```
from google.colab import files

uploaded = files.upload()
```

```
Choose Files No file chosen  Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
```

```
import pandas as pd

df= pd.read_csv("Sample_for_AI.csv")
print(df)
```

```
         ID Purchased Do Not Email Do Not Call  TotalVi
0         1        No           No          No
1         2        No           No          No
2         3       Yes           No          No
3         4        No           No          No
4         5       Yes           No          No
...     ...       ...          ...         ...
9235   9236       Yes          Yes          No
9236   9237        No           No          No
9237   9238        No          Yes          No
9238   9239       Yes           No          No
9239   9240       Yes           No          No

      Total Time Spent on Website  Page Views Per Visi
0                               0                 0.0
1                             674                 2.5
2                            1532                 2.0
3                             305                 1.0
4                            1428                 1.0
...                           ...                  ..
9235                         1845                 2.6
9236                          238                 2.0
9237                          199                 2.0
9238                          499                 3.0
9239                         1279                 3.0

                 Specialization                   Orig
0                           NaN                      A
1                           NaN                      A
2       Business Administration  Landing Page Submissi
3         Media and Advertising  Landing Page Submissi
4                           NaN  Landing Page Submissi
...                         ...                    .
9235     IT Projects Management  Landing Page Submissi
9236      Media and Advertising  Landing Page Submissi
9237    Business Administration  Landing Page Submissi
9238  Human Resource Management  Landing Page Submissi
9239    Supply Chain Management  Landing Page Submissi

[9240 rows x 9 columns]
```

```
df.duplicated().values.any()
```

```
False
```

```
df.nunique()
```

Sample_for_AI.csv

1 to 10 of 9240 entries Filter

ID	Purchased	Do Not Email	Do Not Call	TotalVisits	Total Time Sp
1	No	No	No	0	0
2	No	No	No	5	674
3	Yes	No	No	2	1532
4	No	No	No	1	305
5	Yes	No	No	2	1428
6	No	No	No	0	0
7	Yes	No	No	2	1640
8	No	No	No	0	0
9	No	No	No	2	71
10	No	No	No	4	58

Show 10 per page

1 2 10 100 900 920 924

	0
ID	9240
Purchased	2
Do Not Email	2
Do Not Call	2
TotalVisits	41
Total Time Spent on Website	1731
Page Views Per Visit	114
Specialization	18
Origin	5

dtype: int64

```
print(df.head())
```

```
   ID Purchased Do Not Email Do Not Call  TotalVisits
0   1        No           No          No          0.0
1   2        No           No          No          5.0
2   3       Yes           No          No          2.0
3   4        No           No          No          1.0
4   5       Yes           No          No          2.0

   Total Time Spent on Website  Page Views Per Visit
0                            0                   0.0
1                          674                   2.5
2                         1532                   2.0
3                          305                   1.0
4                         1428                   1.0

                    Origin
0                      API
1                      API
2  Landing Page Submission
3  Landing Page Submission
4  Landing Page Submission
```

```
print(df.isnull().sum())
```

```
ID                             0
Purchased                      0
Do Not Email                   0
Do Not Call                    0
TotalVisits                  137
Total Time Spent on Website    0
Page Views Per Visit         137
Specialization              3380
Origin                         0
dtype: int64
```

```
df.dropna(inplace=True)
```

```
df = pd.get_dummies(df, drop_first=True)
```

```
df
```

	ID	TotalVisits	Total Time Spent on Website	Page Views Per Visit	Purchased_Yes
2	3	2.0	1532	2.00	True
3	4	1.0	305	1.00	False
6	7	2.0	1640	2.00	True
8	9	2.0	71	2.00	False
9	10	4.0	58	4.00	False
...	...	...	...	...	...
9235	9236	8.0	1845	2.67	True
9236	9237	2.0	238	2.00	False
9237	9238	2.0	199	2.00	False
9238	9239	3.0	499	3.00	True
9239	9240	6.0	1279	3.00	True

5799 rows × 27 columns

```
print(df.isnull().sum())
```

```
ID                                        0
TotalVisits                               0
Total Time Spent on Website               0
Page Views Per Visit                      0
Purchased_Yes                             0
Do Not Email_Yes                          0
Do Not Call_Yes                           0
Specialization_Business Administration    0
Specialization_E-Business                 0
Specialization_E-COMMERCE                 0
Specialization_Finance Management         0
Specialization_Healthcare Management      0
Specialization_Hospitality Management     0
Specialization_Human Resource Management  0
Specialization_IT Projects Management     0
Specialization_International Business     0
Specialization_Marketing Management       0
Specialization_Media and Advertising      0
Specialization_Operations Management      0
Specialization_Retail Management          0
Specialization_Rural and Agribusiness     0
Specialization_Services Excellence        0
Specialization_Supply Chain Management    0
Specialization_Travel and Tourism         0
Origin_Landing Page Submission            0
Origin_Lead Add Form                      0
Origin_Lead Import                        0
dtype: int64
```

```
X = df.drop([col for col in df.columns if 'Purchased' in c
y = df['Purchased_Yes']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
```

```
!pip install scikit-learn

from sklearn.ensemble import RandomForestClassifier

X_train, X_test, y_train, y_test = train_test_split(X, y,

model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)
```

```
Requirement already satisfied: scikit-learn in /usr/lo
Requirement already satisfied: numpy>=1.19.5 in /usr/l
Requirement already satisfied: scipy>=1.6.0 in /usr/lo
Requirement already satisfied: joblib>=1.2.0 in /usr/l
Requirement already satisfied: threadpoolctl>=3.1.0 in
```

```
RandomForestClassifier
RandomForestClassifier(random_state=42)
```

```
y_pred = model.predict(X_test)
```

```
!pip install scikit-learn

from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report

X_train, X_test, y_train, y_test = train_test_split(X, y,

model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

print(classification_report(y_test, y_pred))
```

```
Requirement already satisfied: scikit-learn in /usr/lo
Requirement already satisfied: numpy>=1.19.5 in /usr/l
Requirement already satisfied: scipy>=1.6.0 in /usr/lo
Requirement already satisfied: joblib>=1.2.0 in /usr/l
Requirement already satisfied: threadpoolctl>=3.1.0 in
              precision    recall  f1-score   support

       False       0.77      0.78      0.77       624
        True       0.74      0.73      0.73       536

    accuracy                           0.76      1160
   macro avg       0.75      0.75      0.75      1160
weighted avg       0.75      0.76      0.75      1160
```

```
!pip install scikit-learn

from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix # import the

X_train, X_test, y_train, y_test = train_test_split(X, y,

model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

print(classification_report(y_test, y_pred))

conf_matrix = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", conf_matrix)
```

```
Requirement already satisfied: scikit-learn in /usr/lo
Requirement already satisfied: numpy>=1.19.5 in /usr/l
Requirement already satisfied: scipy>=1.6.0 in /usr/lo
Requirement already satisfied: joblib>=1.2.0 in /usr/l
```

```
Requirement already satisfied: threadpoolctl>=3.1.0 in
              precision    recall  f1-score   support

       False       0.77      0.78      0.77       624
        True       0.74      0.73      0.73       536
```