```
# Import required libraries
import numpy as np, pandas as pd
import os, sys, glob
import matplotlib
matplotlib.rcParams['figure.figsize'] = (10, 6)
import seaborn as sns
from sklearn.model_selection import train_test, cross_validation,metrics
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import GradientBoostingClassifier, AdaBoostClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier, export_graphviz
```

```
# Input data files are available in the "../input/" directory.
# For example, running this (by clicking run or pressing Shift+Enter) will list the files in the input directory
ROOT = os.path.join('./')
print(os.listdir("../input/"))
```

```
['Bank_Marketing_Campaign_Results.csv', 'Bank_Marketing_Campaign_Project.', '.ipynb', 'Marital_Status%2C%20Total%20Pop.csv']
```

```
data_frame = pd.read_csv('../input/Bank_Marketing_Campaign_Results.csv')
data_frame.head(10)
```

	cust_age	job	marital	edu	def	bal	hous	loan	cont	day	mon	dur	camp	pdays	prev	out	y
0	58	4	1	3	0	2143	1	0	2	5	8	261	1	-1	0	0	0
1	44	9	2	2	0	29	1	0	2	5	8	151	1	-1	0	0	0
2	33	2	1	2	0	2	1	1	2	5	8	76	1	-1	0	0	0
3	47	1	1	4	0	1506	1	0	2	5	8	92	1	-1	0	0	0
4	33	11	2	4	0	1	0	0	2	5	8	198	1	-1	0	0	0
5	35	5	1	6	0	231	1	0	2	5	8	139	1	-1	0	0	0
6	28	5	1	6	0	447	1	1	2	5	8	217	1	-1	0	0	0
7	42	2	2	6	0	2	1	0	2	5	8	380	1	-1	0	0	0
8	58	4	1	3	0	121	1	1	2	5	8	50	1	-1	0	0	0
9	43	9	2	2	0	693	1	0	2	5	8	55	1	-1	0	0	0

9 rows × 17 columns

```
print("Number of rows in dataset",len(data_frame))
data_frame.shape
```

```
Number of rows in dataset 45211
(45211, 17)
```

```
# Count missing values in each column in the data
data_frame.isnull().sum(axis=0)
```

```
cust_age    0
job         0
marital     0
edu         0
def         0
bal         0
hous        0
loan        0
cont        0
day         0
mon         0
dur         0
camp        0
pdays       0
prev        0
out         0
y           0
dtype: int64
```

```
# Visualize the target variable
sns.countplot(data=data_frame, x='y', palette='hls')
plt.title('Distribution of Scores')
plt.show()
```

```
# Visualize some variable distributions
category_counts = data_frame.groupby('y').size()
print(category_counts)
sns.barplot(category_counts.index, category_counts.values, palette=['hls', 'husl'])
plt.title('Distribution of Scores')
plt.xlabel('y')
plt.show()
```

```
y
0    39922
1     5289
dtype: int64
```

```
category_counts
```

```
y
0    39922
1     5289
dtype: int64
```

```
# Visualize some variable distributions
data_df = data_frame[['cust_age', 'job', 'marital', 'edu', 'def', 'bal', 'hous', 'loan', 'cont', 'day', 'mon', 'dur', 'camp', 'pdays', 'prev', 'out', 'y']]
data_df.head(10)
```

	cust_age	job	marital	edu	def	bal	hous	loan	cont	day	mon	dur	camp	pdays	prev	out	y
0	58	4	1	3	0	2143	1	0	2	5	8	261	1	-1	0	0	0
1	44	9	2	2	0	29	1	0	2	5	8	151	1	-1	0	0	0
2	33	2	1	2	0	2	1	1	2	5	8	76	1	-1	0	0	0
3	47	1	1	4	0	1506	1	0	2	5	8	92	1	-1	0	0	0
4	33	11	2	4	0	1	0	0	2	5	8	198	1	-1	0	0	0
5	35	5	1	6	0	231	1	0	2	5	8	139	1	-1	0	0	0
6	28	5	1	6	0	447	1	1	2	5	8	217	1	-1	0	0	0
7	42	2	2	6	0	2	1	0	2	5	8	380	1	-1	0	0	0
8	58	4	1	3	0	121	1	1	2	5	8	50	1	-1	0	0	0
9	43	9	2	2	0	693	1	0	2	5	8	55	1	-1	0	0	0

```
# Count occurrences of each value in the 'cust_age' column
print("Count of each value in 'cust_age':")
print(data_df['cust_age'].value_counts())
```

```
Count of each value in 'cust_age':
32    2085
31    1996
33    1962
34    1930
30    1879
35    1801
36    1728
29    1626
37    1623
38    1475
39    1353
28    1293
40    1267
41    1080
27    1035
42    1012
43    1000
26    949
44    901
45    887
25    744
46    744
47    704
24    627
48    621
49    596
23    557
50    548
51    509
22    477
52    475
53    462
21    436
54    403
55    387
20    351
56    334
57    319
19    274
58    268
59    256
18    168
60    156
61    111
62    99
63    39
64    26
65    22
66    19
67    20
68    14
69    9
70    6
71    7
72    4
73    4
74    3
75    2
76    2
77    1
78    2
79    1
80    1
83    1
85    1
87    1
95    1
Name: cust_age, dtype: int64
```

```
# Count occurrences of each value in the 'job' column
print("Count of each value in 'job':")
print(data_df['job'].value_counts())
```

```
Count of each value in 'job':
1    9732
0    8307
9    7597
4    6848
7    5171
2    4954
3    4621
6    3816
5    3768
8    1487
10    938
11    168
Name: job, dtype: int64
```

```
# Count occurrences of each value in the 'marital' column
print("Count of each value in 'marital':")
print(data_df['marital'].value_counts())
```

```
Count of each value in 'marital':
1    27214
2    12790
0     5207
Name: marital, dtype: int64
```

```
# Count occurrences of each value in the 'edu' column
print("Count of each value in 'edu':")
print(data_df['edu'].value_counts())
```

```
Count of each value in 'edu':
1    23202
2    13301
3     6851
4     1857
0     1750
Name: edu, dtype: int64
```

```
# Count occurrences of each value in the 'def' column
print("Count of each value in 'def':")
print(data_df['def'].value_counts())
```

```
Count of each value in 'def':
0    44396
1      815
Name: def, dtype: int64
```

```
# Count occurrences of each value in the 'hous' column
print("Count of each value in 'hous':")
print(data_df['hous'].value_counts())
```

```
Count of each value in 'hous':
1    25130
0    20081
Name: hous, dtype: int64
```

```
# Count occurrences of each value in the 'loan' column
print("Count of each value in 'loan':")
print(data_df['loan'].value_counts())
```

```
Count of each value in 'loan':
0    37967
1     7244
Name: loan, dtype: int64
```

```
# Count occurrences of each value in the 'cont' column
print("Count of each value in 'cont':")
print(data_df['cont'].value_counts())
```

```
Count of each value in 'cont':
1    29285
2    13020
0     2906
Name: cont, dtype: int64
```

```
# Count occurrences of each value in the 'day' column
print("Count of each value in 'day':")
print(data_df['day'].value_counts())
```

```
Count of each value in 'day':
20    2752
18    2308
21    2026
17    1943
6     1932
5     1910
14    1603
8     1601
28    1393
7     1351
11    1303
13    1274
30    1262
9     1223
12    1080
15    1035
19    1035
2     1028
29     970
4      909
16     877
27     842
3      741
26     733
22     690
25     689
23     621
1      619
24     615
31     447
10     392
Name: day, dtype: int64
```

```
# Count occurrences of each value in the 'mon' column
print("Count of each value in 'mon':")
print(data_df['mon'].value_counts())
```

```
Count of each value in 'mon':
5    13766
7     6895
8     6352
6     5341
11    3970
4     3914
2     2932
10    2651
1     1303
9     1100
3     1035
12     477
Name: mon, dtype: int64
```

```
# Count occurrences of each value in the 'camp' column
print("Count of each value in 'camp':")
print(data_df['camp'].value_counts())
```

```
Count of each value in 'camp':
1    17544
2    12505
3     5521
4     3267
5     1773
6     1233
7      843
8      543
9      405
10     301
11     207
12     148
13     106
14      70
15      61
16      44
17      35
18      30
19      19
20      13
21      12
22      12
23      10
24       6
25       5
26       4
27       3
28       2
29       2
31       2
32       1
33       1
Name: camp, dtype: int64
```

```
# Count occurrences of each value in the 'pdays' column
print("Count of each value in 'pdays':")
print(data_df['pdays'].value_counts())
```

```
Count of each value in 'pdays':
-1    36954
182     167
92      147
183     126
91      116
184     113
181     112
370     110
365     107
364     105
183     104
185     103
90      102
186     101
187     100
188     100
189     100
190     100
191     100
192     100
193     100
194     100
195     100
196     100
197     100
198     100
199     100
200     100
201     100
202     100
203     100
204     100
205     100
206     100
```