In [7]:
```
# Import necessary libraberies
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [8]:
```
# load the dataset
df=pd.read_csv("gender_submission.csv")
```

In [9]:
```
#basic info
print ("dataset info")
print (df.info(), "\n")
```

```
dataset info
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 2 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  418 non-null    int64
 1   Survived     418 non-null    int64
dtypes: int64(2)
memory usage: 6.7 KB
None
```

In [10]:
```
# display 1st few rows
print("first 5 records")
print(df.head(), "\n")
```

```
first 5 records
   PassengerId  Survived
0          892         0
1          893         1
2          894         0
3          895         0
4          896         1
```

In [11]:
```
# display the statistics
print("statistical summary:")
print(df.describe(), "\n")
```

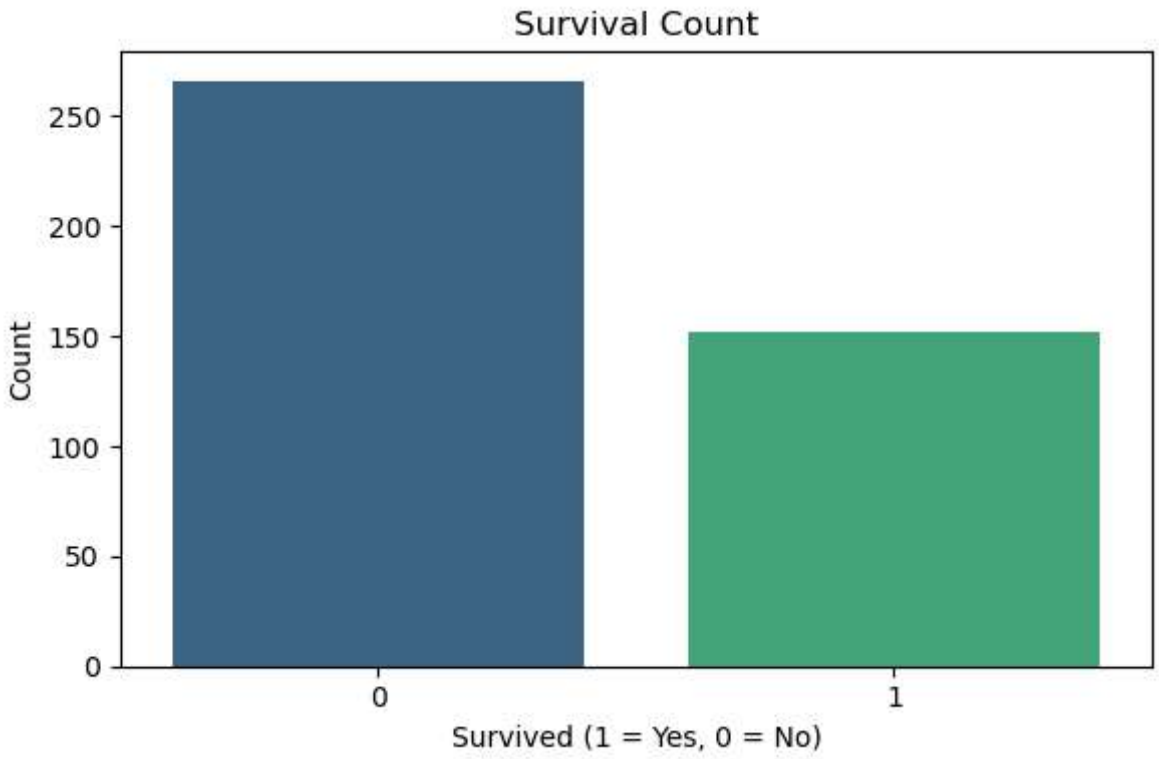
```
statistical summary:
       PassengerId    Survived
count   418.000000  418.000000
mean   1100.500000    0.363636
std     120.810458    0.481622
min     892.000000    0.000000
25%     996.250000    0.000000
50%    1100.500000    0.000000
75%    1204.750000    1.000000
max    1309.000000    1.000000
```

In [12]:
```
# Countplot of Survived
plt.figure(figsize=(6, 4))
sns.countplot(x='Survived', data=df, palette='viridis')
plt.title('Survival Count')
plt.xlabel('Survived (1 = Yes, 0 = No)')
plt.ylabel('Count')
plt.tight_layout()
plt.show()
```

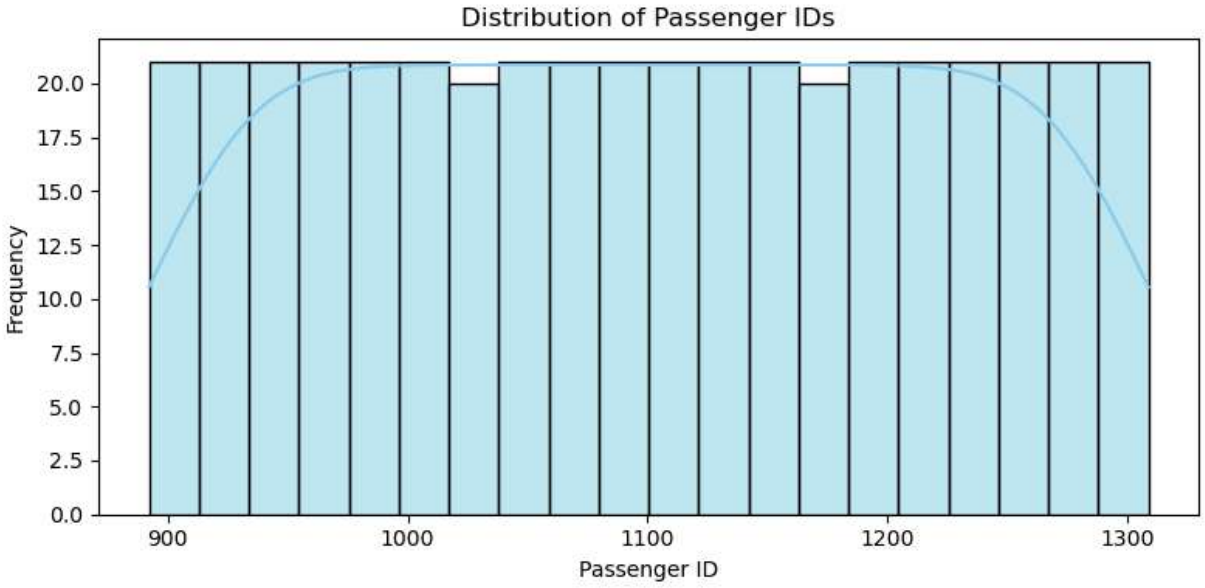
```
C:\Users\Admins\AppData\Local\Temp\ipykernel_15868\1653393873.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.1
4.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.countplot(x='Survived', data=df, palette='viridis')
```



In [13]:
```
# Histogram of PassengerId
plt.figure(figsize=(8, 4))
sns.histplot(df['PassengerId'], bins=20, kde=True, color='skyblue')
plt.title('Distribution of Passenger IDs')
plt.xlabel('Passenger ID')
plt.ylabel('Frequency')
plt.tight_layout()
plt.show()
```

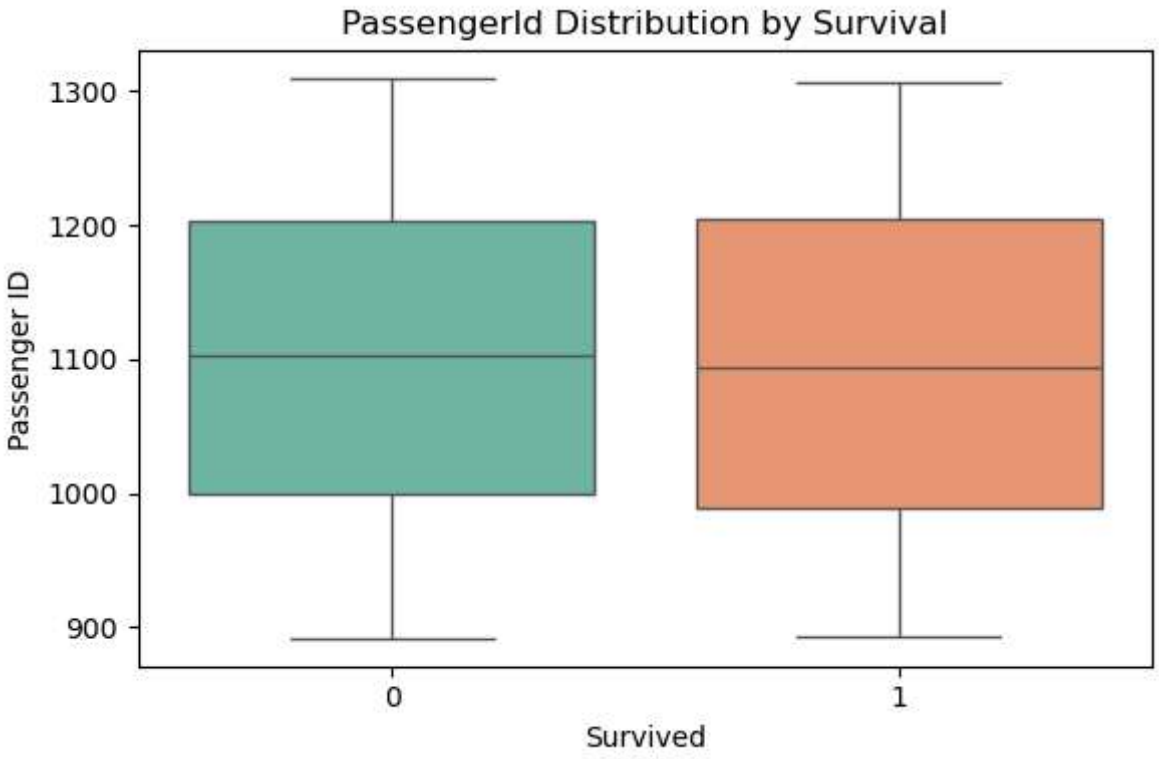


In [14]:
```
# Boxplot of PassengerId grouped by Survived
plt.figure(figsize=(6, 4))
sns.boxplot(x='Survived', y='PassengerId', data=df, palette='Set2')
plt.title('PassengerId Distribution by Survival')
plt.xlabel('Survived')
plt.ylabel('Passenger ID')
plt.tight_layout()
plt.show()
```

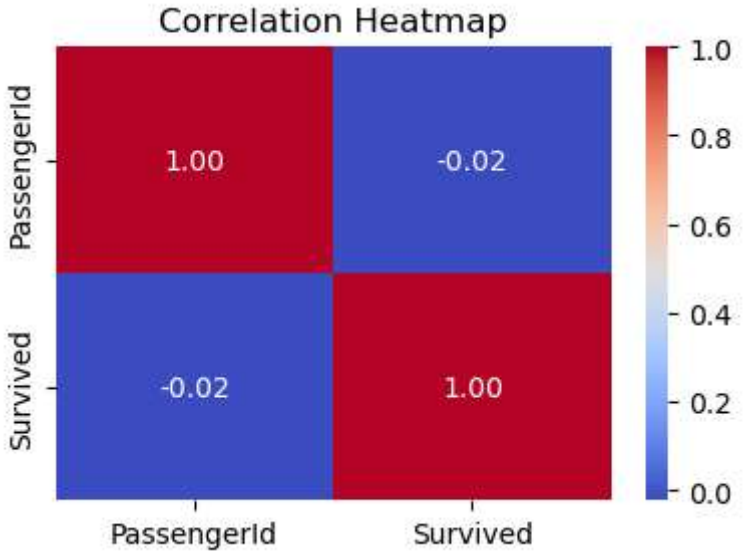
```
C:\Users\Admins\AppData\Local\Temp\ipykernel_15868\3897565764.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.1
4.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.boxplot(x='Survived', y='PassengerId', data=df, palette='Set2')
```



In [15]:
```
# Heatmap of correlation (limited here)
plt.figure(figsize=(4, 3))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap')
plt.tight_layout()
plt.show()
```



In [16]:
```
# Summary of findings
print("📝 Summary of Insights:")
print("""
1. Around 36% of passengers in this test set are predicted to have survived.
2. The survival count plot shows that non-survivors are significantly more.
3. No direct pattern in PassengerId and survival, but ID range is evenly spread.
4. Correlation heatmap shows minimal correlation due to only 2 numeric fields.
```

```
Note: For deeper insights, include additional features like Age, Sex, Pclass etc.
""")
```

📝 Summary of Insights:

1. Around 36% of passengers in this test set are predicted to have survived.
2. The survival count plot shows that non-survivors are significantly more.
3. No direct pattern in PassengerId and survival, but ID range is evenly spread.
4. Correlation heatmap shows minimal correlation due to only 2 numeric fields.

Note: For deeper insights, include additional features like Age, Sex, Pclass etc.

In [ ]: