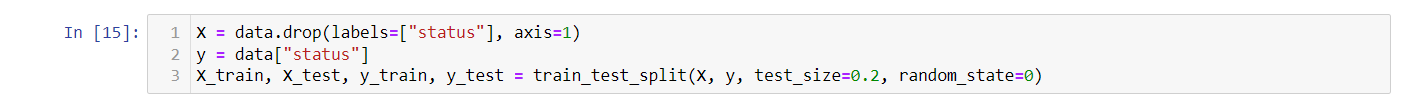
***Pipelining and Modeling***

**PREPARATION:**

After data cleaning and transformation, the next step in a data science project is to create a machine learning model, but as we proceed there are some steps in which we prepare dataset before we train a model on that dataset.

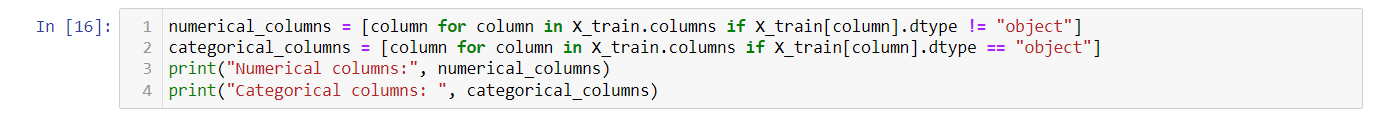
***Splitting dataset:***

Train Test Split. Keeping aside twenty percent of the data for test set.



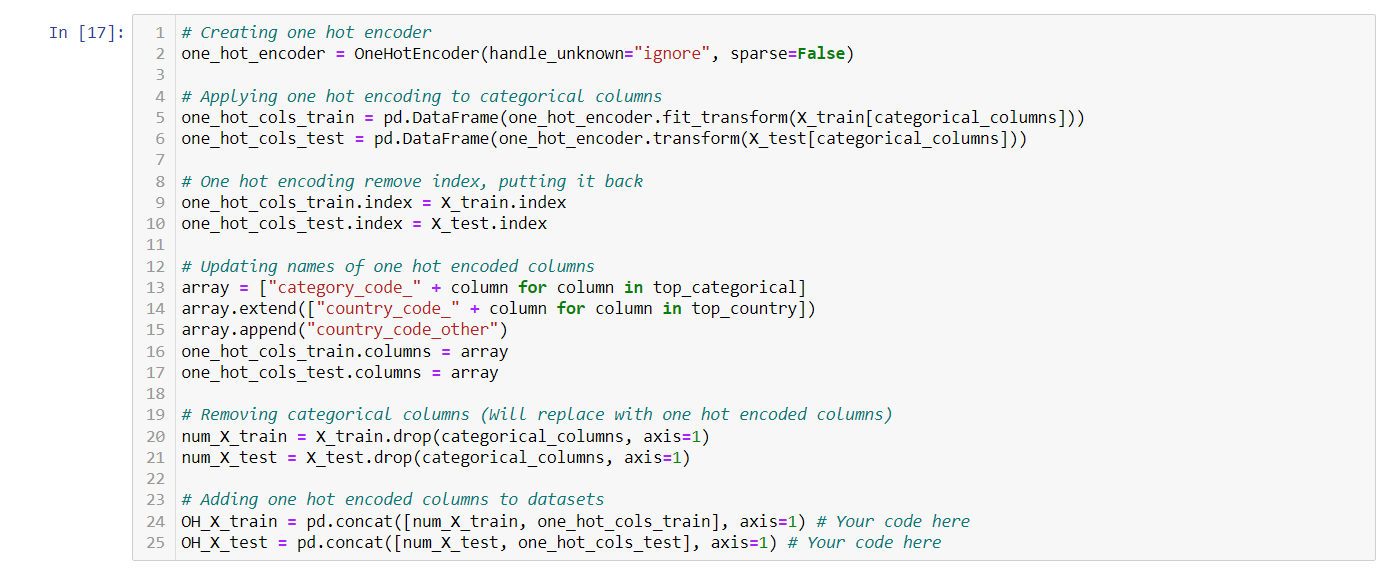
***Separating numerical and categorical columns:***

Separating numerical and categorical columns for processing on them separately.

****

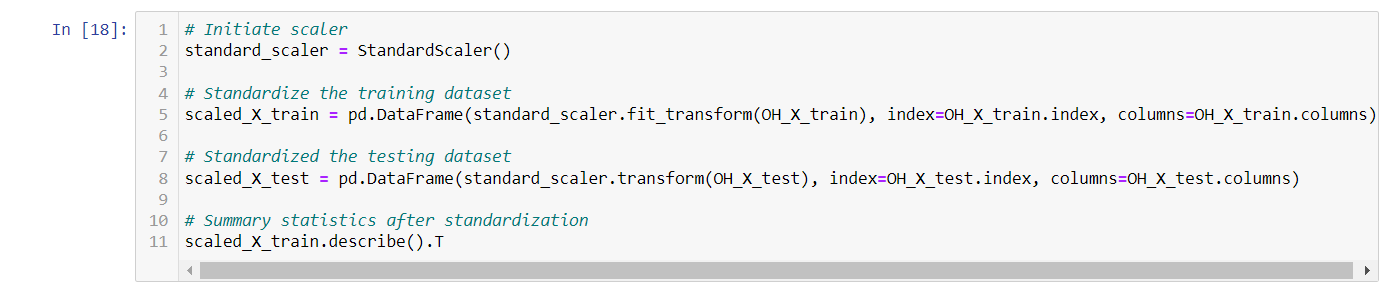
***One-Hot Encoding (Categorical Columns):***

One hot encoding is applied to all categorical columns, separated in previous step.



***Data Scaling (Numerical Columns):***

Numerical columns are scaled using standard scalar.

****

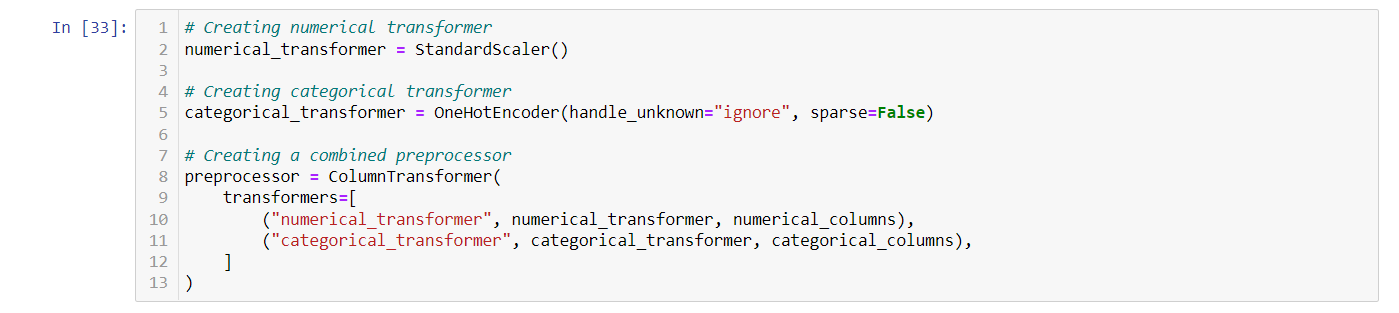
After passing data through above steps now we have final data on which we can train our model.

**PIPELINES CREATION:**

In data preparation steps explained above, we have done some tasks manually such as encoding of categorical columns and scaling of numerical columns. In some scenarios we have to apply more than one transformation to target columns, so we can not do all manually. We can make transformers for numerical and categorical columns so we can process them in a pipeline.

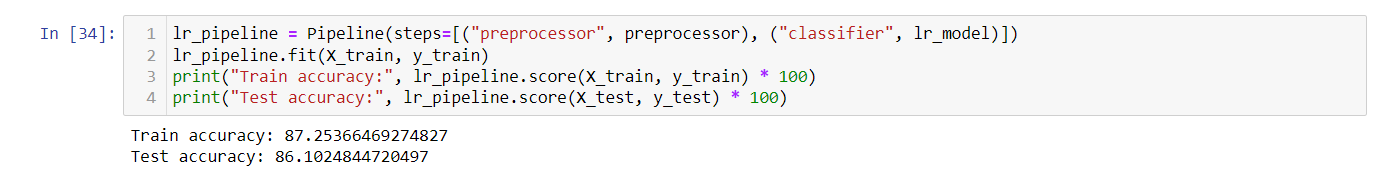
***Transformers for Numerical and Categorical columns:***

Two transformers are created for numerical and categorical columns separately and then a combined preprocessor is created for passing in a pipeline.

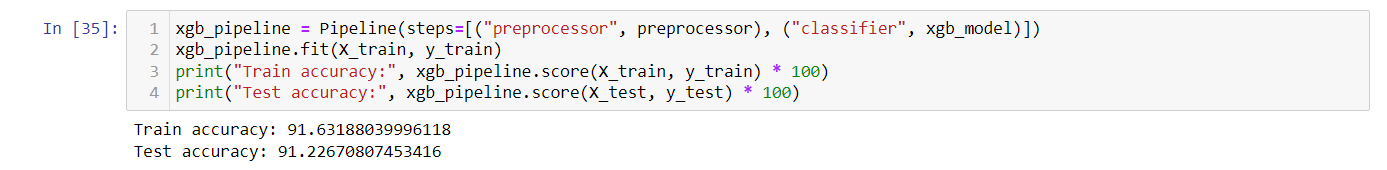


Following are the pipelines created for different models implemented after hyperparameter tuning:

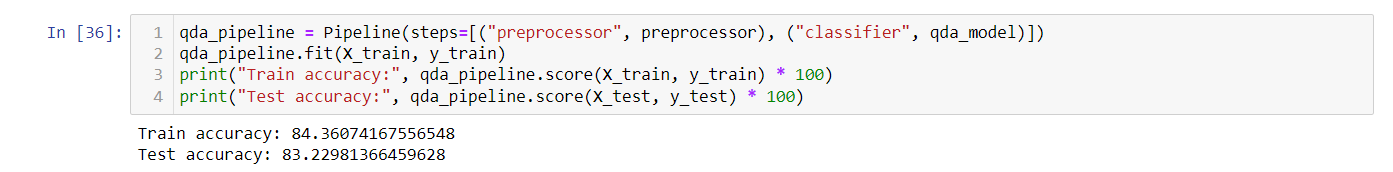
***Logistic Regression Pipeline:***



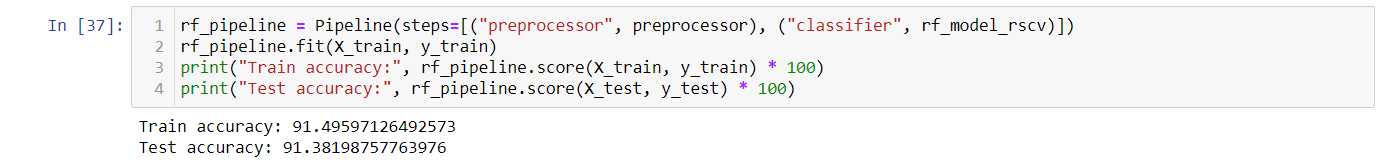
***XGBoost Classifier Pipeline:***

******

***Quadratic Discriminant Analysis Pipeline:***

******

***Random Forest Classifier Pipeline:***

******