REPORT ON DATA ANALYSIS

04-09-2025

CNC Machine ,VMC Machine ,Bending Machine, Injection Moulding

Dataset Description:

DeviceId, TimeStamp, DigitalInput1, NumberPulses_DigitalInput1, DigitalInput2, NumberPulses_DigitalInput2, AnalogInput1, AnalogInput2,VRN,VYN,VBN,Current_R, Current_B, Current_Y, dvcStatus, R_PowerFactor, Y_PowerFactor, B_PowerFactor, R_PowerAngle, Y_PowerAngle, B_PowerAngle, Frequency, Total_ImportEnergyConsumption, Total_ExportEnergyConsumption,Current

i. Data Preparation
ii. Load and Inspect
iii. Check missing values, describe data
iv. Univariate data analysis: Histograms
v. Bivariate data analysis: Scatterplot and boxplot
vi. Multivariate data analysis: Pairplot
vii. Outlier detection: Boxplot

Data Preprocessing

- Handling Missing Values
  - Used .isnull().sum() to check missing values.
  - Replaced missing numeric values with mean/median.
  - Filled categorical values using mode.
- Data Type Conversions
  - Converted Timestamp to datetime.
  - Converted categorical columns into proper categories.
- Outlier Detection
  - Used Boxplots and IQR method to identify unusual spikes.

Descriptive Statistics

- Used df.info() → to check data types and nulls.
- Used df.describe() → to summarize mean, std, min, max, percentiles.

Univariate Analysis

- Plotted histograms and boxplots for each variable.

Bivariate & Multivariate Analysis

- Correlation Analysis
  - Heatmap of numeric variables
- Groupby Analysis
- Time Series Trends

Hydraulic Press EDA

Dataset Description: TimeStamp, Current

i. Load and Inspect
ii. Convert TimeStamp to datetime(Easier to resample, plot, and detect patterns over time)
iii. Check missing values, describe data
iv. Univariate data analysis
v. Bivariate data analysis
vi. Multivariate data analysis
vii. Rolling Statistics (Trend Detection)
viii. Correlations
ix. Outlier & Event Detection

Data Preprocessing

- Handling Missing Values
  - Used .isnull().sum() to check missing values.
  - Replaced missing numeric values with mean/median.
  - Filled categorical values using mode.
- Data Type Conversions
  - Converted Timestamp to datetime.
  - Converted categorical columns into proper categories.
- Outlier Detection
  - Used Boxplots and IQR method to identify unusual spikes.

Descriptive Statistics

- Used df.info() → to check data types and nulls.
- Used df.describe() → to summarize mean, std, min, max, percentiles.

Univariate Analysis

- Plotted histograms and boxplots for each variable.

Bivariate & Multivariate Analysis

- Correlation Analysis
  - Heatmap of numeric variables
- Groupby Analysis
- Time Series Trends

Takeaways:

- VRN → Voltage between R-phase and Neutral
- VBN → Voltage between B-phase and Neutral
- VYN → Voltage between Y-phase and Neutral
- CNC machines require stable 3-phase supply.
- If VRN, VYN, or VBN **deviates too much, it indicates unbalanced voltage**, which can cause **machine errors or motor overheating**.
- Sudden drop in one phase voltage → possible line fault or fuse blown.

- Consistent imbalance → inefficient operation and vibration issues.
- Power Factor (PF) = Ratio of Real Power (kW) to Apparent Power (kVA).
- Low PF (<0.9) means more current is drawn for the same load → higher losses, heating.
- Imbalanced PF across R, Y, B phases can indicate motor drive issues or supply imbalance.
- Power Angle (δ) tells how much the **current is out of sync with voltage.**
- Total Import Energy Consumption: **cumulative amount of electrical energy (kWh)** that the **CNC machine has drawn** (imported) from the power grid or supply source over time. It **increases as the machine runs** and consumes power.
- Total Export Energy Consumption → **energy sent back to the grid** (e.g., during braking in CNC motors).

- Correlation between **current & pulses** → measure how load varies with production.
- Correlation between **Power Factor & Power Angle** → expected relationship (PF = cos δ).
- Correlation between **Analog Inputs** (temperature/pressure sensors) and **energy/current** → overheating issues.

EDA Libraries

Analysed data sets using EDA libraries like ydata-profiling and Autoviz.

The profiling report generates multiple correlation heatmaps (Pearson, Spearman, and Phik):

- Strong correlation was observed between **phase currents (R, Y, B)** and **energy consumption**, which validates the physics of electrical machines.
- **Power Factors** showed correlation with **Power Angles**, which is expected in electrical systems.
- Weak or no correlation was observed between unrelated signals (e.g., digital inputs and analog sensor readings).

This step helps in feature selection and reducing redundancy.

Insights from AutoViz

AutoViz automatically generated multiple plots such as distribution plots, scatter plots, and box plots for both numerical and categorical variables.

- **Distribution Analysis:**
  AutoViz showed that variables like voltages (VRN, VYN, VBN) and currents (IR, IY, IB) follow approximately normal-like distributions, but with occasional spikes indicating machine load variations.
- **Outlier Detection:**
  Box plots highlighted the presence of outliers in current signals (especially IR and IB), likely corresponding to sudden tool load or operational changes in the CNC machine.
- **Correlation Plots:**
  AutoViz correlation matrices indicated that voltage and current values strongly influence **total import energy consumption**, reinforcing electrical power relationships.
  Power Factor (Y_PowerFactor) also showed moderate correlation with energy usage and power angle, which is consistent with expected machine behavior.
- **Categorical Variable Impact:**
  Feature Importance:
  AutoViz highlighted key variables contributing most to variability in energy consumption, which supports feature selection for further machine learning or predictive maintenance models.

This step provided a quick **visual summary of trends, anomalies, and dependencies** in the CNC machine dataset, making it easier to interpret sensor relationships before applying advanced models.