HW1
Summary of Titanic dataset

Ting-Ni, Chen

2025-02-26

目錄

一、讀取資料/敘述統計

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from tabulate import tabulate

column_types={'PassengerId':'category',
              'Survived':'category',
              'Pclass':int,
              'Name':'category',
              'Sex':'category',
              'Age':float,
              'SibSp':int,
              'Parch':int,
              'Fare':float,
              'Cabin':'category',
              'Embarked':'category'}
df = pd.read_csv('titanic.csv', dtype=column_types)
summary_table = tabulate(df.describe().transpose(), headers='keys', tablefmt='pipe')

print(summary_table)
```

|	count	mean	std	min	25%	50%	75%	max
Pclass	891	2.30864	0.836071	1	2	3	3	3
Age	714	29.6991	14.5265	0.42	20.125	28	38	80
SibSp	891	0.523008	1.10274	0	0	0	1	8
Parch	891	0.381594	0.806057	0	0	0	0	6
Fare	891	32.2042	49.6934	0	7.9104	14.4542	31	512.329

二、各Pclass對應的生存率

```python
df["Pclass"] = df["Pclass"].astype("category")

df["Survived"] = df["Survived"].astype(int)

survival_rate = df.groupby("Pclass")["Survived"].mean().reset_index()

plt.figure(figsize=(6, 4))
sns.barplot(x="Pclass", y="Survived", data=survival_rate, order=[1, 2, 3], palette='pastel')

plt.title("Survival Rate by Pclass")
plt.xlabel("Pclass")
plt.ylabel("Survival Rate")
plt.ylim(0, 1)
plt.tight_layout()
plt.show()
```
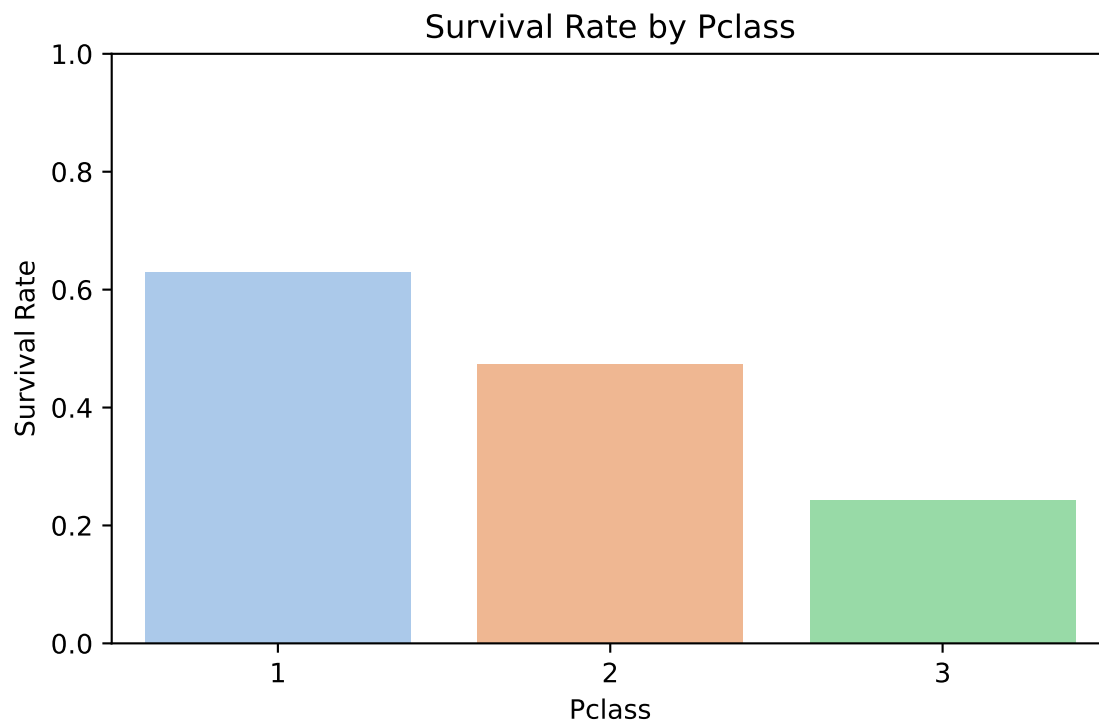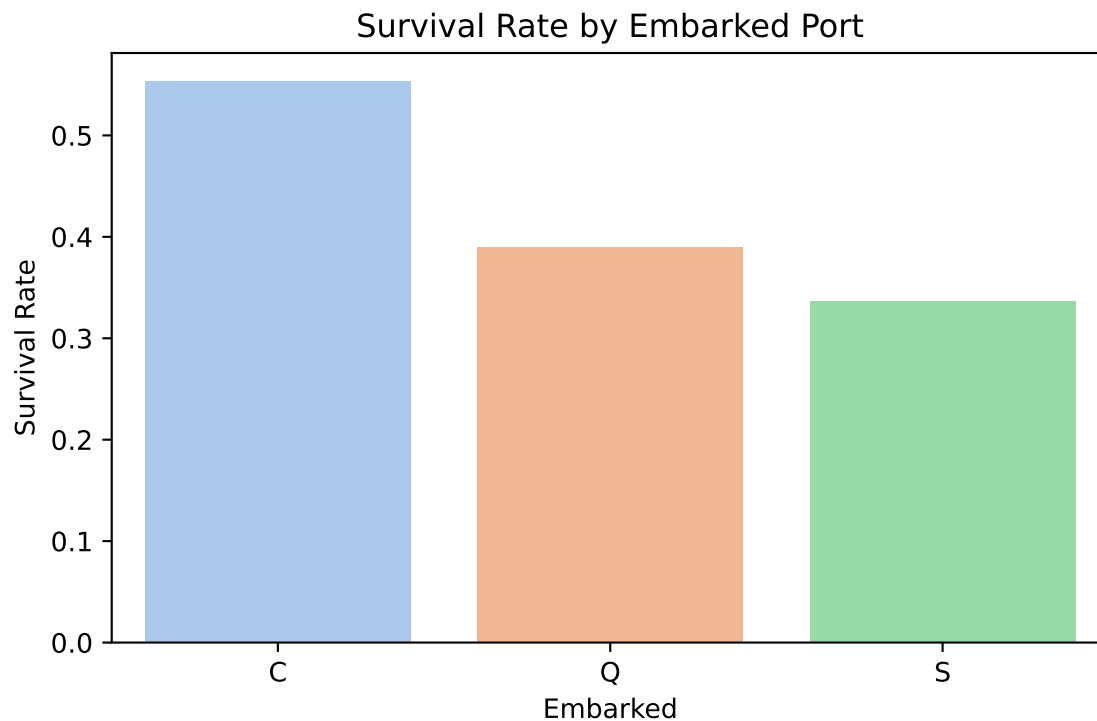

三、各港口上船的乘客對應的生存率

```python
survival_rate_embarked = df.groupby("Embarked")["Survived"].mean().reset_index()

plt.figure(figsize=(6, 4))
sns.barplot(x="Embarked", y="Survived", data=survival_rate_embarked, palette="pastel")

plt.title("Survival Rate by Embarked Port")
plt.xlabel("Embarked")
plt.ylabel("Survival Rate")
plt.tight_layout()
```
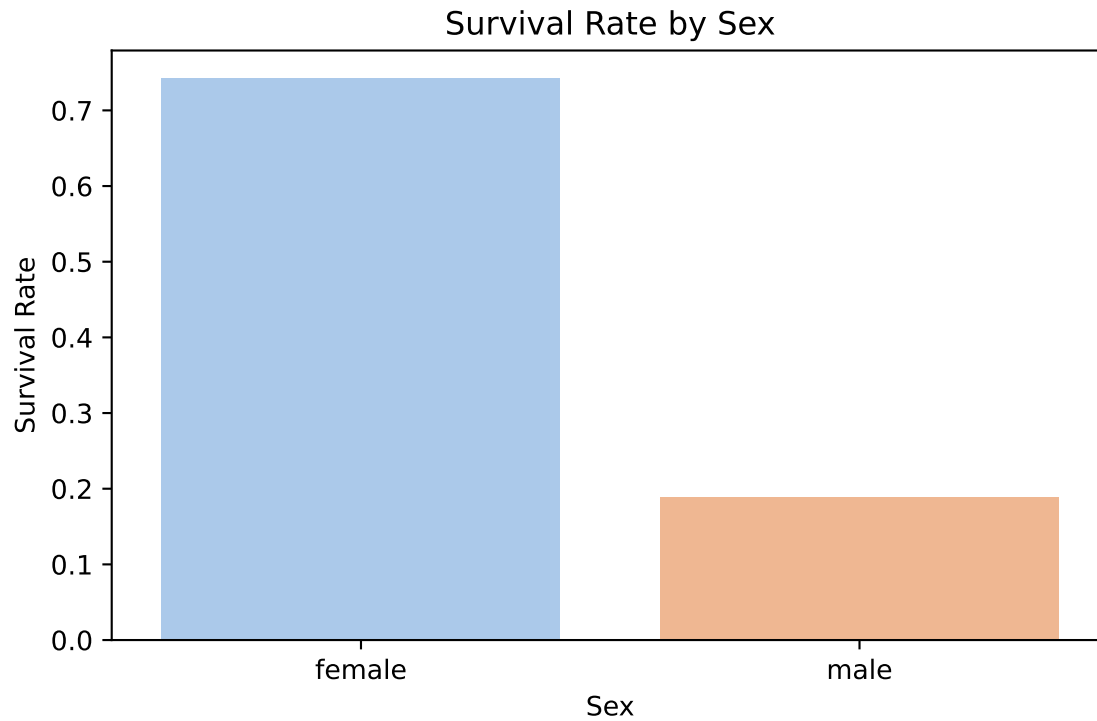
```
plt.show()
```

Survival Rate by Embarked Port

四、性別對應的生存率

```
survival_rate_sex = df.groupby("Sex")["Survived"].mean().reset_index()

plt.figure(figsize=(6, 4))
sns.barplot(x="Sex", y="Survived", data=survival_rate_sex, palette="pastel")

plt.title("Survival Rate by Sex")
plt.xlabel("Sex")
plt.ylabel("Survival Rate")
plt.tight_layout()
plt.show()
```
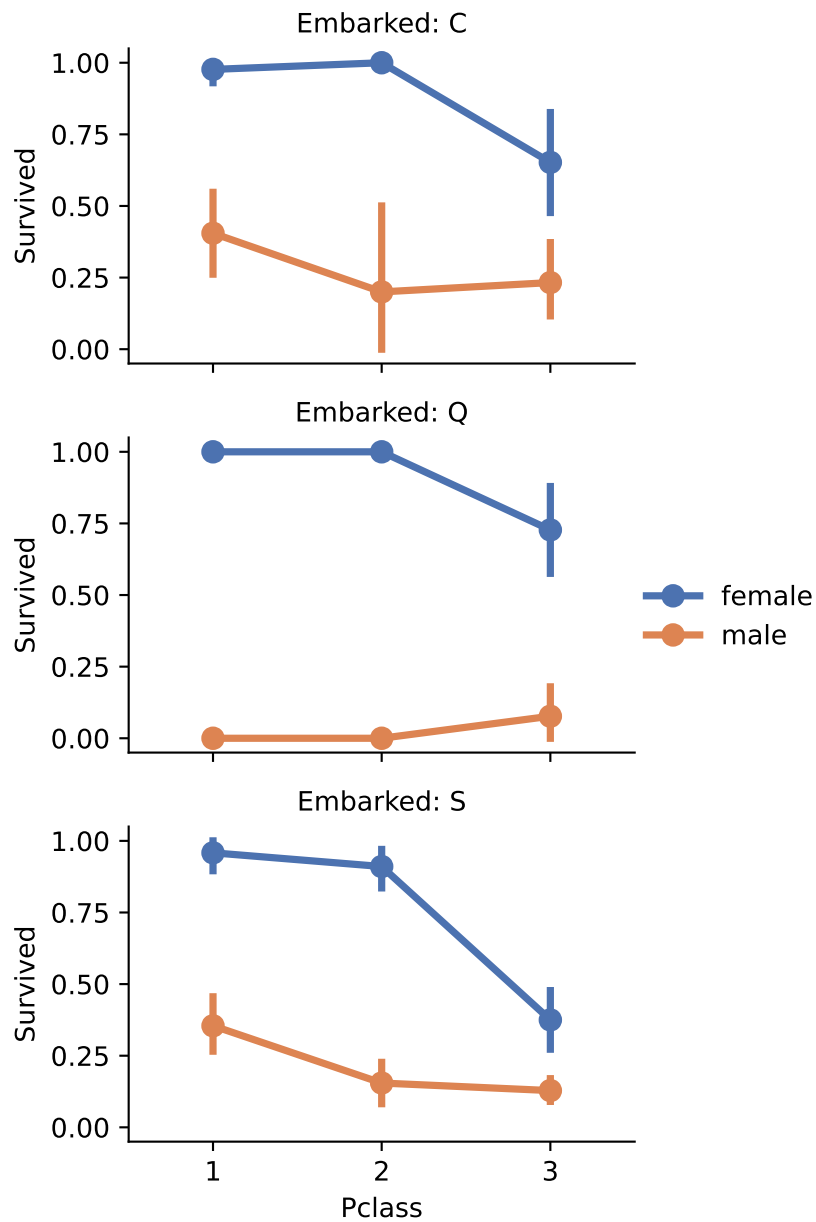
Survival Rate by Sex

五、按 Embarked 分面顯示的存活率 (Survived) 點圖，並根據 Pclass 和 Sex 進行比較

```
grid = sns.FacetGrid(df, row="Embarked", height=2.2, aspect=1.6)
grid.map_dataframe(sns.pointplot, x="Pclass", y="Survived", hue="Sex",
                   palette="deep", hue_order=["female", "male"], order=[1, 2, 3])

grid.add_legend()
grid.set_titles("Embarked: {row_name}")
plt.show()
```

六、不同艙等 (Pclass) 下的乘客年齡 (Age) 分布，並根據存活 (Survived) 與否來區分

```python
df["Pclass"] = df["Pclass"].astype("category")
df["Survived"] = df["Survived"].astype(int)
grid = sns.FacetGrid(df, hue="Survived", row="Pclass", height=3, aspect=2, palette="coolwarm")
grid.map(plt.hist, "Age", bins=20, alpha=0.6, edgecolor="black")
grid.add_legend(title="Survived")
plt.show()
```