

```
import numpy as np
import pandas as pd
import os
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import svm
from sklearn.metrics import accuracy_score
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import StandardScaler
import joblib
from sklearn.metrics import accuracy_score
```

```
[2] df=pd.read_csv(r"/content/collegePlace.csv")
df.head()
```

	Age	Gender	Stream	Internships	CGPA	Hostel	HistoryOfBacklogs	PlacedOrNot
0	22	Male	Electronics And Communication	1	8	1	1	1
1	21	Female	Computer Science	0	7	1	1	1
2	22	Female	Information Technology	1	6	0	0	1
3	21	Male	Information Technology	0	8	0	1	1
4	22	Male	Mechanical	0	8	1	0	1

1s completed at 4:26 AM

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2966 entries, 0 to 2965
Data columns (total 8 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   Age                 2966 non-null   int64
 1   Gender              2966 non-null   object
 2   Stream              2966 non-null   object
 3   Internships         2966 non-null   int64
 4   CGPA                2966 non-null   int64
 5   Hostel              2966 non-null   int64
 6   HistoryOfBacklogs   2966 non-null   int64
 7   PlacedOrNot         2966 non-null   int64
dtypes: int64(6), object(2)
memory usage: 185.5+ KB

[4] df.isnull().sum()

Age                0
Gender             0
Stream            0
Internships        0
CGPA               0
Hostel             0
HistoryOfBacklogs  0
PlacedOrNot        0
```

+ Code + Text

RAM
Disk

↑ ↓ ↶ ↷ ⚙️ 📄 🗑️ ⋮

0s

CPA
Hostel
HistoryOfBacklogs
PlacedOrNot
dtype: int64

0
0
0
0

0s
[5] def transformationplot(feature):
plt.figure(figsize=(12,5))
plt.subplot(1,2,1)
sns.distplot(feature)


2s
[6] transformationplot(np.log(df['Age']))

<ipython-input-5-4e023cd5df85>:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

sns.distplot(feature)




Task2 colab

File Edit View Insert Runtime Tools Help [All changes saved](#)

Comment

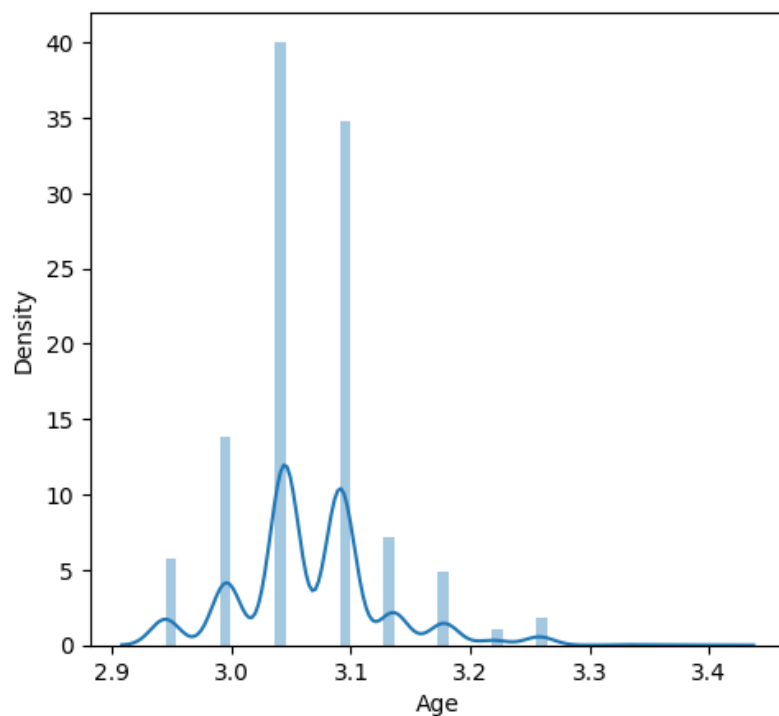
Share



+ Code + Text

✓ RAM  Disk

✓ [6] sns.distplot(feature)



✓ [7] df=df.replace(['Male'],[0])

1s completed at 4:26 AM

```
[7] df=df.replace(['Male'],[0])
df=df.replace(['Female'],[1])
```

```
[8] df=df.replace(['Computer Science','Information Technology','Electronics and Communication','Mechanical','Electrical','Civil'],[0,1,2,3,4,5])
```

```
[9] df = df.drop(['Hostel'],axis=1)
```

df

	Age	Gender	Stream	Internships	CGPA	HistoryOfBacklogs	PlacedOrNot
0	22	0	Electronics And Communication	1	8	1	1
1	21	1	0	0	7	1	1
2	22	1	1	1	6	0	1
3	21	0	1	0	8	1	1
4	22	0	3	0	8	0	1
...	...	...	...	...	...	...	...
2961	23	0	1	0	7	0	0
2962	23	0	3	1	7	0	0

+ Code + Text

✓ 0s

2962	23	0	3	1	7	0	0
2963	22	0	1	1	7	0	0
2964	22	0	0	1	7	0	0
2965	23	0	5	0	8	0	1

2966 rows x 7 columns

RAM Disk

```
[11] plt.figure(figsize=(12,5))
plt.subplot(121)
sns.distplot(df['CGPA'],color='r')
```

<ipython-input-11-f92659182652>:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(df['CGPA'],color='r')
<Axes: xlabel='CGPA', ylabel='Density'>
```