**Pre-processing**

**Introduction**

I have started to research log analysis methods after checked the log csv file, documents and requirements.

After carefully researching the methods and languages for log analysis, I have decided to work with Python to analyse the logs. Because of Python more useful for these types of analysis. It could be done by using SQL or Excel but It would be more complex to parse. I thought that It'd be unnecessary complexity so I tried to answer the questions by using Python.

**Data Cleaning**

In this phase, I chose focusing to the task questions rather than creating specific and meaningful table (cleaning/parsing) from log.csv.

**Data Transformation**

I have started by importing the data and used some parsing, converting, filtering, grouping, regex exp methods for to complete the questions. My approach was converting to the timestamp column to the right format and picking right data from the log message column.

**Task Questions**

* I have noted and explained my steps in py scripts.
* I have successfully done the first two task. Third task will be added later.

**Answers of the Questions**

1. How many products are assembled per shift?

* Important parts: Changing timezone, finding and filtering according to the shift, finding right message for assemblyed product, grouping by Shift, counting

|  |  |
| --- | --- |
| Shifts | Assemblyed Product |
| 24.08.20 | 2 |
| 25.08.20 | 14 |
| 26.08.20 | 10 |

1. What’s the average cycle time?

* Important parts: writing a function which is finding ‘whole cycle seconds’ for each the assembled product, extracting seconds information from the message column, calculating average
* I haven’t added the filter for working hour due to the question. It wasn’t written as per a shift that is why, I haven’t filtered.

|  |  |  |
| --- | --- | --- |
| Sum of Seconds | Count of Assembled Product | Average Cycle Time |
| 78063 | 63 | 1239.09 |

1. How much time does the application spend in each work phase?

My approach was writing a function which extract the seconds information from rows according to each phase, after joined work phase txt file. Third task will be added later.