium-ion has become extremely common in EVs

- Segmenting analysis by battery type can reveal differences in their safety performance

**Descriptive Statistics**

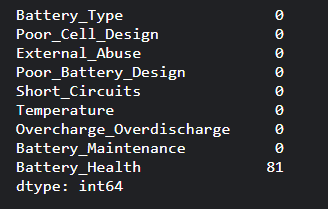
You can easily tell just by looking at the dataset that it contains data about different types of features related to EV Battery. Let us start by looking at descriptive statistic parameters for the dataset. We will use describe() for this.



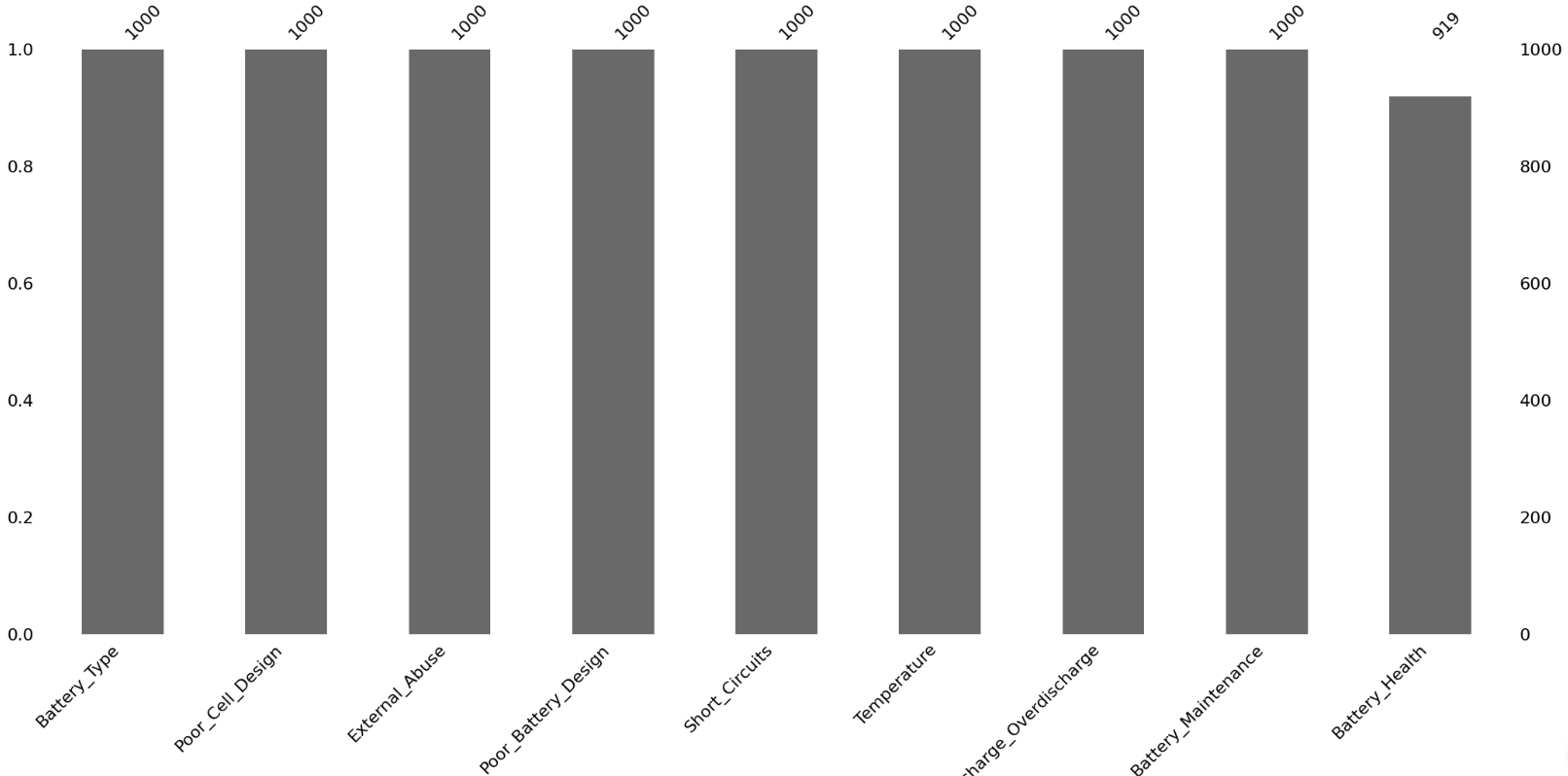
**Pre-Processing:**

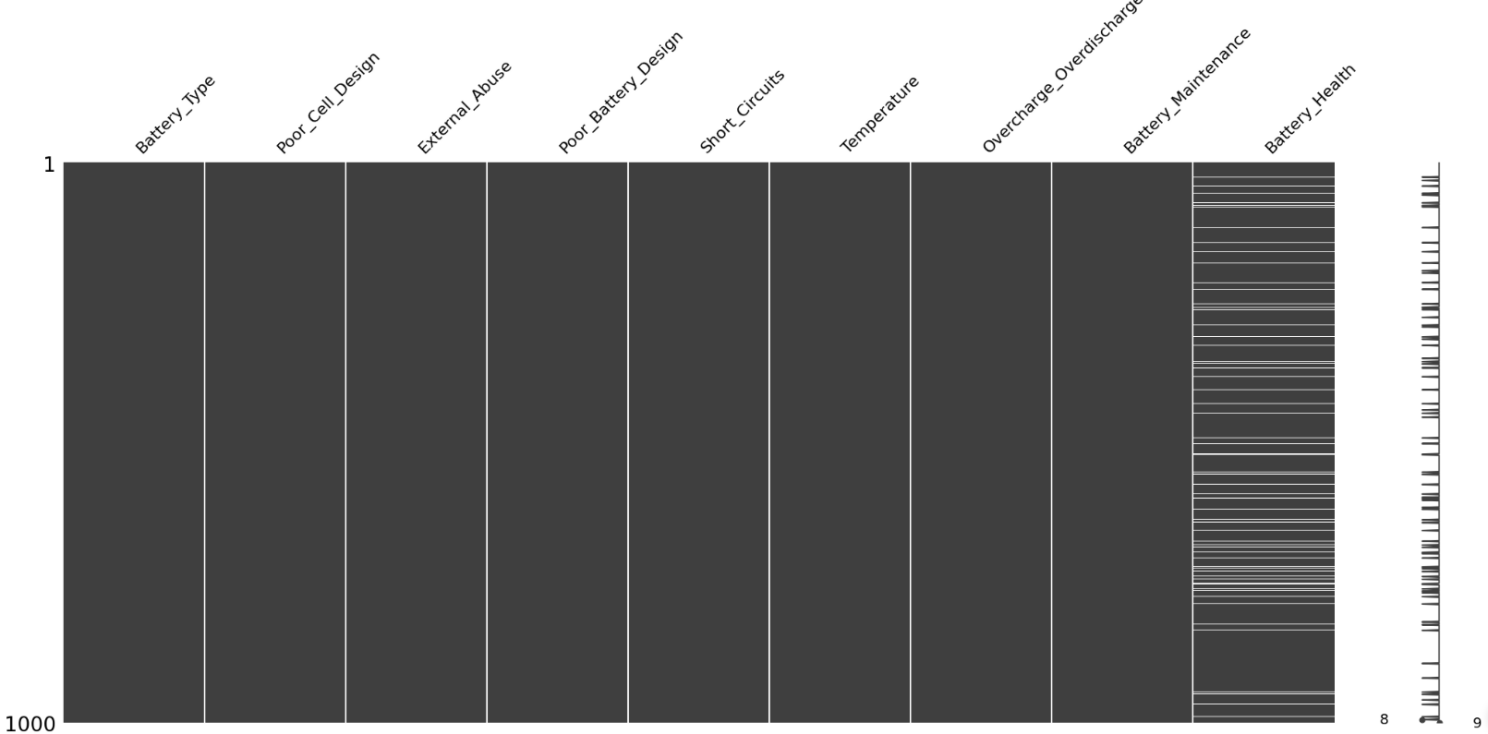
1. **Handling Missing Values:**

In case there are any missing entries,We will use the isnull() function for this purpose.



So,lets visualize the missing values in the dataset using msno.bar(df),msno.matrix(df)

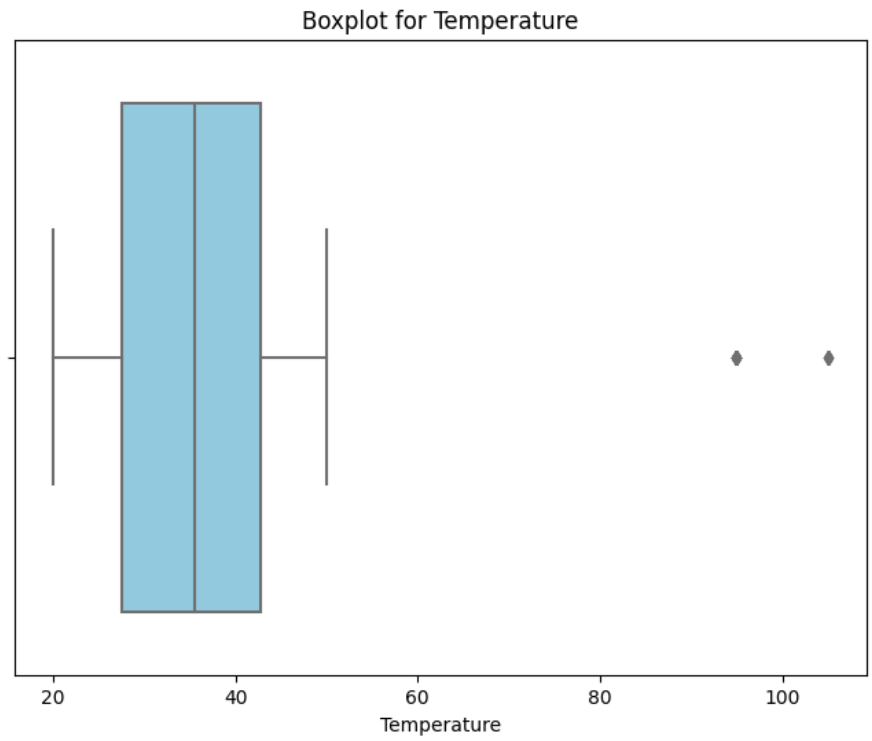




To properly handle the missing values in the EV blast dataset. Simply excluding those records could bias the analysis. By creating a new category "unknown" for features with null entries using code df["Battery\_Health"].fillna("Unknown", inplace = True).

1. **Handling Outliers:**

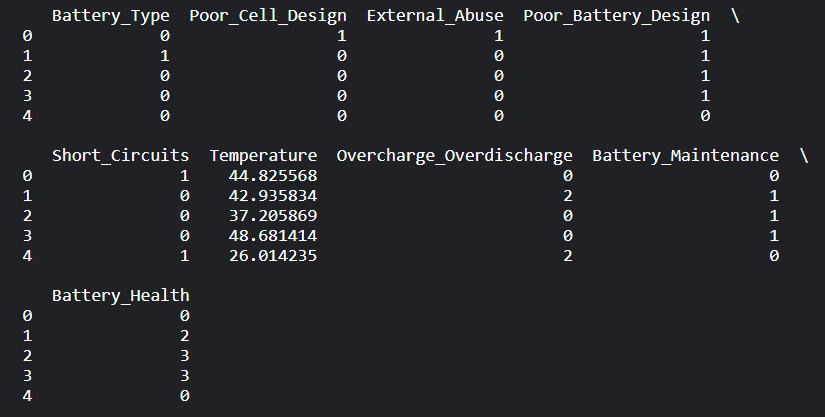
Outliers are data points that are unusually far outside the normal range expected for a variable. When analyzing the EV battery dataset using boxplot, revealed some potential outlier observations in temperature feature.



Removed all the observations using quartile.

1. **Encoding:**

Employed label encoding technique to transform categorical variables into a numerical format, facilitating comprehensive analysis like Battery Health,Battery Type,Battery Maintenance,External Abuse.



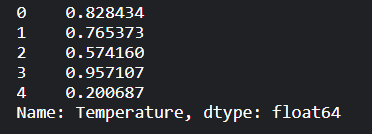
For Battery Health,0 - Blast,1-Chance\_of\_blast,2-Good,3-moderate,4-unkown

For OverCharge\_Overdischarge,0-Normal,1-Over\_discharge,2-Over\_charge

For Battery\_Type,0-lead\_acid,1-Lithium\_ion

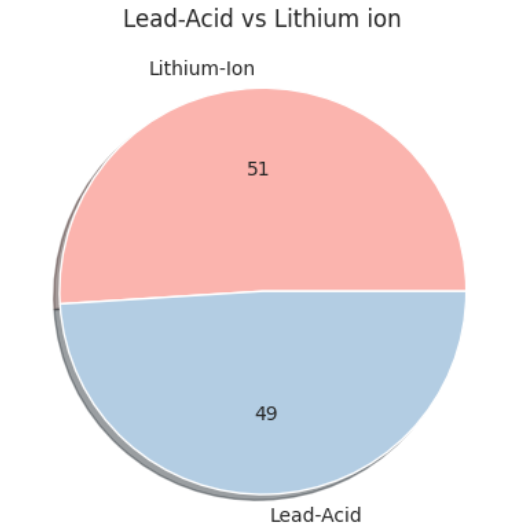
1. **Normalization:**

The EV battery dataset includes a number of continuous variables like temperature ranges that vary substantially in their ranges and distributions. To facilitate effective modeling, these features were normalized using a min-max scaler to transform the values to a common scale without distorting differences in the distributions.



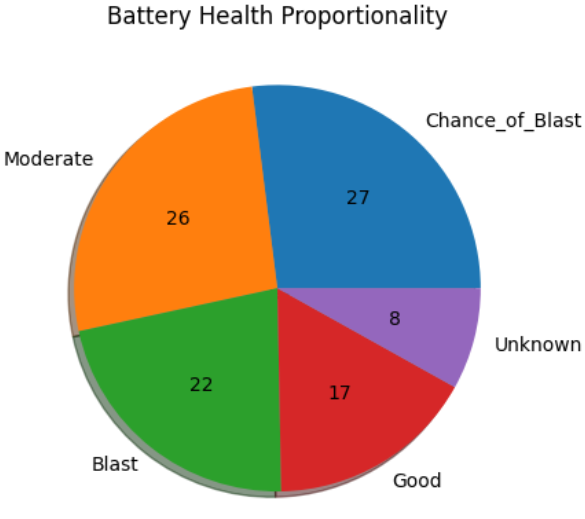
### Graphical Visualizations

1. Battery type Proportionality Using pie chart



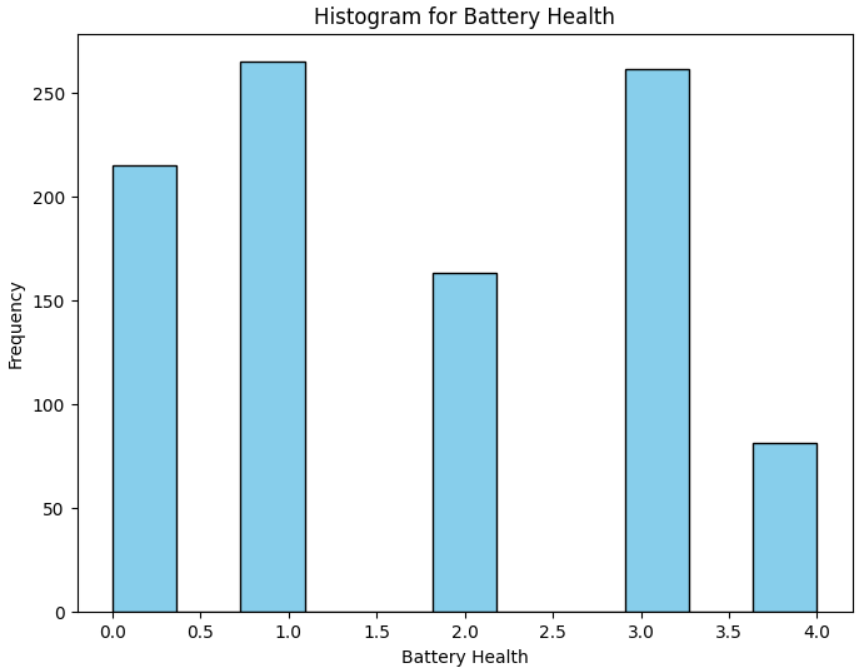
The resulting pie chart will show the proportionality of different battery types in dataset. Each wedge represents a battery type, and the size of the wedge indicates its proportion relative to the total number of batteries in the dataset.

1. Battery Health Proportionality Using Pie Chart



The resulting pie chart visually represents the proportionality of different battery health categories in dataset. Each wedge represents a health category, and the size of the wedge indicates its proportion relative to the total number of batteries in the dataset.

1. Histogram for Battery Health



The rectangular bars represent the number of data points that fall within a certain class interval. The frequency of the data that fall into these groups will then be determined after the data have been grouped based on this interval. Most batteries have a health rating of either 0.5 or 3.5, with fewer batteries having a health rating of 0.0(Blast) to 3.0(Moderate), and the least number of batteries having a health rating of 2.0(good)

1. Correlation Matrix of All Matrix

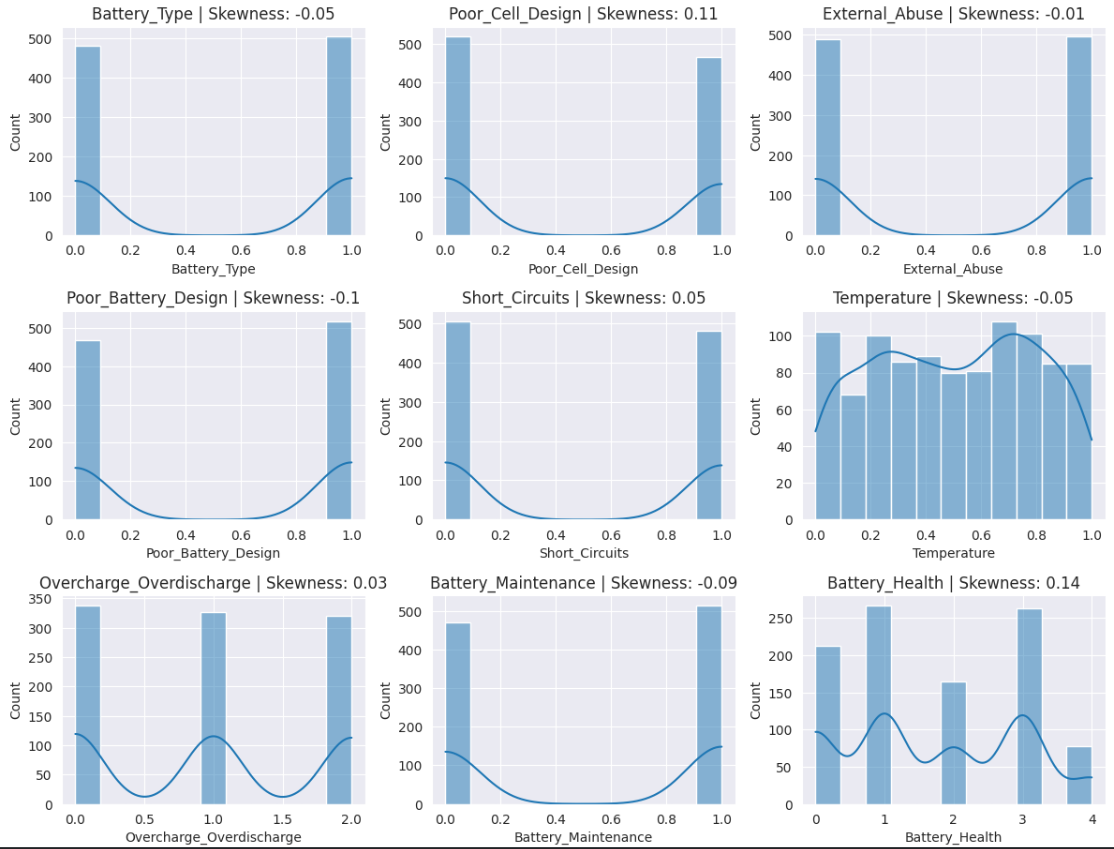


A correlation matrix and heatmap provide a concise and visual representation of the relationships within a dataset. Overall,this give us the relativity between the each and every features.

Battery maintenance and battery health are also highly positively correlated (0.89), which means that batteries that are well-maintained tend to have better health and performance.

Short circuits and temperature are moderately negatively correlated (-0.07), which means that batteries that experience short circuits tend to have lower temperatures than normal. This could be due to the loss of energy or the activation of safety mechanisms.

1. Histogram for all features

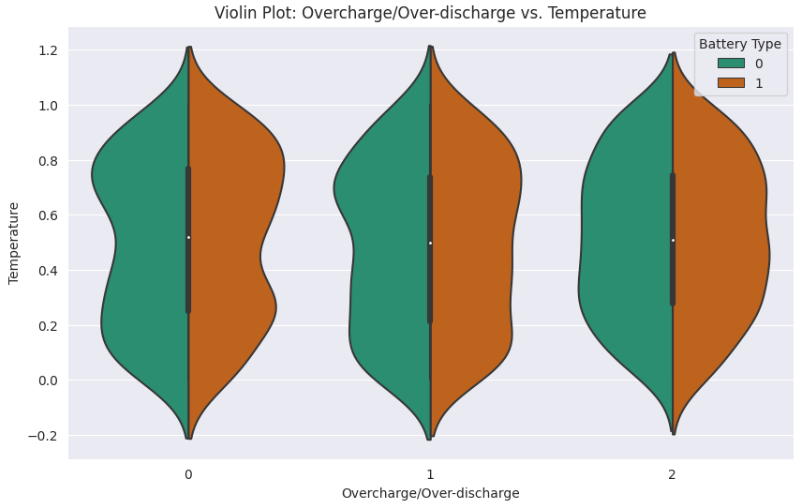


Here We Observe the general shape of each histogram to understand the overall distribution of features.Skewness is indicated in the title of each subplot. A skewness close to 0 suggests a symmetric distribution,while positive or negative skewness indicates skew to the right or left, respectively.

The poor cell design histogram shows that most batteries have a low score for poor cell design, meaning that they have good cell design. The data is slightly skewed to the right.

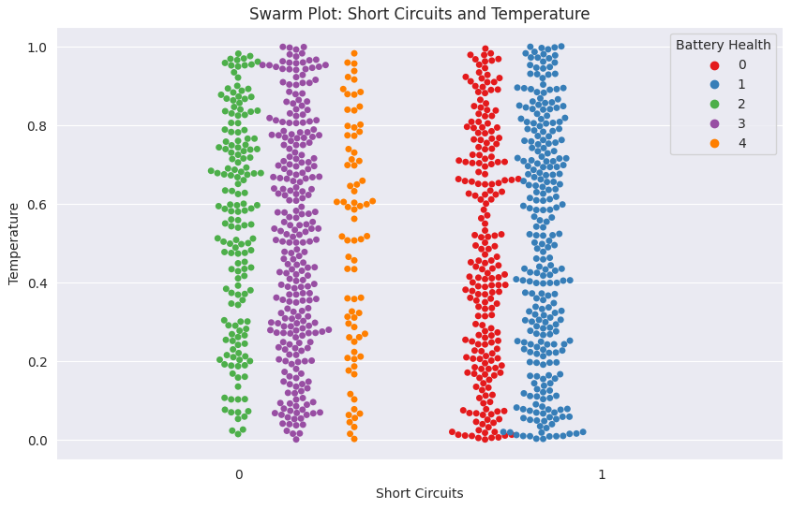
The battery type histogram shows that most batteries are of type 1, followed by type 2 and type 3. The data is slightly skewed to the left, indicating that type 1 batteries are more common than type 2 and type 3 batteries.

1. Violin Plot for Overcharge/Over-discharge vs. Temperature



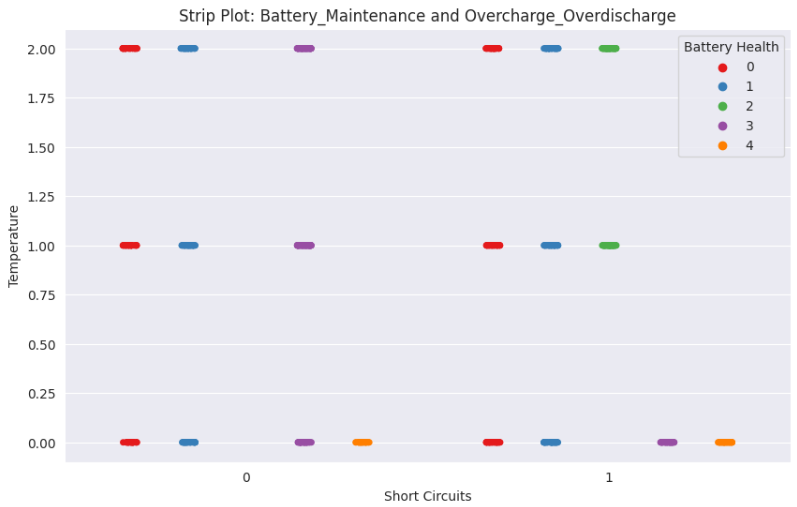
The graph providing insights into the distribution and density of Overcharge/Over-discharge vs. Temperature.The horizontal line inside each violin represents the median or mean temperature.Overlapping regions indicate temperature ranges where both categories have substantial representation, while non-overlapping regions suggest distinct temperature preferences or conditions.

1. Swarm Plot: Short Circuits and Temperature



We can observe that for batteries with zero no short circuits,They are mostly in good,Moderate health (purple and green).Their temperatures range broadly but are generally lower.As the number of short circuits increases from 0 to 1, there is a noticeable increase in temperature.

1. Strip Plot: Battery\_Maintenance and Overcharge\_Overdischarge

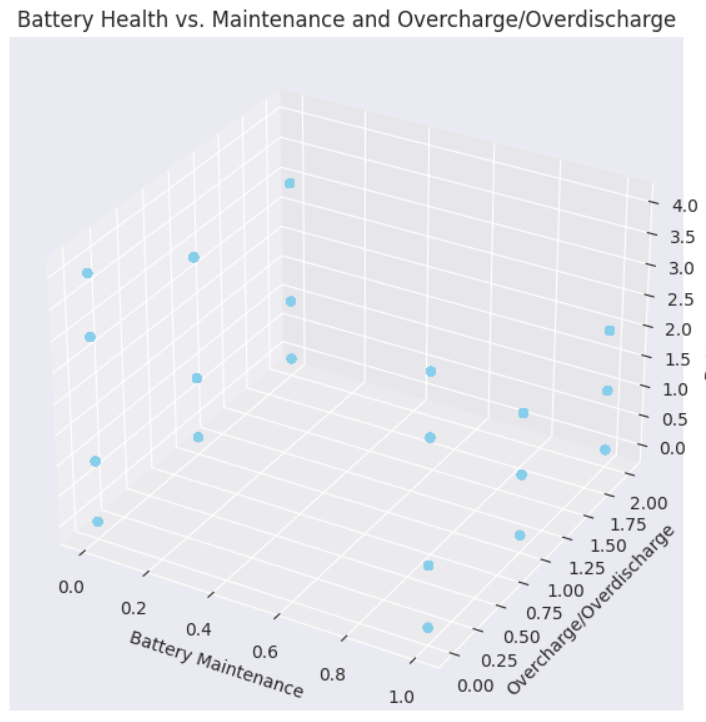


There are fewer instances of batteries experiencing short circuits than not, as shown by the number of data points at “0” and “1” on the x-axis.

Different levels of battery health are observed at both “0” and “1” short circuits, but there is a higher concentration of low battery health (red and purple) at “1” than at “0”.

There is no clear pattern between temperature and battery health, as the data points are scattered across the y-axis. However, there seems to be a slight trend of higher battery health (green and yellow) at lower temperatures than at higher temperatures.

9.Battery Health vs. Maintenance and Overcharge/Overdischarge - 3D Scatter Plot:



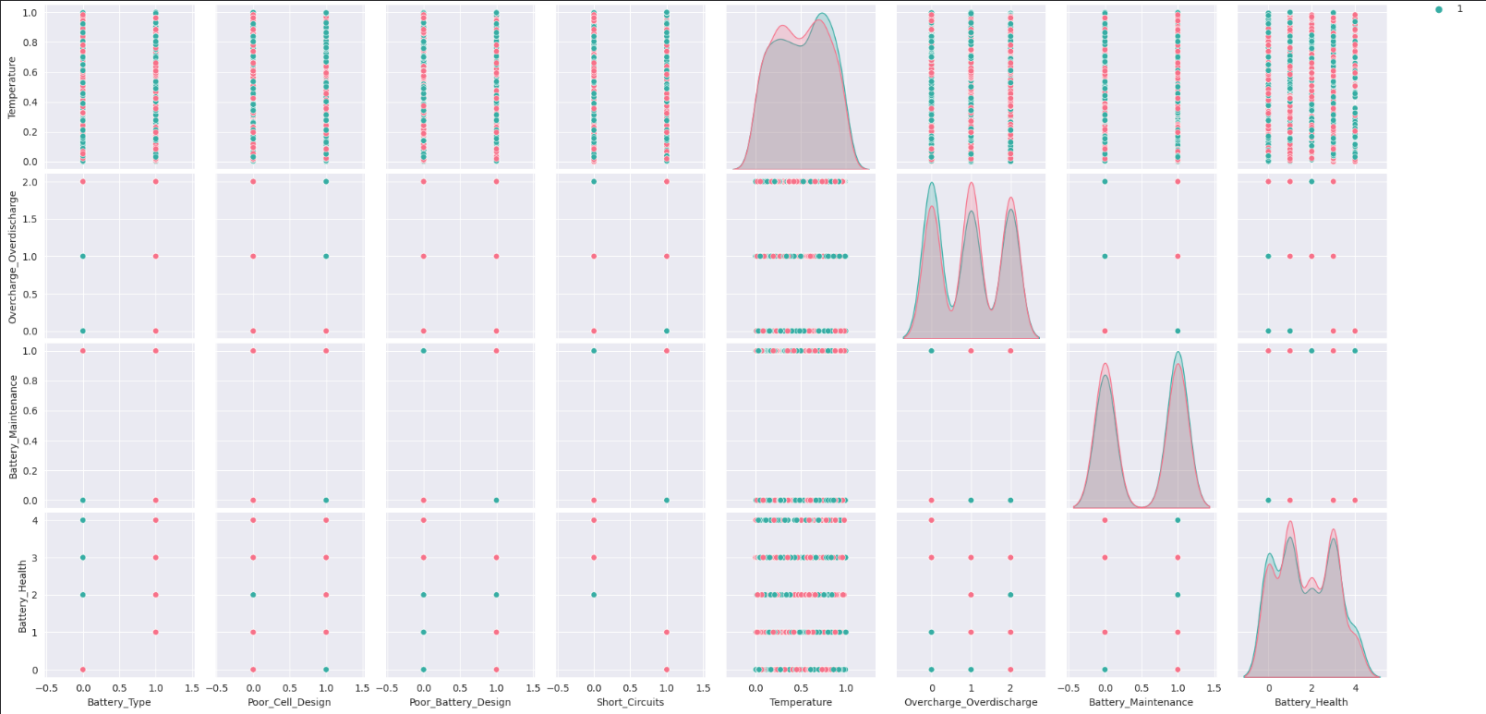
Higher maintenance and lower overcharge/overdischarge levels contribute to better battery health.

There is a negative correlation between overcharge/overdischarge and battery health, meaning that excessive charging or discharging can damage the battery.

There is a positive correlation between maintenance and battery health, meaning that regular check-ups and repairs can extend the battery life and performance.

10.Pairwise Scatter Plots with Hue for External\_Abuse





The strength and direction of correlations vary based on 'External\_Abuse' levels. Strong correlations in certain categories may indicate a more pronounced relationship between variables under specific external abuse conditions.

Pattern Discovery:

From Swarm plot,a pattern is observed:

* If a lead-acid battery experiences a short circuit, it is prone to a blast.
* In the case of a lithium-ion battery facing a short circuit, there is a chance of a blast occurring.

From Violin Plot,a pattern is observed:

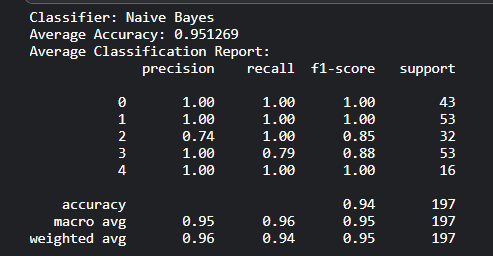
* When a battery undergoes over-discharge, the severity is categorized as moderate.
* Similarly, overcharging a battery also results in a moderate impact.

Model Fitting:

Naive Bayes Model:

Naive Bayes is a classification algorithm based on Bayes' theorem, which is a probabilistic approach to solve classification problems. It is called "naive" because it makes the assumption that the features used to describe an observation are conditionally independent

Used K fold Stratified for training the dataset,is a robust approach to assess the generalization performance of a model, especially when faced with imbalanced datasets.



The Avg Accuracy is 95.1269

End Notes:

In conclusion, our journey through the Exploratory Data Analysis of Electric Vehicle battery health has illuminated crucial aspects of the intricate relationship between various factors and the overall health of these essential components.

Analyzing the dataset reveals that opting for lithium-ion batteries enhances both efficiency and safety. Additionally, maintaining the battery is crucial for optimizing health and overall performance. Emphasizing the importance of preventing overheating emerges as a key strategy to boost efficiency levels.