

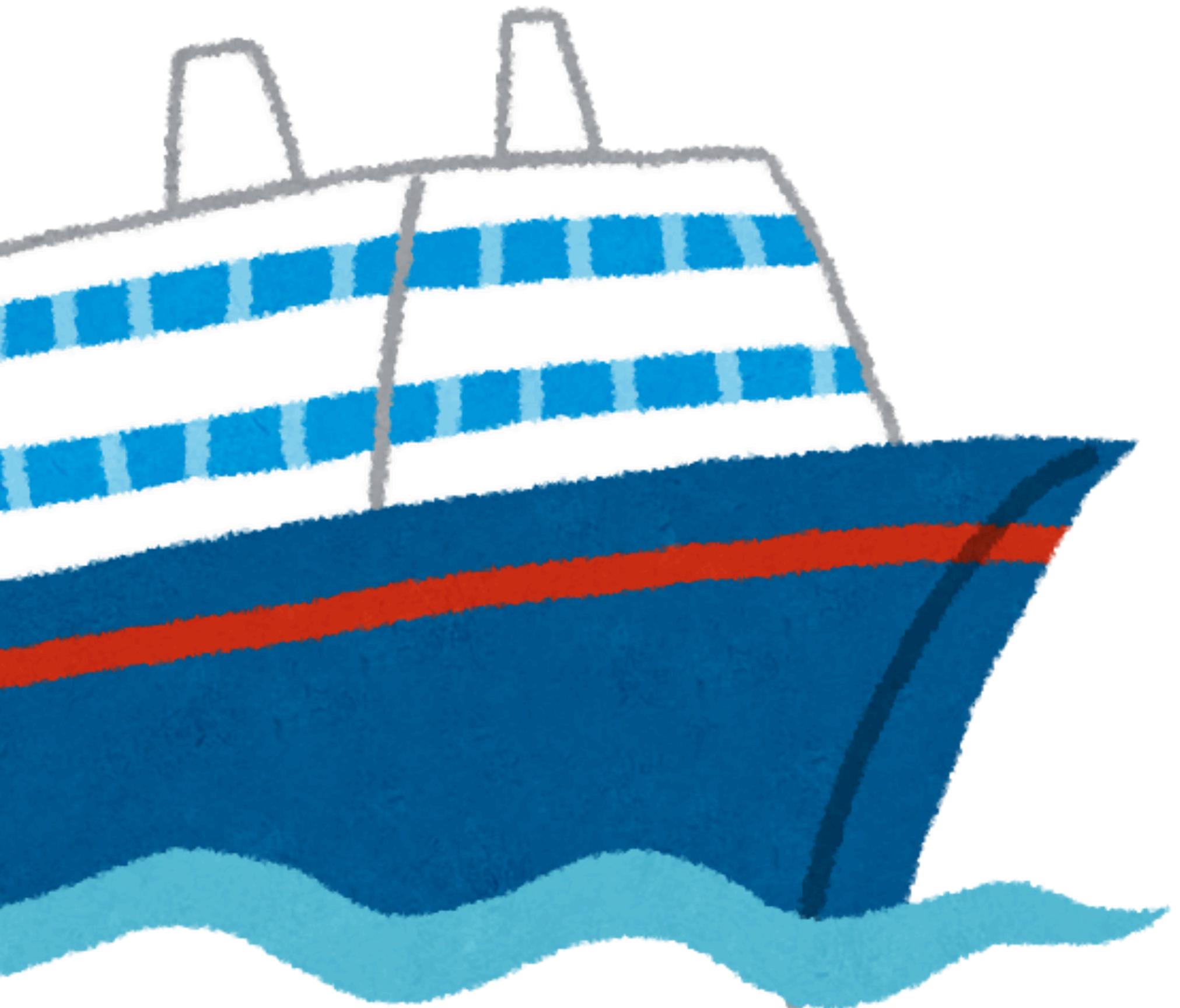
Titanic

Explanatory Data Analysis

by

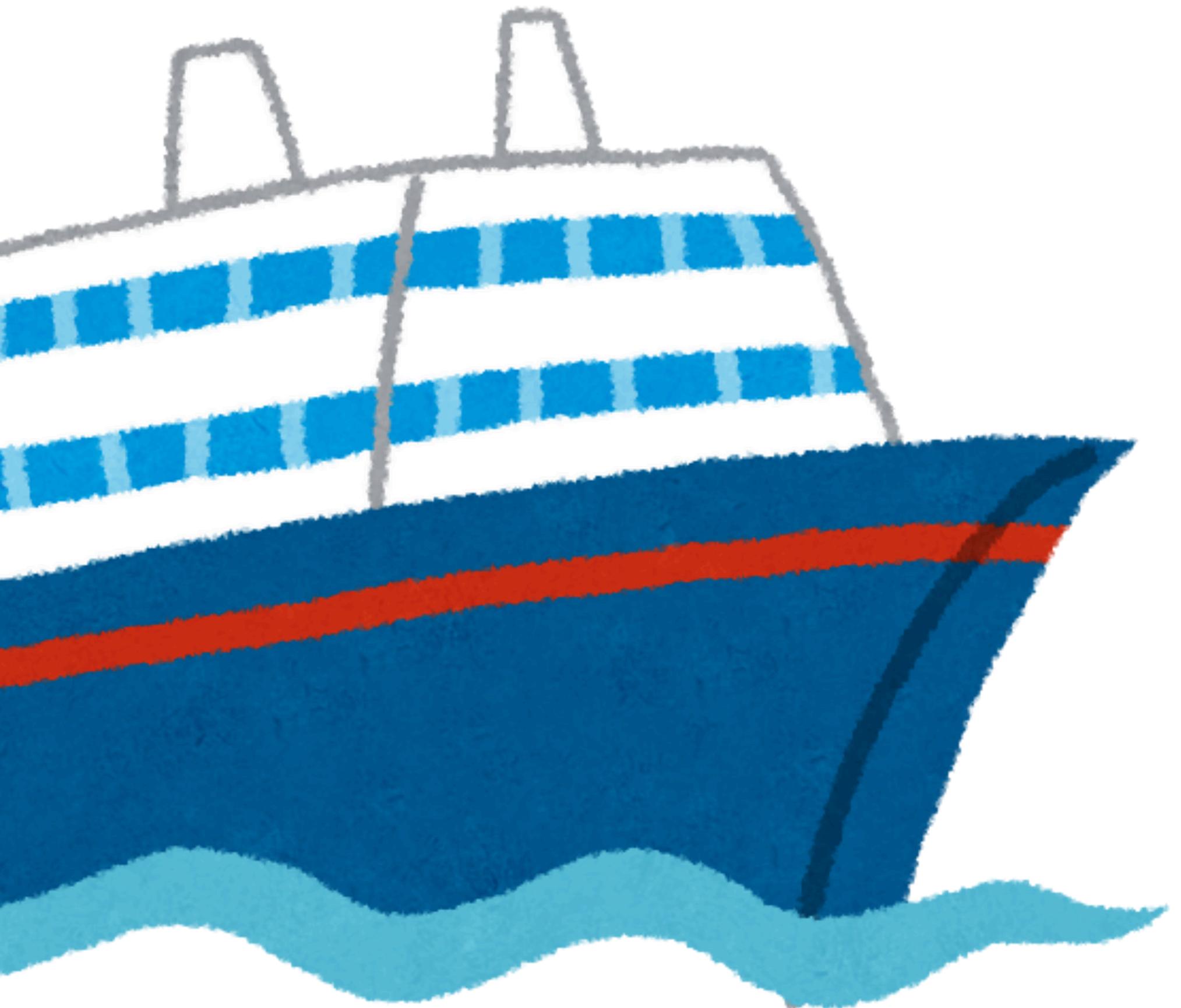
ARIFATUL FATHINAH ESSA





Titanic data

The Titanic dataset contains information about the passengers of the RMS Titanic which sank in 1912. This dataset is generally used for classification analysis and data exploration because it contains demographic variables and passenger safety status.



Titanic data

variables

- Passenger ID
- Survived
- P-class
- Name
- Sex
- Age
- SibSp
- Parch
- Ticket
- Fare
- Cabin
- Embarked

Overview

Step-by-step EDA process

Dataset

The dataset i got from kaggle.
you can type titanic dataset
on google search tab, and
you'll see that kaggle has it

Data preparation

there are several variables
from the data, and “survived”
will be X axis for make the
visualization.

Data visualization

data visualization for
demography and also for
survived passengers.



Dataset

you can check here (<https://www.kaggle.com/datasets/yass erh/titanic-dataset>)

Import the libraries

here's a few libraries that will help for make the EDA

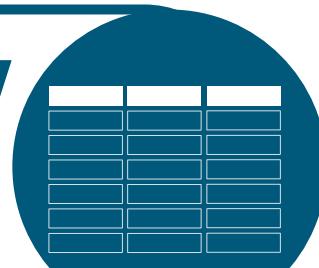
```
# import library

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```



```
#show dataset
titanic = pd.read_csv('train.csv')
titanic.head()
```

Dataset view
the dataset contain 891 rows
and 12 columns.



PassengerId	Survived	Unnamed: 2	Pclass		Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	Non Survived	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	Survived	1	Cumings, Mrs. John Bradley (Florence Briggs Th... er)	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	Survived	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	Survived	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	Non Survived	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Data Preparation

because we will see about survived passengers, so we will prepare for:

1

X Axis

“Survived” will be the X axis

2

Y Axis

“Sex”, “P-class”, and “Embarked” will be the Y axis

3

Crosstab

crosstabulation will be make for survived by sex and embarked

note: cross tabulation isn't part of visualization purpose.
it's just desire to see in numerical form



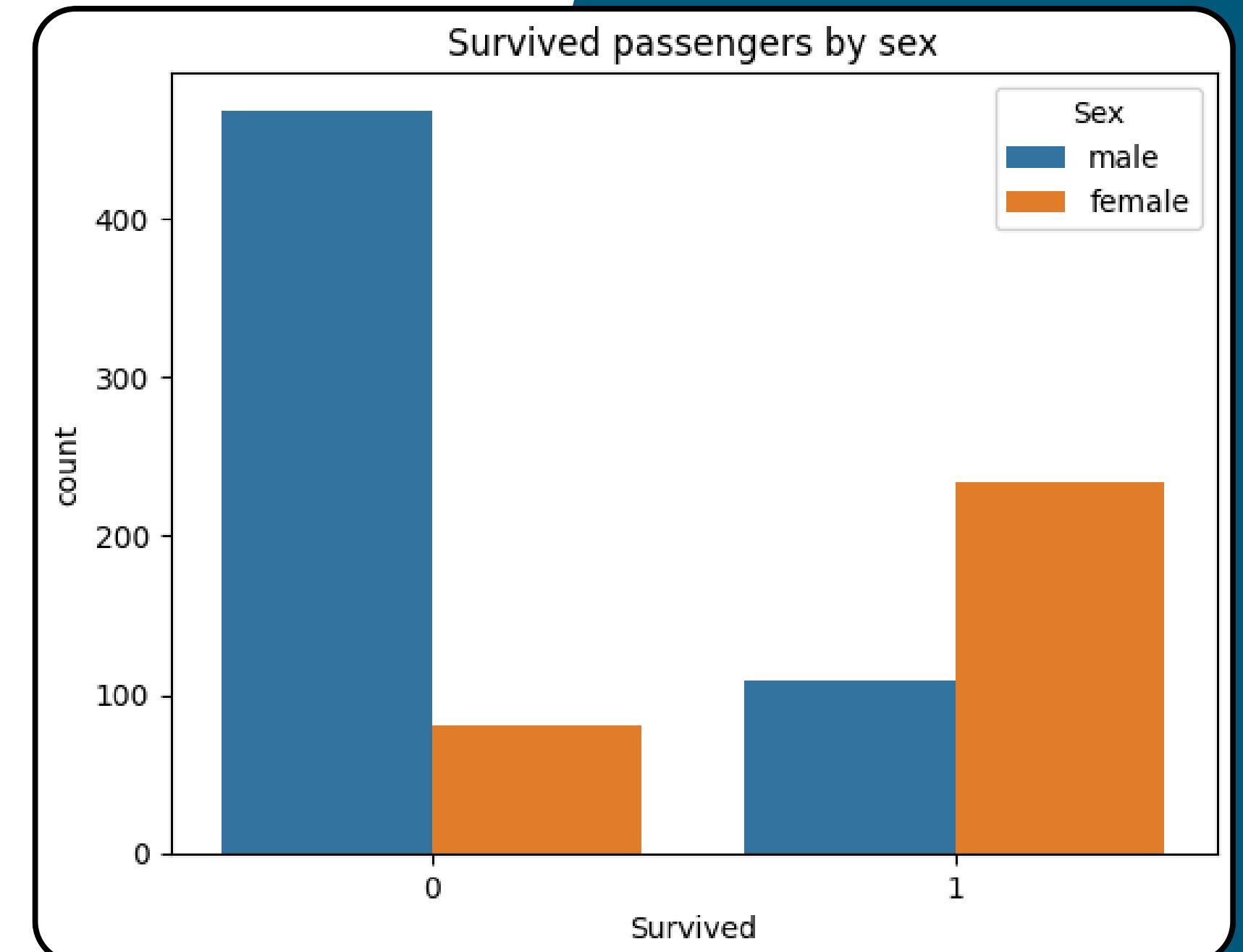
Data Visualization

```
# survived rate by sex  
  
sns.countplot(x='Survived', hue='Sex', data=titanic)  
plt.title('Survived passengers by sex')  
plt.show()
```



- Among passengers who did not survive (0), the majority were male, with only a small number of females.
- Among those who survived (1), the number of female survivors is much higher than male survivors.

This indicates that female passengers had a significantly higher chance of survival compared to males.



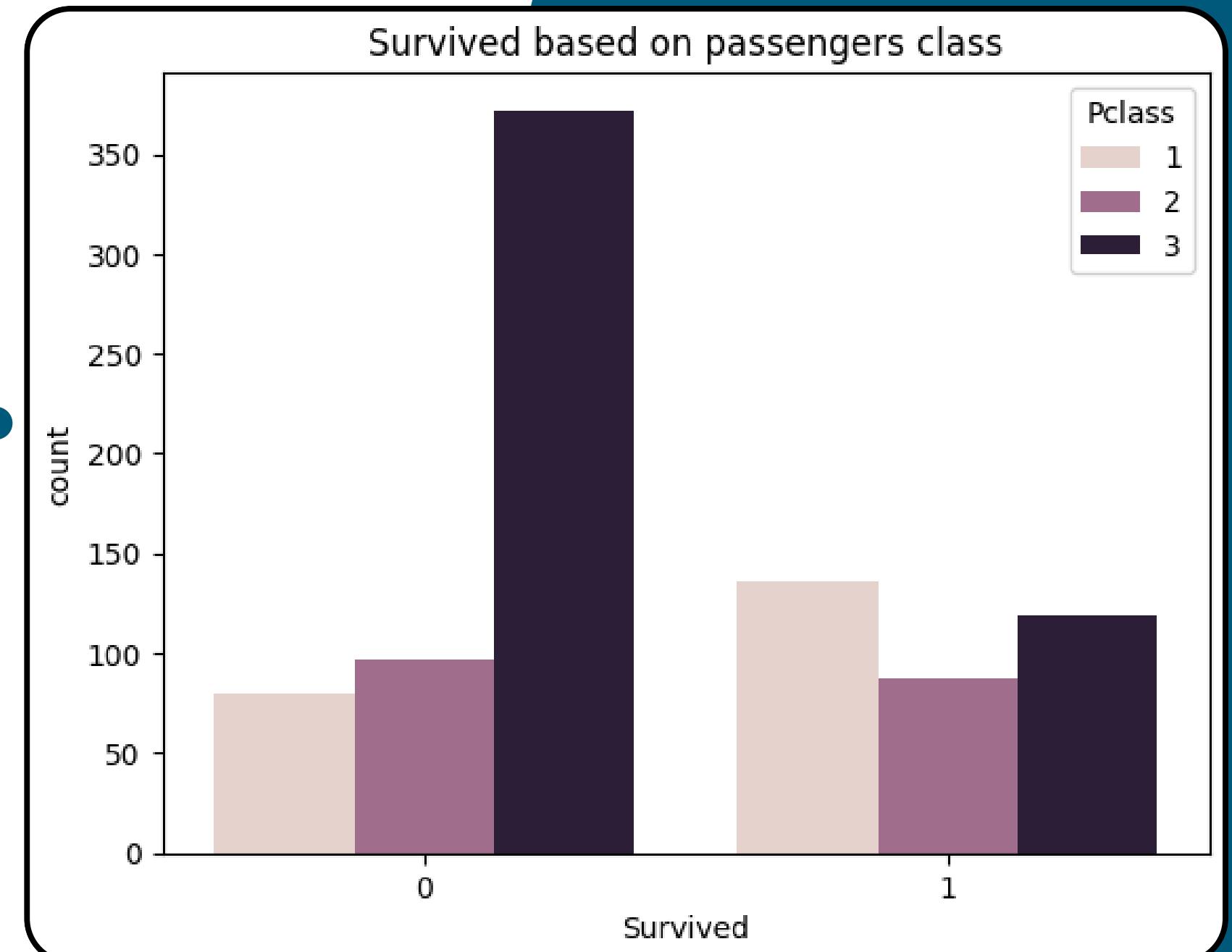
Data Visualization

```
# survival rate by passengers class  
  
sns.countplot(x='Survived', hue='Pclass', data=titanic)  
plt.title('Survived based on passengers class')  
plt.show()
```



- Among those who did not survive (0), the largest group came from 3rd class, followed by 2nd and then 1st class. This shows that 3rd-class passengers had the highest number of fatalities.
- Among those who survived (1), the number of survivors from 1st class is the highest, followed by 3rd class and then 2nd class.

Overall, the chart indicates that passengers in 1st class had a much better chance of survival, while 3rd-class passengers were the most disadvantaged



Data Visualization

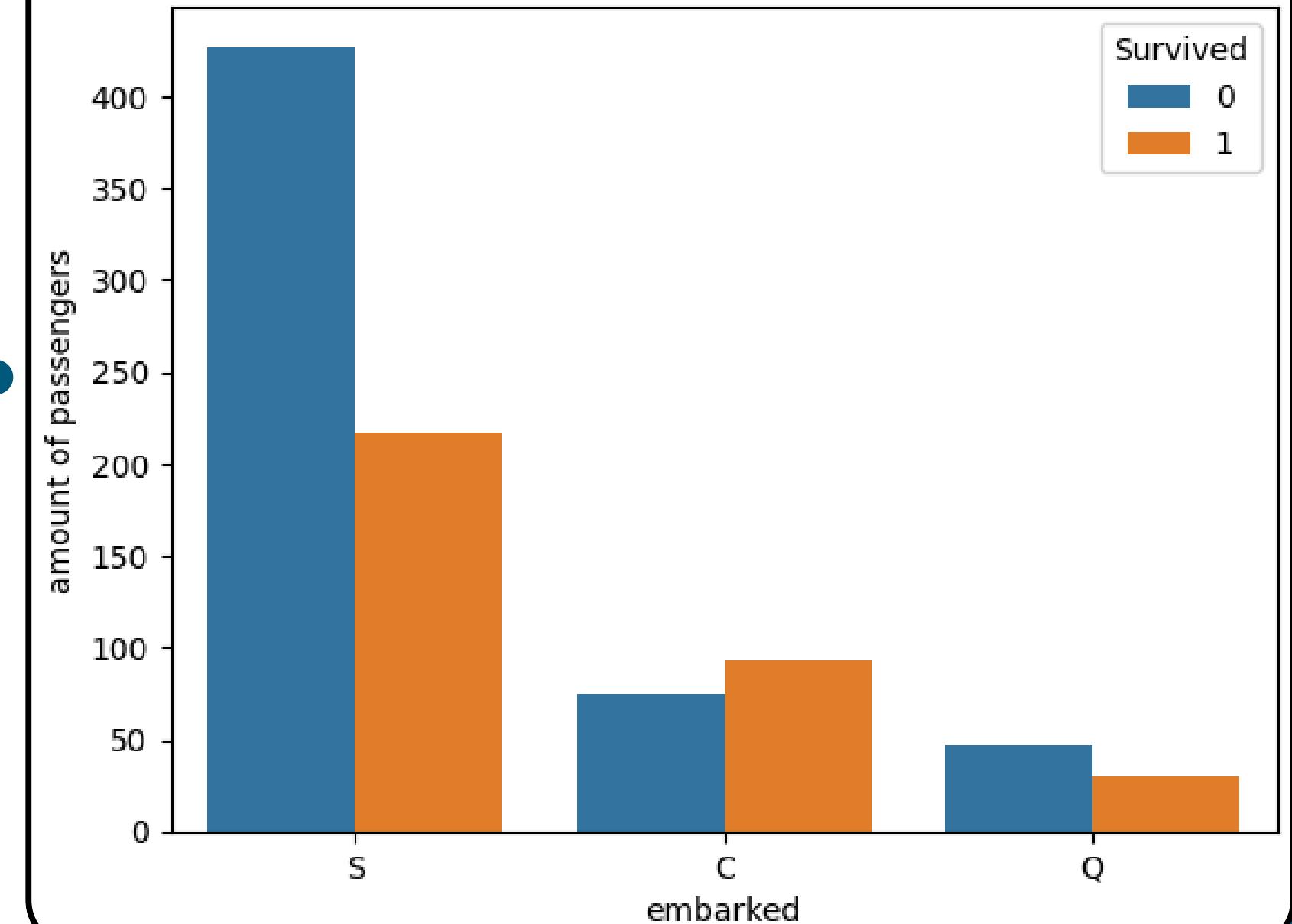
```
# Survival rate by embarked
```

```
sns.countplot(x='Embarked', hue='Survived', data=titanic)
plt.title('Survived Passengers by Embarked')
plt.xlabel('embarked')
plt.ylabel('amount of passengers')
plt.show()
```



- Southampton (S) has the highest number of passengers overall. Most of them did not survive, but a significant portion still survived.
- Cherbourg (C) shows a more balanced distribution, with the number of survivors being slightly higher than non-survivors. This suggests passengers from Cherbourg had a relatively better chance of survival.
- Queenstown (Q) has the fewest passengers. Both survivors and non-survivors are low in number, but non-survivors are slightly higher than survivors.

Survived Passengers by Embarked



Cross-tab



#Cross tabulation

```
pd.crosstab([titanic['Sex'], titanic['Embarked']], titanic['Survived'])
```

Female Passengers

- Cherbourg (C): 64 survived and only 9 did not survive — indicating a very high survival rate for females from Cherbourg.
- Queenstown (Q): 27 survived compared to 9 who did not survive — also showing a high survival rate.
- Southampton (S): 140 survived while 63 did not survive — still a strong survival rate, though lower than Cherbourg.

Male Passengers

- Cherbourg (C): 29 survived and 66 did not survive — meaning most male passengers did not survive.
- Queenstown (Q): Only 3 survived while 38 did not — indicating a very low survival rate.
- Southampton (S): 77 survived compared to 364 who did not survive — this shows the lowest survival success among all groups, especially for males from Southampton.

Sex	Embarked	Survived	
		0	1
female	C	9	64
	Q	9	27
	S	63	140
male	C	66	29
	Q	38	3
	S	364	77

Thank You!

Contact me for collaboration



628 527 1435 939



arifatul fathinah essa



arifatulfathinahessa@gmail.com



github.com/rifaessa-code

