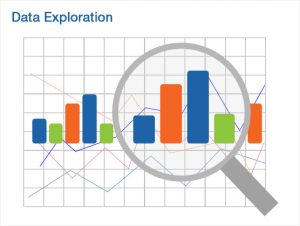
Machine Learning Challenge: Day 6

Welcome to the sixth day of our 30 days Machine learning Challenge.

**Exploring and Understanding Your Data: Techniques and Tools for Machine Learning**



### Machine learning is a powerful tool that allows us to make predictions and decisions based on data. One of the first steps in any machine learning project is data exploration, which involves analyzing and understanding the data that will be used to train and evaluate the model.

### Data Size & Summary Stats:

### Data size is an important consideration in data exploration, as the amount of data available can impact the performance and accuracy of a machine learning model. Summary statistics such as mean, median and standard deviation can provide a quick overview of the data and help identify any outliers or anomalies.

### Histogram

### Matplotlib Sample Histogram – Timo Denk's Blog

### Histograms are a useful tool for visualizing the distribution of a dataset. They provide a graphical representation of the frequency of different values in a dataset, and can help identify patterns or trends.

### Scatter Plot & Joint Plot

Scatter plots allow us to visualize the relationship between two variables, and can be useful for identifying patterns or outliers in the data. Joint plots are similar to scatter plots, but also include histograms of the individual variables.

### Pair Grid & Box and Violin Plots

Pair grids and box and violin plots are also useful tools for visualizing the relationship between variables. Pair grids allow us to plot multiple variables at once, while box and violin plots provide a more detailed view of the distribution of a variable.

* **Comparing Two Ordinal Values**

Comparing two ordinal values, such as categorical variables, can be done using a variety of tools such as bar plots, or stacked bar plots.

* **Correlation, RadViz & Parallel Coordinates**

Correlation is a measure of the relationship between two variables, and can be useful for identifying patterns or trends in the data. RadViz is a visualization tool that is particularly useful for identifying patterns in high-dimensional data.

Finally, parallel coordinates plots are used to visualize high dimensional data, by plotting each variable on a separate axis and connecting the observations by lines. It helps to identify the relationship between different variables and also to identify outliers.