

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from warnings import filterwarnings
filterwarnings(action="ignore")

In [2]: pd.set_option('display.max_columns',10,'display.width',1000)
train = pd.read_csv("C:\Users\hagendra.Varma\Documents\train.csv")
test = pd.read_csv("C:\Users\hagendra.Varma\Documents\test.csv")
train.head()
```

Out[2]:

	PassengerId	Survived	Pclass		Name	Sex	...	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3									
1	2	1	1	Cummings, Mrs. John Bradley	Florence Briggs Th.	female	...	0	A/5 21171	7.2500	NaN	S
2	3	1	3		Heikkinen, Mrs. Laina	female	...	0	STON/O2 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath	(Lily May Peck)	female	...	0	113803	53.1000	C123	S
4	5	0	3		Allen, Mr. William Henry	male	...	0	373450	8.0500	NaN	S
5 rows x 12 columns												

```
In [3]: train.shape
Out[3]: (891, 12)
In [4]: test.shape
Out[4]: (418, 11)
In [5]: train.isnull().sum()
Out[5]: PassengerId    0
Survived          0
Pclass            0
Name              0
Sex              177
Age              177
SibSp            0
Parch            0
Fare             0
Cabin           687
Embarked         2
dtype: int64
In [6]: test.isnull().sum()
Out[6]: PassengerId    0
Pclass            0
Name              0
Sex              86
Age              86
SibSp            0
Parch            0
Ticket           0
Fare             1
Cabin           327
Embarked         0
dtype: int64
In [7]: train.describe(include="all")
Out[7]:
```

	PassengerId	Survived	Pclass		Name	Sex	...	Parch	Ticket	Fare	Cabin	Embarked
count	891.000000	891.000000	891.000000		891	891	...	891.000000	891	891.000000	204	889
unique	NaN	NaN	NaN		891	2	...	NaN	347062	NaN	NaN	3
top	NaN	NaN	NaN	NaN	Braund, Mr. Owen Harris	male	...	NaN	347062	NaN	B96 B96	S
freq	NaN	NaN	NaN	NaN	1	577	...	NaN	7	NaN	4	644
mean	446.000000	0.383838	2.308642		NaN	NaN	...	0.381594	NaN	32.204208	NaN	NaN
std	257.353842	0.486902	0.836071		NaN	NaN	...	0.806067	NaN	49.693429	NaN	NaN
min	1.000000	0.000000	1.000000		NaN	NaN	...	0.000000	NaN	0.000000	NaN	NaN
25%	223.500000	0.000000	2.000000		NaN	NaN	...	0.000000	NaN	7.910400	NaN	NaN
50%	446.000000	0.000000	3.000000		NaN	NaN	...	0.000000	NaN	14.454200	NaN	NaN
75%	668.500000	1.000000	3.000000		NaN	NaN	...	0.000000	NaN	31.000000	NaN	NaN
max	891.000000	1.000000	3.000000		NaN	NaN	...	6.000000	NaN	512.329200	NaN	NaN
15 rows x 12 columns												

```
In [8]: import numpy as np
numeric_columns=train.select_dtypes(include=[np.number]).columns
train(numeric_columns).groupby('Survived').mean()
Out[8]:
```

	Survived
0	447.016393 2.531876 30.626179 0.553734 0.329690 22.117887
1	444.359421 1.902092 28.343690 0.473694 0.464912 48.385408

```
In [9]: numeric_columns=train.select_dtypes(include=[np.number])
correlation_matrix=numeric_columns.corr()
Out[9]:
```

	PassengerId	Survived	Pclass		Age	SibSp		Parch	Fare
PassengerId	1.000000	-0.000007	-0.030144		0.036847	-0.057527		-0.001652	0.012598
Survived	-0.000007	1.000000	-0.338481		-0.372221	-0.035322		0.081629	0.257307
Pclass	-0.030144	-0.338481	1.000000		-0.360226	0.083081		0.018443	-0.549600
Age	0.036847	-0.077221	-0.360226		1.000000	0.308247		-0.189119	0.096067
SibSp	-0.057527	-0.035322	0.083081		0.308247	1.000000		0.414030	0.159651
Parch	-0.001652	0.081629	0.018443		-0.189119	0.414030		1.000000	0.216225
Fare	0.012598	0.257307	-0.549600		0.096067	0.159651		0.216225	1.000000

```
In [10]: male_ind = len(train[train['Sex'] == 'male'])
print("No of Males in Titanic:", male_ind)
No of Males in Titanic: 577
In [11]: female_ind = len(train[train['Sex'] == 'female'])
print("No of females in Titanic:", female_ind)
No of Females in Titanic: 314
In [12]: fig = plt.figure()
ax = fig.add_subplot(2,1,1)
gender = ['Male','Female']
index = [577,314]
ax.bar(gender,index)
plt.xlabel('Gender')
plt.ylabel('No of people onboarding ship')
plt.show()
```

Out[12]:

```
In [13]: alive = len(train[train['Survived'] == 1])
dead = len(train[train['Survived'] == 0])
In [14]: train.groupby('Sex')[['Survived']].mean()
Out[14]:
```

	Survived
Sex	
female	0.742038
male	0.188808

```
In [15]: fig = plt.figure()
ax = fig.add_subplot(2,1,1)
status = ['Survived','Dead']
ind = [alive,dead]
ax.bar(status,ind)
plt.xlabel('Status')
plt.ylabel('No of people onboarding ship')
plt.show()
```

Out[15]:

```
In [16]: plt.figure()
train.loc[train['Survived'] == 1, 'Pclass'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people survive')
plt.figure()
train.loc[train['Survived'] == 0, 'Pclass'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[16]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[16]:

```
In [17]: plt.figure()
age = train.loc[train.Survived == 1, 'Age']
plt.title('The histogram of the age groups of the people that had survived')
plt.hist(age, np.arange(0,100,10))
plt.xticks(np.arange(0,100,10))
plt.figure()
age = train.loc[train.Survived == 0, 'Age']
plt.title('The histogram of the age groups of the people that couldn't survive')
plt.hist(age, np.arange(0,100,10))
plt.xticks(np.arange(0,100,10))
Out[17]:
```

Out[17]:

```
In [18]: train[train['Survived'] == 1, 'Pclass'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[18]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[18]:

```
In [19]: plt.figure()
age = train.loc[train.Survived == 1, 'Age']
plt.title('The histogram of the age groups of the people that had survived')
plt.hist(age, np.arange(0,100,10))
plt.xticks(np.arange(0,100,10))
plt.figure()
age = train.loc[train.Survived == 0, 'Age']
plt.title('The histogram of the age groups of the people that couldn't survive')
plt.hist(age, np.arange(0,100,10))
plt.xticks(np.arange(0,100,10))
Out[19]:
```

Out[19]:

```
In [20]: train[train['Survived'] == 1, 'Pclass'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[20]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[20]:

```
In [21]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[21]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[21]:

```
In [22]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[22]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[22]:

```
In [23]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[23]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[23]:

```
In [24]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[24]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[24]:

```
In [25]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[25]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[25]:

```
In [26]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[26]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[26]:

```
In [27]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[27]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[27]:

```
In [28]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[28]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[28]:

```
In [29]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[29]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[29]:

```
In [30]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[30]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[30]:

```
In [31]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[31]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[31]:

```
In [32]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[32]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[32]:

```
In [33]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[33]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[33]:

```
In [34]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[34]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[34]:

```
In [35]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[35]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[35]:

```
In [36]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[36]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[36]:

```
In [37]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[37]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[37]:

```
In [38]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[38]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[38]:

```
In [39]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[39]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[39]:

```
In [40]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[40]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[40]:

```
In [41]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[41]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[41]:

```
In [42]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[42]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[42]:

```
In [43]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[43]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[43]:

```
In [44]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[44]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[44]:

```
In [45]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[45]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[45]:

```
In [46]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[46]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[46]:

```
In [47]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[47]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[47]:

```
In [48]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[48]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[48]:

```
In [49]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[49]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[49]:

```
In [50]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[50]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[50]:

```
In [51]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[51]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[51]:

```
In [52]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[52]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[52]:

```
In [53]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[53]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[53]:

```
In [54]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[54]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[54]:

```
In [55]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[55]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[55]:

```
In [56]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[56]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[56]:

```
In [57]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[57]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[57]:

```
In [58]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar()
plt.title('Bar graph of people according to ticket class in which people couldn't survive')
Out[58]: Text(0.5, 1.0, 'Bar graph of people according to ticket class in which people couldn't survive')
```

Out[58]:

```
In [59]: train[train['Survived'] == 1, 'Age'].value_counts().sort_index().plot.bar
```

