Praktikum Data Preprocessing

Data Manipulation & Visualisation

Ulima Inas Shabrina(2110181048)

Assignment

1. dataset ← titanic.csv, dan tampilkan
2. rows, cols ← jumlah baris dan kolom pada dataset, dan tampilkan
3. data ← ambil dataset kolom fitur (Name, Sex, Age, Pclass, Fare), dan tampilkan
4. class ← ambil dataset kolom kelas (Survived), dan tampilkan
5. data ← data + fitur Relatives (jumlah akumulasi dari kolom fitur SibSp+Parch dari dataset), dan tampilkan
6. Hitunglah berapa penumpang pada masing-masing nilai Pclass
7. Hitunglah berapa penumpang pada masing-masing nilai Sex
8. Hitunglah berapa penumpang yang tidak selamat (Survived=0) dan selamat (Survived=1) untuk masing-masing Pclass
9. Visualisasi dengan warna yang berbeda untuk class dimana sumbu x=urutan data dan sumbu y=Sex
10. Visualisasi dengan warna yang berbeda untuk class dimana sumbu x=urutan data dan sumbu y=Age (abaikan data yang terdapat missing value)

Menampilkan dataset titanic.csv menggunakan tools Jupyter Notebook

In [1]:
```
import pandas as pd

dataset = pd.read_csv('titanic.csv')

dataset
```

Out[1]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

Menampilkan jumlah baris dan kolom pada dataset

In [2]:
```
rows, cols = dataset.shape

print('Jumlah Baris', rows)
print('Jumlah Kolom', cols)
```

```
Jumlah Baris 891
Jumlah Kolom 12
```

Menampilkan dataset kolom fitur Name, Sex, Age, Pclass, Fare

In [3]:
```
data = pd.DataFrame(dataset, columns = ['Name', 'Sex', 'Age', 'Pclass', 'Fare'])

data
```

Out[3]:

	Name	Sex	Age	Pclass	Fare
0	Braund, Mr. Owen Harris	male	22.0	3	7.2500
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	71.2833
2	Heikkinen, Miss. Laina	female	26.0	3	7.9250
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	53.1000
4	Allen, Mr. William Henry	male	35.0	3	8.0500
...	...	...	...	...	...
886	Montvila, Rev. Juozas	male	27.0	2	13.0000
887	Graham, Miss. Margaret Edith	female	19.0	1	30.0000
888	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	3	23.4500
889	Behr, Mr. Karl Howell	male	26.0	1	30.0000
890	Dooley, Mr. Patrick	male	32.0	3	7.7500

891 rows × 5 columns

Menambahkan dataset kolom kelas Survived

In [4]:
```
survived_task = pd.DataFrame(dataset, columns = ['Survived'])

survived_task
```

Out[4]:

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

891 rows × 1 columns

Menampilkan data dan fitur relatives(akumulasi SibSp + Parch)

In [5]:
```
data['Relatives'] = dataset['SibSp'] + dataset['Parch']

data
```

Out[5]:

	Name	Sex	Age	Pclass	Fare	Relatives
0	Braund, Mr. Owen Harris	male	22.0	3	7.2500	1
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	71.2833	1
2	Heikkinen, Miss. Laina	female	26.0	3	7.9250	0
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	53.1000	1
4	Allen, Mr. William Henry	male	35.0	3	8.0500	0
...	...	...	...	...	...	...
886	Montvila, Rev. Juozas	male	27.0	2	13.0000	0
887	Graham, Miss. Margaret Edith	female	19.0	1	30.0000	0
888	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	3	23.4500	3
889	Behr, Mr. Karl Howell	male	26.0	1	30.0000	0
890	Dooley, Mr. Patrick	male	32.0	3	7.7500	0

891 rows × 6 columns

Menghitung berapa penumpang pada masing masing nilai Pclass

```
In [15]: dataset['Pclass'].value_counts()
```

```
Out[15]: 3    491
         1    216
         2    184
         Name: Pclass, dtype: int64
```

Menghitung berapa penumpang pada masing masing nilai Sex

```
In [16]: dataset['Sex'].value_counts()
```

```
Out[16]: male      577
         female    314
         Name: Sex, dtype: int64
```

Menampilkan jumlah penumpang yang tidak selamat dan selamat untuk masing masing Pclass

In [27]:
```
dataset.groupby(['Pclass', 'Survived']).size().reset_index(name='counts')
```

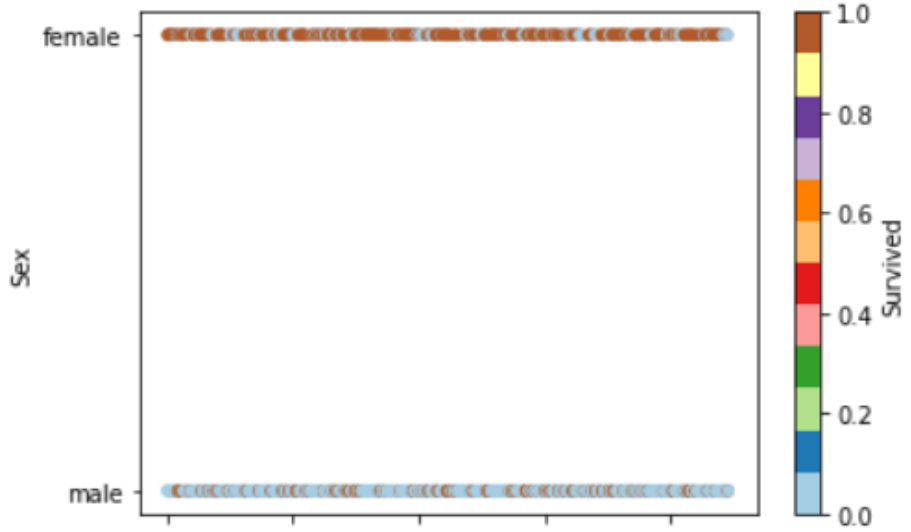
Out[27]:

	Pclass	Survived	counts
0	1	0	80
1	1	1	136
2	2	0	97
3	2	1	87
4	3	0	372
5	3	1	119

Memvisualisasikan dengan warna yang berbeda untuk class dimana sumbu X = urutan Data dan sumbu Y = Sex

In [36]:
```
dataset.plot(x='PassengerId', y='Sex', kind='scatter', c='Survived', colormap='Paired')
```

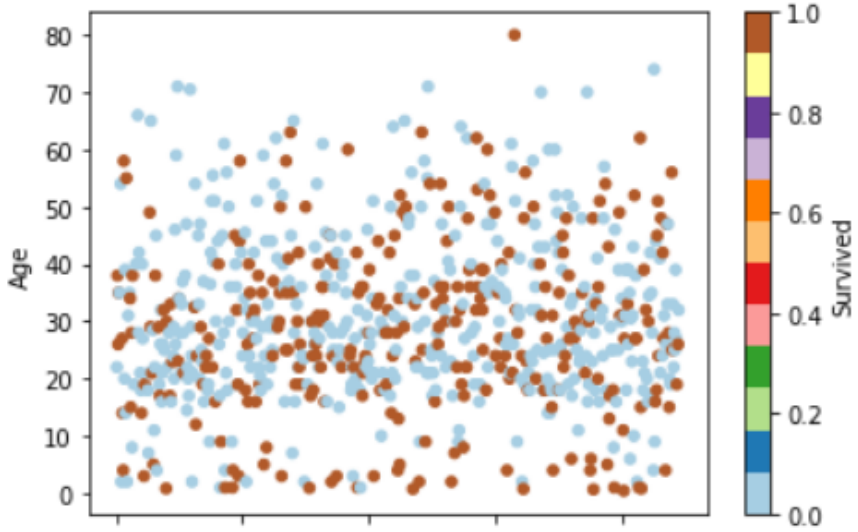
Out[36]: <AxesSubplot:xlabel='PassengerId', ylabel='Sex'>



Memvisualisasikan dengan warna yang berbeda untuk class dimana sumbu X = urutan Data dan sumbu Y = Age

```
In [37]: dataset.plot(x='PassengerId', y='Age', kind='scatter', c='Survived', colormap='Paired')
```

```
Out[37]: <AxesSubplot:xlabel='PassengerId', ylabel='Age'>
```



Terimakasih