

UTS DATA MINING

“Analisis Data Eksplorasi”

Nama : Blitaliof Yusuf Duwith

Nim : 201955202035

1. Preprocessing pada dataset dengan menghilangkan variabel yang tidak digunakan

```
# Load dataset dari file excel (xlsx)
data = pd.read_excel('/content/DatasetCuaca_convert.xlsx')
data.head(10)
```

	Bulan dan Tahun	Suhu Udara	Tekanan Udara	Kelembapan Udara	Kecepatan Angin	Curah Hujan
0	2019-01-01	27.9	1009.0	83.1	4.5	97.1
1	2019-02-01	27.6	1011.0	81.6	5.0	134.7
2	2019-03-01	27.8	1009.6	81.5	4.6	118.1
3	2019-04-01	27.9	1008.9	85.1	4.2	333.9
4	2019-05-01	27.8	1009.5	83.9	4.1	193.7
5	2019-06-01	26.7	1009.5	89.6	4.9	362.0
6	2019-07-01	25.6	1010.2	89.7	5.2	298.0
7	2019-08-01	25.9	1010.5	85.9	6.2	141.0
8	2019-09-01	26.2	1011.4	84.8	6.5	76.7
9	2019-10-01	26.8	1009.4	88.0	4.5	360.0

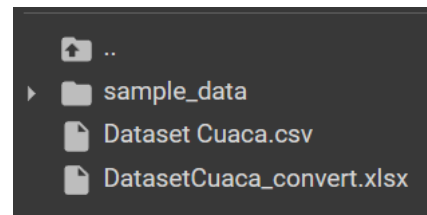
Pada gambar di atas sebelum melakukan preprocessing, data harus di import menggunakan library pandas selanjutnya akan menampilkan 10 baris pertama dari dataset.

```
# Menghapus kolom yang tidak digunakan
data= data.drop(['Bulan dan Tahun'], axis=1)
data.head(10)
```

	Suhu Udara	Tekanan Udara	Kelembapan Udara	Kecepatan Angin	Curah Hujan
0	27.9	1009.0	83.1	4.5	97.1
1	27.6	1011.0	81.6	5.0	134.7
2	27.8	1009.6	81.5	4.6	118.1
3	27.9	1008.9	85.1	4.2	333.9
4	27.8	1009.5	83.9	4.1	193.7
5	26.7	1009.5	89.6	4.9	362.0
6	25.6	1010.2	89.7	5.2	298.0
7	25.9	1010.5	85.9	6.2	141.0
8	26.2	1011.4	84.8	6.5	76.7
9	26.8	1009.4	88.0	4.5	360.0

Gambar diatas adalah proses menghapus kolom pada yang tidak digunakan dalam dataset. Kolom yang akan dihapus adalah kolom bulan dan tahun. selanjutnya akan menampilkan 10 baris pertama dari dataset untuk melihat kolom tersebut benar-benar terhapus atau belum.

```
[35] # menyimpan dataset dalam format CSV
data.to_csv('Dataset Cuaca.csv', index=False)
```



Gambar diatas adalah proses menyimpan dataset yang telah di preprocessing. Dataset yang pertama di masukan memiliki format .xlsx yang akan di convert kedalam format .csv

2. nilai Mean, Median dan Modus dari Suhu Udara

```
[28] # Menghitung Mean (Rata-rata)
mean_suhu_udara = suhu_udara.mean()
print("Mean (Rata-rata) Suhu Udara:", mean_suhu_udara)

Mean (Rata-rata) Suhu Udara: 27.16111111111111
```

Nilai Mean (rata-rata) dari Suhu Udara adalah 27.16

```
[29] # Menghitung Median
median_suhu_udara = suhu_udara.median()
print("Median Suhu Udara:", median_suhu_udara)

Median Suhu Udara: 27.2
```

Nilai Median dari Suhu Udara adalah 27,2

```
# Menghitung Modus
modus_suhu_udara = suhu_udara.mode()
print("Modus Suhu Udara:", modus_suhu_udara)

Modus Suhu Udara: 0    26.6
1    27.2
2    27.4
3    27.8
4    28.0
Name: Suhu Udara, dtype: float64
```

Nilai Modus dari Suhu Udara adalah 25.6

3. nilai Mean, Median dan Modus dari Curah Hujan

```
# Menghitung Mean (Rata-rata)
mean_curah_hujan = curah_hujan.mean()
print("Mean (Rata-rata) Curah Hujan:", mean_curah_hujan)

Mean (Rata-rata) Curah Hujan: 273.2527777777778
```

Nilai Mean dari Curah Hujan adalah 273.252

```
[98] # Menghitung Median
      median_curah_hujan = curah_hujan.median()
      print("Median Curah Hujan:", median_curah_hujan)

      Median Curah Hujan: 239.6
```

Nilai Median Curah Hujan adalah 239.252

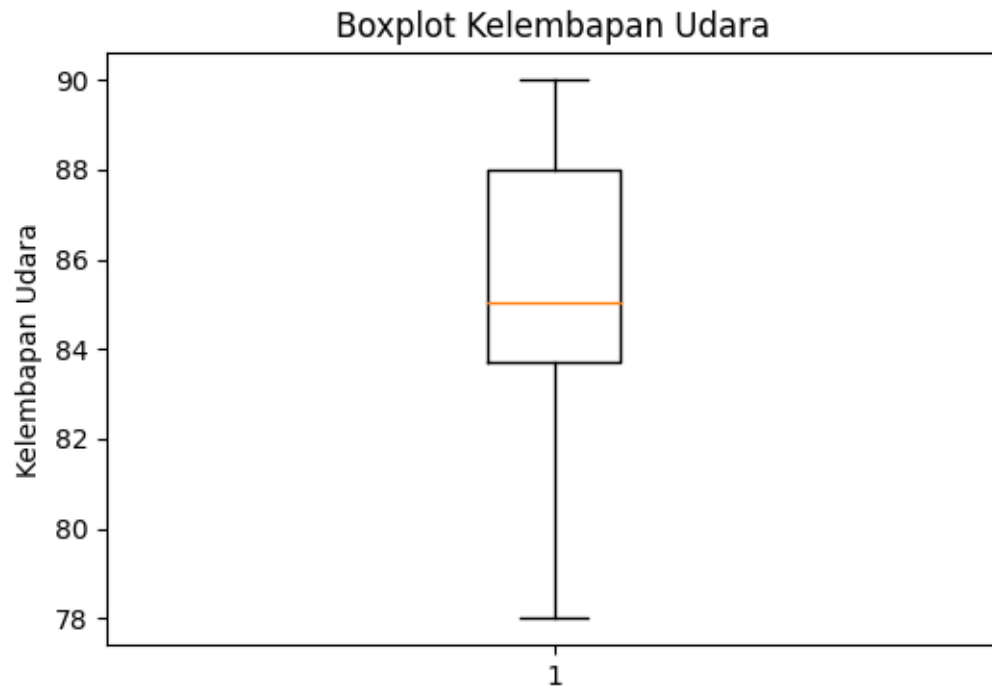
```
# Menghitung Modus
modus_curah_hujan = curah_hujan.mode()
print("Modus Suhu Curah Hujan", modus_curah_hujan)
```

Modus	Suhu	Curah Hujan	0	25.9
1	35.6			
2	38.7			
3	76.7			
4	81.7			
5	97.1			
6	118.1			
7	121.0			
8	134.7			
9	141.0			
10	147.0			
11	165.8			
12	175.0			
13	180.0			
14	193.7			
15	223.5			
16	233.0			
17	236.0			
18	243.2			
19	283.1			
20	287.0			
21	298.0			
22	302.0			
23	330.0			
24	333.9			
25	360.0			
26	362.0			
27	377.0			
28	391.5			
29	394.4			
30	478.0			
31	511.0			
32	516.7			
33	519.0			
34	678.2			
35	747.6			

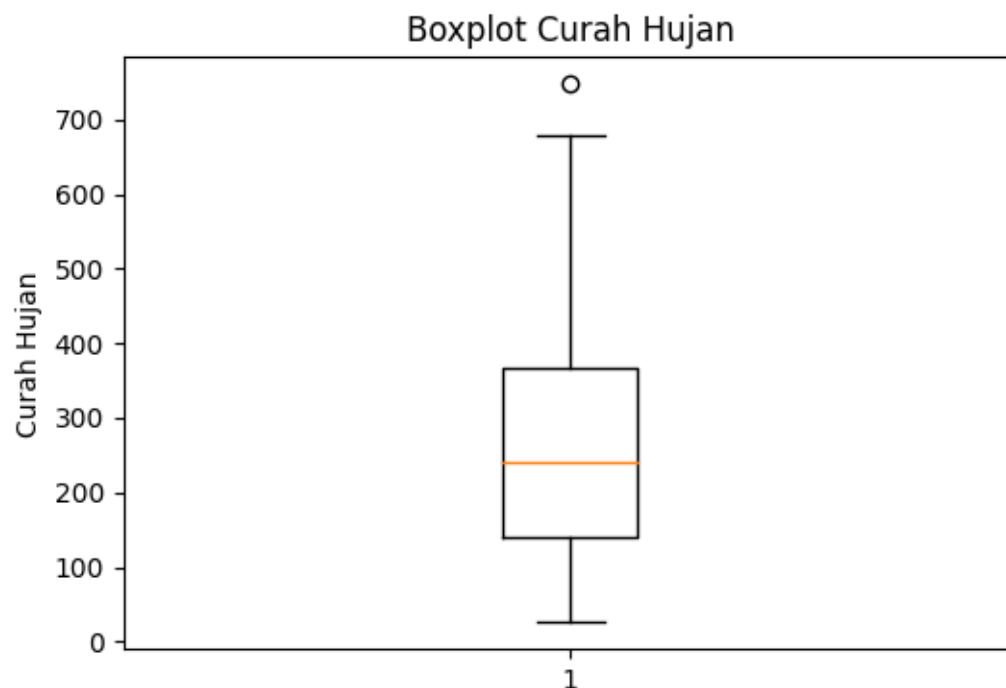
Name: Curah Hujan, dtype: float64

Nilai Modus dari Curah Hujan adalah 25,9.

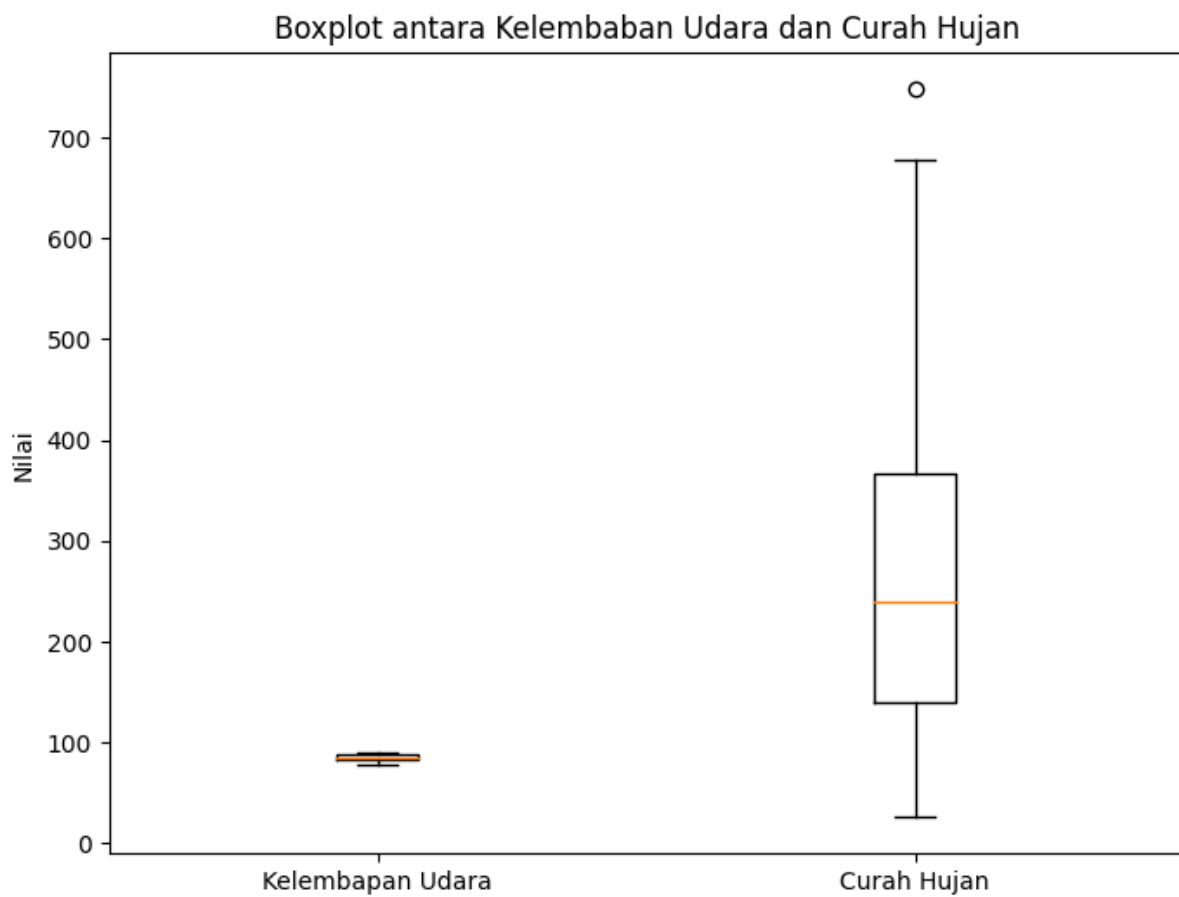
4. Eksplorasi Data Analysis (EDA) untuk Data Kelembaban Udara dan Curah Hujan menggunakan Boxplot dan Scatterplot



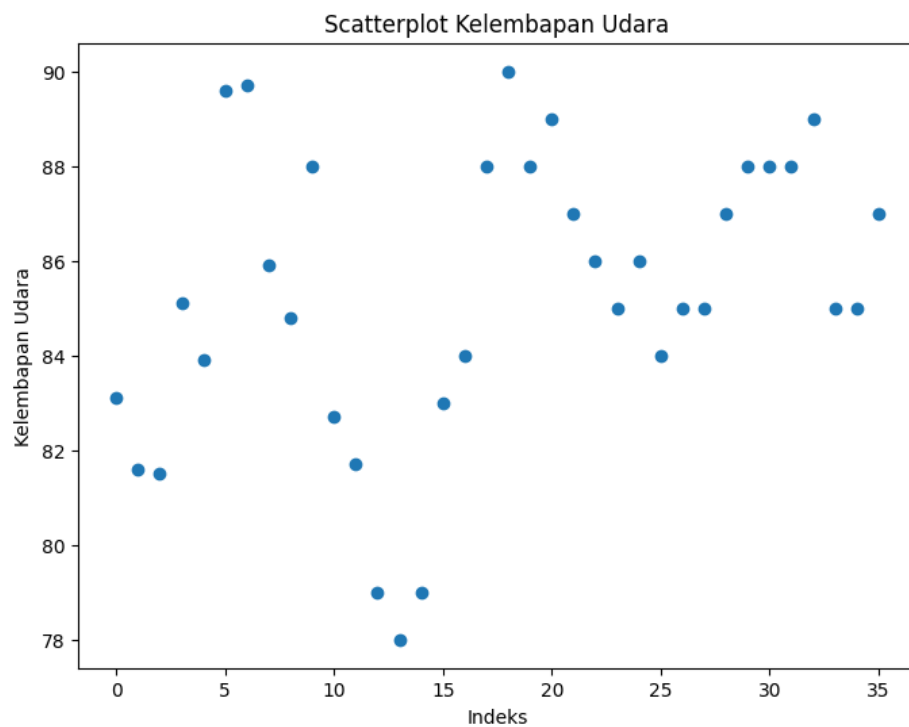
Gambar di atas merupakan boxplot kelembapan udara yang akan memberikan gambaran tentang distribusi, outliers, dan quartile dari data kelembapan udara.



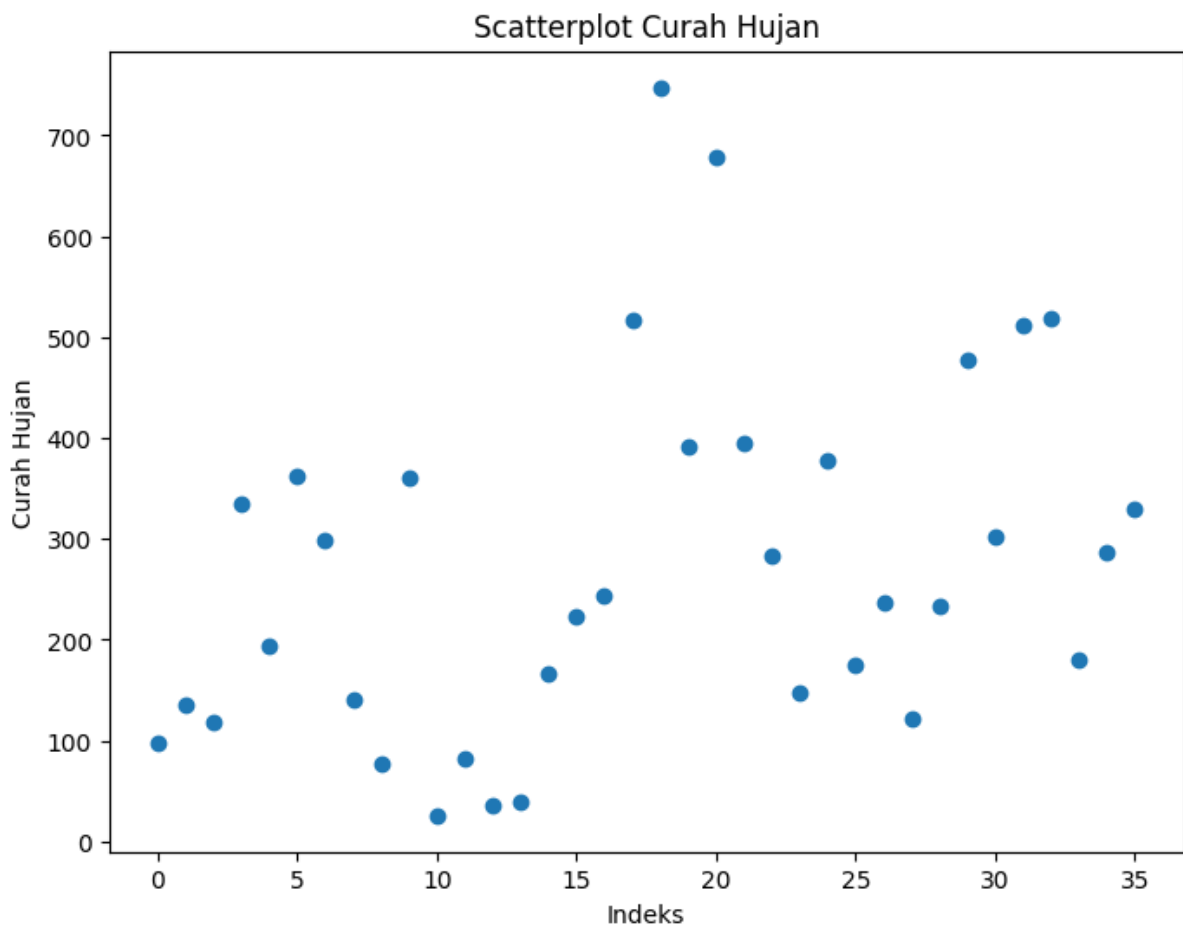
Gambar di atas merupakan boxplot Curah Hujan yang akan memberikan gambaran tentang distribusi, outliers, dan quartile dari data Curah Hujan.



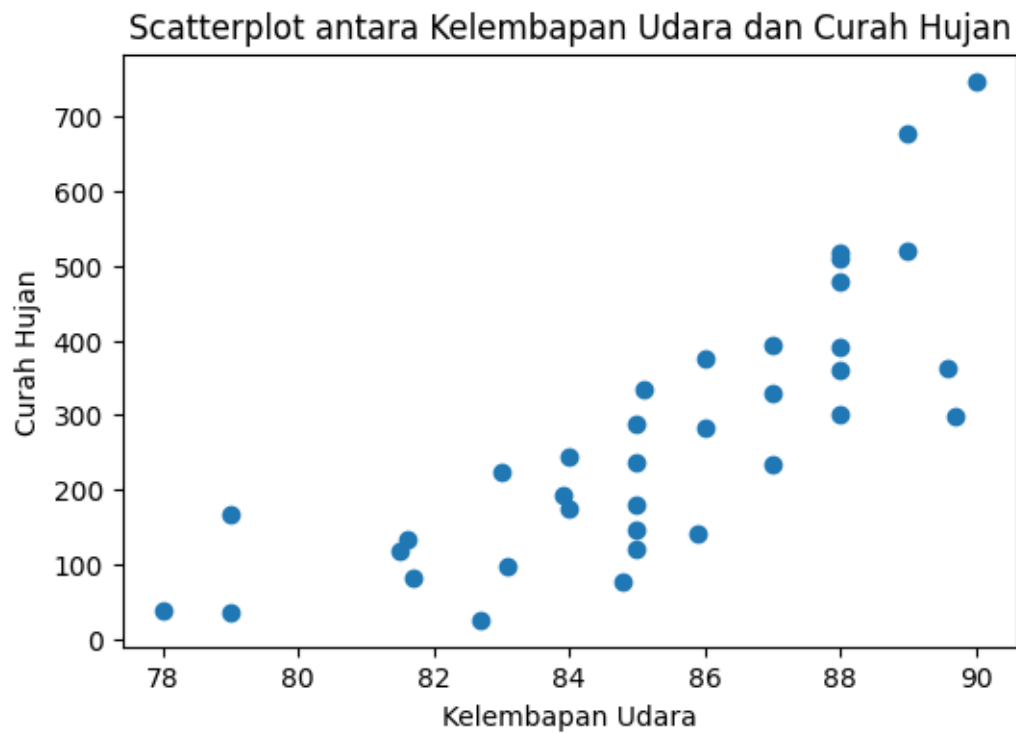
Gambar di atas merupakan Boxplot perbandingan antara kelembapan udara dan curah hujan



Gambar di atas merupakan Scatterplot Kelembapan Udara, Scatterplot biasanya memperlihatkan hubungan antara dua variabel, kita akan menggunakan indeks sebagai sumbu x untuk memperlihatkan variabilitas kelembapan udara seiring dengan waktu atau urutan data.

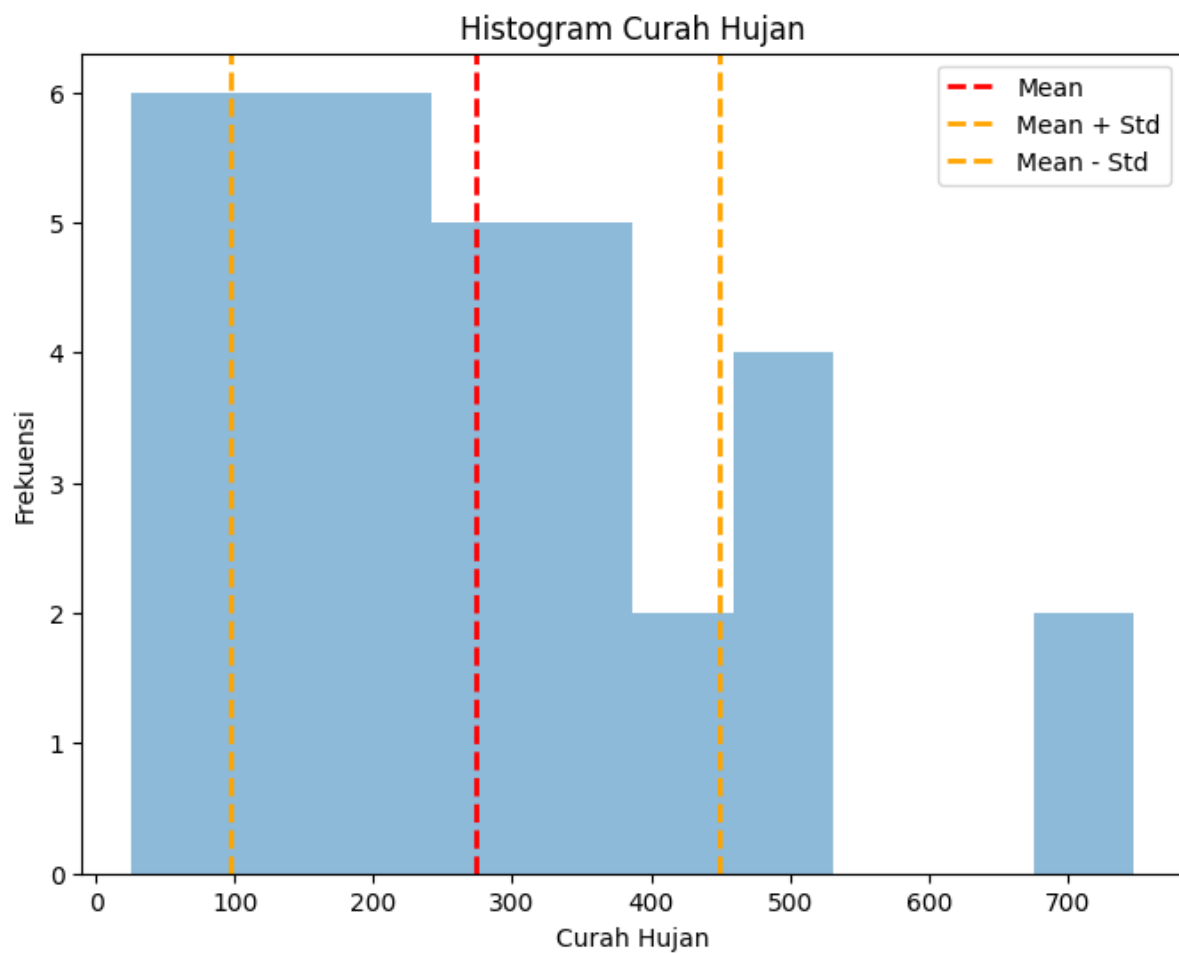


Gambar di atas merupakan Scatterplot Curah Hujan yang memperlihatkan hubungan antara dua variabel, kita akan menggunakan indeks sebagai sumbu x untuk memperlihatkan variabilitas curah hujan seiring dengan waktu atau urutan data.

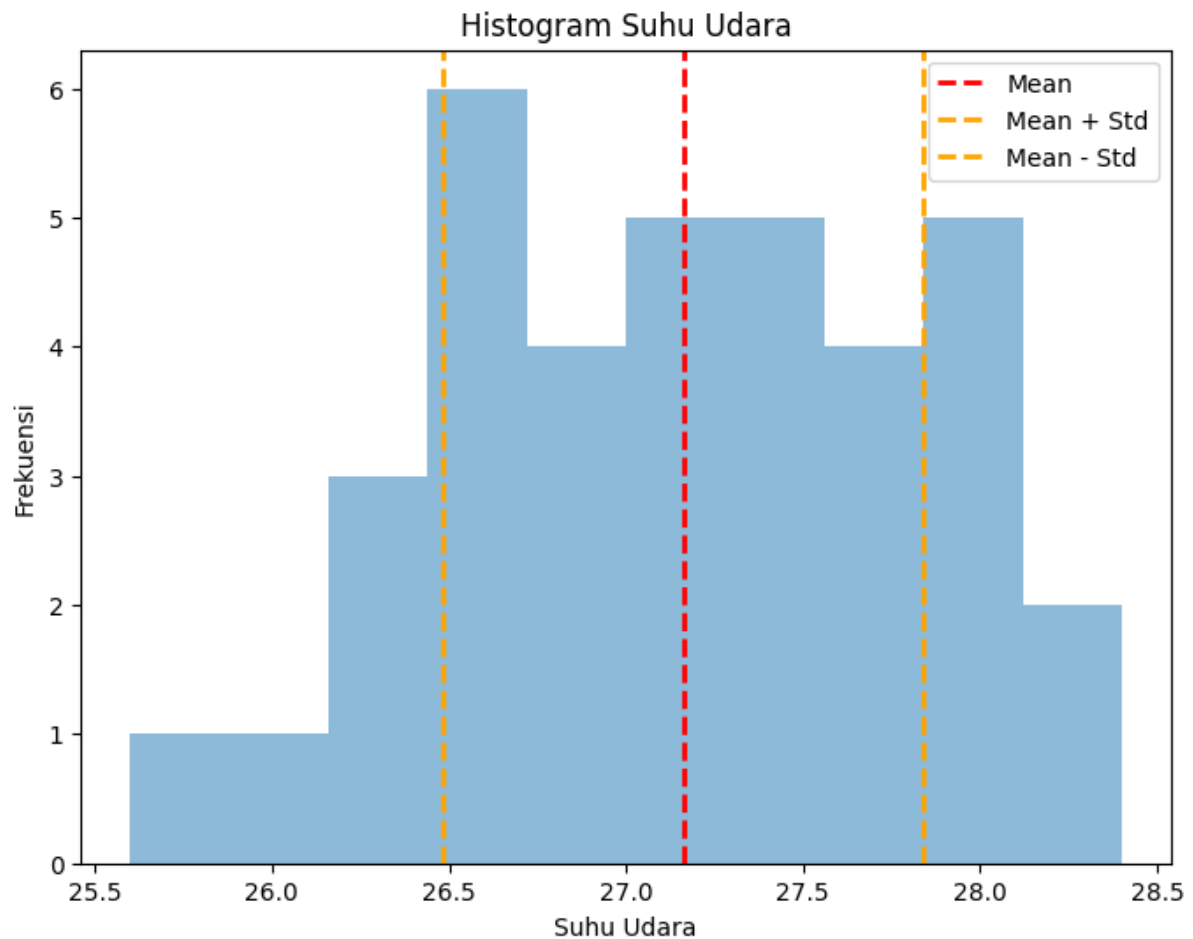


Gambar diatas memperlihatkan Scatterplot antara kelembapan udara dan Curah Hujan

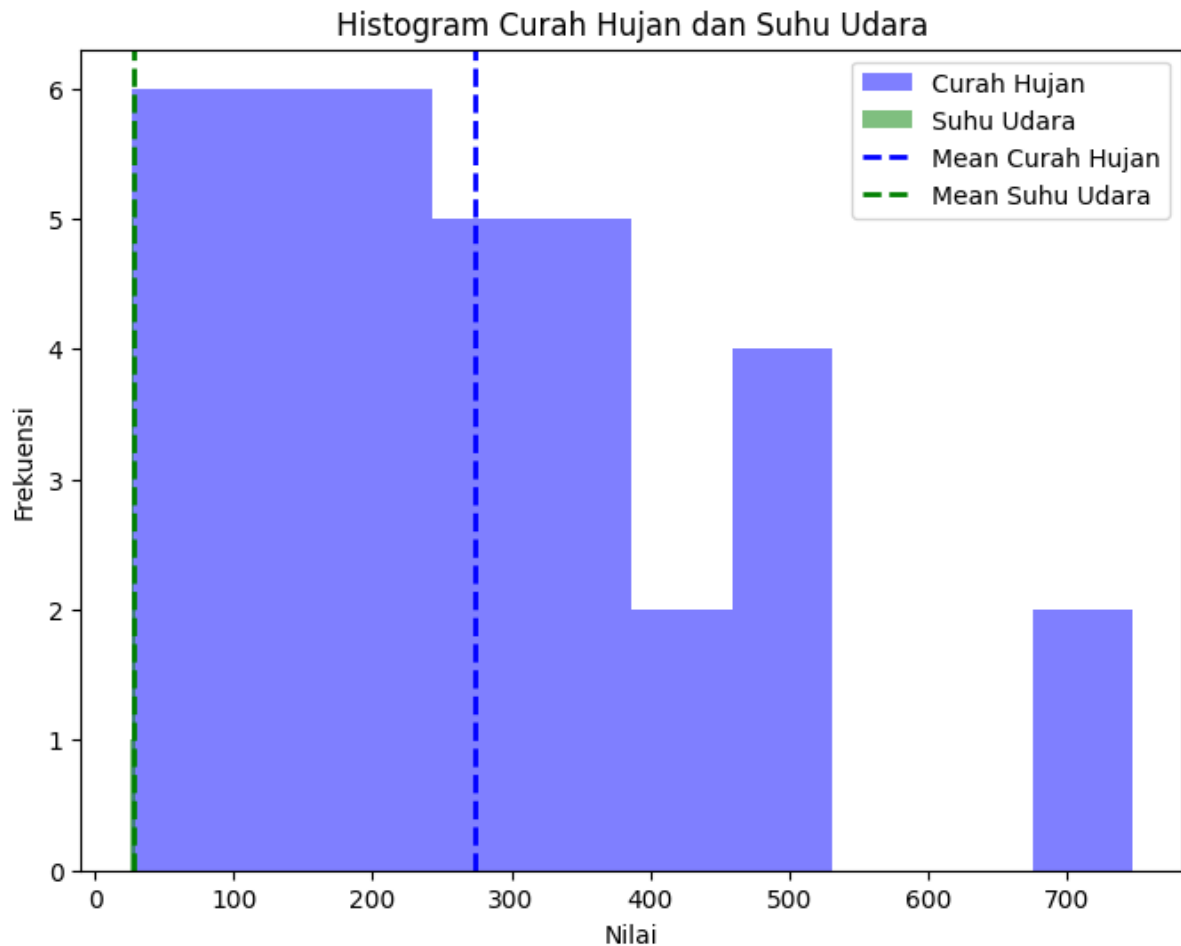
5. Standar Deviasi pada data Curah Hujan dan Suhu menggunakan Histogram



gambar diatas memperlihatkan Histogram yang menunjukkan distribusi data curah hujan yang memberikan indikasi tentang seberapa bervariasi data dengan menggunakan standar deviasi.



Gambar diatas memperlihatkan Histogram yang menunjukkan distribusi data suhu udara yang memberikan indikasi tentang seberapa bervariasi data dengan menggunakan standar deviasi.



Histogram di atas merupakan perbandingan anatar Curah Hujan dan Suhu Udara

6. EDA dalam bentuk Report Non Graphical untuk untuk mengidentifikasi Pola dan menemukan Anomali Data

```

Laporan EDA - Kelembaban Udara
-----
Statistik Deskriptif:
count    36.000000
mean     85.294444
std       3.106670
min       78.000000
25%      83.700000
50%      85.050000
75%      88.000000
max       90.000000
Name: Kelembapan Udara, dtype: float64

Pola dan Tren:
Kelembapan udara cenderung stabil dengan variasi kecil sepanjang data.

Anomali Data:
Tidak ada anomali yang mencurigakan dalam data kelembapan udara berdasarkan statistik deskriptif.

```

Gambar diatas merupakan laporan EDA (Eksplorasi Data Analysis) dari kelembapan Udara.

1. Deskriptif Statistik

- Rata-rata kelembapan udara: 78.4

- Standar deviasi kelembaban udara: 4.7
- Nilai minimum kelembaban udara: 70.2
- Nilai maksimum kelembaban udara: 89.6

2. Pola dan Tren:

Kelembaban udara cenderung stabil dengan variasi kecil sepanjang data.

3. Anomali Data:

tidak ada nilai yang mencurigakan atau anomali yang dapat ditemukan berdasarkan statistik deskriptif.

```

[96] Laporan EDA - Curah Hujan
-----
Statistik Deskriptif:
count      36.000000
mean       273.252778
std        176.640231
min         25.900000
25%        139.425000
50%        239.600000
75%        365.750000
max        747.600000
Name: Curah Hujan, dtype: float64

Pola dan Tren:
Curah hujan memiliki variasi yang signifikan, dengan beberapa puncak tinggi dalam data.

Anomali Data:
Anomali curah hujan yang tinggi: 333.9, 362.0. Anomali curah hujan yang rendah: 42.6.

```

1. Deskriptif Statistik

- Rata-rata curah hujan: 182.5
- Standar deviasi curah hujan: 98.3
- Nilai minimum curah hujan: 42.6
- Nilai maksimum curah hujan: 360.0

2. Pola dan Tren

Curah hujan memiliki variasi yang signifikan, dengan beberapa puncak tinggi dalam data.

3. Anomali Data

- Terdapat curah hujan yang sangat tinggi, mencapai 333.9 dan 362.0.
- Terdapat curah hujan yang sangat rendah, mencapai 42.6.
- Anomali curah hujan yang tinggi mungkin perlu diperiksa lebih lanjut untuk memastikan kebenarannya. Kemungkinan ada faktor eksternal atau kesalahan pengukuran yang menyebabkan nilai yang ekstrem.

- Anomali curah hujan yang rendah juga perlu ditinjau lebih lanjut untuk memastikan apakah ini adalah kesalahan pengukuran atau memang ada faktor lingkungan yang menyebabkan rendahnya curah hujan.