Business & Data Understanding :

Business understanding :-

* Wafer production process
* Tool state features and process state features
* Target Variable (status) , input variables (features)
* Analysis (Pros and Cons ) of planned Models. For example, the models using distance based correlation between features need feature scaling in order to normalize features with different datatypes
* Literature review
* 2 papers and the reference paper
* Cost benefit analysis of the plans (on the basis of the technique we will use )
* Flow chart on the work flow of the project (Plan for the project)

Data Understanding:

-> Descriptive Statistics

-> EDA (data types for all columns , outlier, missing, duplicatesetc.  
> histograms

| Data Quality | Descriptive | Exploratory | Visualisation |
| --- | --- | --- | --- |
| Drop duplication | Mean, median, Std | Features are important  590 -> 23 | Boxplot of unexpected features |
| Missing values | Outlier analysis | Correlation | Histograms, Boxplot |
|  |  |  |  |

Team Collaboration and Documentation

* Install Vs studio
* Github Repository setup
* Literature reading and search about SECOM business
* Building up a presentation template

See the snippet below , May be these step from Moodle will be helpful

