

Preliminary analysis on TERMINET SCHN data

```
In [34]: import pandas as pd
from scipy.stats import pearsonr
```

```
In [23]: df = pd.read_csv("./SOE_28-03-2022.xlsx - Sheet2.csv")
```

```
In [30]: df["NAME"].unique()
```

```
Out[30]: array(['CPU_USE', 'FAIL_CONF', 'FAIL_RTU', 'MEM_USE', 'PS1_V',
        'RTU_HEALTH', 'TEMP'], dtype=object)
```

```
In [93]: CPUs = df[df["NAME"] == "CPU_USE"]["VAL"].dropna()[:-1]
FAIL_CONFs = df[df["NAME"] == "FAIL_CONF"]["VAL"].dropna()
FAIL_RTUs = df[df["NAME"] == "FAIL_RTU"]["VAL"].dropna()
MEM_USEs = df[df["NAME"] == "MEM_USE"]["VAL"].dropna()
PS1_Vs = df[df["NAME"] == "PS1_V"]["VAL"].dropna()
RTU_HEALTHs = df[df["NAME"] == "RTU_HEALTH"]["VAL"].dropna()
TEMPs = df[df["NAME"] == "TEMP"]["VAL"].dropna()
```

```
In [95]: print("len(CPUUs):", len(CPUUs))
print("len(FAIL_CONFs):", len(FAIL_CONFs))
print("len(FAIL_RTUs):", len(FAIL_RTUs))
print("len(MEM_USEs):", len(MEM_USEs))
print("len(PS1_Vs):", len(PS1_Vs))
print("len(RTU_HEALTHs):", len(RTU_HEALTHs))
print("len(TEMPs):", len(TEMPs))
```

```
len(CPUUs): 377
len(FAIL_CONFs): 377
len(FAIL_RTUs): 377
len(MEM_USEs): 377
len(PS1_Vs): 377
len(RTU_HEALTHs): 377
len(TEMPs): 377
```

ERROR distribution

```
In [97]: print("FAIL_CONFs.unique", FAIL_CONFs.unique())
print("FAIL_RTUs.unique", FAIL_RTUs.unique())
```

```
FAIL_CONFs.unique [0.]
FAIL_RTUs.unique [0.]
```

-----> NB: there are no errors in the dataset!!!

Correlations

CPU

```
In [98]: print("pearsonr(CPUUs, TEMPs): ", pearsonr(CPUUs, TEMPs)[1])
print("pearsonr(CPUUs, MEM_USEs): ", pearsonr(CPUUs, MEM_USEs)[1])
print("pearsonr(CPUUs, PS1_Vs): ", pearsonr(CPUUs, PS1_Vs)[1])
print("pearsonr(CPUUs, RTU_HEALTHs): ", pearsonr(CPUUs, RTU_HEALTHs)[1])
```

```
pearsonr(CPUs, TEMPs): 0.2505619822815974
pearsonr(CPUs, MEM_USEs): 8.786607940359446e-36
pearsonr(CPUs, PS1_Vs): 0.33321768947957675
pearsonr(CPUs, RTU_HEALTHs): 0.12338897563973102
```

MEMORY USAGE

In [99]:

```
print("pearsonr(MEM_USEs, TEMPs): ", pearsonr(MEM_USEs, TEMPs)[1])
print("pearsonr(MEM_USEs, PS1_Vs): ", pearsonr(MEM_USEs, PS1_Vs)[1])
print("pearsonr(MEM_USEs, RTU_HEALTHs): ", pearsonr(MEM_USEs, RTU_HEALTHs)[1])
```

```
pearsonr(MEM_USEs, TEMPs): 0.0009284286611734351
pearsonr(MEM_USEs, PS1_Vs): 0.5632157774318161
pearsonr(MEM_USEs, RTU_HEALTHs): 0.4720148305912593
```

PS1_vs

In [100...]

```
print("pearsonr(PS1_Vs, TEMPs): ", pearsonr(PS1_Vs, TEMPs)[1])
print("pearsonr(PS1_Vs, RTU_HEALTHs): ", pearsonr(PS1_Vs, RTU_HEALTHs)[1])
```

```
pearsonr(PS1_Vs, TEMPs): 0.23315064103529948
pearsonr(PS1_Vs, RTU_HEALTHs): 0.7334088000294624
```

Conclusions from a preliminary analysis

There is no correlation with errors, since there are not errors in the dataset. There is a weak corellation between:

- PS1_Vs and MEM_USEs (0.56)
- PS1_Vs and RTU_HEALTHs (0.73)

In []: