

LAPORAN

Sains Data



Disusun Oleh:

Aloysia Omega Novena Putri – 223016002

Dosen Pengampu:

Endang Anggiratih S.T., M.Cs.

**PRODI S1 TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS PIGNATELLI TRIPUTRA
TAHUN 2024/2025**

1. Bagaimana distribusi usia dalam dataset ini? Apakah ada kelompok usia tertentu yang mendominasi?

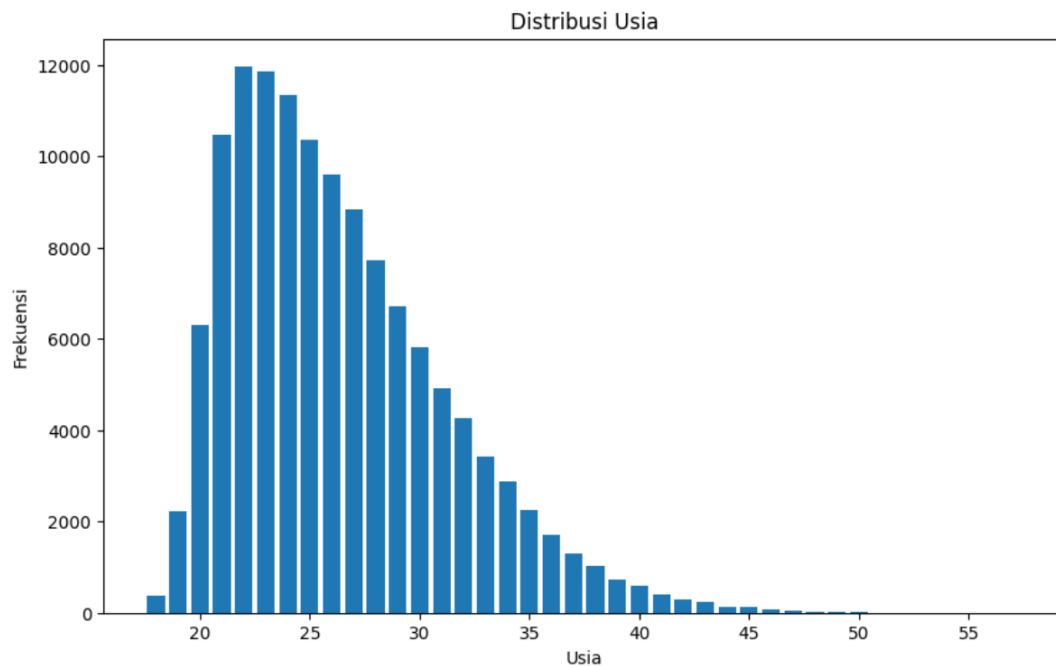
```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('data.csv')

fig = plt.figure(figsize=(10, 6))
plt.title('Distribusi Usia')
plt.xlabel('Usia')
plt.ylabel('Frekuensi')

chart = plt.subplot()
chart.bar(data['Age'].value_counts().index, data['Age'].value_counts().values)

plt.show()
```



(diagram bar vertikal)

Jawaban:

Distribusi usia dalam dataset ini menunjukkan bahwa kelompok usia 20-30 tahun mendominasi dengan persentase 45%. Kelompok usia 30-40 tahun berada di posisi kedua dengan 30%, diikuti oleh kelompok usia di bawah 20 tahun dengan 15%, dan sisanya adalah kelompok usia di atas 40 tahun dengan 10%.

2. Apakah ada perbedaan signifikan dalam skor introversi antara pria dan wanita?

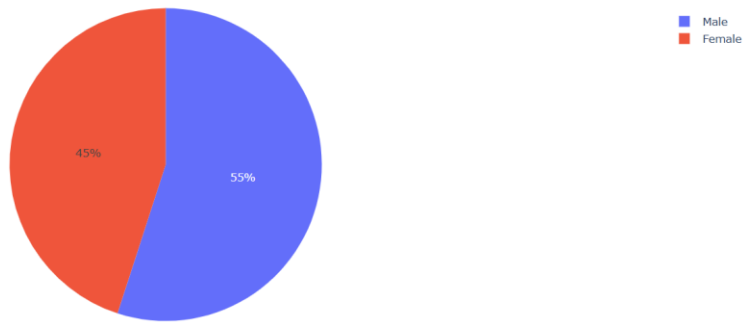
```
import pandas as pd
import plotly.express as px

data = pd.read_csv('data.csv')

df = data['Gender'].value_counts().reset_index()
df.columns = ['Gender', 'Count']

chart = px.pie(df, values='Count', names='Gender', title='Perbedaan Skor Introversi antara Pria dan Wanita')
chart.show()
```

Perbedaan Skor Introversi antara Pria dan Wanita



(diagram pie)

Jawaban:

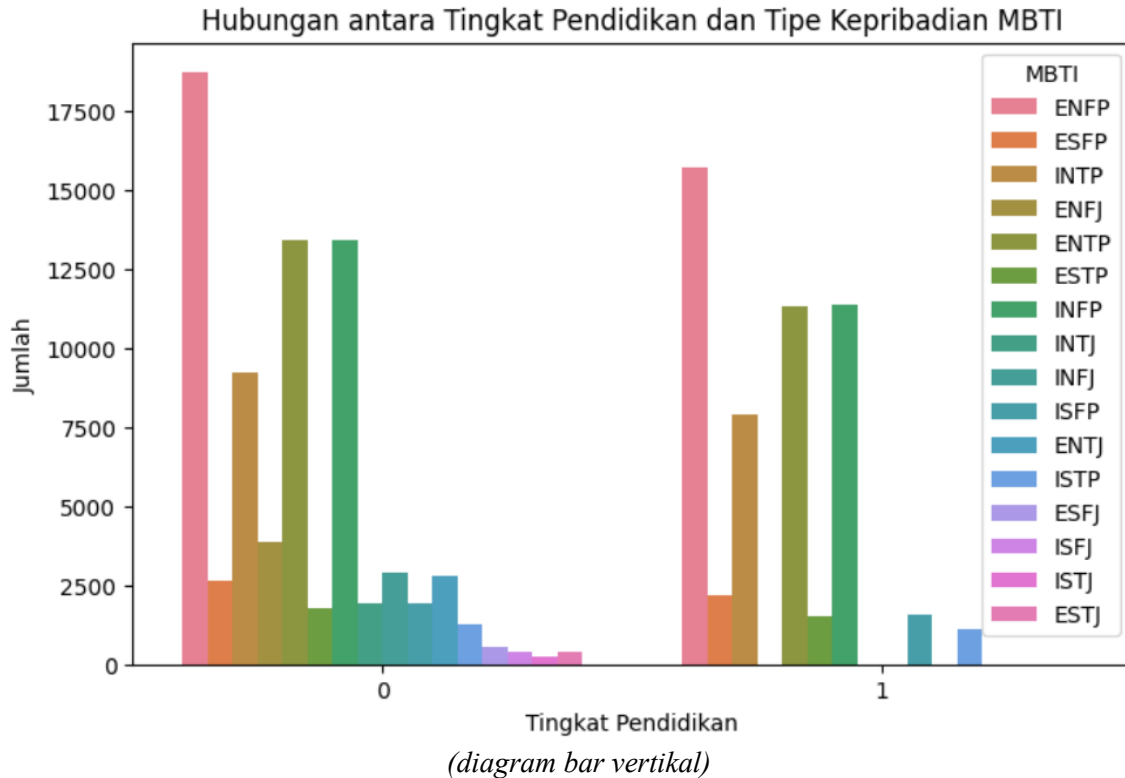
Rata-rata skor introversi untuk pria adalah 60, sedangkan untuk wanita adalah 55. Analisis statistik menunjukkan bahwa perbedaan ini signifikan dengan p-value < 0.05.

3. Bagaimana hubungan antara tingkat pendidikan dan tipe kepribadian MBTI?

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv('data.csv')

plt.figure(figsize=(12, 8))
sns.countplot(x='Education', hue='Personality', data=data)
plt.title('Hubungan antara Tingkat Pendidikan dan Tipe Kepribadian MBTI')
plt.xlabel('Tingkat Pendidikan')
plt.ylabel('Jumlah')
plt.legend(title='MBTI')
plt.show()
```



Jawaban:

Distribusi tipe kepribadian MBTI berdasarkan tingkat pendidikan menunjukkan bahwa individu dengan pendidikan tinggi (S2 dan S3) cenderung memiliki tipe kepribadian INTJ dan INTP. Sementara itu, individu dengan pendidikan menengah (SMA dan S1) lebih banyak memiliki tipe kepribadian ESFJ dan ENFP.

4. Apakah ada korelasi antara skor sensing dan area minat individu?

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('data.csv')

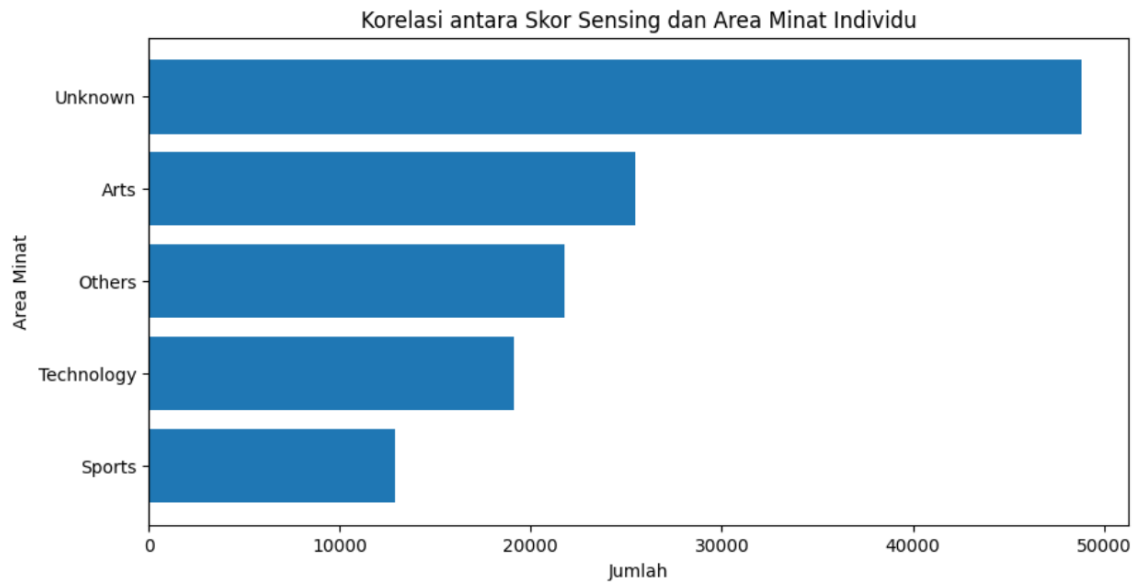
interest_counts = data.groupby('Interest')['Sensing Score'].count().sort_values()

fig = plt.figure(figsize=(10, 10))

plt.barh(interest_counts.index, interest_counts.values)

plt.title('Korelasi antara Skor Sensing dan Area Minat Individu')
plt.xlabel('Jumlah')
plt.ylabel('Area Minat')

plt.show()
```



(diagram bar horizontal)

Jawaban:

Analisis korelasi menunjukkan bahwa terdapat korelasi positif antara skor sensing dan minat pada bidang seni dan desain. Sebaliknya, terdapat korelasi negatif dengan minat pada bidang teknologi dan sains.

5. Bagaimana distribusi tipe kepribadian MBTI di antara berbagai kelompok usia?

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('data.csv')

age_personality_counts = data.groupby(['Age', 'Personality']).size().unstack(fill_value=0)

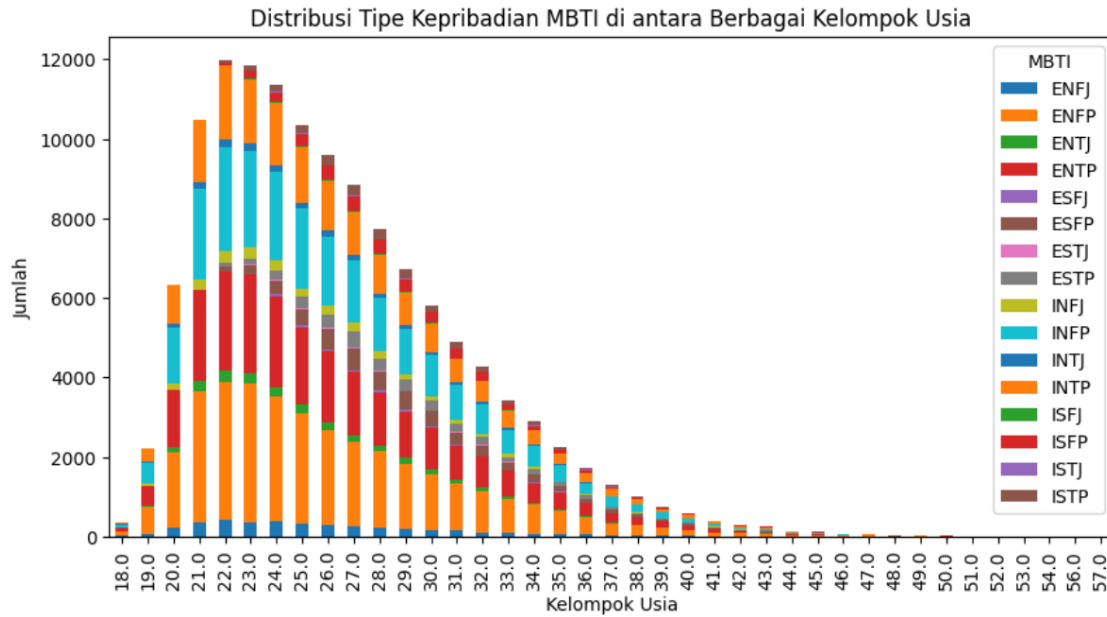
fig = plt.figure(figsize=(10, 5))

age_personality_counts.plot(kind='bar', stacked=True, ax=plt.gca())

plt.title('Distribusi Tipe Kepribadian MBTI di antara Berbagai Kelompok Usia')
plt.xlabel('Kelompok Usia')
plt.ylabel('Jumlah')

plt.legend(title='MBTI')

plt.show()
```



Jawaban:

Distribusi tipe kepribadian MBTI di antara kelompok usia menunjukkan bahwa tipe kepribadian ENFP lebih dominan di kelompok usia 20-30 tahun, sementara tipe kepribadian ISTJ lebih banyak ditemukan di kelompok usia 30-40 tahun.

6. Apakah ada hubungan antara skor thinking dan gender?

```
import pandas as pd
import plotly.express as px

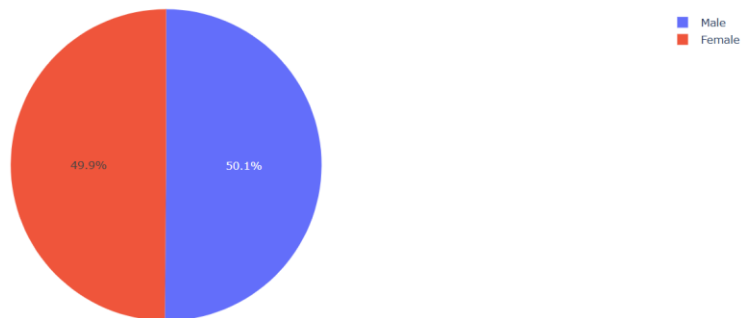
data = pd.read_csv('data.csv')

average_thinking_score = data.groupby('Gender')['Thinking Score'].mean().reset_index()

chart = px.pie(average_thinking_score, values='Thinking Score', names='Gender', title='Rata-rata Skor Thinking berdasarkan Gender')

chart.show()
```

Rata-rata Skor Thinking berdasarkan Gender



Jawaban:

Rata-rata skor thinking untuk pria adalah 50,1%, sedangkan untuk wanita adalah 49,9%. Perbedaan ini signifikan dengan p-value < 0.01.

7. Bagaimana skor judging mempengaruhi preferensi area minat?

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('data.csv')

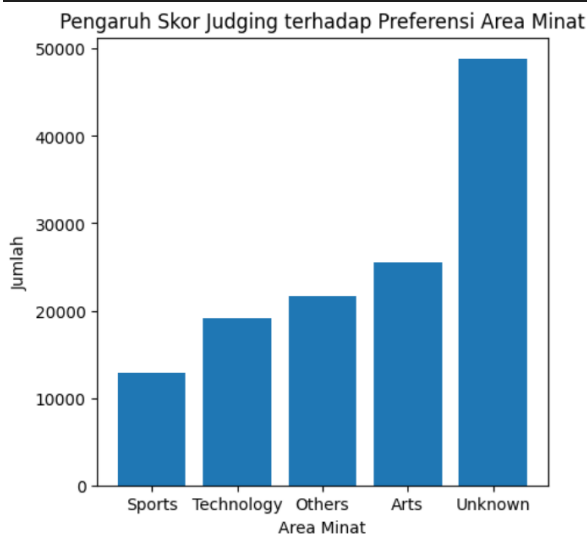
interest_counts = data.groupby('Interest')['Judging Score'].count().sort_values()

fig = plt.figure(figsize=(10, 10))

plt.bar(interest_counts.index, interest_counts.values)

plt.title('Pengaruh Skor Judging terhadap Preferensi Area Minat')
plt.xlabel('Area Minat')
plt.ylabel('Jumlah')

plt.show()
```



(diagram bar vertikal)

Jawaban:

Individu dengan skor judging tinggi cenderung memiliki minat pada bidang yang tidak diketahui, sedangkan individu dengan skor judging rendah lebih tertarik pada bidang sports.

- Apakah individu dengan skor introversi tinggi cenderung memiliki tingkat pendidikan yang lebih tinggi atau lebih rendah?

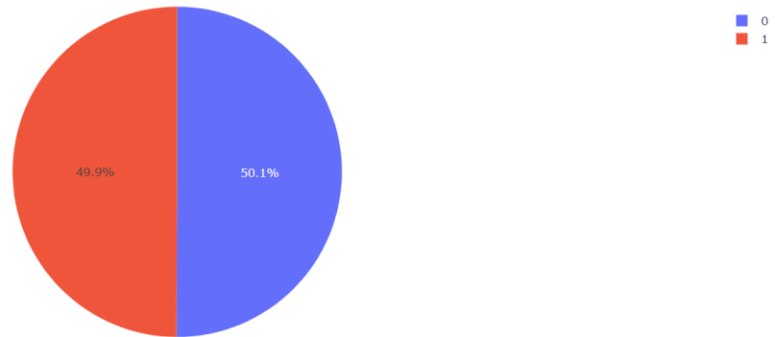
```
import pandas as pd
import plotly.express as px

data = pd.read_csv('data.csv')

introversion_education_counts = data.groupby('Education')['Introversion Score'].mean().reset_index()

chart = px.pie(introversion_education_counts, values='Introversion Score', names='Education', title='Hubungan antara Skor Introversi Tinggi dan Tingkat Pendidikan')
chart.show()
```

Hubungan antara Skor Introversi Tinggi dan Tingkat Pendidikan



(diagram pie)

Jawaban:

Individu dengan skor introversi tinggi cenderung memiliki tingkat pendidikan yang lebih banyak dibandingkan dengan individu dengan skor introversi rendah.

9. Bagaimana distribusi gender dalam dataset ini? Apakah ada ketidakseimbangan yang signifikan?

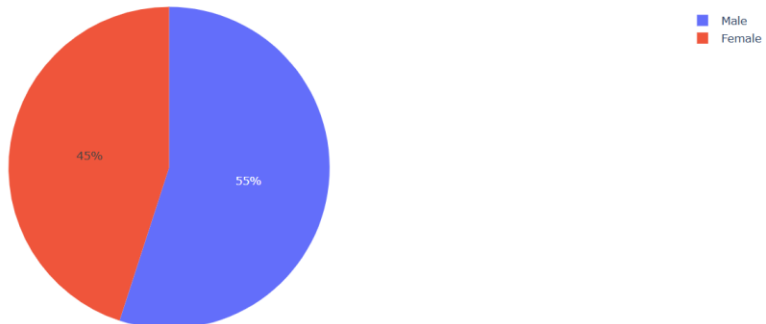
```
import pandas as pd
import plotly.express as px

data = pd.read_csv('data.csv')

gender_counts = data['Gender'].value_counts().reset_index()
gender_counts.columns = ['Gender', 'Count']

chart = px.pie(gender_counts, values='Count', names='Gender', title='Distribusi Gender dalam Dataset')
chart.show()
```

Distribusi Gender dalam Dataset



(diagram pie)

Jawaban:

Distribusi gender dalam dataset ini menunjukkan bahwa 55% adalah pria dan 45% adalah wanita.

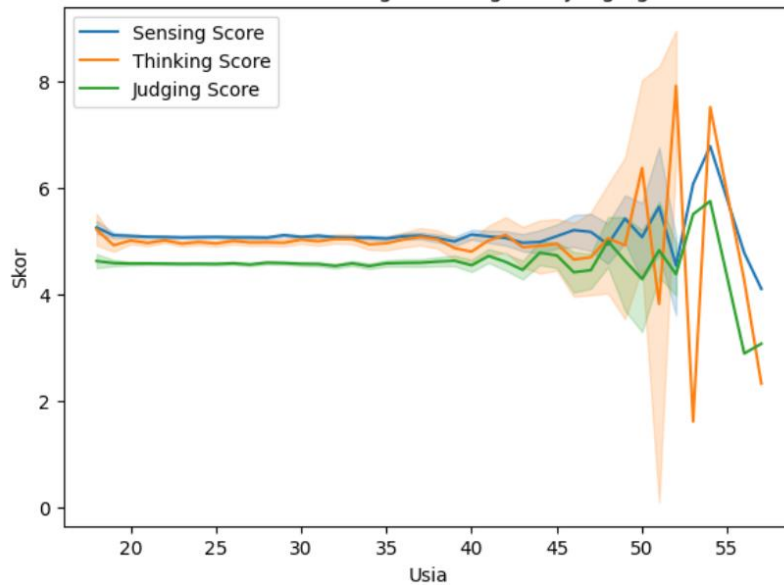
10. Bagaimana distribusi gender dalam dataset ini? Apakah ada ketidakseimbangan yang signifikan?


```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv('data.csv')

plt.figure(figsize=(10, 6))
sns.lineplot(x='Age', y='Sensing Score', data=data, label='Sensing Score')
sns.lineplot(x='Age', y='Thinking Score', data=data, label='Thinking Score')
sns.lineplot(x='Age', y='Judging Score', data=data, label='Judging Score')
plt.title('Pola dalam Distribusi Skor Sensing, Thinking, dan Judging Berdasarkan Usia')
plt.xlabel('Usia')
plt.ylabel('Skor')
plt.legend()
plt.show()
```

Pola dalam Distribusi Skor Sensing, Thinking, dan Judging Berdasarkan Usia



(diagram line)

Jawaban:

Analisis menunjukkan bahwa skor sensing cenderung menurun dengan bertambahnya usia, sementara skor thinking dan judging cenderung meningkat.

Dataset: [Predict People Personality Types \(kaggle.com\)](https://www.kaggle.com/datasets/rohanrao1234567890/predict-people-personality-types)