

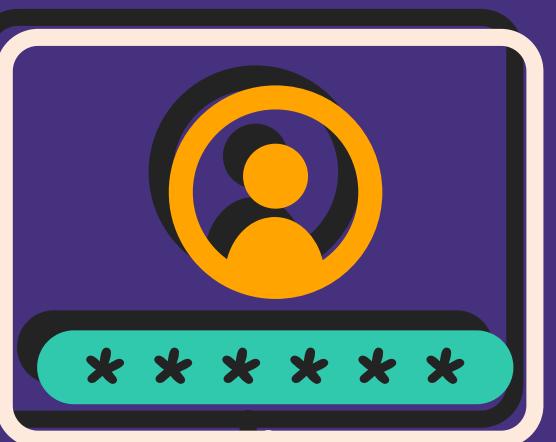
DATA MANIPULATION & VISUALISATION)

NADILA IMAARAH 3323600015 2D4 SDT A

DATASET TITANIC

DATASET TITANIC INI SAYA DAPATKAN DARI WEB KAGGLE. DATA INI BERISI INFORMASI MENGENAI PENUMPANG PADA KAPAL TITANIC.

PADA DATASET INI TERDIRI DARI BEBERAPA VARIABLE YAKNI PASSENGERID. SURVIVED, PCLASS DAN LAIN SEBAGAINYA



MENGHITUNG JUMLAH BARIS DAN KOLOM

```
rows, cols = df.shape  
print(f"Jumlah baris: {rows}, Jumlah kolom: {cols}")
```

Jumlah baris: 1349, Jumlah kolom: 14



Pada tahap ini saya menampilkan jumlah baris dan jumlah kolom dari dataset titanic, didapatkan jumlah kolom sebanyak 14 dan jumlah baris sebanyak 1349

MENGAMBIL DATASET KOLOM NAME,SEX, AGE,PCLASS DAN FARE



```
data = pd.DataFrame(df,columns=['name', 'sex', 'age', 'pclass', 'fare'])  
data.head()
```



	name	sex	age	pclass	fare
0	Allen, Miss. Elisabeth Walton	female	29.0000	1	211.3375
1	Allison, Master. Hudson Trevor	male	0.9167	1	151.5500
2	Allison, Miss. Helen Loraine	female	2.0000	1	151.5500
3	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1	151.5500
4	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1	151.5500



Pada tahap ini saya mengambil dan menampilkan yang hanya terdiri dari kolom name, sex, age , pclass dan fare

MENGAMBIL DATA SET KOLOM KELAS (SURVIVED)



```
data = pd.DataFrame(df, columns=['survived'])  
data.head()
```

	survived
0	1
1	1
2	0
3	0
4	0

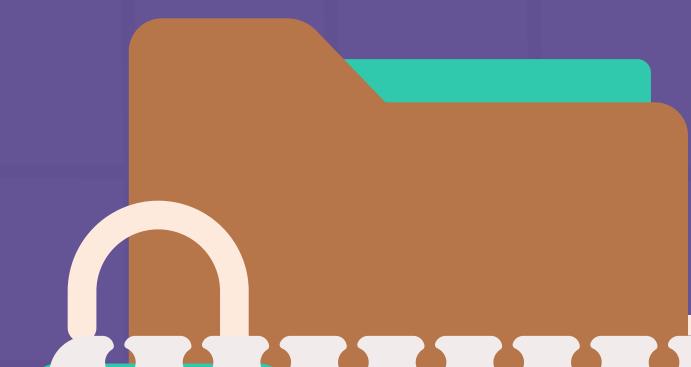
Pada tahap ini saya mengambil dan menampilkan kolom survived

DATA + FITUR RELATIVES



```
# Menambahkan fitur 'Relatives'  
df['relatives'] = df['sibsp'] + df['parch']  
  
# Tampilkan dataset dengan fitur baru  
df[['name', 'sex', 'age', 'pclass', 'fare', 'relatives']].head()
```

		name	sex	age	pclass	fare	relatives
0		Allen, Miss. Elisabeth Walton	female	29.0000	1	211.3375	0
1		Allison, Master. Hudson Trevor	male	0.9167	1	151.5500	3
2		Allison, Miss. Helen Loraine	female	2.0000	1	151.5500	3
3		Allison, Mr. Hudson Joshua Creighton	male	30.0000	1	151.5500	3
4		Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1	151.5500	3



Koleksi data adalah proses mengumpulkan informasi dari survei, wawancara, sensor, media sosial, dan database untuk dianalisis.

HITUNG PENUMPANG MASING-MASING PCLASS



```
pclass_counts = df['pclass'].value_counts()  
print(pclass_counts)
```

```
pclass  
3    747  
1    324  
2    278  
Name: count, dtype: int64
```

Pada tahap ini saya menghitung dan menampulkan jumlah penumpang berdasarkan pclass, dan didapatkan hasil bahwa :

1. pclass 3 memiliki penumpang sebanyak 747
2. pclass 1 memiliki penumpang sebanyak 324
3. pclass 2 memiliki penumpang sebanyak 278



HITUNG JUMLAH PENUMPANG BERDASARKAN SEX



```
sex_counts = df['sex'].value_counts()  
print(sex_counts)
```

sex	
male	874
female	475

Pada tahap ini saya mengitung dan menampilkan jumlah penumpang berdasarkan sex dan didapatkan hasilnya yakni :

1. terdapat penumpang dengan sex male berjumlah 874
2. terdapat penumpang dengan sex female berjumlah 475

HITUNG JUMLAH PENUMPANG SELAMAT (1) DAN TIDAK SELAMAT (0) BERDASARKAN PCCLASS



```
survival_counts = df.groupby(['pclass', 'survived']).size().unstack()  
print(survival_counts)
```

Pada tahap ini menampilkan jumlah penumpang selamat dan tidak selamat berdasarkan pclass dan didapatkan bahwa :

1. pclass 1 yang selamat sebanyak 201 dan tidak selamat sebanyak 123
2. pclass 2 yang selamat sebanyak 120 dan tidak selamat sebanyak 158
3. pclass 3 yang selamat sebanyak 181 dan tidak selamat 566

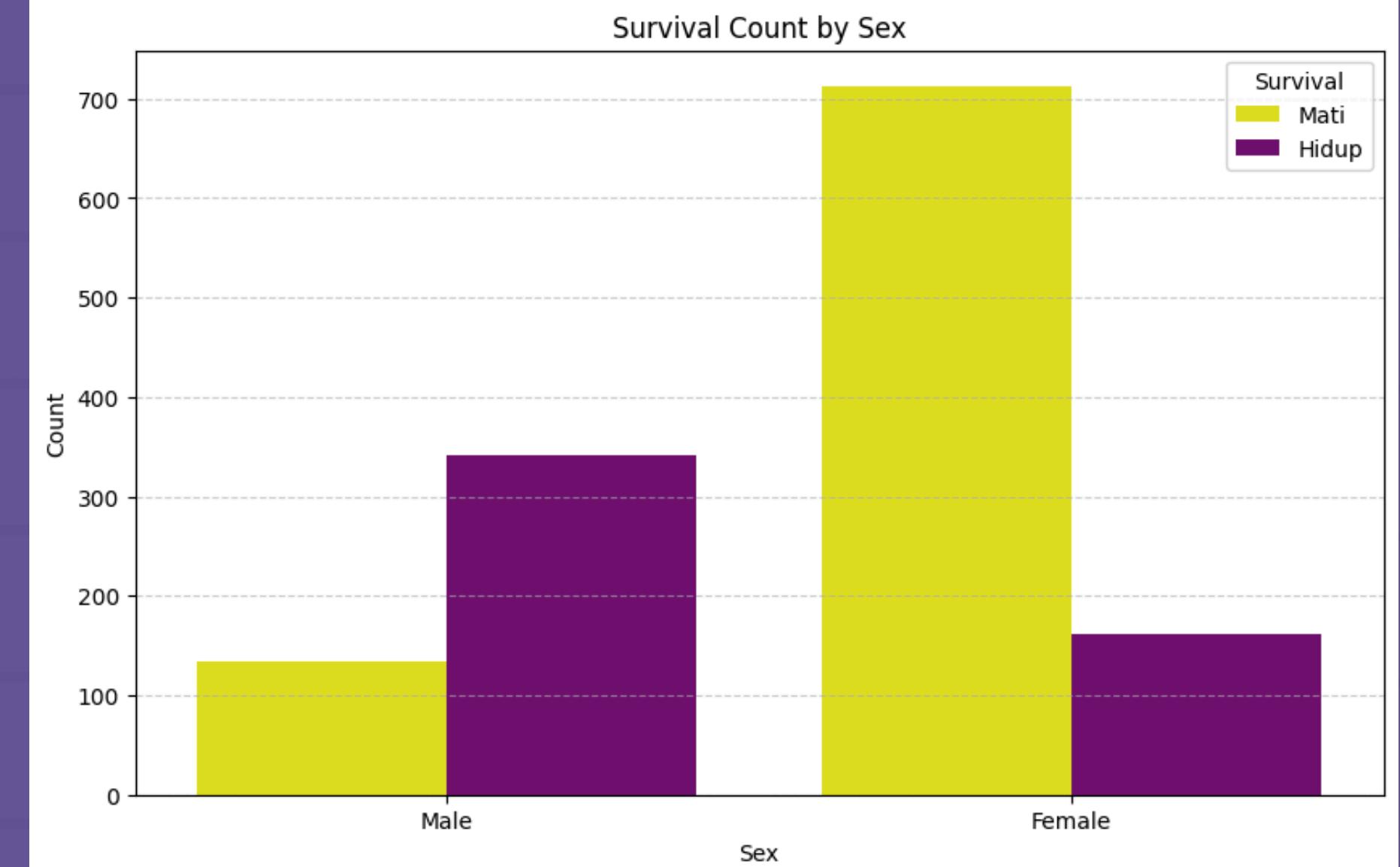
survived	0	1
pclass		
1	123	201
2	158	120
3	566	181

VISUALISASI DENGAN WARNA YANG BERBEDA UNTUK CLASS DIMANA SUMBU X=URUTAN DATA DAN SUMBU Y=SEX



```
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='sex',
               hue='survived', palette={0: 'yellow', 1: 'purple'})
plt.xlabel('Sex')
plt.ylabel('Count')
plt.title('Survival Count by Sex')
plt.legend(title='Survival', labels=['Mati', 'Hidup'])
plt.xticks(ticks=[0, 1], labels=['Male', 'Female'])
plt.grid(axis='y', linestyle='--', alpha=0.6)

plt.show()
```

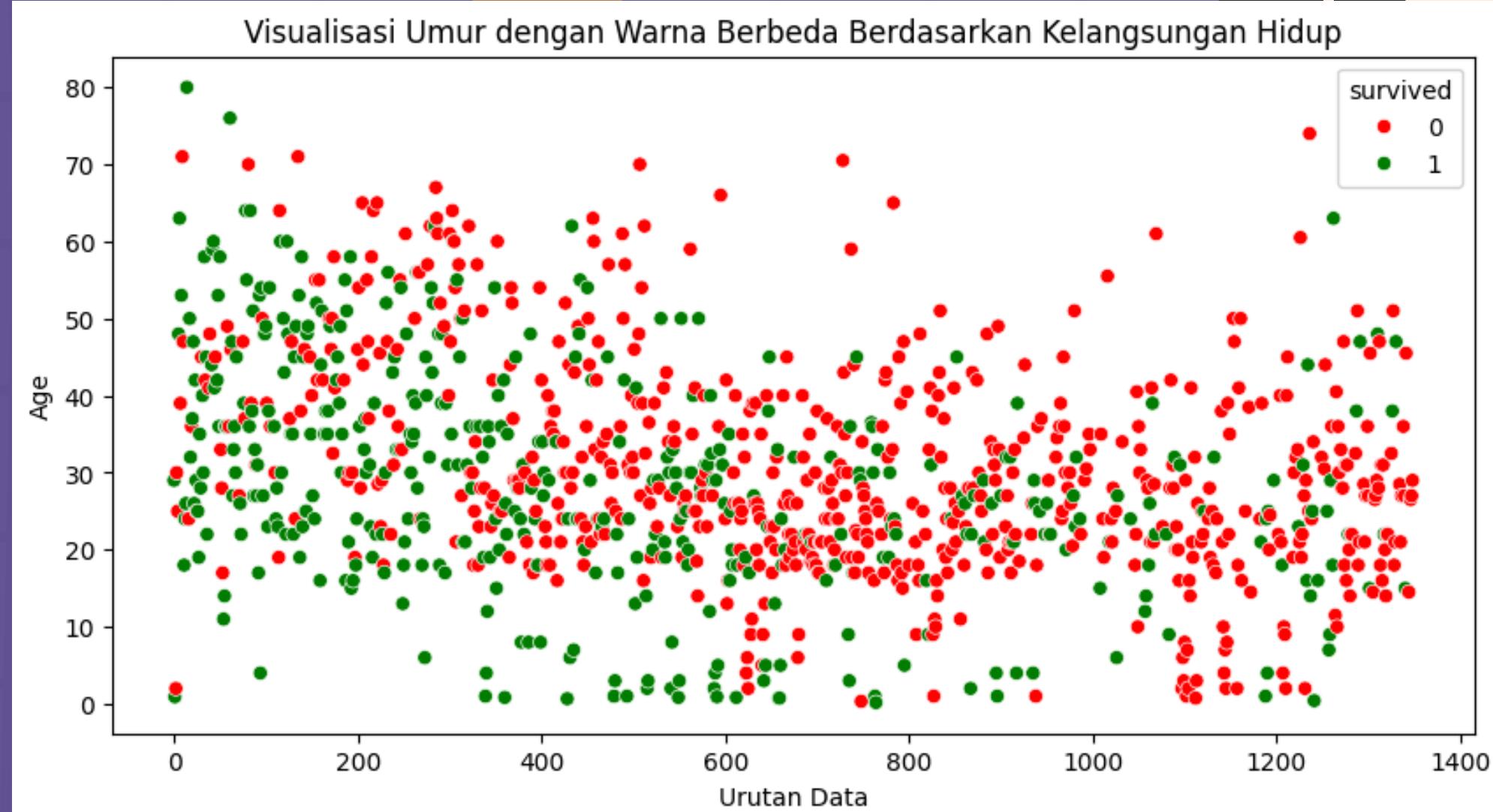


VISUALISASI DENGAN WARNA YANG BERBEDA UNTUK CLASS DIMANA
SUMBU X=URUTAN DATA DAN SUMBU Y=AGE (ABAIKAN DATA YANG
TERDAPAT MISSING VALUE)



```
df_cleaned = df.dropna(subset=['age'])

# Visualisasi dengan warna berbeda berdasarkan Survived
plt.figure(figsize=(10, 5))
sns.scatterplot(x=df_cleaned.index, y=df_cleaned['age'],
                 hue=df_cleaned['survived'], palette={0: "red", 1: "green"})
plt.xlabel("Urutan Data")
plt.ylabel("Age")
plt.title("Visualisasi Umur dengan Warna Berbeda Berdasarkan Kelangsungan Hidup")
plt.show()
```



TERIMA KASIH!

