Ejercicio IA Titanic

Análisis Exploratorio de Datos

Carga Dataset

In [6]:
```
import pandas as pd

# Ruta del archivo (usa raw string o dobles barras)
ruta = r'C:\Users\Dc\Documents\Maestría_Data_Science\Fundamentos_IA\Data_Sets\ti
df_titanic = pd.read_csv(ruta)

# Mostrar las primeras filas
print(df_titanic.head())
```

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S
```

Análisis

In [7]:
```
# Mostrar las primeras filas
df_titanic.describe()
print(df_titanic.describe())
```

```
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  714.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   14.526497    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   20.125000    0.000000
50%     446.000000    0.000000    3.000000   28.000000    0.000000
75%     668.500000    1.000000    3.000000   38.000000    1.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000

            Parch        Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000  512.329200
```

In [9]:
```python
# Identificar datos nulos y calcular el porcentaje
nulos = df_titanic.isnull().sum()
total = len(df_titanic)
porcentaje_nulos = (nulos / total) * 100

# Mostrar los resultados
print("Datos nulos por columna:")
print(nulos)
print("\nPorcentaje de datos nulos por columna:")
print(porcentaje_nulos)
```

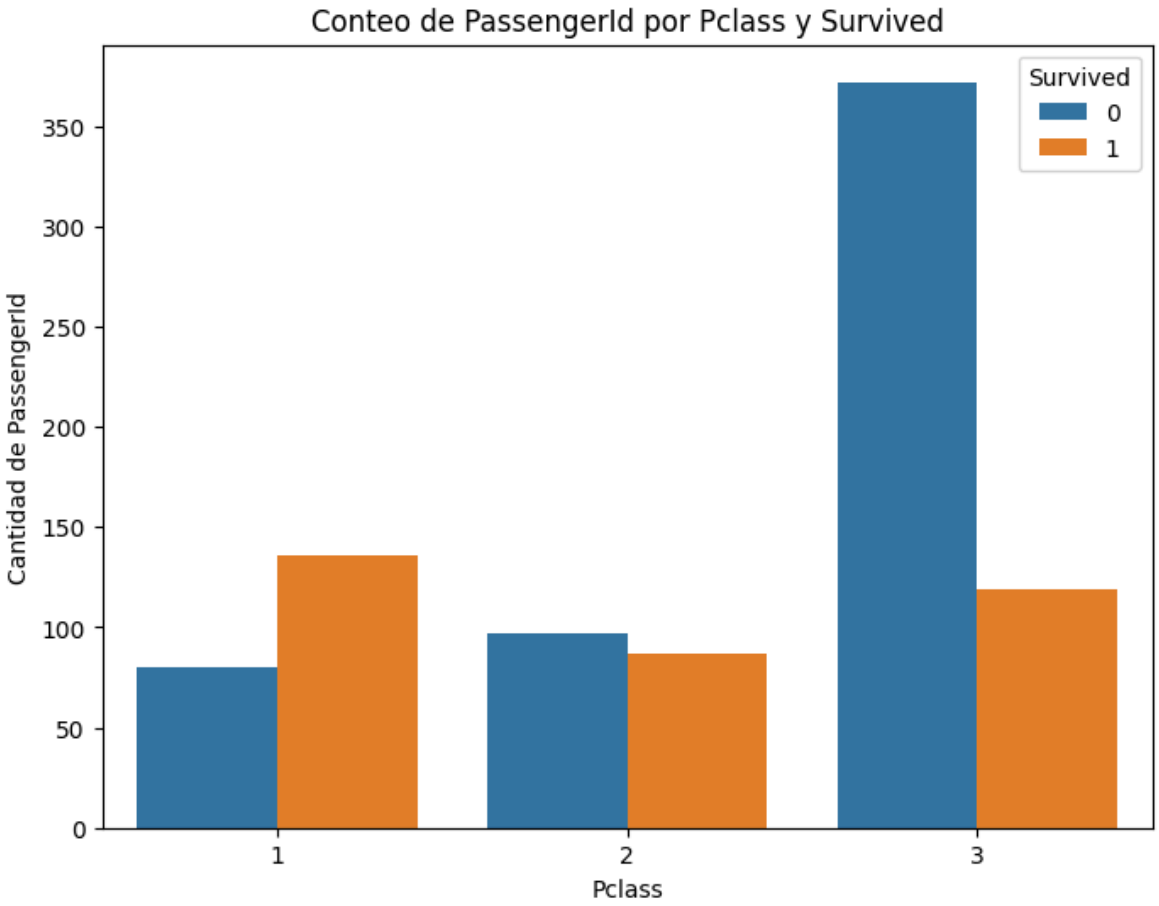
```
Datos nulos por columna:
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64

Porcentaje de datos nulos por columna:
PassengerId     0.000000
Survived        0.000000
Pclass          0.000000
Name            0.000000
Sex             0.000000
Age            19.865320
SibSp           0.000000
Parch           0.000000
Ticket          0.000000
Fare            0.000000
Cabin          77.104377
Embarked        0.224467
dtype: float64
```

In [ ]:
```
# Mostramos la información del DataFrame
df_titanic.info()
```

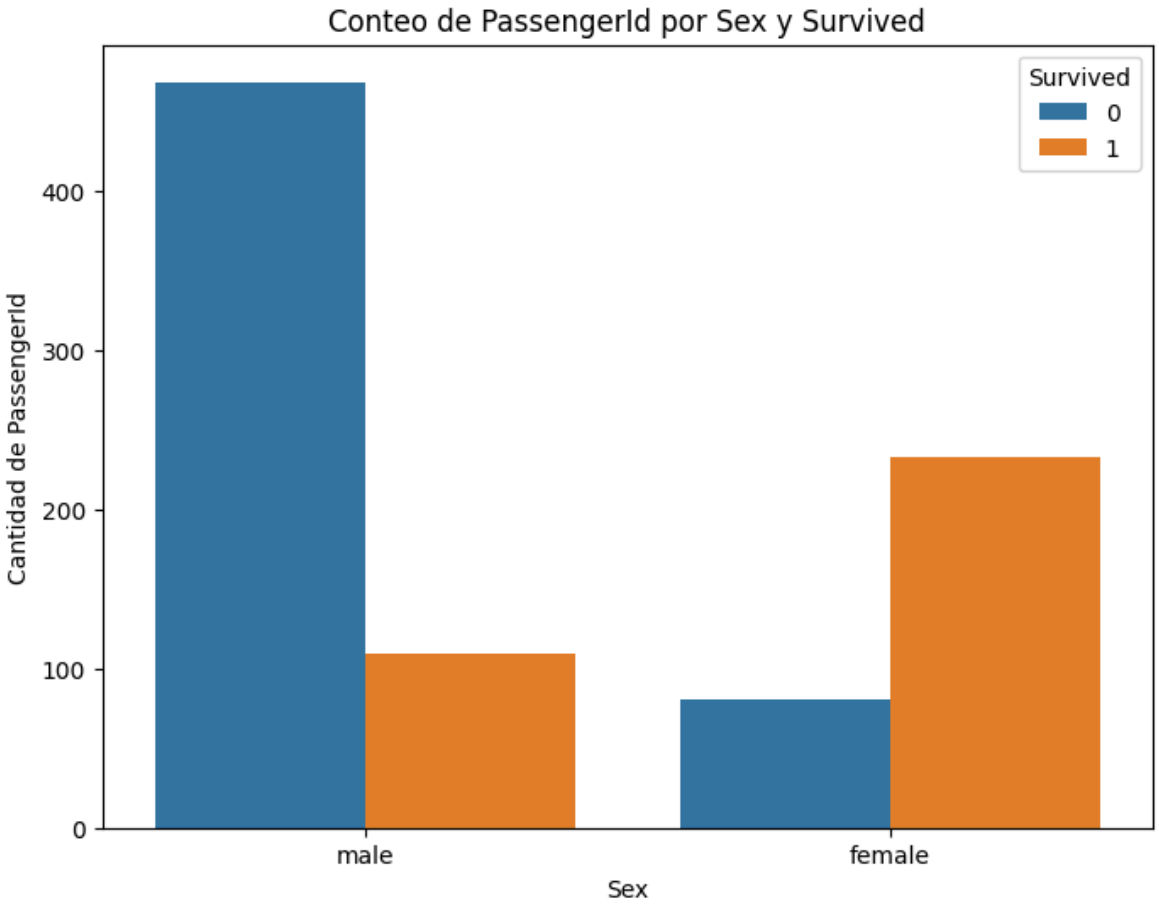
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

Análisis Gráfico

In [10]:
```
import seaborn as sns

import matplotlib.pyplot as plt

# Crear el gráfico
plt.figure(figsize=(8, 6))
sns.countplot(data=df_titanic, x='Pclass', hue='Survived')

# Personalizar el gráfico
plt.title('Conteo de PassengerId por Pclass y Survived')
plt.xlabel('Pclass')
plt.ylabel('Cantidad de PassengerId')
plt.legend(title='Survived', loc='upper right')

# Mostrar el gráfico
plt.show()
```



In [11]:
```
# Crear el gráfico
plt.figure(figsize=(8, 6))
sns.countplot(data=df_titanic, x='Sex', hue='Survived')

# Personalizar el gráfico
plt.title('Conteo de PassengerId por Sex y Survived')
plt.xlabel('Sex')
plt.ylabel('Cantidad de PassengerId')
plt.legend(title='Survived', loc='upper right')

# Mostrar el gráfico
plt.show()
```


Conteo de PassengerId por Sex y Survived
Survived
0
1
Cantidad de PassengerId
400
300
200
100
0
male
female
Sex