AIM(Problem try to reduce)-

The data consists of 10,000 observations of space taken by the SDSS. Every observation is described by 17 feature columns and 1 class column which identifies it to be either a star, galaxy or quasar. The main purpose of that dataset is to make these scattered dataset in proper understandable form. The SDSS are used for exposing and developing photographic plates—a slow and laborious process—and the data, owned by the scientists who gathered it, was difficult for others to access. Images were taken using a [photometric system](https://en.wikipedia.org/wiki/Photometric_system) of five filters (named *u*, *g*, *r*, *i* and *z*). These images are processed to produce lists of objects observed and various parameters. For imaging observations, the SDSS telescope used the [drift scanning](https://en.wikipedia.org/wiki/Drift-scanning) technique, which tracks the telescope along a [great circle](https://en.wikipedia.org/wiki/Great_circle) on the sky and continuously records small strips of the sky.[[8]](https://en.wikipedia.org/wiki/Sloan_Digital_Sky_Survey#cite_note-8)

So our aim is to make the SDSS best possible way to understandable and make their feature to use and understand identifies it to be either star galaxy or quasar.

Analysis-

The class identifies an object to be either a galaxy, star or quasar. SDSS maps of galaxies, quasars, and intergalactic gas have become a central pillar for tests of the standard cosmological model that describes our understanding of the history and future of the Universe. SDSS data have helped to demonstrate that the Universe is dominated by unseen dark matter and pervasive dark energy, and seeded with structure by quantum fluctuations in the infant cosmos. Mapping of streams of stars left from galaxy mergers in the Milky Way, as well as the discovery of many new dwarf companion galaxies of the Milky Way. and we work according to target(Class) feature with others.

Working with dataset-

The construction work of these dataset is that first we find out the shape, columns. convert variable in required changes which are used to work and also understand the types of feature. And we find out the null vales it is not their and we understand also null values by use of heatmap. The EDA Process is included their first we find the correlation between variable and it is also understand by heatmap. Then the columns variables are correlated between that the values(numbers,colours) are used to understand properly then the target variable are shows the diffrences between the varible. after that we convert category of target variable. Then we check all the values by their graphs of variable and drop the some column and see the type of variable. Then we perform train and test split, decision tree classifier , Prediction on test data accuracy. That are the construction work for data set which are included prediction, working ,construction of dataset.

Concluding Remark-

The conclusion is that which are understand the prediction and train - test the dataset then their the regression(logistic) are also understandable to find out the best possible dataset which include the minimum loss of dataset and give most probable result. The class variable star, galaxy and type of logistic regression are included and their result .