Evidence of realization

Giedre Zalaite, r0844971

3ACS, IT Factory

Academic year 20xx-20xx

Campus Geel, Kleinhoefstraat 4, BE-2440 Geel

LID VAN ASSOCIATIE KU LEUVEN

THOMAS MORE

Contents

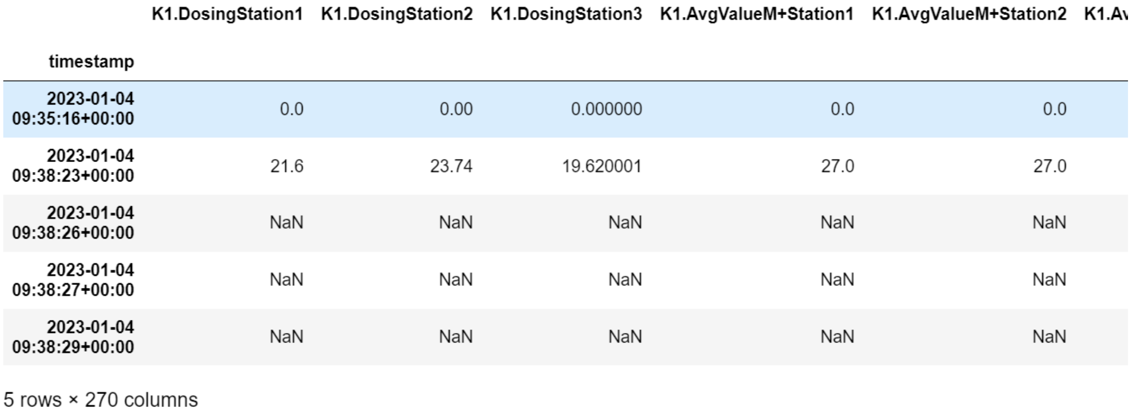
1 INTRODUCTION

This document serves as evidence or realization for my internship at Duracell, where I worked on the analysis of a machine responsible for manufacturing rings used in battery composition. The manufacturing process involves harsh materials that put a strain on the equipment, necessitating regular equipment maintenance and replacement. The objective of the project was to analyse the machine's behaviour, identify influential measures and conditions, and develop machine learning models to predict optimal equipment maintenance timing. During my internship, I actively participated in data cleaning, data visualization, and drawing meaningful conclusions to lay the groundwork for model development.

The project involved working with datasets obtained from iHistorian and a SQL server, aiming to gain valuable insights into machine performance and identify critical factors for further analysis. This document provides a comprehensive overview of the data cleaning techniques employed, the visualizations created, and the conclusions derived from the analysis. By presenting this evidence, I aim to demonstrate my understanding of data analysis methodologies, visualization techniques, and the ability to draw meaningful insights to support the development of machine learning models for predictive maintenance.

2 DATA CLEANING

The initial dataset was obtained from iHistorian. This dataset is updated only when a value is changed, with the updated measure entered and other measures filled with NA. To clean this dataset using Python, I created a pandas data frame, standardized the measure names to English, and removed any duplicate entries, if present.



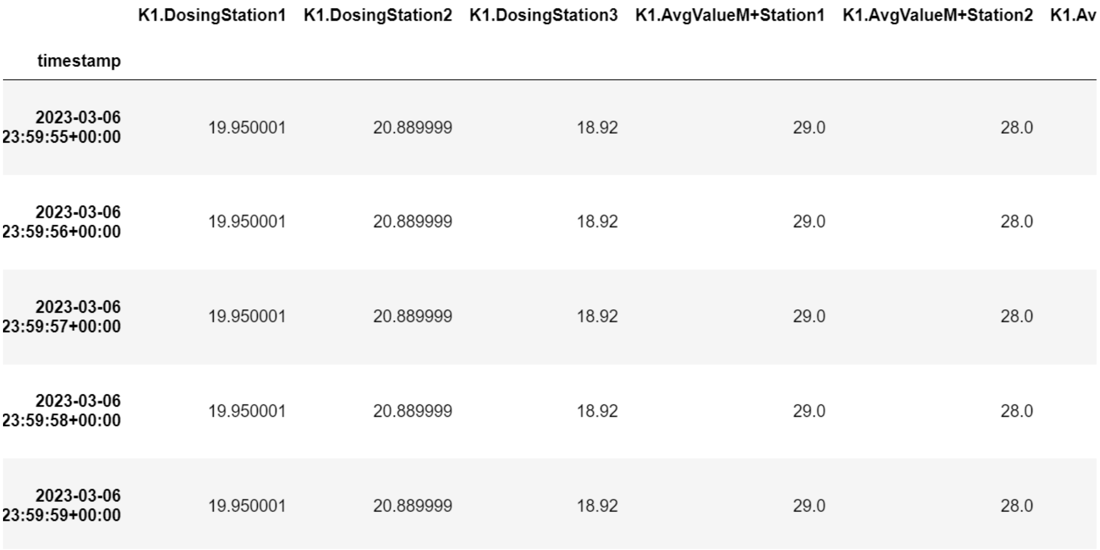
	K1.DosingStation1	K1.DosingStation2	K1.DosingStation3	K1.AvgValueM+Station1	K1.AvgValueM+Station2	K1.Av
timestamp						
2023-01-04 09:35:16+00:00	0.0	0.00	0.000000	0.0	0.0	
2023-01-04 09:38:23+00:00	21.6	23.74	19.620001	27.0	27.0	
2023-01-04 09:38:26+00:00	NaN	NaN	NaN	NaN	NaN	
2023-01-04 09:38:27+00:00	NaN	NaN	NaN	NaN	NaN	
2023-01-04 09:38:29+00:00	NaN	NaN	NaN	NaN	NaN	

5 rows × 270 columns

To ensure data accuracy during calculations and analysis, I proceeded to adjust the intervals between entries to be consistent at one second. Additionally, I filled all NA values with the corresponding values from the preceding rows.

Another dataset was obtained from a SQL server, which provides information about machine status, including downtime, uptime, and the material present in the machine at specific time periods. While most of this data is irrelevant to the project, the line of the machine, material, downtime, and date with time are of interest. During machine downtime, the data related to it is not significant. Hence, the first dataset was filtered based on downtime, material name, and the machine's line, and this information was added to the first dataset.

In order to filter out downtime and merge the necessary values, it was necessary to identify the rows in the first dataset that fell within the time period covered by the second dataset and match the required values. Once the datasets were connected, the rows with non-zero downtime could be removed.



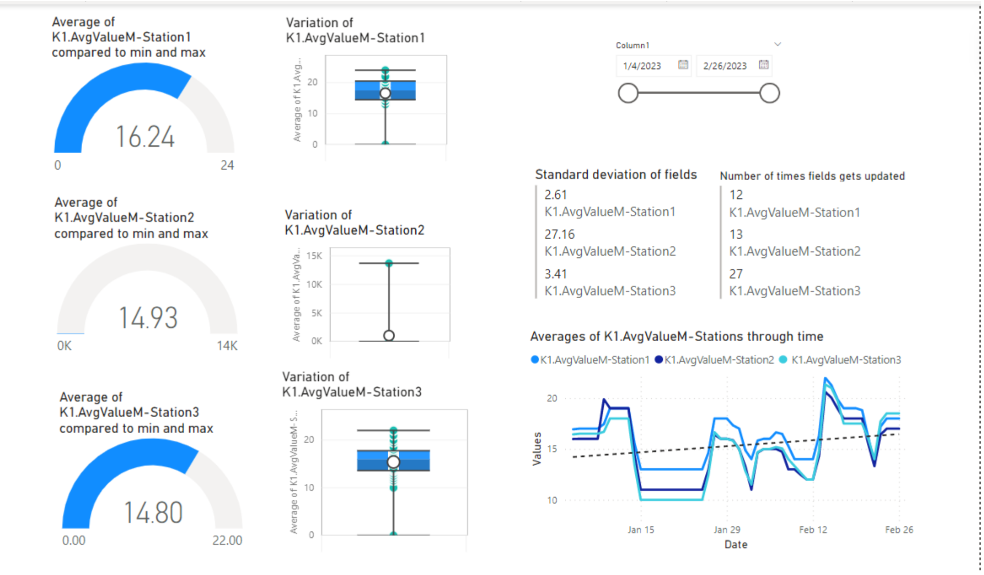
	K1.DosingStation1	K1.DosingStation2	K1.DosingStation3	K1.AvgValueM+Station1	K1.AvgValueM+Station2	K1.Av
timestamp						
2023-03-06 23:59:55+00:00	19.950001	20.889999	18.92	29.0	28.0	
2023-03-06 23:59:56+00:00	19.950001	20.889999	18.92	29.0	28.0	
2023-03-06 23:59:57+00:00	19.950001	20.889999	18.92	29.0	28.0	
2023-03-06 23:59:58+00:00	19.950001	20.889999	18.92	29.0	28.0	
2023-03-06 23:59:59+00:00	19.950001	20.889999	18.92	29.0	28.0	

3 DATA VISUALIZATION

The subsequent phase involved examining the behaviour of the machines, necessitating the creation of visualizations. I will present a range of graphs from each group and explain their significance to the project or noteworthy aspects. I will also share the findings and challenges encountered during the visualization process.

3.1 Power BI

Initially, I utilized Power BI to create graphs that focused on a longer time period, specifically from the beginning of January to the middle of March. During this phase, it was crucial to explore the data further and determine the key factors for future analysis. Graphs were generated for each measure to observe their changes over time, frequency of updates, and key performance indicators such as minimum, maximum, average values, and standard deviations. These insights helped in better understanding the measures. After several meetings, specific measures were selected for long-term importance, and certain daily trends were identified, such as spikes at the beginning of the day. Consequently, the decision was made to focus on shorter time periods for subsequent analysis.


Average of K1.AvgValueM-Station1 compared to min and max
16.24
0
24
Average of K1.AvgValueM-Station2 compared to min and max
14.93
0K
14K
Average of K1.AvgValueM-Station3 compared to min and max
14.80
0.00
22.00
Variation of K1.AvgValueM-Station1
Average of K1.Avg...
0
10
20
Variation of K1.AvgValueM-Station2
Average of K1.AvgVa...
0K
5K
10K
15K
Variation of K1.AvgValueM-Station3
Average of K1.AvgValueM-S...
0
10
20
Averages of K1.AvgValueM-Stations through time
● K1.AvgValueM-Station1
● K1.AvgValueM-Station2
● K1.AvgValueM-Station3
Values
10
15
20
Jan 15
Jan 29
Feb 12
Feb 26
Date

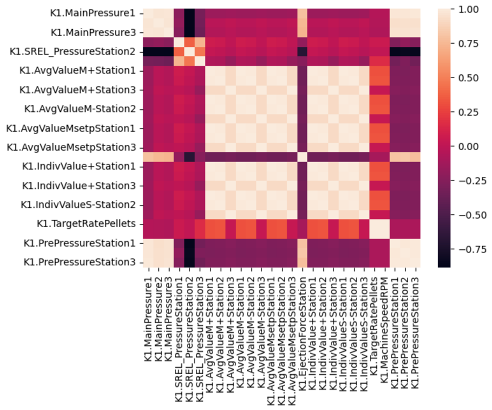
Column1

1/4/2023 2/26/2023

Standard deviation of fields	Number of times fields gets updated
2.61 K1.AvgValueM-Station1	12 K1.AvgValueM-Station1
27.16 K1.AvgValueM-Station2	13 K1.AvgValueM-Station2
3.41 K1.AvgValueM-Station3	27 K1.AvgValueM-Station3

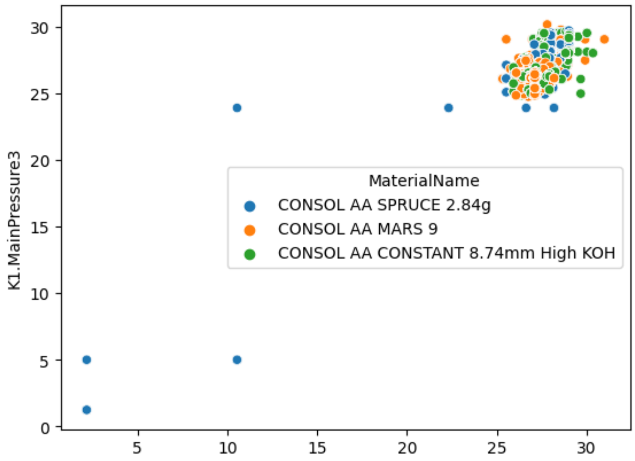
3.2 Correlation matrixes

To address memory limitations, I transitioned to using Python and a virtual environment with a Jupyter server that had been set up earlier. The analysis continued by examining correlations between measures to determine which ones could potentially be eliminated and which ones were crucial to retain. Any changes in correlation indicated potential issues. Following a meeting with the production team, it was concluded that all stations of main pressure and SREL pressure should be retained, as consistent correlations among the stations were crucial. Deviations in these correlations could indicate faults in the machine.

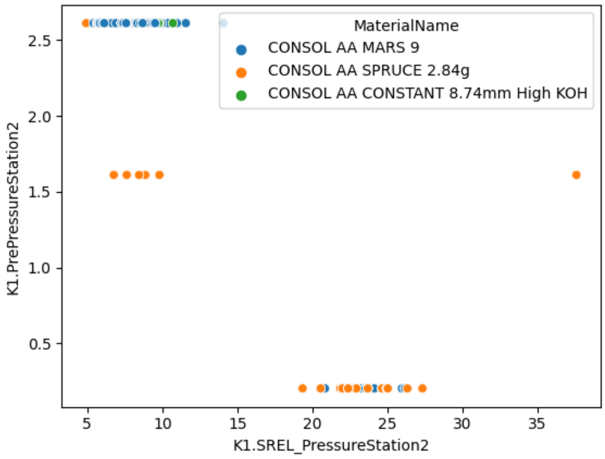


3.3 Scatter plots

The next step involved creating scatter plots for multiple measures, using different colours to represent the material in the machine. This analysis aimed to assess the impact of different materials on the measures. Certain materials exhibited distinct trends and had more outliers compared to others. These observations were noted and considered during the model development.

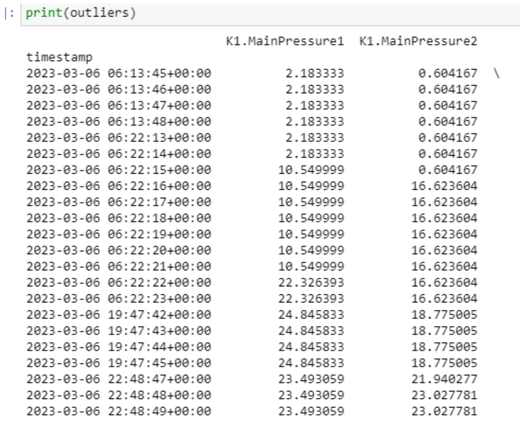


4]: [ ]



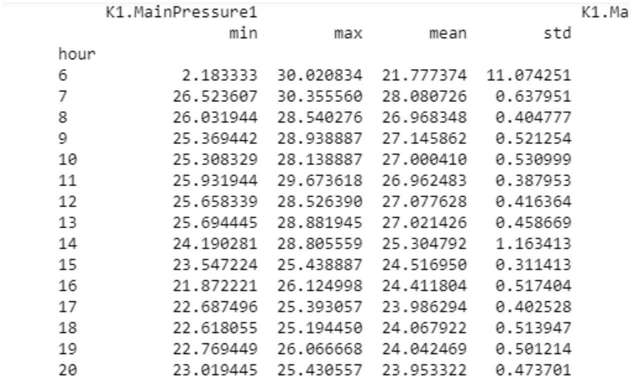
3.4 Over time analysis

Finally, line graphs were generated to visualize the changes in values over time. This time, only a short 24-hour period was considered, focusing solely on the most important measures affecting the machine. Outliers were plotted on top of the line graphs, and calculations were performed to determine minimum, maximum and average values, and standard deviations for each hour. Moving averages were also calculated with a window size of 1 minute to smooth the line graphs and facilitate comparison. These calculations and graphs provided additional insights, such as the incorrect values observed at the beginning of the day when the machine was just being turned on, the decreasing trend in most measures with occasional exceptions, and more. All these observations were taken into account when deciding which measures to retain for the model, which measures to combine, and which trends, materials, and times should be considered.

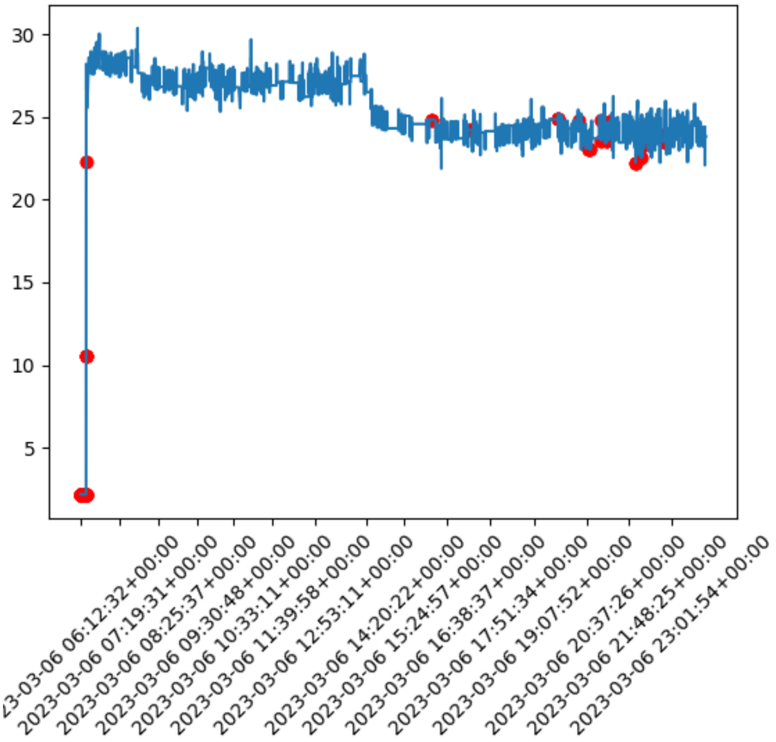


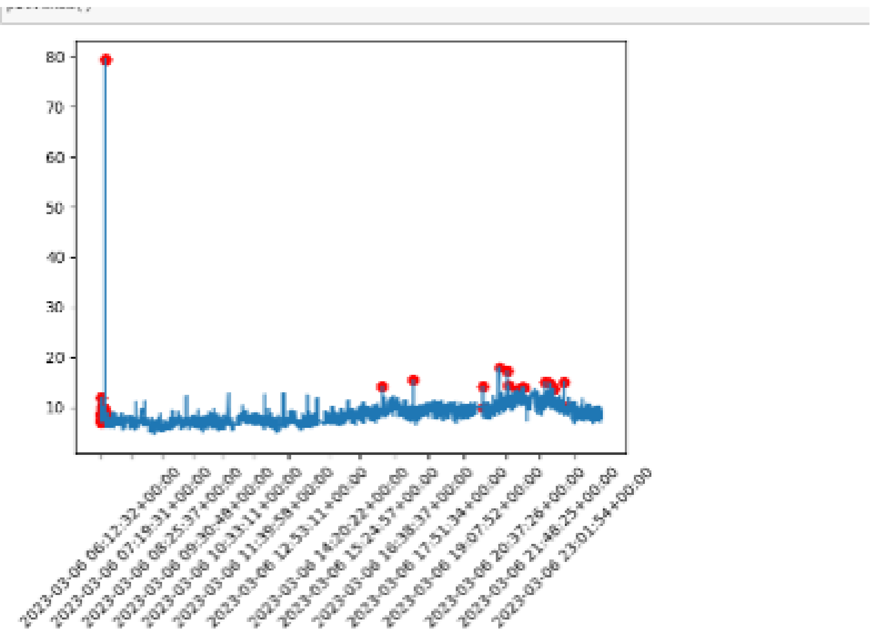
```
|: print(outliers)

                           K1.MainPressure1  K1.MainPressure2
timestamp
2023-03-06 06:13:45+00:00          2.183333          0.604167  \
2023-03-06 06:13:46+00:00          2.183333          0.604167
2023-03-06 06:13:47+00:00          2.183333          0.604167
2023-03-06 06:13:48+00:00          2.183333          0.604167
2023-03-06 06:22:13+00:00          2.183333          0.604167
2023-03-06 06:22:14+00:00          2.183333          0.604167
2023-03-06 06:22:15+00:00         10.549999          0.604167
2023-03-06 06:22:16+00:00         10.549999         16.623604
2023-03-06 06:22:17+00:00         10.549999         16.623604
2023-03-06 06:22:18+00:00         10.549999         16.623604
2023-03-06 06:22:19+00:00         10.549999         16.623604
2023-03-06 06:22:20+00:00         10.549999         16.623604
2023-03-06 06:22:21+00:00         10.549999         16.623604
2023-03-06 06:22:22+00:00         22.326393         16.623604
2023-03-06 06:22:23+00:00         22.326393         16.623604
2023-03-06 19:47:42+00:00         24.845833         18.775005
2023-03-06 19:47:43+00:00         24.845833         18.775005
2023-03-06 19:47:44+00:00         24.845833         18.775005
2023-03-06 19:47:45+00:00         24.845833         18.775005
2023-03-06 22:48:47+00:00         23.493059         21.940277
2023-03-06 22:48:48+00:00         23.493059         23.027781
2023-03-06 22:48:49+00:00         23.493059         23.027781
```

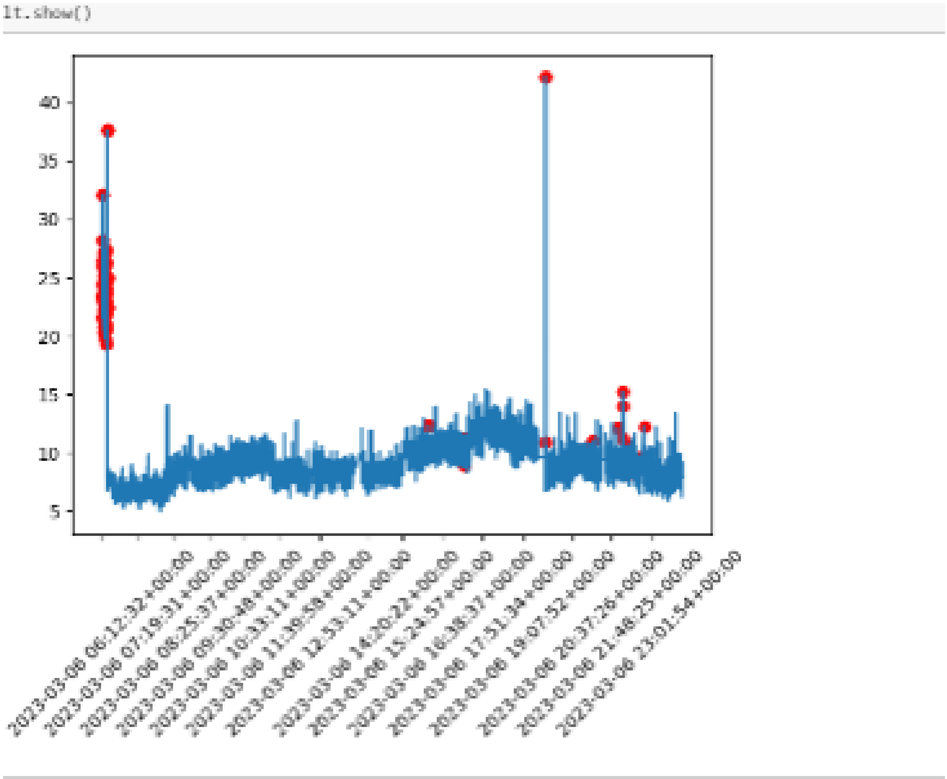


```
     K1.MainPressure1                                         K1.Ma
                  min        max       mean        std
hour
6            2.183333  30.020834  21.777374  11.074251
7           26.523607  30.355560  28.080726   0.637951
8           26.031944  28.540276  26.968348   0.404777
9           25.369442  28.938887  27.145862   0.521254
10          25.308329  28.138887  27.000410   0.530999
11          25.931944  29.673618  26.962483   0.387953
12          25.658339  28.526390  27.077628   0.416364
13          25.694445  28.881945  27.021426   0.458669
14          24.190281  28.805559  25.304792   1.163413
15          23.547224  25.438887  24.516950   0.311413
16          21.872221  26.124998  24.411804   0.517404
17          22.687496  25.393057  23.986294   0.402528
18          22.618055  25.194450  24.067922   0.513947
19          22.769449  26.066668  24.042469   0.501214
20          23.019445  25.430557  23.953322   0.473701
```





```
plt.show()
```



```
lt.show()
```

4 CONCLUSION

In conclusion, this document not only demonstrates the successful application of data cleaning techniques and data visualization processes for machine behaviour analysis but also highlights the valuable skills I have acquired during my internship experience. By leveraging tools such as Power BI and Python, I have gained a comprehensive understanding of data analysis, including measures, trends, and correlations, which has enhanced my abilities as an IT specialist.

Throughout this project, I have learned how to set up a virtual environment and conduct in-depth data analysis, providing me with a solid foundation for future endeavours. These acquired skills will undoubtedly contribute to my professional growth and serve as a strong starting point for my career upon graduating.

Moreover, this internship experience has provided me with invaluable opportunities to work with data, Python, perform a thorough analysis, and draw meaningful conclusions. Through these tasks, I have gained practical insights into machine performance and identified potential faults, equipping me with the ability to make informed decisions regarding equipment maintenance. The knowledge and skills I have acquired through this internship will undoubtedly serve me well in my future career. The ability to effectively work with data, perform analysis, and draw insightful conclusions is a strong foundation for continued growth and a career in the field of data analysis and machine learning.

In addition to technical skills, this internship has enhanced my time management abilities as I worked on various tasks within specified deadlines. The experience has also fostered adaptability as I navigated through different datasets, software tools, and analysis techniques. Moreover, I have sharpened my problem-solving skills by identifying and addressing challenges encountered during the data cleaning and visualization processes.

In summary, this project has not only provided valuable experience during my internship but has also equipped me with the necessary skills and knowledge to excel in my future endeavours. The groundwork established through the virtual environment setup and the thorough analysis conducted will undoubtedly contribute to ongoing analysis and future advancements to work with data.