

```
# Importing necessary libraries
import pandas as pd
import numpy as np
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.metrics import classification_report, confusion_matrix
```

```
df = pd.read_csv("/content/2011-Q1-cabi-trip-history-data.csv")
df.head()
```

	Duration	Start date	End date	Start station	End station	Bike#	Member Type
0	0h 1min. 50sec.	3/31/2011 23:58	4/1/2011 0:00	14th & Harvard St NW (31105)	16th & Harvard St NW (31103)	W00749	Registered
1	0h 16min. 21sec.	3/31/2011 23:52	4/1/2011 0:08	19th & L St NW (31224)	7th & Water St SW / SW Waterfront (31609)	W01048	Casual

Lincoln Park /

```
df["Duration"] = df['Duration'].str.replace('min', '')
df["Duration"] = df['Duration'].str.replace('sec', '')
df["Duration"] = df['Duration'].str.replace('h', '')
df["Duration"] = df['Duration'].str.replace(' ', '')
df["Duration"] = df['Duration'].str.replace('.', '').astype(float)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:5: FutureWarning: The default value of regex will change from True to
""
```

```
address=df['End station'].unique()
print(address)
df = df.dropna()
```

'10th & Monroe St NE (31504)' 'C & O Canal & Wisconsin Ave NW (31225)'
'18th & M St NW (31221)' '11th & Kenyon St NW (31102)'
'3rd & H St NE (31616)' 'Park Rd & Holmead Pl NW (31602)'
'Calvert St & Woodley Pl NW (31106)'
'Georgia & New Hampshire Ave NW (31400)' '10th & U St NW (31111)'
'Potomac & Pennsylvania Ave SE (31606)'
'Columbus Circle / Union Station (31623)' 'Florida Ave & R St NW (31503)'
'14th & D St SE (31607)' 'McPherson Square - 14th & H St NW (31216)'
'Lincoln Park / 13th & East Capitol St NE (31619)'
'13th & D St NE (31622)' '21st & I St NW (31205)'
'16th & U St NW (31229)' '12th & Army Navy Dr (31008)'
'14th & R St NW (31202)' '5th St & K St NW (31600)'
'19th St & Pennsylvania Ave NW (31100)'
'Wisconsin Ave & Macomb St NW (31302)' '4th & East Capitol St NE (31618)'
'3rd & D St SE (31605)' 'Van Ness Metro / UDC (31300)'
'Tenleytown / Wisconsin Ave & Albemarle St NW (31303)'
'1st & M St NE (31603)' '36th & Calvert St NW / Glover Park (31304)'
'S Glebe & Potomac Ave (31010)' '8th & H St NW (31228)'
'New York Ave & 15th St NW (31222)' '13th St & New York Ave NW (31227)'
'Lamont & Mt Pleasant NW (31107)' 'Eckington Pl & Q St NE (31505)'
'4th St & Massachusetts Ave NW (31604)'
'Eastern Market / 7th & North Carolina Ave SE (31610)'
'Convention Center / 7th & M St NW (31223)'
'M St & New Jersey Ave SE (31208)' 'Kennedy Center (31211)'
'14th & Harvard St NW (31105)'
'Connecticut Ave & Newark St NW / Cleveland Park (31305)'
'17th & K St NW [formerly 17th & L St NW] (31213)'
'14th & G St NW (31238)' '4th & M St SW (31108)'
'Eastern Market Metro / Pennsylvania Ave & 7th St SE (31613)'
'L'Enfant Plaza / 7th & C St SW (31218)' '19th & L St NW (31224)'
'20th St & Florida Ave NW (31110)'
'Harvard St & Adams Mill Rd NW (31112)'
'Ward Circle / American University (31301)'
'19th St & Constitution Ave NW (31235)'
'Bladensburg Rd & Benning Rd NE (31617)' '5th & F St NW (31620)'

```

12th & Hayes St (31001)' John McCormack Dr & Michigan Ave NE (31502)
'10th St & Constitution Ave NW (31219)'
'Georgetown Harbor / 30th St NW (31215)'
'19th & East Capitol St SE (31601)' '1st & N St SE (31209)'
'19th & E Street NW (31206)' 'Pennsylvania & Minnesota Ave SE (31805)'
'Anacostia Library (31804)' '8th & Eye St SE / Barracks Row (31608)'
'23rd & Eads (31013)' '26th & Crystal Dr (31012)'
'Anacostia Metro (31801)' '15th & Hayes St (31005)'
'4th & Adams St NE (31500)' '20th & Bell St (31000)'
'14th St Heights / 14th & Crittenden St NW (31402)' nan
'Minnesota Ave Metro/DOES (31703)'
'Randle Circle & Minnesota Ave NE (31702)'
'Good Hope Rd & MLK Ave SE (31802)' 'Benning Branch Library (31705)'
'Nannie Helen Burroughs & Minnesota Ave NE (31704)'
'Alta Bicycle Share Demonstration Station (31999)']

```

```

from sklearn.preprocessing import LabelEncoder
gle = LabelEncoder()
address_labels = gle.fit_transform(df['End station'])
address_labels = gle.fit_transform(df['Start station'])
address_mappings = {index: label for index, label in
                    enumerate(gle.classes_)}

address_mappings

48: '37th & O St NW / Georgetown University (31236)',
49: '3rd & D St SE (31605)',
50: '3rd & H St NE (31616)',
51: '4th & Adams St NE (31500)',
52: '4th & East Capitol St NE (31618)',
53: '4th & M St SW (31108)',
54: '4th St & Massachusetts Ave NW (31604)',
55: '5th & F St NW (31620)',
56: '5th St & K St NW (31600)',
57: '7th & T St NW (31109)',
58: '7th & Water St SW / SW Waterfront (31609)',
59: '8th & Eye St SE / Barracks Row (31608)',
60: '8th & H St NW (31228)',
61: 'Adams Mill & Columbia Rd NW (31104)',
62: 'Alta Bicycle Share Demonstration Station (31999)',
63: 'Anacostia Library (31804)',
64: 'Anacostia Metro (31801)',
65: 'Benning Branch Library (31705)',
66: 'Bladensburg Rd & Benning Rd NE (31617)',
67: 'C & O Canal & Wisconsin Ave NW (31225)',
68: 'Calvert St & Woodley Pl NW (31106)',
69: 'Columbus Circle / Union Station (31623)',
70: 'Connecticut Ave & Newark St NW / Cleveland Park (31305)',
71: 'Convention Center / 7th & M St NW (31223)',
72: 'Eastern Market / 7th & North Carolina Ave SE (31610)',
73: 'Eastern Market Metro / Pennsylvania Ave & 7th St SE (31613)',
74: 'Eckington Pl & Q St NE (31505)',
75: 'Florida Ave & R St NW (31503)',
76: 'Georgetown Harbor / 30th St NW (31215)',
77: 'Georgia & New Hampshire Ave NW (31400)',
78: 'Georgia Ave and Fairmont St NW (31207)',
79: 'Good Hope & Naylor Rd SE (31700)',
80: 'Good Hope Rd & MLK Ave SE (31802)',
81: 'Harvard St & Adams Mill Rd NW (31112)',
82: 'John McCormack Dr & Michigan Ave NE (31502)',
83: 'Kennedy Center (31211)',
84: 'L'Enfant Plaza / 7th & C St SW (31218)',
85: 'Lamont & Mt Pleasant NW (31107)',
86: 'Lincoln Park / 13th & East Capitol St NE (31619)',
87: 'M St & New Jersey Ave SE (31208)',
88: 'Massachusetts Ave & Dupont Circle NW (31200)',
89: 'McPherson Square - 14th & H St NW (31216)',
90: 'Minnesota Ave Metro/DOES (31703)',
91: 'Nannie Helen Burroughs & Minnesota Ave NE (31704)',
92: 'New York Ave & 15th St NW (31222)',
93: 'North Capitol St & F St NW (31624)',
94: 'Park Rd & Holmead Pl NW (31602)',
95: 'Pennsylvania & Minnesota Ave SE (31805)',
96: 'Potomac & Pennsylvania Ave SE (31606)',
97: 'Randle Circle & Minnesota Ave NE (31702)',
98: 'S Glebe & Potomac Ave (31010)',
99: 'S Joyce & Army Navy Dr (31006)',
100: 'Tenleytown / Wisconsin Ave & Albemarle St NW (31303)',
101: 'US Dept of State / Virginia Ave & 21st St NW (31220)',
102: 'USDA / 12th & Independence Ave SW (31217)',
103: 'Van Ness Metro / UDC (31300)',
104: 'Ward Circle / American University (31301)',
105: 'Wisconsin Ave & Macomb St NW (31302)'}

```

```
df['End station'] = address_labels
```

```

gle = LabelEncoder()
address_labels = gle.fit_transform(df['Start station'])
address_mappings = {index: label for index, label in

```

```

        enumerate(gle.classes_))
address_mappings

48: '37th & O St NW / Georgetown University (31236)',
49: '3rd & D St SE (31605)',
50: '3rd & H St NE (31616)',
51: '4th & Adams St NE (31500)',
52: '4th & East Capitol St NE (31618)',
53: '4th & M St SW (31108)',
54: '4th St & Massachusetts Ave NW (31604)',
55: '5th & F St NW (31620)',
56: '5th St & K St NW (31600)',
57: '7th & T St NW (31109)',
58: '7th & Water St SW / SW Waterfront (31609)',
59: '8th & Eye St SE / Barracks Row (31608)',
60: '8th & H St NW (31228)',
61: 'Adams Mill & Columbia Rd NW (31104)',
62: 'Alta Bicycle Share Demonstration Station (31999)',
63: 'Anacostia Library (31804)',
64: 'Anacostia Metro (31801)',
65: 'Benning Branch Library (31705)',
66: 'Bladensburg Rd & Benning Rd NE (31617)',
67: 'C & O Canal & Wisconsin Ave NW (31225)',
68: 'Calvert St & Woodley Pl NW (31106)',
69: 'Columbus Circle / Union Station (31623)',
70: 'Connecticut Ave & Newark St NW / Cleveland Park (31305)',
71: 'Convention Center / 7th & M St NW (31223)',
72: 'Eastern Market / 7th & North Carolina Ave SE (31610)',
73: 'Eastern Market Metro / Pennsylvania Ave & 7th St SE (31613)',
74: 'Eckington Pl & Q St NE (31505)',
75: 'Florida Ave & R St NW (31503)',
76: 'Georgetown Harbor / 30th St NW (31215)',
77: 'Georgia & New Hampshire Ave NW (31400)',
78: 'Georgia Ave and Fairmont St NW (31207)',
79: 'Good Hope & Naylor Rd SE (31700)',
80: 'Good Hope Rd & MLK Ave SE (31802)',
81: 'Harvard St & Adams Mill Rd NW (31112)',
82: 'John McCormack Dr & Michigan Ave NE (31502)',
83: 'Kennedy Center (31211)',
84: 'L'Enfant Plaza / 7th & C St SW (31218)",
85: 'Lamont & Mt Pleasant NW (31107)',
86: 'Lincoln Park / 13th & East Capitol St NE (31619)',
87: 'M St & New Jersey Ave SE (31208)',
88: 'Massachusetts Ave & Dupont Circle NW (31200)',
89: 'McPherson Square - 14th & H St NW (31216)',
90: 'Minnesota Ave Metro/DOES (31703)',
91: 'Nannie Helen Burroughs & Minnesota Ave NE (31704)',
92: 'New York Ave & 15th St NW (31222)',
93: 'North Capitol St & F St NW (31624)',
94: 'Park Rd & Holmead Pl NW (31602)',
95: 'Pennsylvania & Minnesota Ave SE (31805)',
96: 'Potomac & Pennsylvania Ave SE (31606)',
97: 'Randle Circle & Minnesota Ave NE (31702)',
98: 'S Glebe & Potomac Ave (31010)',
99: 'S Joyce & Army Navy Dr (31006)',
100: 'Tenleytown / Wisconsin Ave & Albemarle St NW (31303)',
101: 'US Dept of State / Virginia Ave & 21st St NW (31220)',
102: 'USDA / 12th & Independence Ave SW (31217)',
103: 'Van Ness Metro / UDC (31300)',
104: 'Ward Circle / American University (31301)',
105: 'Wisconsin Ave & Macomb St NW (31302)'}

df['Start station'] = address_labels

df.head()
```

	Duration	Start date	End date	Start station	End station	Bike#	Member Type
0	150.0	3/31/2011 23:58	4/1/2011 0:00	12	21	W00749	Registered
1	1621.0	3/31/2011 23:52	4/1/2011 0:08	30	58	W01048	Casual
2	319.0	3/31/2011 23:47	3/31/2011 23:50	86	8	W00340	Registered

```

gle = LabelEncoder()
bike_labels = gle.fit_transform(df['Bike#'])
bike_mappings = {index: label for index, label in
                  enumerate(gle.classes_)}
bike_mappings

{0: '? (0x26B3BBA8)',
1: '? (0x3EB026B9)',
2: '? (0x4752DD3A)',
```

```

3: '? (0x7C120F6A)',
4: '? (0x9A5FEA16)',
5: '? (0xAAC5A4C0)',
6: '? (0xEBA95C18)',
7: '? (0xEE8DADF1)',
8: '? (0xFAF0B948)',
9: 'W00005',
10: 'W00006',
11: 'W00007',
12: 'W00008',
13: 'W00009',
14: 'W00010',
15: 'W00011',
16: 'W00012',
17: 'W00013',
18: 'W00014',
19: 'W00015',
20: 'W00017',
21: 'W00018',
22: 'W00019',
23: 'W00021',
24: 'W00024',
25: 'W00025',
26: 'W00026',
27: 'W00027',
28: 'W00028',
29: 'W00030',
30: 'W00031',
31: 'W00032',
32: 'W00033',
33: 'W00035',
34: 'W00036',
35: 'W00037',
36: 'W00038',
37: 'W00039',
38: 'W00040',
39: 'W00041',
40: 'W00042',
41: 'W00043',
42: 'W00044',
43: 'W00045',
44: 'W00046',
45: 'W00048',
46: 'W00049',
47: 'W00050',
48: 'W00051',
49: 'W00055',
50: 'W00057',
51: 'W00058',
52: 'W00060',
53: 'W00061',
54: 'W00062',
55: 'W00063',
56: 'W00064',
57: 'W00065',

```

```
df['Bike#'] = bike_labels
```

```
# Assign X and y
```

```
X = df.iloc[:, [0, 3, 5]].values
```

```
y = df.iloc[:, -1].values
```

```
print(y)
```

```
['Registered' 'Casual' 'Registered' ... 'Registered' 'Casual' 'Registered']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=0)
```

```
tree = DecisionTreeClassifier(max_leaf_nodes=3, random_state=0)
```

```
tree.fit(X_train, y_train)
```

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

```
# Check the Accuracy
```

```
score = metrics.accuracy_score(y_test, y_pred)
```

```
print("Accuracy of our model is: {:.1f}%".format(score*100))
```

```
print(confusion_matrix(y_test, y_pred))
```

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

```
Accuracy of our model is: 88.4%
```

```
[[ 1264  3100]
```

```
 [  379 25269]]
```

```
precision    recall  f1-score   support
```

```
Casual      0.77      0.29      0.42      4364
```

Registered	0.89	0.99	0.94	25648
accuracy			0.88	30012
macro avg	0.83	0.64	0.68	30012
weighted avg	0.87	0.88	0.86	30012