


```
#importing neccesary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
#loading the dataset
df=pd.read_excel('/content/drive/MyDrive/marketin_ research_data.xlsx')
df
```



	No	Age	earnings from work	studies	credit score	purchase of goods
0	1	21	high	no	good	no
1	2	28	high	no	very good	no
2	3	25	high	no	good	yes
3	4	42	average	no	good	yes
4	5	41	low	yes	good	yes
5	6	45	low	yes	very good	no
6	7	35	low	yes	very good	yes
7	8	29	average	no	good	no
8	9	27	low	yes	good	yes
9	10	25	low	yes	good	yes
10	11	20	average	yes	good	yes
11	12	20	low	yes	very good	no
12	13	18	high	no	very good	yes
13	14	20	low	yes	good	no
14	15	43	average	yes	good	yes
15	16	17	average	yes	very good	yes
16	17	36	average	no	very good	yes
17	18	34	high	yes	good	yes
18	19	48	average	no	very good	no

```
#dropping unnecessary columns
df1=df.drop('No',axis=1)
df1
```

index	Age	earnings from work	studies	credit score	purchase of goods
0	1.05	high	no	good	no
1	1.4	high	no	very good	no
2	1.25	high	no	good	yes
3	2.1	average	no	good	yes
4	2.05	low	yes	good	yes
5	2.25	low	yes	very good	no
6	1.75	low	yes	very good	yes
7	1.45	average	no	good	no
8	1.35	low	yes	good	yes
9	1.25	low	yes	good	yes
10	1.0	average	yes	good	yes
11	1.0	low	yes	very good	no
12	0.9	high	no	very good	yes
13	1.0	low	yes	good	no
14	2.15	average	yes	good	yes
15	0.85	average	yes	very good	yes
16	1.8	average	no	very good	yes

```
#downscaling the age column for better analysis
df1['Age']=df1['Age']/20
df2=df1.copy()
df2
```

	Age	earnings from work	studies	credit score	purchase of goods
0	1.05	high	no	good	no
1	1.40	high	no	very good	no
2	1.25	high	no	good	yes



```
#converting the categorical data columns into numerical data columns
df3=df2.replace(['low','average','high'],[0,1,2])
df3.head()
```

	Age	earnings from work	studies	credit score	purchase of goods
0	1.05	2	no	good	no
1	1.40	2	no	very good	no
2	1.25	2	no	good	yes
3	2.10	1	no	good	yes
4	2.05	0	yes	good	yes
43	0.00	high	no	very good	yes



```
df4=df3.replace(['no','yes'],[0,1])
df4.head()
```

	Age	earnings from work	studies	credit score	purchase of goods
0	1.05	2	0	good	0
1	1.40	2	0	very good	0
2	1.25	2	0	good	1
3	2.10	1	0	good	1
4	2.05	0	1	good	1



```
df5=df4.replace(['good','very good'],[0,1])
df5
```

	Age	earnings from work	studies	credit score	purchase of goods	
0	1.05	2	0	0	0	
1	1.40	2	0	1	0	
2	1.25	2	0	0	1	
3	2.10	1	0	0	1	
4	2.05	0	1	0	1	
5	2.25	0	1	1	0	
6	1.75	0	1	1	1	
7	1.45	1	0	0	0	
8	1.35	0	1	0	1	
9	1.25	0	1	0	1	
10	1.00	1	1	0	1	

```
#obtaining the x_train and y_train(target) values
x_train=df5.drop('purchase of goods',axis='columns')
x_train.tail()
```

	Age	earnings from work	studies	credit score	
14	2.15	1	1	0	
15	0.85	1	1	1	
16	1.80	1	0	1	
17	1.70	2	1	0	
18	2.40	1	0	1	

```
y_train=df5['purchase of goods']
y_train.sample(4)
```

```
9      1
12     1
8      1
5      0
Name: purchase of goods, dtype: int64
```

```
##importing the naive bayes classifier algorithm libraries
from sklearn.naive_bayes import GaussianNB
```

```
#creating the model and training the model
model=GaussianNB()
model.fit(x_train,y_train)
```

```
GaussianNB()
```

```
#answering the question given problem
#(we need to tell if a person under the age of 25,with average income,graduate degree,good credit score,so we need to note that a person under age of 25 is equivalent to lets say
##for that reason we need to note that a person under age of 25 is equivalent to lets say
# to 1.2*20=24,so we will set age as =1.2
#average income is represented by =1
#graduate degree is represented by =1
#credit score is represented by =0
#we are now ready to go,
```

```
#we now call our model to predict if the above described individual is likely to purchase a commodity
#note that if we get a zero (0) value it means that the person is not likely to purchase a commodity
#likely to purchase a commodity
model.predict([[1.2,1,1,0]])
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:451: UserWarning: X does not have valid feature names, but
array([1])
```

```
#there we go,so the above described individual will purchase the commodity and thus has a probability of 1.0
```

END OF DATACLEANING,MODEL CREATION,TRAINING AND EVALUATION.THANK YOU!!!