Some conclusion(In each csv file, blue color stands for outliers and red stand for important values):

1. Main status which have no corresponding machine: 32,33,39,40,201,203,205.

|  |  |
| --- | --- |
| Status\_pair | Time(average,max,min) |
| 200,201 | (72.19,224,1) |
| 202,203 | (30.69,99,1) |
| 204,205 | (7.78, 30, 1) |
| 3,4 | (6.75, 23, 1) |
| 5,6 | (4.21,14,1) |
| 430,431 | (3.58, 12, 1) |
| 490,491 | (7.62, 21, 2) |
| 480,481 | (3.60, 12, 1) |
| 310,311 | (2.59, 8, 1) |
| 444,445 | (14.43, 48, 1) |
| 366,367 | (2.81, 7, 1) |
| 502,503 | (6.37, 22, 1) |
| 488,489 | (9.45, 27, 2) |
| 424,425 | (3, 7, 1) |
| 506,507 | (6.52, 20, 1) |
| 344,345 | (7, 17,2) |
| 372,373 | (39.36, 123, 7) |
| 478,479 | (22.5, 45, 6) |
| 12,13 | (4.31, 14, 1) |

1. Drop all the outliers and count the average.
2. N00090 can do all the status in orange color but it will usually take more time.
3. T00010 and T00009 are two same type of machine and T00010 is faster.
4. As for the correlation coefficient of “l\_size” and “d\_size” is 0.517, which means that the two are significantly positively correlated.
5. Average “d\_size” and “l\_size” on each machine corresponds to what have been described in the machine\_processing\_time.xlsx.
6. In mac1.csv, you can see the trend that some machine tend to process with large size of product and vice versa.

Future work:

Machine learning in choosing specific machine:

Input: product\_id(d\_size,l\_size)

Output: mac\_cd1, mac\_cd2

Then combine it to status\_sequence.