



CORRELATION

DESCRIPTIVE STATISTICS

Prepared by:

Gyro A. Madrona

Electronics Engineer

L14-correlation-exercise.ipynb

L14-correlation-exercise.ipynb > M Air Quality > # Load air quality dataset

Code + Markdown | Run All | Restart | Clear All Outputs | View data | Jupyter Variables | Outline

Python 3.13.1

Correlation

Data Analyst: Gyro A. Madrona

Department: Electrical Engineering

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from scipy import stats
```

Python

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

Python

```
1 # Summary of Statistics
2 df.describe()
```

Python

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

Python

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

Python

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

Python

Air Quality

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

Python

NOx-Humidity

```
1 # Regression line
2 x = air['NOx']
3 y = air['Relative_Humidity']
4 slope, intercept, r_value, p_value, std_err = stats.linregress(x,y)
5 line = slope*x + intercept
6
7 # Scatter plot
8 plt.figure(figsize=(7,5))
9 plt.scatter(x,y)
10 plt.plot(x,line)
11 plt.show()
12
```

Python