

**LAPORAN**  
**PRAKTIKUM DATA WAREHOUSING DAN DATA MINING**  
**( MODUL 7 )**



**Disusun oleh :**

**NAMA : CINDI DILA APRILIANA**

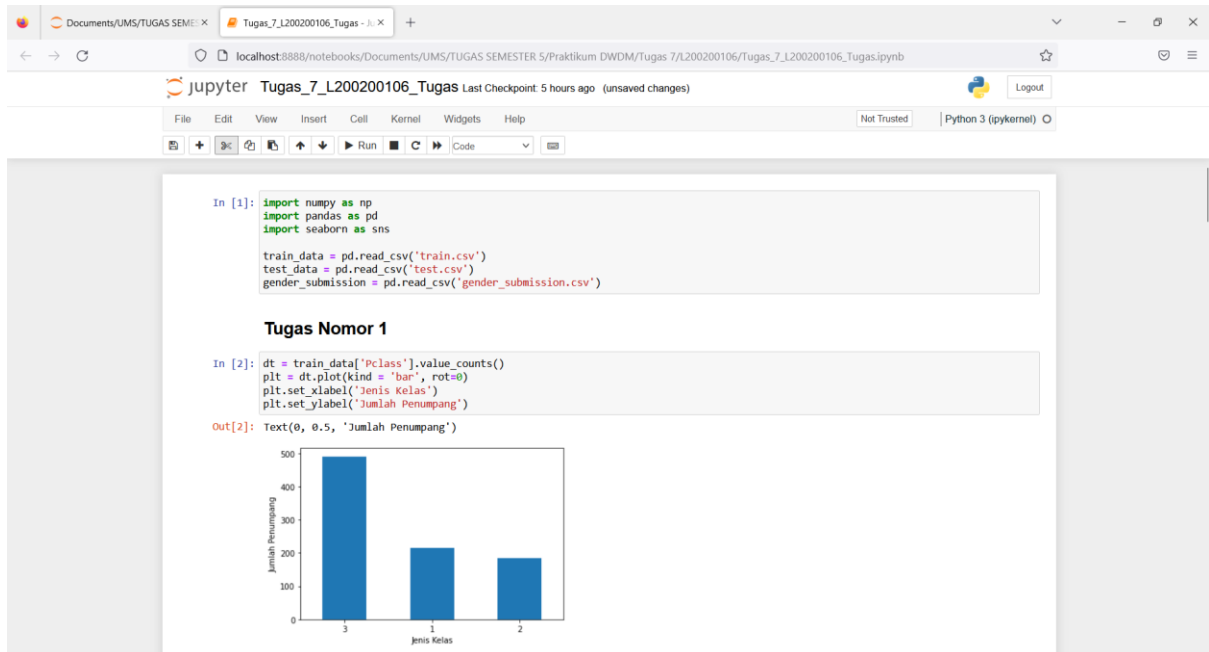
**NIM : L200200106**

**KELAS : D**

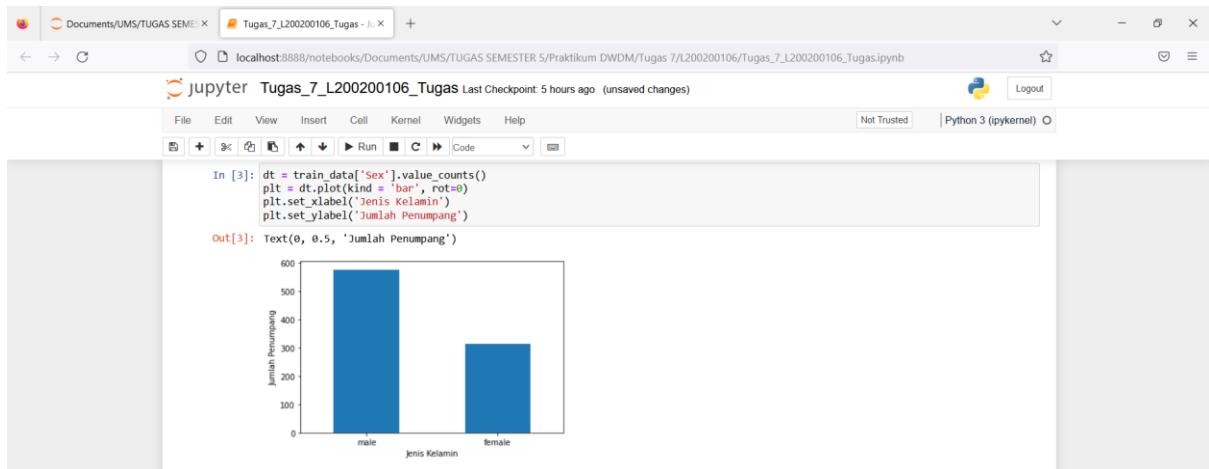
**INFORMATIKA**  
**FAKULTAS KOMUNIKASI DAN INFORMATIKA**  
**UNIVERSITAS MUHAMMADIYAH SURAKARTA**  
**TAHUN 2022/2023**

1. Lakukan kembali langkah 13 pada prosedur praktikum untuk melihat data atribut lainnya dengan grafik batang

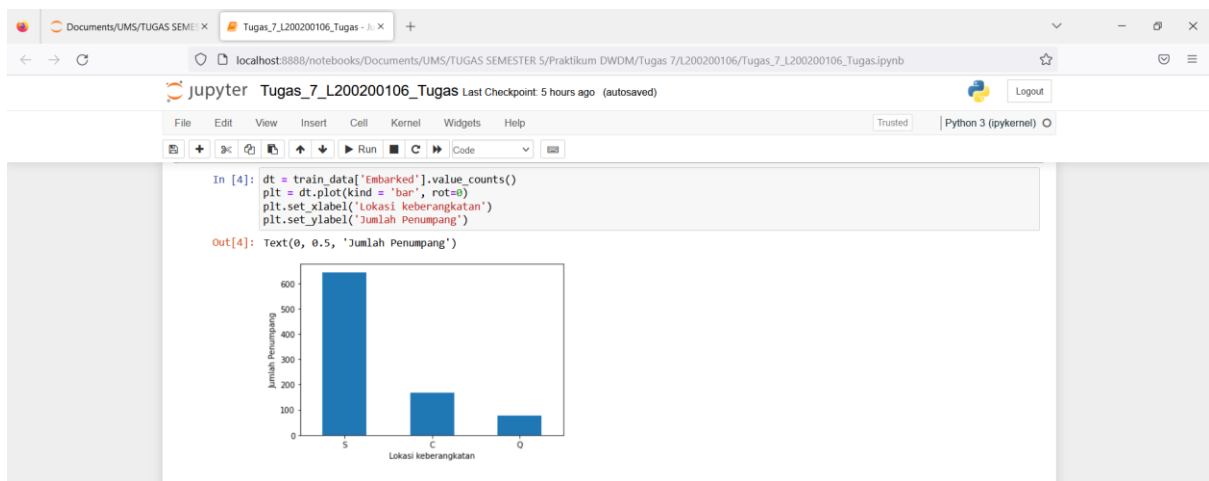
- **Pclass**



- **Sex**



- **Embarked**



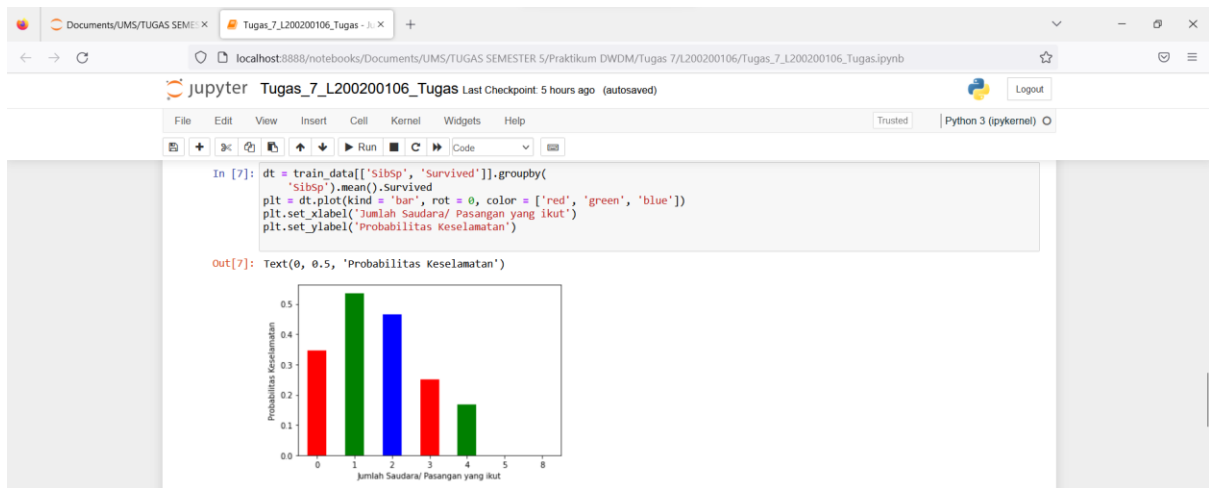
2. Lakukan kembali langkah 14 pada prosedur praktikum untuk melihat probabilitas keselamatan (Survived) berdasarkan :
  - **jenis kelamin (Sex)**



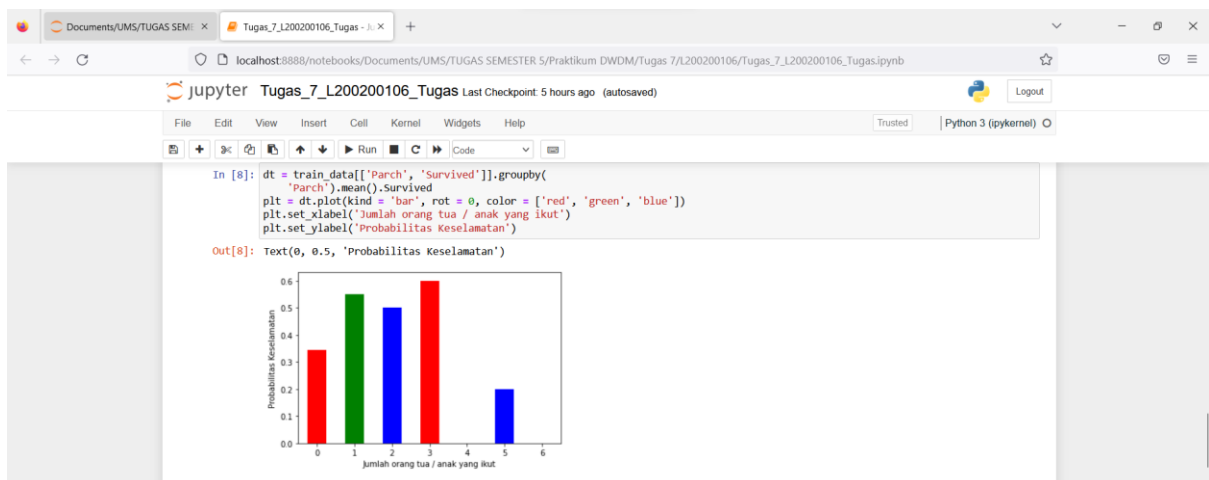
- lokasi keberangkatan (Embarked)



- Jumlah saudara/pasangan yang ikut (SibSp)

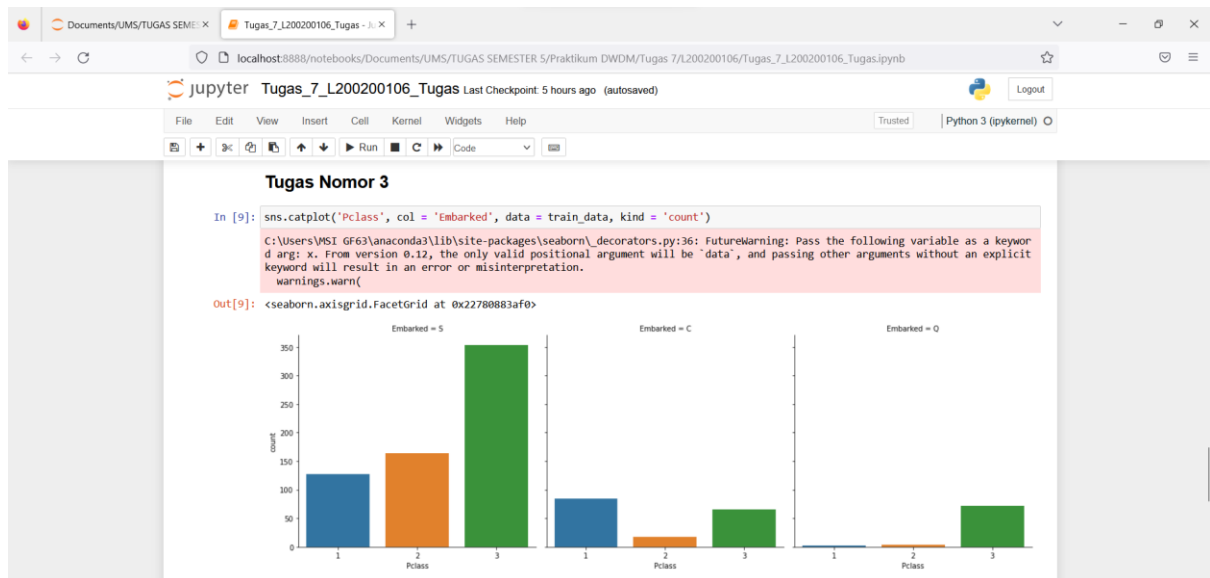


- Jumlah orang tua/anak yang ikut (Parch)

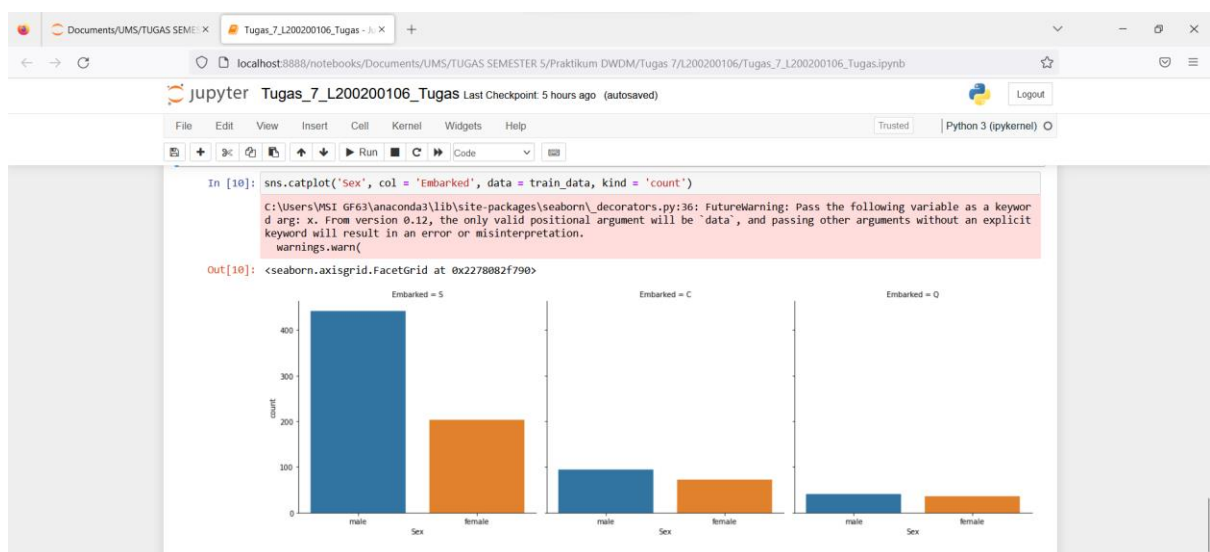


3. Ulangi kembali langkah 15 pada prosedur praktikum untuk melihat multidimensi terhadap atribut

- **Kelas penumpang (Pclass) vs lokasi keberangkatan (Embarked)**



- **Jenis kelamin (Sex) vs lokasi keberangkatan (Embarked)!**



- Ubahlah data sebutan/gelar penumpang (Title) menjadi data angka dengan ketentuan sebagai berikut: Master: 0, Miss: 1, Mr: 2, Mrs: 3, Others: 4

**Tugas Nomer 4**

```
In [11]: train_data['Title'] = train_data['Name'].str.extract('([A-Za-z]+)\.', expand=False)
train_data = train_data.drop(columns='Name')

train_data['Title'] = train_data['Title'].replace(['Dr', 'Rev', 'Col', 'Major', 'Countess', 'Sir', 'Jonkheer', 'Lady', 'Capt', 'Don'], 'ot')
train_data['Title'] = train_data['Title'].replace('Ms', 'Miss')
train_data['Title'] = train_data['Title'].replace('Mme', 'Mrs')
train_data['Title'] = train_data['Title'].replace('Mlle', 'Miss')
```

```
In [12]: train_data.head()
```

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Title
0	1	0	3	male	22.0	1	0	A/5 21171	7.2500	NaN	S	Mr
1	2	1	1	female	38.0	1	0	PC 17599	71.2833	C85	C	Mrs
2	3	1	3	female	26.0	0	0	STON/O2 3101282	7.9250	NaN	S	Miss
3	4	1	1	female	35.0	1	0	113803	53.1000	C123	S	Mrs
4	5	0	3	male	35.0	0	0	373450	8.0500	NaN	S	Mr

```
In [13]: train_data['Title'].unique()
Out[13]: array(['Mr', 'Mrs', 'Miss', 'Master', 'others'], dtype=object)
```

```
In [14]: train_data['Title'] = train_data['Title'].map({'Master':0, 'Miss':1, 'Mr':2, 'Mrs':3, 'others':4})
train_data.head()
```

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Title
0	1	0	3	male	22.0	1	0	A/5 21171	7.2500	NaN	S	2
1	2	1	1	female	38.0	1	0	PC 17599	71.2833	C85	C	3
2	3	1	3	female	26.0	0	0	STON/O2 3101282	7.9250	NaN	S	1
3	4	1	1	female	35.0	1	0	113803	53.1000	C123	S	3
4	5	0	3	male	35.0	0	0	373450	8.0500	NaN	S	2

- Carilah nilai korelasi antar atribut termasuk atribut Title setelah diubah menjadi data angka dengan menggunakan heatmap

**Tugas Nomer 5**

```
In [15]: corr_matrix = train_data.corr()
```

```
In [16]: import matplotlib.pyplot as plt
corr_matrix = train_data.corr()
plt.figure(figsize=(9,8))
sns.heatmap(data = corr_matrix, cmap='BrBG', annot=True, linewidths=0.2)
```

```
Out[16]: <AxesSubplot>
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Title
PassengerId	1.000000	-0.005	-0.035	0.037	-0.058	-0.0017	0.013	0.079
Survived	-0.005	1.000000	0.34	-0.077	-0.035	0.082	0.26	-0.071
Pclass	-0.035	0.34	1.000000	-0.37	0.083	0.018	-0.55	-0.18
Age	0.037	-0.077	-0.37	1.000000	-0.31	-0.19	0.096	0.51
SibSp	-0.058	-0.035	0.083	-0.31	1.000000	0.41	0.16	-0.21
Parch	-0.0017	0.082	0.018	-0.19	0.41	1.000000	0.22	-0.12
Fare	0.013	0.26	-0.55	0.096	0.16	0.22	1.000000	-0.013
Title	0.079	-0.071	-0.18	0.51	-0.21	-0.12	-0.013	1.000000