**Assignment 3:**

**Installation**

1. To install all the relevant python packages, use the command on the command prompt as ‘pip install –r requirements.txt’ after extracting the zip folder
2. To run the file, on a command prompt type ‘python a3.py’

**Explanation**

1. Load the Training and testing data

train\_original = pd.read\_csv('train\_\_titanic.csv')  
test\_original = pd.read\_csv('test\_titanic.csv')

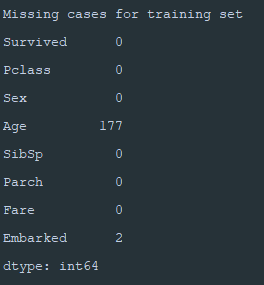
1. Exclude some features to reduce data dimensionality

*# Exclude some features to reduce data dimension*train = train\_original.drop(['PassengerId', 'Name', 'Ticket', 'Cabin'], axis=1)  
test = test\_original.drop(['PassengerId', 'Name', 'Ticket', 'Cabin'], axis=1)  
total = [train, test]

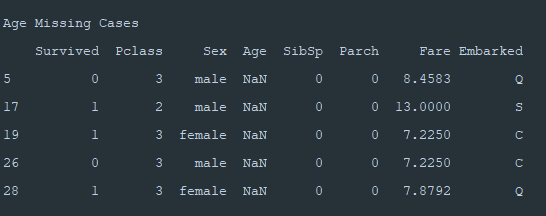
1. Check the shape of the data



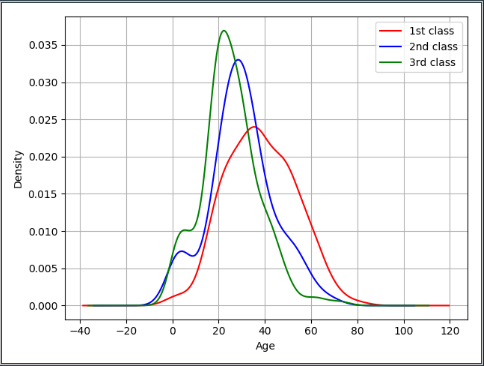
1. Get Missing cases for training set



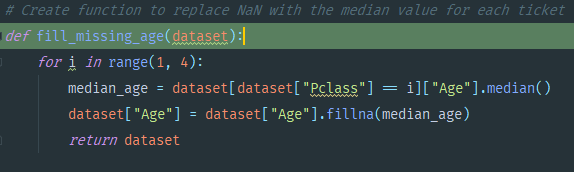
1. Get Age Missing cases



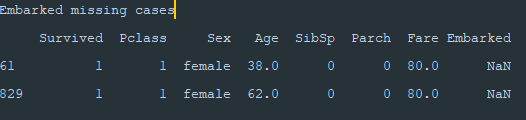
1. Distribution of Age Missing set

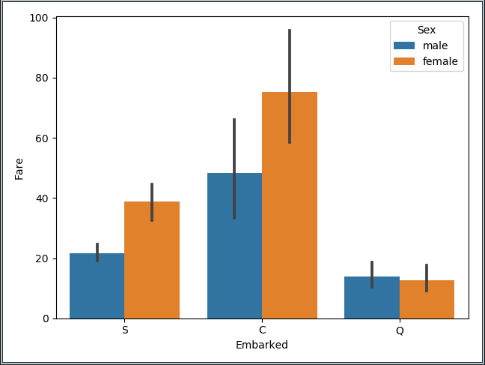


1. Fill the missing age

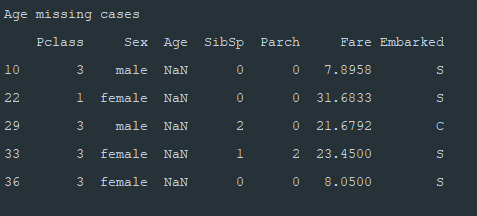


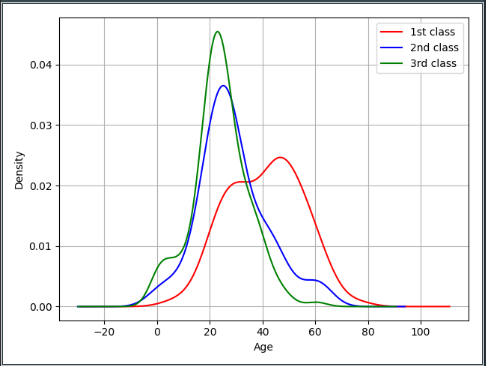
1. Get journey embarked missing cases

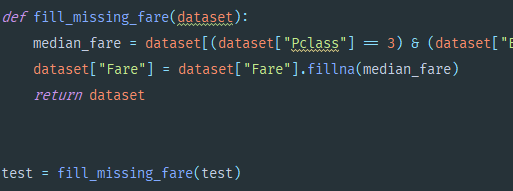


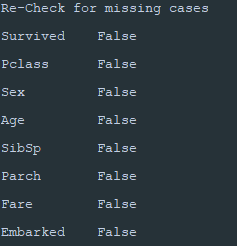


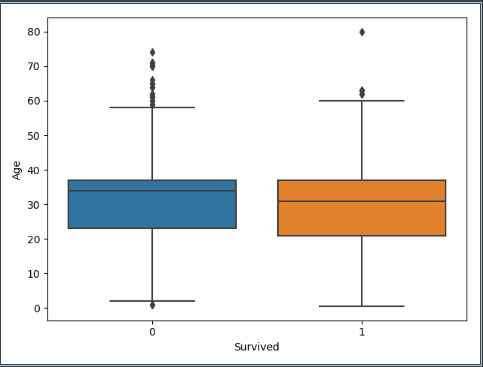
1. Considering Sex=female and Fare=80, Ports of Embarkation (Embarked) for two missing cases can be assumed to be Cherbourg (C).
2. Get missing and impure data and triage them in training dataset



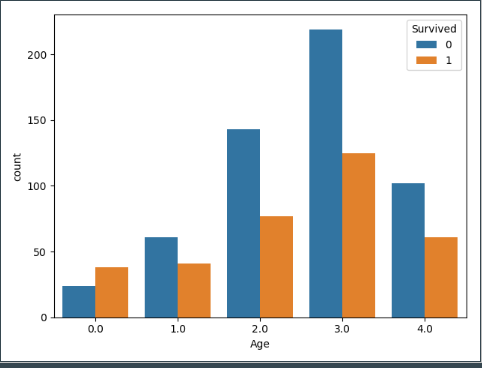




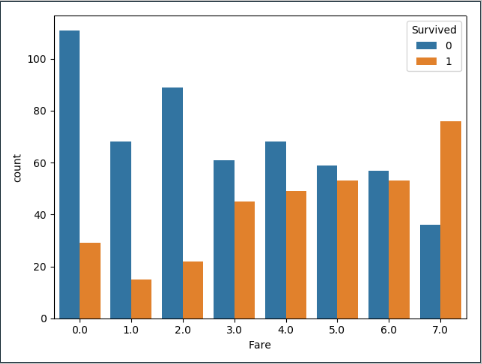




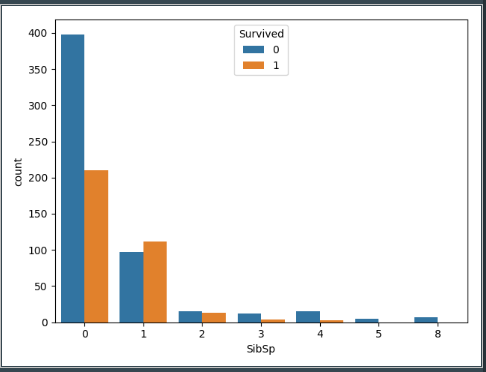
1. Discretize Age Feature



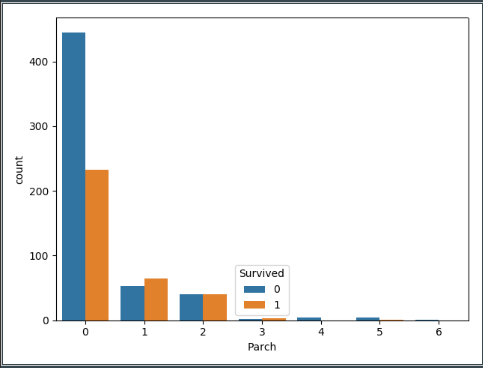
1. Discretize Fare



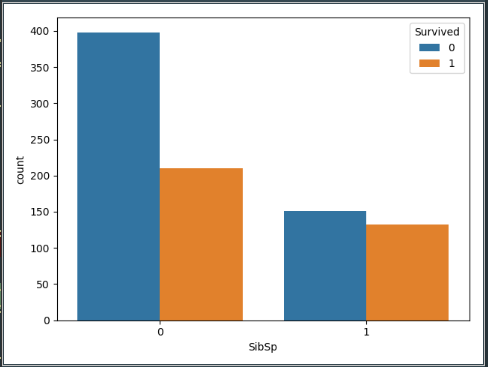
1. Countplot for the number of siblings/spouse



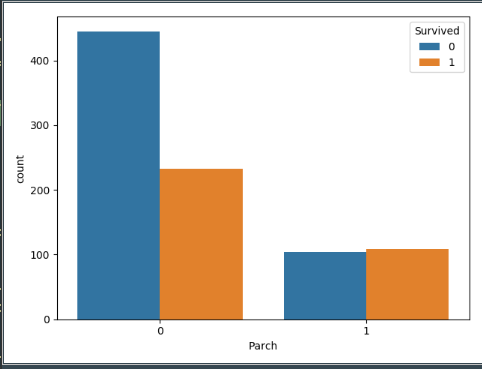
1. Countplot for number of parents/siblings



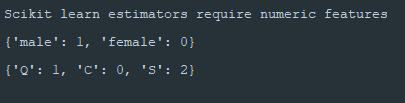
1. Convert SIbSp to binary feature

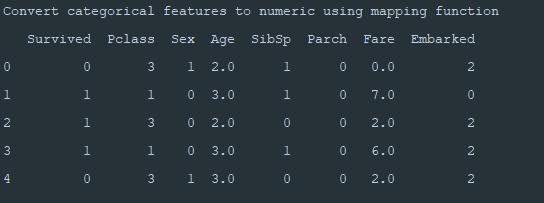


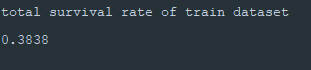
1. Convert Parch into Binary Feature



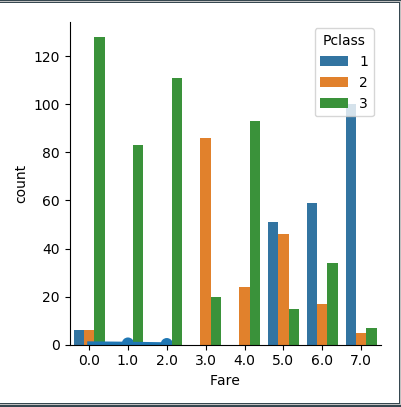
1. Declare Scikit learn estimators



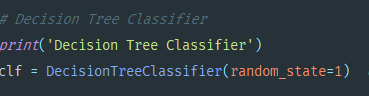
1. 
2. Get total survival rate

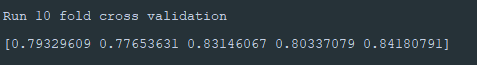
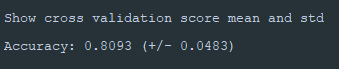


1. Get interrelation between fair and PClass

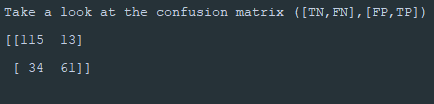
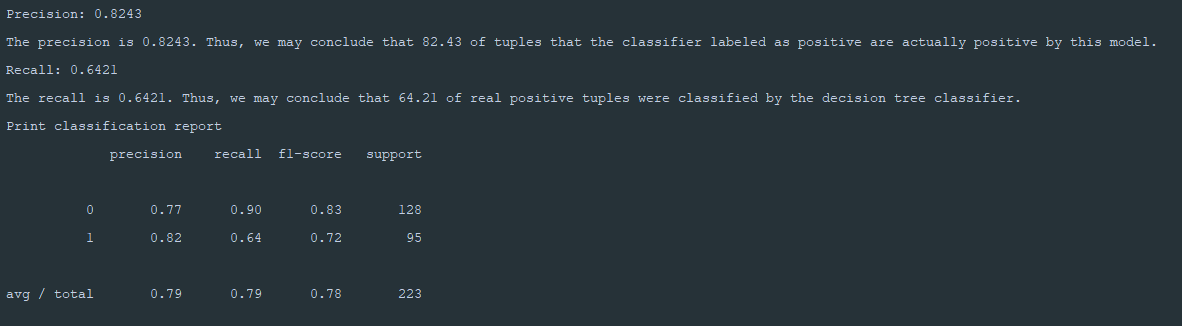


1. Declare DT classifier



1. 
2. 
3. Accuracy of Decision Tree



1. 
2. 
3. **Conclusion**

*# To sum up, passengers had higher chance of survival:  
#  
# - if they belonged to the first class (or hold expensive ticket)  
#  
# - if they were female  
#  
# - if they were young  
#  
# - if they had family  
#  
# - if they came from Cherbourg  
#  
# Among these five conditions, ticket class, sex, and age were the most influential on survival.*