

```

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv(r'C:/dsbda/social_network_ads.csv')
df["Gender"].replace({"Male":1,"Female":0}, inplace=True)
df

```

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	1	19	19000	0
1	15810944	1	35	20000	0
2	15668575	0	26	43000	0
3	15603246	0	27	57000	0
4	15804002	1	19	76000	0
...	...	...	...	...	...
395	15691863	0	46	41000	1
396	15706071	1	51	23000	1
397	15654296	0	50	20000	1
398	15755018	1	36	33000	0
399	15594041	0	49	36000	1

[400 rows x 5 columns]

```

#remove user id column
df = df.drop(columns='User ID')

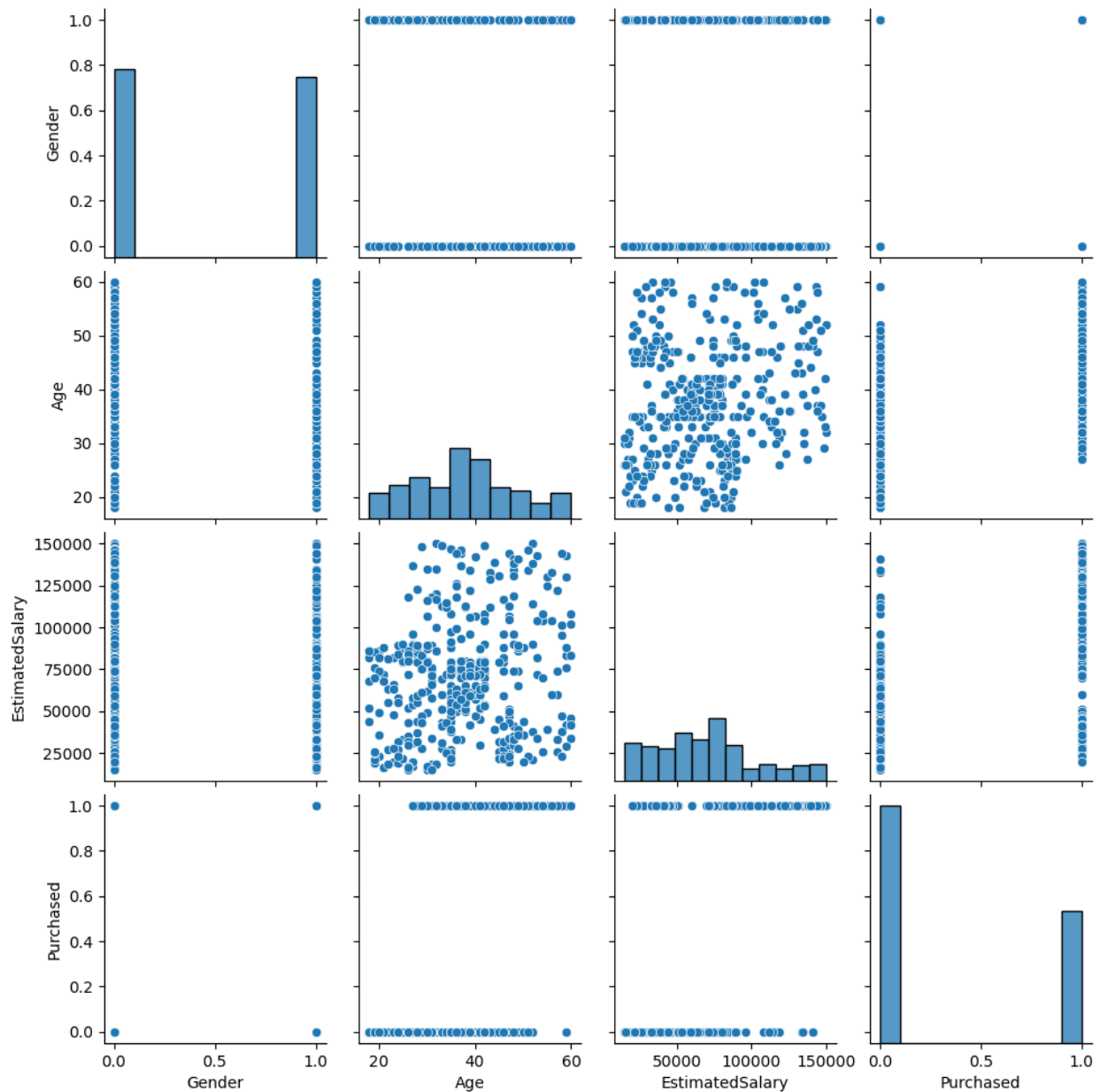
#pairplot using seaborn
sns.pairplot(df)
plt.show()

```

```

C:\ProgramData\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
self._figure.tight_layout(*args, **kwargs)

```



*#display the list of columns in dataframes*

```
df.columns
```

```
Index(['Gender', 'Age', 'EstimatedSalary', 'Purchased'],
      dtype='object')
```

```
x=df[['Gender', 'Age', 'EstimatedSalary']]
```

```
y=df['Purchased']
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.linear_model import LogisticRegression
```

```
x_train,x_test,y_train,y_test =  
train_test_split(x,y,test_size=0.25,random_state=0)
```

x\_train

	Gender	Age	EstimatedSalary
250	0	44	39000
63	1	32	120000
312	0	38	50000
159	0	32	135000
283	0	52	21000
..	...	...	...
323	0	48	30000
192	1	29	43000
117	1	36	52000
47	0	27	54000
172	0	26	118000

[300 rows x 3 columns]

```
print(y)
```

0	0
1	0
2	0
3	0
4	0
..	..
395	1
396	1
397	1
398	0
399	1

Name: Purchased, Length: 400, dtype: int64

y\_train

250	0
63	1
312	0
159	1
283	1
..	..
323	1
192	0
117	0
47	0
172	0

Name: Purchased, Length: 300, dtype: int64

```

model=LogisticRegression()
model.fit(x_train,y_train)

LogisticRegression()

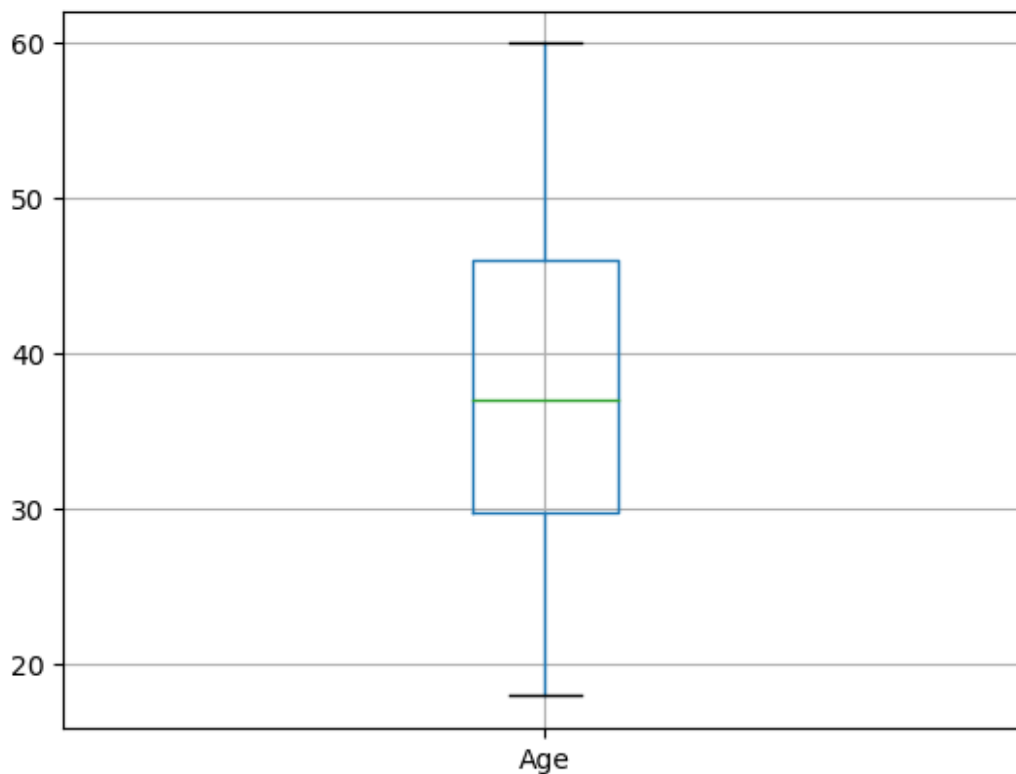
y_pred = model.predict(x_test)

y_pred
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int64)

col = ['Age']
df.boxplot(col)

<Axes: >

```



```

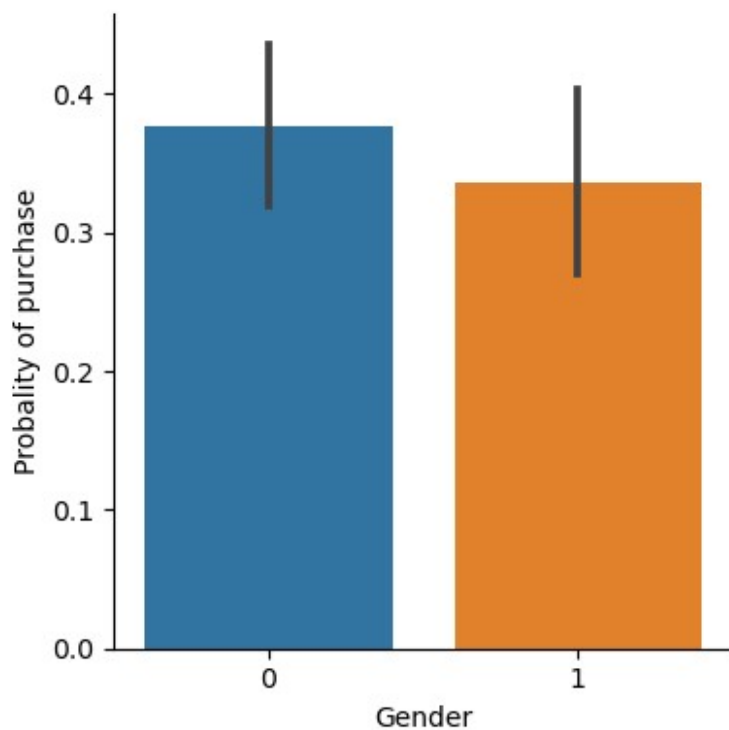
model.score(x_train,y_train)

```

```

0.63
model.score(x_test,y_test)
0.68
import seaborn as sns
g=sns.catplot(x='Gender',y='Purchased',data=df,kind='bar', height=4)
g.set_ylabels('Probability of purchase')
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
<seaborn.axisgrid.FacetGrid at 0x28577d69910>

```

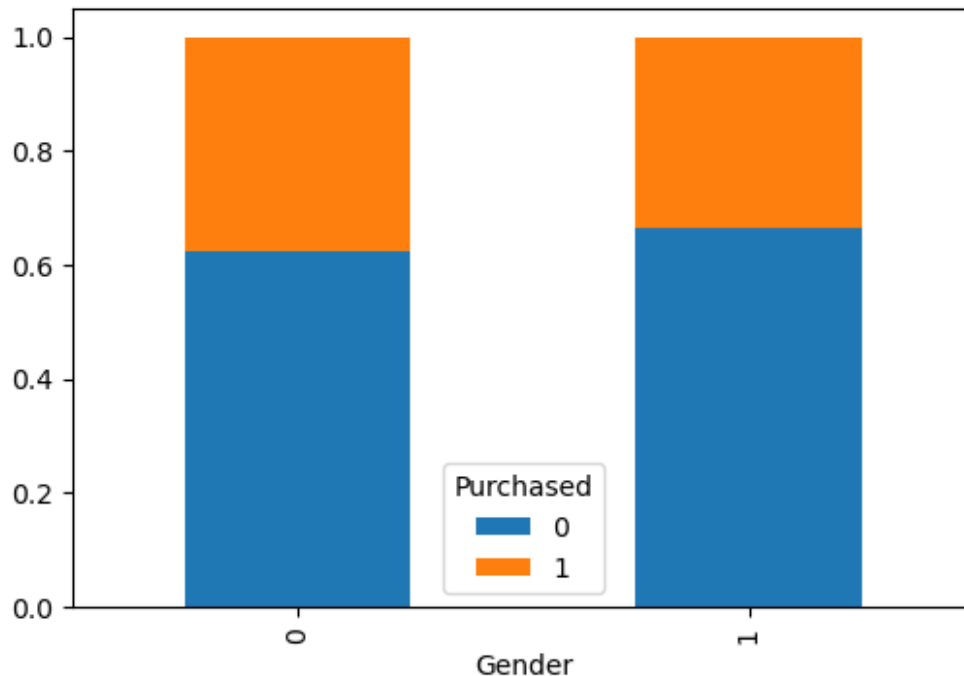


```

ct = pd.crosstab(df['Gender'], df['Purchased'], normalize="index" )
print(ct)
ct.plot.bar(figsize=(6,4), stacked =True)
plt.show()

```

Purchased	0	1
Gender		
0	0.622549	0.377451
1	0.663265	0.336735



```
from sklearn.metrics import
precision_score,confusion_matrix,accuracy_score,recall_score

cm= confusion_matrix(y_test, y_pred)
cm
array([[68,  0],
       [32,  0]], dtype=int64)

#unpack the values from the confusion matrix into four variables
tn, fp, tp, fn = confusion_matrix(y_test,y_pred).ravel()

a = accuracy_score(y_test,y_pred)
print("Accuracy Score:", a)

Accuracy Score: 0.68

r = recall_score(y_test, y_pred)
print("Recall Score:", r)

Recall Score: 0.0

p = precision_score(y_test, y_pred)
print("Precision Score:", p)

Precision Score: 0.0
```

```
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\metrics\
_classification.py:1469: UndefinedMetricWarning: Precision is ill-
defined and being set to 0.0 due to no predicted samples. Use
`zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))
```

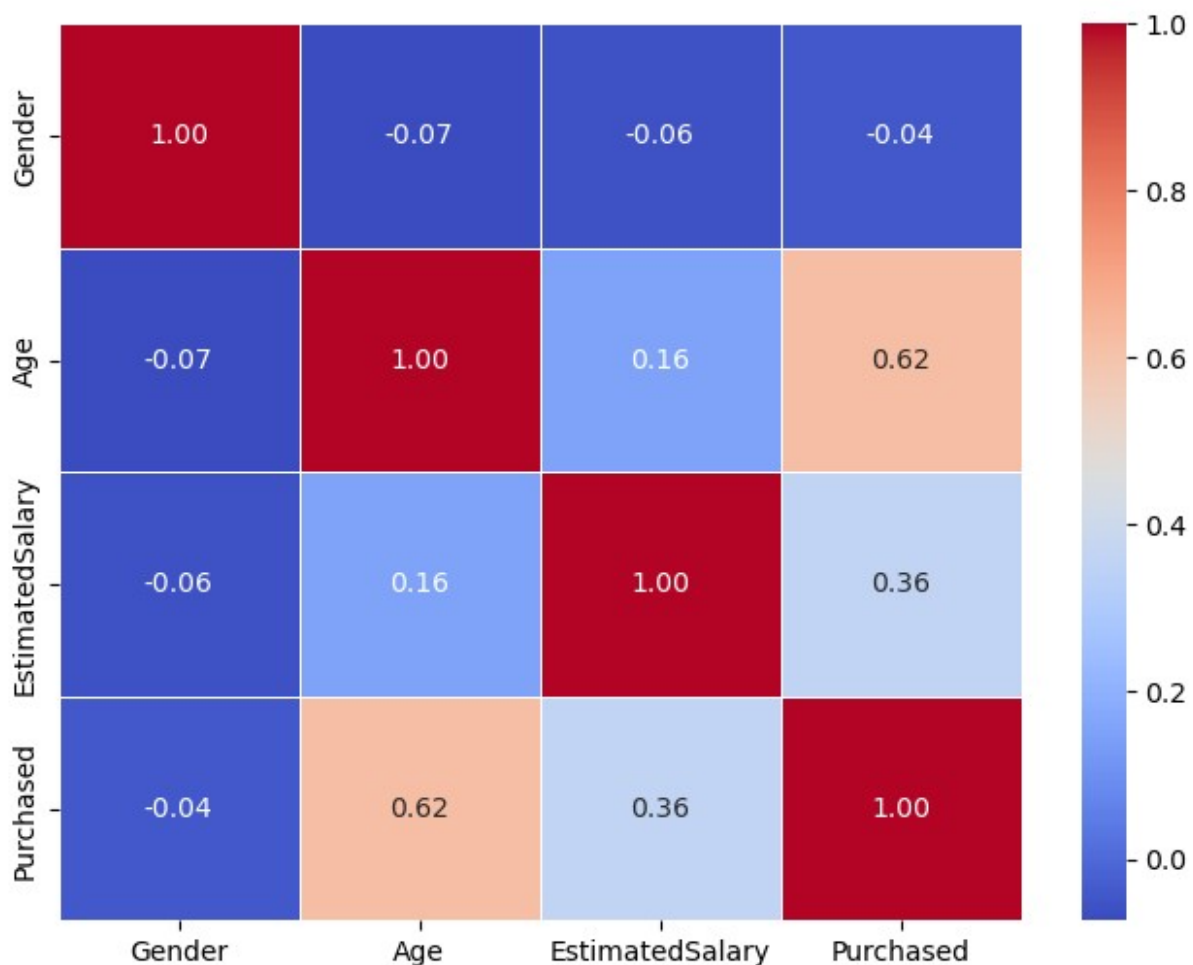
```
e = 1 - a
print("Error Rate:", e)
```

```
Error Rate: 0.31999999999999995
```

```
corr_matrix = df.corr()
```

```
plt.figure(figsize=(8,6))
sns.heatmap(corr_matrix, annot=True, cmap="coolwarm", fmt=".2f",
linewidths=0.5)
```

```
<Axes: >
```



```
sns.displot(df['Purchased'])
```

```
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118:  
UserWarning: The figure layout has changed to tight  
    self._figure.tight_layout(*args, **kwargs)
```

```
<seaborn.axisgrid.FacetGrid at 0x285778e3650>
```

