

**0% complete**

0% complete

0% complete

60% complete

this is complete

80% complete

this is complete

**30% complete**

0% complete

**0% complete**

this is complete

**88% complete**

**this is complete**

1. Activities and Schedule

The following are the main project activities, including five main components: understanding the dataset, literature review, data visualization and preprocessing, exploring data analysis, modelling and algorithms, and communicating the report, respectively.

**-Understanding the dataset (wk1)**

*Acquisition of the dataset and understanding the representation of different variables.*

**-Literature review (wk2-wk8)**

*Searching the literature is conducted during the whole project—those important references keep providing creative ideas to work.*

**-Data visualization and preprocessing (wk2-wk5)**

*This is the most time-consumption part of our project. We aim to test the accuracy of different imputations methods to obtain the entire dataset without missing values. It includes the following steps:*

1. Dealing with the invalid data (wk1)
2. Distribution of response variables (wk1 to wk4)
3. Handling the missing data and outliers (wk2 to wk5)
4. Plot the variable correlation diagram (wk1 to wk3)
5. Observation and corresponding data manipulation (wk2 to wk5)

**-Explore data analysis, modelling and algorithms (wk6-wk8)**

*Selecting the related modelling and tuning parameters to increase accuracy. Choose the fittest models after the assessment.*

1. Model selecting and fitting (wk6-wk7)
2. Tune parameters (wk6-wk8)
3. Diagnostic and model assessment (wk7-wk8)

**-Communicate visualize report (wk5-wk9)**