

Ishika Prasad

ip1262@rit.edu (<mailto:ip1262@rit.edu>)

Import Libraries

```
In [134]: ▶ import pandas as pd
          from pandas import ExcelWriter
          from pandas import ExcelFile
          from matplotlib import pyplot as plt
```

Read Excel file

```
In [135]: ▶ df = pd.read_excel('powerplant.xlsx', sheet_name='Sheet1')
```

Print the data read from excel file

```
In [136]: ▶ print(df)
```

	Unnamed: 0	Plant 1	Plant 2
0	month 1	18.4	20.1
1	month 2	15.7	21.1
2	month 3	11.9	26.4
3	month 4	23.6	16.5
4	month 5	13.8	28.9
5	month 6	20.9	28.1
6	month 7	14.0	25.7
7	month 8	15.8	30.9
8	month 9	19.6	24.2
9	month 10	13.3	21.5
10	month 11	14.7	22.5
11	month 12	16.0	19.4
12	month 13	14.6	21.5
13	month 14	16.2	23.4

Print the data for Plant 1

In [137]: `print(df['Plant 1'])`

```
0    18.4
1    15.7
2    11.9
3    23.6
4    13.8
5    20.9
6    14.0
7    15.8
8    19.6
9    13.3
10   14.7
11   16.0
12   14.6
13   16.2
Name: Plant 1, dtype: float64
```

Print the data for Plant 2

In [138]: `print(df['Plant 2'])`

```
0    20.1
1    21.1
2    26.4
3    16.5
4    28.9
5    28.1
6    25.7
7    30.9
8    24.2
9    21.5
10   22.5
11   19.4
12   21.5
13   23.4
Name: Plant 2, dtype: float64
```

Sum of plant 1

In [139]: `sum_plant1 = df['Plant 1'].sum()
print('Sum of plant 1 =', sum_plant1)`

Sum of plant 1 = 228.49999999999997

Sum of plant 2

```
In [140]: ▶ sum_plant2 = df['Plant 2'].sum()  
          print('Sum of plant 2 =', sum_plant2)
```

Sum of plant 2 = 330.19999999999993

Mean of plant 1

```
In [141]: ▶ mean_plant1 = df['Plant 1'].mean()  
          print('Mean of plant 1 =', mean_plant1)
```

Mean of plant 1 = 16.32142857142857

Mean of plant 2

```
In [142]: ▶ mean_plant2 = df['Plant 2'].mean()  
          print('Mean of plant 2 =', mean_plant2)
```

Mean of plant 2 = 23.585714285714282

Median of plant 1

```
In [143]: ▶ median_plant1 = df['Plant 1'].median()  
          print('Median of plant 1 =', median_plant1)
```

Median of plant 1 = 15.75

Median of plant 2

```
In [144]: ▶ median_plant2 = df['Plant 2'].median()  
          print('Median of plant 2 =', median_plant2)
```

Median of plant 2 = 22.95

Mode of plant 1

```
In [145]: mode_plant1 = df['Plant 1'].mode()
print('Mode of plant 1 =\n', mode_plant1)
```

```
Mode of plant 1 =
0      11.9
1      13.3
2      13.8
3      14.0
4      14.6
5      14.7
6      15.7
7      15.8
8      16.0
9      16.2
10     18.4
11     19.6
12     20.9
13     23.6
dtype: float64
```

Mode of plant 2

```
In [146]: mode_plant2 = df['Plant 2'].mode()
print('Mode of plant 2 =\n', mode_plant2)
```

```
Mode of plant 2 =
0      21.5
dtype: float64
```

Standard deviation of plant 1

```
In [147]: std_plant1 = df['Plant 1'].std()
print('Standard deviation of plant 1 =', std_plant1)
```

```
Standard deviation of plant 1 = 3.2320935276425056
```

Standard deviation of plant 2

```
In [148]: std_plant2 = df['Plant 2'].std()
print('Standard deviation of plant 2 =', std_plant2)
```

```
Standard deviation of plant 2 = 4.02737882744847
```

Variance of plant 1

```
In [149]: ▶ var_plant1 = df['Plant 1'].var()
print('Variance of plant 1 =', var_plant1)
```

Variance of plant 1 = 10.446428571428575

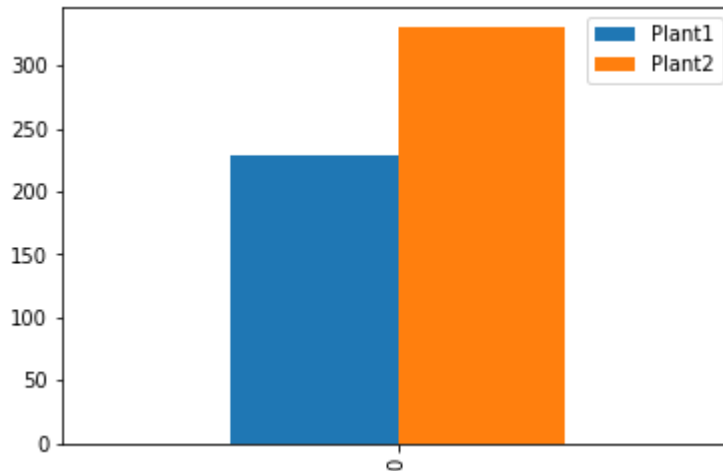
Variance of plant 2

```
In [150]: ▶ var_plant2 = df['Plant 2'].var()
print('Variance of plant 2 =', var_plant2)
```

Variance of plant 2 = 16.219780219780215

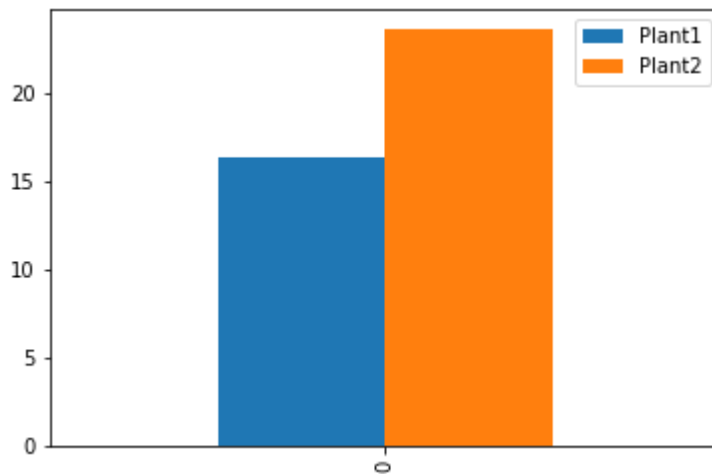
Plot for sum of plant 1 and plant 2

```
In [151]: ▶ data = [[sum_plant1, sum_plant2]]
df = pd.DataFrame(data, columns = ['Plant1', 'Plant2'])
df[['Plant1', 'Plant2']].plot(kind = 'bar')
plt.show()
```



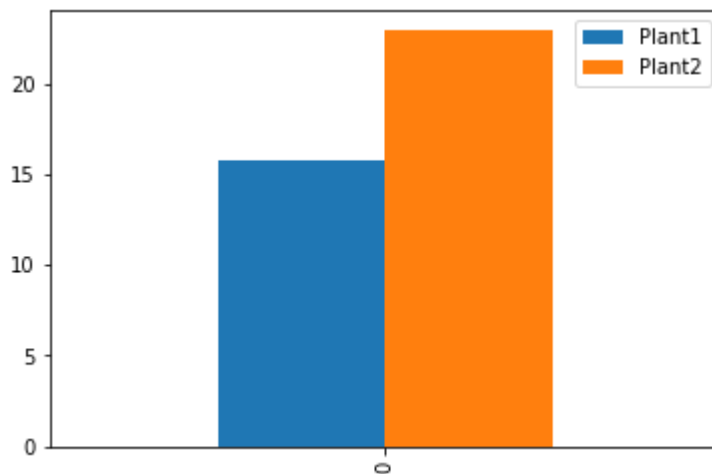
Plot for mean of plant 1 and plant 2

```
In [152]: data = [[mean_plant1, mean_plant2]]  
df = pd.DataFrame(data, columns = ['Plant1', 'Plant2'])  
df[['Plant1', 'Plant2']].plot(kind = 'bar')  
plt.show()
```



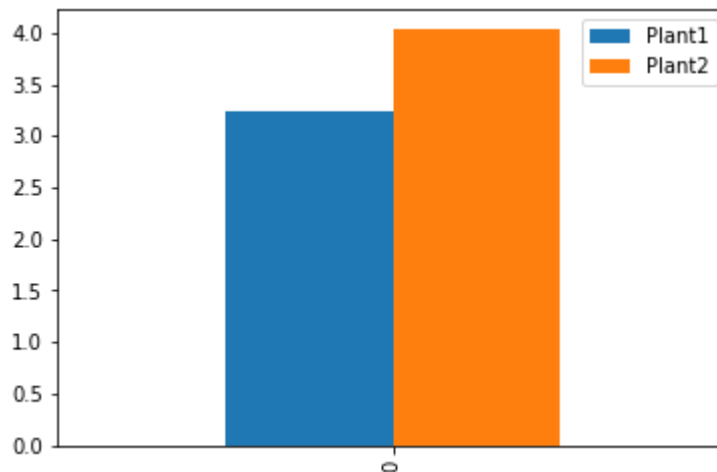
Plot for median of plant 1 and plant 2

```
In [153]: data = [[median_plant1, median_plant2]]  
df = pd.DataFrame(data, columns = ['Plant1', 'Plant2'])  
df[['Plant1', 'Plant2']].plot(kind = 'bar')  
plt.show()
```



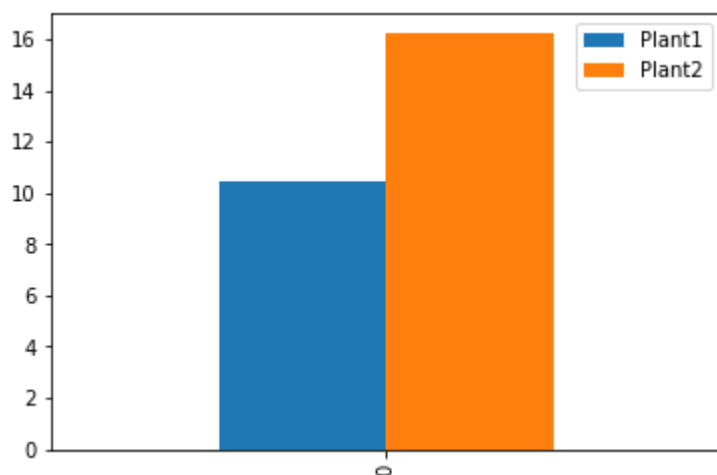
Plot for standard deviation of plant 1 and plant 2

```
In [154]: data = [[std_plant1, std_plant2]]
df = pd.DataFrame(data, columns = ['Plant1', 'Plant2'])
df[['Plant1', 'Plant2']].plot(kind = 'bar')
plt.show()
```



Plot for variance of plant 1 and plant 2

```
In [155]: data = [[var_plant1, var_plant2]]
df = pd.DataFrame(data, columns = ['Plant1', 'Plant2'])
df[['Plant1', 'Plant2']].plot(kind = 'bar')
plt.show()
```



Conclusion

Descriptive statistics comparison of plant 1 and plant 2 inferred from the plotted graph:

- 1) The mean of plant 1 is comparatively less than plant 2
- 2) The median of plant 1 is comparatively less than plant 2
- 3) The standard deviation of plant 1 is comparatively less than plant 2
- 4) The variance of plant 1 is comparatively less than plant 2

In conclusion, average sulfur dioxide emissions for plant 1 is less than plant 2.

