Leyan:

Read some research papers about bladder cancer

Have got an understanding of the data variables from the dataset provided by Gorden

Learned intermediate Python on DataCamp and get ready for project coding

Built an unit test module template for our project

Created a comparison table to display how K-mean clustering and hierarchical clustering works, what’s the difference between these two methods. Provided this comparison table to our team, so we can better understand these two methods.

Preparing Python code to express the different statistic models (on going)

Questions/need to be discussed with the team:

Need to figure out a better way to handle “N/A” values and “uncertain” values for the same variable?

Since we will use K mean clustering model, how do we define K?