```python
import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style="whitegrid")

from scipy import stats
import statsmodels.api as sm
```

```
C:\Users\msi\anaconda3\lib\site-packages\statsmodels\tsa\base\
tsa_model.py:7: FutureWarning: pandas.Int64Index is deprecated and
will be removed from pandas in a future version. Use pandas.Index with
the appropriate dtype instead.
  from pandas import (to_datetime, Int64Index, DatetimeIndex, Period,
C:\Users\msi\anaconda3\lib\site-packages\statsmodels\tsa\base\
tsa_model.py:7: FutureWarning: pandas.Float64Index is deprecated and
will be removed from pandas in a future version. Use pandas.Index with
the appropriate dtype instead.
  from pandas import (to_datetime, Int64Index, DatetimeIndex, Period,
```

```python
leads = pd.read_csv("Leads.csv")
```

```python
leads.shape
```

```
(9240, 37)
```

```python
leads.columns
```

```
Index(['Prospect ID', 'Lead Number', 'Lead Origin', 'Lead Source',
       'Do Not Email', 'Do Not Call', 'Converted', 'TotalVisits',
       'Total Time Spent on Website', 'Page Views Per Visit', 'Last
Activity',
       'Country', 'Specialization', 'How did you hear about X
Education',
       'What is your current occupation',
       'What matters most to you in choosing a course', 'Search',
'Magazine',
       'Newspaper Article', 'X Education Forums', 'Newspaper',
       'Digital Advertisement', 'Through Recommendations',
       'Receive More Updates About Our Courses', 'Tags', 'Lead
Quality',
       'Update me on Supply Chain Content', 'Get updates on DM
Content',
       'Lead Profile', 'City', 'Asymmetrique Activity Index',
       'Asymmetrique Profile Index', 'Asymmetrique Activity Score',
       'Asymmetrique Profile Score',
       'I agree to pay the amount through cheque',
       'A free copy of Mastering The Interview', 'Last Notable
Activity'],
      dtype='object')
```

```
leads.info

<bound method DataFrame.info of
Prospect ID  Lead Number  \
0     7927b2df-8bba-4d29-b9a2-b6e0beafe620         660737
1     2a272436-5132-4136-86fa-dcc88c88f482         660728
2     8cc8c611-a219-4f35-ad23-fdfd2656bd8a         660727
3     0cc2df48-7cf4-4e39-9de9-19797f9b38cc         660719
4     3256f628-e534-4826-9d63-4a8b88782852         660681
...                                        ...            ...
9235  19d6451e-fcd6-407c-b83b-48e1af805ea9         579564
9236  82a7005b-7196-4d56-95ce-a79f937a158d         579546
9237  aac550fe-a586-452d-8d3c-f1b62c94e02c         579545
9238  5330a7d1-2f2b-4df4-85d6-64ca2f6b95b9         579538
9239  571b5c8e-a5b2-4d57-8574-f2ffb06fdeff         579533
```

	Lead Origin	Lead Source	Do Not Email	Do Not Call
0	API	Olark Chat	No	No
1	API	Organic Search	No	No
2	Landing Page Submission	Direct Traffic	No	No
3	Landing Page Submission	Direct Traffic	No	No
4	Landing Page Submission	Google	No	No
...	...	...	...	...
9235	Landing Page Submission	Direct Traffic	Yes	No
9236	Landing Page Submission	Direct Traffic	No	No
9237	Landing Page Submission	Direct Traffic	Yes	No
9238	Landing Page Submission	Google	No	No
9239	Landing Page Submission	Direct Traffic	No	No

```
      Converted  TotalVisits  Total Time Spent on Website  \
0             0          0.0                            0
1             0          5.0                          674
2             1          2.0                         1532
3             0          1.0                          305
4             1          2.0                         1428
...         ...          ...                          ...
9235          1          8.0                         1845
9236          0          2.0                          238
9237          0          2.0                          199
```

```
9238            1        3.0                                    499
9239            1        6.0                                   1279

      Page Views Per Visit  ...  Get updates on DM Content       Lead
Profile  \
0                     0.00  ...                         No
Select
1                     2.50  ...                         No
Select
2                     2.00  ...                         No   Potential
Lead
3                     1.00  ...                         No
Select
4                     1.00  ...                         No
Select
...                    ...  ...                        ...
...
9235                  2.67  ...                         No   Potential
Lead
9236                  2.00  ...                         No   Potential
Lead
9237                  2.00  ...                         No   Potential
Lead
9238                  3.00  ...                         No
NaN
9239                  3.00  ...                         No   Potential
Lead

                        City Asymmetrique Activity Index  \
0                     Select                   02.Medium
1                     Select                   02.Medium
2                     Mumbai                   02.Medium
3                     Mumbai                   02.Medium
4                     Mumbai                   02.Medium
...                      ...                         ...
9235                  Mumbai                   02.Medium
9236                  Mumbai                   02.Medium
9237                  Mumbai                   02.Medium
9238    Other Metro Cities                   02.Medium
9239          Other Cities                   02.Medium

     Asymmetrique Profile Index Asymmetrique Activity Score  \
0                     02.Medium                        15.0
1                     02.Medium                        15.0
2                       01.High                        14.0
3                       01.High                        13.0
4                       01.High                        15.0
...                         ...                         ...
9235                    01.High                        15.0
9236                    01.High                        14.0
```

```
9237                          01.High                        13.0
9238                          02.Medium                      15.0
9239                          01.High                        15.0

     Asymmetrique Profile Score I agree to pay the amount through
cheque  \
0                            15.0
No
1                            15.0
No
2                            20.0
No
3                            17.0
No
4                            18.0
No
...                           ...                                     .
..
9235                         17.0
No
9236                         19.0
No
9237                         20.0
No
9238                         16.0
No
9239                         18.0
No

     A free copy of Mastering The Interview Last Notable Activity
0                                             No              Modified
1                                             No         Email Opened
2                                            Yes         Email Opened
3                                             No              Modified
4                                             No              Modified
...                                          ...                  ...
9235                                          No   Email Marked Spam
9236                                         Yes             SMS Sent
9237                                         Yes             SMS Sent
9238                                          No             SMS Sent
9239                                         Yes             Modified

[9240 rows x 37 columns]>

leads.isnull().sum()

Prospect ID                                              0
Lead Number                                              0
Lead Origin                                              0
Lead Source                                             36
```

```
Do Not Email                                         0
Do Not Call                                          0
Converted                                            0
TotalVisits                                        137
Total Time Spent on Website                          0
Page Views Per Visit                               137
Last Activity                                      103
Country                                           2461
Specialization                                    1438
How did you hear about X Education                2207
What is your current occupation                   2690
What matters most to you in choosing a course     2709
Search                                               0
Magazine                                             0
Newspaper Article                                    0
X Education Forums                                    0
Newspaper                                            0
Digital Advertisement                                0
Through Recommendations                              0
Receive More Updates About Our Courses               0
Tags                                              3353
Lead Quality                                      4767
Update me on Supply Chain Content                    0
Get updates on DM Content                            0
Lead Profile                                      2709
City                                              1420
Asymmetrique Activity Index                       4218
Asymmetrique Profile Index                        4218
Asymmetrique Activity Score                       4218
Asymmetrique Profile Score                        4218
I agree to pay the amount through cheque             0
A free copy of Mastering The Interview               0
Last Notable Activity                                0
dtype: int64
```

```python
for c in leads.columns:
    if leads[c].isnull().sum()>3000:
        leads.drop(c, axis=1,inplace=True)

leads.isnull().sum()
```

```
Prospect ID                                          0
Lead Number                                          0
Lead Origin                                          0
Lead Source                                         36
Do Not Email                                         0
Do Not Call                                          0
Converted                                            0
TotalVisits                                        137
Total Time Spent on Website                          0
```

```
Page Views Per Visit                              137
Last Activity                                     103
Country                                          2461
Specialization                                   1438
How did you hear about X Education               2207
What is your current occupation                  2690
What matters most to you in choosing a course    2709
Search                                              0
Magazine                                            0
Newspaper Article                                   0
X Education Forums                                  0
Newspaper                                           0
Digital Advertisement                               0
Through Recommendations                             0
Receive More Updates About Our Courses              0
Update me on Supply Chain Content                   0
Get updates on DM Content                           0
Lead Profile                                     2709
City                                             1420
I agree to pay the amount through cheque            0
A free copy of Mastering The Interview              0
Last Notable Activity                               0
dtype: int64
```

```python
leads.drop(['City'], axis = 1, inplace = True)

leads.drop(['Country'], axis = 1, inplace = True)

round(100*(leads.isnull().sum()/len(leads.index)), 2)
```

```
Prospect ID                                      0.00
Lead Number                                      0.00
Lead Origin                                      0.00
Lead Source                                      0.39
Do Not Email                                     0.00
Do Not Call                                      0.00
Converted                                        0.00
TotalVisits                                      1.48
Total Time Spent on Website                      0.00
Page Views Per Visit                             1.48
Last Activity                                    1.11
Specialization                                  15.56
How did you hear about X Education              23.89
What is your current occupation                29.11
What matters most to you in choosing a course  29.32
Search                                           0.00
Magazine                                         0.00
Newspaper Article                                0.00
X Education Forums                               0.00
Newspaper                                        0.00
```

```
Digital Advertisement                          0.00
Through Recommendations                        0.00
Receive More Updates About Our Courses         0.00
Update me on Supply Chain Content              0.00
Get updates on DM Content                      0.00
Lead Profile                                  29.32
I agree to pay the amount through cheque       0.00
A free copy of Mastering The Interview         0.00
Last Notable Activity                          0.00
dtype: float64
```

```python
leads.drop(['Lead Profile', 'How did you hear about X Education'],
axis = 1, inplace = True)

from matplotlib import pyplot as plt
import seaborn as sns
sns.pairplot(leads,diag_kind='kde',hue='Converted')
plt.show()
```
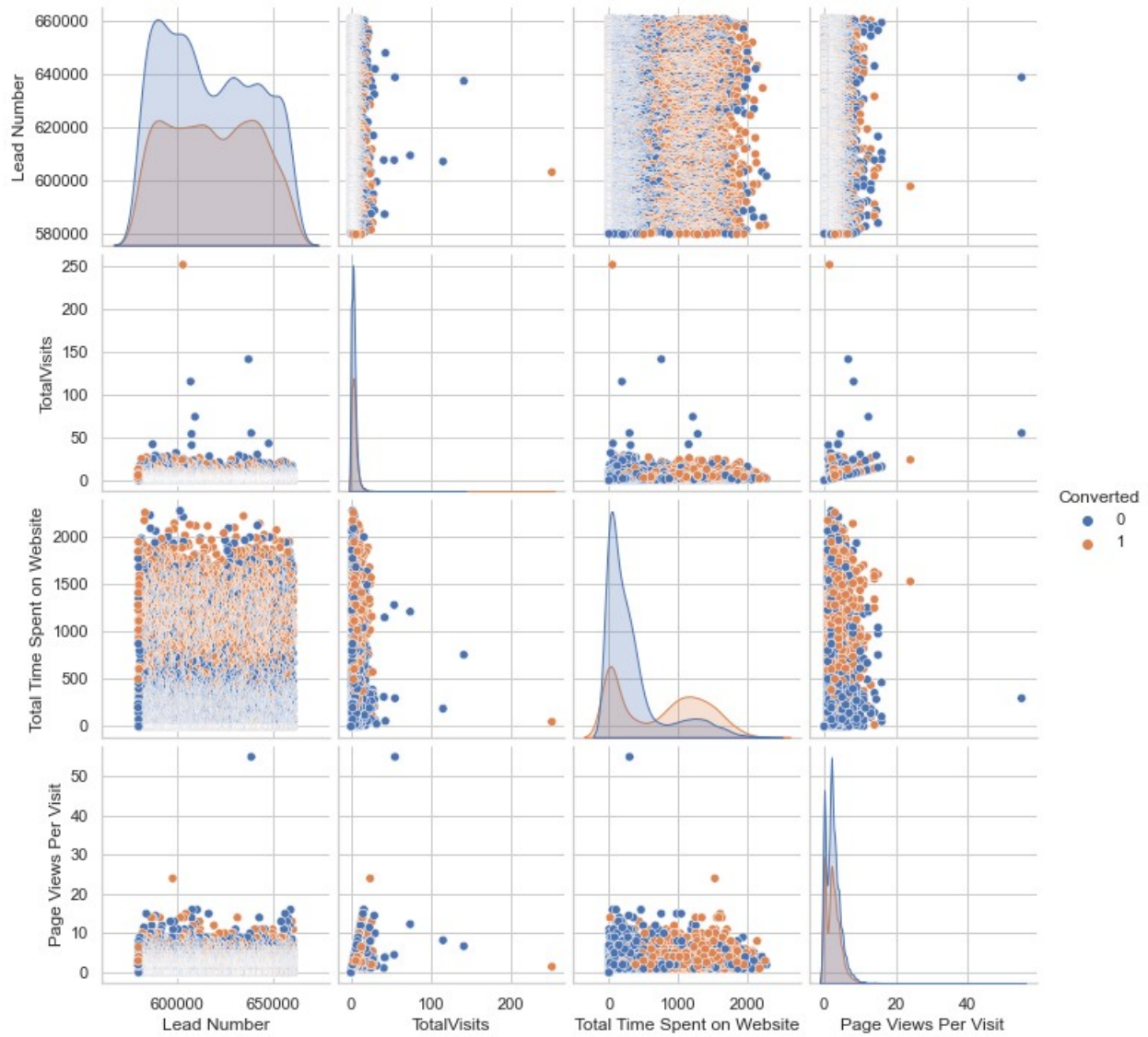
```
x_edu = leads[['TotalVisits','Total Time Spent on Website','Page Views
Per Visit','Converted']]
sns.pairplot(x_edu,diag_kind='kde',hue='Converted')
plt.show()
```
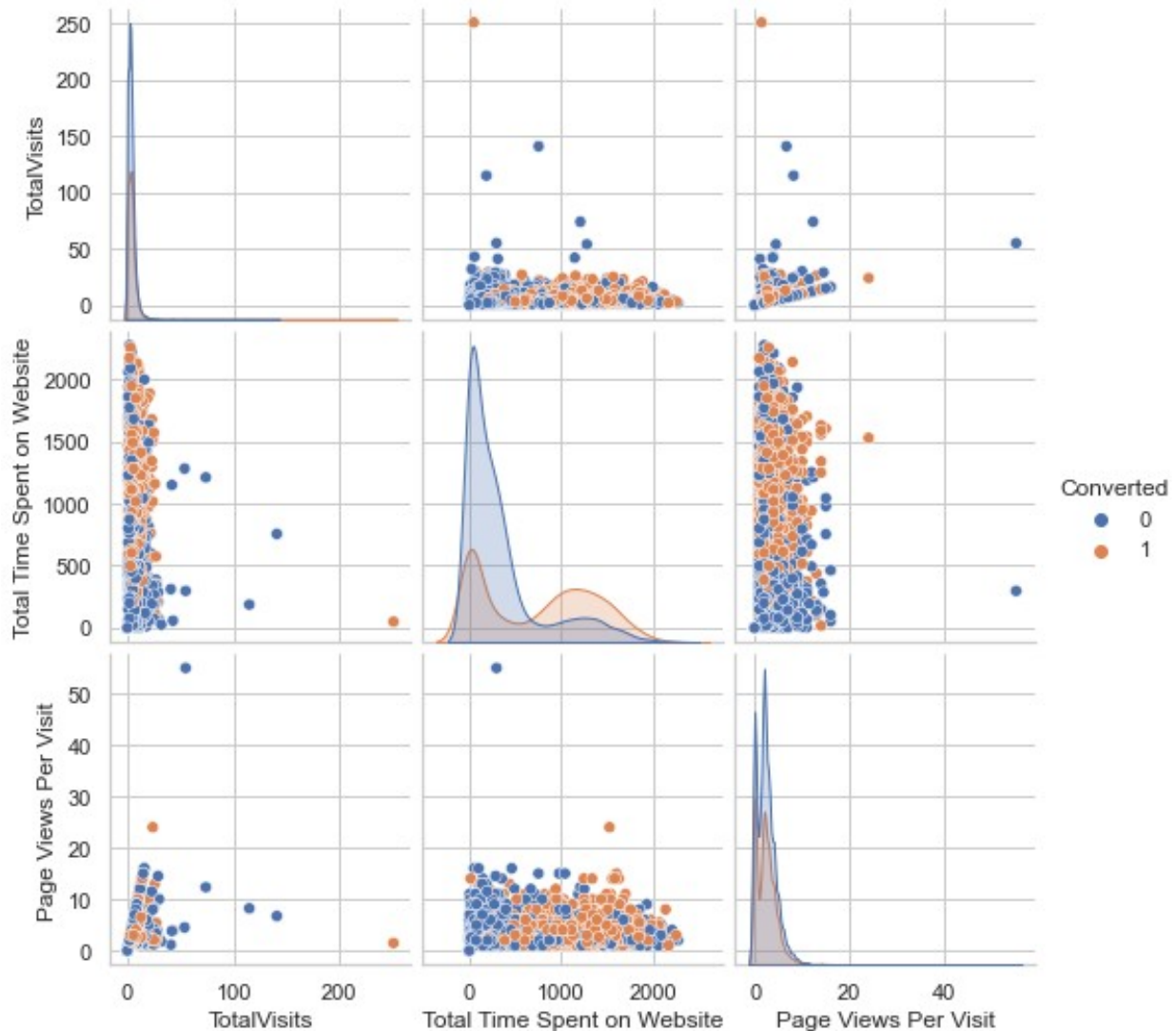
```python
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.feature_selection import RFE
from statsmodels.stats.outliers_influence import
variance_inflation_factor
from sklearn import metrics
from sklearn.metrics import
classification_report,recall_score,roc_auc_score,roc_curve,accuracy_sc
ore,precision_score,precision_recall_curve,confusion_matrix
from sklearn.preprocessing import LabelEncoder

from sklearn.preprocessing import PowerTransformer
pt = PowerTransformer()
transformedx_edu = pd.DataFrame(pt.fit_transform(x_edu))
```

```python
transformedx_edu.columns = x_edu.columns
transformedx_edu.head()
```

```
   TotalVisits  Total Time Spent on Website  Page Views Per Visit
Converted
0   -1.457907                    -1.473767             -1.454706  -
0.791863
1    0.747918                     0.729628              0.308534  -
0.791863
2   -0.141636                     1.306093              0.065574
1.262845
3   -0.640428                     0.264936             -0.536967  -
0.791863
4   -0.141636                     1.252499             -0.536967
1.262845
```

```python
leads.drop(['Do Not Call', 'Search', 'Magazine', 'Newspaper Article',
'X Education Forums', 'Newspaper',
            'Digital Advertisement', 'Through Recommendations',
'Receive More Updates About Our Courses',
            'Update me on Supply Chain Content', 'Get updates on DM
Content',
            'I agree to pay the amount through cheque'], axis = 1,
inplace = True)
```

```python
leads.drop(['What matters most to you in choosing a course'], axis =
1, inplace=True)
```

```python
leads = leads[~pd.isnull(leads['What is your current occupation'])]
```

```python
leads = leads[~pd.isnull(leads['TotalVisits'])]
```

```python
leads = leads[~pd.isnull(leads['Lead Source'])]
```

```python
leads = leads[~pd.isnull(leads['Specialization'])]
```

```python
leads.drop(['Prospect ID', 'Lead Number'], 1, inplace = True)
```

```
C:\Users\msi\AppData\Local\Temp/ipykernel_5192/1994502933.py:1:
FutureWarning: In a future version of pandas all arguments of
DataFrame.drop except for the argument 'labels' will be keyword-only.
  leads.drop(['Prospect ID', 'Lead Number'], 1, inplace = True)
```

```python
df = pd.DataFrame({'P': ['p', 'q', 'p']})
df
```

```
   P
0  p
1  q
2  p
```

```python
pd.get_dummies(df)
```

```
      P_p  P_q
0      1    0
1      0    1
2      1    0

pd.get_dummies(df)

      P_p  P_q
0      1    0
1      0    1
2      1    0

dummy = pd.get_dummies(leads[['Lead Origin', 'Lead Source', 'Do Not
Email', 'Last Activity',
                              'What is your current occupation','A
free copy of Mastering The Interview',
                              'Last Notable Activity']],
drop_first=True)

leads = pd.concat([leads, dummy], axis=1)

dummy_spl = pd.get_dummies(leads['Specialization'], prefix =
'Specialization')
dummy_spl = dummy_spl.drop(['Specialization_Select'], 1)
leads = pd.concat([leads, dummy_spl], axis = 1)

C:\Users\msi\AppData\Local\Temp/ipykernel_5192/1153367888.py:2:
FutureWarning: In a future version of pandas all arguments of
DataFrame.drop except for the argument 'labels' will be keyword-only.
  dummy_spl = dummy_spl.drop(['Specialization_Select'], 1)

leads = leads.drop(['Lead Origin', 'Lead Source', 'Do Not Email',
'Last Activity',
                    'Specialization', 'What is your current
occupation',
                    'A free copy of Mastering The Interview', 'Last
Notable Activity'], 1)

C:\Users\msi\AppData\Local\Temp/ipykernel_5192/3047035120.py:1:
FutureWarning: In a future version of pandas all arguments of
DataFrame.drop except for the argument 'labels' will be keyword-only.
  leads = leads.drop(['Lead Origin', 'Lead Source', 'Do Not Email',
'Last Activity',

X = leads.drop(['Converted'], 1)
X.head()

C:\Users\msi\AppData\Local\Temp/ipykernel_5192/550316025.py:1:
FutureWarning: In a future version of pandas all arguments of
DataFrame.drop except for the argument 'labels' will be keyword-only.
  X = leads.drop(['Converted'], 1)
```

```
   TotalVisits  Total Time Spent on Website  Page Views Per Visit  \
0         0.0                            0                   0.0
1         5.0                          674                   2.5
2         2.0                         1532                   2.0
3         1.0                          305                   1.0
4         2.0                         1428                   1.0

   Lead Origin_Landing Page Submission  Lead Origin_Lead Add Form  \
0                                    0                          0
1                                    0                          0
2                                    1                          0
3                                    1                          0
4                                    1                          0

   Lead Origin_Lead Import  Lead Source_Direct Traffic  Lead
Source_Facebook  \
0                        0                           0
0
1                        0                           0
0
2                        0                           1
0
3                        0                           1
0
4                        0                           0
0

   Lead Source_Google  Lead Source_Live Chat  ...  \
0                   0                      0  ...
1                   0                      0  ...
2                   0                      0  ...
3                   0                      0  ...
4                   1                      0  ...

   Specialization_IT Projects Management  \
0                                       0
1                                       0
2                                       0
3                                       0
4                                       0

   Specialization_International Business  Specialization_Marketing
Management  \
0                                       0
0
1                                       0
0
2                                       0
0
3                                       0
```

```
0
4                                                    0
0

    Specialization_Media and Advertising  Specialization_Operations
Management  \
0                                                    0
0
1                                                    0
0
2                                                    0
0
3                                                    1
0
4                                                    0
0

    Specialization_Retail Management  Specialization_Rural and
Agribusiness  \
0                                 0
0
1                                 0
0
2                                 0
0
3                                 0
0
4                                 0
0

    Specialization_Services Excellence  Specialization_Supply Chain
Management  \
0                                    0
0
1                                    0
0
2                                    0
0
3                                    0
0
4                                    0
0

    Specialization_Travel and Tourism
0                                  0
1                                  0
2                                  0
3                                  0
4                                  0
```

```
[5 rows x 74 columns]

y = leads['Converted']

y.head()

0    0
1    0
2    1
3    0
4    1
Name: Converted, dtype: int64

X_train, X_test, y_train, y_test = train_test_split(X, y,
train_size=0.7, test_size=0.3, random_state=100)

scaler = MinMaxScaler()

X_train[['TotalVisits', 'Page Views Per Visit', 'Total Time Spent on
Website']] = scaler.fit_transform(X_train[['TotalVisits', 'Page Views
Per Visit', 'Total Time Spent on Website']])

X_train.head()

C:\Users\msi\AppData\Local\Temp/ipykernel_5192/1538083051.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  X_train[['TotalVisits', 'Page Views Per Visit', 'Total Time Spent on
Website']] = scaler.fit_transform(X_train[['TotalVisits', 'Page Views
Per Visit', 'Total Time Spent on Website']])
```

	TotalVisits	Total Time Spent on Website	Page Views Per Visit
8003	0.015936	0.029489	0.125
218	0.015936	0.082306	0.250
4171	0.023904	0.034331	0.375
4037	0.000000	0.000000	0.000
3660	0.000000	0.000000	0.000

```
      Lead Origin_Landing Page Submission  Lead Origin_Lead Add
Form  \
```

8003	1	0
218	1	0
4171	1	0
4037	0	0
3660	0	1

	Lead Origin_Lead Import	Lead Source_Direct Traffic
8003	0	1
218	0	1
4171	0	1
4037	0	0
3660	0	0

	Lead Source_Facebook	Lead Source_Google	Lead Source_Live Chat	
8003	0	0	0	...
218	0	0	0	...
4171	0	0	0	...
4037	0	0	0	...
3660	0	0	0	...

	Specialization_IT Projects Management
8003	1
218	0
4171	0
4037	0
3660	0

	Specialization_International Business
8003	0
218	0
4171	0
4037	0
3660	0

	Specialization_Marketing Management
8003	0
218	0
4171	0
4037	0

```
3660                                                          0

        Specialization_Media and Advertising  \
8003                                         0
218                                          0
4171                                         0
4037                                         0
3660                                         0

        Specialization_Operations Management  Specialization_Retail
Management  \
8003                                     0
0
218                                      0
0
4171                                     0
0
4037                                     0
0
3660                                     0
0

        Specialization_Rural and Agribusiness  \
8003                                         0
218                                          0
4171                                         0
4037                                         0
3660                                         0

        Specialization_Services Excellence  \
8003                                       0
218                                        0
4171                                       0
4037                                       0
3660                                       0

        Specialization_Supply Chain Management  \
8003                                          0
218                                           0
4171                                          0
4037                                          0
3660                                          0

        Specialization_Travel and Tourism
8003                                     0
218                                      0
4171                                     1
4037                                     0
3660                                     0
```

```
[5 rows x 74 columns]

logreg = LogisticRegression()
rfe = RFE(logreg, 15)
rfe = rfe.fit(X_train, y_train)
```

C:\Users\msi\anaconda3\lib\site-packages\sklearn\utils\
validation.py:70: FutureWarning: Pass n_features_to_select=15 as
keyword args. From version 1.0 (renaming of 0.25) passing these as
positional arguments will result in an error
  warnings.warn(f"Pass {args_msg} as keyword args. From version "
C:\Users\msi\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:763: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
C:\Users\msi\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:763: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
C:\Users\msi\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:763: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression

```
  n_iter_i = _check_optimize_result(
C:\Users\msi\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:763: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
C:\Users\msi\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:763: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
C:\Users\msi\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:763: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
C:\Users\msi\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:763: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
```

```
https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
C:\Users\msi\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:763: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
C:\Users\msi\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:763: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
C:\Users\msi\anaconda3\lib\site-packages\sklearn\linear_model\
_logistic.py:763: ConvergenceWarning: lbfgs failed to converge
(status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as
shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(

list(zip(X_train.columns, rfe.support_, rfe.ranking_))

[('TotalVisits', True, 1),
 ('Total Time Spent on Website', True, 1),
 ('Page Views Per Visit', False, 23),
 ('Lead Origin_Landing Page Submission', False, 8),
 ('Lead Origin_Lead Add Form', True, 1),
```

```
('Lead Origin_Lead Import', False, 52),
('Lead Source_Direct Traffic', False, 24),
('Lead Source_Facebook', False, 51),
('Lead Source_Google', False, 36),
('Lead Source_Live Chat', False, 44),
('Lead Source_Olark Chat', True, 1),
('Lead Source_Organic Search', False, 35),
('Lead Source_Pay per Click Ads', False, 43),
('Lead Source_Press_Release', False, 53),
('Lead Source_Reference', True, 1),
('Lead Source_Referral Sites', False, 37),
('Lead Source_Social Media', False, 58),
('Lead Source_WeLearn', False, 42),
('Lead Source_Welingak Website', True, 1),
('Lead Source_bing', False, 33),
('Lead Source_testone', False, 38),
('Do Not Email_Yes', True, 1),
('Last Activity_Converted to Lead', False, 25),
('Last Activity_Email Bounced', False, 4),
('Last Activity_Email Link Clicked', False, 49),
('Last Activity_Email Marked Spam', False, 57),
('Last Activity_Email Opened', False, 41),
('Last Activity_Email Received', False, 54),
('Last Activity_Form Submitted on Website', False, 28),
('Last Activity_Had a Phone Conversation', True, 1),
('Last Activity_Olark Chat Conversation', False, 5),
('Last Activity_Page Visited on Website', False, 26),
('Last Activity_SMS Sent', True, 1),
('Last Activity_Unreachable', False, 47),
('Last Activity_Unsubscribed', False, 40),
('Last Activity_View in browser link Clicked', False, 34),
('Last Activity_Visited Booth in Tradeshow', False, 48),
('What is your current occupation_Housewife', True, 1),
('What is your current occupation_Other', False, 46),
('What is your current occupation_Student', True, 1),
('What is your current occupation_Unemployed', True, 1),
('What is your current occupation_Working Professional', True, 1),
('A free copy of Mastering The Interview_Yes', False, 50),
('Last Notable Activity_Email Bounced', False, 3),
('Last Notable Activity_Email Link Clicked', False, 20),
('Last Notable Activity_Email Marked Spam', False, 59),
('Last Notable Activity_Email Opened', False, 27),
('Last Notable Activity_Email Received', False, 60),
('Last Notable Activity_Had a Phone Conversation', True, 1),
('Last Notable Activity_Modified', False, 2),
('Last Notable Activity_Olark Chat Conversation', False, 32),
('Last Notable Activity_Page Visited on Website', False, 31),
('Last Notable Activity_SMS Sent', False, 45),
('Last Notable Activity_Unreachable', True, 1),
```

```
 ('Last Notable Activity_Unsubscribed', False, 39),
 ('Last Notable Activity_View in browser link Clicked', False, 29),
 ('Specialization_Banking, Investment And Insurance', False, 6),
 ('Specialization_Business Administration', False, 15),
 ('Specialization_E-Business', False, 11),
 ('Specialization_E-COMMERCE', False, 9),
 ('Specialization_Finance Management', False, 14),
 ('Specialization_Healthcare Management', False, 10),
 ('Specialization_Hospitality Management', False, 55),
 ('Specialization_Human Resource Management', False, 16),
 ('Specialization_IT Projects Management', False, 18),
 ('Specialization_International Business', False, 22),
 ('Specialization_Marketing Management', False, 12),
 ('Specialization_Media and Advertising', False, 21),
 ('Specialization_Operations Management', False, 19),
 ('Specialization_Retail Management', False, 30),
 ('Specialization_Rural and Agribusiness', False, 7),
 ('Specialization_Services Excellence', False, 56),
 ('Specialization_Supply Chain Management', False, 13),
 ('Specialization_Travel and Tourism', False, 17)]
```

```python
col = X_train.columns[rfe.support_]

X_train = X_train[col]

X_train_sm = sm.add_constant(X_train)
logm2 = sm.GLM(y_train, X_train_sm, family = sm.families.Binomial())
res = logm2.fit()
res.summary()
```

```
C:\Users\msi\anaconda3\lib\site-packages\statsmodels\tsa\
tsatools.py:142: FutureWarning: In a future version of pandas all
arguments of concat except for the argument 'objs' will be keyword-
only.
  x = pd.concat(x[::order], 1)
```

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                  Generalized Linear Model Regression Results

==============================================================================
========
Dep. Variable:                  Converted   No. Observations:
4461
Model:                                GLM   Df Residuals:
4445
Model Family:                    Binomial   Df Model:
15
Link Function:                      logit   Scale:
1.0000
```

```
Method:                          IRLS    Log-Likelihood:
-2072.8
Date:                 Tue, 28 May 2024    Deviance:
4145.5
Time:                        19:23:55    Pearson chi2:
4.84e+03
No. Iterations:                     22

Covariance Type:              nonrobust
```

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.0061	0.600	-1.677	0.094	-2.182	0.170
TotalVisits	11.3439	2.682	4.230	0.000	6.088	16.600
Total Time Spent on Website	4.4312	0.185	23.924	0.000	4.068	4.794
Lead Origin_Lead Add Form	2.9483	1.191	2.475	0.013	0.614	5.283
Lead Source_Olark Chat	1.4584	0.122	11.962	0.000	1.219	1.697
Lead Source_Reference	1.2994	1.214	1.070	0.285	-1.080	3.679
Lead Source_Welingak Website	3.4159	1.558	2.192	0.028	0.362	6.470
Do Not Email_Yes	-1.5053	0.193	-7.781	0.000	-1.884	-1.126
Last Activity_Had a Phone Conversation	1.0397	0.983	1.058	0.290	-0.887	2.966
Last Activity_SMS Sent	1.1827	0.082	14.362	0.000	1.021	1.344
What is your current occupation_Housewife	22.6492	2.45e+04	0.001	0.999	-4.8e+04	4.8e+04
What is your current occupation_Student	-1.1544	0.630	-1.831	0.067	-2.390	0.081
What is your current occupation_Unemployed	-1.3395	0.594	-2.254	0.024	-2.505	-0.175
What is your current occupation_Working Professional	1.2743	0.623	2.045	0.041	0.053	2.496
Last Notable Activity_Had a Phone Conversation	23.1932	2.08e+04	0.001	0.999	-4.08e+04	4.08e+04
Last Notable Activity_Unreachable	2.7868	0.807	3.453	0.001	1.205	4.369

```
==================================================
"""

y_train_pred = res.predict(sm.add_constant(X_train))
y_train_pred[:10]

C:\Users\msi\anaconda3\lib\site-packages\statsmodels\tsa\
tsatools.py:142: FutureWarning: In a future version of pandas all
arguments of concat except for the argument 'objs' will be keyword-
only.
  x = pd.concat(x[::order], 1)

8003    0.299132
218     0.141834
4171    0.127602
4037    0.291679
3660    0.956262
207     0.194437
2044    0.177750
6411    0.952857
6498    0.075665
2085    0.982340
dtype: float64

y_train_pred = y_train_pred.values.reshape(-1)
y_train_pred[:10]

array([0.29913183, 0.14183438, 0.12760155, 0.29167908, 0.95626237,
       0.19443676, 0.17775018, 0.95285672, 0.07566484, 0.98233985])

y_train_pred_final = pd.DataFrame({'Converted':y_train.values,
'Conversion_Prob':y_train_pred})
y_train_pred_final.head()

    Converted  Conversion_Prob
0           0          0.299132
1           0          0.141834
2           1          0.127602
3           1          0.291679
4           1          0.956262

y_train_pred_final['Predicted'] =
y_train_pred_final.Conversion_Prob.map(lambda x: 1 if x > 0.5 else 0)

y_train_pred_final.head()

    Converted  Conversion_Prob  Predicted
0           0          0.299132          0
1           0          0.141834          0
2           1          0.127602          0
```

| 3 | 1 | 0.291679 | 0 |
| 4 | 1 | 0.956262 | 1 |