36

3.3 3.3 Project Task 2: Executive Summary

student: Rachel Nelson

class: DSC640-T302 Data Presentation & Visualization (2221-1)

37
```
# Import Packages
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

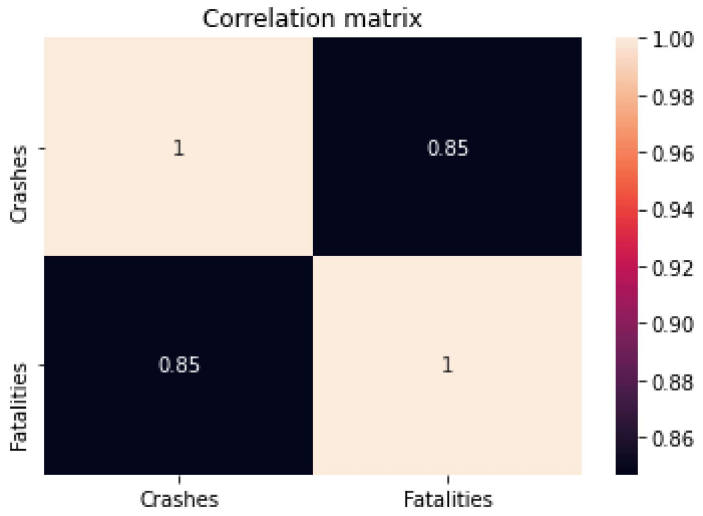
38
```
# Import data into dataframe and get header information
df = pd.read_csv('combined crash and fatality.csv')
```

39
```
df = df.rename(columns={"Nb of Crashes": "Crashes"},errors="raise")
df = df.drop(columns=['Year'])
```

40
```
print(df.head(5))
print(df.describe())
```

```
   Crashes  Fatalities
0       12           4
1       62          33
2       97          54
3       98          42
4       80          69
          Crashes   Fatalities
count  106.000000   106.000000
mean   266.764151  1529.235849
std    146.963070  1016.321167
min     12.000000     4.000000
25%    182.000000   621.250000
50%    262.000000  1609.500000
75%    330.000000  2212.000000
max    944.000000  4700.000000
```

42
```
# Correlation Matrix
# taking all rows but only 6 columns
correlation_mat = df.corr()
sns.heatmap(correlation_mat, annot = True)
plt.title("Correlation matrix")
plt.show()
```


Correlation matrix
Crashes
Fatalities
1
0.85
0.85
1
Crashes
Fatalities
1.00
0.98
0.96
0.94
0.92
0.90
0.88
0.86