Muley_Tushar_Project_1

January 9, 2022

Name: Tushar Muley

Assignment: Project 1 - Human and Economic Cost of Hurricanes

Date: Jan 9, 2022

[1]:
```
# Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
import math
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

[2]:
```
# Setting changes
pd.set_option('display.max_columns', None)
```

[3]:
```
# Load data into a dataframe
file = "natural-disasters.csv"

disaster_df = pd.read_csv(file)
```

[5]:
```
# Check the dimension of the table/look at the data
print("The dimension of the table is: ", disaster_df.shape)
```

```
The dimension of the table is:  (5569, 169)
```

[6]:
```
# View the information on the data
disaster_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5569 entries, 0 to 5568
Columns: 169 entries, Number of deaths from drought to
total_damages_pct_gdp_glacial_lake
```

```
dtypes: float64(159), int64(9), object(1)
memory usage: 7.2+ MB
```

[7]:
```python
# Print the columns names
print("Column names: ", disaster_df.columns)
```

```
Column names:  Index(['Number of deaths from drought',
       'Number of people injured from drought',
       'Number of people affected from drought',
       'Number of people left homeless from drought',
       'Number of total people affected by drought',
       'Reconstruction costs from drought', 'Insured damages against drought',
       'Total economic damages from drought', 'Death rates from drought',
       'Injury rates from drought',
       …
       'Total economic damages from extreme temperatures as a share of GDP',
       'Total economic damages from floods as a share of GDP',
       'Total economic damages from landslides as a share of GDP',
       'Total economic damages from mass movements as a share of GDP',
       'Total economic damages from storms as a share of GDP',
       'Total economic damages from volcanic activity as a share of GDP',
       'Total economic damages from volcanic activity as a share of GDP.1',
       'Entity', 'Year', 'total_damages_pct_gdp_glacial_lake'],
      dtype='object', length=169)
```

[8]:
```python
# What type of variables are in the table before dropping variables.
print("Describe Data")
print(disaster_df.describe())
```

```
Describe Data
       Number of deaths from drought  Number of people injured from drought  \
count                   8.120000e+02                             812.000000
mean                    2.889513e+04                               0.078818
std                     2.234236e+05                               1.587154
min                     0.000000e+00                               0.000000
25%                     0.000000e+00                               0.000000
50%                     0.000000e+00                               0.000000
75%                     0.000000e+00                               0.000000
max                     3.000000e+06                              32.000000

       Number of people affected from drought  \
count                            8.120000e+02
mean                             6.791441e+06
std                              3.268972e+07
min                              0.000000e+00
25%                              0.000000e+00
50%                              1.975000e+05
75%                              1.890100e+06
max                              3.829851e+08
```

```
       Number of people left homeless from drought  \
count                                        812.0
mean                                           0.0
std                                            0.0
min                                            0.0
25%                                            0.0
50%                                            0.0
75%                                            0.0
max                                            0.0

       Number of total people affected by drought  \
count                                8.120000e+02
mean                                 6.791441e+06
std                                  3.268972e+07
min                                  0.000000e+00
25%                                  0.000000e+00
50%                                  1.975000e+05
75%                                  1.890100e+06
max                                  3.829851e+08

       Reconstruction costs from drought  Insured damages against drought  \
count                              812.0                     8.120000e+02
mean                                 0.0                     5.161084e+04
std                                  0.0                     8.027164e+05
min                                  0.0                     0.000000e+00
25%                                  0.0                     0.000000e+00
50%                                  0.0                     0.000000e+00
75%                                  0.0                     0.000000e+00
max                                  0.0                     1.600000e+07

       Total economic damages from drought  Death rates from drought  \
count                         8.120000e+02                805.000000
mean                          4.554066e+05                121.478088
std                           1.909631e+06               1239.616752
min                           0.000000e+00                  0.000000
25%                           0.000000e+00                  0.000000
50%                           0.000000e+00                  0.000000
75%                           1.570000e+04                  0.000000
max                           2.548120e+07              17699.115044

       Injury rates from drought  \
count                 805.000000
mean                    0.000094
std                     0.002523
min                     0.000000
25%                     0.000000
50%                     0.000000
```

```
75%                     0.000000
max                     0.071459

       Number of people affected by drought per 100,000  \
count                                         805.000000
mean                                         8788.592155
std                                         17878.585188
min                                             0.000000
25%                                             0.000000
50%                                          1446.591830
75%                                          8838.902115
max                                        120967.741935

       Homelessness rate from drought  \
count                           805.0
mean                              0.0
std                               0.0
min                               0.0
25%                               0.0
50%                               0.0
75%                               0.0
max                               0.0

       Total number of people affected by drought per 100,000  \
count                                         805.000000
mean                                         8788.592250
std                                         17878.585143
min                                             0.000000
25%                                             0.000000
50%                                          1446.591830
75%                                          8838.902115
max                                        120967.741935

       Number of deaths from earthquakes  \
count                        1148.000000
mean                         4070.986063
std                         21293.115522
min                             0.000000
25%                             2.000000
50%                            27.000000
75%                           360.250000
max                        277005.000000

       Number of people injured from earthquakes  \
count                                1148.000000
mean                                 4886.367596
std                                 26869.490001
min                                     0.000000
```

```
25%                                       0.000000
50%                                      55.500000
75%                                     513.500000
max                                  370939.000000

       Number of people affected by earthquakes  \
count                                 1.148000e+03
mean                                  3.027026e+05
std                                   2.240501e+06
min                                   0.000000e+00
25%                                   0.000000e+00
50%                                   5.000000e+02
75%                                   2.252475e+04
max                                   4.672439e+07

       Number of people left homeless from earthquakes  \
count                                        1.148000e+03
mean                                         4.410635e+04
std                                          2.659460e+05
min                                          0.000000e+00
25%                                          0.000000e+00
50%                                          0.000000e+00
75%                                          3.038000e+03
max                                          5.151700e+06

       Number of total people affected by earthquakes  \
count                                       1.148000e+03
mean                                        3.516953e+05
std                                         2.325692e+06
min                                         0.000000e+00
25%                                         1.200000e+02
50%                                         3.316500e+03
75%                                         4.457050e+04
max                                         4.751247e+07

       Reconstruction costs from earthquakes  \
count                              1.148000e+03
mean                               1.151556e+05
std                                1.606288e+06
min                                0.000000e+00
25%                                0.000000e+00
50%                                0.000000e+00
75%                                0.000000e+00
max                                3.500000e+07

       Insured damages against earthquakes  \
count                            1.148000e+03
mean                             1.785274e+05
```

```
std                              2.065232e+06
min                              0.000000e+00
25%                              0.000000e+00
50%                              0.000000e+00
75%                              0.000000e+00
max                              5.169400e+07

       Total economic damages from earthquakes  Death rates from earthquakes  \
count                              1.148000e+03                   1121.000000
mean                               1.457443e+06                     10.163333
std                                1.125521e+07                     80.413816
min                                0.000000e+00                      0.000000
25%                                0.000000e+00                      0.005047
50%                                0.000000e+00                      0.071841
75%                                5.000000e+04                      1.041749
max                                2.302998e+08                   2237.109257

       Injury rates from earthquakes  \
count                    1121.000000
mean                       11.667370
std                       113.956132
min                         0.000000
25%                         0.000000
50%                         0.122903
75%                         1.047073
max                      3015.378430

       Number of people affected by earthquakes per 100,000  \
count                                        1121.000000
mean                                          437.986043
std                                          2759.493505
min                                             0.000000
25%                                             0.000000
50%                                             1.257862
75%                                            46.355138
max                                         56826.791938

       Homelessness rate from earthquakes  \
count                         1121.000000
mean                            82.779595
std                            723.743783
min                              0.000000
25%                              0.000000
50%                              0.000000
75%                              4.382275
max                          17669.343840

       Total number of people affected by earthquakes per 100,000  \
```

```
count                                        1121.000000
mean                                          532.433009
std                                          3317.209327
min                                             0.000000
25%                                             0.274970
50%                                             8.816992
75%                                            93.174936
max                                         75662.979239

       Number of deaths from disasters  \
count                     5.569000e+03
mean                      8.194785e+03
std                       1.193417e+05
min                       0.000000e+00
25%                       1.000000e+00
50%                       1.600000e+01
75%                       1.150000e+02
max                       3.718976e+06

       Number of people injured from disasters  \
count                             5.569000e+03
mean                              2.754367e+03
std                               4.035970e+04
min                               0.000000e+00
25%                               0.000000e+00
50%                               0.000000e+00
75%                               3.500000e+01
max                               1.884112e+06

       Number of people affected by disasters  \
count                            5.569000e+03
mean                             2.882401e+06
std                              2.270438e+07
min                              0.000000e+00
25%                              0.000000e+00
50%                              5.000000e+03
75%                              1.100000e+05
max                              6.570457e+08

       Number of people left homeless from disasters  \
count                                   5.569000e+03
mean                                    6.375854e+04
std                                     6.909609e+05
min                                     0.000000e+00
25%                                     0.000000e+00
50%                                     0.000000e+00
75%                                     5.000000e+02
max                                     2.942487e+07
```

```
       Number of total people affected by disasters  \
count                                  5.569000e+03
mean                                   2.948914e+06
std                                    2.305264e+07
min                                    0.000000e+00
25%                                    8.500000e+01
50%                                    8.250000e+03
75%                                    1.333620e+05
max                                    6.574532e+08

       Reconstruction costs from disasters  Insured damages against disasters  \
count                         5.569000e+03                       5.569000e+03
mean                          2.663307e+04                       3.119986e+05
std                           7.623283e+05                       3.612870e+06
min                           0.000000e+00                       0.000000e+00
25%                           0.000000e+00                       0.000000e+00
50%                           0.000000e+00                       0.000000e+00
75%                           0.000000e+00                       0.000000e+00
max                           3.751550e+07                       1.270846e+08

       Total economic damages from disasters  Death rates from disasters  \
count                           5.569000e+03                 5491.000000
mean                            1.327784e+06                   20.630260
std                             1.173243e+07                  421.357312
min                             0.000000e+00                    0.000000
25%                             0.000000e+00                    0.005711
50%                             0.000000e+00                    0.098312
75%                             6.000000e+04                    0.510515
max                             3.640952e+08                17699.115044

       Injury rates from disasters  \
count                  5491.000000
mean                      6.268850
std                     124.020006
min                       0.000000
25%                       0.000000
50%                       0.000000
75%                       0.129953
max                    6516.094700

       Number of people affected by disasters per 100,000  \
count                                        5491.000000
mean                                         2777.972841
std                                          9884.964674
min                                             0.000000
25%                                             0.000000
50%                                            45.266823
```

```
75%                                            823.929487
max                                         120967.741935

       Homelessness rate from disasters  \
count                       5491.000000
mean                         144.208504
std                         1930.735467
min                            0.000000
25%                            0.000000
50%                            0.000000
75%                            1.877319
max                       109090.909091

       Total number of people affected by disasters per 100,000  \
count                                        5491.000000
mean                                         2928.450194
std                                         10264.056117
min                                             0.000000
25%                                             0.500841
50%                                            72.728819
75%                                           953.976696
max                                        155865.957447

       Number of deaths from volcanic activity  \
count                               293.000000
mean                                592.505119
std                                3420.146402
min                                   0.000000
25%                                   0.000000
50%                                   0.000000
75%                                  43.000000
max                               38690.000000

       Number of people injured from volcanic activity  \
count                                       293.000000
mean                                        176.771331
std                                        1243.347136
min                                           0.000000
25%                                           0.000000
50%                                           0.000000
75%                                           0.000000
max                                       14114.000000

       Number of people affected by volcanic activity  \
count                                    2.930000e+02
mean                                     6.141873e+04
std                                      2.054935e+05
min                                      0.000000e+00
```

```
25%                                        0.000000e+00
50%                                        5.000000e+03
75%                                        3.084500e+04
max                                        1.901284e+06

       Number of people left homeless from volcanic activity  \
count                                         293.000000
mean                                         2565.119454
std                                         12073.160049
min                                             0.000000
25%                                             0.000000
50%                                             0.000000
75%                                             0.000000
max                                        110000.000000

       Number of total people affected by volcanic activity  \
count                                       2.930000e+02
mean                                        6.416062e+04
std                                         2.067275e+05
min                                         0.000000e+00
25%                                         2.000000e+02
50%                                         5.600000e+03
75%                                         3.995800e+04
max                                         1.915398e+06

       Reconstruction costs from volcanic activity  \
count                                        293.0
mean                                           0.0
std                                            0.0
min                                            0.0
25%                                            0.0
50%                                            0.0
75%                                            0.0
max                                            0.0

       Insured damages against volcanic activity  \
count                                      293.0
mean                                         0.0
std                                          0.0
min                                          0.0
25%                                          0.0
50%                                          0.0
75%                                          0.0
max                                          0.0

       Total economic damages from volcanic activity  \
count                                     293.000000
mean                                    33207.590444
```

```
std                                       136252.752850
min                                            0.000000
25%                                            0.000000
50%                                            0.000000
75%                                            0.000000
max                                      1000000.000000

       Death rates from volcanic activity  \
count                            290.000000
mean                              16.662702
std                              201.111584
min                                0.000000
25%                                0.000000
50%                                0.000000
75%                                0.085050
max                             3365.519021

       Injury rates from volcanic activity  \
count                             290.000000
mean                                0.237749
std                                 1.368176
min                                 0.000000
25%                                 0.000000
50%                                 0.000000
75%                                 0.000000
max                                16.693933

       Number of people affected by volcanic activity per 100,000  \
count                                         290.000000
mean                                         1154.709266
std                                          5912.139165
min                                             0.000000
25%                                             0.000000
50%                                            16.598252
75%                                           102.389926
max                                         50000.000000

       Homelessness rate from volcanic activity  \
count                                  290.000000
mean                                    18.290761
std                                    142.349076
min                                      0.000000
25%                                      0.000000
50%                                      0.000000
75%                                      0.000000
max                                   1811.594203

       Total number of people affected by volcanic activity per 100,000  \
```

```
count                                         290.000000
mean                                         1173.237776
std                                          5919.423063
min                                             0.000000
25%                                             0.770507
50%                                            19.473166
75%                                           112.055458
max                                         50000.000000

       Number of deaths from floods  Number of people injured from floods  \
count                  3.056000e+03                             3056.000000
mean                   4.575145e+03                              897.586387
std                    1.084255e+05                            10942.935683
min                    0.000000e+00                                0.000000
25%                    1.000000e+00                                0.000000
50%                    1.300000e+01                                0.000000
75%                    6.200000e+01                                0.000000
max                    3.700000e+06                           252827.000000

       Number of people affected by floods  \
count                         3.056000e+03
mean                          2.474700e+06
std                           1.588563e+07
min                           0.000000e+00
25%                           3.650000e+02
50%                           1.004450e+04
75%                           1.050000e+05
max                           2.755356e+08

       Number of people left homeless from floods  \
count                                3.056000e+03
mean                                 6.093493e+04
std                                  5.934142e+05
min                                  0.000000e+00
25%                                  0.000000e+00
50%                                  0.000000e+00
75%                                  2.525000e+02
max                                  1.800310e+07

       Number of total people affected by floods  \
count                               3.056000e+03
mean                                2.536533e+06
std                                 1.627298e+07
min                                 0.000000e+00
25%                                 1.000000e+03
50%                                 1.403500e+04
75%                                 1.238645e+05
max                                 2.936627e+08
```

```
       Reconstruction costs from floods  Insured damages against floods  \
count                      3.056000e+03                    3.056000e+03
mean                       2.169498e+03                    5.432668e+04
std                        4.520501e+04                    4.607332e+05
min                        0.000000e+00                    0.000000e+00
25%                        0.000000e+00                    0.000000e+00
50%                        0.000000e+00                    0.000000e+00
75%                        0.000000e+00                    0.000000e+00
max                        1.440751e+06                    1.128800e+07

       Total economic damages from floods  Death rates from floods  \
count                        3.056000e+03              3032.000000
mean                         5.766453e+05                 1.664863
std                          3.550686e+06                31.996619
min                          0.000000e+00                 0.000000
25%                          0.000000e+00                 0.003909
50%                          0.000000e+00                 0.052546
75%                          3.000000e+04                 0.192962
max                          7.075905e+07              1348.379905

       Injury rates from floods  \
count               3032.000000
mean                   0.489515
std                    8.296810
min                    0.000000
25%                    0.000000
50%                    0.000000
75%                    0.000000
max                  379.987334

       Number of people affected by floods per 100,000  \
count                                      3032.000000
mean                                       1025.080836
std                                        3437.191660
min                                           0.000000
25%                                           1.576583
50%                                          65.446396
75%                                         487.460884
max                                       52369.730296

       Homelessness rate from floods  \
count                    3032.000000
mean                       67.255193
std                       938.132184
min                         0.000000
25%                         0.000000
50%                         0.000000
```

```
75%                         0.785303
max                     49295.774648

       Total number of people affected by floods per 100,000  \
count                                        3032.000000
mean                                         1092.825544
std                                          3631.016410
min                                             0.000000
25%                                             5.557691
50%                                            83.860588
75%                                           538.652468
max                                         55279.159757

       Number of deaths from mass movements  \
count                             77.000000
mean                             120.623377
std                              320.823900
min                                0.000000
25%                               16.000000
50%                               45.000000
75%                               76.000000
max                             2000.000000

       Number of people injured from mass movements  \
count                                     77.000000
mean                                      10.779221
std                                       25.087019
min                                        0.000000
25%                                        0.000000
50%                                        0.000000
75%                                        5.000000
max                                      115.000000

       Number of people affected by mass movements  \
count                                    77.000000
mean                                    549.038961
std                                    1542.815134
min                                      0.000000
25%                                      0.000000
50%                                      0.000000
75%                                    200.000000
max                                   8000.000000

       Number of people left homeless from mass movements  \
count                                           77.000000
mean                                           155.350649
std                                            428.730647
min                                              0.000000
```

```
25%                                                  0.000000
50%                                                  0.000000
75%                                                  0.000000
max                                               2000.000000

       Number of total people affected by mass movements  \
count                                            77.000000
mean                                            715.168831
std                                            1598.227731
min                                               0.000000
25%                                               0.000000
50%                                               0.000000
75%                                             697.000000
max                                            8000.000000

       Reconstruction costs from mass movements  \
count                                      77.0
mean                                        0.0
std                                         0.0
min                                         0.0
25%                                         0.0
50%                                         0.0
75%                                         0.0
max                                         0.0

       Insured damages against mass movements  \
count                                    77.0
mean                                      0.0
std                                       0.0
min                                       0.0
25%                                       0.0
50%                                       0.0
75%                                       0.0
max                                       0.0

       Total economic damages from mass movements  \
count                                   77.000000
mean                                  5428.571429
std                                  32007.576641
min                                      0.000000
25%                                      0.000000
50%                                      0.000000
75%                                      0.000000
max                                 200000.000000

       Death rates from mass movements  Injury rates from mass movements  \
count                        74.000000                         74.000000
mean                          0.880867                          0.026166
```

```
std                           3.060634                           0.080224
min                           0.000000                           0.000000
25%                           0.015363                           0.000000
50%                           0.113817                           0.000000
75%                           0.443672                           0.000374
max                          18.606382                           0.392519

       Number of people affected by mass movements per 100,000  \
count                                          74.000000
mean                                            1.224912
std                                             3.839277
min                                             0.000000
25%                                             0.000000
50%                                             0.000000
75%                                             0.000000
max                                            22.732439

       Homelessness rate from mass movements  \
count                              74.000000
mean                                0.768331
std                                 2.557920
min                                 0.000000
25%                                 0.000000
50%                                 0.000000
75%                                 0.000000
max                                12.759985

       Total number of people affected by mass movements per 100,000  \
count                                          74.000000
mean                                            2.019409
std                                             4.979931
min                                             0.000000
25%                                             0.000000
50%                                             0.000000
75%                                             0.660369
max                                            22.732439

       Number of deaths from storms  Number of people injured from storms  \
count                   2332.000000                           2332.000000
mean                    1200.503431                           1193.894511
std                    11718.621158                          18488.618227
min                        0.000000                              0.000000
25%                        1.000000                              0.000000
50%                       11.000000                              0.000000
75%                      100.000000                             33.000000
max                   304495.000000                         601939.000000

       Number of people affected by storms  \
```

```
count                         2.332000e+03
mean                          1.008452e+06
std                           5.911747e+06
min                           0.000000e+00
25%                           0.000000e+00
50%                           2.000000e+02
75%                           3.128750e+04
max                           1.111119e+08

       Number of people left homeless from storms  \
count                                2.332000e+03
mean                                 4.629072e+04
std                                  4.196207e+05
min                                  0.000000e+00
25%                                  0.000000e+00
50%                                  0.000000e+00
75%                                  0.000000e+00
max                                  1.117116e+07

       Number of total people affected by storms  \
count                               2.332000e+03
mean                                1.055937e+06
std                                 6.004925e+06
min                                 0.000000e+00
25%                                 0.000000e+00
50%                                 1.000000e+03
75%                                 4.099550e+04
max                                 1.111626e+08

       Reconstruction costs from storms  Insured damages against storms  \
count                      2.332000e+03                    2.332000e+03
mean                       3.515866e+03                    5.174520e+05
std                        7.857675e+04                    4.398694e+06
min                        0.000000e+00                    0.000000e+00
25%                        0.000000e+00                    0.000000e+00
50%                        0.000000e+00                    0.000000e+00
75%                        0.000000e+00                    0.000000e+00
max                        2.512500e+06                    1.087646e+08

       Total economic damages from storms  Death rates from storms  \
count                        2.332000e+03              2302.000000
mean                         1.369009e+06                 8.026993
std                          1.024531e+07               134.027837
min                          0.000000e+00                 0.000000
25%                          0.000000e+00                 0.003264
50%                          5.000000e+02                 0.045400
75%                          1.000000e+05                 0.308354
max                          2.729390e+08              4915.816640
```

```
       Injury rates from storms  \
count              2302.000000
mean                  5.697505
std                 108.746322
min                   0.000000
25%                   0.000000
50%                   0.000000
75%                   0.084705
max                4166.666667

       Number of people affected by storms per 100,000  \
count                                      2302.000000
mean                                       1812.438867
std                                        8663.247482
min                                           0.000000
25%                                           0.000000
50%                                           0.876989
75%                                         281.042629
max                                      106382.978723

       Homelessness rate from storms  \
count                    2302.000000
mean                      211.151924
std                      2731.769660
min                         0.000000
25%                         0.000000
50%                         0.000000
75%                         0.009410
max                    109090.909091

       Total number of people affected by storms per 100,000  \
count                                        2302.000000
mean                                         2029.288296
std                                          9496.501755
min                                             0.000000
25%                                             0.000000
50%                                             6.658198
75%                                           379.148989
max                                        155865.957447

       Number of deaths from landslides  \
count                        655.000000
mean                         204.320611
std                          786.877243
min                            0.000000
25%                           17.000000
50%                           41.000000
```

```
75%                            125.500000
max                          12000.000000

       Number of people injured from landslides  \
count                                655.000000
mean                                  36.983206
std                                  181.817358
min                                    0.000000
25%                                    0.000000
50%                                    0.000000
75%                                   14.500000
max                                 3000.000000

       Number of people affected by landslides  \
count                             6.550000e+02
mean                              3.169583e+04
std                               2.592852e+05
min                               0.000000e+00
25%                               0.000000e+00
50%                               0.000000e+00
75%                               5.075000e+02
max                               4.000000e+06

       Number of people left homeless from landslides  \
count                                    6.550000e+02
mean                                     1.300046e+04
std                                      1.512823e+05
min                                      0.000000e+00
25%                                      0.000000e+00
50%                                      0.000000e+00
75%                                      0.000000e+00
max                                      2.500366e+06

       Number of total people affected by landslides  \
count                                   6.550000e+02
mean                                    4.473327e+04
std                                     3.001016e+05
min                                     0.000000e+00
25%                                     0.000000e+00
50%                                     3.700000e+01
75%                                     1.976500e+03
max                                     4.000000e+06

       Reconstruction costs from landslides  \
count                            655.000000
mean                               9.160305
std                              165.646902
min                                0.000000
```

```
25%                                   0.000000
50%                                   0.000000
75%                                   0.000000
max                                3000.000000

       Insured damages against landslides  \
count                            655.000000
mean                            1355.419847
std                            14495.822358
min                                0.000000
25%                                0.000000
50%                                0.000000
75%                                0.000000
max                           200000.000000

       Total economic damages from landslides  Death rates from landslides  \
count                            6.550000e+02                   646.000000
mean                             3.361507e+04                     0.832527
std                              1.442824e+05                     5.732425
min                              0.000000e+00                     0.000000
25%                              0.000000e+00                     0.018526
50%                              0.000000e+00                     0.071530
75%                              0.000000e+00                     0.286839
max                              1.277078e+06                    94.307848

       Injury rates from landslides  \
count                    646.000000
mean                       0.084109
std                        0.529415
min                        0.000000
25%                        0.000000
50%                        0.000000
75%                        0.011163
max                        7.808944

       Number of people affected by landslides per 100,000  \
count                                         646.000000
mean                                           36.587989
std                                           283.779783
min                                             0.000000
25%                                             0.000000
50%                                             0.000000
75%                                             0.831151
max                                          4667.607968

       Homelessness rate from landslides  \
count                         646.000000
mean                            9.387541
```

```
std                              99.013506
min                               0.000000
25%                               0.000000
50%                               0.000000
75%                               0.000000
max                            2158.273381

       Total number of people affected by landslides per 100,000  \
count                                         646.000000
mean                                           46.059639
std                                           302.820861
min                                             0.000000
25%                                             0.000000
50%                                             0.075773
75%                                             2.999416
max                                          4667.607968

       Number of deaths from fog  Number of people injured from fog  \
count                        2.0                                2.0
mean                      4000.0                                0.0
std                          0.0                                0.0
min                       4000.0                                0.0
25%                       4000.0                                0.0
50%                       4000.0                                0.0
75%                       4000.0                                0.0
max                       4000.0                                0.0

       Number of people affected by fog  \
count                               2.0
mean                                0.0
std                                 0.0
min                                 0.0
25%                                 0.0
50%                                 0.0
75%                                 0.0
max                                 0.0

       Number of people left homeless from fog  \
count                                      2.0
mean                                       0.0
std                                        0.0
min                                        0.0
25%                                        0.0
50%                                        0.0
75%                                        0.0
max                                        0.0

       Number of total people affected by fog  Reconstruction costs from fog  \
```

count	2.0	2.0
mean	0.0	0.0
std	0.0	0.0
min	0.0	0.0
25%	0.0	0.0
50%	0.0	0.0
75%	0.0	0.0
max	0.0	0.0

	Insured damages against fog	Total economic damages from fog \
count	2.0	2.0
mean	0.0	0.0
std	0.0	0.0
min	0.0	0.0
25%	0.0	0.0
50%	0.0	0.0
75%	0.0	0.0
max	0.0	0.0

	Death rates from fog	Injury rates from fog \
count	2.000000	2.0
mean	7.897179	0.0
std	0.000000	0.0
min	7.897179	0.0
25%	7.897179	0.0
50%	7.897179	0.0
75%	7.897179	0.0
max	7.897179	0.0

	Number of people affected by fog per 100,000 \
count	2.0
mean	0.0
std	0.0
min	0.0
25%	0.0
50%	0.0
75%	0.0
max	0.0

	Homelessness rate from fog \
count	2.0
mean	0.0
std	0.0
min	0.0
25%	0.0
50%	0.0
75%	0.0
max	0.0

```
       Total number of people affected by fog per 100,000  \
count                                                2.0
mean                                                 0.0
std                                                  0.0
min                                                  0.0
25%                                                  0.0
50%                                                  0.0
75%                                                  0.0
max                                                  0.0

       Number of deaths from wildfires  \
count                       407.000000
mean                         21.852580
std                          78.697669
min                           0.000000
25%                           0.000000
50%                           1.000000
75%                          14.000000
max                        1000.000000

       Number of people injured from wildfires  \
count                               407.000000
mean                                 55.208845
std                                 209.573572
min                                   0.000000
25%                                   0.000000
50%                                   0.000000
75%                                   9.500000
max                                2292.000000

       Number of people affected by wildfires  \
count                            4.070000e+02
mean                             8.375322e+04
std                              7.395070e+05
min                              0.000000e+00
25%                              0.000000e+00
50%                              0.000000e+00
75%                              1.804500e+03
max                              1.004653e+07

       Number of people left homeless from wildfires  \
count                                      407.000000
mean                                      1179.592138
std                                       4434.044965
min                                          0.000000
25%                                          0.000000
50%                                          0.000000
```

```
75%                                              280.500000
max                                            50540.000000

       Number of total people affected by wildfires  \
count                                      4.070000e+02
mean                                       8.498802e+04
std                                        7.399048e+05
min                                        0.000000e+00
25%                                        0.000000e+00
50%                                        2.000000e+02
75%                                        4.000000e+03
max                                        1.005676e+07

       Reconstruction costs from wildfires  Insured damages against wildfires  \
count                           407.000000                       4.070000e+02
mean                           3159.705160                       2.581622e+05
std                           42539.307212                       1.540124e+06
min                               0.000000                       0.000000e+00
25%                               0.000000                       0.000000e+00
50%                               0.000000                       0.000000e+00
75%                               0.000000                       0.000000e+00
max                          643000.000000                       1.650000e+07

       Total economic damages from wildfires  Death rates from wildfires  \
count                           4.070000e+02                  403.000000
mean                            5.852544e+05                    0.055872
std                             2.286453e+06                    0.173391
min                             0.000000e+00                    0.000000
25%                             0.000000e+00                    0.000000
50%                             0.000000e+00                    0.001097
75%                             1.000000e+05                    0.020800
max                             2.280200e+07                    1.078981

       Injury rates from wildfires  \
count                   403.000000
mean                      0.120676
std                       0.532678
min                       0.000000
25%                       0.000000
50%                       0.000000
75%                       0.006502
max                       5.702338

       Number of people affected by wildfires per 100,000  \
count                                         403.000000
mean                                          165.828932
std                                          2427.198194
min                                             0.000000
```

```
25%                                                0.000000
50%                                                0.000000
75%                                                3.946414
max                                            48426.150121

       Homelessness rate from wildfires  \
count                        403.000000
mean                           6.532515
std                           52.878282
min                            0.000000
25%                            0.000000
50%                            0.000000
75%                            0.190189
max                          986.193294

       Total number of people affected by wildfires per 100,000  \
count                                         403.000000
mean                                          172.482123
std                                          2427.353627
min                                             0.000000
25%                                             0.000000
50%                                             0.503728
75%                                             9.263234
max                                         48426.150121

       Number of deaths from extreme temperatures  \
count                                  572.000000
mean                                   674.101399
std                                   4779.906351
min                                      0.000000
25%                                      4.000000
50%                                     32.000000
75%                                    182.250000
max                                  74698.000000

       Number of people injured from extreme temperatures  \
count                                       5.720000e+02
mean                                        7.172993e+03
std                                         1.064309e+05
min                                         0.000000e+00
25%                                         0.000000e+00
50%                                         0.000000e+00
75%                                         0.000000e+00
max                                         1.800413e+06

       Number of people affected by extreme temperatures  \
count                                       5.720000e+02
mean                                        3.542953e+05
```

```
std                                            4.624017e+06
min                                            0.000000e+00
25%                                            0.000000e+00
50%                                            0.000000e+00
75%                                            0.000000e+00
max                                            7.917050e+07

       Number of people left homeless from extreme temperatures  \
count                                            572.000000
mean                                             893.660839
std                                            13796.864858
min                                                0.000000
25%                                                0.000000
50%                                                0.000000
75%                                                0.000000
max                                           233000.000000

       Number of total people affected by extreme temperatures  \
count                                            5.720000e+02
mean                                             3.623619e+05
std                                              4.625901e+06
min                                              0.000000e+00
25%                                              0.000000e+00
50%                                              0.000000e+00
75%                                              3.595000e+02
max                                              7.917120e+07

       Reconstruction costs from extreme temperatures  \
count                                           572.0
mean                                              0.0
std                                               0.0
min                                               0.0
25%                                               0.0
50%                                               0.0
75%                                               0.0
max                                               0.0

       Insured damages against extreme temperatures  \
count                                  5.720000e+02
mean                                   2.094825e+04
std                                    1.451951e+05
min                                    0.000000e+00
25%                                    0.000000e+00
50%                                    0.000000e+00
75%                                    0.000000e+00
max                                    1.600000e+06

       Total economic damages from extreme temperatures  \
```

```
count                                          5.720000e+02
mean                                           2.209313e+05
std                                            1.457786e+06
min                                            0.000000e+00
25%                                            0.000000e+00
50%                                            0.000000e+00
75%                                            0.000000e+00
max                                            2.194000e+07

       Death rates from extreme temperatures  \
count                               566.000000
mean                                  0.830026
std                                   3.885326
min                                   0.000000
25%                                   0.010742
50%                                   0.048426
75%                                   0.168456
max                                  38.862828

       Injury rates from extreme temperatures  \
count                                566.000000
mean                                  12.407698
std                                  273.920010
min                                    0.000000
25%                                    0.000000
50%                                    0.000000
75%                                    0.000000
max                                 6516.072980

       Number of people affected by extreme temperatures per 100,000  \
count                                         566.000000
mean                                          410.186358
std                                          3120.655302
min                                             0.000000
25%                                             0.000000
50%                                             0.000000
75%                                             0.000000
max                                         48169.556840

       Homelessness rate from extreme temperatures  \
count                                     566.000000
mean                                        0.187956
std                                         2.302845
min                                         0.000000
25%                                         0.000000
50%                                         0.000000
75%                                         0.000000
max                                        45.117845
```

```
       Total number of people affected by extreme temperatures per 100,000  \
count                                         566.000000
mean                                          422.782012
std                                          3135.553750
min                                             0.000000
25%                                             0.000000
50%                                             0.000000
75%                                             1.135364
max                                         48169.556840

       Number of deaths from glacial lake outbursts  \
count                                           0.0
mean                                            NaN
std                                             NaN
min                                             NaN
25%                                             NaN
50%                                             NaN
75%                                             NaN
max                                             NaN

       Number of people injured from glacial lake outbursts  \
count                                                0.0
mean                                                 NaN
std                                                  NaN
min                                                  NaN
25%                                                  NaN
50%                                                  NaN
75%                                                  NaN
max                                                  NaN

       Number of people affected by glacial lake outbursts  \
count                                                0.0
mean                                                 NaN
std                                                  NaN
min                                                  NaN
25%                                                  NaN
50%                                                  NaN
75%                                                  NaN
max                                                  NaN

       Number of people left homeless from glacial lake outbursts  \
count                                                0.0
mean                                                 NaN
std                                                  NaN
min                                                  NaN
25%                                                  NaN
50%                                                  NaN
```

```
75%                                                     NaN
max                                                     NaN

       Number of total people affected by glacial lake outbursts  \
count                                                   0.0
mean                                                    NaN
std                                                     NaN
min                                                     NaN
25%                                                     NaN
50%                                                     NaN
75%                                                     NaN
max                                                     NaN

       Reconstruction costs from glacial lake outbursts  \
count                                                  0.0
mean                                                   NaN
std                                                    NaN
min                                                    NaN
25%                                                    NaN
50%                                                    NaN
75%                                                    NaN
max                                                    NaN

       Insured damages against glacial lake outbursts  \
count                                                0.0
mean                                                 NaN
std                                                  NaN
min                                                  NaN
25%                                                  NaN
50%                                                  NaN
75%                                                  NaN
max                                                  NaN

       Total economic damages from glacial lake outbursts  \
count                                                   0.0
mean                                                    NaN
std                                                     NaN
min                                                     NaN
25%                                                     NaN
50%                                                     NaN
75%                                                     NaN
max                                                     NaN

       Death rates from glacial lake outbursts  \
count                                       0.0
mean                                        NaN
std                                         NaN
min                                         NaN
```

```
25%                                        NaN
50%                                        NaN
75%                                        NaN
max                                        NaN

       Injury rates from glacial lake outbursts  \
count                                       0.0
mean                                        NaN
std                                         NaN
min                                         NaN
25%                                         NaN
50%                                         NaN
75%                                         NaN
max                                         NaN

       Number of people affected by glacial lake outbursts per 100,000  \
count                                                0.0
mean                                                 NaN
std                                                  NaN
min                                                  NaN
25%                                                  NaN
50%                                                  NaN
75%                                                  NaN
max                                                  NaN

       Homelessness rate from glacial lake outbursts  \
count                                            0.0
mean                                             NaN
std                                              NaN
min                                              NaN
25%                                              NaN
50%                                              NaN
75%                                              NaN
max                                              NaN

       Total number of people affected by glacial lake outbursts per 100,000  \
count                                                0.0
mean                                                 NaN
std                                                  NaN
min                                                  NaN
25%                                                  NaN
50%                                                  NaN
75%                                                  NaN
max                                                  NaN

       Total economic damages from disasters as a share of GDP  \
count                                        4485.000000
mean                                            1.142727
```

```
std                                                   9.431985
min                                                   0.000000
25%                                                   0.000000
50%                                                   0.000000
75%                                                   0.140063
max                                                 280.087777

       Total economic damages from drought as a share of GDP  \
count                                               722.000000
mean                                                  0.288242
std                                                   1.303854
min                                                   0.000000
25%                                                   0.000000
50%                                                   0.000000
75%                                                   0.037417
max                                                  17.598331

       Total economic damages from earthquakes as a share of GDP  \
count                                               769.000000
mean                                                  0.798568
std                                                   5.104692
min                                                   0.000000
25%                                                   0.000000
50%                                                   0.001745
75%                                                   0.070481
max                                                  95.930842

       Total economic damages from extreme temperatures as a share of GDP  \
count                                               545.000000
mean                                                  0.061597
std                                                   0.772519
min                                                   0.000000
25%                                                   0.000000
50%                                                   0.000000
75%                                                   0.000000
max                                                  16.274852

       Total economic damages from floods as a share of GDP  \
count                                              2755.000000
mean                                                  0.207195
std                                                   1.402456
min                                                   0.000000
25%                                                   0.000000
50%                                                   0.000000
75%                                                   0.039742
max                                                  56.383921

       Total economic damages from landslides as a share of GDP  \
```

```
count                                               579.000000
mean                                                  0.061634
std                                                   0.566196
min                                                   0.000000
25%                                                   0.000000
50%                                                   0.000000
75%                                                   0.000000
max                                                   9.061466

       Total economic damages from mass movements as a share of GDP  \
count                                                61.000000
mean                                                  0.199554
std                                                   1.092590
min                                                   0.000000
25%                                                   0.000000
50%                                                   0.000000
75%                                                   0.000000
max                                                   6.084995

       Total economic damages from storms as a share of GDP  \
count                                              1854.000000
mean                                                  1.885051
std                                                  13.769081
min                                                   0.000000
25%                                                   0.000000
50%                                                   0.002782
75%                                                   0.112616
max                                                 280.087777

       Total economic damages from volcanic activity as a share of GDP  \
count                                               228.000000
mean                                                  0.061460
std                                                   0.313755
min                                                   0.000000
25%                                                   0.000000
50%                                                   0.000000
75%                                                   0.000862
max                                                   2.865788

       Total economic damages from volcanic activity as a share of GDP.1  \
count                                               377.000000
mean                                                  0.486609
std                                                   6.719902
min                                                   0.000000
25%                                                   0.000000
50%                                                   0.000000
75%                                                   0.014476
max                                                 127.277641
```

```
              Year  total_damages_pct_gdp_glacial_lake
count  5569.000000                                 0.0
mean   1989.666906                                 NaN
std      25.467663                                 NaN
min    1900.000000                                 NaN
25%    1978.000000                                 NaN
50%    1997.000000                                 NaN
75%    2008.000000                                 NaN
max    2020.000000                                 NaN
```

[9]:

```python
# Check of missing data values.
for a in disaster_df.columns:
    miss=disaster_df[a].isnull().sum()
    if miss > 0:
        print('{} has {} missing values'.format(a,miss))
    else:
        print('{} has NO missing values'.format(a))
```

```
Number of deaths from drought has 4757 missing values
Number of people injured from drought has 4757 missing values
Number of people affected from drought has 4757 missing values
Number of people left homeless from drought has 4757 missing values
Number of total people affected by drought has 4757 missing values
Reconstruction costs from drought has 4757 missing values
Insured damages against drought has 4757 missing values
Total economic damages from drought has 4757 missing values
Death rates from drought has 4764 missing values
Injury rates from drought has 4764 missing values
Number of people affected by drought per 100,000 has 4764 missing values
Homelessness rate from drought has 4764 missing values
Total number of people affected by drought per 100,000 has 4764 missing values
Number of deaths from earthquakes has 4421 missing values
Number of people injured from earthquakes has 4421 missing values
Number of people affected by earthquakes has 4421 missing values
Number of people left homeless from earthquakes has 4421 missing values
Number of total people affected by earthquakes has 4421 missing values
Reconstruction costs from earthquakes has 4421 missing values
Insured damages against earthquakes has 4421 missing values
Total economic damages from earthquakes has 4421 missing values
Death rates from earthquakes has 4448 missing values
Injury rates from earthquakes has 4448 missing values
Number of people affected by earthquakes per 100,000 has 4448 missing values
Homelessness rate from earthquakes has 4448 missing values
Total number of people affected by earthquakes per 100,000 has 4448 missing
values
Number of deaths from disasters has NO missing values
Number of people injured from disasters has NO missing values
```

```
Number of people affected by disasters has NO missing values
Number of people left homeless from disasters has NO missing values
Number of total people affected by disasters has NO missing values
Reconstruction costs from disasters has NO missing values
Insured damages against disasters has NO missing values
Total economic damages from disasters has NO missing values
Death rates from disasters has 78 missing values
Injury rates from disasters has 78 missing values
Number of people affected by disasters per 100,000 has 78 missing values
Homelessness rate from disasters has 78 missing values
Total number of people affected by disasters per 100,000 has 78 missing values
Number of deaths from volcanic activity has 5276 missing values
Number of people injured from volcanic activity has 5276 missing values
Number of people affected by volcanic activity has 5276 missing values
Number of people left homeless from volcanic activity has 5276 missing values
Number of total people affected by volcanic activity has 5276 missing values
Reconstruction costs from volcanic activity has 5276 missing values
Insured damages against volcanic activity has 5276 missing values
Total economic damages from volcanic activity has 5276 missing values
Death rates from volcanic activity has 5279 missing values
Injury rates from volcanic activity has 5279 missing values
Number of people affected by volcanic activity per 100,000 has 5279 missing
values
Homelessness rate from volcanic activity has 5279 missing values
Total number of people affected by volcanic activity per 100,000 has 5279
missing values
Number of deaths from floods has 2513 missing values
Number of people injured from floods has 2513 missing values
Number of people affected by floods has 2513 missing values
Number of people left homeless from floods has 2513 missing values
Number of total people affected by floods has 2513 missing values
Reconstruction costs from floods has 2513 missing values
Insured damages against floods has 2513 missing values
Total economic damages from floods has 2513 missing values
Death rates from floods has 2537 missing values
Injury rates from floods has 2537 missing values
Number of people affected by floods per 100,000 has 2537 missing values
Homelessness rate from floods has 2537 missing values
Total number of people affected by floods per 100,000 has 2537 missing values
Number of deaths from mass movements has 5492 missing values
Number of people injured from mass movements has 5492 missing values
Number of people affected by mass movements has 5492 missing values
Number of people left homeless from mass movements has 5492 missing values
Number of total people affected by mass movements has 5492 missing values
Reconstruction costs from mass movements has 5492 missing values
Insured damages against mass movements has 5492 missing values
Total economic damages from mass movements has 5492 missing values
Death rates from mass movements has 5495 missing values
```

```
Injury rates from mass movements has 5495 missing values
Number of people affected by mass movements per 100,000 has 5495 missing values
Homelessness rate from mass movements has 5495 missing values
Total number of people affected by mass movements per 100,000 has 5495 missing
values
Number of deaths from storms has 3237 missing values
Number of people injured from storms has 3237 missing values
Number of people affected by storms has 3237 missing values
Number of people left homeless from storms has 3237 missing values
Number of total people affected by storms has 3237 missing values
Reconstruction costs from storms has 3237 missing values
Insured damages against storms has 3237 missing values
Total economic damages from storms has 3237 missing values
Death rates from storms has 3267 missing values
Injury rates from storms has 3267 missing values
Number of people affected by storms per 100,000 has 3267 missing values
Homelessness rate from storms has 3267 missing values
Total number of people affected by storms per 100,000 has 3267 missing values
Number of deaths from landslides has 4914 missing values
Number of people injured from landslides has 4914 missing values
Number of people affected by landslides has 4914 missing values
Number of people left homeless from landslides has 4914 missing values
Number of total people affected by landslides has 4914 missing values
Reconstruction costs from landslides has 4914 missing values
Insured damages against landslides has 4914 missing values
Total economic damages from landslides has 4914 missing values
Death rates from landslides has 4923 missing values
Injury rates from landslides has 4923 missing values
Number of people affected by landslides per 100,000 has 4923 missing values
Homelessness rate from landslides has 4923 missing values
Total number of people affected by landslides per 100,000 has 4923 missing
values
Number of deaths from fog has 5567 missing values
Number of people injured from fog has 5567 missing values
Number of people affected by fog has 5567 missing values
Number of people left homeless from fog has 5567 missing values
Number of total people affected by fog has 5567 missing values
Reconstruction costs from fog has 5567 missing values
Insured damages against fog has 5567 missing values
Total economic damages from fog has 5567 missing values
Death rates from fog has 5567 missing values
Injury rates from fog has 5567 missing values
Number of people affected by fog per 100,000 has 5567 missing values
Homelessness rate from fog has 5567 missing values
Total number of people affected by fog per 100,000 has 5567 missing values
Number of deaths from wildfires has 5162 missing values
Number of people injured from wildfires has 5162 missing values
Number of people affected by wildfires has 5162 missing values
```

```
Number of people left homeless from wildfires has 5162 missing values
Number of total people affected by wildfires has 5162 missing values
Reconstruction costs from wildfires has 5162 missing values
Insured damages against wildfires has 5162 missing values
Total economic damages from wildfires has 5162 missing values
Death rates from wildfires has 5166 missing values
Injury rates from wildfires has 5166 missing values
Number of people affected by wildfires per 100,000 has 5166 missing values
Homelessness rate from wildfires has 5166 missing values
Total number of people affected by wildfires per 100,000 has 5166 missing values
Number of deaths from extreme temperatures has 4997 missing values
Number of people injured from extreme temperatures has 4997 missing values
Number of people affected by extreme temperatures has 4997 missing values
Number of people left homeless from extreme temperatures has 4997 missing values
Number of total people affected by extreme temperatures has 4997 missing values
Reconstruction costs from extreme temperatures has 4997 missing values
Insured damages against extreme temperatures has 4997 missing values
Total economic damages from extreme temperatures has 4997 missing values
Death rates from extreme temperatures has 5003 missing values
Injury rates from extreme temperatures has 5003 missing values
Number of people affected by extreme temperatures per 100,000 has 5003 missing
values
Homelessness rate from extreme temperatures has 5003 missing values
Total number of people affected by extreme temperatures per 100,000 has 5003
missing values
Number of deaths from glacial lake outbursts has 5569 missing values
Number of people injured from glacial lake outbursts has 5569 missing values
Number of people affected by glacial lake outbursts has 5569 missing values
Number of people left homeless from glacial lake outbursts has 5569 missing
values
Number of total people affected by glacial lake outbursts has 5569 missing
values
Reconstruction costs from glacial lake outbursts has 5569 missing values
Insured damages against glacial lake outbursts has 5569 missing values
Total economic damages from glacial lake outbursts has 5569 missing values
Death rates from glacial lake outbursts has 5569 missing values
Injury rates from glacial lake outbursts has 5569 missing values
Number of people affected by glacial lake outbursts per 100,000 has 5569 missing
values
Homelessness rate from glacial lake outbursts has 5569 missing values
Total number of people affected by glacial lake outbursts per 100,000 has 5569
missing values
Total economic damages from disasters as a share of GDP has 1084 missing values
Total economic damages from drought as a share of GDP has 4847 missing values
Total economic damages from earthquakes as a share of GDP has 4800 missing
values
Total economic damages from extreme temperatures as a share of GDP has 5024
missing values
```

```
Total economic damages from floods as a share of GDP has 2814 missing values
Total economic damages from landslides as a share of GDP has 4990 missing values
Total economic damages from mass movements as a share of GDP has 5508 missing
values
Total economic damages from storms as a share of GDP has 3715 missing values
Total economic damages from volcanic activity as a share of GDP has 5341 missing
values
Total economic damages from volcanic activity as a share of GDP.1 has 5192
missing values
Entity has NO missing values
Year has NO missing values
total_damages_pct_gdp_glacial_lake has 5569 missing values
```

Note: Missing values are in the range of 4,421 and 5,279

[ ]:
```
# Check out a few rows
disaster_df.head()
```

[4]:
```
# Filter data down to US only
disaster_us_df = disaster_df[disaster_df['Entity'] == 'United States']
```

[5]:
```
# Check the dimension of the table/look at the data
print("The dimension of the table is: ", disaster_us_df.shape)
```

```
The dimension of the table is:  (99, 169)
```

[12]:
```
# Drop all NaN Values from dataset
disaster_us_df. dropna()
```

[12]:
```
Empty DataFrame
Columns: [Number of deaths from drought, Number of people injured from drought,
Number of people affected from drought, Number of people left homeless from
drought, Number of total people affected by drought, Reconstruction costs from
drought, Insured damages against drought, Total economic damages from drought,
Death rates from drought, Injury rates from drought, Number of people affected
by drought per 100,000, Homelessness rate from drought, Total number of people
affected by drought per 100,000, Number of deaths from earthquakes, Number of
people injured from earthquakes, Number of people affected by earthquakes,
Number of people left homeless from earthquakes, Number of total people affected
by earthquakes, Reconstruction costs from earthquakes, Insured damages against
earthquakes, Total economic damages from earthquakes, Death rates from
earthquakes, Injury rates from earthquakes, Number of people affected by
earthquakes per 100,000, Homelessness rate from earthquakes, Total number of
people affected by earthquakes per 100,000, Number of deaths from disasters,
Number of people injured from disasters, Number of people affected by disasters,
Number of people left homeless from disasters, Number of total people affected
by disasters, Reconstruction costs from disasters, Insured damages against
disasters, Total economic damages from disasters, Death rates from disasters,
Injury rates from disasters, Number of people affected by disasters per 100,000,
```

```
Homelessness rate from disasters, Total number of people affected by disasters
per 100,000, Number of deaths from volcanic activity, Number of people injured
from volcanic activity, Number of people affected by volcanic activity, Number
of people left homeless from volcanic activity, Number of total people affected
by volcanic activity, Reconstruction costs from volcanic activity, Insured
damages against volcanic activity, Total economic damages from volcanic
activity, Death rates from volcanic activity, Injury rates from volcanic
activity, Number of people affected by volcanic activity per 100,000,
Homelessness rate from volcanic activity, Total number of people affected by
volcanic activity per 100,000, Number of deaths from floods, Number of people
injured from floods, Number of people affected by floods, Number of people left
homeless from floods, Number of total people affected by floods, Reconstruction
costs from floods, Insured damages against floods, Total economic damages from
floods, Death rates from floods, Injury rates from floods, Number of people
affected by floods per 100,000, Homelessness rate from floods, Total number of
people affected by floods per 100,000, Number of deaths from mass movements,
Number of people injured from mass movements, Number of people affected by mass
movements, Number of people left homeless from mass movements, Number of total
people affected by mass movements, Reconstruction costs from mass movements,
Insured damages against mass movements, Total economic damages from mass
movements, Death rates from mass movements, Injury rates from mass movements,
Number of people affected by mass movements per 100,000, Homelessness rate from
mass movements, Total number of people affected by mass movements per 100,000,
Number of deaths from storms, Number of people injured from storms, Number of
people affected by storms, Number of people left homeless from storms, Number of
total people affected by storms, Reconstruction costs from storms, Insured
damages against storms, Total economic damages from storms, Death rates from
storms, Injury rates from storms, Number of people affected by storms per
100,000, Homelessness rate from storms, Total number of people affected by
storms per 100,000, Number of deaths from landslides, Number of people injured
from landslides, Number of people affected by landslides, Number of people left
homeless from landslides, Number of total people affected by landslides,
Reconstruction costs from landslides, Insured damages against landslides, Total
economic damages from landslides, Death rates from landslides, …]
Index: []
```

[14]:
```python
# Check the dimension of the table/look at the data
print("The dimension of the table is: ", disaster_us_df.shape)
```

```
The dimension of the table is:  (99, 169)
```

[14]:
```python
# Write file down to csv
disaster_us_df.to_csv('disaster_us.csv', index=False)
```

[4]:
```python
# Reload data
file2 = "disaster_us_storm.csv"
disaster_us_storm_df = pd.read_csv(file2)
```

```
# Additional file
file3= "north_atlantic_hurricanes_stats.csv"
hurricanes_us_df = pd.read_csv(file3)
```

Disasters US Storms

[20]:
```
# Check the dimension of the table/look at the data
print("The dimension of disaster_us_storm_df is: ", disaster_us_storm_df.shape)
```

```
The dimension of disaster_us_storm_df is:  (99, 16)
```

[21]:
```
# What type of variables are in the table before dropping variables.
print("Describe Data disaster_us_storm_df")
print(disaster_us_storm_df.describe())
```

```
Describe Data disaster_us_storm_df
       no_deaths_storms  no_people_injured_storms  no_people_affected_storms  \
count         95.000000                 95.000000               9.500000e+01
mean         326.189474                158.926316               1.063053e+06
std          656.820570                420.195325               8.731234e+06
min            1.000000                  0.000000               0.000000e+00
25%          103.500000                  0.000000               0.000000e+00
50%          190.000000                  0.000000               0.000000e+00
75%          303.000000                165.000000               9.184000e+03
max         6000.000000               3593.000000               8.501880e+07

       no_people_left_homeless_storms  no_total_people_affected_storms  \
count                       95.000000                     9.500000e+01
mean                      5622.442105                     1.068834e+06
std                      28403.032587                     8.730827e+06
min                          0.000000                     0.000000e+00
25%                          0.000000                     0.000000e+00
50%                          0.000000                     0.000000e+00
75%                         81.000000                     1.080450e+04
max                     250000.000000                     8.501947e+07

       reconstruction_costs_storms  insured_damages_storms  \
count                    95.000000            9.500000e+01
mean                    657.894737            4.525446e+06
std                    6412.364701            1.249115e+07
min                       0.000000            0.000000e+00
25%                       0.000000            0.000000e+00
50%                       0.000000            0.000000e+00
75%                       0.000000            3.690250e+06
max                   62500.000000            8.296500e+07

       total_economic_damages_storms  death_rates_storms  injury_rates_storms  \
count                   9.500000e+01           95.000000            95.000000
mean                    9.607543e+06            0.232532             0.056396
```

```
std                     2.677974e+07           0.801661            0.142446
min                     0.000000e+00           0.000412            0.000000
25%                     1.950000e+04           0.042962            0.000000
50%                     4.080000e+05           0.077257            0.000000
75%                     5.472500e+06           0.207161            0.065360
max                     1.711100e+08           7.713919            1.153140

       no_people_affected_storms_per_100000  homelessness_rate_storms  \
count                             95.000000                 95.000000
mean                             334.708298                  2.069297
std                             2704.323998                 10.746575
min                                0.000000                  0.000000
25%                                0.000000                  0.000000
50%                                0.000000                  0.000000
75%                                3.012068                  0.027080
max                            26320.306110                 97.279668

       total_no_people_affected_storms_per_100000  \
count                                   95.000000
mean                                   336.833992
std                                   2704.172894
min                                      0.000000
25%                                      0.000000
50%                                      0.000000
75%                                      3.885984
max                                  26320.512920

       total_economic_damages_storms_share_GDP         Year
count                                60.000000    99.000000
mean                                  0.129685  1968.959596
std                                   0.208821    32.009858
min                                   0.000000  1900.000000
25%                                   0.026670  1946.500000
50%                                   0.062040  1971.000000
75%                                   0.133962  1995.500000
max                                   1.213737  2020.000000
```

Hurricanes stats

[22]:
```
# Check the dimension of the table/look at the data
print("The dimension of north_atlantic_hurricanes_stats is: ", hurricanes_us_df.
 ↪shape)
```

```
The dimension of north_atlantic_hurricanes_stats is:  (169, 13)
```

[23]:
```
# What type of variables are in the table before dropping variables.
print("Describe Data north_atlantic_hurricanes_stats")
print(hurricanes_us_df.describe())
```

```
Describe Data north_atlantic_hurricanes_stats
```

```
              Year  no_us_hurricanes_HUDRAT_NOAA  \
count   169.000000                    169.000000
mean   1935.000000                      1.804734
std      48.930222                      1.524710
min    1851.000000                      0.000000
25%    1893.000000                      1.000000
50%    1935.000000                      2.000000
75%    1977.000000                      3.000000
max    2019.000000                      8.000000

       no_major_us_hurricanes_HUDRAT_ NOAA  \
count                           169.000000
mean                              0.568047
std                               0.777157
min                               0.000000
25%                               0.000000
50%                               0.000000
75%                               1.000000
max                               4.000000

       no_major_north_atlantic_hurricanes_HUDRAT_NOAA  \
count                                      169.000000
mean                                         1.875740
std                                          1.604292
min                                          0.000000
25%                                          1.000000
50%                                          2.000000
75%                                          3.000000
max                                          7.000000

       no_noth_atlantic_hurricanes_HUDRAT_NOAA  \
count                               169.000000
mean                                  5.473373
std                                   2.540306
min                                   0.000000
25%                                   4.000000
50%                                   5.000000
75%                                   7.000000
max                                  15.000000

       accumulated_cyclone_energy_ACE_HUDRAT_NOAA  \
count                                  169.000000
mean                                    88.331361
std                                     52.976042
min                                      3.000000
25%                                     50.000000
50%                                     74.000000
75%                                    113.000000
```

```
max                                      259.000000

       cyclone_power_dissipation_index_PDI_HUDRAT_NOAA  \
count                                       163.000000
mean                                          0.983188
std                                           1.386486
min                                           0.000000
25%                                           0.000000
50%                                           0.000000
75%                                           2.025500
max                                           6.003000

       hurricane_fatality_rate         ACE  deaths_hurricanes_us  \
count               165.000000  169.000000            120.000000
mean                  0.095924   88.764290            256.775000
std                   0.365453   53.285074            599.146262
min                   0.000000    2.530000              0.000000
25%                   0.000000   49.710000             30.000000
50%                   0.000000   76.062500            150.500000
75%                   0.041278  115.837500            272.750000
max                   3.438047  258.570000           6000.000000

       total_economic_damages_storms   death_bin
count                   1.200000e+02  120.000000
mean                    7.115763e+06    2.541667
std                     2.366521e+07    2.007950
min                     0.000000e+00    1.000000
25%                     0.000000e+00    1.000000
50%                     8.650000e+04    2.000000
75%                     2.549200e+06    3.000000
max                     1.711100e+08   10.000000
```

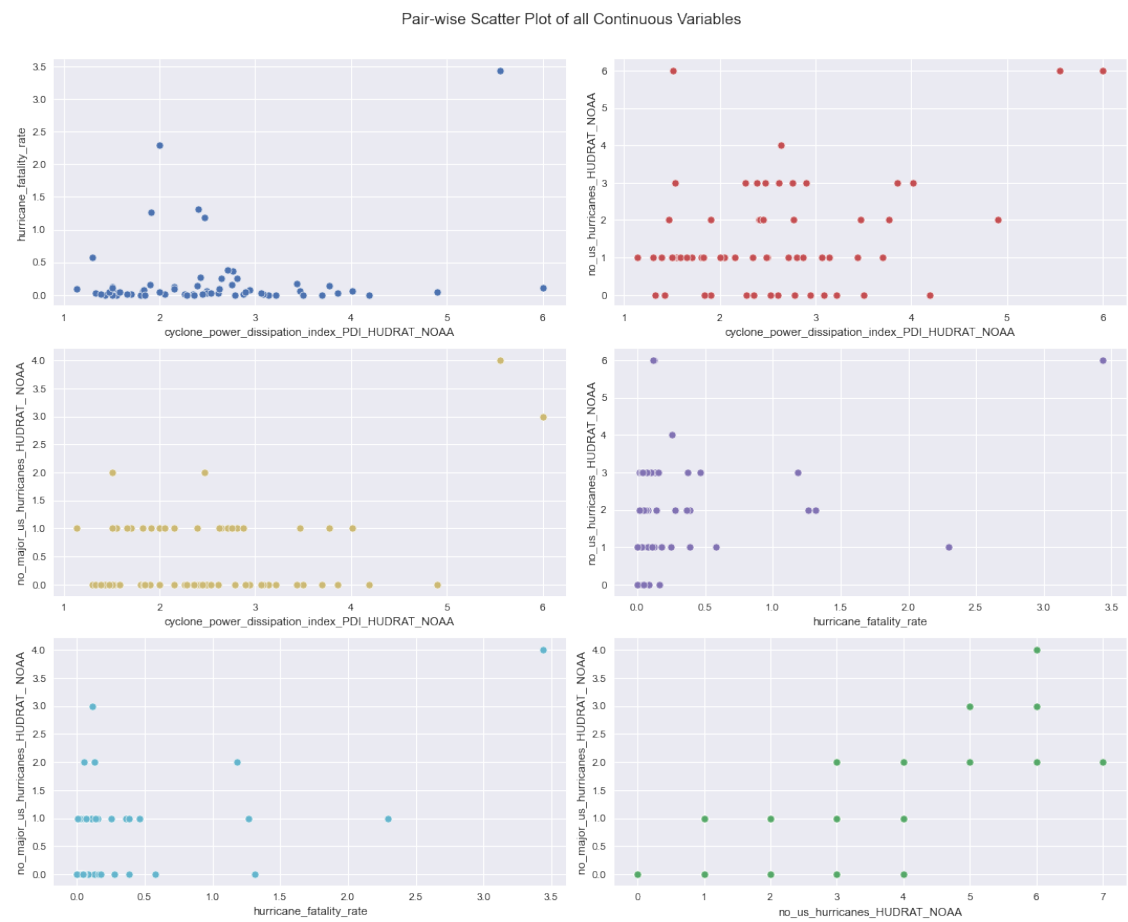
[11]:
```python
# Importing Autoviz class
from autoviz.AutoViz_Class import AutoViz_Class#Instantiate the AutoViz class
AV = AutoViz_Class()
```

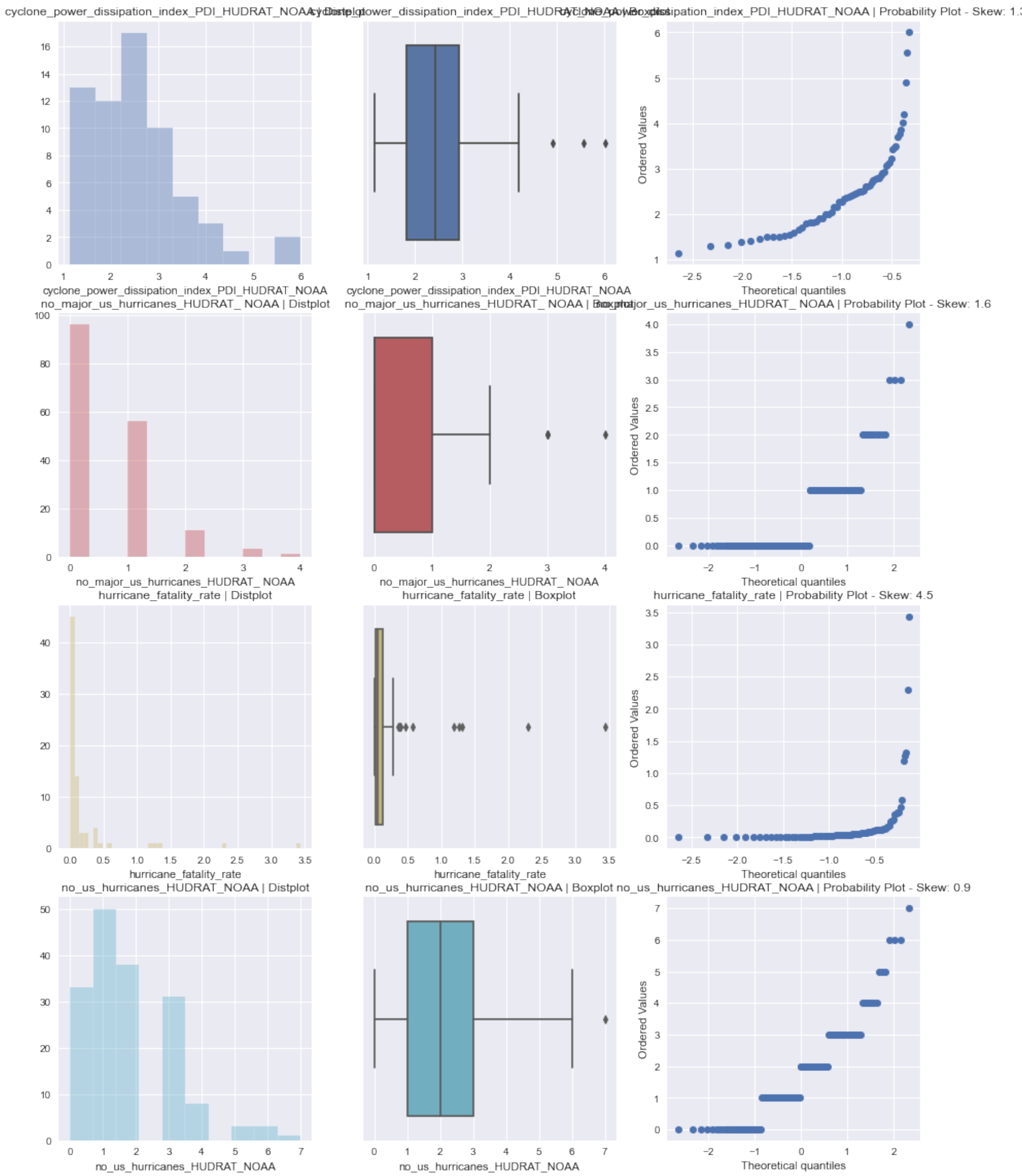
```
Imported AutoViz_Class version: 0.0.81. Call using:
    from autoviz.AutoViz_Class import AutoViz_Class
    AV = AutoViz_Class()
    AV.AutoViz(filename, sep=',', depVar='', dfte=None, header=0, verbose=0,
lowess=False,chart_format='svg',max_rows_analyzed=150000,max_cols_analyzed=30)
Note: verbose=0 or 1 generates charts and displays them in your local Jupyter
notebook.
      verbose=2 saves plots in your local machine under AutoViz_Plots directory
and does not display charts.
```

[22]:
```python
#north_atlantic_hurricanes_stats is: hurricanes_us_df
#Trying this to see what happens
```

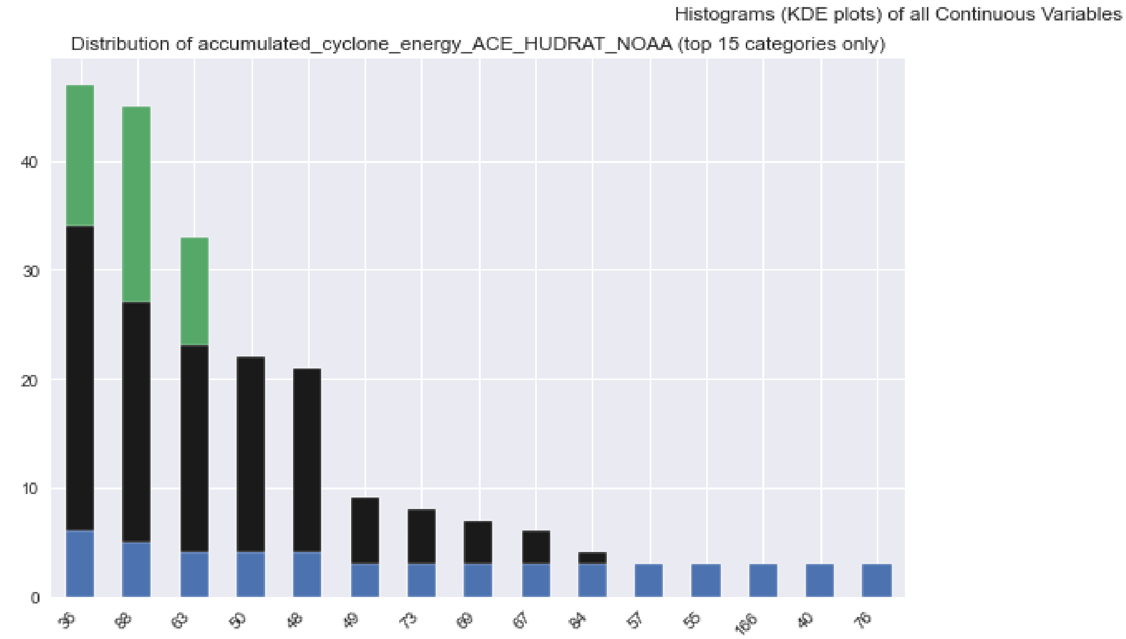
```
df = AV.AutoViz('north_atlantic_hurricanes_stats.csv')
```

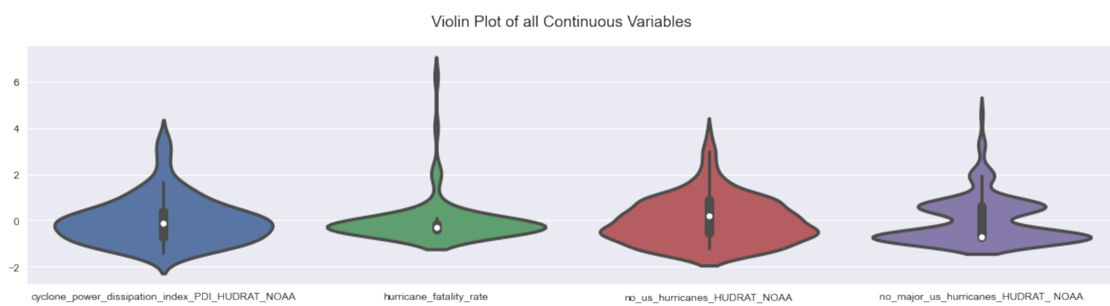
```
Shape of your Data Set: (168, 10)
############## C L A S S I F Y I N G  V A R I A B L E S  ####################
Classifying variables in data set…
    Number of Numeric Columns =  4
    Number of Integer-Categorical Columns =  3
    Number of String-Categorical Columns =  0
    Number of Factor-Categorical Columns =  0
    Number of String-Boolean Columns =  0
    Number of Numeric-Boolean Columns =  0
    Number of Discrete String Columns =  0
    Number of NLP String Columns =  0
    Number of Date Time Columns =  0
    Number of ID Columns =  1
    Number of Columns to Delete =  2
    10 Predictors classified…
        This does not include the Target column(s)
        3 variables removed since they were ID or low-information variables
Number of All Scatter Plots = 10
```

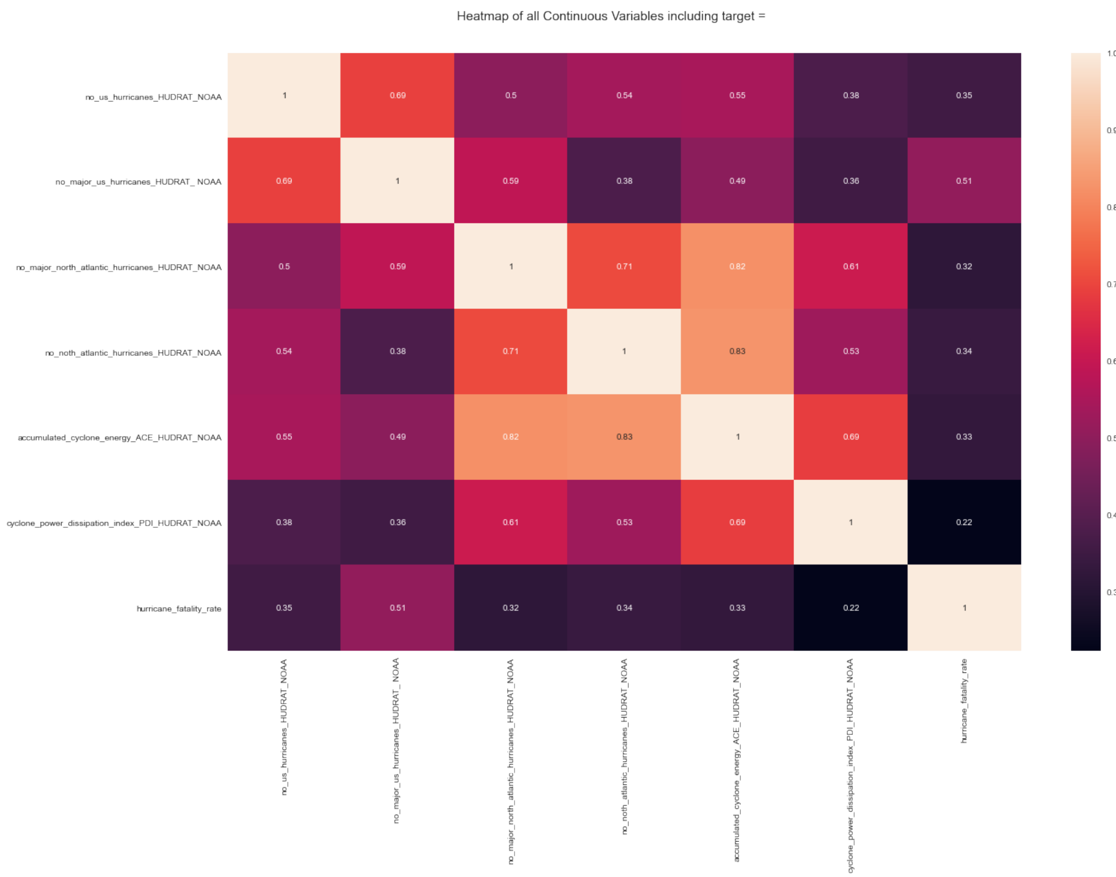




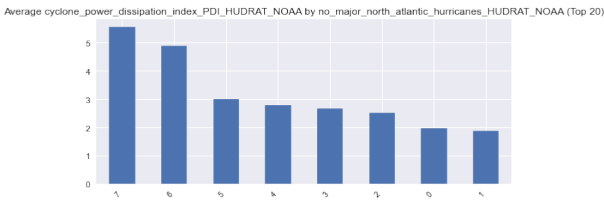
cyclone_power_dissipation_index_PDI_HUDRAT_NOAA | Distplot
cyclone_power_dissipation_index_PDI_HUDRAT_NOAA | Boxplot
cyclone_power_dissipation_index_PDI_HUDRAT_NOAA | Probability Plot - Skew: 1.3
cyclone_power_dissipation_index_PDI_HUDRAT_NOAA
Ordered Values
Theoretical quantiles
no_major_us_hurricanes_HUDRAT_ NOAA | Distplot
no_major_us_hurricanes_HUDRAT_ NOAA | Boxplot
no_major_us_hurricanes_HUDRAT_ NOAA | Probability Plot - Skew: 1.6
no_major_us_hurricanes_HUDRAT_ NOAA
hurricane_fatality_rate | Distplot
hurricane_fatality_rate | Boxplot
hurricane_fatality_rate | Probability Plot - Skew: 4.5
hurricane_fatality_rate
no_us_hurricanes_HUDRAT_NOAA | Distplot
no_us_hurricanes_HUDRAT_NOAA | Boxplot
no_us_hurricanes_HUDRAT_NOAA | Probability Plot - Skew: 0.9
no_us_hurricanes_HUDRAT_NOAA

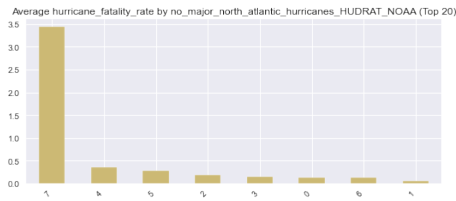

Histograms (KDE plots) of all Continuous Variables
Distribution of accumulated_cyclone_energy_ACE_HUDRAT_NOAA (top 15 categories only)
0
10
20
30
40

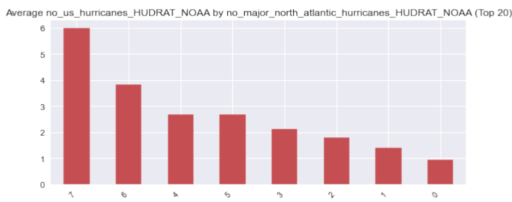

Violin Plot of all Continuous Variables
-2
0
2
4
6
cyclone_power_dissipation_index_PDI_HUDRAT_NOAA
hurricane_fatality_rate
no_us_hurricanes_HUDRAT_NOAA
no_major_us_hurricanes_HUDRAT_ NOAA

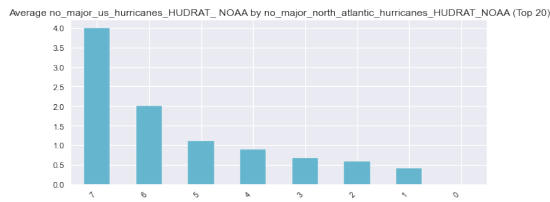

Heatmap of all Continuous Variables including target =
no_us_hurricanes_HUDRAT_NOAA
no_major_us_hurricanes_HUDRAT_ NOAA
no_major_north_atlantic_hurricanes_HUDRAT_NOAA
no_noth_atlantic_hurricanes_HUDRAT_NOAA
accumulated_cyclone_energy_ACE_HUDRAT_NOAA
cyclone_power_dissipation_index_PDI_HUDRAT_NOAA
hurricane_fatality_rate
1 0.69 0.5 0.54 0.55 0.38 0.35
0.69 1 0.59 0.38 0.49 0.36 0.51
0.5 0.59 1 0.71 0.82 0.61 0.32
0.54 0.38 0.71 1 0.83 0.53 0.34
0.55 0.49 0.82 0.83 1 0.69 0.33
0.38 0.36 0.61 0.53 0.69 1 0.22
0.35 0.51 0.32 0.34 0.33 0.22 1
1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3

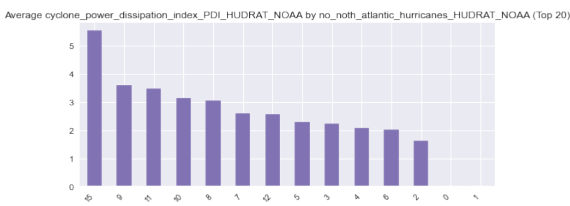
Bar plots for each Continuous by each Categorical variable

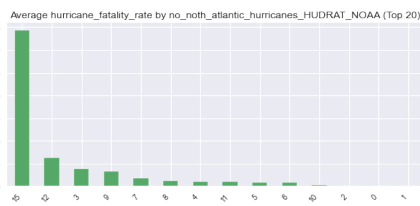


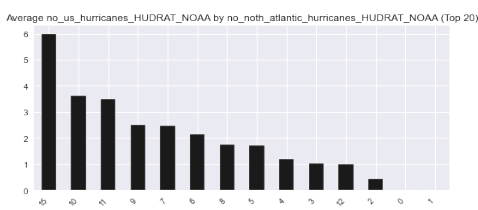


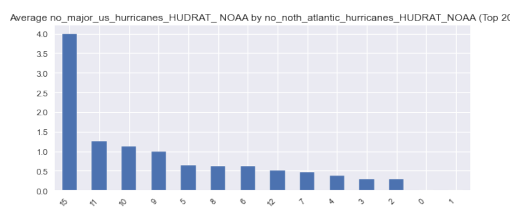


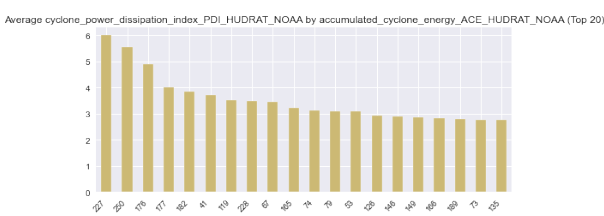


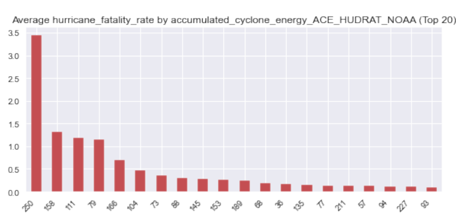


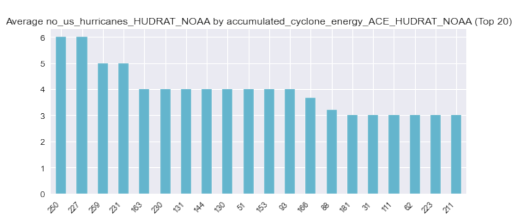


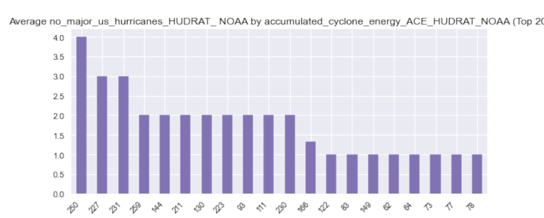










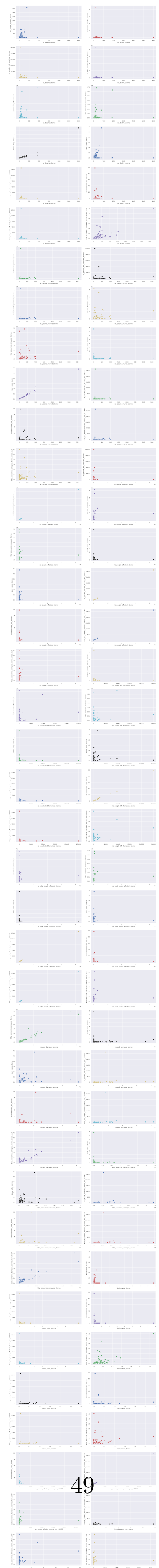


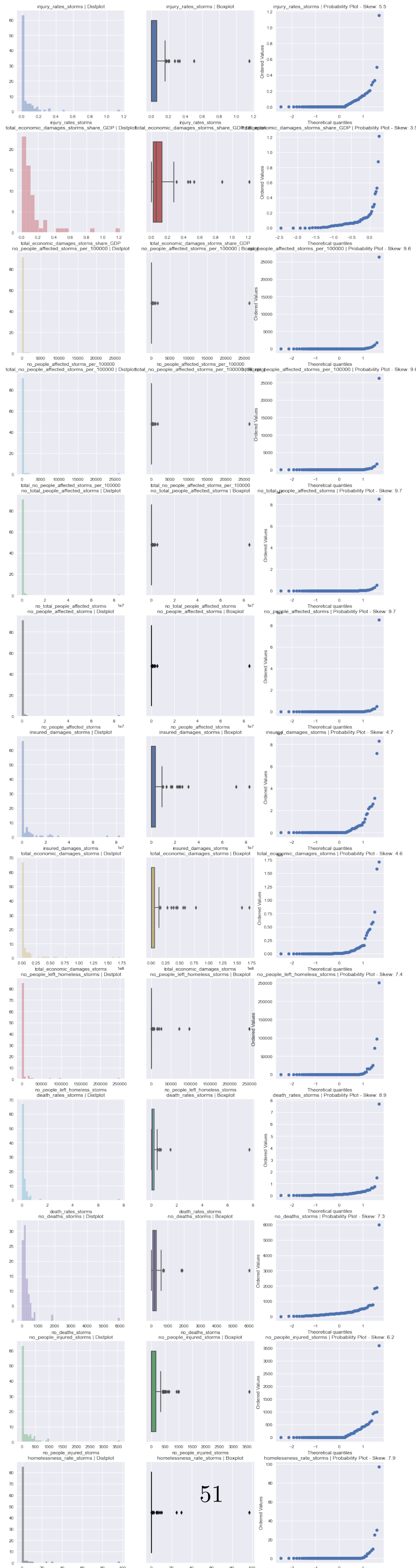
```
Time to run AutoViz (in seconds) = 13.330

 ###################### VISUALIZATION Completed ########################
```

[12]:
```
#disaster_us_storm_df is:, disaster_us_storm_df
#Trying this to see what happens
df2 = AV.AutoViz("disaster_us_storm.csv")
```

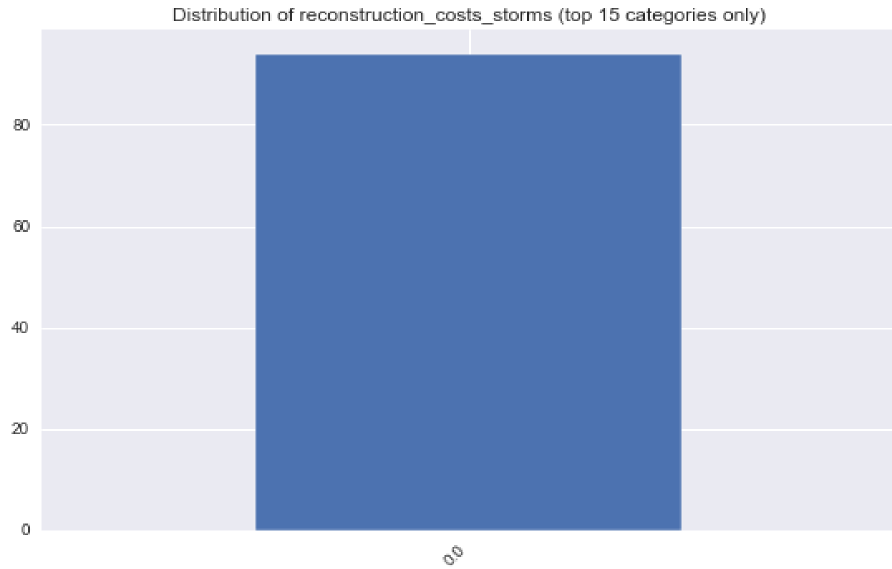
```
Shape of your Data Set: (99, 16)
############## C L A S S I F Y I N G  V A R I A B L E S  ####################
Classifying variables in data set…
    Number of Numeric Columns =  13
    Number of Integer-Categorical Columns =  0
    Number of String-Categorical Columns =  0
    Number of Factor-Categorical Columns =  0
    Number of String-Boolean Columns =  0
    Number of Numeric-Boolean Columns =  1
    Number of Discrete String Columns =  0
    Number of NLP String Columns =  0
    Number of Date Time Columns =  1
    Number of ID Columns =  0
    Number of Columns to Delete =  1
    16 Predictors classified…
        This does not include the Target column(s)
        1 variables removed since they were ID or low-information variables
Number of All Scatter Plots = 91
```

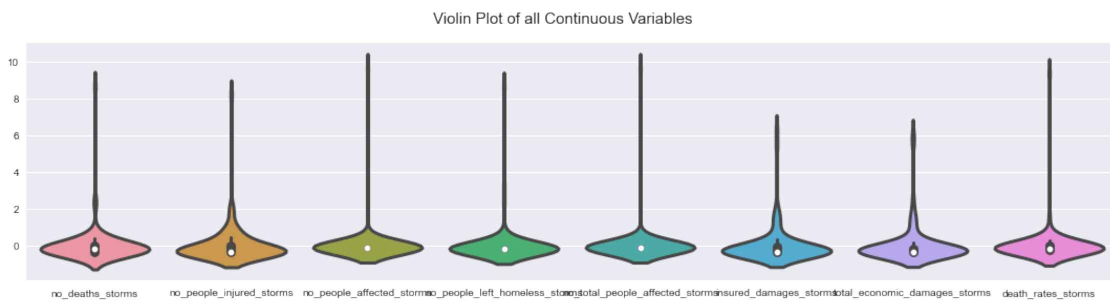
Pair-wise Scatter Plot of all Continuous Variables

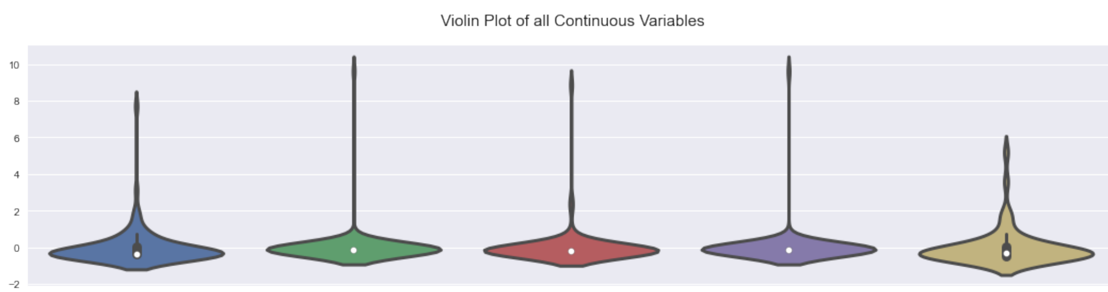


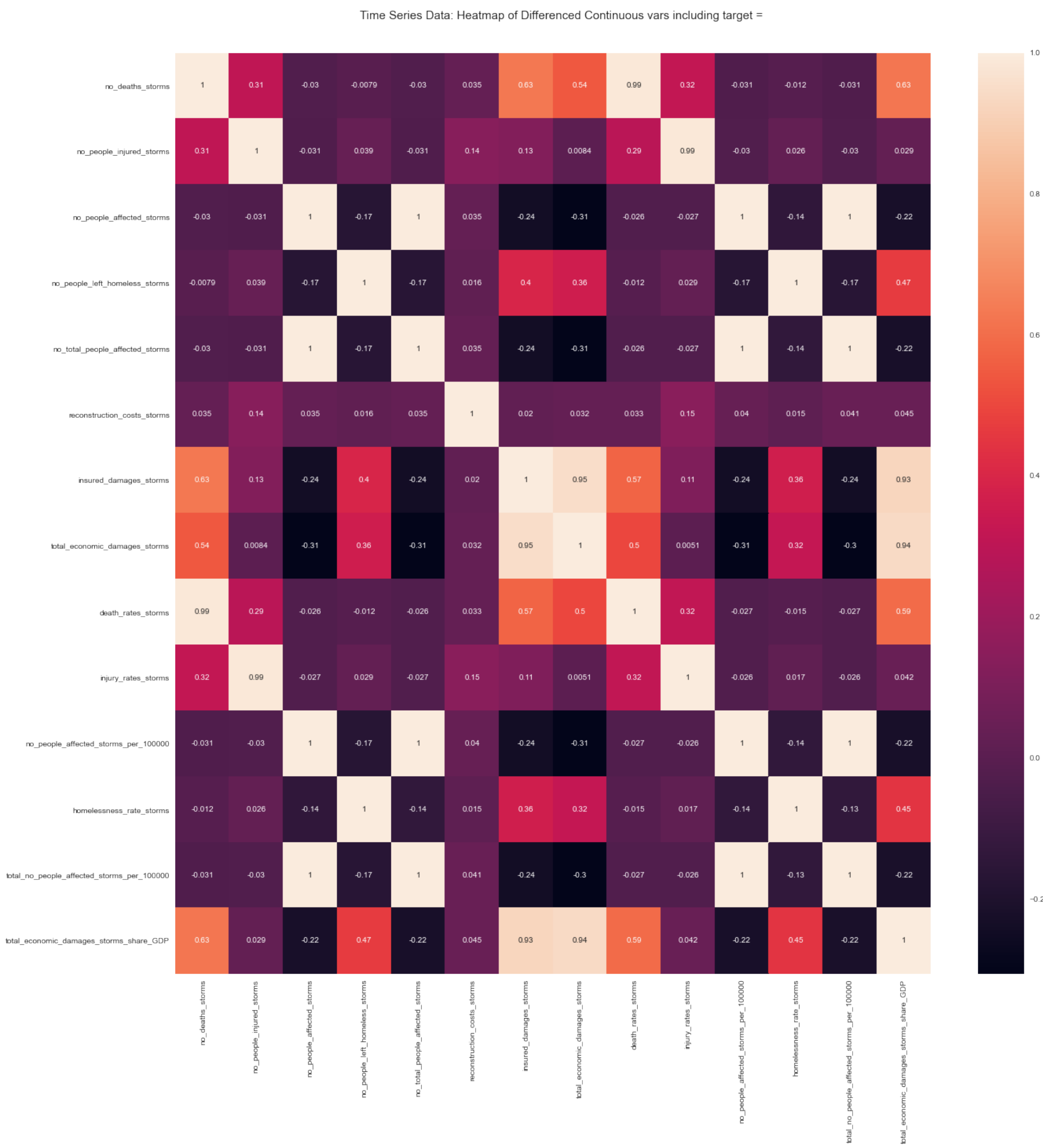

injury_rates_storms | Distplot
injury_rates_storms | Boxplot
injury_rates_storms | Probability Plot - Skew: 5.5
injury_rates_storms
total_economic_damages_storms_share_GDP | Distplot
total_economic_damages_storms_share_GDP | Boxplot
total_economic_damages_storms_share_GDP | Probability Plot - Skew: 3.5
total_economic_damages_storms_share_GDP
no_people_affected_storms_per_100000 | Distplot
no_people_affected_storms_per_100000 | Boxplot
no_people_affected_storms_per_100000 | Probability Plot - Skew: 9.6
no_people_affected_storms_per_100000
total_no_people_affected_storms_per_100000 | Distplot
total_no_people_affected_storms_per_100000 | Boxplot
total_no_people_affected_storms_per_100000 | Probability Plot - Skew: 9.6
total_no_people_affected_storms_per_100000
no_total_people_affected_storms | Distplot
no_total_people_affected_storms | Boxplot
no_total_people_affected_storms | Probability Plot - Skew: 9.7
no_total_people_affected_storms
no_people_affected_storms | Distplot
no_people_affected_storms | Boxplot
no_people_affected_storms | Probability Plot - Skew: 9.7
no_people_affected_storms
insured_damages_storms | Distplot
insured_damages_storms | Boxplot
insured_damages_storms | Probability Plot - Skew: 4.7
insured_damages_storms
total_economic_damages_storms | Distplot
total_economic_damages_storms | Boxplot
total_economic_damages_storms | Probability Plot - Skew: 4.6
total_economic_damages_storms
no_people_left_homeless_storms | Distplot
no_people_left_homeless_storms | Boxplot
no_people_left_homeless_storms | Probability Plot - Skew: 7.4
no_people_left_homeless_storms
death_rates_storms | Distplot
death_rates_storms | Boxplot
death_rates_storms | Probability Plot - Skew: 8.9
death_rates_storms
no_deaths_storms | Distplot
no_deaths_storms | Boxplot
no_deaths_storms | Probability Plot - Skew: 7.3
no_deaths_storms
no_people_injured_storms | Distplot
no_people_injured_storms | Boxplot
no_people_injured_storms | Probability Plot - Skew: 6.2
no_people_injured_storms
homelessness_rate_storms | Distplot
homelessness_rate_storms | Boxplot
homelessness_rate_storms | Probability Plot - Skew: 7.9
homelessness_rate_storms
Ordered Values
Theoretical quantiles

Histograms (KDE plots) of all Continuous Variables



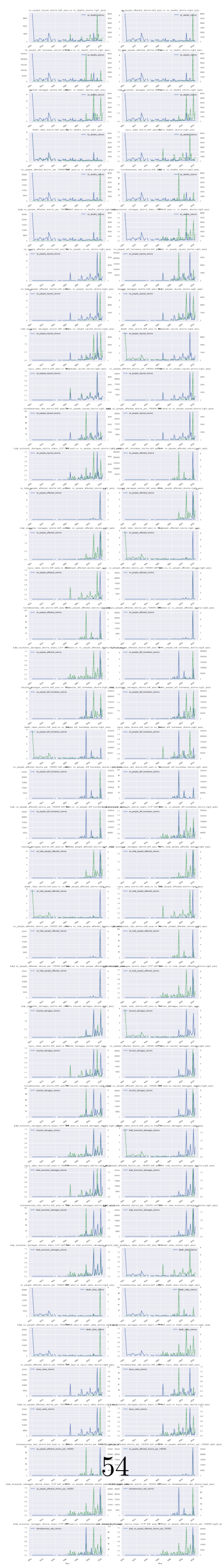







```
No categorical or numeric vars in data set. Hence no bar charts.
Time to run AutoViz (in seconds) = 49.251

 ###################### VISUALIZATION Completed ########################
```

Time Series Plot by Year: Pairwise Continuous Variables



[9]:
```
# Column names
hurricanes_us_df.columns
```

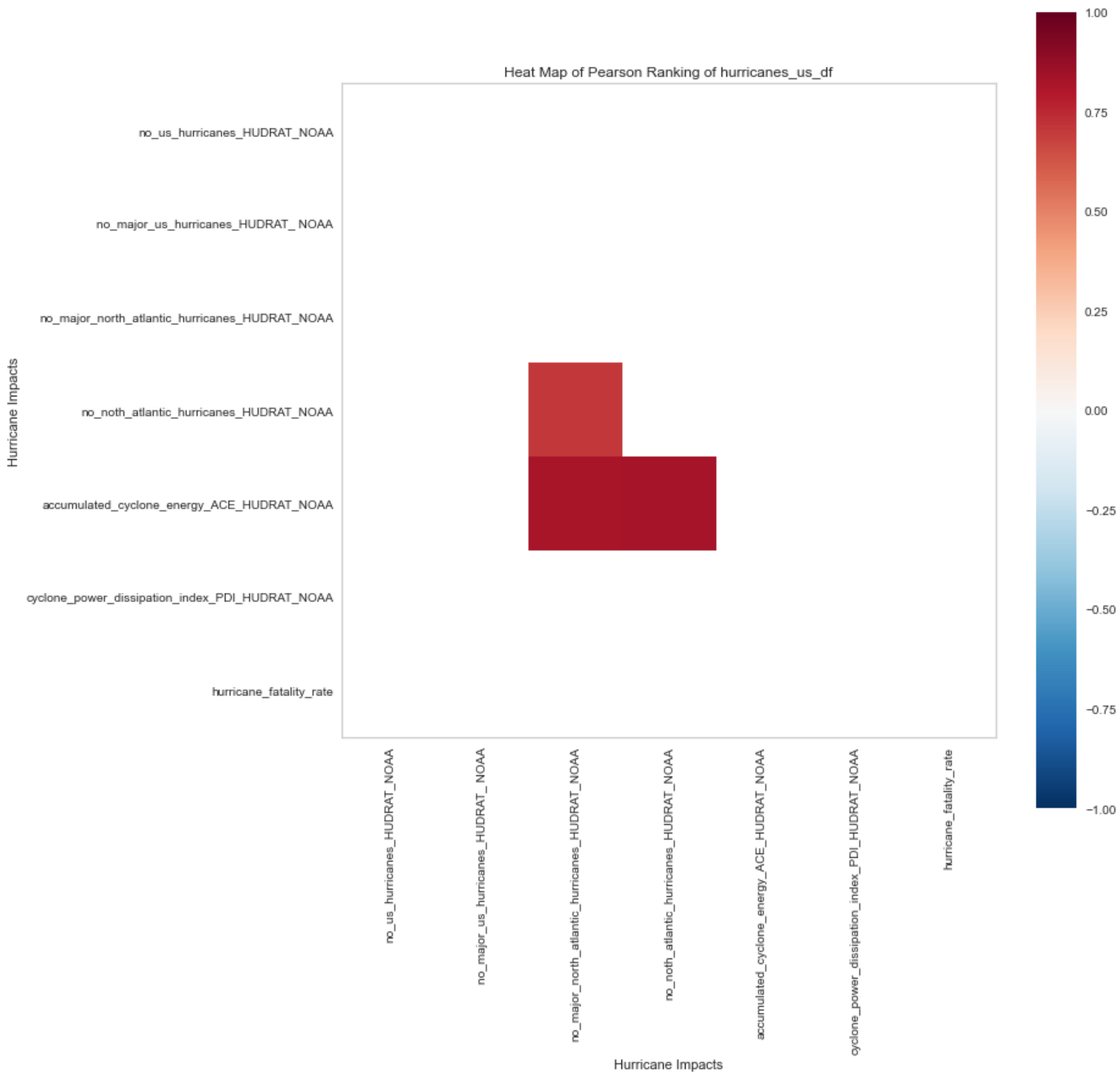
[9]:
```
Index(['Entity', 'Year', 'no_us_hurricanes_HUDRAT_NOAA',
       'no_major_us_hurricanes_HUDRAT_ NOAA',
       'no_major_north_atlantic_hurricanes_HUDRAT_NOAA',
       'no_noth_atlantic_hurricanes_HUDRAT_NOAA',
       'accumulated_cyclone_energy_ACE_HUDRAT_NOAA',
       'cyclone_power_dissipation_index_PDI_HUDRAT_NOAA',
       'hurricane_fatality_rate', 'ACE', 'deaths_hurricanes_us',
       'total_economic_damages_storms', 'death_bin'],
      dtype='object')
```

[10]:
```
# Second file column names
disaster_us_storm_df.columns
```

[10]:
```
Index(['no_deaths_storms', 'no_people_injured_storms',
       'no_people_affected_storms', 'no_people_left_homeless_storms',
       'no_total_people_affected_storms', 'reconstruction_costs_storms',
       'insured_damages_storms', 'total_economic_damages_storms',
       'death_rates_storms', 'injury_rates_storms',
       'no_people_affected_storms_per_100000', 'homelessness_rate_storms',
       'total_no_people_affected_storms_per_100000',
       'total_economic_damages_storms_share_GDP', 'Entity', 'Year'],
      dtype='object')
```

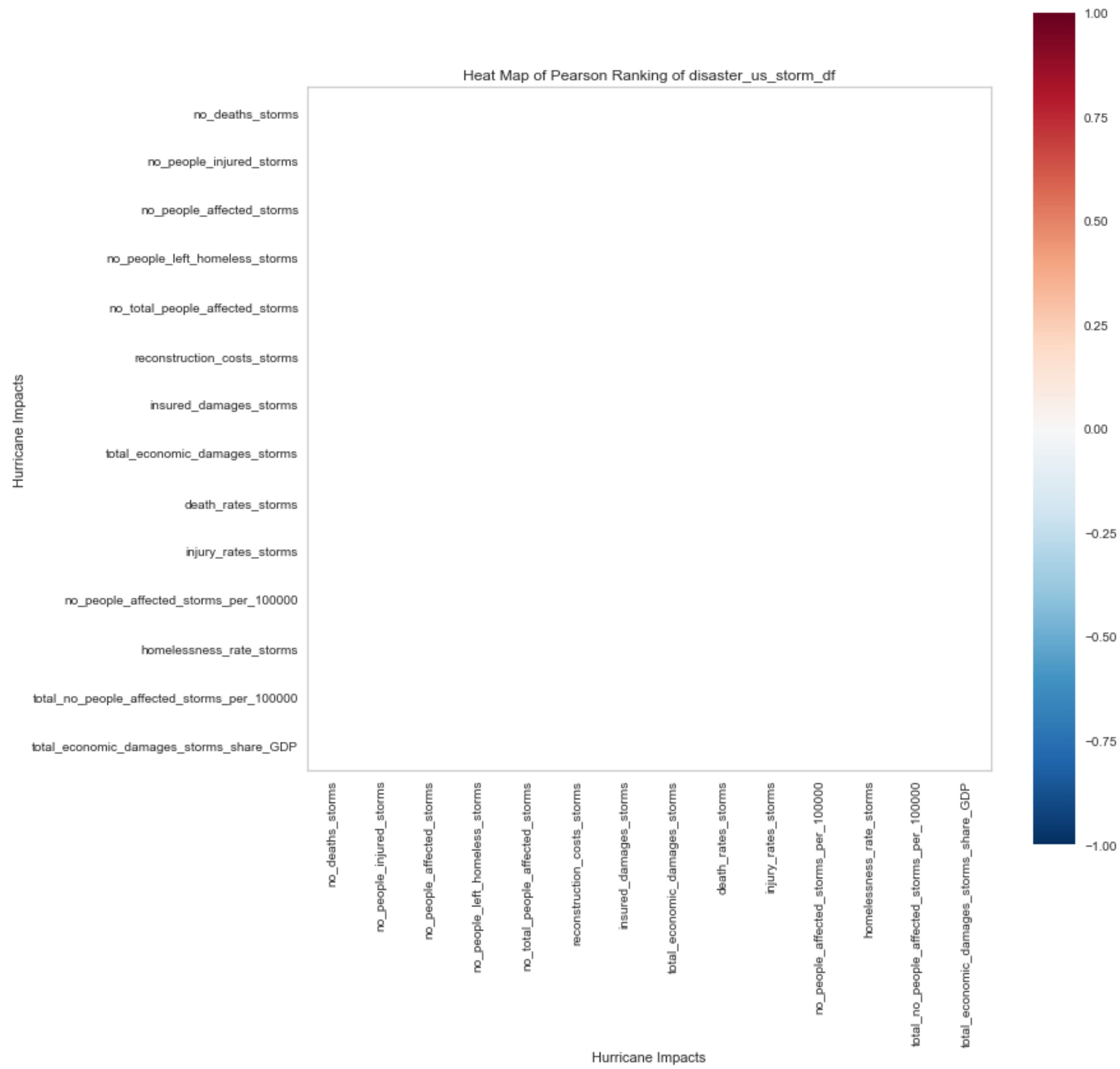
[19]:
```
# Pearson Ranking
# set up the figure size

# New features
num_features = ['no_us_hurricanes_HUDRAT_NOAA','no_major_us_hurricanes_HUDRAT_␣
 ↪NOAA',
               ␣
 ↪'no_major_north_atlantic_hurricanes_HUDRAT_NOAA','no_noth_atlantic_hurricanes_HUDRAT_NOAA',
                'accumulated_cyclone_energy_ACE_HUDRAT_NOAA',
               ␣
 ↪'cyclone_power_dissipation_index_PDI_HUDRAT_NOAA','hurricane_fatality_rate']
plt.rcParams['figure.figsize'] = (12, 12)

# Import the package for visulization of the correlation
from yellowbrick.features import Rank2D

# Extract the numpy arrays from the data frame
X = hurricanes_us_df[num_features].to_numpy()
```

```python
# instantiate the visualizer with the Covariance ranking algorithm
visualizer = Rank2D(features=num_features, algorithm='pearson')
visualizer.fit(X)                    # Fit the data to the visualizer
visualizer.transform(X)              # Transform the data
plt.title("Heat Map of Pearson Ranking of hurricanes_us_df")
plt.xlabel("Hurricane Impacts")
plt.ylabel("Hurricane Impacts")
plt.show()
```



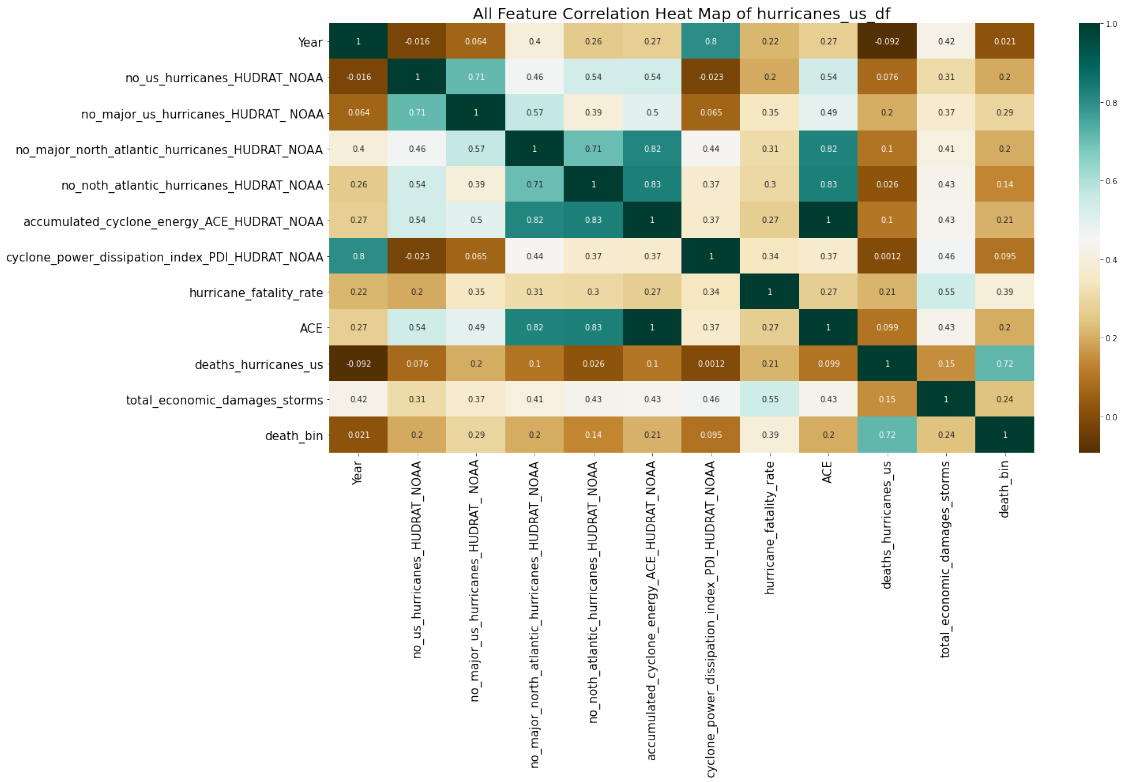
[20]:
```python
# Pearson Ranking
# set up the figure size

# New features
```

```
num_features = ['no_deaths_storms',␣
↪'no_people_injured_storms','no_people_affected_storms',␣
↪'no_people_left_homeless_storms',
                'no_total_people_affected_storms',␣
↪'reconstruction_costs_storms',
                'insured_damages_storms',␣
↪'total_economic_damages_storms','death_rates_storms', 'injury_rates_storms',
                'no_people_affected_storms_per_100000',␣
↪'homelessness_rate_storms',
               ␣
↪'total_no_people_affected_storms_per_100000','total_economic_damages_storms_share_GDP']
plt.rcParams['figure.figsize'] = (12, 12)

# Import the package for visulization of the correlation
from yellowbrick.features import Rank2D

# Extract the numpy arrays from the data frame
X = disaster_us_storm_df[num_features].to_numpy()

# instantiate the visualizer with the Covariance ranking algorithm
visualizer = Rank2D(features=num_features, algorithm='pearson')
visualizer.fit(X)                # Fit the data to the visualizer
visualizer.transform(X)            # Transform the data
plt.title("Heat Map of Pearson Ranking of disaster_us_storm_df")
plt.xlabel("Hurricane Impacts")
plt.ylabel("Hurricane Impacts")
plt.show()
```



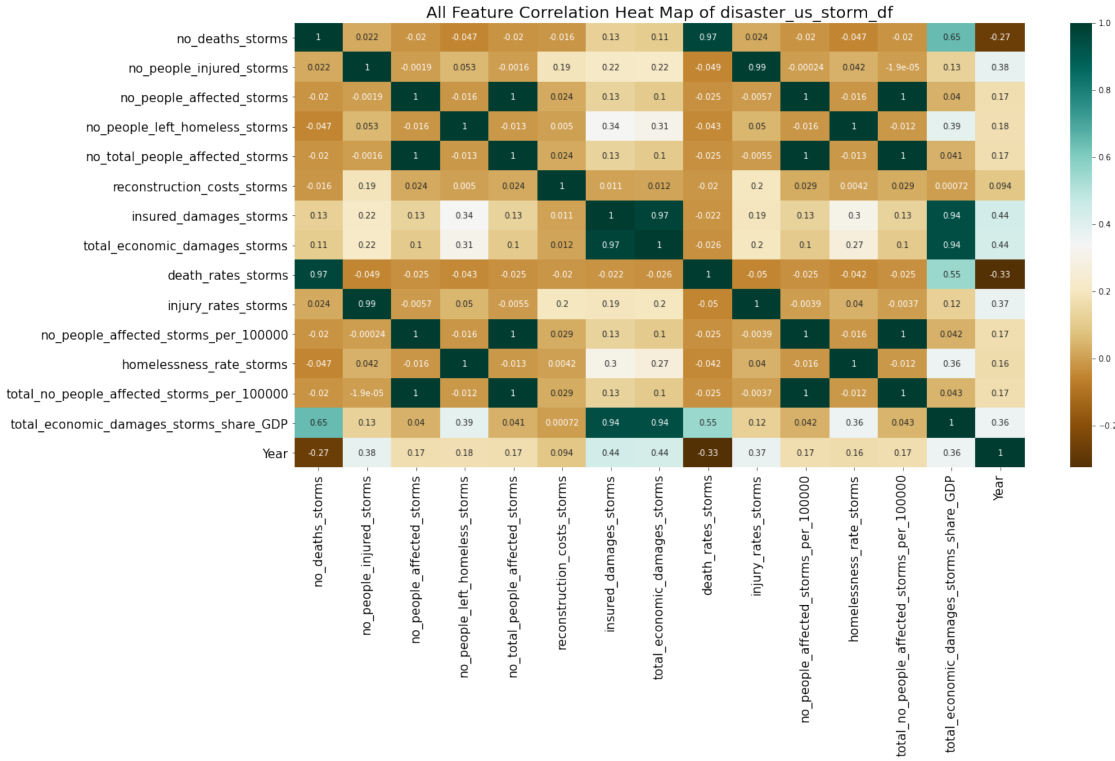
[24]:
```
# Run full correlation on the data set using heat map
plt.figure(figsize=(20,10))
c= hurricanes_us_df.corr()
plt.title("All Feature Correlation Heat Map of hurricanes_us_df",fontsize=20)
plt.xlabel("Hurricanes Impact")
plt.ylabel("Hurricanes Impact")
plt.xticks(fontsize= 15)
plt.yticks(fontsize= 15)
sns.heatmap(c,cmap="BrBG",annot=True)
```

[24]:
```
<AxesSubplot:title={'center':'All Feature Correlation Heat Map of
hurricanes_us_df'}>
```

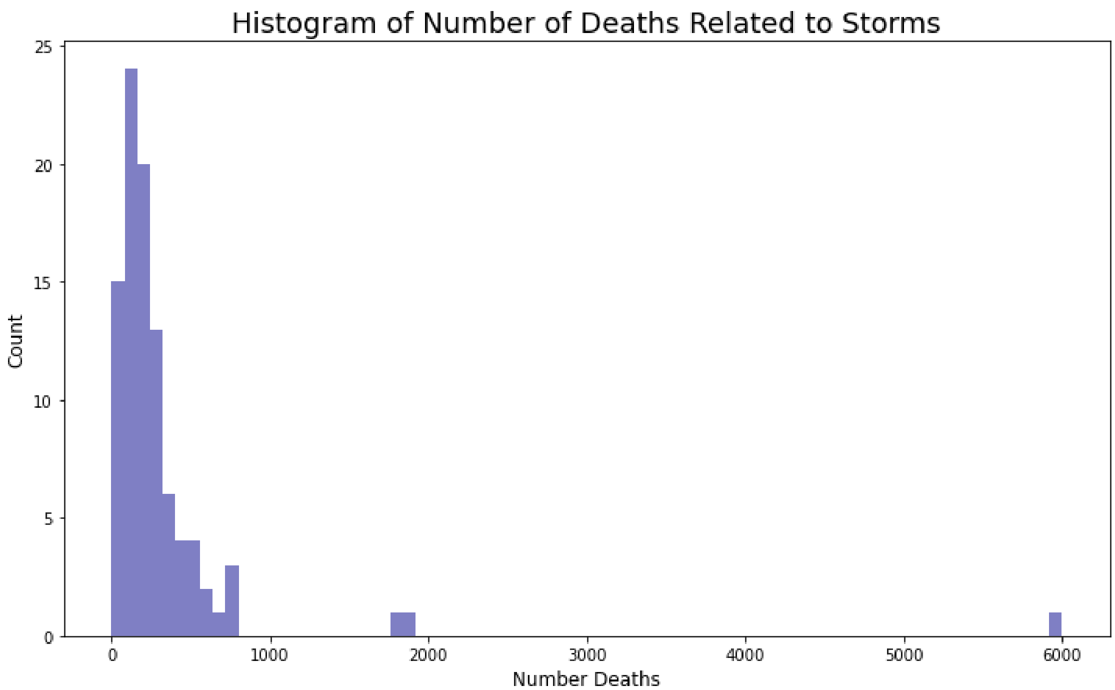



```
[22]: # Run full correlation on the data set using heat map
plt.figure(figsize=(20,10))
c= disaster_us_storm_df.corr()
plt.title("All Feature Correlation Heat Map of disaster_us_storm_df",␣
 ↪fontsize=20)
plt.xlabel("Hurricanes Impact")
plt.ylabel("Hurricanes Impact")
plt.xticks(fontsize= 15)
plt.yticks(fontsize= 15)
sns.heatmap(c,cmap="BrBG",annot=True)
```

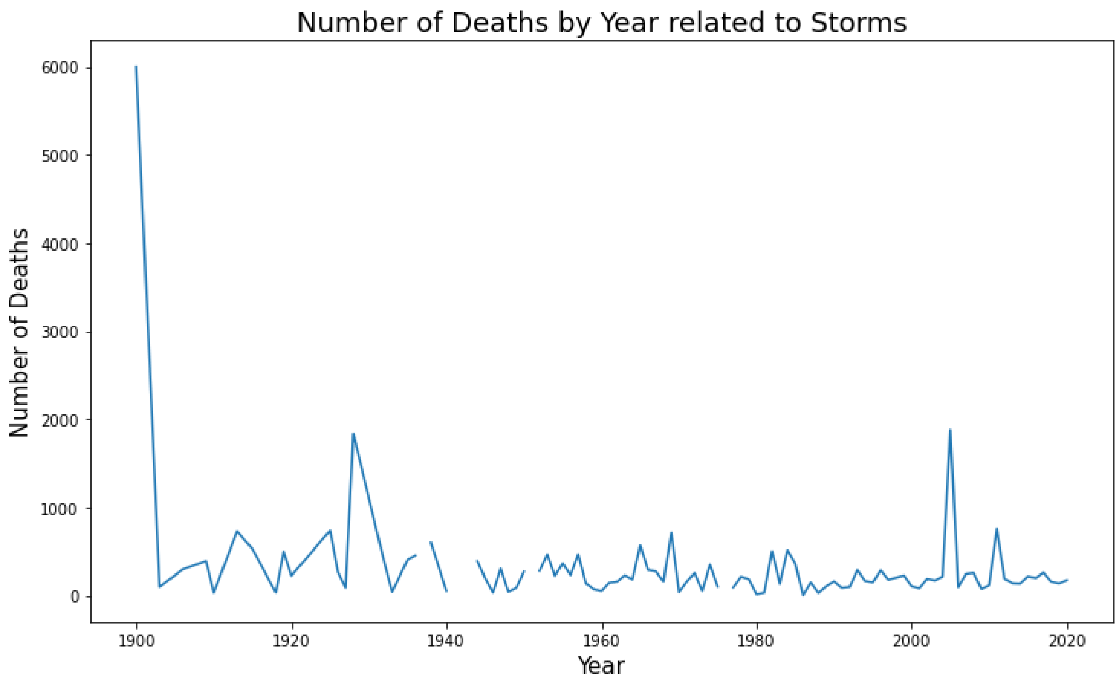
```
[22]: <AxesSubplot:title={'center':'All Feature Correlation Heat Map of
disaster_us_storm_df'}>
```



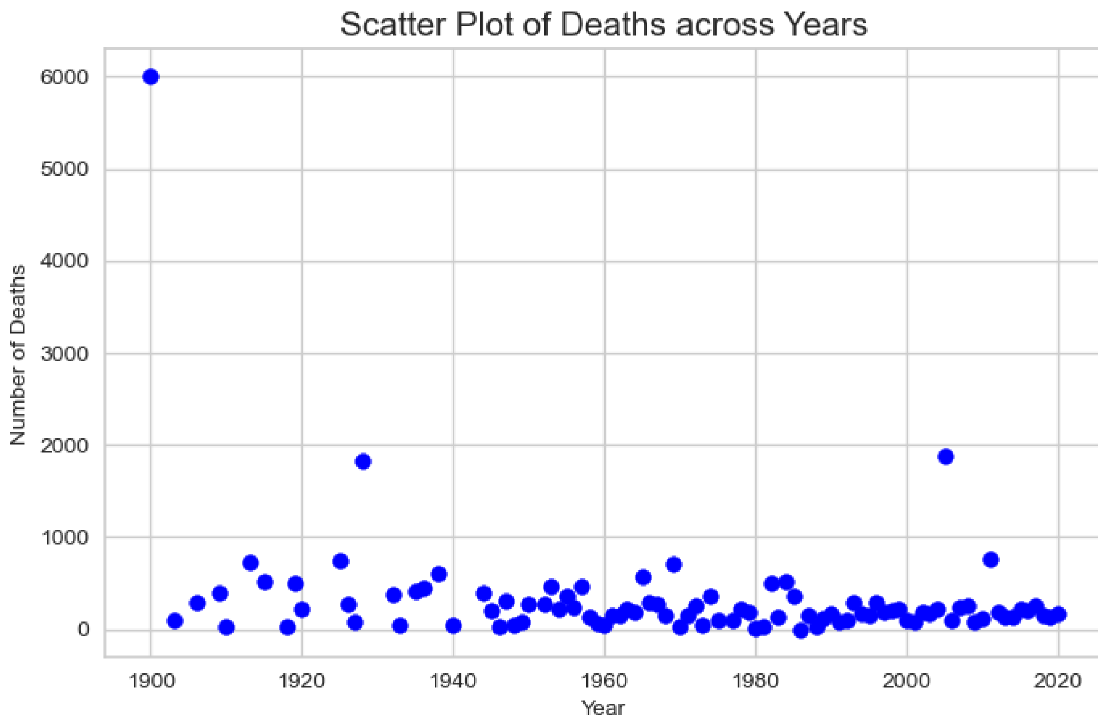

```
[25]: # Histograme of death by year

plt.rcParams['figure.figsize'] = (12, 7)
plt.hist(disaster_us_storm_df['no_deaths_storms'].dropna(), bins=75,␣
 ↪facecolor='darkblue', alpha=0.5)
#Labels
plt.title('Histogram of Number of Deaths Related to Storms', fontsize = 18)
plt.xlabel('Number Deaths', fontsize = 12)
plt.ylabel('Count', fontsize = 12)
plt.ticklabel_format(axis='x', style='plain')
plt.show()
```



[12]:
```
# line chart number of deaths realated to storms

x1 = disaster_us_storm_df['Year']
y1 = disaster_us_storm_df['no_deaths_storms']

plt.rcParams['figure.figsize'] = (12, 7)
plt.plot(x1, y1)
plt.title('Number of Deaths by Year related to Storms', fontsize = 18)
plt.xlabel('Year', fontsize = 15)
plt.ylabel('Number of Deaths', fontsize = 15)
plt.show()
```



[51]:
```
# Scatterplot - To check data
x = disaster_us_storm_df['Year']
y = disaster_us_storm_df['no_deaths_storms']

# Plot
plt.scatter(x,y,color='blue')
plt.rcParams.update({'figure.figsize':(8,5), 'figure.dpi':100})

# Labels
plt.title('Scatter Plot of Deaths across Years',fontsize=15)
plt.xlabel('Year',fontsize=10)
plt.ylabel('Number of Deaths',fontsize=10)
plt.ticklabel_format(axis='x', style='plain')
plt.ticklabel_format(axis='y', style='plain')
plt.show()
```

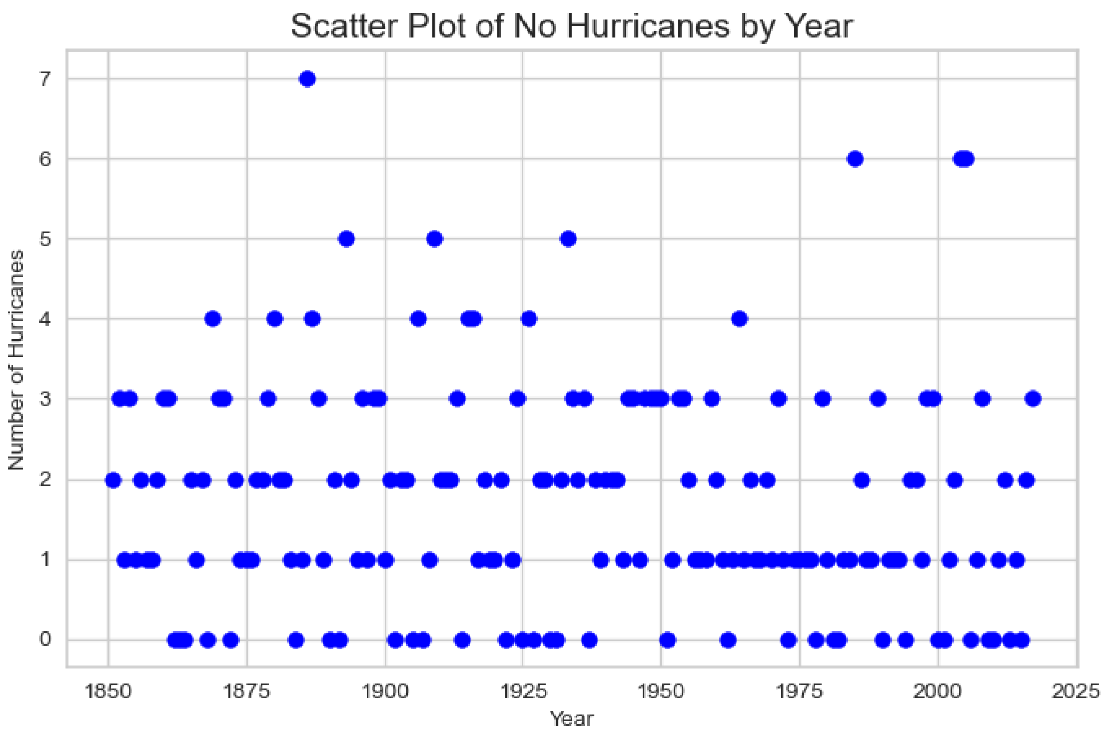


[26]:
```
hurricanes_us_df.columns
```

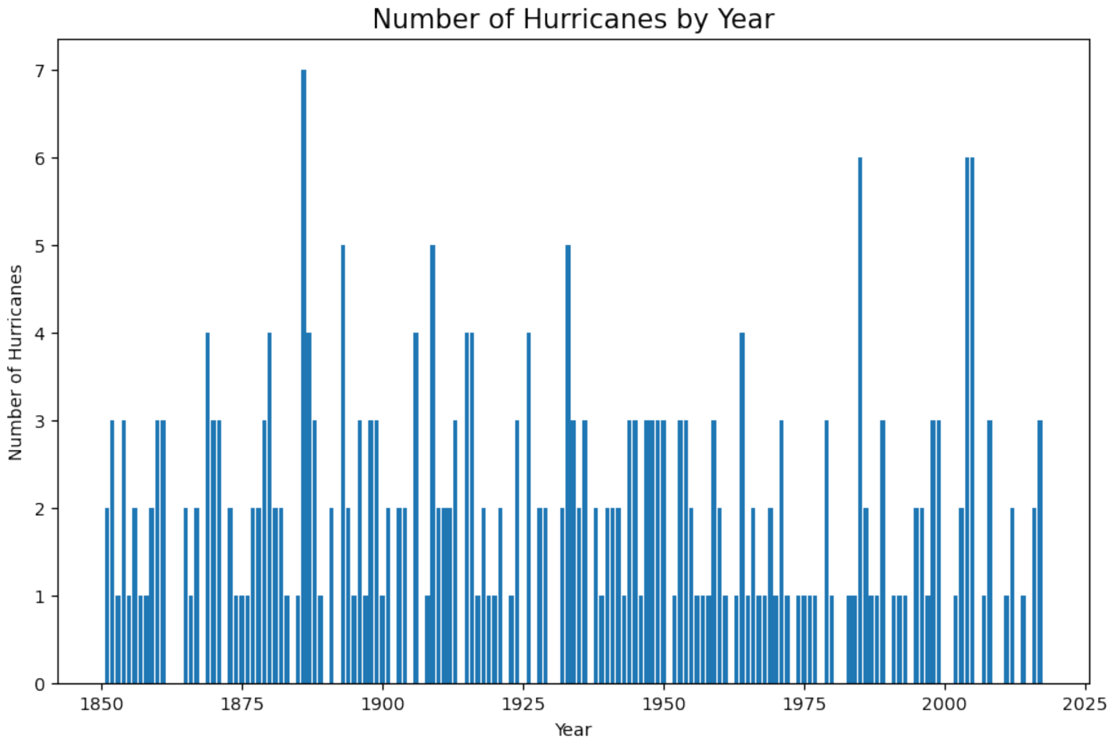
[26]:
```
Index(['Entity', 'Year', 'no_us_hurricanes_HUDRAT_NOAA',
       'no_major_us_hurricanes_HUDRAT_ NOAA',
       'no_major_north_atlantic_hurricanes_HUDRAT_NOAA',
       'no_noth_atlantic_hurricanes_HUDRAT_NOAA',
       'accumulated_cyclone_energy_ACE_HUDRAT_NOAA',
       'cyclone_power_dissipation_index_PDI_HUDRAT_NOAA',
       'hurricane_fatality_rate', 'ACE', 'deaths_hurricanes_us',
       'total_economic_damages_storms', 'death_bin'],
      dtype='object')
```

[53]:
```
# Scatterplot - To check data
x = hurricanes_us_df['Year']
y = hurricanes_us_df['no_us_hurricanes_HUDRAT_NOAA']

# Plot
plt.scatter(x,y,color='blue')
plt.rcParams.update({'figure.figsize':(8,5), 'figure.dpi':100})

#Labels
plt.title('Scatter Plot of No Hurricanes by Year',fontsize=15)
plt.xlabel('Year',fontsize=10)
```

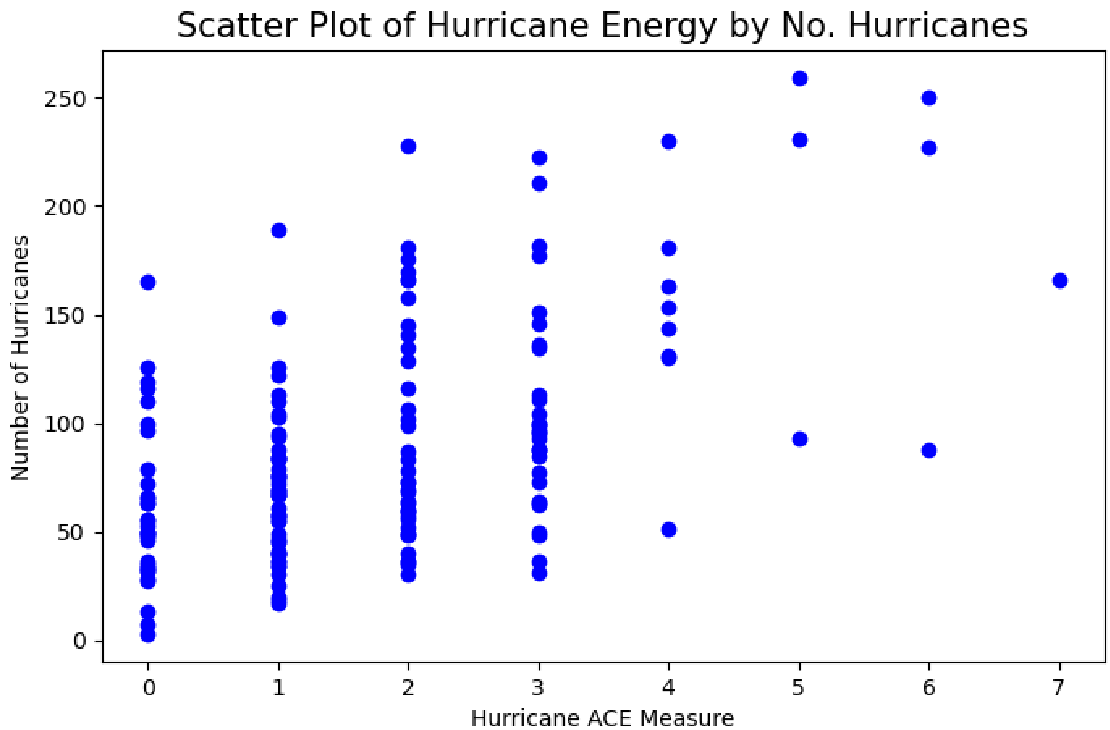
```
plt.ylabel('Number of Hurricanes',fontsize=10)
plt.ticklabel_format(axis='x', style='plain')
plt.ticklabel_format(axis='y', style='plain')
plt.show()
```



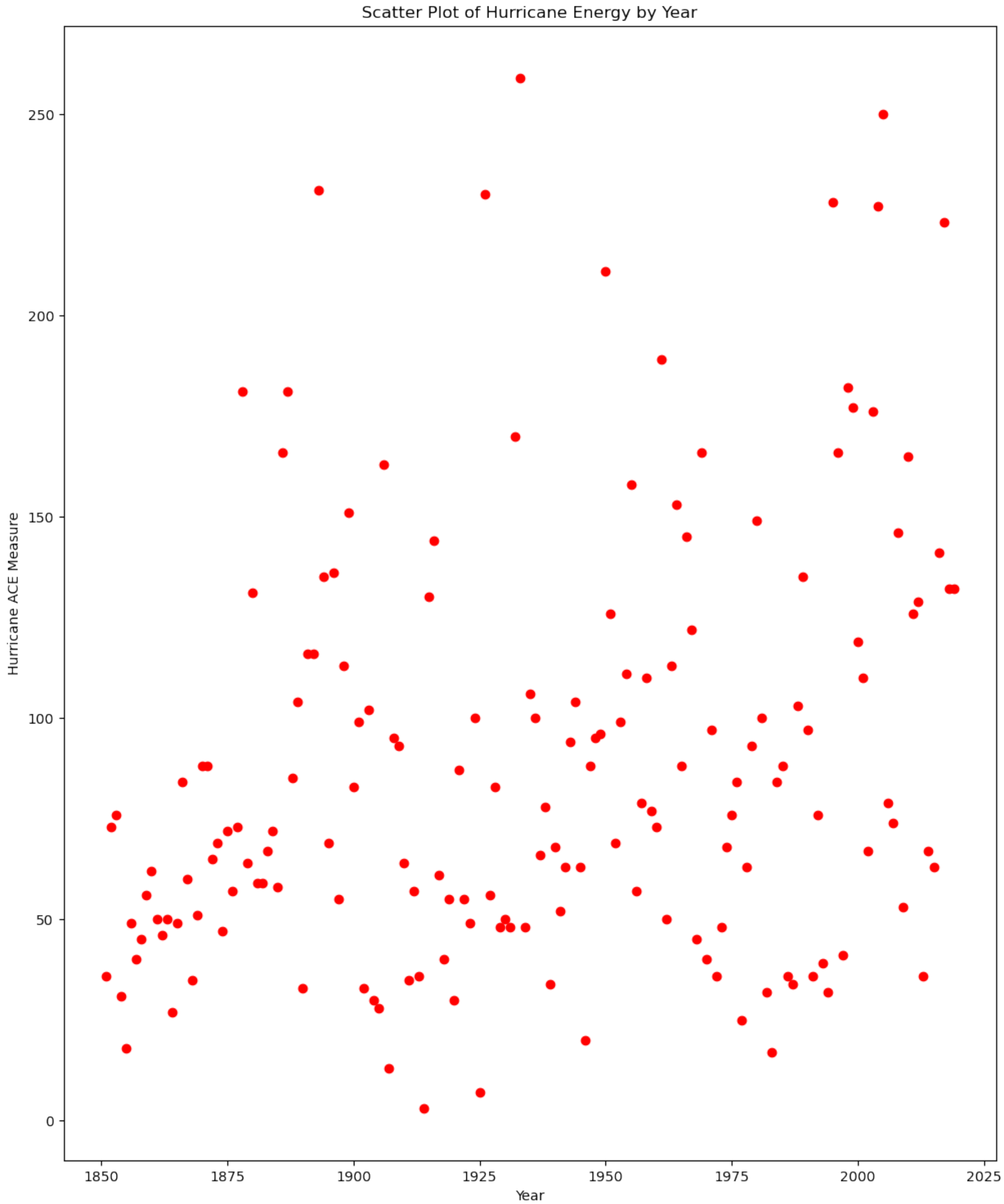
[11]:
```
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
year = hurricanes_us_df['Year']
no_h = hurricanes_us_df['no_us_hurricanes_HUDRAT_NOAA']
ax.bar(year,no_h)
plt.title('Number of Hurricanes by Year',fontsize=15)
plt.xlabel('Year',fontsize=10)
plt.ylabel('Number of Hurricanes',fontsize=10)
plt.show()
```



[9]:
```
# Scatterplot - To check data
plt.clf()

x = hurricanes_us_df['no_us_hurricanes_HUDRAT_NOAA']
y = hurricanes_us_df['accumulated_cyclone_energy_ACE_HUDRAT_NOAA']

# Plot
plt.scatter(x,y,color='blue')
plt.rcParams.update({'figure.figsize':(8,5), 'figure.dpi':100})

#Labels
plt.title('Scatter Plot of Hurricane Energy by No. Hurricanes',fontsize=15)
plt.xlabel('Hurricane ACE Measure',fontsize=10)
plt.ylabel('Number of Hurricanes',fontsize=10)
plt.ticklabel_format(axis='x', style='plain')
plt.ticklabel_format(axis='y', style='plain')
plt.show()
```



[16]:
```
# Scatterplot - To check data
y = hurricanes_us_df['accumulated_cyclone_energy_ACE_HUDRAT_NOAA']
x = hurricanes_us_df['Year']

# Plot
plt.scatter(x,y,color='red')
plt.rcParams.update({'figure.figsize':(8,8), 'figure.dpi':100})

#Labels
plt.title('Scatter Plot of Hurricane Energy by Year',fontsize=12)
plt.ylabel('Hurricane ACE Measure',fontsize=10)
plt.xlabel('Year',fontsize=10)
plt.ticklabel_format(axis='x', style='plain')
plt.ticklabel_format(axis='y', style='plain')
plt.show()
```



Note: reducing the data to 1950 and greater since economic damages have not been record before 1950.

[17]:
```
# Filter data down to US only
hurricanes_us_1900_df = hurricanes_us_df[hurricanes_us_df['Year'] >= 1900]
```

```
[18]: # Drop hurricane_fatality_rate
hurricanes_us_1900_df.drop(['hurricane_fatality_rate'], axis=1)
```

```
[18]:             Entity  Year  no_us_hurricanes_HUDRAT_NOAA  \
49   North Atlantic  1900                             1
50   North Atlantic  1901                             2
51   North Atlantic  1902                             0
52   North Atlantic  1903                             2
53   North Atlantic  1904                             2
..              …    …                            …
164  North Atlantic  2015                             0
165  North Atlantic  2016                             2
166  North Atlantic  2017                             3
167  North Atlantic  2018                             8
168  North Atlantic  2019                             6

     no_major_us_hurricanes_HUDRAT_ NOAA  \
49                                  1
50                                  0
51                                  0
52                                  0
53                                  0
..                                …
164                                 0
165                                 0
166                                 2
167                                 2
168                                 3

     no_major_north_atlantic_hurricanes_HUDRAT_NOAA  \
49                                               2
50                                               0
51                                               0
52                                               1
53                                               0
..                                             …
164                                              2
165                                              4
166                                              6
167                                              2
168                                              2

     no_noth_atlantic_hurricanes_HUDRAT_NOAA  \
49                                        3
50                                        6
51                                        3
52                                        7
```

```
53                                          4
..                                        …
164                                         4
165                                         7
166                                        10
167                                         8
168                                         8

     accumulated_cyclone_energy_ACE_HUDRAT_NOAA  \
49                                           83
50                                           99
51                                           33
52                                          102
53                                           30
..                                          …
164                                          63
165                                         141
166                                         223
167                                         132
168                                         132

     cyclone_power_dissipation_index_PDI_HUDRAT_NOAA        ACE  \
49                                               0.0   83.3450
50                                               0.0   98.9750
51                                               0.0   32.6500
52                                               0.0  102.0700
53                                               0.0   30.3450
..                                               …        …
164                                              NaN   62.6850
165                                              NaN  141.2525
166                                              NaN  224.8775
167                                              NaN  132.5825
168                                              NaN  132.2025

     deaths_hurricanes_us  total_economic_damages_storms  death_bin
49                 6000.0                        30000.0       10.0
50                    0.0                            0.0        1.0
51                    0.0                            0.0        1.0
52                   98.0                            0.0        1.0
53                    0.0                            0.0        1.0
..                    …                           …          …
164                 218.0                     15720000.0        3.0
165                 199.0                     28050000.0        2.0
166                 261.0                    171110000.0        3.0
167                 158.0                     38875000.0        2.0
168                 139.0                     11325000.0        2.0
```

```
[120 rows x 12 columns]
```

[28]:
```
hurricanes_us_1900_df.columns
```

[28]:
```
Index(['Entity', 'Year', 'no_us_hurricanes_HUDRAT_NOAA',
       'no_major_us_hurricanes_HUDRAT_ NOAA',
       'no_major_north_atlantic_hurricanes_HUDRAT_NOAA',
       'no_noth_atlantic_hurricanes_HUDRAT_NOAA',
       'accumulated_cyclone_energy_ACE_HUDRAT_NOAA',
       'cyclone_power_dissipation_index_PDI_HUDRAT_NOAA',
       'hurricane_fatality_rate', 'ACE', 'deaths_hurricanes_us',
       'total_economic_damages_storms', 'death_bin'],
      dtype='object')
```

0.1 Modeling

[19]:
```
# Import libraries
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

Notes: Reduce years from 1850 to 1900 due to a lot of missing data.

[20]:
```
# filter for future use
hurricanes_1900_df = hurricanes_us_df[hurricanes_us_df['Year'] >= 1900]
```

[21]:
```
hurricanes_1900_df.head
```

[21]:
```
<bound method NDFrame.head of              Entity  Year
no_us_hurricanes_HUDRAT_NOAA  \
49   North Atlantic  1900                             1
50   North Atlantic  1901                             2
51   North Atlantic  1902                             0
52   North Atlantic  1903                             2
53   North Atlantic  1904                             2
..              …    …                          …
164  North Atlantic  2015                             0
165  North Atlantic  2016                             2
166  North Atlantic  2017                             3
167  North Atlantic  2018                             8
168  North Atlantic  2019                             6

     no_major_us_hurricanes_HUDRAT_ NOAA  \
49                                     1
```

```
50                                    0
51                                    0
52                                    0
53                                    0
..                                  ...
164                                   0
165                                   0
166                                   2
167                                   2
168                                   3

     no_major_north_atlantic_hurricanes_HUDRAT_NOAA  \
49                                                2
50                                                0
51                                                0
52                                                1
53                                                0
..                                              ...
164                                               2
165                                               4
166                                               6
167                                               2
168                                               2

     no_noth_atlantic_hurricanes_HUDRAT_NOAA  \
49                                         3
50                                         6
51                                         3
52                                         7
53                                         4
..                                       ...
164                                        4
165                                        7
166                                       10
167                                        8
168                                        8

     accumulated_cyclone_energy_ACE_HUDRAT_NOAA  \
49                                           83
50                                           99
51                                           33
52                                          102
53                                           30
..                                          ...
164                                          63
165                                         141
166                                         223
```

```
167                                             132
168                                             132

     cyclone_power_dissipation_index_PDI_HUDRAT_NOAA  hurricane_fatality_rate  \
49                                               0.0                 0.000000
50                                               0.0                 0.000000
51                                               0.0                 0.000000
52                                               0.0                 0.000000
53                                               0.0                 0.000000
..                                               …                   …
164                                              NaN                 0.043557
165                                              NaN                      NaN
166                                              NaN                      NaN
167                                              NaN                      NaN
168                                              NaN                      NaN

          ACE  deaths_hurricanes_us  total_economic_damages_storms  death_bin
49    83.3450                6000.0                        30000.0       10.0
50    98.9750                   0.0                            0.0        1.0
51    32.6500                   0.0                            0.0        1.0
52   102.0700                  98.0                            0.0        1.0
53    30.3450                   0.0                            0.0        1.0
..        …                     …                              …      …
164   62.6850                 218.0                     15720000.0        3.0
165  141.2525                 199.0                     28050000.0        2.0
166  224.8775                 261.0                    171110000.0        3.0
167  132.5825                 158.0                     38875000.0        2.0
168  132.2025                 139.0                     11325000.0        2.0

[120 rows x 13 columns]>
```

[22]:
```
# Change data types for modeling

hurricanes_1900_df['no_us_hurricanes_HUDRAT_NOAA'] =␣
↪hurricanes_1900_df['no_us_hurricanes_HUDRAT_NOAA'].astype(int)
hurricanes_1900_df['no_major_us_hurricanes_HUDRAT_ NOAA'] =␣
↪hurricanes_1900_df['no_major_us_hurricanes_HUDRAT_ NOAA'].astype(int)
hurricanes_1900_df['no_major_north_atlantic_hurricanes_HUDRAT_NOAA'] =␣
↪hurricanes_1900_df['no_major_north_atlantic_hurricanes_HUDRAT_NOAA'].
↪astype(int)
hurricanes_1900_df['no_noth_atlantic_hurricanes_HUDRAT_NOAA'] =␣
↪hurricanes_1900_df['no_noth_atlantic_hurricanes_HUDRAT_NOAA'].astype(int)
hurricanes_1900_df['accumulated_cyclone_energy_ACE_HUDRAT_NOAA'] =␣
↪hurricanes_1900_df['accumulated_cyclone_energy_ACE_HUDRAT_NOAA'].astype(int)
hurricanes_1900_df['death_bin'] = hurricanes_1900_df['death_bin'].astype(int)
```

```
<ipython-input-22-ac42eb629da4>: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
  hurricanes_1900_df['no_us_hurricanes_HUDRAT_NOAA'] =
hurricanes_1900_df['no_us_hurricanes_HUDRAT_NOAA'].astype(int)
<ipython-input-22-ac42eb629da4>:4: 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
  hurricanes_1900_df['no_major_us_hurricanes_HUDRAT_ NOAA'] =
hurricanes_1900_df['no_major_us_hurricanes_HUDRAT_ NOAA'].astype(int)
<ipython-input-22-ac42eb629da4>:5: 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
  hurricanes_1900_df['no_major_north_atlantic_hurricanes_HUDRAT_NOAA'] =
hurricanes_1900_df['no_major_north_atlantic_hurricanes_HUDRAT_NOAA'].astype(int)
<ipython-input-22-ac42eb629da4>:6: 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
  hurricanes_1900_df['no_noth_atlantic_hurricanes_HUDRAT_NOAA'] =
hurricanes_1900_df['no_noth_atlantic_hurricanes_HUDRAT_NOAA'].astype(int)
<ipython-input-22-ac42eb629da4>:7: 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
  hurricanes_1900_df['accumulated_cyclone_energy_ACE_HUDRAT_NOAA'] =
hurricanes_1900_df['accumulated_cyclone_energy_ACE_HUDRAT_NOAA'].astype(int)
<ipython-input-22-ac42eb629da4>:8: 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
  hurricanes_1900_df['death_bin'] = hurricanes_1900_df['death_bin'].astype(int)
```

[23]:

```
# Change year to to int
hurricanes_1900_df['Year'] = hurricanes_1900_df['Year'].astype(int)
```

```
<ipython-input-23-b1ff7401944b>:2: 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
  hurricanes_1900_df['Year'] = hurricanes_1900_df['Year'].astype(int)
```

[34]:

```
hurricanes_1900_df.head()
```

[34]:

```
           Entity  Year  no_us_hurricanes_HUDRAT_NOAA  \
49  North Atlantic  1900                             1
50  North Atlantic  1901                             2
51  North Atlantic  1902                             0
52  North Atlantic  1903                             2
53  North Atlantic  1904                             2

    no_major_us_hurricanes_HUDRAT_ NOAA  \
49                                    1
50                                    0
51                                    0
52                                    0
53                                    0

    no_major_north_atlantic_hurricanes_HUDRAT_NOAA  \
49                                               2
50                                               0
51                                               0
52                                               1
53                                               0

    no_noth_atlantic_hurricanes_HUDRAT_NOAA  \
49                                        3
50                                        6
51                                        3
52                                        7
53                                        4

    accumulated_cyclone_energy_ACE_HUDRAT_NOAA  \
49                                          83
50                                          99
51                                          33
52                                         102
53                                          30
```

```
    cyclone_power_dissipation_index_PDI_HUDRAT_NOAA  hurricane_fatality_rate  \
49                                              0.0                      0.0
50                                              0.0                      0.0
51                                              0.0                      0.0
52                                              0.0                      0.0
53                                              0.0                      0.0

        ACE  deaths_hurricanes_us  total_economic_damages_storms  death_bin
49   83.345                6000.0                        30000.0         10
50   98.975                   0.0                            0.0          1
51   32.650                   0.0                            0.0          1
52  102.070                  98.0                            0.0          1
53   30.345                   0.0                            0.0          1
```

0.1.1 Linear regression model

[36]:
```
# split the data in training and Test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=0)
```

[37]:
```
# train the model using linear regression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

[37]:
```
LinearRegression()
```

[69]:
```
# used deaths_hurricanes_us
coeff = pd.DataFrame(regressor.coef_, X.columns, columns=['Coefficient'])
coeff
```

[69]:

	Coefficient
no_us_hurricanes_HUDRAT_NOAA	-53.974540
no_major_us_hurricanes_HUDRAT_ NOAA	262.029082
no_major_north_atlantic_hurricanes_HUDRAT_NOAA	-42.694718
no_noth_atlantic_hurricanes_HUDRAT_NOAA	-44.326837
accumulated_cyclone_energy_ACE_HUDRAT_NOAA	3.175403

[70]:
```
# used deaths_hurricanes_us

y_pred = regressor.predict(X_test)
```

[71]:
```
# used deaths_hurricanes_us
print(regressor.score(X_test, y_test))
```

```
-2.4876948515430244
```

[72]:
```
# calculate rmse and r2
# used deaths_hurricanes_us
```

```
from sklearn.metrics import mean_squared_error, r2_score
rmse = np.sqrt(mean_squared_error(y_test,y_pred))
r2 = r2_score(y_test,y_pred)
```

[73]:
```
# print rmse
rmse
```

[73]: 312.24188966237466

[74]:
```
# print r
r2
```

[74]: -2.4876948515430244

[75]:
```
# Calculation of Mean Squared Error (MSE)
mean_squared_error(y_test,y_pred)
```

[75]: 97494.99765993055

[76]:
```
mean_absolute_error(y_test,y_pred)
```

[76]: 270.6704113740778

[77]:
```
df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})

df.head(25)
```

[77]:

	Actual	Predicted
97	42.0	251.255390
143	165.0	254.858900
144	150.0	463.229674
57	0.0	225.259542
146	180.0	186.768272
71	0.0	285.198459
56	0.0	327.506747
59	30.0	205.828009
94	196.0	541.387837
138	108.0	419.334123
82	40.0	619.077922
99	272.0	574.607641
51	0.0	258.034303
109	50.0	409.414166
168	139.0	727.615619
123	352.0	447.511690
79	0.0	270.953562
92	0.0	223.716258
160	762.0	151.286056
125	0.0	147.635387

```
112   229.0  152.700530
108    71.0  235.160728
65      0.0  394.902630
73      0.0  134.819598
```

[47]:

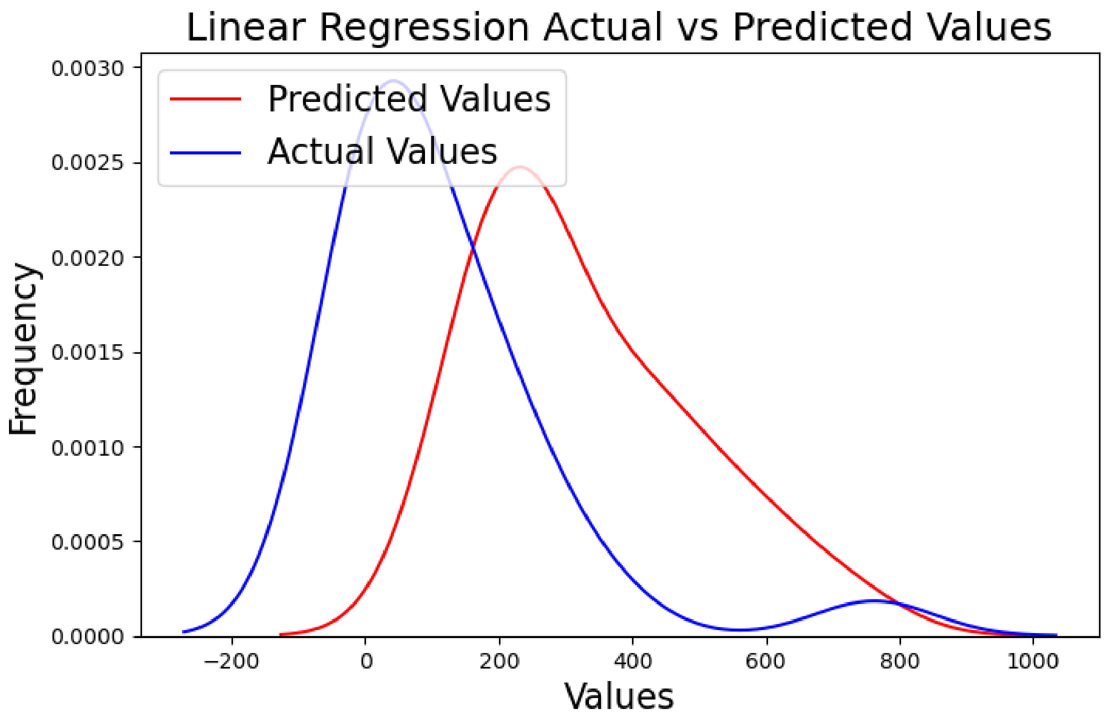
```
print(('R-Squared :'), regressor.score(X_test, y_test))
```

```
R-Squared : -2.4876948515430244
```

[48]:

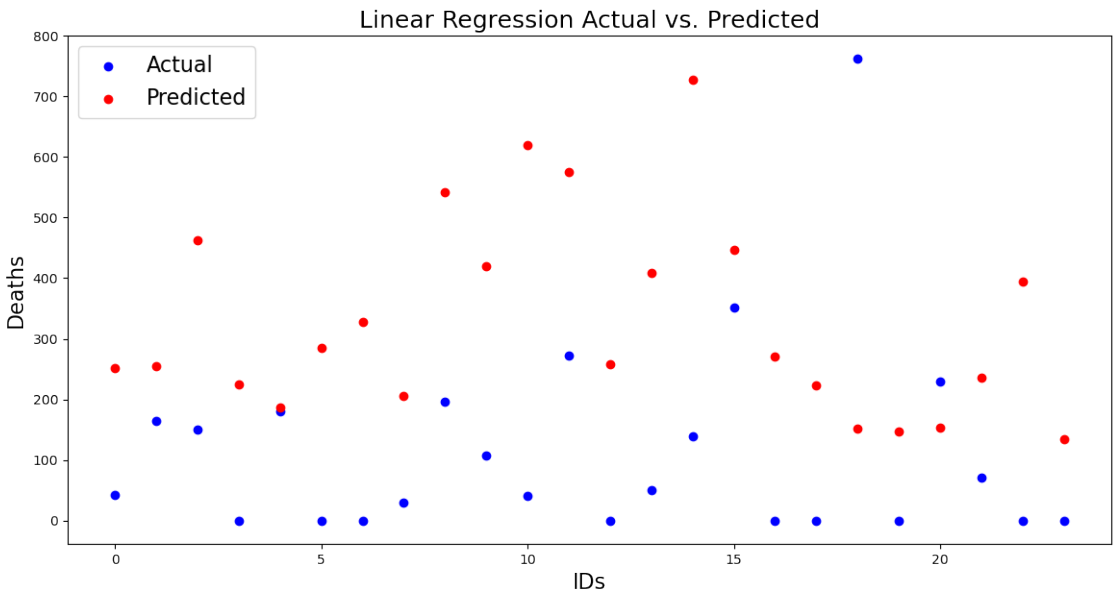
```
sns.distplot(y_pred, hist = False, color = 'r', label = 'Predicted Values')
sns.distplot(y_test, hist = False, color = 'b', label = 'Actual Values')
plt.title('Linear Regression Actual vs Predicted Values', fontsize = 18)
plt.xlabel('Values', fontsize = 16)
plt.ylabel('Frequency', fontsize = 16)
plt.legend(loc = 'upper left', fontsize = 16)
plt.ticklabel_format(style='plain', axis='x')


plt.savefig('ap.png')
```

```
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
```



[49]:

```
plt.figure(figsize=(14,7))
plt.scatter(range(len(y_test)), y_test, color='blue', label='Actual')
plt.scatter(range(len(y_pred)), y_pred, color='red', label='Predicted')
plt.title('Linear Regression Actual vs. Predicted', fontsize = 18)
plt.xlabel('IDs', fontsize = 16)
plt.ylabel('Deaths', fontsize = 16)
plt.legend(loc = 'upper left', fontsize = 16)
plt.ticklabel_format(style='plain', axis='x')
plt.show()
```



Using death_bin as alternative to using actual number of deaths. Note: below is done with bining of deaths to see if that increases model accuracy

[24]:
```
# split out X and y
# Updated from deaths_hurricanes_us to death_bin

X = hurricanes_1900_df[['no_us_hurricanes_HUDRAT_NOAA',
       'no_major_us_hurricanes_HUDRAT_ NOAA',
       'no_major_north_atlantic_hurricanes_HUDRAT_NOAA',
       'no_noth_atlantic_hurricanes_HUDRAT_NOAA',
       'accumulated_cyclone_energy_ACE_HUDRAT_NOAA']]

y = hurricanes_1900_df['death_bin']
```

[25]:
```
# split the data in training and Test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=0)
```

[26]:
```
# train the model using linear regression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

[26]: LinearRegression()

[39]:
```
# used death_bin
coeff2 = pd.DataFrame(regressor.coef_, X.columns, columns=['Coefficient'])
coeff2
```

[39]:

	Coefficient
no_us_hurricanes_HUDRAT_NOAA	-0.011755
no_major_us_hurricanes_HUDRAT_ NOAA	0.922708
no_major_north_atlantic_hurricanes_HUDRAT_NOAA	-0.154419
no_noth_atlantic_hurricanes_HUDRAT_NOAA	-0.099639
accumulated_cyclone_energy_ACE_HUDRAT_NOAA	0.010755

[40]:
```
y_pred = regressor.predict(X_test)
```

[41]:
```
print(regressor.score(X_test, y_test))
```

```
-1.0763589168854542
```

Note: The number improved with using binning. Original r2 was -2.47

[42]:
```
from sklearn.metrics import mean_squared_error, r2_score
rmse = np.sqrt(mean_squared_error(y_test,y_pred))
r2 = r2_score(y_test,y_pred)
```

[43]:
```
# print rmse
rmse
```

[43]: 1.9501418472777152

[44]:
```
r2
```

[44]: -1.0763589168854542

[45]:
```
mean_squared_error(y_test,y_pred)
```

[45]: 3.80305322450374

[46]:
```
mean_absolute_error(y_test,y_pred)
```

[46]: 1.6580823413052255

[53]:
```
df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})

df.head(25)
```

[53]:

	Actual	Predicted
97	1	2.830455
143	2	2.182054
144	2	3.619981
57	1	2.394514
146	2	2.112673
71	1	2.274995
56	1	2.276630
59	1	2.348278

```
94         2   3.817487
138        2   3.469845
82         1   4.686403
99         3   4.193684
51         1   2.192809
109        1   3.113721
168        2   5.148093
123        4   3.071702
79         1   2.166441
92         1   2.328979
160        7   2.165016
125        1   2.121792
112        3   2.179623
108        1   2.846068
65         1   3.715417
73         1   2.369998
```

[48]:
```python
df.columns
```

[48]: `Index(['Actual', 'Predicted'], dtype='object')`

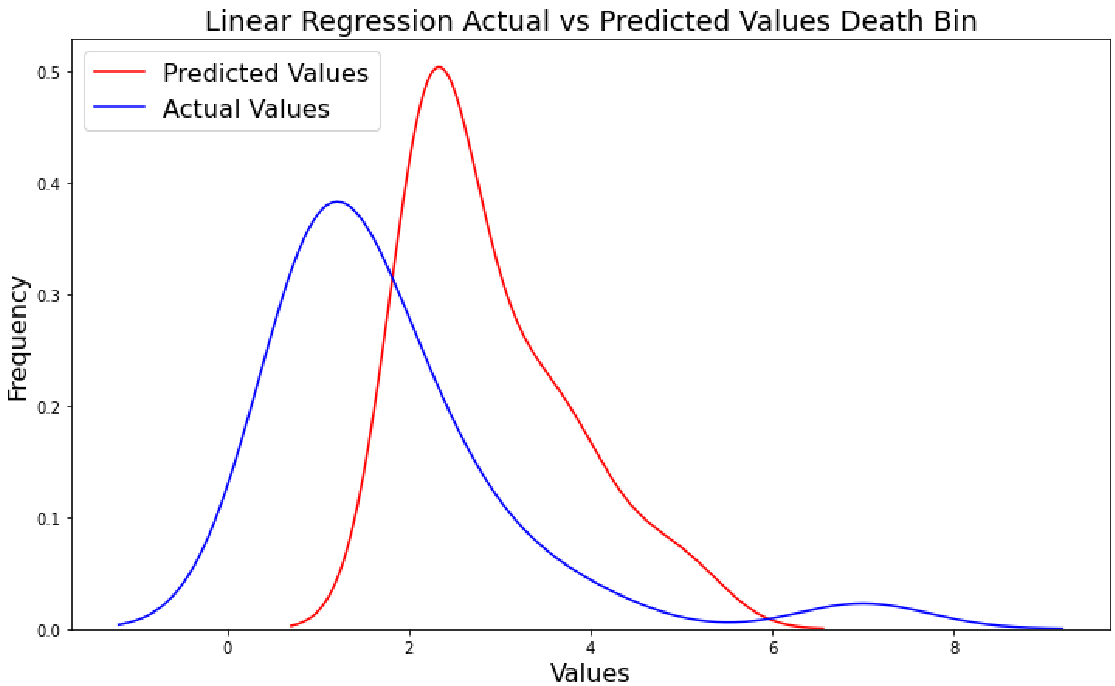
[50]:
```python
print(('R-Squared :'), regressor.score(X_test, y_test))
```

```
R-Squared : -1.0763589168854542
```

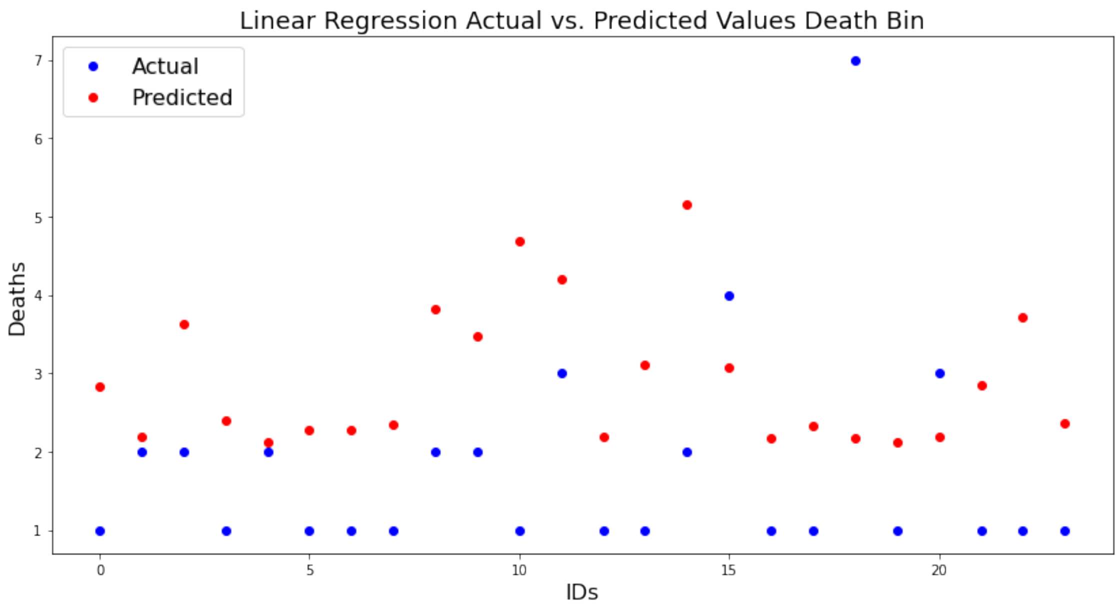
[51]:
```python
sns.distplot(y_pred, hist = False, color = 'r', label = 'Predicted Values')
sns.distplot(y_test, hist = False, color = 'b', label = 'Actual Values')
plt.title('Linear Regression Actual vs Predicted Values Death Bin', fontsize =␣
 ↪18)
plt.xlabel('Values', fontsize = 16)
plt.ylabel('Frequency', fontsize = 16)
plt.legend(loc = 'upper left', fontsize = 16)
plt.ticklabel_format(style='plain', axis='x')


plt.savefig('ap.png')
```

```
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
```

```
warnings.warn(msg, FutureWarning)
```



[52]:
```
plt.figure(figsize=(14,7))
plt.scatter(range(len(y_test)), y_test, color='blue', label='Actual')
plt.scatter(range(len(y_pred)), y_pred, color='red', label='Predicted')
plt.title('Linear Regression Actual vs. Predicted Values Death Bin', fontsize =␣
↪18)
plt.xlabel('IDs', fontsize = 16)
plt.ylabel('Deaths', fontsize = 16)
plt.legend(loc = 'upper left', fontsize = 16)
plt.ticklabel_format(style='plain', axis='x')
plt.show()
```



0.1.2 Decision Tree Model

[55]:
```
# using decision tree model
from sklearn.tree import DecisionTreeRegressor
dtr = DecisionTreeRegressor()
dtr.fit(X_train, y_train)
```

[55]:
```
DecisionTreeRegressor()
```

[79]:
```
y_pred = dtr.predict(X_test)
```

[80]:
```
df2 = pd.DataFrame({'Real Values':y_test, 'Predicted Values':y_pred})
```

[81]:
```
df2.head(25)
```

[81]:

	Real Values	Predicted Values
97	42.0	391.0
143	165.0	503.0
144	150.0	227.0
57	0.0	162.0
146	180.0	87.0
71	0.0	87.0
56	0.0	739.0
59	30.0	0.0
94	196.0	0.0
138	108.0	87.0
82	40.0	1882.0

```
99       272.0        261.0
51         0.0        503.0
109       50.0         97.0
168      139.0        158.0
123      352.0          0.0
79         0.0          0.0
92         0.0        162.0
160      762.0          0.0
125        0.0        517.0
112      229.0        105.0
108       71.0        102.0
65         0.0        183.0
73         0.0        455.0
```

[82]:
```
X_test.shape
```

[82]: (24, 5)

[83]:
```
y_test.shape
```

[83]: (24,)

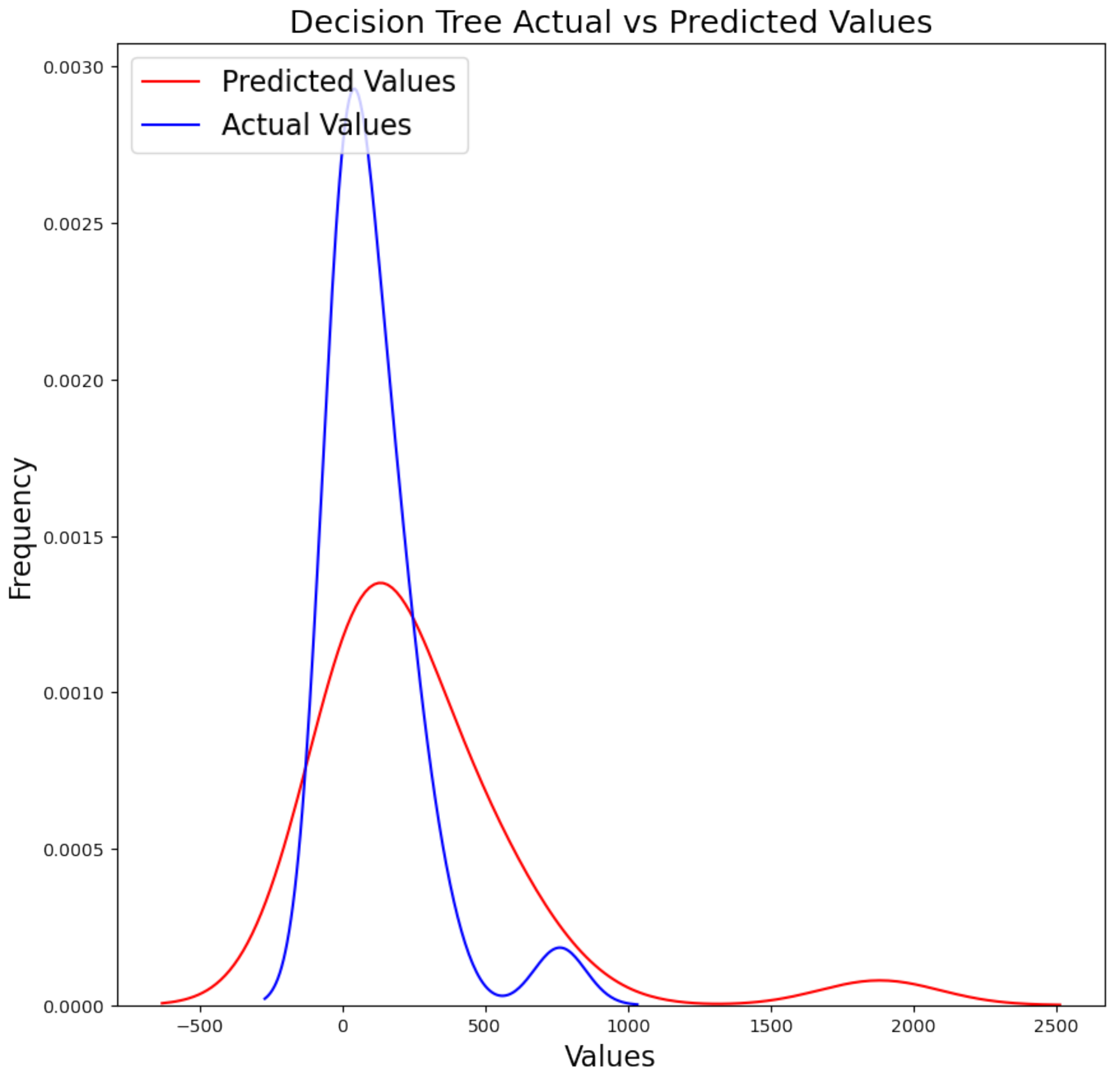
[84]:
```
y_pred.shape
```

[84]: (24,)

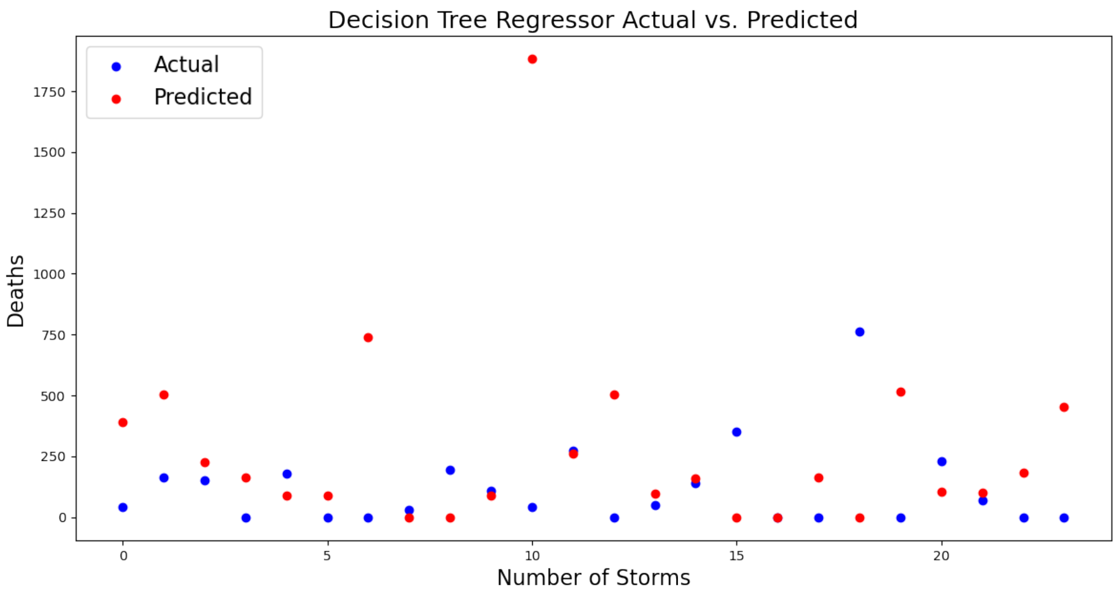
[85]:
```
# Visualising the Decision Tree Regression Results
plt.figure(figsize=(10,10))

sns.distplot(y_pred, hist = False, color = 'r', label = 'Predicted Values')
sns.distplot(y_test, hist = False, color = 'b', label = 'Actual Values')
plt.title('Decision Tree Actual vs Predicted Values', fontsize = 18)
plt.xlabel('Values', fontsize = 16)
plt.ylabel('Frequency', fontsize = 16)
plt.legend(loc = 'upper left', fontsize = 16)
plt.ticklabel_format(style='plain', axis='x')
```

```
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
```

```
warnings.warn(msg, FutureWarning)
```



[93]:
```
plt.figure(figsize=(14,7))
plt.scatter(range(len(y_test)), y_test, color='blue', label='Actual')
plt.scatter(range(len(y_pred)), y_pred, color='red', label='Predicted')
plt.title('Decision Tree Regressor Actual vs. Predicted', fontsize = 18)
plt.xlabel('Number of Storms', fontsize = 16)
plt.ylabel('Deaths', fontsize = 16)
plt.legend(loc = 'upper left', fontsize = 16)
plt.ticklabel_format(style='plain', axis='x')
plt.show()
```



[87]:
```
# calculate rmse and r2
#rmse = np.sqrt(mean_squared_error(y_test,y_pred))
#r2 = r2_score(y_test,y_pred)

corr_matrix = np.corrcoef(y_test, y_pred)
corr = corr_matrix[0,1]
R_sq = corr**2
```

[88]:
```
R_sq
```

[88]: 0.07097932327541905

[89]:
```
# Calculation of Mean Squared Error (MSE)
mean_squared_error(y_test,y_pred)
```

[89]: 240579.58333333334

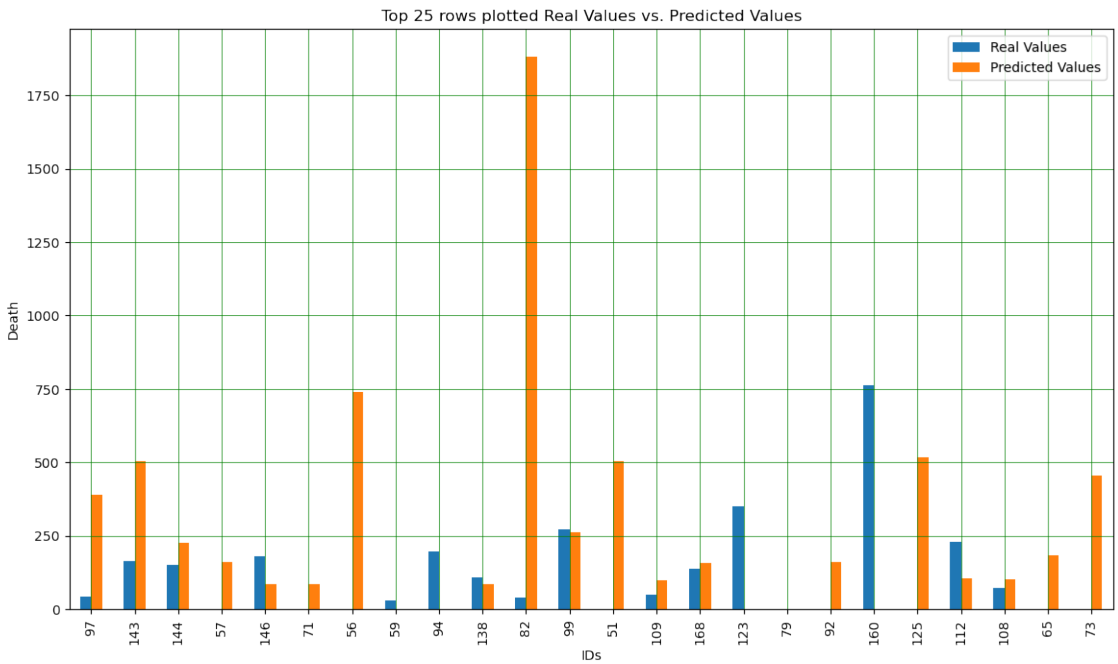
[90]:
```
# rmse metric
math.sqrt(mean_squared_error(y_test,y_pred))
```

[90]: 490.4891266209001

[91]:
```
# mae
mean_absolute_error(y_test,y_pred)
```

[91]: 295.8333333333333

[92]:
```
df = df2.head(25)
df.plot(kind='bar',figsize=(14,8))
plt.grid(which='major', linestyle='-', linewidth='0.5', color='green')
plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black')
plt.title('Top 25 rows plotted Real Values vs. Predicted Values')
plt.xlabel('IDs')
plt.ylabel('Death')

plt.show()
```



0.1.3 Decision Tree using death_bin

[56]:
```
y_pred_death_bin = dtr.predict(X_test)
```

[57]:
```
df_death_bin = pd.DataFrame({'Real Values':y_test, 'Predicted Values':
↪y_pred_death_bin})
```

[58]:
```
df_death_bin.head(25)
```

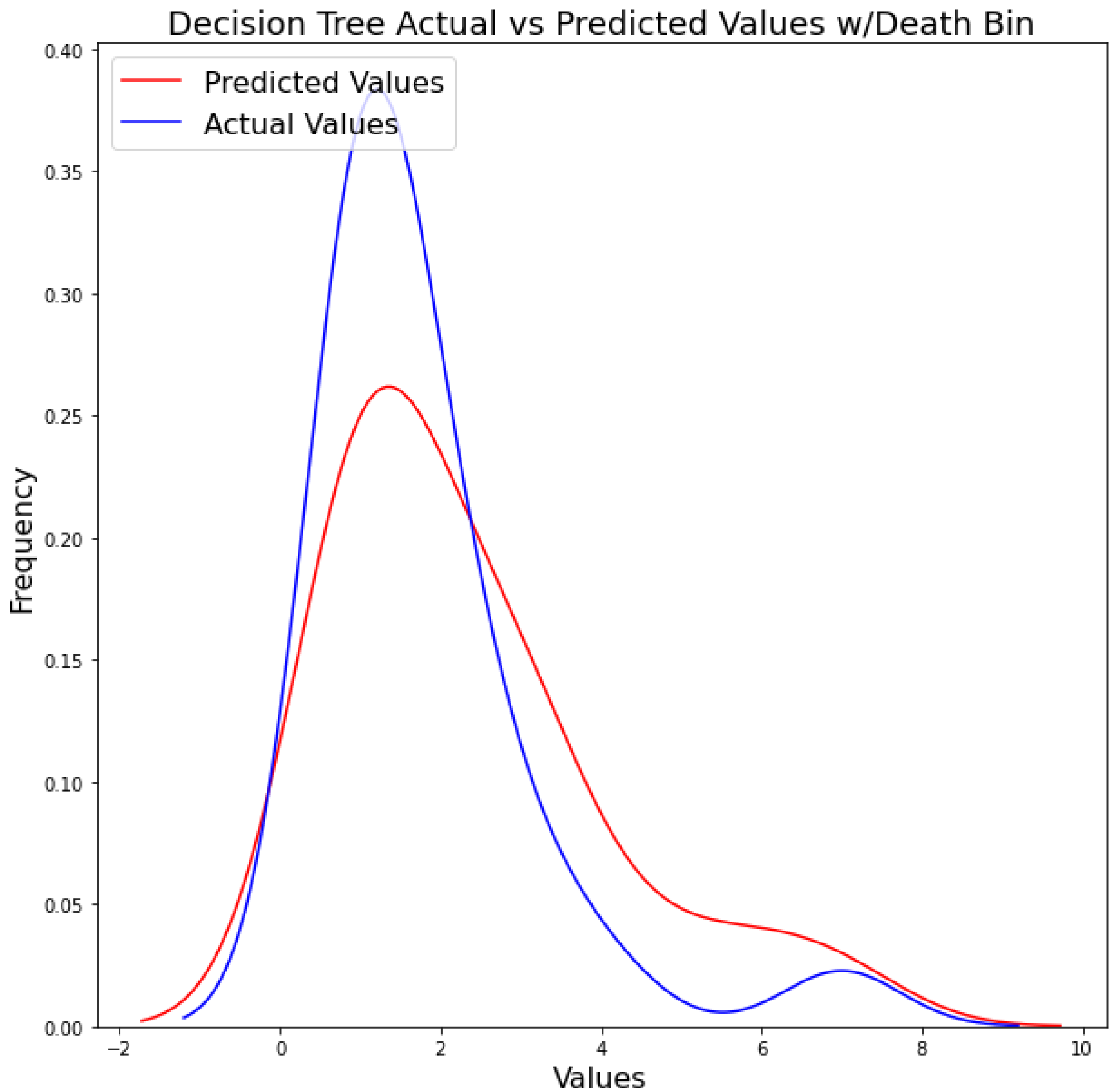
[58]:

	Real Values	Predicted Values
97	1	4.0
143	2	3.0
144	2	3.0
57	1	1.0
146	2	1.0

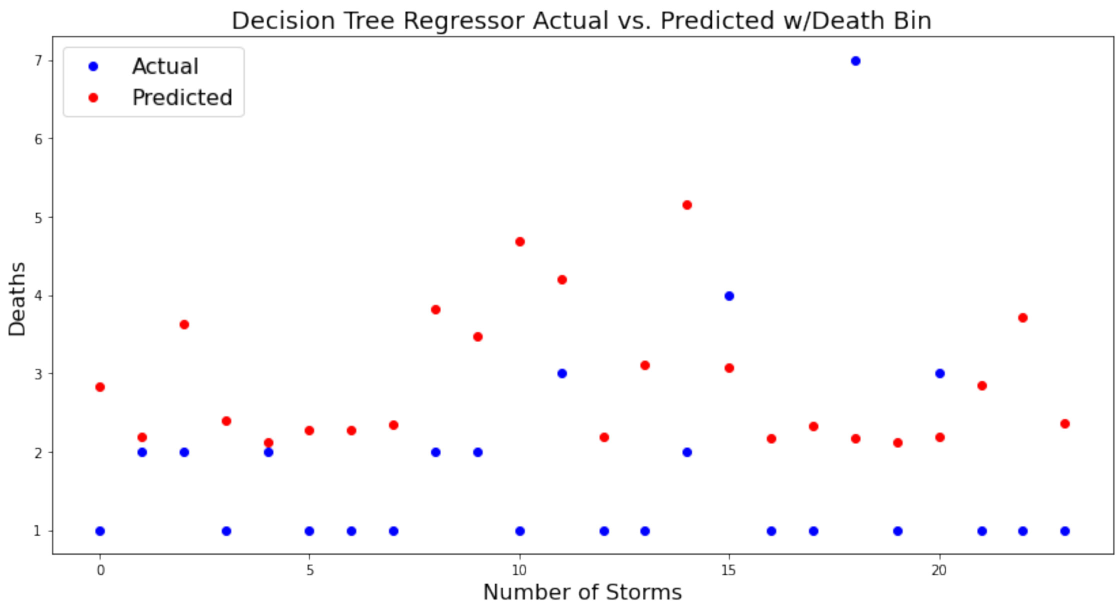
71	1	1.0
56	1	7.0
59	1	1.0
94	2	1.0
138	2	2.0
82	1	3.0
99	3	3.0
51	1	3.0
109	1	1.0
168	2	2.0
123	4	2.0
79	1	1.0
92	1	1.0
160	7	1.0
125	1	6.0
112	3	1.0
108	1	1.0
65	1	2.0
73	1	5.0

[59]:
```python
# Visualising the Decision Tree Regression Results
plt.figure(figsize=(10,10))

sns.distplot(y_pred_death_bin, hist = False, color = 'r', label = 'Predicted␣
↪Values')
sns.distplot(y_test, hist = False, color = 'b', label = 'Actual Values')
plt.title('Decision Tree Actual vs Predicted Values w/Death Bin', fontsize = 18)
plt.xlabel('Values', fontsize = 16)
plt.ylabel('Frequency', fontsize = 16)
plt.legend(loc = 'upper left', fontsize = 16)
plt.ticklabel_format(style='plain', axis='x')
```

```
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
```



[60]:
```
plt.figure(figsize=(14,7))
plt.scatter(range(len(y_test)), y_test, color='blue', label='Actual')
plt.scatter(range(len(y_pred)), y_pred, color='red', label='Predicted')
plt.title('Decision Tree Regressor Actual vs. Predicted w/Death Bin', fontsize␣
 ↪= 18)
plt.xlabel('Number of Storms', fontsize = 16)
plt.ylabel('Deaths', fontsize = 16)
plt.legend(loc = 'upper left', fontsize = 16)
plt.ticklabel_format(style='plain', axis='x')
plt.show()
```



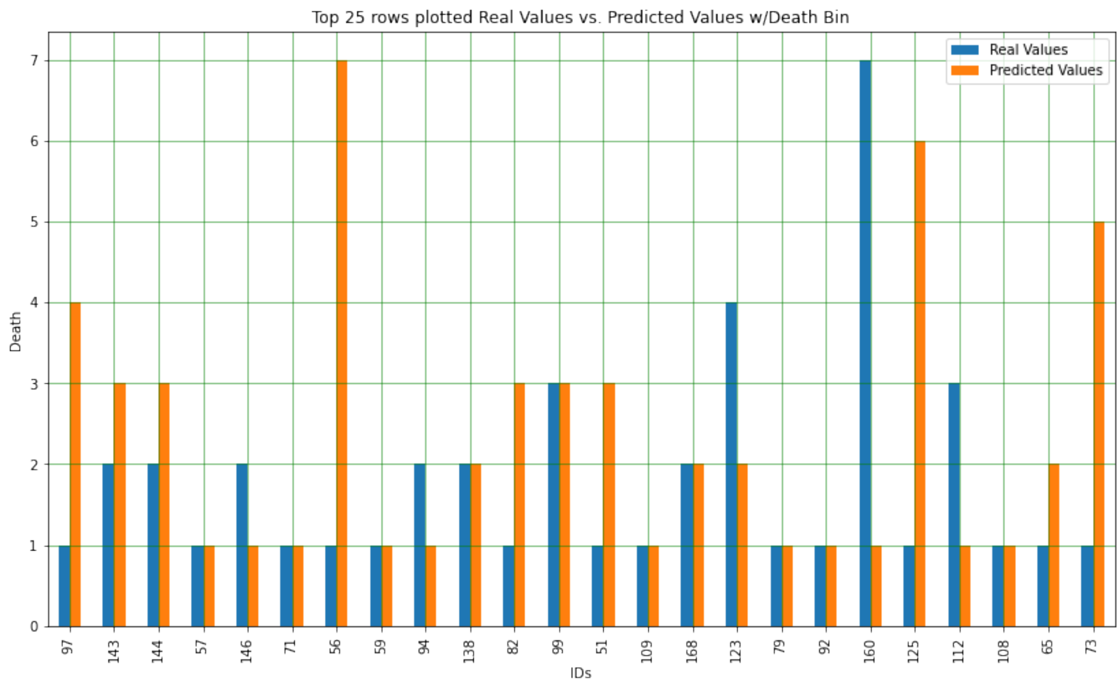
[61]:
```
# calculate rmse and r2
#rmse = np.sqrt(mean_squared_error(y_test,y_pred))
#r2 = r2_score(y_test,y_pred)

corr_matrix = np.corrcoef(y_test, y_pred)
corr = corr_matrix[0,1]
R_sq = corr**2
```

[62]:
```
R_sq
```

[62]: 4.10226712746072e-05

R_sq is 0.0000410226712746072

[63]:
```
df = df_death_bin.head(25)
df.plot(kind='bar',figsize=(14,8))
plt.grid(which='major', linestyle='-', linewidth='0.5', color='green')
plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black')
plt.title('Top 25 rows plotted Real Values vs. Predicted Values w/Death Bin')
plt.xlabel('IDs')
plt.ylabel('Death')

plt.show()
```



OSL with Death Bin

[64]:
```
#add the column of ones to the inputs to calculate the intercept
X = sm.add_constant(X)
```

[65]:
```
#create regression model based on ordinary least squares
model = sm.OLS(y, X)
```

[66]:
```
#variable results
results = model.fit()

print(results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              death_bin   R-squared:                       0.095
Model:                            OLS   Adj. R-squared:                  0.056
Method:                 Least Squares   F-statistic:                     2.401
Date:                Thu, 30 Dec 2021   Prob (F-statistic):             0.0413
Time:                        21:54:23   Log-Likelihood:                -247.42
No. Observations:                 120   AIC:                             506.8
Df Residuals:                     114   BIC:                             523.6
Df Model:                           5
Covariance Type:            nonrobust
================================================================================
==================================
```

```
                                                     coef    std err          t
P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
----------------------------------
const                                              2.1189      0.429      4.942
0.000       1.269       2.968
no_us_hurricanes_HUDRAT_NOAA                      -0.0638      0.181     -0.353
0.725      -0.422       0.295
no_major_us_hurricanes_HUDRAT_ NOAA                0.6875      0.350      1.962
0.052      -0.007       1.382
no_major_north_atlantic_hurricanes_HUDRAT_NOAA    -0.0305      0.226     -0.135
0.893      -0.478       0.417
no_noth_atlantic_hurricanes_HUDRAT_NOAA           -0.0992      0.137     -0.726
0.469      -0.370       0.172
accumulated_cyclone_energy_ACE_HUDRAT_NOAA         0.0081      0.008      0.990
0.324      -0.008       0.024
==============================================================================
Omnibus:                       45.914   Durbin-Watson:                   2.113
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               91.062
Skew:                           1.664   Prob(JB):                     1.68e-20
Kurtosis:                       5.672   Cond. No.                         265.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

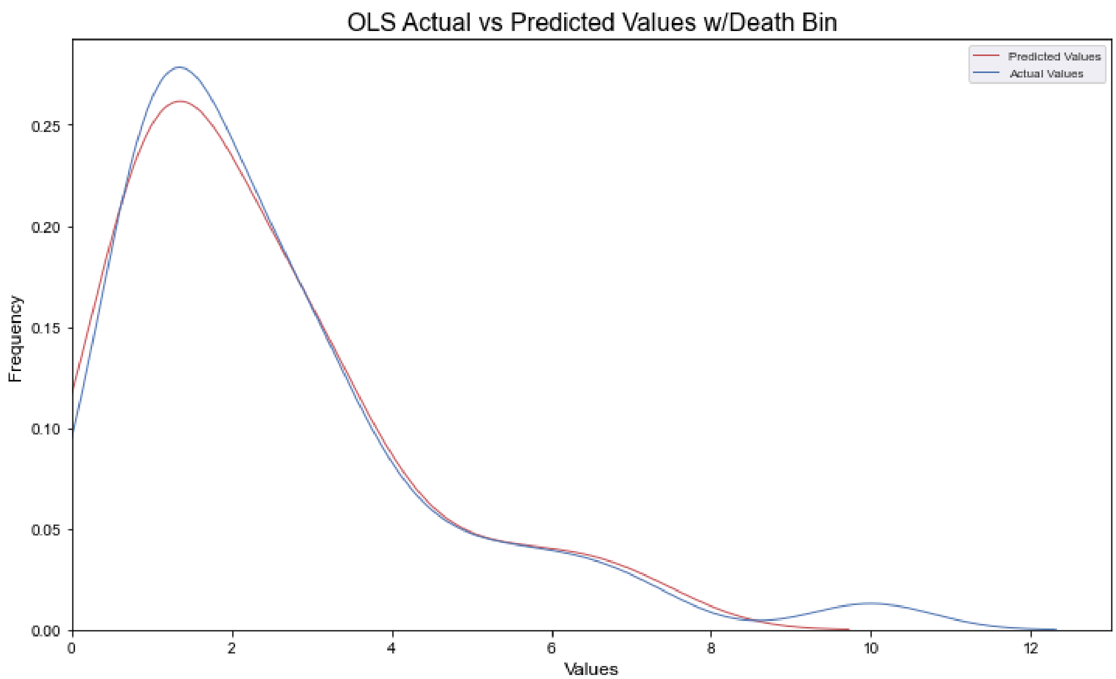
[69]:

```
sns.distplot(y_pred_death_bin, hist = False, color = 'r', label = 'Predicted␣
 ↪Values',  kde_kws=dict(linewidth=1))
sns.distplot(y, hist = False, color = 'b', label = 'Actual Values', ␣
 ↪kde_kws=dict(linewidth=1))
sns.set(rc = {'figure.figsize':(15,15)})
plt.title('OLS Actual vs Predicted Values w/Death Bin', fontsize = 16)
plt.xlabel('Values', fontsize = 12)
plt.ylabel('Frequency', fontsize = 12)
plt.legend(loc = 'upper right', fontsize = 8)
plt.ticklabel_format(style='plain', axis='x')
plt.xlim(0)
```

```
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
```

```
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
```

[69]: (0.0, 13.013805010457247)



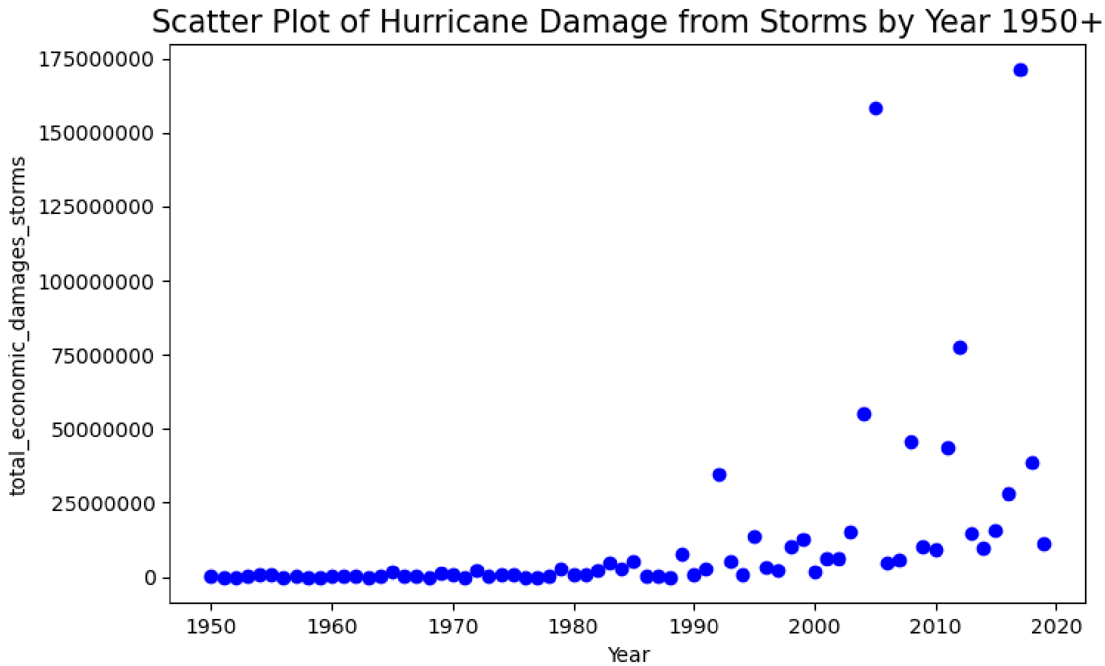
0.2 ****************************

Model with 1950 data including ecomonic damage

[27]:
```
# Filter data down to US only
hurricanes_us_df_1950 = hurricanes_us_df[hurricanes_us_df['Year'] >= 1950]
```

[32]:
```
# For graph
y = hurricanes_us_df_1950['total_economic_damages_storms']
x = hurricanes_us_df_1950['Year']

# Plot
plt.scatter(x,y,color='blue')
plt.rcParams.update({'figure.figsize':(8,5), 'figure.dpi':100})

#Labels
plt.title('Scatter Plot of Hurricane Damage from Storms by Year␣
 ↪1950+',fontsize=15)
plt.ylabel('total_economic_damages_storms',fontsize=10)
plt.xlabel('Year',fontsize=10)
```

```
plt.ticklabel_format(axis='x', style='plain')
plt.ticklabel_format(axis='y', style='plain')
plt.show()
```



[19]:
```
hurricanes_us_df_1950.columns
```

[19]:
```
Index(['Entity', 'Year', 'no_us_hurricanes_HUDRAT_NOAA',
       'no_major_us_hurricanes_HUDRAT_ NOAA',
       'no_major_north_atlantic_hurricanes_HUDRAT_NOAA',
       'no_noth_atlantic_hurricanes_HUDRAT_NOAA',
       'accumulated_cyclone_energy_ACE_HUDRAT_NOAA',
       'cyclone_power_dissipation_index_PDI_HUDRAT_NOAA',
       'hurricane_fatality_rate', 'ACE', 'deaths_hurricanes_us',
       'total_economic_damages_storms'],
      dtype='object')
```

[30]:
```
hurricanes_us_df_1950['no_us_hurricanes_HUDRAT_NOAA'] =␣
↪hurricanes_us_df_1950['no_us_hurricanes_HUDRAT_NOAA'].astype(int)
hurricanes_us_df_1950['no_major_us_hurricanes_HUDRAT_ NOAA'] =␣
↪hurricanes_us_df_1950['no_major_us_hurricanes_HUDRAT_ NOAA'].astype(int)
hurricanes_us_df_1950['no_major_north_atlantic_hurricanes_HUDRAT_NOAA'] =␣
↪hurricanes_us_df_1950['no_major_north_atlantic_hurricanes_HUDRAT_NOAA'].
↪astype(int)
hurricanes_us_df_1950['no_noth_atlantic_hurricanes_HUDRAT_NOAA'] =␣
↪hurricanes_us_df_1950['no_noth_atlantic_hurricanes_HUDRAT_NOAA'].astype(int)
```

```python
hurricanes_us_df_1950['accumulated_cyclone_energy_ACE_HUDRAT_NOAA'] =␣
↪hurricanes_us_df_1950['accumulated_cyclone_energy_ACE_HUDRAT_NOAA'].
↪astype(int)
hurricanes_us_df_1950['Year'] = hurricanes_us_df_1950['Year'].astype(int)
hurricanes_us_df_1950['cyclone_power_dissipation_index_PDI_HUDRAT_NOAA'] =␣
↪hurricanes_us_df_1950['cyclone_power_dissipation_index_PDI_HUDRAT_NOAA']
```

```
<ipython-input-30-e6864baa302a>:1: 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
  hurricanes_us_df_1950['no_us_hurricanes_HUDRAT_NOAA'] =
hurricanes_us_df_1950['no_us_hurricanes_HUDRAT_NOAA'].astype(int)
<ipython-input-30-e6864baa302a>:2: 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
  hurricanes_us_df_1950['no_major_us_hurricanes_HUDRAT_ NOAA'] =
hurricanes_us_df_1950['no_major_us_hurricanes_HUDRAT_ NOAA'].astype(int)
<ipython-input-30-e6864baa302a>: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
  hurricanes_us_df_1950['no_major_north_atlantic_hurricanes_HUDRAT_NOAA'] = hurr
icanes_us_df_1950['no_major_north_atlantic_hurricanes_HUDRAT_NOAA'].astype(int)
<ipython-input-30-e6864baa302a>:4: 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
  hurricanes_us_df_1950['no_noth_atlantic_hurricanes_HUDRAT_NOAA'] =
hurricanes_us_df_1950['no_noth_atlantic_hurricanes_HUDRAT_NOAA'].astype(int)
<ipython-input-30-e6864baa302a>:5: 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
  hurricanes_us_df_1950['accumulated_cyclone_energy_ACE_HUDRAT_NOAA'] =
hurricanes_us_df_1950['accumulated_cyclone_energy_ACE_HUDRAT_NOAA'].astype(int)
<ipython-input-30-e6864baa302a>:6: 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
  hurricanes_us_df_1950['Year'] = hurricanes_us_df_1950['Year'].astype(int)
<ipython-input-30-e6864baa302a>:7: 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
  hurricanes_us_df_1950['cyclone_power_dissipation_index_PDI_HUDRAT_NOAA'] =
hurricanes_us_df_1950['cyclone_power_dissipation_index_PDI_HUDRAT_NOAA']
```

[31]:

```
# split out X and y
X = hurricanes_us_df_1950[['no_us_hurricanes_HUDRAT_NOAA',
       'no_major_us_hurricanes_HUDRAT_ NOAA',
       'no_major_north_atlantic_hurricanes_HUDRAT_NOAA',
       'no_noth_atlantic_hurricanes_HUDRAT_NOAA',
       'accumulated_cyclone_energy_ACE_HUDRAT_NOAA',
                           'deaths_hurricanes_us']]

y = hurricanes_us_df_1950['total_economic_damages_storms']
```

[32]:

```
# split the data in training and Test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=0)
```

[39]:

```
# using decision tree model
from sklearn.tree import DecisionTreeRegressor
dtr = DecisionTreeRegressor()
dtr.fit(X_train, y_train)
```

[39]:

```
DecisionTreeRegressor()
```

[40]:

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

[41]:

```
df2 = pd.DataFrame({'Real Values':y_test, 'Predicted Values':y_pred})
```

[42]:

```
df2.head(25)
```

[42]:

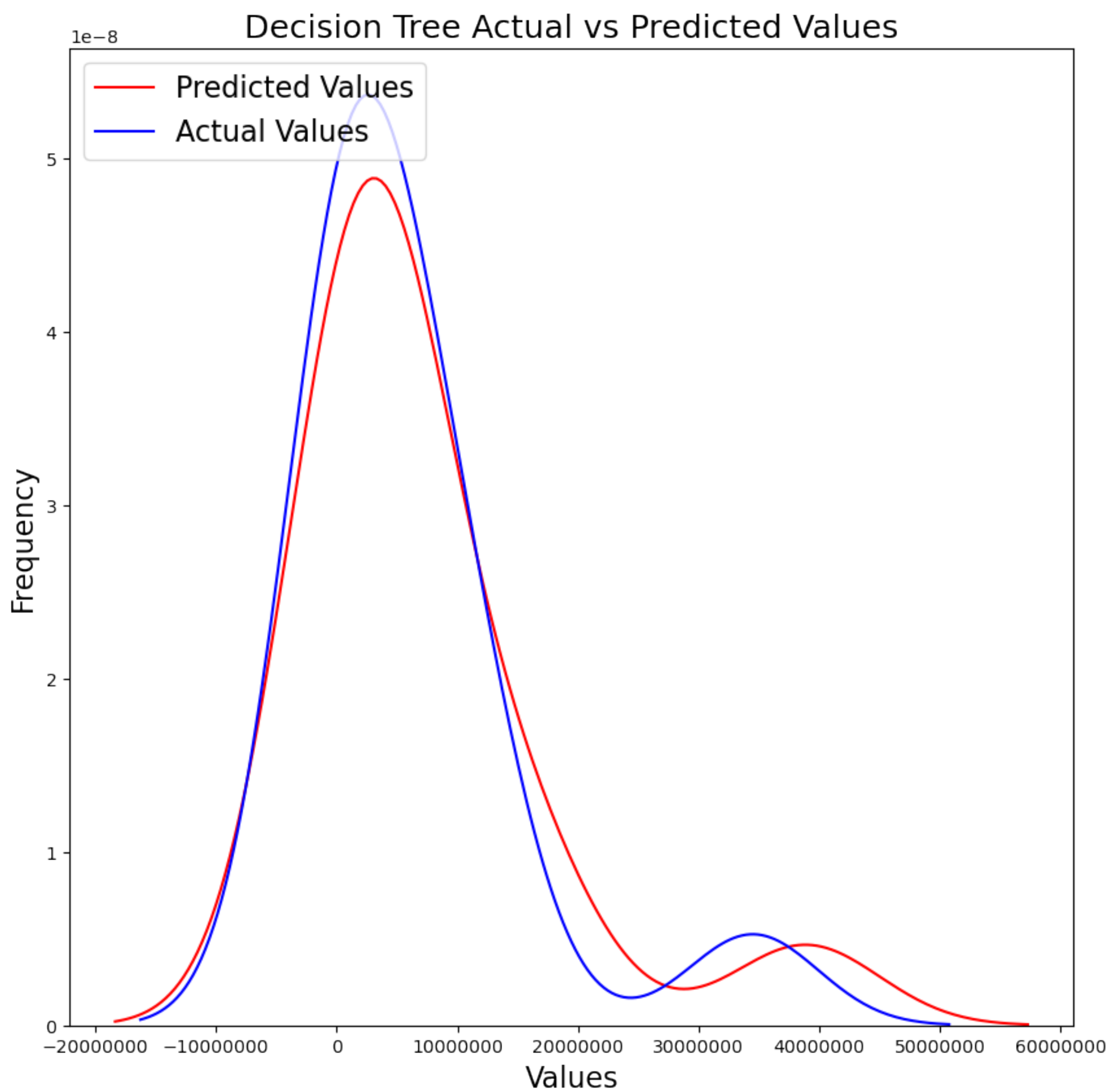
	Real Values	Predicted Values
125	0.0	50000.0
126	0.0	4655000.0
147	10053450.0	15148400.0
121	2100000.0	5125000.0
129	910000.0	7880000.0

```
150    6328800.0         1865600.0
106     150000.0         1000000.0
158   10390000.0           50000.0
133    2545600.0         1610000.0
168   11325000.0        38875000.0
155    4727860.0         2560000.0
127     100000.0        15720000.0
130     861000.0         1865600.0
141   34500000.0         4655000.0
```

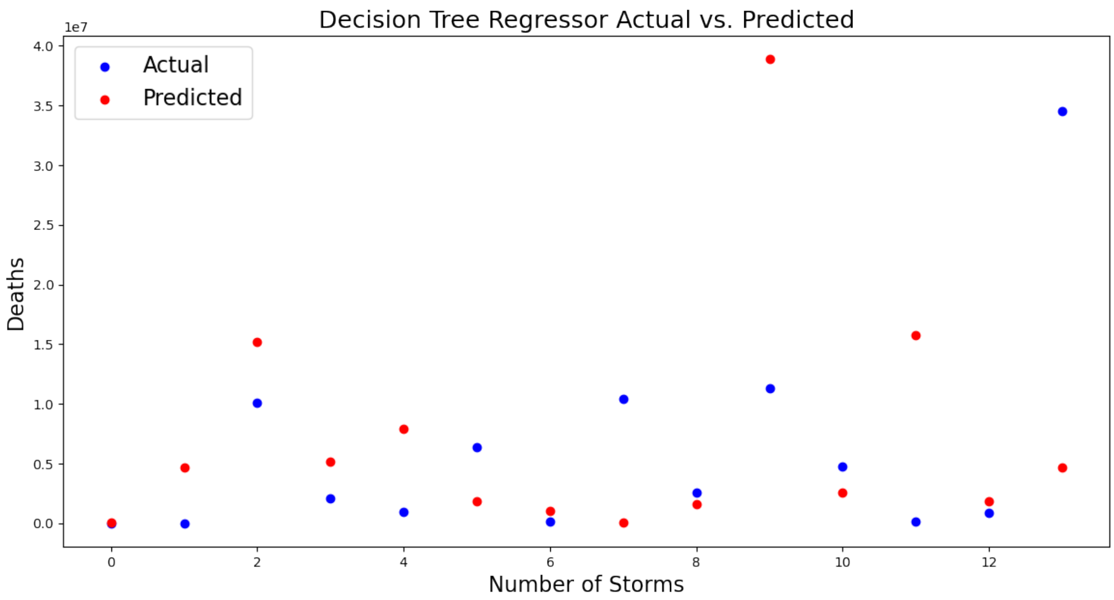
[43]:

```
# Visualising the Decision Tree Regression Results
plt.figure(figsize=(10,10))

sns.distplot(y_pred, hist = False, color = 'r', label = 'Predicted Values')
sns.distplot(y_test, hist = False, color = 'b', label = 'Actual Values')
plt.title('Decision Tree Actual vs Predicted Values', fontsize = 18)
plt.xlabel('Values', fontsize = 16)
plt.ylabel('Frequency', fontsize = 16)
plt.legend(loc = 'upper left', fontsize = 16)
plt.ticklabel_format(style='plain', axis='x')
```

```
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
```




```
[44]: plt.figure(figsize=(14,7))
      plt.scatter(range(len(y_test)), y_test, color='blue', label='Actual')
      plt.scatter(range(len(y_pred)), y_pred, color='red', label='Predicted')
      plt.title('Decision Tree Regressor Actual vs. Predicted', fontsize = 18)
      plt.xlabel('Number of Storms', fontsize = 16)
      plt.ylabel('Deaths', fontsize = 16)
      plt.legend(loc = 'upper left', fontsize = 16)
      plt.ticklabel_format(style='plain', axis='x')
      plt.show()
```

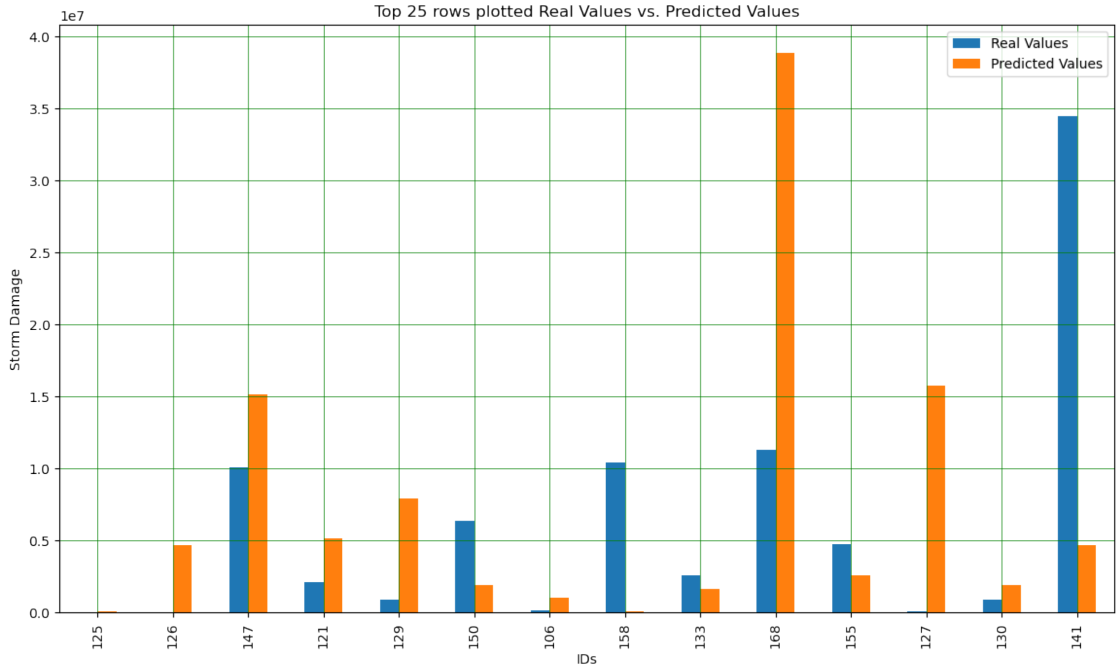



```
[45]: # calculate rmse and r2
      #rmse = np.sqrt(mean_squared_error(y_test,y_pred))
      #r2 = r2_score(y_test,y_pred)

      corr_matrix = np.corrcoef(y_test, y_pred)
      corr = corr_matrix[0,1]
      R_sq = corr**2
```

```
[46]: R_sq
```

```
[46]: 0.02501494202440917
```

```
[48]: df = df2.head(25)
      df.plot(kind='bar',figsize=(14,8))
      plt.grid(which='major', linestyle='-', linewidth='0.5', color='green')
      plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black')
      plt.title('Top 25 rows plotted Real Values vs. Predicted Values')
      plt.xlabel('IDs')
      plt.ylabel('Storm Damage')

      plt.show()
```



OLS Model Ordinary Least Squares

[12]:
```
hurricanes_us_df_1950.columns
```

[12]:
```
Index(['Entity', 'Year', 'no_us_hurricanes_HUDRAT_NOAA',
       'no_major_us_hurricanes_HUDRAT_ NOAA',
       'no_major_north_atlantic_hurricanes_HUDRAT_NOAA',
       'no_noth_atlantic_hurricanes_HUDRAT_NOAA',
       'accumulated_cyclone_energy_ACE_HUDRAT_NOAA',
       'cyclone_power_dissipation_index_PDI_HUDRAT_NOAA',
       'hurricane_fatality_rate', 'ACE', 'deaths_hurricanes_us',
       'total_economic_damages_storms'],
      dtype='object')
```

[33]:
```
X = hurricanes_us_df_1950[['no_us_hurricanes_HUDRAT_NOAA',
       'no_major_us_hurricanes_HUDRAT_ NOAA',
       'no_major_north_atlantic_hurricanes_HUDRAT_NOAA',
       'no_noth_atlantic_hurricanes_HUDRAT_NOAA',
       'accumulated_cyclone_energy_ACE_HUDRAT_NOAA',
                           'deaths_hurricanes_us']]

y = hurricanes_us_df_1950['total_economic_damages_storms']
```

[34]:
```
hurricanes_us_df_1950
```

[34]:
```
             Entity  Year  no_us_hurricanes_HUDRAT_NOAA  \
99   North Atlantic  1950                             3
100  North Atlantic  1951                             0
101  North Atlantic  1952                             1
102  North Atlantic  1953                             3
103  North Atlantic  1954                             3
..              …     …                           …
164  North Atlantic  2015                             0
165  North Atlantic  2016                             2
166  North Atlantic  2017                             3
167  North Atlantic  2018                             8
168  North Atlantic  2019                             6

     no_major_us_hurricanes_HUDRAT_ NOAA  \
99                                     2
100                                    0
101                                    0
102                                    0
103                                    2
..                                  …
164                                    0
165                                    0
166                                    2
167                                    2
168                                    3

     no_major_north_atlantic_hurricanes_HUDRAT_NOAA  \
99                                                6
100                                               3
101                                               2
102                                               3
103                                               3
..                                             …
164                                               2
165                                               4
166                                               6
167                                               2
168                                               2

     no_noth_atlantic_hurricanes_HUDRAT_NOAA  \
99                                        11
100                                        8
101                                        5
102                                        7
103                                        7
..                                      …
164                                        4
```

```
165                                      7
166                                     10
167                                      8
168                                      8

     accumulated_cyclone_energy_ACE_HUDRAT_NOAA  \
99                                          211
100                                         126
101                                          69
102                                          99
103                                         111
..                                          …
164                                          63
165                                         141
166                                         223
167                                         132
168                                         132

     cyclone_power_dissipation_index_PDI_HUDRAT_NOAA  hurricane_fatality_rate  \
99                                            0.0000                 0.124777
100                                           2.7846                 0.000000
101                                           2.3445                 0.019041
102                                           2.2639                 0.012486
103                                           2.4730                 1.183861
..                                               …                      …
164                                              NaN                 0.043557
165                                              NaN                      NaN
166                                              NaN                      NaN
167                                              NaN                      NaN
168                                              NaN                      NaN

          ACE  deaths_hurricanes_us  total_economic_damages_storms  death_bin
99   211.2825                 272.0                       100000.0        3.0
100  126.3250                   0.0                            0.0        1.0
101   69.0800                 279.0                            0.0        3.0
102   98.5075                 468.0                        52000.0        5.0
103  110.8800                 223.0                       731000.0        3.0
..        …                     …                             …        …
164   62.6850                 218.0                     15720000.0        3.0
165  141.2525                 199.0                     28050000.0        2.0
166  224.8775                 261.0                    171110000.0        3.0
167  132.5825                 158.0                     38875000.0        2.0
168  132.2025                 139.0                     11325000.0        2.0

[70 rows x 13 columns]
```

[35]:
```
#add the column of ones to the inputs to calculate the intercept
X = sm.add_constant(X)
```

[36]:
```
#create regression model based on ordinary least squares
model = sm.OLS(y, X)
```

[37]:
```
#variable results
results = model.fit()

print(results.summary())
```

```
                                        OLS Regression Results
================================================================================
=========
Dep. Variable:     total_economic_damages_storms   R-squared:
0.410
Model:                                       OLS   Adj. R-squared:
0.354
Method:                            Least Squares   F-statistic:
7.302
Date:                           Mon, 03 Jan 2022   Prob (F-statistic):
6.00e-06
Time:                                   22:14:19   Log-Likelihood:
-1285.6
No. Observations:                             70   AIC:
2585.
Df Residuals:                                 63   BIC:
2601.
Df Model:                                      6
Covariance Type:                       nonrobust
================================================================================
==================================
                                                      coef    std err          t
P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
----------------------------------
const                                           -1.727e+07   7.91e+06     -2.182
0.033   -3.31e+07   -1.46e+06
no_us_hurricanes_HUDRAT_NOAA                     6.438e+05   2.66e+06      0.242
0.810   -4.67e+06    5.96e+06
no_major_us_hurricanes_HUDRAT_ NOAA              5.942e+06   5.36e+06      1.108
0.272   -4.77e+06    1.67e+07
no_major_north_atlantic_hurricanes_HUDRAT_NOAA  2.477e+05   3.48e+06      0.071
0.943    -6.7e+06    7.19e+06
no_noth_atlantic_hurricanes_HUDRAT_NOAA          1.432e+05   2.39e+06      0.060
0.952   -4.63e+06    4.91e+06
accumulated_cyclone_energy_ACE_HUDRAT_NOAA      1.368e+05   1.36e+05      1.009
0.317   -1.34e+05    4.08e+05
```

```
deaths_hurricanes_us                              4.017e+04    1.31e+04      3.075
0.003     1.41e+04     6.63e+04
==============================================================================
Omnibus:                       69.945   Durbin-Watson:                   1.411
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              640.020
Skew:                           2.817   Prob(JB):                    1.05e-139
Kurtosis:                      16.700   Cond. No.                         988.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[39]:
```
ypred =  results.predict(X)
print('predicted response:',ypred, sep='\n')
```

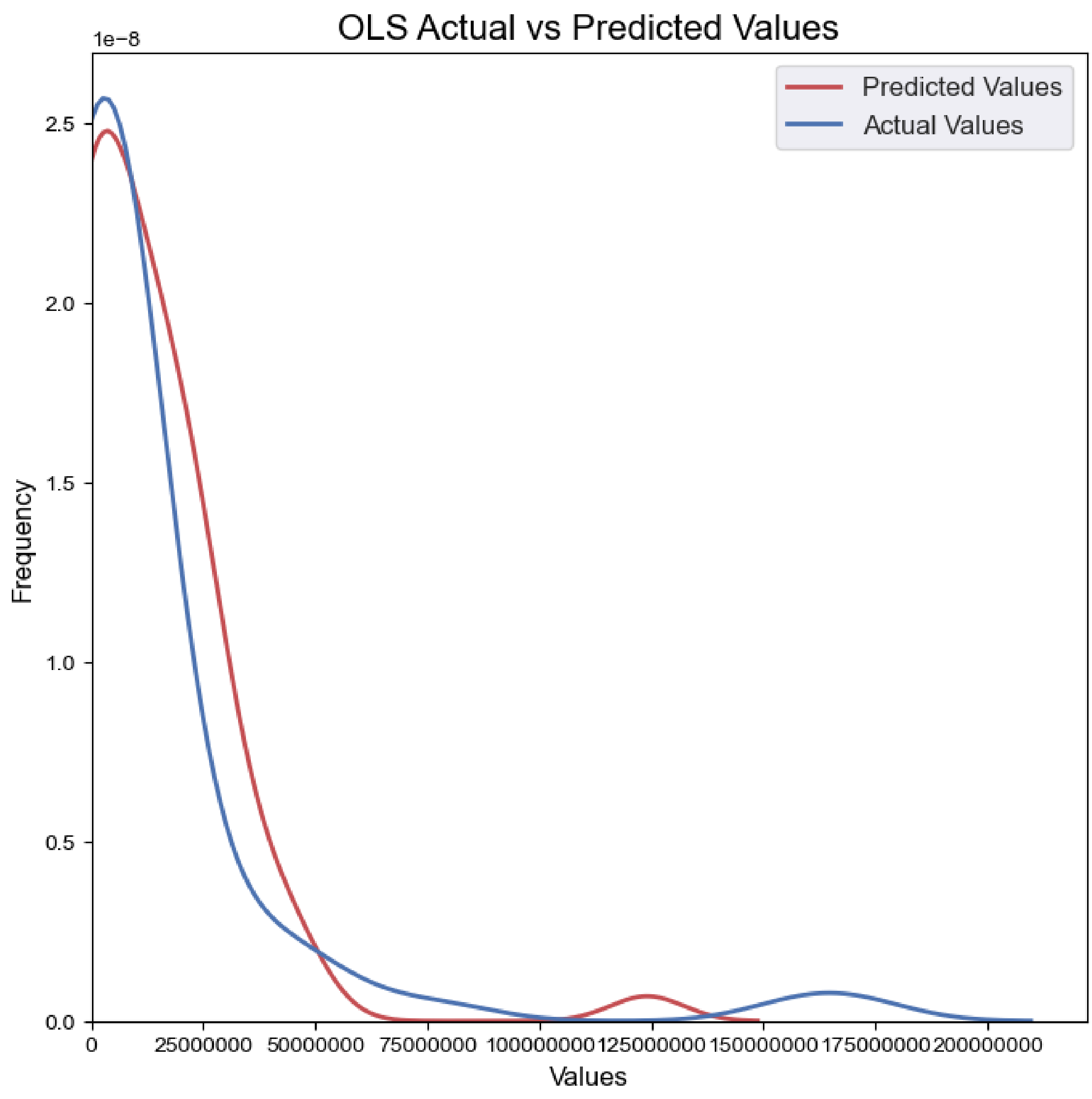
```
predicted response:
99     3.939034e+07
100    1.852374e+06
101    5.229120e+06
102    1.874506e+07
103    2.242950e+07
           …
164    1.171444e+06
165    1.328883e+07
166    4.044656e+07
167    2.580538e+07
168    2.969652e+07
Length: 70, dtype: float64
```

[40]:
```
sns.distplot(ypred, hist = False, color = 'r', label = 'Predicted Values', ␣
 ↪kde_kws=dict(linewidth=2))
sns.distplot(y, hist = False, color = 'b', label = 'Actual Values', ␣
 ↪kde_kws=dict(linewidth=2))
sns.set(rc = {'figure.figsize':(15,15)})
plt.title('OLS Actual vs Predicted Values', fontsize = 16)
plt.xlabel('Values', fontsize = 12)
plt.ylabel('Frequency', fontsize = 12)
plt.legend(loc = 'upper right', fontsize = 12)
plt.ticklabel_format(style='plain', axis='x')
plt.xlim(0)
```

```
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
```

```
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
```

[40]: (0.0, 222070229.6450136)



[44]:
```
ols_df = pd.DataFrame({'Real Values':y, 'Predicted Values':ypred})
```

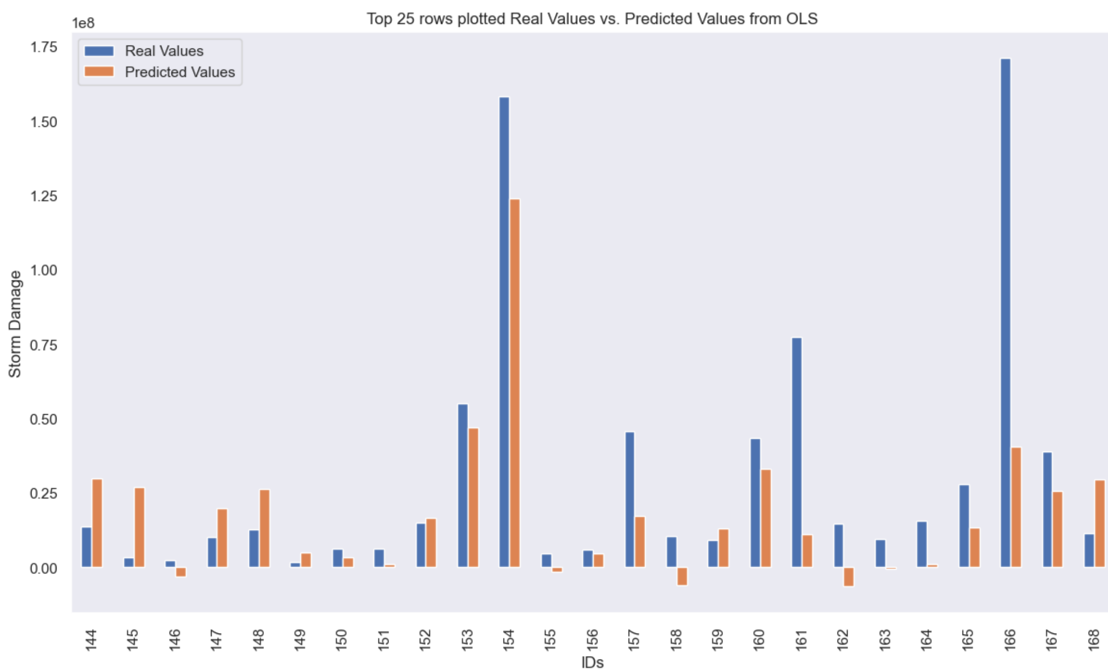
[45]:
```
ols_df
```

[45]:

```
       Real Values  Predicted Values
99        100000.0      3.939034e+07
100            0.0      1.852374e+06
101            0.0      5.229120e+06
102        52000.0      1.874506e+07
103       731000.0      2.242950e+07
..             ...               ...
164     15720000.0      1.171444e+06
165     28050000.0      1.328883e+07
166    171110000.0      4.044656e+07
167     38875000.0      2.580538e+07
168     11325000.0      2.969652e+07

[70 rows x 2 columns]
```

[48]:

```python
ols = ols_df.tail(25)
ols.plot(kind='bar',figsize=(14,8))
plt.grid(which='major', linestyle='-', linewidth='0.5', color='green')
plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black')
plt.title('25 rows plotted Real Values vs. Predicted Values from OLS')
plt.xlabel('IDs')
plt.ylabel('Storm Damage')

plt.show()
```



Deaths by OLS

[49]:
```
# split out X and y
X = hurricanes_us_df_1950[['no_us_hurricanes_HUDRAT_NOAA',
       'no_major_us_hurricanes_HUDRAT_ NOAA',
       'no_major_north_atlantic_hurricanes_HUDRAT_NOAA',
       'no_noth_atlantic_hurricanes_HUDRAT_NOAA',
       'accumulated_cyclone_energy_ACE_HUDRAT_NOAA']]

y = hurricanes_us_df_1950['deaths_hurricanes_us']
```

[50]:
```
#add the column of ones to the inputs to calculate the intercept
X = sm.add_constant(X)
```

[51]:
```
#create regression model based on ordinary least squares
model = sm.OLS(y, X)
```

[52]:
```
#variable results
results = model.fit()

print(results.summary())
```

```
                            OLS Regression Results
================================================================================
Dep. Variable:     deaths_hurricanes_us   R-squared:                       0.224
Model:                              OLS   Adj. R-squared:                  0.163
Method:                   Least Squares   F-statistic:                     3.695
Date:                  Mon, 03 Jan 2022   Prob (F-statistic):            0.00532
Time:                          22:30:56   Log-Likelihood:                -477.21
No. Observations:                    70   AIC:                             966.4
Df Residuals:                        64   BIC:                             979.9
Df Model:                             5
Covariance Type:              nonrobust
================================================================================
==================================
                                                    coef    std err          t
P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
----------------------------------
const                                            54.4223     75.425      0.722
0.473     -96.256     205.101
no_us_hurricanes_HUDRAT_NOAA                     -6.8552     25.447     -0.269
0.788     -57.691      43.980
no_major_us_hurricanes_HUDRAT_ NOAA             107.4614     49.510      2.171
0.034       8.554     206.369
no_major_north_atlantic_hurricanes_HUDRAT_NOAA   25.7825     33.102      0.779
0.439     -40.346      91.911
no_noth_atlantic_hurricanes_HUDRAT_NOAA          23.6061     22.647      1.042
0.301     -21.636      68.848
```

```
accumulated_cyclone_energy_ACE_HUDRAT_NOAA        -0.8826      1.293     -0.683
0.497      -3.465       1.700
==============================================================================
Omnibus:                       52.265   Durbin-Watson:                   2.239
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              242.618
Skew:                           2.190   Prob(JB):                     2.07e-53
Kurtosis:                      11.000   Cond. No.                         330.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[53]:

```
ypred =  results.predict(X)
print('predicted response:',ypred, sep='\n')
```

```
predicted response:
99     476.919593
100    209.415030
101    156.265427
102    189.072613
103    393.404592
          …
164    144.809959
165    184.642379
166    442.722593
167    338.417981
168    459.589881
Length: 70, dtype: float64
```

[54]:

```
sns.distplot(ypred, hist = False, color = 'r', label = 'Predicted Values', ␣
 ↪kde_kws=dict(linewidth=2))
sns.distplot(y, hist = False, color = 'b', label = 'Actual Values', ␣
 ↪kde_kws=dict(linewidth=2))
sns.set(rc = {'figure.figsize':(15,15)})
plt.title('OLS Model Actual vs Predicted Values for Deaths ', fontsize = 16)
plt.xlabel('Values', fontsize = 12)
plt.ylabel('Frequency', fontsize = 12)
plt.legend(loc = 'upper right', fontsize = 12)
plt.ticklabel_format(style='plain', axis='x')
plt.xlim(0)
```

```
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
```

```
C:\Users\Tushar\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `kdeplot` (an axes-level function for
kernel density plots).
  warnings.warn(msg, FutureWarning)
```

[54]: (0.0, 2332.700029085593)