

```

In [241]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

In [242]: df = pd.read_csv('accident data.csv')

In [243]: df

Out[243]:
   Date AccidentTime WeekDay RegistrationDateForSubjectData AccidentCategory NumberOfUnits RoadType LocationC
0  28/6/2009      21:12   Sunday           11/24/2010           Car accident             2.0   Ordinary
   street road
1  5/6/2012         1:14   Tuesday           6/5/2012           Lone vehicle
   overturned in the
2  26/6/2018         7:50   Tuesday           26/6/2018           Lone vehicle
   overturned in the
3  10/7/2018        23:01   Tuesday           10/7/2018           Lone vehicle
   overturned in the
4  6/5/2019         15:47   Tuesday           5/6/2019           Lone vehicle
   overturned in the
5  1/3/2021         NaN    Monday           1/3/2021             NaN             NaN    Ordinary
   street road

6 rows x 32 columns

In [244]: df=df.drop(['AccidentTime','WeekDay','RegistrationDateForSubjectData','AccidentCategory','Nu
numberOfUnits','RoadType','Weather','NumberOfLanes','Town','LocationAccuracy','AccidentType',
'AccidentTypeSubcategory','NumberOfCars'], axis=1)

In [245]: df

Out[245]:
   Date LocationConditions TireType Conditions LightingConditions RoadWidthInMeters SpeedLimitInKilometerPerHour
0  28/6/2009           Tunnel   Asphalt   Dry   Dark with road
   lighting             10             80
1  5/6/2012           Tunnel   Asphalt   Dry   Dark with road
   lighting             7             80
2  26/6/2018           Tunnel   Asphalt   Dry   Dark with road
   lighting             7             80
3  10/7/2018           Tunnel   Asphalt   Dry   Dark with road
   lighting             7             80
4  6/5/2019           Tunnel   Asphalt   Dry   Dark with road
   lighting             7             80
5  1/3/2021           Tunnel   Asphalt   Dry   Clear light             10             80

In [246]: df=df.drop(['LightingConditions'],axis=1)

In [247]: df=df.drop(['Conditions'],axis=1)

In [248]: df=df.drop(['LocationConditions','TireType'],axis=1)

In [249]: df=df.drop(['Date'],axis=1)

In [251]: df=df.drop(['MunicipalityName','MunicipalityName'], axis=1)

In [253]: df=df.drop(['CountyName'], axis=1)

In [254]: df

Out[254]:
   RoadWidthInMeters SpeedLimitInKilometerPerHour MunicipalityNumber Year ADT CountyNumber TemperatureInDegreeCel
0             10             80             1516  2009  1600             15
1              7             80             1577  2012  2030             15
2              7             80             1520  2018  3130             15
3              7             80             1520  2018  3130             15
4              7             80             1520  2019  3130             15
5             10             80             1516  2021  3100             15

In [255]: df.size

Out[255]: 72

In [256]: df.shape

Out[256]: (6, 12)

In [257]: x = df.iloc[:, 0:11]

In [258]: x

Out[258]:
   RoadWidthInMeters SpeedLimitInKilometerPerHour MunicipalityNumber Year ADT CountyNumber TemperatureInDegreeCel
0             10             80             1516  2009  1600             15
1              7             80             1577  2012  2030             15
2              7             80             1520  2018  3130             15
3              7             80             1520  2018  3130             15
4              7             80             1520  2019  3130             15
5             10             80             1516  2021  3100             15

In [259]: y =df.iloc[:, -1]

In [260]: y

Out[260]:
0    1
1    1
2    1
3    1
4    1
5    0
Name: Result, dtype: int64

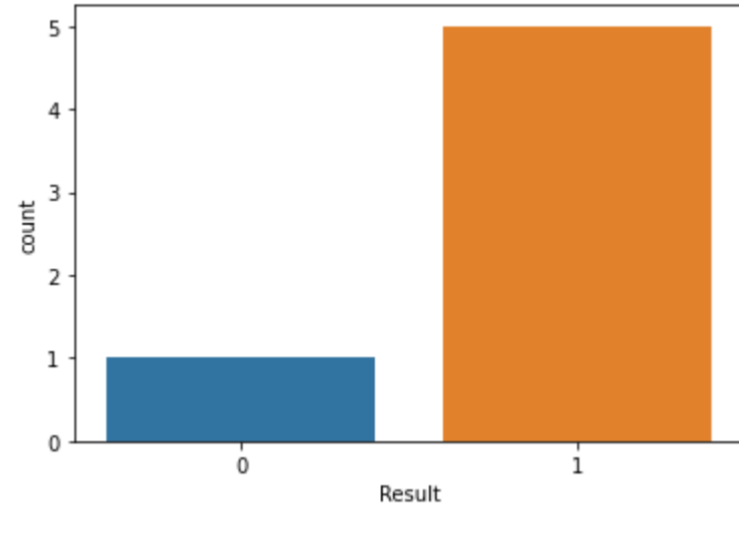
In [261]: df['Result'].value_counts()

Out[261]:
1    5
0    1
Name: Result, dtype: int64

In [262]: import seaborn as sns

In [263]: sns.countplot(x='Result',data=df)

Out[263]: <matplotlib.axes._subplots.AxesSubplot at 0x23d47968f40>



In [264]: x = df.iloc[:, 0:11]
y =df.iloc[:, -1]

In [265]: from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test=train_test_split(x,y, test_size=0.2, random_state=1)

In [266]: from sklearn.linear_model import LogisticRegression
logmodel=LogisticRegression()
logmodel.fit(X_train,Y_train)

Out[266]: LogisticRegression()

In [268]: Y_pred=logmodel.predict(X_test)
print("Accuracy", (logmodel.score(X_test, Y_test)))
from sklearn.metrics import confusion_matrix
confusion_matrix=confusion_matrix(Y_test,Y_pred)
print(confusion_matrix)

Accuracy 0.5
<function confusion_matrix at 0x000023D47C5DC10>

In [246]: df=df.drop(['LightingConditions'],axis=1)

In [247]: df=df.drop(['Conditions'],axis=1)

In [248]: df=df.drop(['LocationConditions','TireType'],axis=1)

In [249]: df=df.drop(['Date'],axis=1)

In [251]: df=df.drop(['MunicipalityName','MunicipalityName'], axis=1)

In [253]: df=df.drop(['CountyName'], axis=1)

In [254]: df

Out[254]:
   RoadWidthInMeters SpeedLimitInKilometerPerHour MunicipalityNumber Year ADT CountyNumber TemperatureInDegreeCel
0             10             80             1516  2009  1600             15
1              7             80             1577  2012  2030             15
2              7             80             1520  2018  3130             15
3              7             80             1520  2018  3130             15
4              7             80             1520  2019  3130             15
5             10             80             1516  2021  3100             15

In [255]: df.size

Out[255]: 72

In [256]: df.shape

Out[256]: (6, 12)

In [257]: x = df.iloc[:, 0:11]

In [258]: x

Out[258]:
   RoadWidthInMeters SpeedLimitInKilometerPerHour MunicipalityNumber Year ADT CountyNumber TemperatureInDegreeCel
0             10             80             1516  2009  1600             15
1              7             80             1577  2012  2030             15
2              7             80             1520  2018  3130             15
3              7             80             1520  2018  3130             15
4              7             80             1520  2019  3130             15
5             10             80             1516  2021  3100             15

In [259]: y =df.iloc[:, -1]

In [260]: y

Out[260]:
0    1
1    1
2    1
3    1
4    1
5    0
Name: Result, dtype: int64

In [261]: df['Result'].value_counts()

Out[261]:
1    5
0    1
Name: Result, dtype: int64

In [262]: import seaborn as sns

In [263]: sns.countplot(x='Result',data=df)

Out[263]: <matplotlib.axes._subplots.AxesSubplot at 0x23d47968f40>



In [264]: x = df.iloc[:, 0:11]
y =df.iloc[:, -1]

In [265]: from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test=train_test_split(x,y, test_size=0.2, random_state=1)

In [266]: from sklearn.linear_model import LogisticRegression
logmodel=LogisticRegression()
logmodel.fit(X_train,Y_train)

Out[266]: LogisticRegression()

In [268]: Y_pred=logmodel.predict(X_test)
print("Accuracy", (logmodel.score(X_test, Y_test)))
from sklearn.metrics import confusion_matrix
confusion_matrix=confusion_matrix(Y_test,Y_pred)
print(confusion_matrix)

Accuracy 0.5
<function confusion_matrix at 0x000023D47C5DC10>

In [ ]:

```