PRAKTIKUM DATA MINING 2

FARIS SAIFULLAH
(3124640034)

NO.1

```python
import pandas as pd
dataset = pd.read_csv('titanic.csv')
print(dataset)
```

```
     PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
..           ...       ...     ...
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

                                                  Name     Sex   Age  SibSp  \
0                              Braund, Mr. Owen Harris    male  22.0      1
1    Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                               Heikkinen, Miss. Laina  female  26.0      0
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                             Allen, Mr. William Henry    male  35.0      0
..                                                 ...     ...   ...    ...
886                              Montvila, Rev. Juozas    male  27.0      0
887                       Graham, Miss. Margaret Edith  female  19.0      0
888           Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1
889                              Behr, Mr. Karl Howell    male  26.0      0
890                                Dooley, Mr. Patrick    male  32.0      0

     Parch            Ticket     Fare Cabin Embarked
0        0         A/5 21171   7.2500   NaN        S
1        0          PC 17599  71.2833   C85        C
2        0  STON/O2. 3101282   7.9250   NaN        S
3        0            113803  53.1000  C123        S
4        0            373450   8.0500   NaN        S
..     ...               ...      ...   ...      ...
886      0            211536  13.0000   NaN        S
887      0            112053  30.0000   B42        S
888      2        W./C. 6607  23.4500   NaN        S
889      0            111369  30.0000  C148        C
890      0            370376   7.7500   NaN        Q

[891 rows x 12 columns]
```

NO.2

```
import pandas as pd
dataset = pd.read_csv('titanic.csv')
rows, cols = dataset.shape
print('Jumlah baris', rows)
print('Jumlah kolom', cols)
```

```
Jumlah baris 891
Jumlah kolom 12
```

NO.3

```python
import pandas as pd
dataset = pd.read_csv('titanic.csv')
data = dataset[['Name', 'Sex', 'Age', 'Pclass', 'Fare']]
print(data)
```

```
                                                  Name     Sex   Age  Pclass  \
0                              Braund, Mr. Owen Harris    male  22.0       3
1    Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0       1
2                               Heikkinen, Miss. Laina  female  26.0       3
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0       1
4                             Allen, Mr. William Henry    male  35.0       3
..                                                 ...     ...   ...     ...
886                              Montvila, Rev. Juozas    male  27.0       2
887                       Graham, Miss. Margaret Edith  female  19.0       1
888           Johnston, Miss. Catherine Helen "Carrie"  female   NaN       3
889                              Behr, Mr. Karl Howell    male  26.0       1
890                                Dooley, Mr. Patrick    male  32.0       3

        Fare
0     7.2500
1    71.2833
2     7.9250
3    53.1000
4     8.0500
..       ...
886  13.0000
887  30.0000
888  23.4500
889  30.0000
890   7.7500

[891 rows x 5 columns]
```

NO.4

```
import pandas as pd
dataset = pd.read_csv('titanic.csv')
kelas = dataset[['Survived']]
print(kelas)
```

```
     Survived
0           0
1           1
2           1
3           1
4           0
..        ...
886         0
887         1
888         0
889         1
890         0

[891 rows x 1 columns]
```

NO.5

```python
data = pd.read_csv('titanic.csv')
data['Relatives'] = data['SibSp'] + data['Parch']
print(data.head())
```

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked  Relatives
0      0         A/5 21171   7.2500   NaN        S          1
1      0          PC 17599  71.2833   C85        C          1
2      0  STON/O2. 3101282   7.9250   NaN        S          0
3      0            113803  53.1000  C123        S          1
4      0            373450   8.0500   NaN        S          0
```

NO.6

```
data = pd.read_csv('titanic.csv')
penumpang_per_pclass = data.groupby('Pclass').size()
print(penumpang_per_pclass)
```

```
Pclass
1    216
2    184
3    491
dtype: int64
```

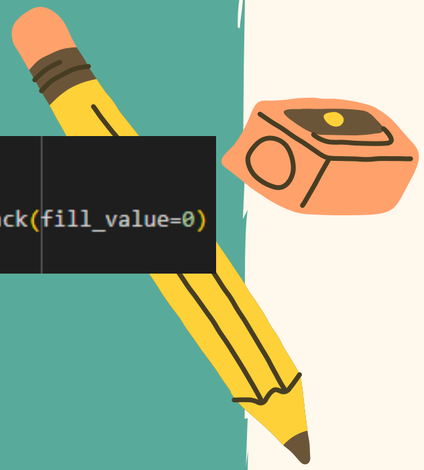
NO.7

```
data = pd.read_csv('titanic.csv')
penumpang_per_sex = data.groupby('Sex').size()
print(penumpang_per_sex)
```

```
Sex
female    314
male      577
dtype: int64
```

NO.8

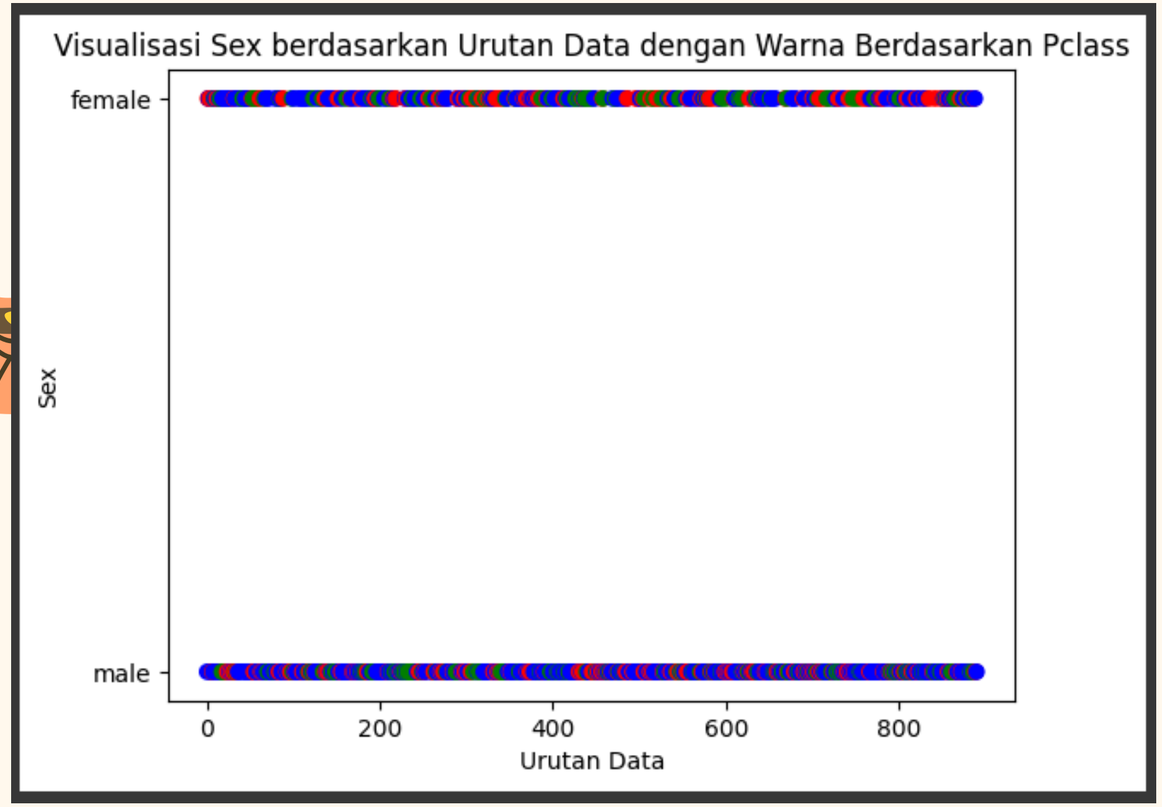
```python
import pandas as pd
dataset = pd.read_csv('titanic.csv')
pclass_survival_counts = dataset.groupby(['Pclass', 'Survived']).size().unstack(fill_value=0)
print(pclass_survival_counts)
```



Survived	0	1
Pclass		
1	80	136
2	97	87
3	372	119

NO.9

```python
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv('titanic.csv')
colors = {1: 'red', 2: 'green', 3: 'blue'}
x = data.index
y = data['Sex']
plt.scatter(x, y, c=data['Pclass'].map(colors))
plt.xlabel('Urutan Data')
plt.ylabel('Sex')
plt.title('Visualisasi Sex berdasarkan Urutan Data dengan Warna Berdasarkan Pclass')
plt.show()
```



NO.10

```python
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv('titanic.csv')
data_clean = data.dropna(subset=['Age'])
colors = {1: 'red', 2: 'green', 3: 'blue'}
x = data_clean.index
y = data_clean['Age']
plt.scatter(x, y, c=data_clean['Pclass'].map(colors))
plt.xlabel('Urutan Data')
plt.ylabel('Age')
plt.title('Visualisasi Age berdasarkan Urutan Data dengan Warna Berdasarkan Pclass')
plt.show()
```