

LAPORAN
PRAKTIKUM DATA WAREHOUSING DAN DATA MINING
(MODUL 7)



Disusun oleh :

NAMA : DIEGO ALFIANDRO

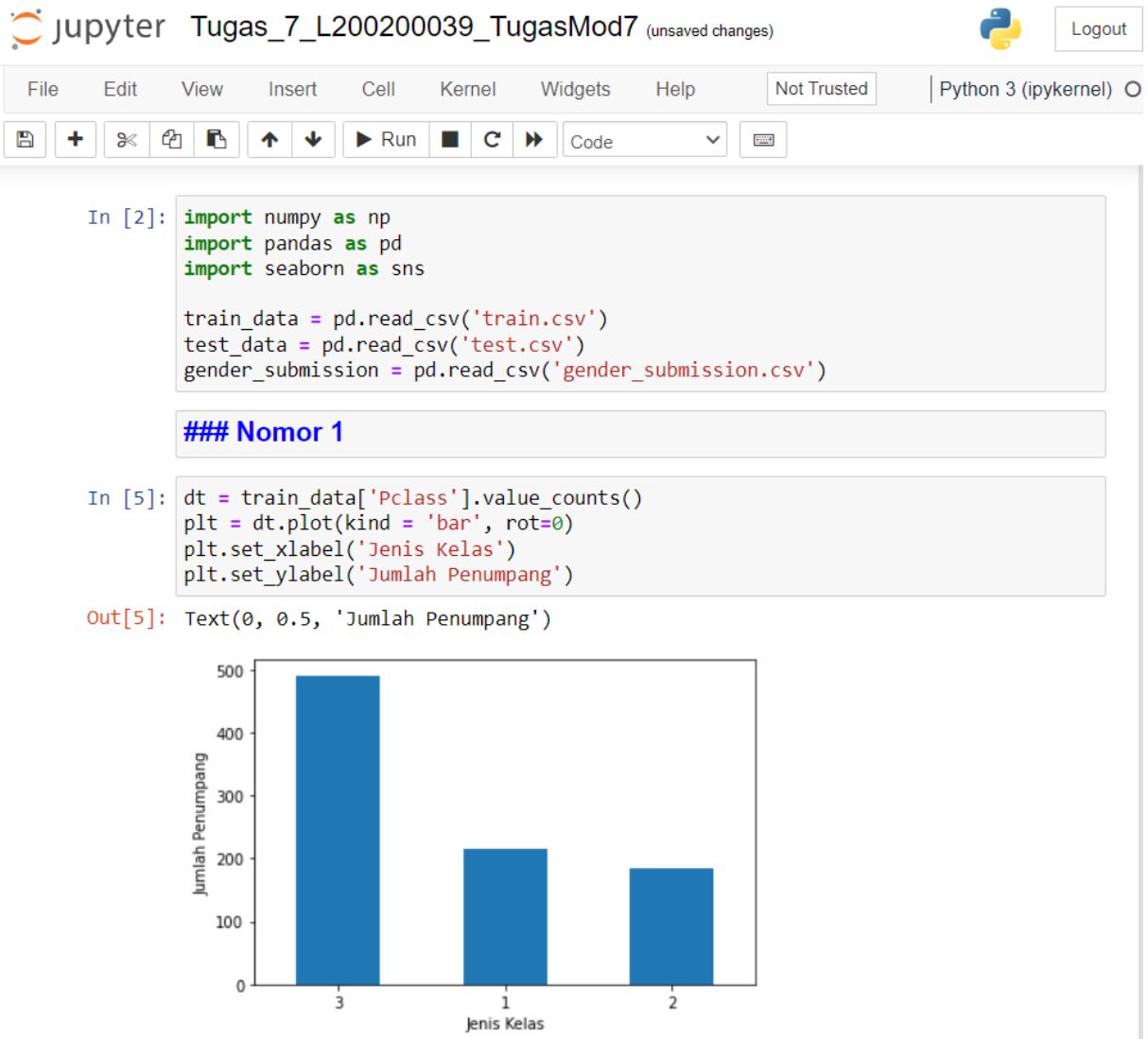
NIM : L200200039

KELAS : D

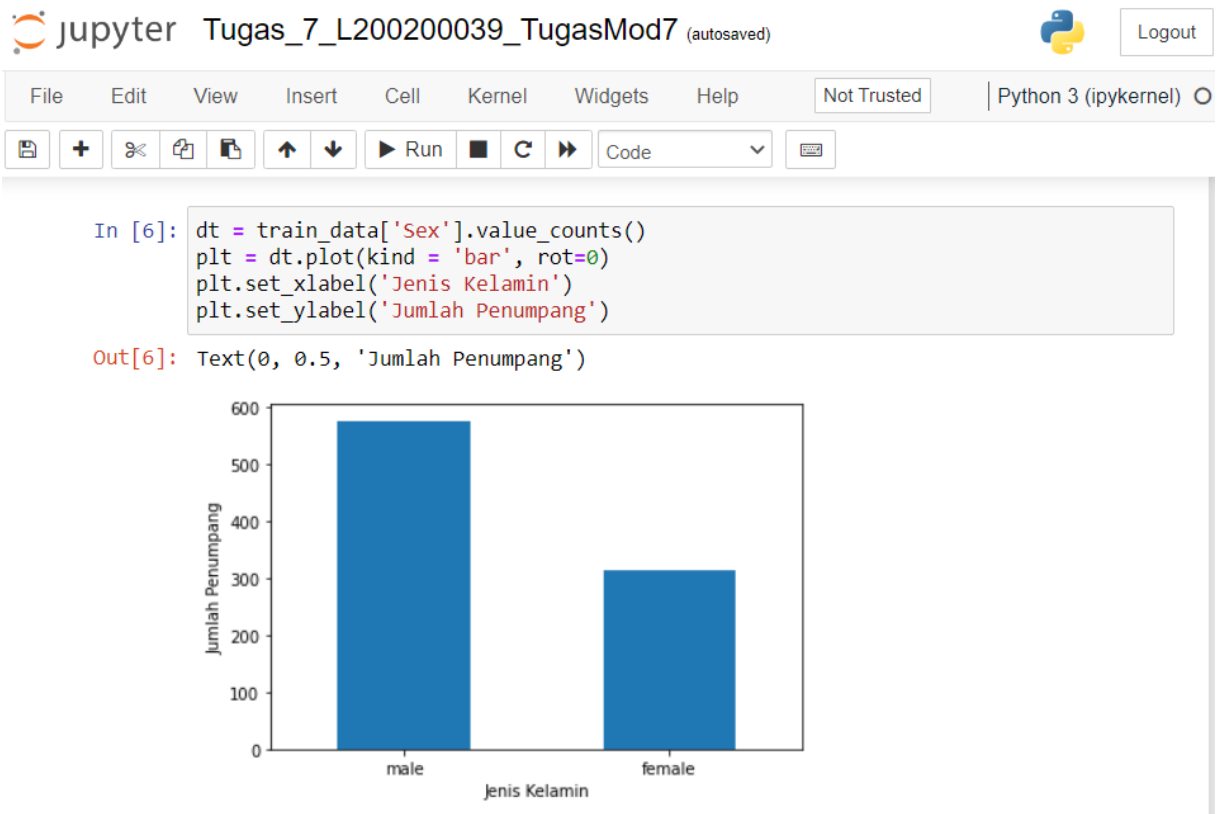
INFORMATIKA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH SURAKARTA
TAHUN 2022/2023

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

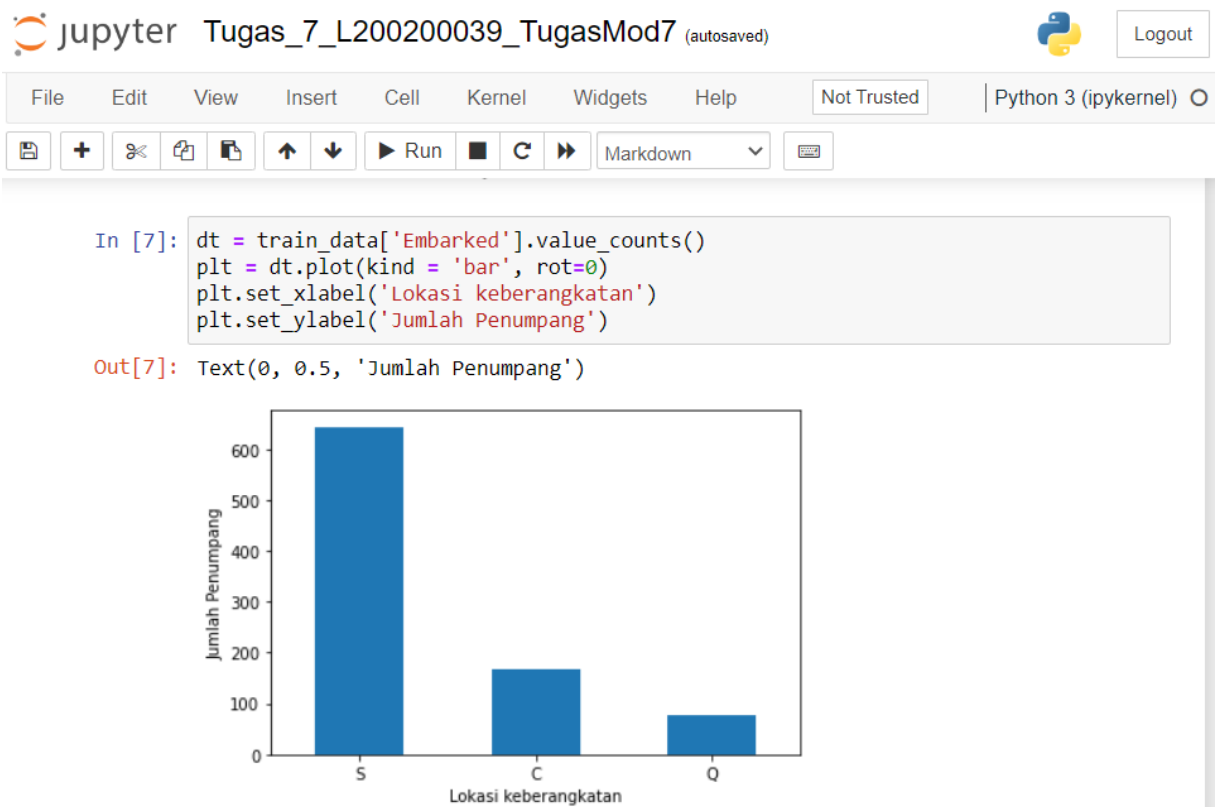
- **Pclass**



- Sex

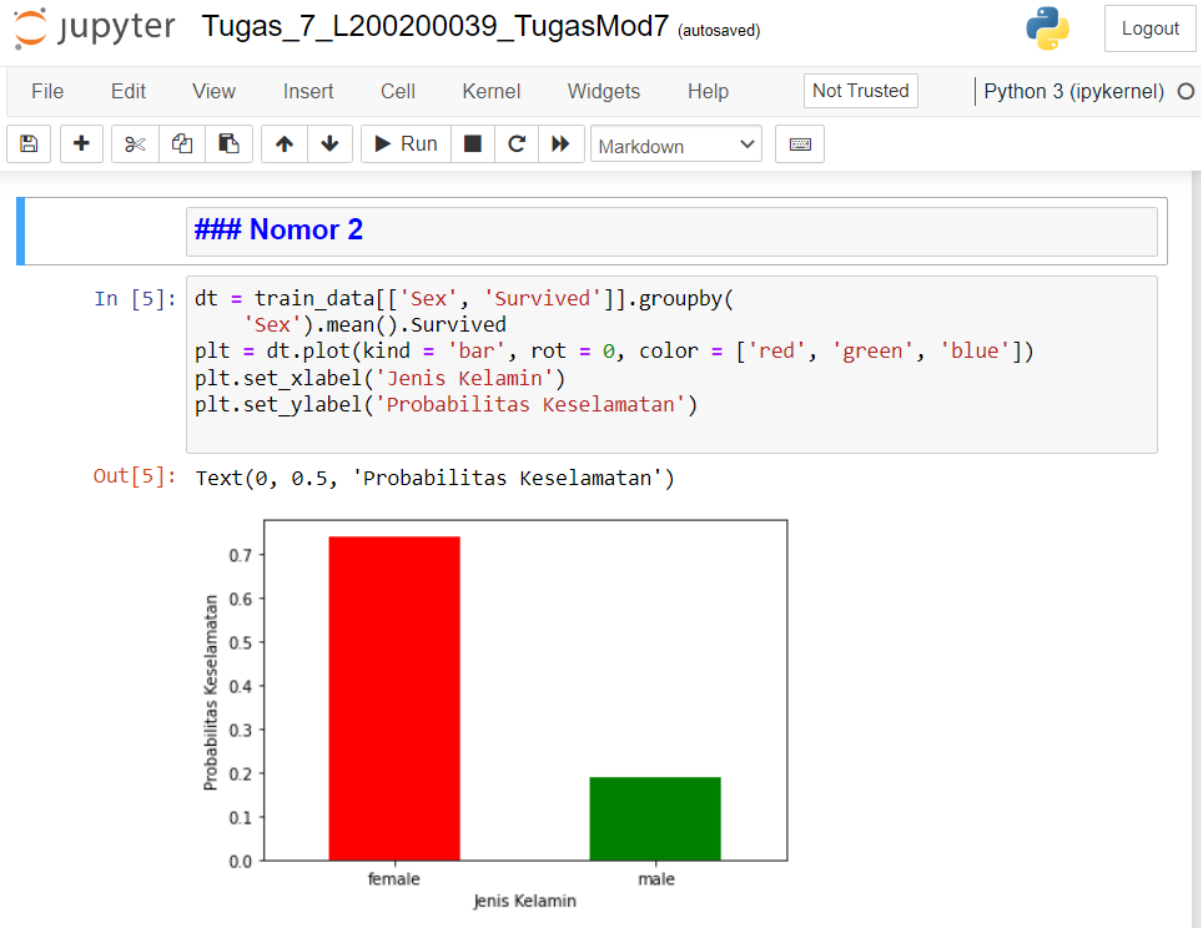


- Embarked

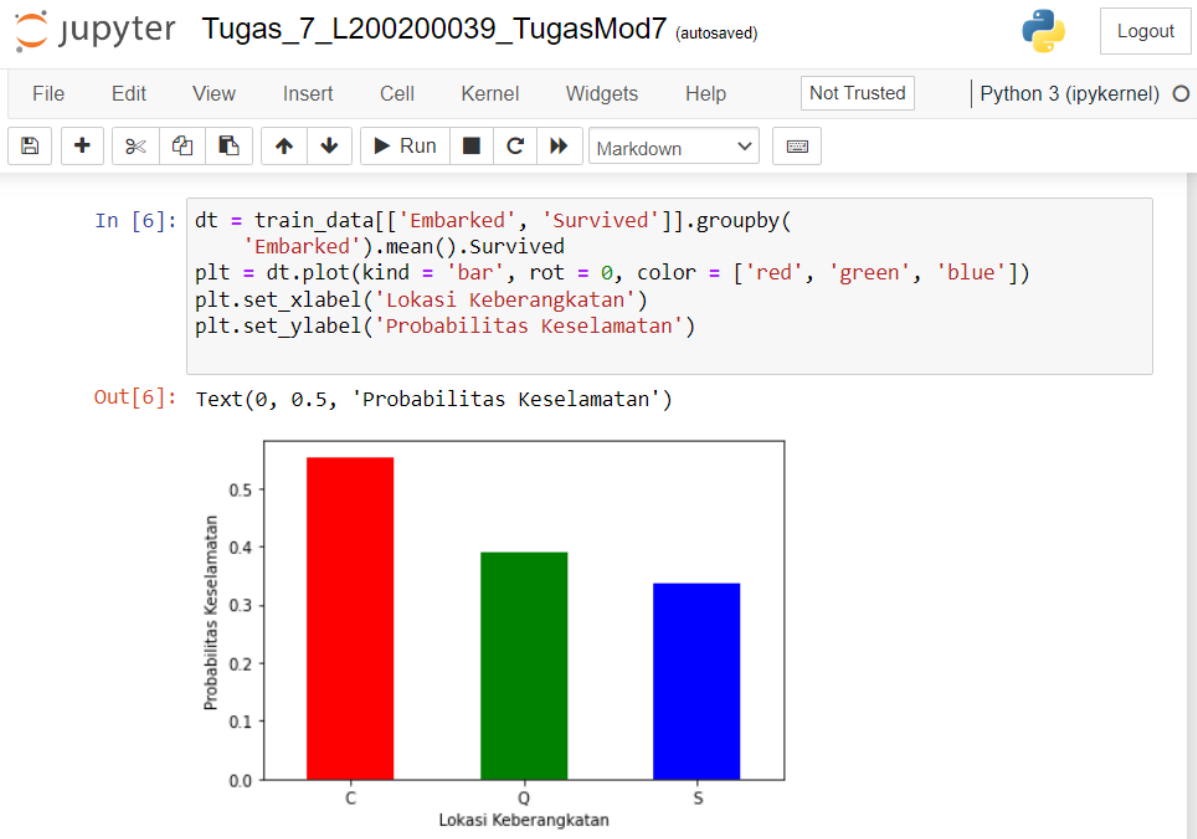


2. Lakukan kembali langkah 14 pada prosedur praktikum untuk melihat probabilitas keselamatan (Survived) berdasarkan :

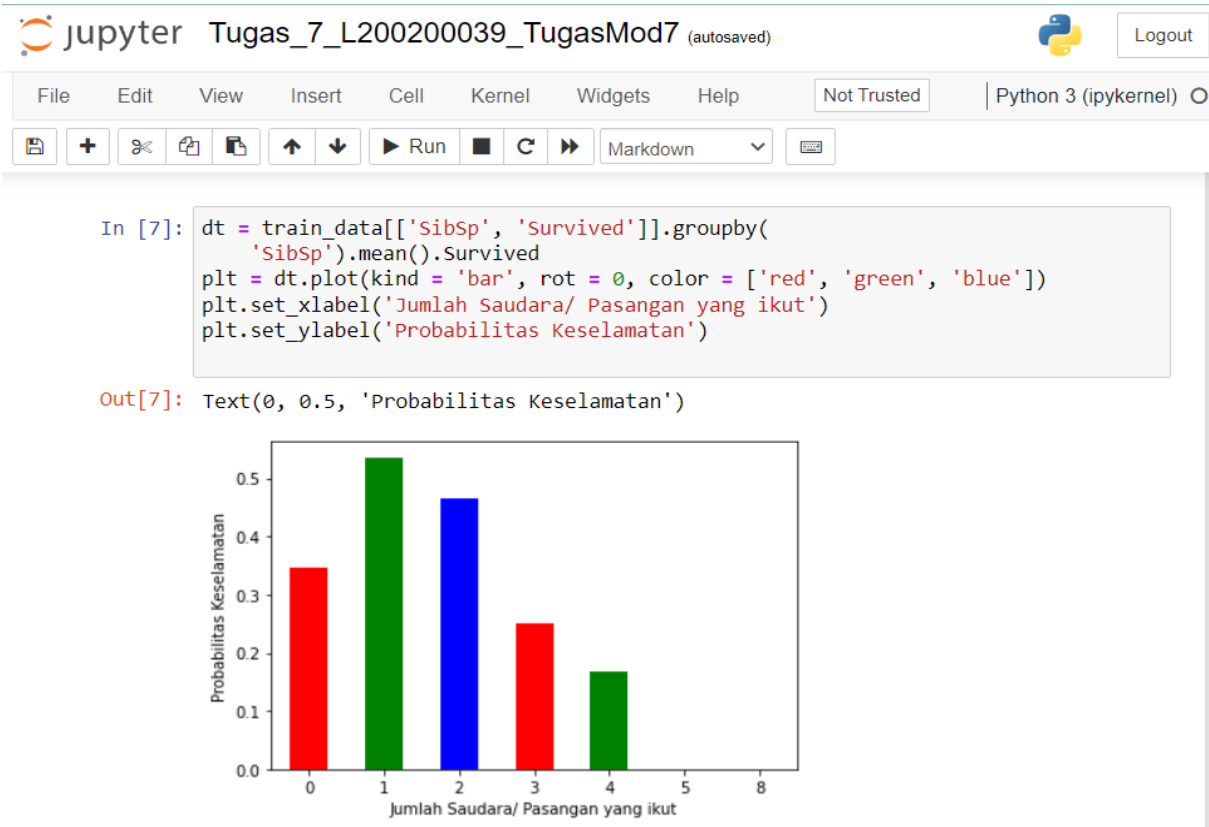
- **jenis kelamin (Sex)**



- lokasi keberangkatan (Embarked)



- Jumlah saudara/pasangan yang ikut (SibSp)



- **Jumlah orang tua/anak yang ikut (Parch)**

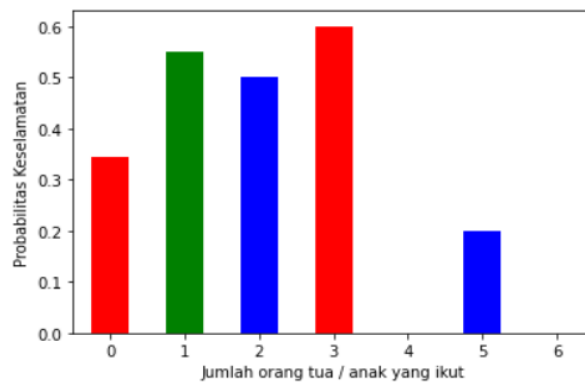
Jupyter Tugas_7_L200200039_TugasMod7 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Save + Close Copy Paste Up Down Run Stop Restart Markdown

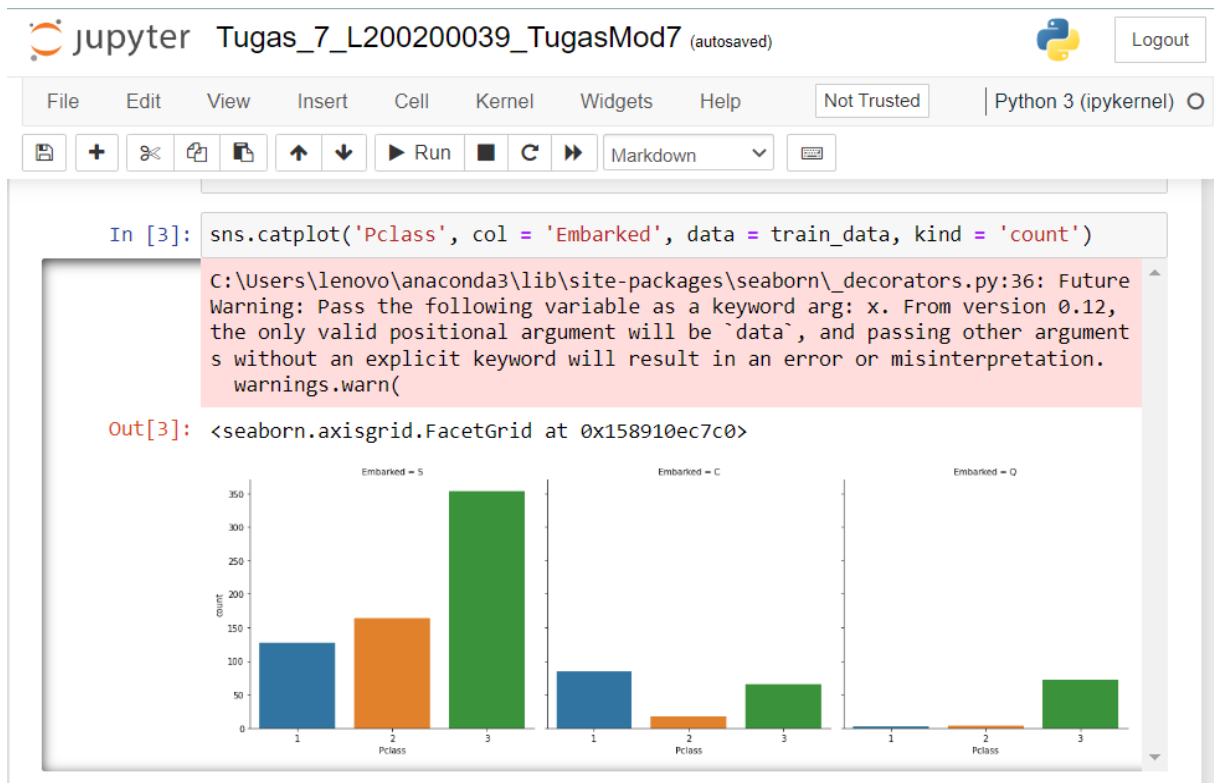
```
In [8]: dt = train_data[['Parch', 'Survived']].groupby(
        'Parch').mean().Survived
        plt = dt.plot(kind = 'bar', rot = 0, color = ['red', 'green', 'blue'])
        plt.set_xlabel('Jumlah orang tua / anak yang ikut')
        plt.set_ylabel('Probabilitas Keselamatan')
```

Out[8]: Text(0, 0.5, 'Probabilitas Keselamatan')

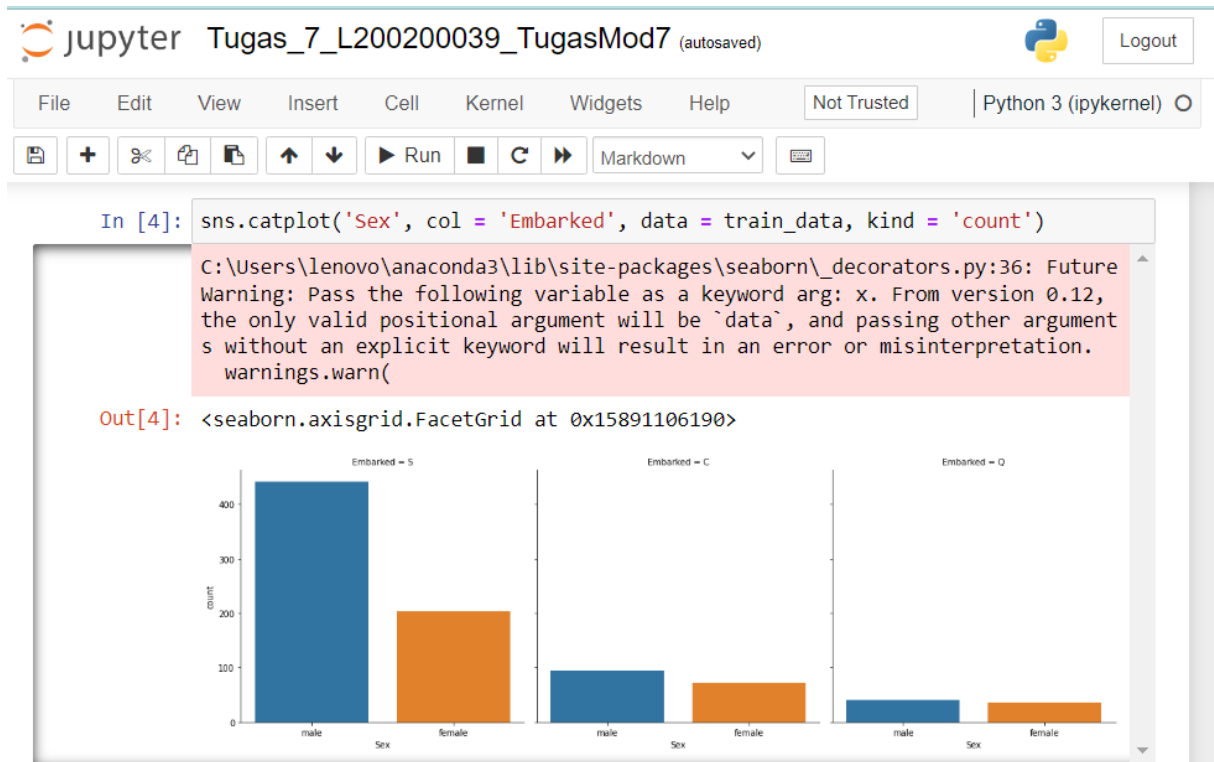


3. Ulangi kembali langkah 15 pada prosedur praktikum untuk melihat multidimensi terhadap atribut

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



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



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

Jupyter Tugas_7_L200200039_TugasMod7 (autosaved) Python 3 (ipykernel)

File Edit View Insert Cell Kernel Widgets Help Not Trusted

Run

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'], 'other')
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

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

