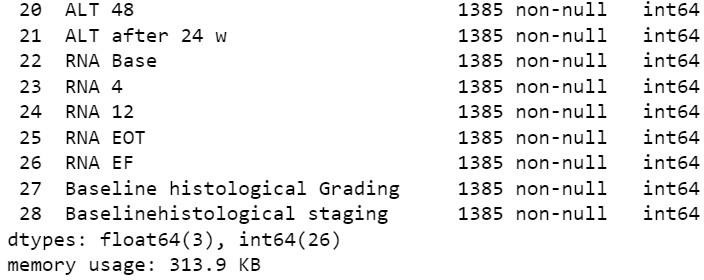
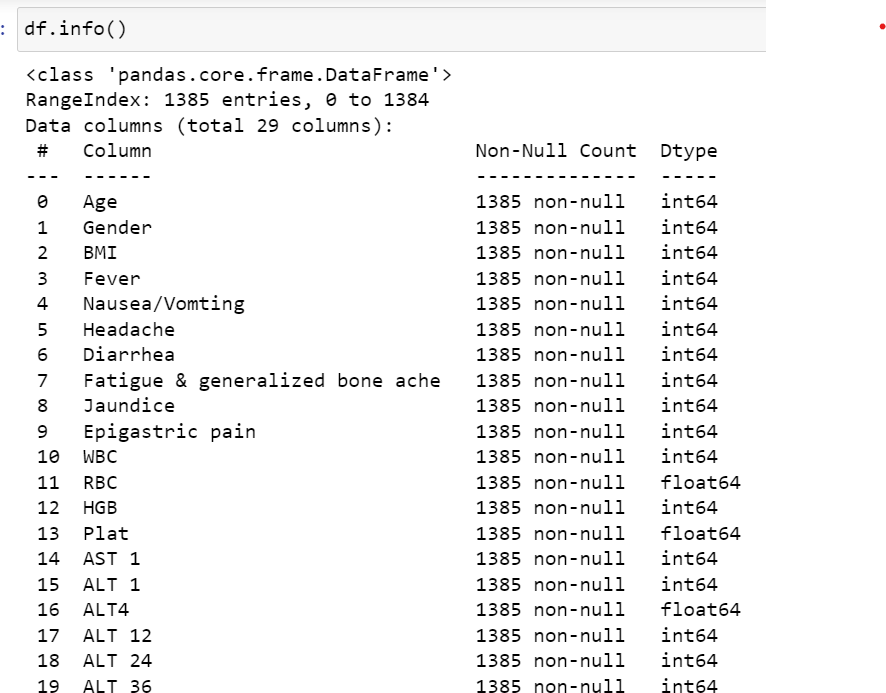
 

Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 15 July 2024 |
| Team ID | 739716 |
| Project Title | Predicting Baseline Histological staging in HCV patients using ML |
| Maximum Marks | 6 Marks |

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data  Overview | Now that we know the nature of the data, let's preprocess the information that was gathered.  The downloaded data set may contain too much randomness to be used for training a  machine learning model, so to get good results, the dataset must be carefully cleaned. The following steps are involved in this activity.  ? Handling missing values  ? Handling categorical data  There are no categorical variables in our datasets  The general procedures for pre-processing data before applying it to machine learning are as follows. Your dataset's state will determine whether or not you need to follow each of  these stages.  The first step will be to find the shape i.e. dimensions of the dataset. To find the shape of our data, the df. shape method is used. To find the data type, the df.info() function is used.    Thus, our dataset contains 1385 rows and 29 columns. |



From the above, we can see that there are no null values in this dataset.

Data exploration and preprocessing report

Data set variables will be stastically analyzed to identify patterns and outliers with python employed for preprocessing tasks like normalization and feature engineering.data cleaning will address missing values and outliers,ensuring quality for subsequent analysis and

modelling and forming a strong foundation for insights and predictions. images will be