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

```
from google.colab import files
uploaded = files.upload()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving titanic.csv to titanic (1).csv
```

```
import io
```

```
df=pd.read_csv(io.BytesIO(uploaded['titanic (1).csv']))
```

```
df.head()
```

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True

```
print("\nDataset Info:")
print(df.info())
```

```

Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   survived     891 non-null    int64  
 1   pclass       891 non-null    int64  
 2   sex          891 non-null    object 
 3   age          714 non-null    float64
 4   sibsp        891 non-null    int64  
 5   parch        891 non-null    int64  
 6   fare         891 non-null    float64
 7   embarked     889 non-null    object 
 8   class        891 non-null    object 
 9   who          891 non-null    object 
 10  adult_male   891 non-null    bool   
 11  deck         203 non-null    object 
 12  embark_town  889 non-null    object 
 13  alive        891 non-null    object 
 14  alone        891 non-null    bool   
dtypes: bool(2), float64(2), int64(4), object(7)
memory usage: 92.4+ KB
None
```

```
print("\nSummary Statistics:")
print(df.describe(include='all'))
```

```

Summary Statistics:
          survived      pclass   sex         age       sibsp       parch  \
count   891.000000  891.000000   891  714.000000  891.000000  891.000000   
unique         NaN         NaN     2         NaN         NaN         NaN   
top            NaN         NaN  male         NaN         NaN         NaN   
freq           NaN         NaN   577         NaN         NaN         NaN   
mean      0.383838    2.308642   NaN   29.699118    0.523008    0.381594   
std       0.486592    0.836071   NaN   14.526497    1.102743    0.806057   
min       0.000000    1.000000   NaN    0.420000    0.000000    0.000000   
25%       0.000000    2.000000   NaN   20.125000    0.000000    0.000000   
50%       0.000000    3.000000   NaN   28.000000    0.000000    0.000000   
75%       1.000000    3.000000   NaN   38.000000    1.000000    0.000000   
max       1.000000    3.000000   NaN   80.000000    8.000000    6.000000   

              fare embarked  class  who adult_male deck  embark_town alive  \
count   891.000000      889    891  891        891  203          889   891   
unique         NaN        3      3    3          2    7            3     2   
top            NaN        S  Third  man       True    C  Southampton    no   
freq           NaN      644    491  537        537   59          644   549   
```

```
mean     32.204208      NaN    NaN  NaN        NaN  NaN          NaN   NaN
std      49.693429      NaN    NaN  NaN        NaN  NaN          NaN   NaN
min       0.000000      NaN    NaN  NaN        NaN  NaN          NaN   NaN
25%       7.910400      NaN    NaN  NaN        NaN  NaN          NaN   NaN
50%      14.454200      NaN    NaN  NaN        NaN  NaN          NaN   NaN
75%      31.000000      NaN    NaN  NaN        NaN  NaN          NaN   NaN
max     512.329200      NaN    NaN  NaN        NaN  NaN          NaN   NaN

       alone
count    891
unique     2
top     True
freq     537
mean     NaN
std      NaN
min      NaN
25%      NaN
50%      NaN
75%      NaN
max      NaN
```

```
print("\nMissing Values:")
print(df.isnull().sum())
```

```
Missing Values:
survived         0
pclass           0
sex              0
age            177
sibsp            0
parch            0
fare             0
embarked         2
class            0
who              0
adult_male       0
deck           688
embark_town      2
alive            0
alone            0
dtype: int64
```

```
df = df.dropna(subset=['survived', 'sex'])
print(df)
```

```
     survived  pclass     sex   age  sibsp  parch     fare embarked   class  \
0           0       3    male  22.0      1      0   7.2500        S   Third   
1           1       1  female  38.0      1      0  71.2833        C   First   
2           1       3  female  26.0      0      0   7.9250        S   Third   
3           1       1  female  35.0      1      0  53.1000        S   First   
4           0       3    male  35.0      0      0   8.0500        S   Third   
..        ...     ...     ...   ...    ...    ...      ...      ...     ...   
886         0       2    male  27.0      0      0  13.0000        S  Second   
887         1       1  female  19.0      0      0  30.0000        S   First   
888         0       3  female   NaN      1      2  23.4500        S   Third   
889         1       1    male  26.0      0      0  30.0000        C   First   
890         0       3    male  32.0      0      0   7.7500        Q   Third   

       who  adult_male deck  embark_town alive  alone  
0      man        True  NaN  Southampton    no  False  
1    woman       False    C    Cherbourg   yes  False  
2    woman       False  NaN  Southampton   yes   True  
3    woman       False    C  Southampton   yes  False  
4      man        True  NaN  Southampton    no   True  
..     ...         ...  ...          ...   ...    ...  
886    man        True  NaN  Southampton    no   True  
887  woman       False    B  Southampton   yes   True  
888  woman       False  NaN  Southampton    no  False  
889    man        True    C    Cherbourg   yes   True  
890    man        True  NaN   Queenstown    no   True  

[891 rows x 15 columns]
```

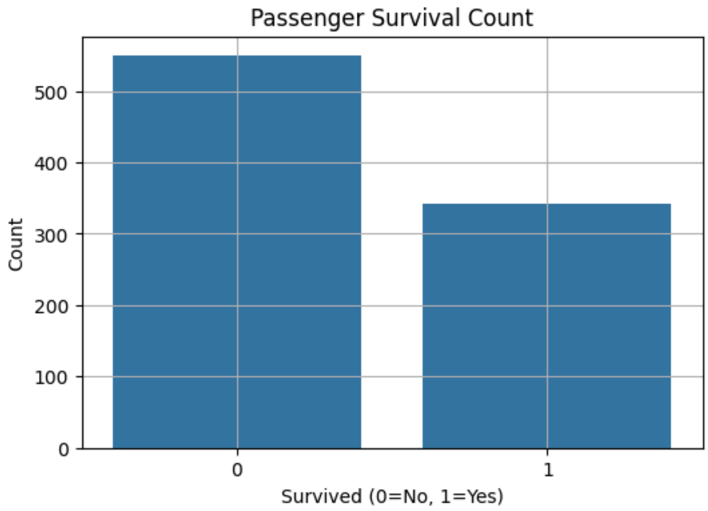
```
questions = [
    "What percentage of passengers survived?",
    "Does gender affect survival rate?",
    "Which class had higher survival rates?",
    "Were older passengers less likely to survive?",
    "Are there missing values that could affect analysis?"
]

print("\nBusiness Questions:")
for q in questions:
    print("-", q)
```

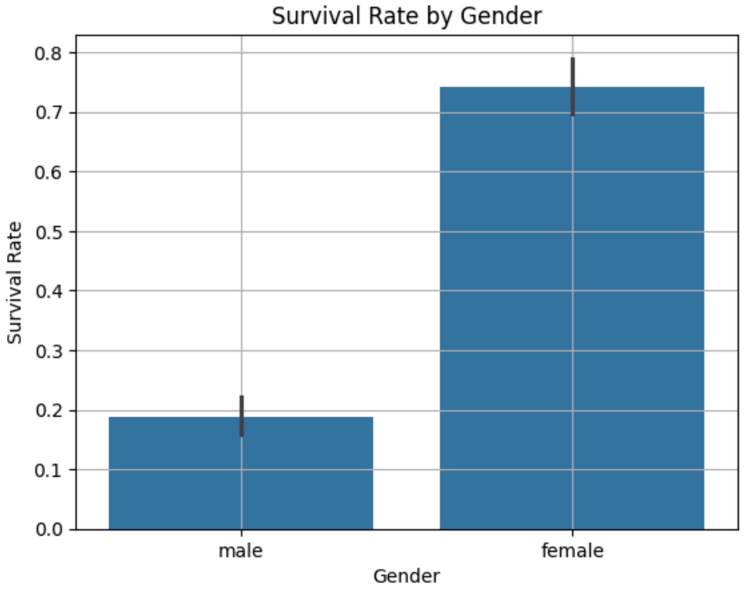
```

Business Questions:
- What percentage of passengers survived?
- Does gender affect survival rate?
- Which class had higher survival rates?
- Were older passengers less likely to survive?
- Are there missing values that could affect analysis?
```

```
plt.figure(figsize=(6,4))
sns.countplot(x='survived', data=df)
plt.title('Passenger Survival Count')
plt.xlabel('Survived (0=No, 1=Yes)')
plt.ylabel('Count')
plt.grid(True)
plt.show()
```



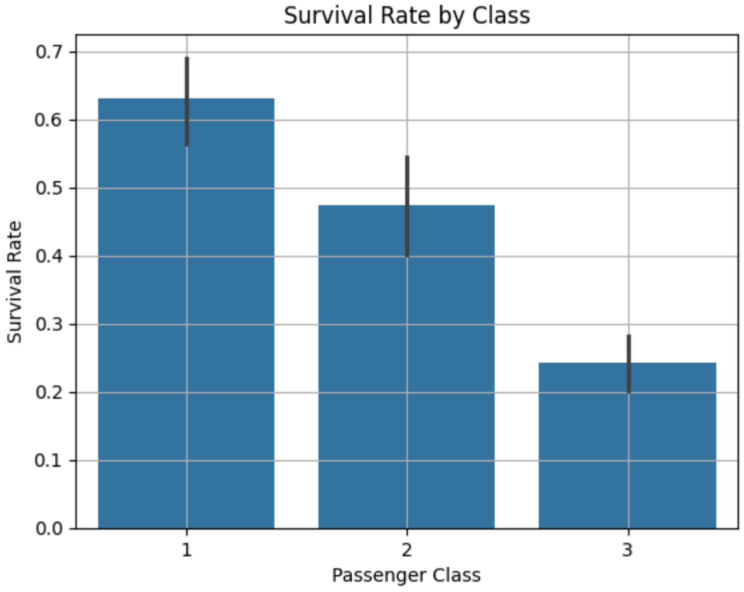
```
sns.barplot(x='sex', y='survived', data=df)
plt.title('Survival Rate by Gender')
plt.xlabel('Gender')
plt.ylabel('Survival Rate')
plt.grid(True)
plt.show()
```



```
class_survival = df.groupby('pclass')['survived'].mean()
print("\nSurvival Rate by Passenger Class:")
print(class_survival)

sns.barplot(x='pclass', y='survived', data=df)
plt.title('Survival Rate by Class')
plt.xlabel('Passenger Class')
plt.ylabel('Survival Rate')
plt.grid(True)
plt.show()
```

```
Survival Rate by Passenger Class:
pclass
1    0.629630
2    0.472826
3    0.242363
Name: survived, dtype: float64
```



```
print("\nMissing Age values:", df['age'].isnull().sum())
print("Missing Embarked values:", df['embarked'].isnull().sum())
```

```
Missing Age values: 177
Missing Embarked values: 2
```

```
female_survival = df[df['sex'] == 'female']['survived'].mean()
male_survival = df[df['sex'] == 'male']['survived'].mean()

print("\nAverage Survival Rate by Gender:")
print("Female:", female_survival)
print("Male:", male_survival)
```

```
Average Survival Rate by Gender:
Female: 0.7420382165605095
Male: 0.18890814558058924
```

```
import scipy.stats as stats

t_stat, p_value = stats.ttest_ind(male_survival, female_survival)

print("\nTwo-Sample T-Test Result:")
print("t-statistic:", t_stat)
print("p-value:", p_value)

alpha = 0.05  # significance level

if p_value < alpha:
    print("\nResult: Reject Null Hypothesis (H0)")
    print("Conclusion: Survival rate differs significantly between males and females.")
else:
    print("\nResult: Fail to Reject Null Hypothesis (H0)")
    print("Conclusion: No significant difference in survival rates between genders.")
```

```
Two-Sample T-Test Result:
t-statistic: -19.297816550123354
```