****

**Superior University Lahore**

***Lab Task # 2***

**Name:** Muhammad Abdullah Kashif

**Roll No**: SU92-BSAIM-F23-070

**Instructor**: Sir Rasikh Ali

# Course: Programming for Artificial Intelligence (Lab)

Spaceship Titanic

**Passenger Transport Prediction Dataset Analysis**

### ****1. Introduction****

The passenger transport prediction dataset is used to determine whether a passenger was transported or not. It contains details about passengers, their living conditions, expenses, and travel destinations. The goal of this analysis is to explore, clean, and model the dataset to predict the Transported variable effectively.

### ****2. Dataset Description****

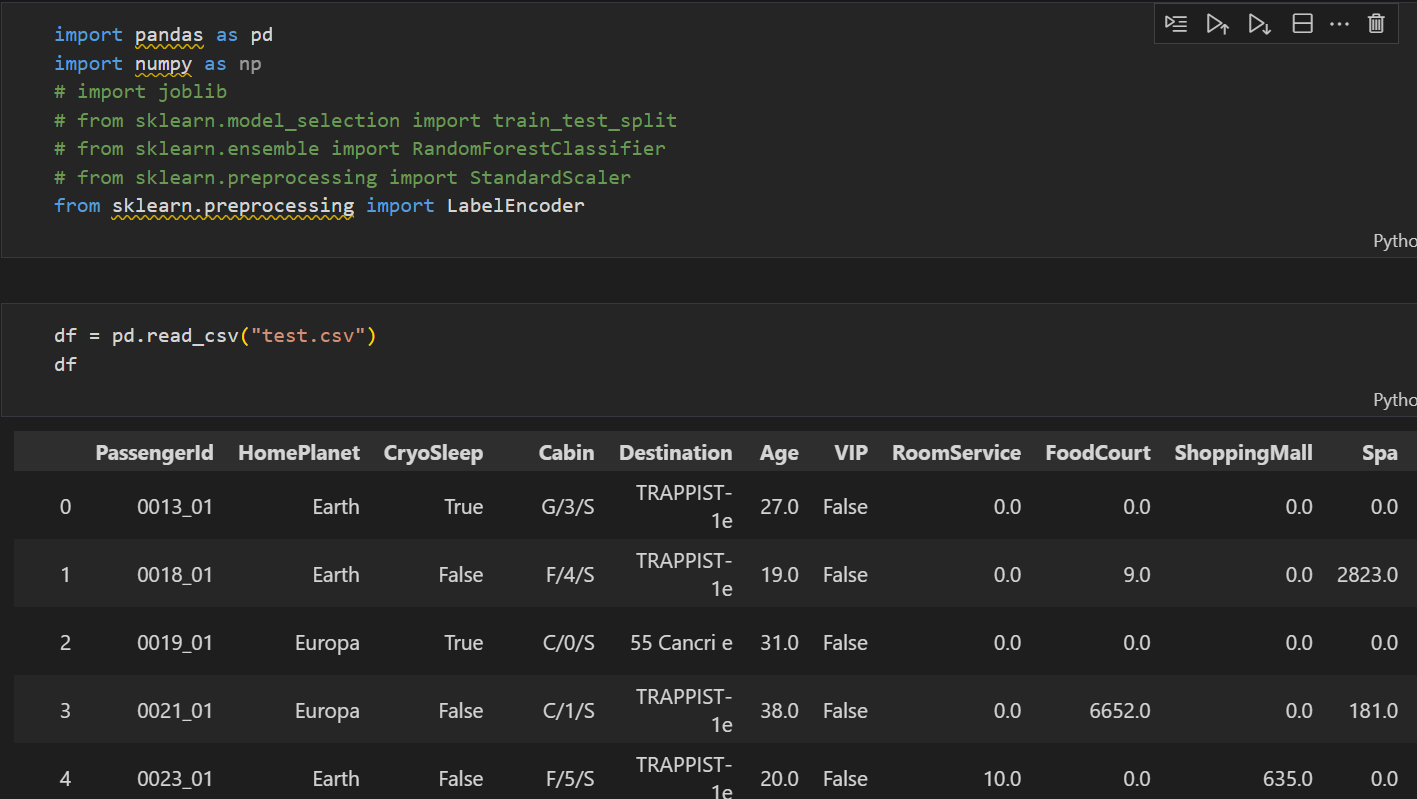
The dataset consists of **8,693 training samples** and **4,277 test samples** with **14 features**. The target variable, Transported, indicates whether a passenger was successfully transported. The key features include:

* **Passenger Details**: PassengerId, HomePlanet, Name
* **Travel Information**: Cabin, Destination, CryoSleep, VIP
* **Financial Features**: RoomService, FoodCourt, ShoppingMall, Spa, VRDeck
* **Demographic Feature**: Age

### ****3. Data Loading & Exploration****

**3.1 Loading the Data**

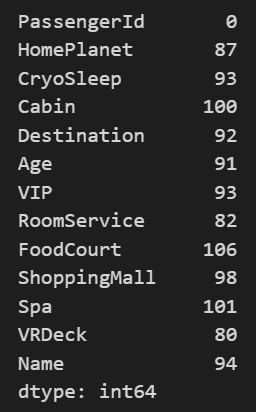
The dataset is loaded using the Pandas library. The training data contains passenger attributes along with the target variable, while the test set excludes Transported.



**3.2 Checking Missing Values**

Several features contain missing values, including:

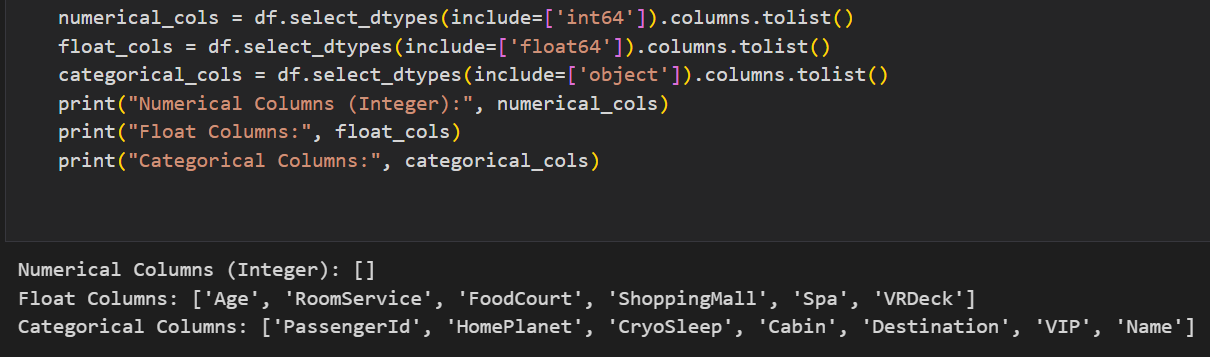
* HomePlanet, CryoSleep, Cabin, and Destination: Missing values suggest incomplete passenger records.
* RoomService, FoodCourt, ShoppingMall, Spa, VRDeck: Missing values might indicate passengers who didn’t spend money on these services.
* Age: Missing values could be imputed using the median or mean age of other passengers.

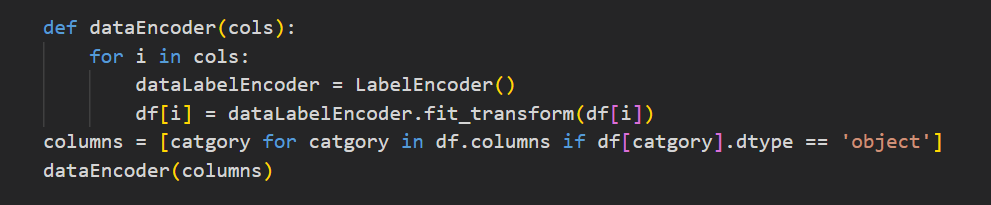


### ****4. Data Preprocessing****

**4.1 Handling Categorical Variables**

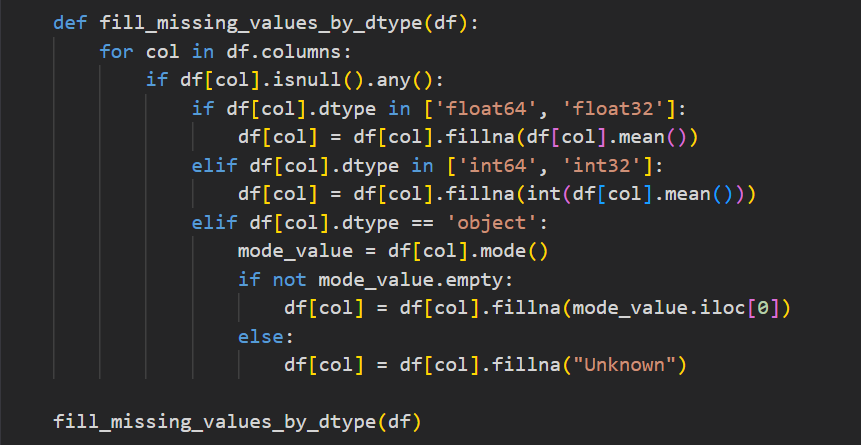
* Features like HomePlanet, Destination, and Cabin are converted into numerical values using one-hot encoding.
* Boolean features (CryoSleep, VIP) are mapped to binary values (0,1).





**4.2 Handling Missing Values**

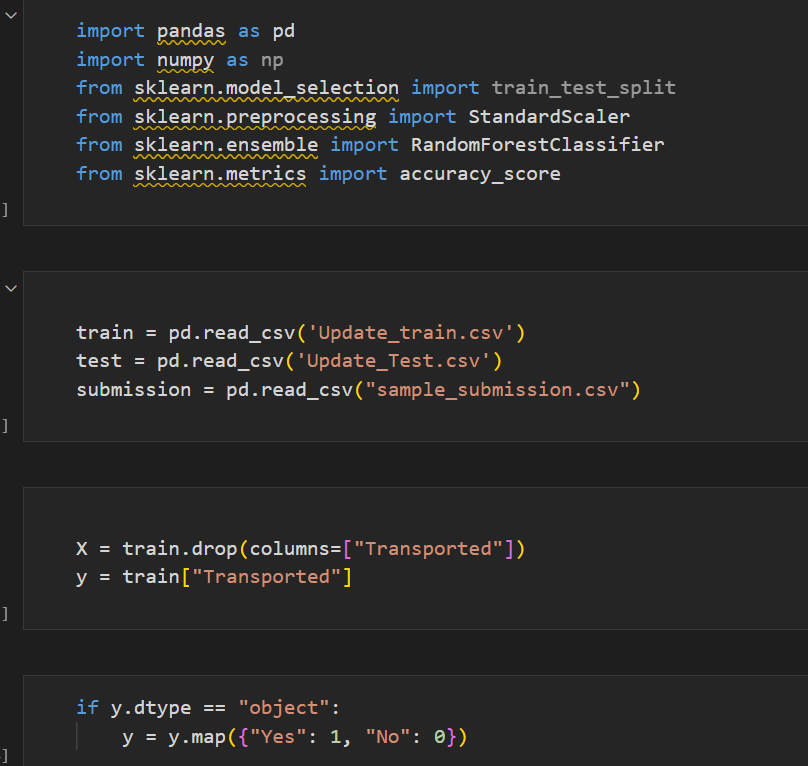
* HomePlanet, Destination, and Cabin are filled based on common values.
* Financial attributes are filled with zeros, assuming missing values indicate no spending.

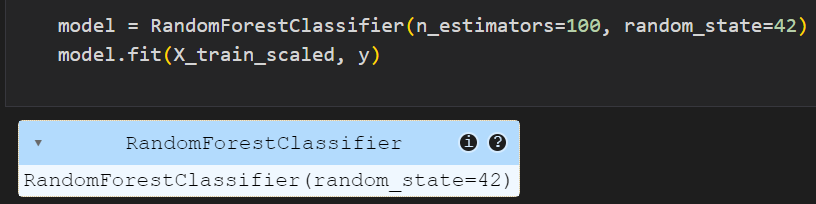


### ****5. Model Implementation****

Various classification models will be tested for passenger transport prediction:

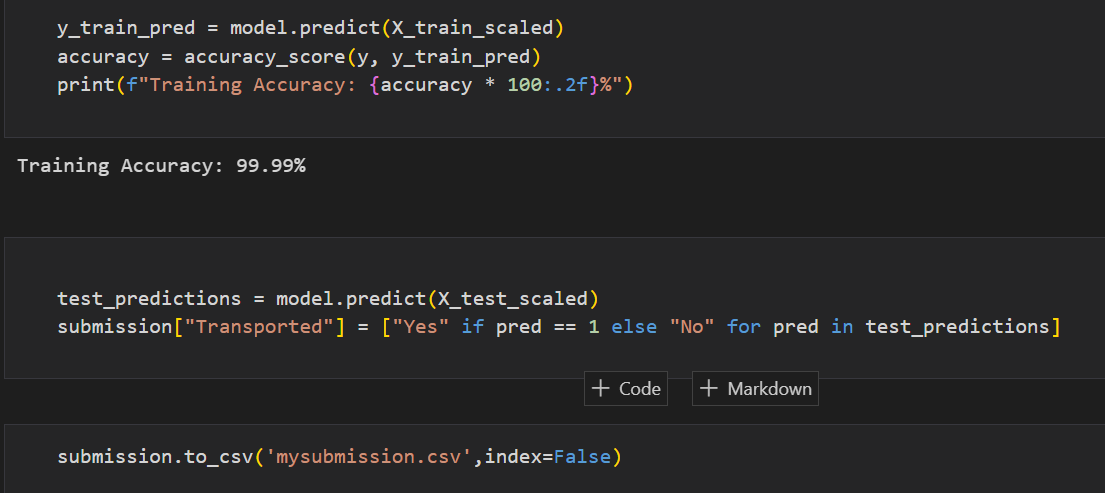
* **Random Forest :** To capture non-linear relationships.





### ****6. Model Evaluation****

The models will be evaluated using:



**7. Conclusion**

This analysis provides an in-depth exploration of the dataset, missing values handling, feature engineering, and modeling approaches. The best-performing model will be chosen based on predictive accuracy and stability to determine passenger transport outcomes effectively.