

Fig1: Comparing imputation of null value mean and most frequent strategy

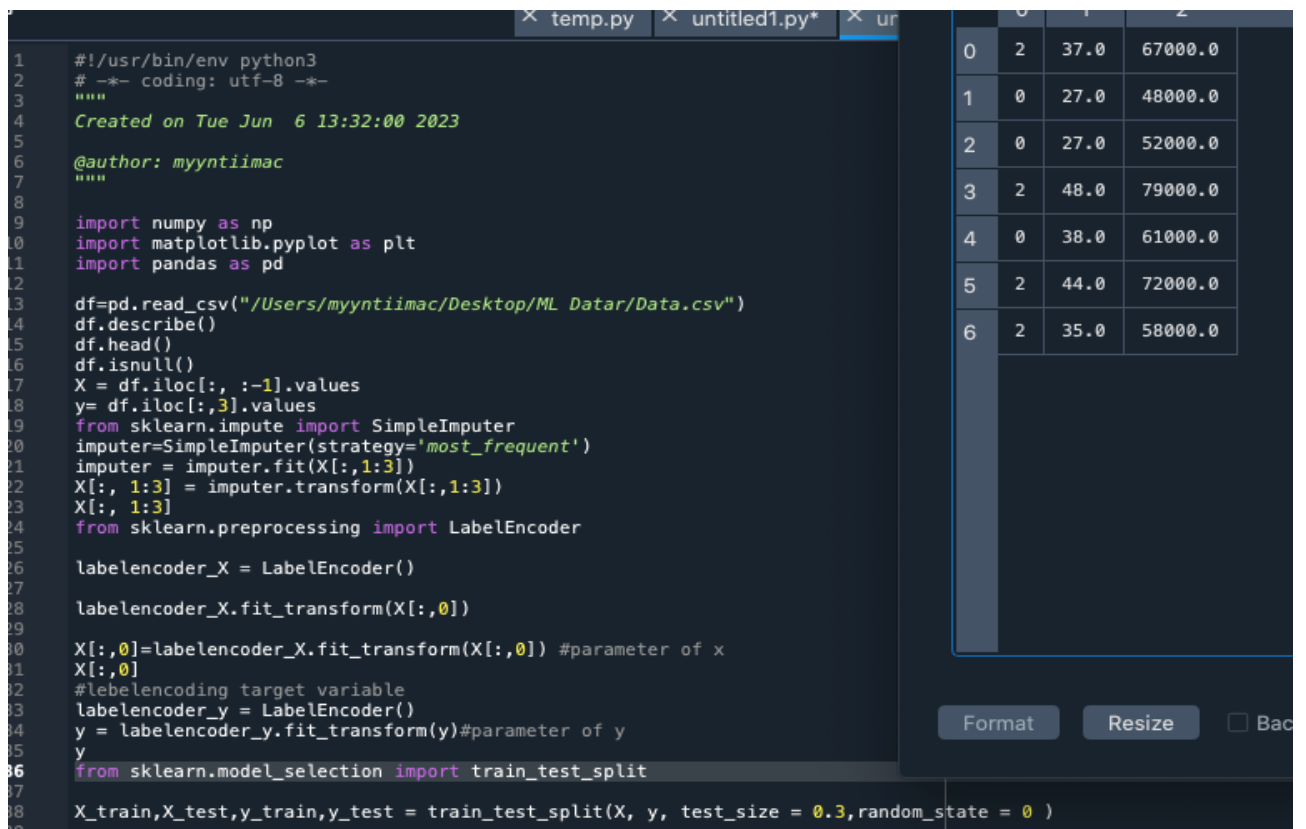


Fig2: Test size 30% and random state=0

temp.pyuntitled1.py*untitled3

```
1#!/usr/bin/env python3
2# -*- coding: utf-8 -*-
3"""
4Created on Tue Jun  6 13:32:00 2023
5
6@author: myyntiimac
7"""
8
9import numpy as np
10import matplotlib.pyplot as plt
11import pandas as pd
12
13df=pd.read_csv("/Users/myyntiimac/Desktop/ML Datar/Data.csv")
14df.describe()
15df.head()
16df.isnull()
17X = df.iloc[:, :-1].values
18y = df.iloc[:,3].values
19from sklearn.impute import SimpleImputer
20imputer=SimpleImputer(strategy='most_frequent')
21imputer = imputer.fit(X[:,1:3])
22X[:, 1:3] = imputer.transform(X[:,1:3])
23X[:, 1:3]
24from sklearn.preprocessing import LabelEncoder
25
26labelencoder_X = LabelEncoder()
27
28labelencoder_X.fit_transform(X[:,0])
29
30X[:,0]=labelencoder_X.fit_transform(X[:,0]) #parameter of x
31X[:,0]
32#lebelencoding target variable
33labelencoder_y = LabelEncoder()
34y = labelencoder_y.fit_transform(y)#parameter of y
35y
36from sklearn.model_selection import train_test_split
37
38X_train,X_test,y_train,y_test = train_test_split(X, y, test_size = 0.25)
```

	0	1	2
0	2	37.0	67000.0
1	1	50.0	83000.0
2	0	27.0	52000.0
3	1	40.0	48000.0
4	0	38.0	61000.0
5	0	27.0	48000.0
6	1	30.0	54000.0

FormatResize☐ Background color

Fig3: Test size 25% and no random state