



CORRELATION

DESCRIPTIVE STATISTICS

Prepared by:

Gyro A. Madrona

Electronics Engineer

Correlation

Data Analyst: Gyro A. Madrona

Department: Electrical Engineering

```
1 # imports and packages
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 from scipy import stats
```

Python

```
1 # load dataset
2 df = pd.read_csv(r'raw\current-power.csv',
3                 delimiter=",")
4 df
```

Open 'df' in Data Wrangler

Python

```
1 # summary of statistics
2 df.describe()
```

Python

```
1 # covariance of current and power
2 df.cov()
```

Python

```
1 # scatter plot
2 x = df['Current']
3 y = df['Power']
4
5 plt.scatter(x,y)
6 plt.show()
```

Python

```
1 # correlation coefficient of current and power
2 df.corr()
```

Python

Air Quality

```
1 # load dataset
2 air = pd.read_csv(r"raw\air-quality.csv",
3                  delimiter=",")
4 air
```

Open 'air' in Data Wrangler

Python

NOx-Humidity

```
1 # regression line
2 x = air['NOx']
3 y = air['Relative_Humidity']
4
5 # scatter plot
6 plt.scatter(x,y,color='#1C1F24')
7 plt.show()
```

Python