# SQL

## EDA (Exploratory Data Analysis)

## Report

EDA (Exploratory Data Analysis) is a data analysis approach or methodology that is not specific to SQL but rather a general process used to summarize, understand, and visualize data in preparation for more advanced analysis. EDA is typically performed using tools, programming languages, or data analysis software, such as Python, R, or specialized data visualization tools. But here we have used SQL which is Structured Query Language to understand and summarize the data for further data analysis process.

We have summarize some of the categories or features of the data like Machine_ID, Assembly_Line_No, Hydraulic_Pressure(bar),Coolant_Pressure(bar),Coolant_Temperature, Hydraulic_Oil_Temperature(Â°C), Spindle_Bearing_Temperature(Â°C),Spindle_Vibration(Âµm),Tool_Vibration(Âµm), Spindle_Speed(RPM), Voltage(volts), Torque(Nm),Cutting(kN) and Downtime.

***Here's a refined summary of our findings:***

**Database and table**

- Database name : machine_data
- Table name : MachineDowntime
- Table count : 2500

**Machine_ID** :

*We identified three types of Machine_ID and their respective counts:*

| Machine_ID | Count |
|---|---|
| Makino-L3-Unit1-2015 | 818 records |
| Makino-L2-Unit1-2015 | 808 records |
| Makino-L1-Unit1-2013 | 874 records |

**Hydraulic_Pressure(bar)** :

*We computed the average Hydraulic Pressure for each Machine_ID*

| Machine_ID | AvgHydraulic_Pressure_bar |
|---|---|
| Makino-L3-Unit1-2015 | 100.855378955018 |
| Makino-L2-Unit1-2015 | 103.403585337761 |
| Makino-L1-Unit1-2013 | 100.220547700202 |

**Spindle_Speed(RPM) :**

*We determined the minimum and maximum Spindle Speed:*

Min Spindle_Speed_RPM is 0

MaxSpindle_Speed_RPM is 27957.

**Cutting(kN)** :

*Calculated the average Cutting force for each Assembly Line.*

| Assembly_Line_No | AvgCutting_kNForce |
|---|---|
| • Shopfloor-L1 | 2.7742268040964 |
| • Shopfloor-L2 | 2.79363975406433 |
| • Shopfloor-L3 | 2.78051533859932 |

### Coolant_Temperature :

*We determined the average Coolant Temperature for records with Machine Failure:*

| Machine_ID | AvgCoolantTempWithFailure |
|---|---|
| • Makino-L3-Unit1-2015 | 19.9192770107683 |
| • Makino-L2-Unit1-2015 | 20.3315655486752 |
| • Makino-L1-Unit1-2013 | 19.7374448524173 |

### Spindle Speed above 25000 RPM.

*We identified the count of records with Machine Failure and Spindle Speed above 25,000 RPM:*

| Machine_ID | RecordsWithFailureAndHighSpeed |
|---|---|
| • Makino-L1-Unit1-2013 | 131 |
| • Makino-L2-Unit1-2015 | 109 |
| • Makino-L3-Unit1-2015 | 106 |

### Moments of Distribution –

**First Moment (Mean):**

The mean is the average value of a dataset. It provides a central value around which the data points are distributed. In business, the mean is often used to determine the average performance, such as average sales, average customer spending, or average production costs, helping in making decisions about pricing, budgeting, and resource allocation.

**Second Moment (Variance):**

The variance measures the dispersion or spread of the data around the mean. It indicates how much the data points vary from the average. Businesses use variance to assess risk and volatility. For example, a high variance in sales might suggest unpredictable market conditions, leading to decisions on risk management strategies or diversification.

**Third Moment (Skewness):**

Skewness measures the asymmetry of the data distribution. Positive skewness indicates a distribution with a long tail on the right, while negative skewness indicates a long tail on the left. Understanding skewness helps businesses identify potential outliers or anomalies in data, such as extremely high or low sales, which can influence decisions on marketing strategies or product development.

**Fourth Moment (Kurtosis):**

Kurtosis measures the "tailedness" of the data distribution, indicating the presence of outliers. High kurtosis means more outliers, while low kurtosis means fewer outliers. Businesses use kurtosis to evaluate the likelihood of extreme events or outliers, such as rare but impactful financial losses, aiding in developing strategies for risk management and contingency planning.

## 1st Business Moment: Measure Of Central Tendency

**1. Mean**

| Category | Mean before Data Cleaning | Mean after Data Cleaning |
|---|---|---|
| Hydraulic_Pressure(bar) | 101.409084 | 101.390501 |
| Coolant_Pressure(bar) | 4.947058 | 4.947005 |
| Air_System_Pressure(bar) | 6.499275 | 6.499315 |
| Coolant_Temperature | 18.559887 | 18.572560 |
| Hydraulic_Oil_Temperature(°C) | 47.618317 | 47.618840 |
| Spindle_Bearing_Temperature(°C) | 35.063698 | 35.063800 |
| Spindle_Vibration(µm) | 1.009334 | 1.009328 |
| Tool_Vibration(µm) | 25.411975 | 25.412164 |
| Spindle_Speed(RPM) | 20274.792302 | 20274.462800 |
| Voltage(volts) | 348.996792 | 348.996800 |
| Torque(Nm) | 25.234968 | 25.230035 |
| Cutting(kN) | 2.782551 | 2.782544 |

**2. Median**

| Category | Median before Data Cleaning | Median after Data Cleaning |
|---|---|---|
| Hydraulic_Pressure(bar) | 96.763486 | 96.763486 |
| Coolant_Pressure(bar) | 4.939960 | 4.939960 |
| Air_System_Pressure(bar) | 6.505191 | 6.505191 |
| Coolant_Temperature | 21.200000 | 21.200000 |
| Hydraulic_Oil_Temperature(°C) | 47.700000 | 47.700000 |
| Spindle_Bearing_Temperature(°C) | 35.100000 | 35.100000 |
| Spindle_Vibration(µm) | 1.008000 | 1.008000 |
| Tool_Vibration(µm) | 25.455000 | 25.455000 |
| Spindle_Speed(RPM) | 20137.500000 | 20137.500000 |
| Voltage(volts) | 349.000000 | 349.000000 |
| Torque(Nm) | 24.647736 | 24.647736 |
| Cutting(kN) | 2.780000 | 2.780000 |

**3. Mode**

| Category | Mode before Data Cleaning | Mode after Data Cleaning |
|---|---|---|
| Date | 15-03-2022 | 15-03-2022 |
| Machine_ID | Makino-L1-Unit1-2013 | Makino-L1-Unit1-2013 |
| Assembly_Line_No | Shopfloor-L1 | Shopfloor-L1 |
| Hydraulic_Pressure(bar) | 88.279134 | 88.279134 |
| Coolant_Pressure(bar) | 4.566854 | 4.566854 |
| Air_System_Pressure(bar) | 5.628717 | 6.505191 |
| Coolant_Temperature | 26.4 | 26.4 |
| Hydraulic_Oil_Temperature(°C) | 47.5 | 47.7 |
| Spindle_Bearing_Temperature(°C) | 34.6 | 34.6 |
| Spindle_Vibration(µm) | 1.231 | 1.008 |
| Tool_Vibration(µm) | 26.736 | 25.455 |
| Spindle_Speed(RPM) | 17726.0 | 17726.0 |
| Voltage(volts) | 337.0 | 349.0 |
| Torque(Nm) | 35.580334 | 35.580334 |
| Cutting(kN) | 3.55 | 3.55 |
| Downtime | Machine_Failure | Machine_Failure |

**2nd Business Moment: Measure Of Dispersion**

**4. Variance**

| Category | Variance before Data Cleaning | Variance after Data Cleaning |
|---|---|---|
| Hydraulic_Pressure(bar) | 9.174417e+02 | 9.138565e+02 |
| Coolant_Pressure(bar) | 9.947213e-01 | 9.871588e-01 |
| Air_System_Pressure(bar) | 1.658765e-01 | 1.647483e-01 |
| Coolant_Temperature | 7.317912e+01 | 7.286103e+01 |
| Hydraulic_Oil_Temperature(°C) | 1.420291e+01 | 1.411201e+01 |
| Spindle_Bearing_Temperature(°C) | 1.417389e+01 | 1.413420e+01 |
| Spindle_Vibration(µm) | 1.175791e-01 | 1.170615e-01 |
| Tool_Vibration(µm) | 4.143679e+01 | 4.125441e+01 |
| Spindle_Speed(RPM) | 1.484299e+07 | 1.480740e+07 |
| Voltage(volts) | 2.058984e+03 | 2.054040e+03 |
| Torque(Nm) | 3.768196e+01 | 3.736818e+01 |
| Cutting(kN) | 3.803044e-01 | 3.792391e-01 |

**5. Standard Deviation**

| Category | Variance before Data Cleaning | Variance after Data Cleaning |
|---|---|---|
| Hydraulic_Pressure(bar) | 30.289301 | 30.230060 |
| Coolant_Pressure(bar) | 0.997357 | 0.993559 |
| Air_System_Pressure(bar) | 0.407279 | 0.405892 |
| Coolant_Temperature | 8.554480 | 8.535867 |
| Hydraulic_Oil_Temperature(°C) | 3.768674 | 3.756596 |
| Spindle_Bearing_Temperature(°C) | 3.764823 | 3.759547 |
| Spindle_Vibration(µm) | 0.342898 | 0.342143 |
| Tool_Vibration(µm) | 6.437142 | 6.422959 |
| Spindle_Speed(RPM) | 3852.660056 | 3848.038101 |
| Voltage(volts) | 45.376024 | 45.321518 |
| Torque(Nm) | 6.138564 | 6.112952 |
| Cutting(kN) | 0.616688 | 0.615824 |

**3rd & 4th Moment Business Decision**

**6. Skewness**

| Category | Variance before Data Cleaning | Variance after Data Cleaning |
|---|---|---|
| Hydraulic_Pressure(bar) | 0.197076 | 0.199274 |
| Coolant_Pressure(bar) | 0.147081 | 0.147806 |
| Air_System_Pressure(bar) | -0.052899 | -0.053378 |
| Coolant_Temperature | 0.108301 | 0.104172 |
| Hydraulic_Oil_Temperature(°C) | -0.002291 | -0.002716 |
| Spindle_Bearing_Temperature(°C) | -0.035942 | -0.036074 |
| Spindle_Vibration(µm) | 0.001532 | 0.001587 |
| Tool_Vibration(µm) | -0.061005 | -0.061228 |
| Spindle_Speed(RPM) | -0.172475 | -0.172424 |
| Voltage(volts) | -0.028658 | -0.028693 |
| Torque(Nm) | 0.030578 | 0.033119 |
| Cutting(kN) | 0.113963 | 0.114157 |

**7. Kurtosis**

| Category | Variance before Data Cleaning | Variance after Data Cleaning |
|---|---|---|
| Hydraulic_Pressure(bar) | -0.911830 | -0.920292 |
| Coolant_Pressure(bar) | 1.201405 | 1.169455 |
| Air_System_Pressure(bar) | 0.018971 | -0.001578 |
| Coolant_Temperature | 1.581860 | 1.564639 |
| Hydraulic_Oil_Temperature(°C) | 0.057724 | 0.038164 |
| Spindle_Bearing_Temperature(°C) | -0.035556 | -0.043862 |
| Spindle_Vibration(µm) | -0.005537 | -0.018718 |
| Tool_Vibration(µm) | 0.019749 | 0.006451 |
| Spindle_Speed(RPM) | 1.355132 | 1.344769 |
| Voltage(volts) | -0.083325 | -0.090328 |
| Torque(Nm) | -0.444431 | -0.465628 |
| Cutting(kN) | -1.081785 | -1.087171 |