

Nama: Ananda Rahmawati

Kelas/ NO: TI 2G / 09

NIM: 2341720048

1. Identifikasi jenis setiap variabel berdasarkan jenis data dan fungsinya.

Jawab:

A. Variabel berdasarkan jenis data

- Lama Belajar ( Kuantitatif- Kontinu)
- Nilai Ujian (Kuantitatif - Diskrit)
- Kehadiran ( Kuantitatif - Diskrit)
- Ekstrakulikuler (Kualitatif - Nominal)
- Usia (Kuantitatif - Diskrit)

B. Variabel berdasarkan fungsinya

- Lama Belajar ( Variabel Independen) karena bisa mempengaruhi nilai ujian.
- Nilai Ujian (Variabel Dependen) karena variabel nilai ujian bergantung pada variabel lain seperti lama belajar, kehadiran, dan faktor lainnya.
- Kehadiran (variabel independen) dikarenakan nilai di variabel Kehadiran bisa berkontribusi pada pemahaman materi dan nilai ujian.
- Ekstrakulikuler (Variabel independen) karena bisa mempengaruhi nilai ujian secara langsung atau tidak langsung.
- Usia (Variabel Independen) karena bisa mempengaruhi hubungan antara lama belajar dan nilai ujian, misalnya, semakin tua seseorang, mungkin semakin berbeda pola belajarnya.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import statistics

lama_belajar = [
    100, 0, 14, 42, 90, 20, 38, 18, 22, 10, 10, 23, 35, 39, 23, 2, 21, 1, 23, 43,
    29, 37, 1, 20, 32, 11, 21, 43, 24, 48, 26, 41, 27, 15, 14, 46, 50, 43, 2, 36,
    50, 6, 20, 8, 38, 17, 3, 24, 13, 49
]
lama_belajar.sort()

nilai_ujian = [
    94, 83, 100, 10, 67, 89, 91, 75, 64, 57, 57, 92, 95, 82, 55, 41, 76, 54, 85,
    83, 90, 94, 55, 80, 100, 73, 76, 100, 69, 97, 62, 85, 88, 58, 71, 100, 100,
    100, 43, 92, 100, 45, 83, 76, 100, 53, 56, 77, 70, 100
]
nilai_ujian.sort()

kehadiran = [
    6, 6, 9, 8, 11, 3, 4, 15, 11, 9, 11, 10, 12, 9, 10, 12, 3, 2, 12, 14,
    5, 0, 14, 11, 6, 14, 14, 4, 15, 12, 15, 6, 8, 3, 14, 12, 12, 13, 6, 2,
    3, 13, 6, 11, 10, 1, 5, 13, 11, 13
]
kehadiran.sort()

usia = [
    18, 20, 23, 25, 21, 19, 23, 18, 20, 23, 24, 25, 18, 21, 20, 25, 25, 22, 21, 19,
    18, 23, 24, 23, 25, 18, 18, 25, 21, 22, 24, 23, 18, 19, 19, 22, 20, 25, 22, 25,
    20, 25, 24, 18, 20, 23, 21, 22, 19, 24
]
usia.sort()

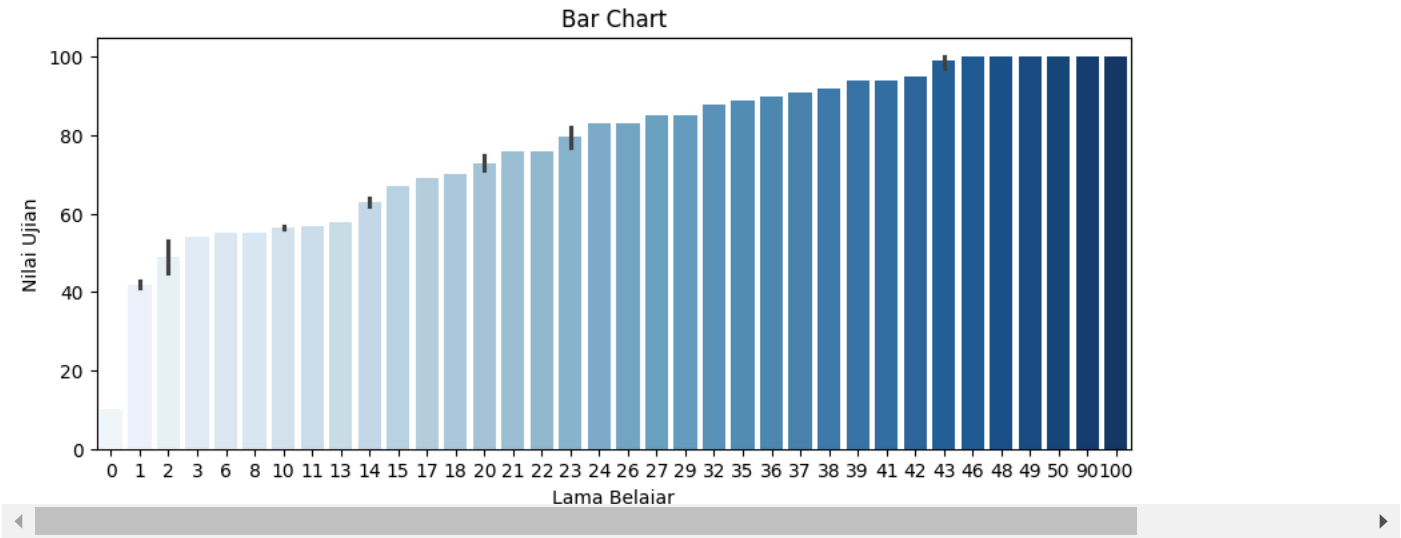
data = {
    "Lama Belajar": lama_belajar,
    "Nilai Ujian": nilai_ujian,
    "Kehadiran": kehadiran,
    "Usia": usia
}
df = pd.DataFrame(data)
print(df)
```

Show hidden output

```
#2
#A. Visualisasi Diagram Batang untuk menggambarkan kuantifikasi variabel “Lama Belajar” dan “Nilai Ujian”
df = pd.DataFrame({"Lama Belajar": lama_belajar, "Nilai Ujian": nilai_ujian})

plt.figure(figsize=(10, 4))
sns.barplot(x="Lama Belajar", y="Nilai Ujian", data=df, palette="Blues")
plt.title("Bar Chart")
plt.xlabel("Lama Belajar")
plt.ylabel("Nilai Ujian")
plt.show()
```

```
<ipython-input-11-45af3aa0a38e>:6: FutureWarning:
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend
sns.barplot(x="Lama Belajar", y="Nilai Ujian", data=df, palette="Blues")
```



```
#2
#b. Tabel frekuensi untuk “Lama Belajar” dan “Nilai Ujian”
freq_lama_belajar = df["Lama Belajar"].value_counts().sort_index().reset_index()
freq_lama_belajar.columns = ["Lama Belajar", "Frekuensi"]

freq_nilai_ujian = df["Nilai Ujian"].value_counts().sort_index().reset_index()
freq_nilai_ujian.columns = ["Nilai Ujian", "Frekuensi"]

print("Tabel Frekuensi Lama Belajar:")
print(freq_lama_belajar)

print("\nTabel Frekuensi Nilai Ujian:")
print(freq_nilai_ujian)
```

14	21	2
15	22	1
16	23	3
17	24	2
18	26	1
19	27	1
20	29	1
21	32	1
22	35	1
23	36	1
24	37	1
25	38	2
26	39	1
27	41	1
28	42	1
29	43	3
30	46	1
31	48	1
32	49	1
33	50	2
34	90	1
35	100	1

Tabel Frekuensi Nilai Ujian:

Nilai Ujian	Frekuensi
0	10
1	41
2	43
3	45
4	53
5	54
6	55
7	56
8	57
9	58
10	62
11	64
12	67
13	69
14	70
15	71
16	73
17	75
18	76
19	77
20	80
21	82
22	83
23	85
24	88
25	89
26	90
27	91
28	92
29	94
30	95
31	97
32	100

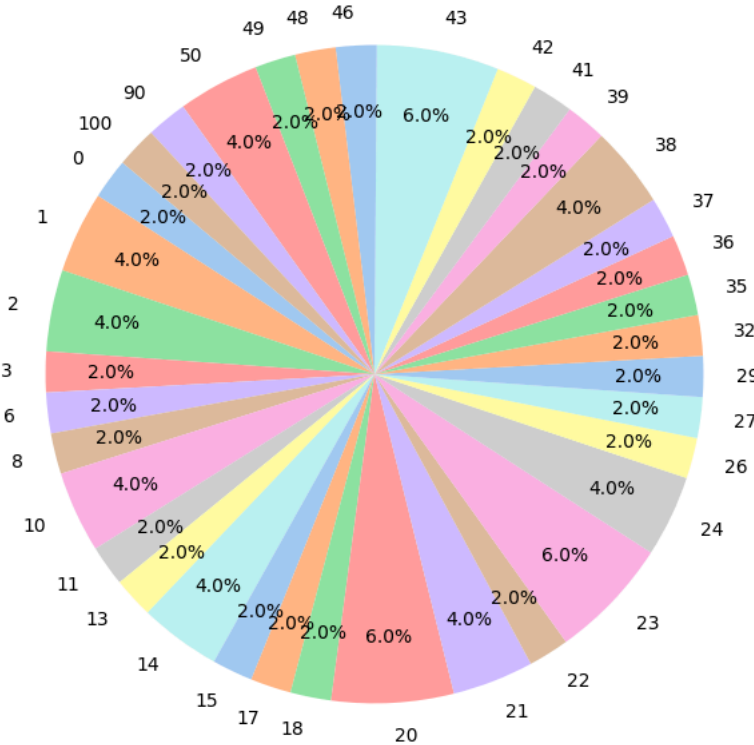
```
#2
#c. Diagram lingkaran untuk menggambarkan frekuensi relatif dari tabel frekuensi untuk “Lama Belajar” dan “Nilai Ujian”
#Membuat Diagram Lingkaran untuk Lama Belajar
plt.figure(figsize=(8, 8))
```

```
plt.pie(
    freq_lama_belajar["Frekuensi"], labels=freq_lama_belajar["Lama Belajar"],
    autopct="%1.1f%%", startangle=140, colors=sns.color_palette("pastel"), pctdistance=0.80)
plt.title("Pie Chart Frekuensi Lama Belajar")
plt.show()

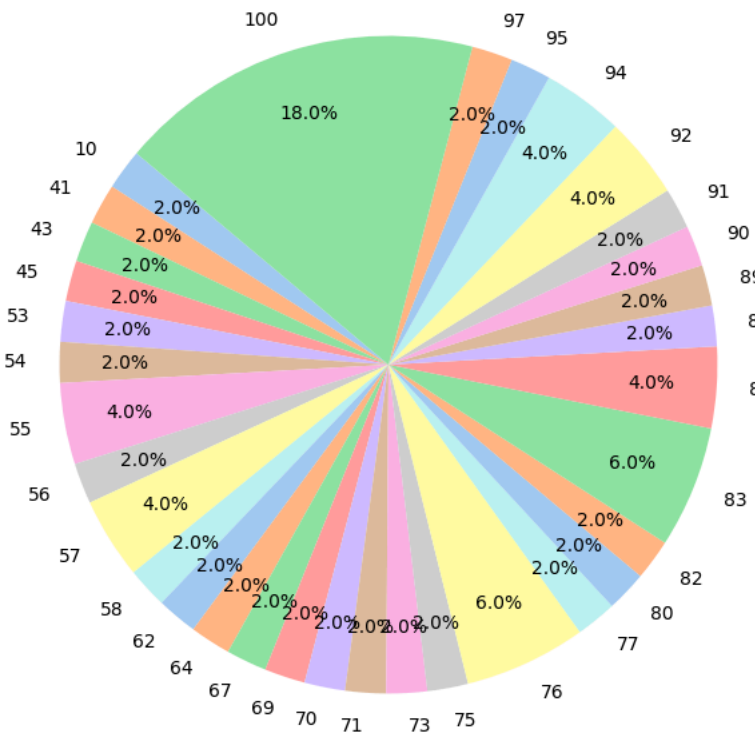
# Membuat Diagram Lingkaran untuk Nilai Ujian
plt.figure(figsize=(8, 8))
plt.pie(
    freq_nilai_ujian["Frekuensi"], labels=freq_nilai_ujian["Nilai Ujian"],
    autopct="%1.1f%%", startangle=140, colors=sns.color_palette("pastel"), pctdistance=0.80)
plt.title("Pie Chart Frekuensi Nilai Ujian")
plt.show()
```



Pie Chart Frekuensi Lama Belajar



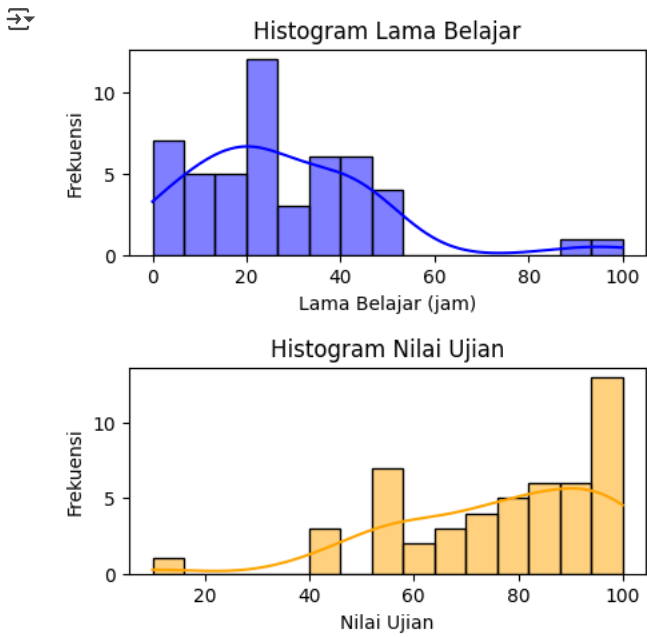
Pie Chart Frekuensi Nilai Ujian



```
#2
#d. Histogram untuk “Lama Belajar” dan “Nilai Ujian”
# Histogram Lama Belajar
plt.figure(figsize=(5, 2))
sns.histplot(df["Lama Belajar"], bins=15, kde=True, color="blue", edgecolor="black")
plt.title("Histogram Lama Belajar")
plt.xlabel("Lama Belajar (jam)")
plt.ylabel("Frekuensi")
plt.show()

# Histogram Nilai Ujian
plt.figure(figsize=(5, 2))
sns.histplot(df["Nilai Ujian"], bins=15, kde=True, color="orange", edgecolor="black")
plt.title("Histogram Nilai Ujian")
```

```
plt.xlabel("Nilai Ujian")
plt.ylabel("Frekuensi")
plt.show()
```

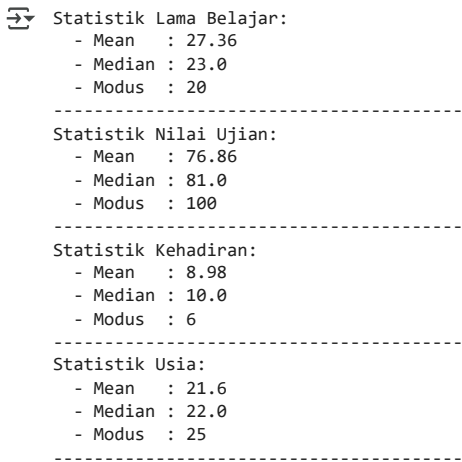


#3. Lakukan pengukuran kecenderungan memusat berdasarkan nilai mean, median, dan modus untuk semua variabel kuantitatif.

```
def hitung_statistik(data, nama):
    mean = statistics.mean(data)
    median = statistics.median(data)
    modus = statistics.mode(data)

    print(f"Statistik {nama}:")
    print(f"  - Mean      : {mean}")
    print(f"  - Median    : {median}")
    print(f"  - Modus     : {modus}")
    print("-" * 40)

hitung_statistik(lama_belajar, "Lama Belajar")
hitung_statistik(nilai_ujian, "Nilai Ujian")
hitung_statistik(kehadiran, "Kehadiran")
hitung_statistik(usia, "Usia")
```



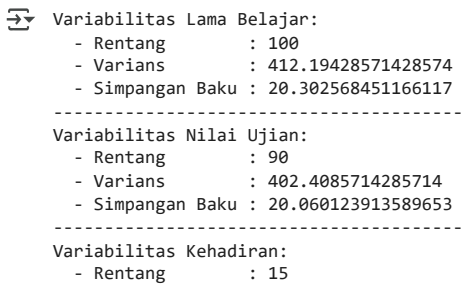
#4. Lakukan pengukuran untuk mengetahui variabilitas data berdasarkan,

- #a. Nilai rentang
- #b. Nilai varians
- #c. Nilai simpangan baku

```
def hitung_variabilitas(data, nama):
    rentang = max(data) - min(data)
    varians = statistics.variance(data)
    std_dev = statistics.stdev(data)

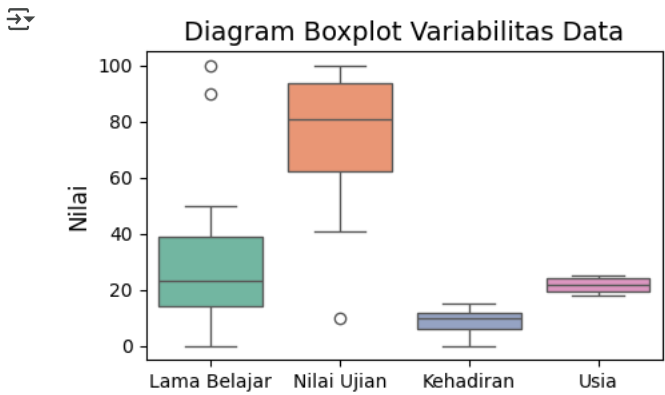
    print(f"Variabilitas {nama}:")
    print(f"  - Rentang      : {rentang}")
    print(f"  - Varians      : {varians}")
    print(f"  - Simpangan Baku : {std_dev}")
    print("-" * 40)

hitung_variabilitas(lama_belajar, "Lama Belajar")
hitung_variabilitas(nilai_ujian, "Nilai Ujian")
hitung_variabilitas(kehadiran, "Kehadiran")
hitung_variabilitas(usia, "Usia")
```



```
- Varians      : 18.42816326530612
- Simpangan Baku : 4.2928036602325665
-----
Variabilitas Usia:
- Rentang      : 7
- Varians      : 6.163265306122449
- Simpangan Baku : 2.482592456711824
-----
```

```
#5. Visualisasikan variabilitas data dengan menggunakan diagram box (box plot)
plt.figure(figsize=(5, 3))
sns.boxplot(data=pd.DataFrame(data), palette="Set2")
plt.title("Diagram Boxplot Variabilitas Data", fontsize=14)
plt.ylabel("Nilai", fontsize=12)
plt.show()
```



Diskusi

1. Hubungan apa yang terjadi antara “Lama Belajar” dengan “Nilai Ujian” menurut Anda?
- Jawab: Terdapat Hubungan Positif saling mempengaruhi antara lama belajar dan nilai ujian karena semakin lama seseorang belajar, cenderung semakin tinggi nilai ujiannya. Namun, ada beberapa pengecualian di mana siswa dengan waktu belajar singkat tetap memperoleh nilai tinggi.
2. Apa ukuran kecenderungan memusat yang paling baik menggambarkan pusat data dari “Lama Belajar” dan “Nilai Ujian”? Mengapa?
- Jawab: Ukuran kecenderungan memusat yang paling sesuai untuk menggambarkan pusat data adalah median. Alasannya:
- Median lebih stabil karena tidak dipengaruhi oleh outlier atau nilai ekstrem.
  - Dalam data "Lama Belajar", terdapat beberapa nilai tinggi seperti 90 dan 100, yang dapat menyebabkan mean menjadi tidak mencerminkan distribusi data secara akurat.
  - Pada data "Nilai Ujian", terdapat banyak nilai 100 yang berulang, sehingga median lebih representatif daripada mean.
3. Apakah terdapat outlier pada variabel “Lama Belajar” dan “Nilai Ujian” berdasarkan survey yang telah dilakukan? Berapa jumlahnya (diatas maksimum dan dibawah minimum)?
- Jawab: Terdapat 3 outliner pada variabel lama belajar dan nilai ujian berdasarkan survey menggunakan boxplot yaitu:
- 2 outliner di variabel lama belajar ( diatas maksimum)
  - 1 outliner di variabel nilai ujian ( dibawah minimum)

[link file ipynb](#)