# Project 1

| Student ID | Student Name | Contribution (%) | Contribution (Descriptions) | Score |
|---|---|---|---|---|
| **2102264** | **Winston Goo** | 25 | Program / Report | |
| **2100924** | **Heng Zhi Xiong** | 25 | Program / Report | |
| **2100880** | **Wong Guo Liang** | 25 | Report | |
| **2102018** | **Mohamed Nurandi** | 25 | Program / Report | |

# TABLE OF CONTENTS

# LIST OF FIGURES

# 1 Introduction

Injection Moulding is widely used for creating plastic components and products as it enables us to produce plastic parts in large quantities with great precision and accuracy. Injection moulding is a manufacturing process that involves using several materials like metals, glasses, and commonly used thermoplastic and thermosetting polymers. It is also utilized in different industries, like the automobile and consumer products sectors.

## 1.1 Injection moulding process

The injection moulding process requires six significant steps to turn raw materials into the wanted design and solidify. The Mould close phase begins with the moulding machine closing and applying clamp pressure. The Injection phase follows, in which the screw advances, pushing molten plastic out of the barrel and into the mould cavities. The Pack/Hold step follows, which packs the part the rest of the way and maintains material pressure until the gate freezes. The material is full up but too hot to be removed from the mould after the Pack/Hold phase, thus the Cooling/Plasticizing phase absorbs heat from the plastic portion to cool it down. Following that, the Mould Open phase, in which the material will be pulled out. Finally, there is the Part Ejection stage, in which the components must be removed. [1]

## 1.2 Problem statement

Injection moulding produces plastic parts and components in most industries. This technology aims to efficiently generate high-quality, accurate, and consistent plastic parts. However, injection moulding may have complicated issues that must be handled for production to be successful. [2] These issues involve tool design, machine operation, and quality assurance. It is also hard to produce consistent results while searching for the correct material, drying it, and combining it. Optimal process parameters like temperature, pressure injection speed, and cooling time are required for perfect parts. Creating and maintaining a dependable moulding process that reduces cycle times and scrap is difficult.

## 1.3 Objectives

This project's objective is to use process monitoring to determine which process setting is within the acceptable range. With this approach, it can easily identify the few main aspects of the production process according to the data provided as shown in the following:

- Prepare the dataset provided for analysis. Hence, data pre-processing techniques is required to make sure that our dataset is appropriate to maintain an accurate and stable analysis.
- With the data provided, we need to understand which parameter is affecting the product quality. Furthermore, we need to determine the correlation coefficient between each parameter and each condition.

# 2 Data Understanding and preprocessing

## 2.1 Understand of the data

The distribution of these classes within our raw dataset is graphically depicted in Figure 1. This comprehensive collection amounts to a total of 111 records, each characterized by a unique set of 12 parameters, linked to the injection moulding process.

| Condition | Count |
|---|---|
| Normal1 | 50 |
| Condition1 | 27 |
| Condition5 | 34 |

*Figure 1 - Records for each condition*

The total null values for each condition under each parameter are shown in figure below:

| | Condition | Cylinder heating zone 1 | Cylinder heating zone 2 | Cylinder heating zone 3 | Cylinder heating zone 4 | Cylinder heating zone 5 | Maximum injection pressure | Mould temperature control unit 1 | Cycle time | Injection time | Dosage time | Switch-over volume | Material cushion |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Normal1 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1 | Condition1 | 2 | 3 | 2 | 3 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 2 | Condition5 | 2 | 3 | 3 | 1 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 |

*Figure 2 - Null values for each condition*

## 2.2 Data preprocessing

Data checking and Data Pre-processing are the most fundamental and critical stage in the data preparation process. This is to ensure in producing accurate, statistically stable, and reliable data outcomes.

In the data checking process, this involves searching in the raw dataset for errors, missing value, and outliers. Firstly, it begins with checking data distribution through both box plot and scatter plot for a better data visualization. With data visualization, we can easily identify our outliers through box and whisker. As a result, we have detected two different outliers that are within the parameters "Switch-over volume" and "Cycle time" as shown in Figure 3.

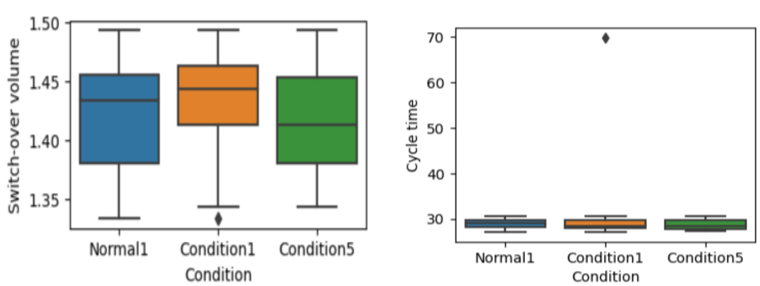


*Figure 3 - Outlier for 'Switch over volume' & 'Cycle time'*

In our next phase, Data Pre-processing phase, we eliminated the outlier from the 'Condition1' class associated with the parameter 'Cycle time'. Due to the outlier having an exceptional value of 68.802, while the rest of the value is within the range of 27~29. This removal was to maintain statistical stability, resulting in a total of 110 records and 26 data points for Condition1, as opposed to the initial count of 27 data points. Moreover, the removal of the outlier not only resolved the statistical stability but also resolved the null value issue under the parameter 'Switch Over Volume' and enabled the use of forward filling, which was not possible previously. However, the outlier under the parameter 'Switch Over time' was retained due to its small margin difference from the average, which computed less than 0.1.

## 2.3 Discussion (preprocessing result)

When filling the null values, we had a choice of either a forward fill, backward fill, or mean fill. Therefore, we decided to try out all three filling methods and compare them with the original data afterwards to check the difference in standard deviation caused by each method. Before we use either forward or backward, we had to check if there were any null values in the first or last rows, this is to prevent any errors when using the fill null function.

After filling all the null values accordingly, we had to calculate its standard deviations, grouped by the different conditions, for each filling method including the original data. This is done to compare newly filled data with the original data by subtracting between each set of data. An example of the formula below:

*Std difference = abs (original data std – forward fill data std)*

For a straightforward comparison between the different methods, we took the mean of the standard deviations of different conditions within its filling methods and a bar plot as seen in Figure 4 to visually represent the difference. From Figure 4, it can clearly be interpreted that the method with the least deviation from the original data is forward fill, with approximately 65% parameters appearing to have the smaller standard deviation compared to the other methods. Hence, from these calculations, we decided to use the forward fill as it generated the most accurate and reliable data filling compared to backwards and mean fill.

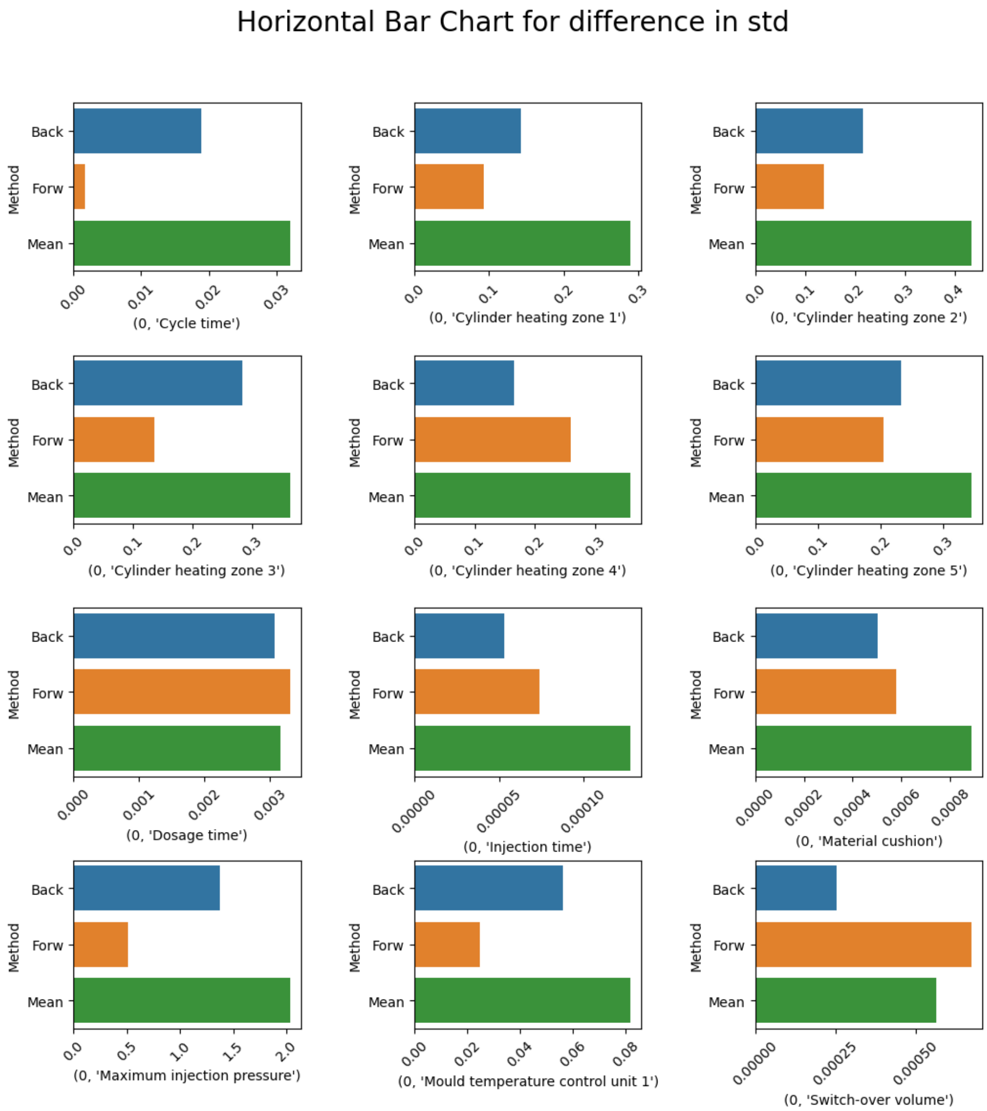


*Figure 4 - Difference in standard deviation*

Figures 5 show before and after being filled with the forward fill method. We will replace the null value with the previous value of the null row on each individual column. Based on Figure 5, under cylinder heating zone 1, the null value will be replaced with ‘288.35’. For cylinder heating zone 4, the null value will be replaced with '308.36’ and for Maximum injection pressure, null value will be replaced with ‘1544.9’

| Condition | Cylinder heating zone 1 | Cylinder heating zone 4 | Maximum injection pressure |
|---|---|---|---|
| Normal1 | 282.86 | 316.12 | 1527.07 |
| Normal1 | 295.22 | 306.33 | 1423.64 |
| Normal1 | 292.08 | 326.91 | 1509.38 |
| Normal1 | 274.11 | 308.36 | 1448.77 |
| Normal1 | 264.64 | 308.36 | 1516.42 |
| Normal1 | 286.2 | null | 1506.39 |
| Normal1 | 273.03 | 310.59 | 1544.9 |
| Normal1 | 288.35 | 325.41 | null |
| Normal1 | null | 310.03 | 1593.95 |
| Normal1 | 266.7 | 292.46 | 1462.78 |

| Condition | Cylinder heating zone 1 | Cylinder heating zone 4 | Maximum injection pressure |
|---|---|---|---|
| Normal1 | 282.86 | 316.12 | 1527.07 |
| Normal1 | 295.22 | 306.33 | 1423.64 |
| Normal1 | 292.08 | 326.91 | 1509.38 |
| Normal1 | 274.11 | 308.36 | 1448.77 |
| Normal1 | 264.64 | 308.36 | 1516.42 |
| Normal1 | 286.2 | 308.36 | 1506.39 |
| Normal1 | 273.03 | 310.59 | 1544.9 |
| Normal1 | 288.35 | 325.41 | 1544.9 |
| Normal1 | 288.35 | 310.03 | 1593.95 |
| Normal1 | 266.7 | 292.46 | 1462.78 |

Figure 5 – Before (Left) and After (Right) filling the null value

# 3 Statistical analysis

In this project, we performed the foundation of statistical analysis, understanding measures such as mean, median, range, and standard deviation. Our primary goal was to correlate variables and reveal patterns and trends in the dataset given.

## 3.1 Mean/ median/ range/ standard deviation

The statistical mean/median is a measurement of the center distribution, providing us with an idea about where the data gathered tends to concentrate. With the computed mean for Condition 1" displays an increment of 10°C in the parameter "Cylinder heating zone 1 to 5" compared to other conditions. With the increment of temperature, this corresponds to the parameters "Maximum Injection Pressure" and "Dosage time" on have the lowest value within three conditions. On top of it, under Condition: "Condition5", parameter "Mould Temperature control unit 1 has approximate increase 10% compared to other conditions, with a value of 99.51.

The statistical mean is shown in the figure below:

| Condition | Cylinder heating zone 1 | Cylinder heating zone 2 | Cylinder heating zone 3 | Cylinder heating zone 4 | Cylinder heating zone 5 | Maximum injection pressure | Mould temperature control unit 1 | Cycle time | Injection time | Dosage time | Switch-over volume | Material cushion |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Normal1 | 280.91 | 296.25 | 311.50 | 309.86 | 312.43 | 1547.75 | 88.89 | 28.95 | 0.23 | 2.94 | 1.42 | 0.95 |
| Condition1 | 288.73 | 303.81 | 319.36 | 321.21 | 321.91 | 1484.78 | 90.03 | 28.63 | 0.23 | 2.86 | 1.43 | 0.91 |
| Condition5 | 280.47 | 293.59 | 313.56 | 314.36 | 319.22 | 1585.44 | 99.51 | 28.69 | 0.23 | 2.93 | 1.41 | 0.97 |

*Figure 6 - Statistical Mean*

The statistical median is a measurement of the middle value in a set of data, arranged from smallest to largest. A Comparison for 'Normal1' to 'Condition1/5' will be discussed further in section 3.2.

The statistical median is shown in the figure below:

| Condition | Cylinder heating zone 1 | Cylinder heating zone 2 | Cylinder heating zone 3 | Cylinder heating zone 4 | Cylinder heating zone 5 | Maximum injection pressure | Mould temperature control unit 1 | Cycle time | Injection time | Dosage time | Switch-over volume | Material cushion |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Normal1 | 283.43 | 297.12 | 311.30 | 310.32 | 311.14 | 1547.39 | 88.66 | 28.99 | 0.23 | 2.95 | 1.43 | 0.95 |
| Condition1 | 285.33 | 303.56 | 319.91 | 321.28 | 318.34 | 1490.92 | 90.15 | 28.36 | 0.23 | 2.84 | 1.44 | 0.90 |
| Condition5 | 280.18 | 292.14 | 314.62 | 316.86 | 323.00 | 1602.04 | 99.65 | 28.37 | 0.23 | 2.90 | 1.41 | 0.97 |

*Figure 7 - Statistical Median*

After calculating the mean and median, we can now determine if the data is skewed or not for each parameter under each condition, as illustrated in Figure 10.

The statistical range is a measurement of the difference between the highest the lowest value of the parameter.

The statistical range is shown in the figure below:

| Condition | Cylinder heating zone 1 | Cylinder heating zone 2 | Cylinder heating zone 3 | Cylinder heating zone 4 | Cylinder heating zone 5 | Maximum injection pressure | Mould temperature control unit 1 | Cycle time | Injection time | Dosage time | Switch-over volume | Material cushion |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Normal1** | 33.14 | 33.12 | 35.66 | 35.85 | 35.84 | 222.55 | 10.47 | 3.35 | 0.02 | 0.47 | 0.16 | 0.13 |
| **Condition1** | 31.84 | 35.02 | 36.53 | 35.80 | 36.34 | 198.29 | 10.62 | 3.32 | 0.02 | 0.37 | 0.16 | 0.16 |
| **Condition5** | 28.57 | 31.20 | 31.88 | 36.60 | 35.67 | 229.89 | 11.69 | 3.33 | 0.02 | 0.37 | 0.15 | 0.10 |

*Figure 8 - Range Difference*

The standard deviation is a statistical measure of the spread or variability of data points around the mean.

The statistical standard deviation is shown in the figure below:

| Condition | Cylinder heating zone 1 | Cylinder heating zone 2 | Cylinder heating zone 3 | Cylinder heating zone 4 | Cylinder heating zone 5 | Maximum injection pressure | Mould temperature control unit 1 | Cycle time | Injection time | Dosage time | Switch-over volume | Material cushion |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Normal1** | 10.38 | 8.41 | 9.48 | 10.32 | 11.19 | 60.85 | 2.78 | 0.96 | 0.01 | 0.11 | 0.05 | 0.03 |
| **Condition1** | 9.53 | 10.47 | 11.48 | 11.42 | 10.40 | 60.65 | 3.33 | 1.04 | 0.01 | 0.11 | 0.05 | 0.04 |
| **Condition5** | 8.56 | 10.50 | 8.81 | 11.72 | 12.44 | 65.25 | 3.77 | 1.09 | 0.01 | 0.11 | 0.05 | 0.03 |

*Figure 9 - Standard Deviation*

## 3.2 Discussion (Comparing variables that change across conditions)

Given the statistical values, we implemented Histogram diagram in Figure 10 to determine whether the parameters and condition under normal distribution or skewed distribution. Our analysis shows that within the 12 parameters, the only parameter that isn't skewed is ‘injection time’. The remaining 11 parameters have at least one of the conditions is positive/negative skewed distribution.


*Figure 10 - Skewness Diagram*

With the understanding of the distribution, we decided to use the median to determine the percentage difference between 'Normal1 to Condition1' and 'Normal1 to Condition5'. With the comparison, we have determined that the parameter defer away from "Normal1" with more than 3% are affected by it. By knowing our threshold, we have further observed with data visualization to further analysis the difference, which will be mentioned in part 4. Hence, our threshold is 3%, and our highest difference is 'Mould Temperature Control Unit 1' at 12.4% as shown in Figure 11.

```
Comparison from Normal1 to Condition1

Cylinder heating zone 1: 0.67 %
Cylinder heating zone 2: 2.17 %
Cylinder heating zone 3: 2.77 %
Cylinder heating zone 4: 3.53 %
Cylinder heating zone 5: 2.31 %
Maximum injection pressure: 3.65 %
Mould temperature control unit 1: 1.69 %
Cycle time: 2.19 %
Injection time: 0.0 %
Dosage time: 3.56 %
Switch-over volume: 0.7 %
Material cushion: 4.36 %
```

```
Comparison from Normal1 to Condition5

Cylinder heating zone 1: 1.15 %
Cylinder heating zone 2: 1.67 %
Cylinder heating zone 3: 1.07 %
Cylinder heating zone 4: 2.11 %
Cylinder heating zone 5: 3.81 %
Maximum injection pressure: 3.53 %
Mould temperature control unit 1: 12.4 %
Cycle time: 2.16 %
Injection time: 0.0 %
Dosage time: 1.66 %
Switch-over volume: 1.4 %
Material cushion: 2.32 %
```

*Figure 11 - Comparison between conditions with median*

# 4 Data visualization

## 4.1 Pair plot and line chart on each column

Data visualization makes it possible to clearly analyze the dataset. From Figure 12 (highlighted in black box) the diagonal histogram density, clearly shows that ‘Cylinder heating zone 4’, ‘Cylinder heating zone 5’, ‘Maximum injection pressure’, ‘Mould temperature control unit’, ‘Dosage time’ and ‘Material cushion’ are the parameter that defer away from the condition ‘Normal1’. While the other parameters are overlapping with the condition ‘Normal1’, do not show significant difference from the threshold. With this evidence, we can conclude that the threshold is effective at a level of 3%.

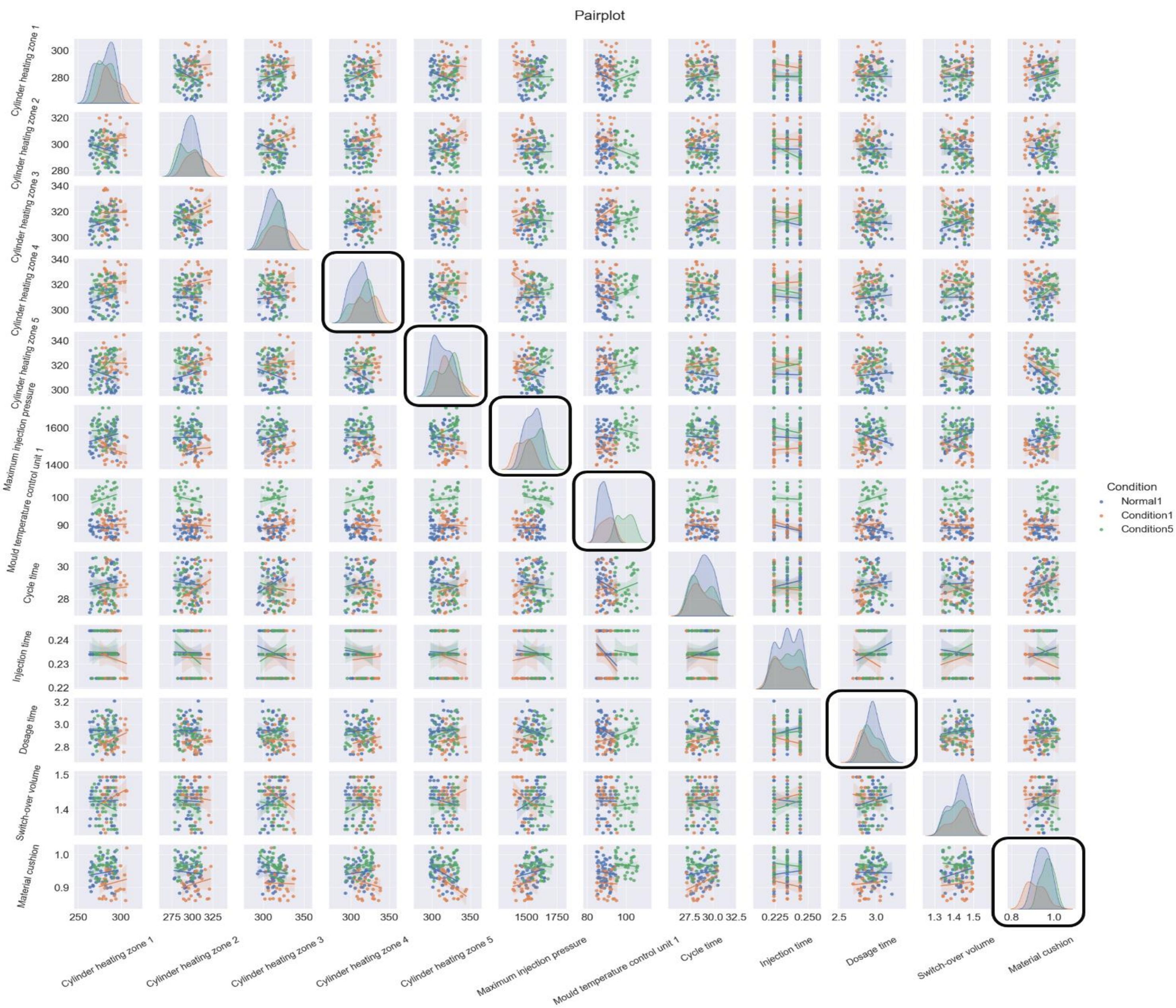


*Figure 12 - Pair Plot*

With the line chart, we could clearly observe that at specific parameters, there is a great difference in median between each condition by observing the median lines. This difference in median implies that the specific parameter is one of the many causes that developed the product quality defect labeled as condition 1 and condition 5. Therefore, this visualizes the calculations done at discussion **3.2**. The line chart for parameters 'Cylinder heating zone 4 and 5', 'Maximum injection pressure', 'Dosage time', 'Material cushion' and 'Mould temperature control unit 1' can be seen in Figure 13. From the characteristics of the median line, we could observe that the comparison between Normal 1 and Condition 5, we could see at parameters **'Mould temperature control unit 1' and 'Cylinder heating zone 5'** there is significant jumps in the median. Also, the comparison between Normal 1 and Condition 1, we could see at parameters '**Cylinder heating zone 4', 'Maximum injection pressure' and 'Material cushion**', there is also significant jumps in the median. In addition, we could imply that the respective parameters from the observations made has influence in its conditions. However, the all other parameters do not produce any visual significance in median similar to the previously noted parameters, therefore supplementary analysis techniques is needed to further minimize our top 4 crucial factors that influence the quality of the final result.

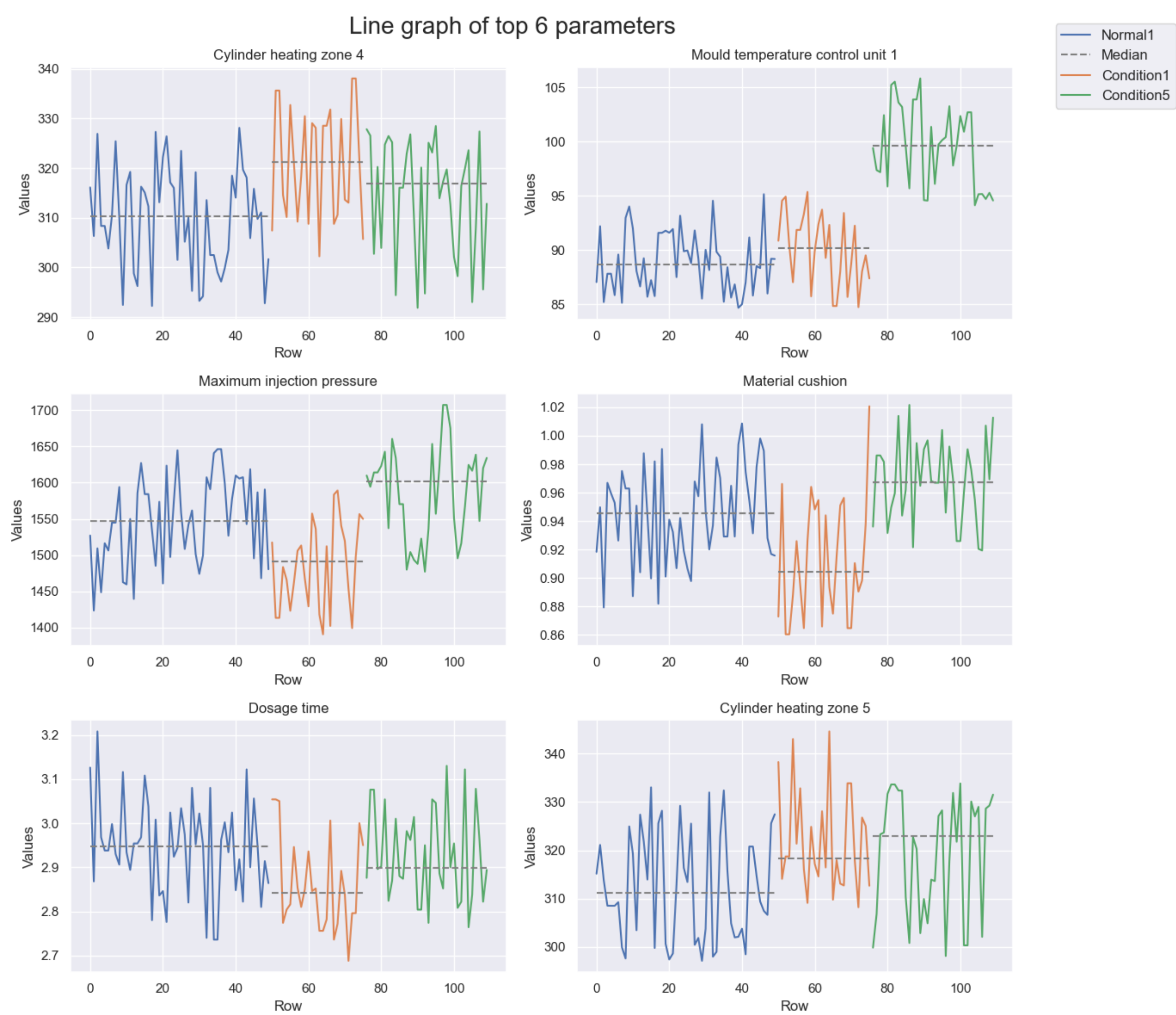


*Figure 13 - Line chart*

## 4.2 Heat Map

With the understanding of classification and regression, the correlation analysis has encountered an issue with classification on the dependent variables: Normal1, Condition1 and Condition5. To solve the issue, we converted the data type 'string' into 'float or integer' and created a correlation matrix (heat map) as shown in Figure 14. This correlation matrix represents the relationship between parameters in a dataset. In the correlation matrix, the cells are color-coded based on their relationship, the darker the color indicates a stronger correlation or the number closest toward 1 or –1. The parameter "Moulding temperature control unit" has the highest correlation of 0.77 which justifies the analysis mentioned at discussion **3.2** with the value of 12.4% comparison.

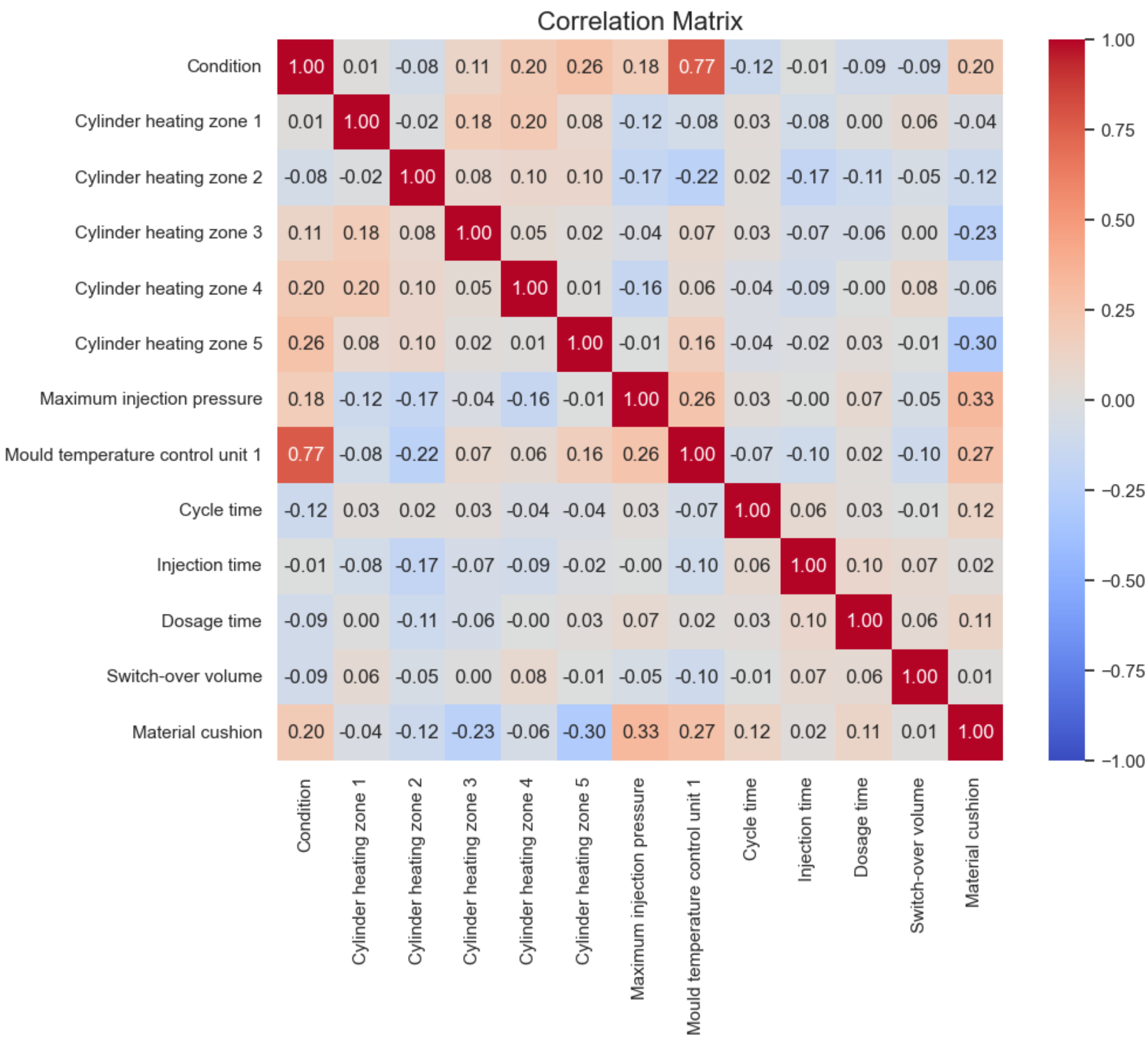


*Figure 14 - Correlation matrix*

With the help of the correlation analysis, parameters that were previously difficult to identify on a line graph are further removed and to double check the correlation matrix. As shown in Figure 15, the correlation analysis is shown in descending order, however, we only take the top 4 parameters into account. With such evidence, we can prove that “Mould temperature control unit 1” has the strongest positive correlation at a value of 0.77.

```
Mould temperature control unit 1    0.770190
Cylinder heating zone 5             0.264772
Material cushion                    0.202276
Cylinder heating zone 4             0.195379
Maximum injection pressure          0.181182
Cycle time                          0.119465
Cylinder heating zone 3             0.114261
Dosage time                         0.091194
Switch-over volume                  0.089220
Cylinder heating zone 2             0.077315
Cylinder heating zone 1             0.013379
Injection time                      0.007206
```

*Figure 15 - Rank of correlation parameters*

Lastly, we can also use pair plot to show a clear visualization to explore the relationship between parameters and condition or parameters and parameters. Figure 16 clearly demonstrates the steep or shallow slope, which denotes a strong or weak linear relationship. A stronger correlation indicates that the metric is having an impact on the product’s quality. With all these findings above, our top 4 parameters consist of “Mould temperature control unit 1, Cylinder heating zone 5, Material Cushion, Cylinder heating zone 4”.

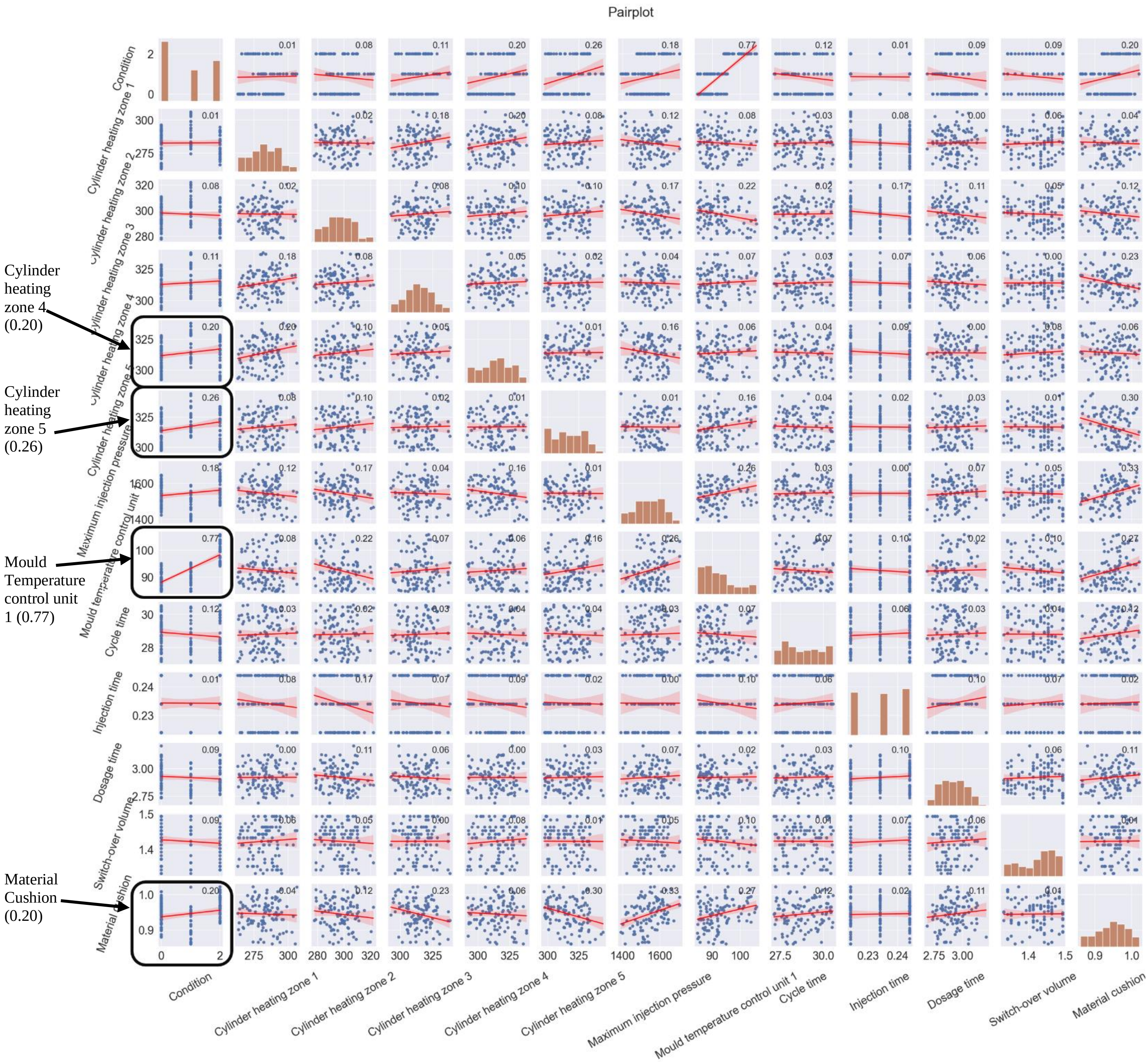


*Figure 16 - Pair plot with conditions*

However, for further investigation of our data by using a heat map using the top 4 parameters correlated to conditions, we have found that there are no good correlations between each of the parameters which can be shown in Figure 17. According to the Rough Rule of Thumb, the most positive correlation from the heat map is +0.27 which is categorized as not correlated and the most negative correlation being -0.3 which is also categorized as not correlated. Hence, we are unable to narrow down the parameters and must observe all four parameters.

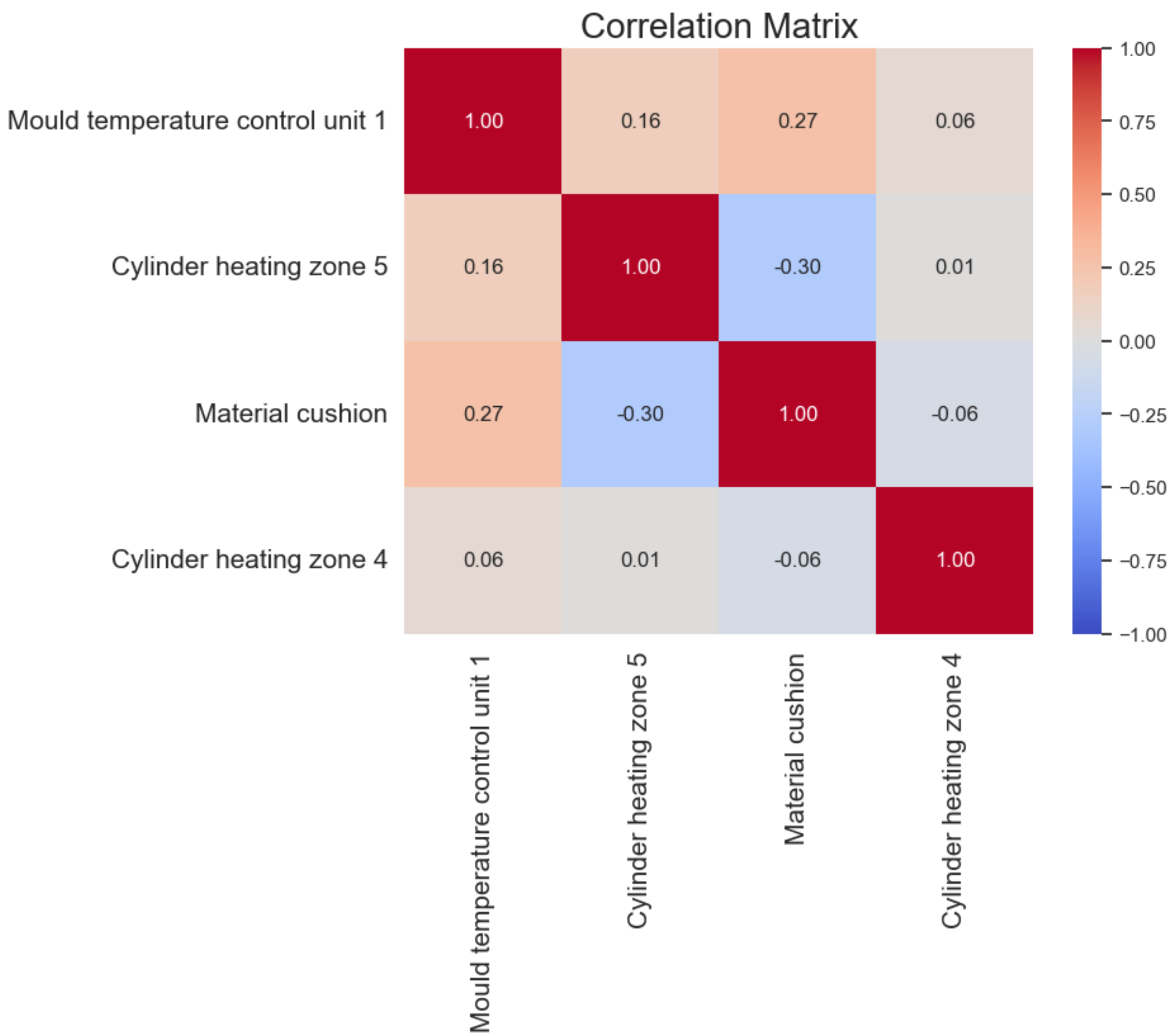


*Figure 17 - Top 4 parameter heatmap*

## 4.3 Discussion (Comparing variables that change across conditions)

After making thorough observations of the dataset using visualization methods like line chart, heat map and Pair plot, we could identify the relationships between process parameters and product quality. From the line chart, we could assess that the underlying parameter that caused an issue in product quality of Condition 5 was “Cylinder heating zone 5”, “Maximum injection pressure” and the parameter that was highly affected “Mould Temperature Control Unit 1”. Therefore, in the future, we could easily detect for Condition 5 by monitoring “Mould Temperature Control Unit 1” as it has the highest positive correlation with the Conditions, and it also will have an average of 12.4% increase in median whenever Condition 5 is found. This could also be said for Condition 1. However, for condition 1, the parameter that is highly affected is the “Material cushion”.

# 5 Discussion and conclusion

When comparing Condition 1 and Condition 5 with Normal 1, we found that Condition 5 has the most significant change in the moulding machine quality. Although Condition 1 has changes that may affect the quality, those differences are not too far from Normal 1 as compared to Condition 5. Therefore, with the data analytics tasks that we have done, we could clearly denote that “Mould Temperature Control Unit 1” has the strongest relationship to product quality. However, observing at Figure 18, since there is no significant dominance of correlation, it is important that we could not only observe “Mould Temperature Control Unit 1” for quality control. Hence, there is a need to observe the 4 parameters “Cylinder heating zone 5”, “Maximum injection pressure”, “Material Cushion” and “Mould Temperature Control Unit 1” in order to maintain a stable and normal quality mould injection process.

Addressing the issue on the influence of missing data, there are a few reasons why this could happen:

Firstly, if the data collection is being done manually, human errors may occur. For example, the operators or technicians may miss recording a certain measurement or when entering data into the spreadsheet, typographical errors can lead to missing values. Secondly, if the data is being transmitted automatically from the injection moulding machine to the data storage system, interferences in system during the data transmission can result in missing data or if the sensors fail to record a particular measurement, it could also result in null values.

Therefore, a few of the solutions to this problem:

Collect more data, as having more data may outweigh the significance of the null values and just drop the recordings with null values. However, it is essential that this does not substantially reduce the sample size too much so that the analysis loses statistical power.

# 6 References

[1] The outline of injection molding, https://www.polyplastics.com/en/support/mold/outline/index.html (accessed Sep. 21, 2023).

[2] J. Schwartz, “Understanding the benefits of injection molding process control,” RevPart, https://revpart.com/injection-molding-process-control/ (accessed Sep. 21, 2023).