

task2-1

July 20, 2024

```
[4]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv("train.csv")

print(data.head())

print(data.isnull().sum())

print(data.isnull().sum())

print(data.describe())

plt.figure(figsize=(8, 6))
sns.countplot(x='Survived', data=data)
plt.title('Distribution of Survival')
plt.xlabel('Survived')
plt.ylabel('Count')
plt.show()

plt.figure(figsize=(8, 6))
sns.countplot(x='Pclass', data=data)
plt.title('Distribution of Passenger Class')
plt.xlabel('Pclass')
plt.ylabel('Count')
plt.show()

plt.figure(figsize=(10, 6))
sns.histplot(data['Age'], kde=True, bins=30)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()

plt.figure(figsize=(8, 6))
```

```

sns.barplot(x='Pclass', y='Survived', data=data)
plt.title('Survival Rate by Passenger Class')
plt.xlabel('Pclass')
plt.ylabel('Survival Rate')
plt.show()

plt.figure(figsize=(8, 6))
sns.barplot(x='Sex', y='Survived', data=data)
plt.title('Survival Rate by Sex')
plt.xlabel('Sex')
plt.ylabel('Survival Rate')
plt.show()

plt.figure(figsize=(12, 10))
sns.pairplot(data[['Survived', 'Pclass', 'Age', 'Fare', 'SibSp', 'Parch']])
plt.show()

plt.figure(figsize=(12, 8))
sns.heatmap(data.corr(), annot=True, cmap='coolwarm', linewidths=0.5)
plt.title('Correlation Heatmap')
plt.show()

plt.figure(figsize=(8, 6))
sns.barplot(x='Embarked', y='Survived', data=data)
plt.title('Survival Rate by Embarkation Point')
plt.xlabel('Embarked')
plt.ylabel('Survival Rate')
plt.show()

plt.figure(figsize=(12, 6))
sns.boxplot(x='Sex', y='Age', hue='Survived', data=data)
plt.title('Survival Rate by Age and Sex')
plt.xlabel('Sex')
plt.ylabel('Age')
plt.show()

```

	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

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

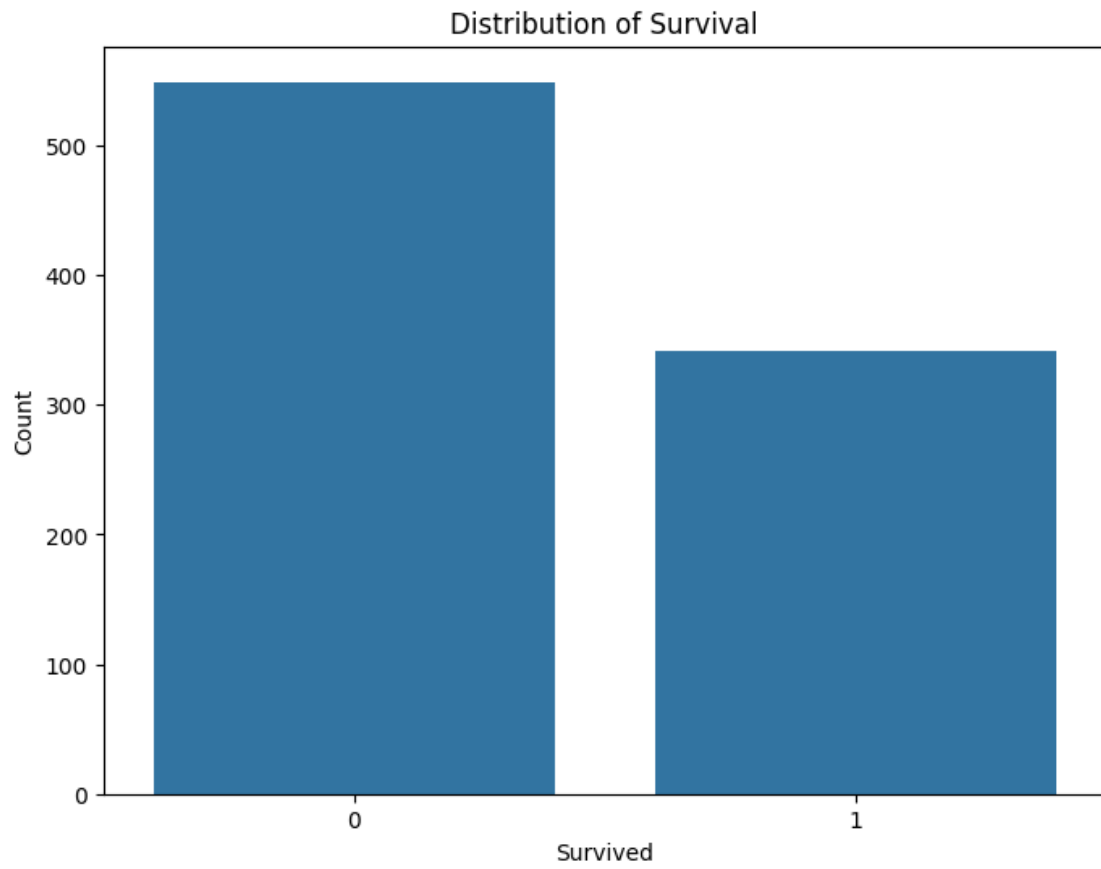
PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0
Parch	0
ticket	0
Fare	0
Cabin	687
Embarked	2

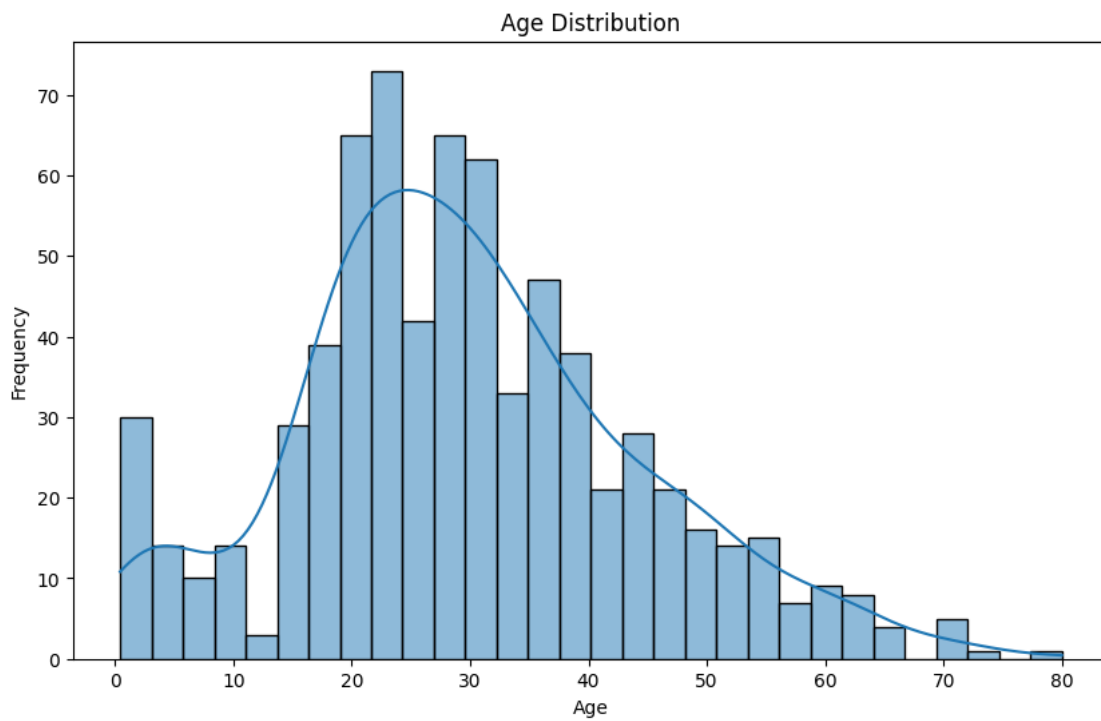
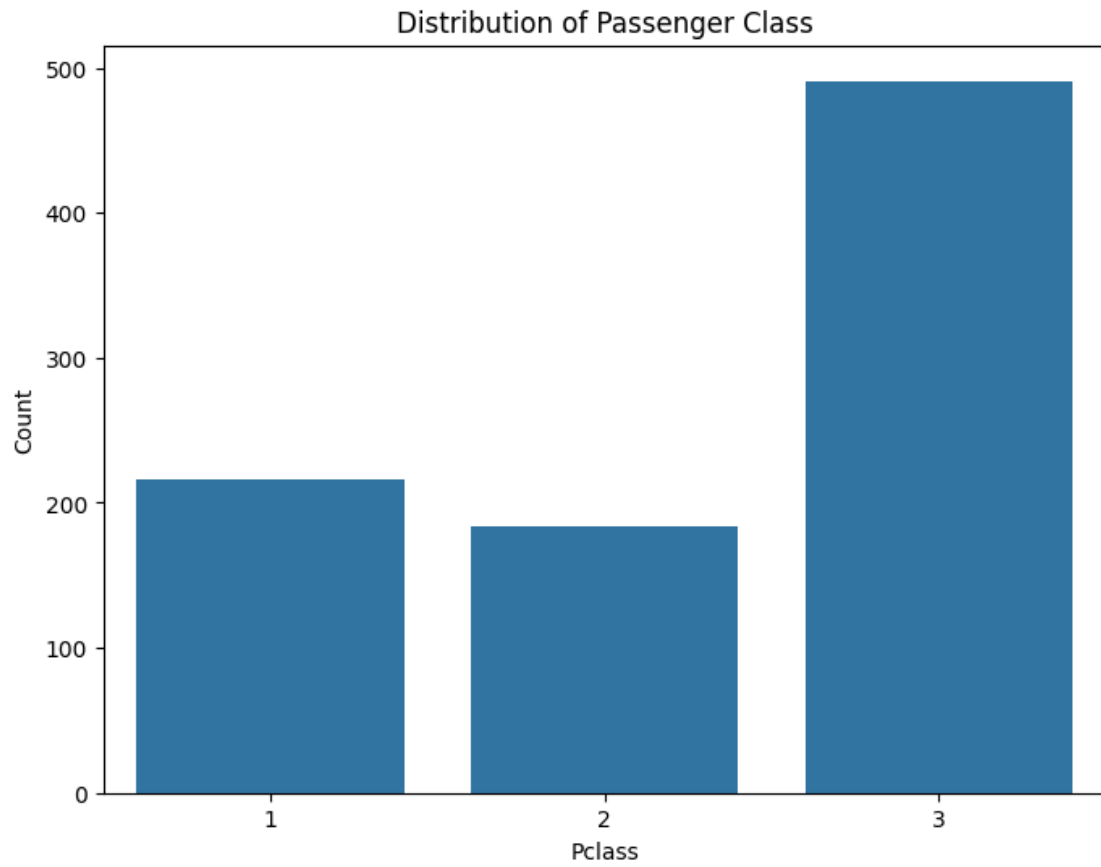
dtype: int64

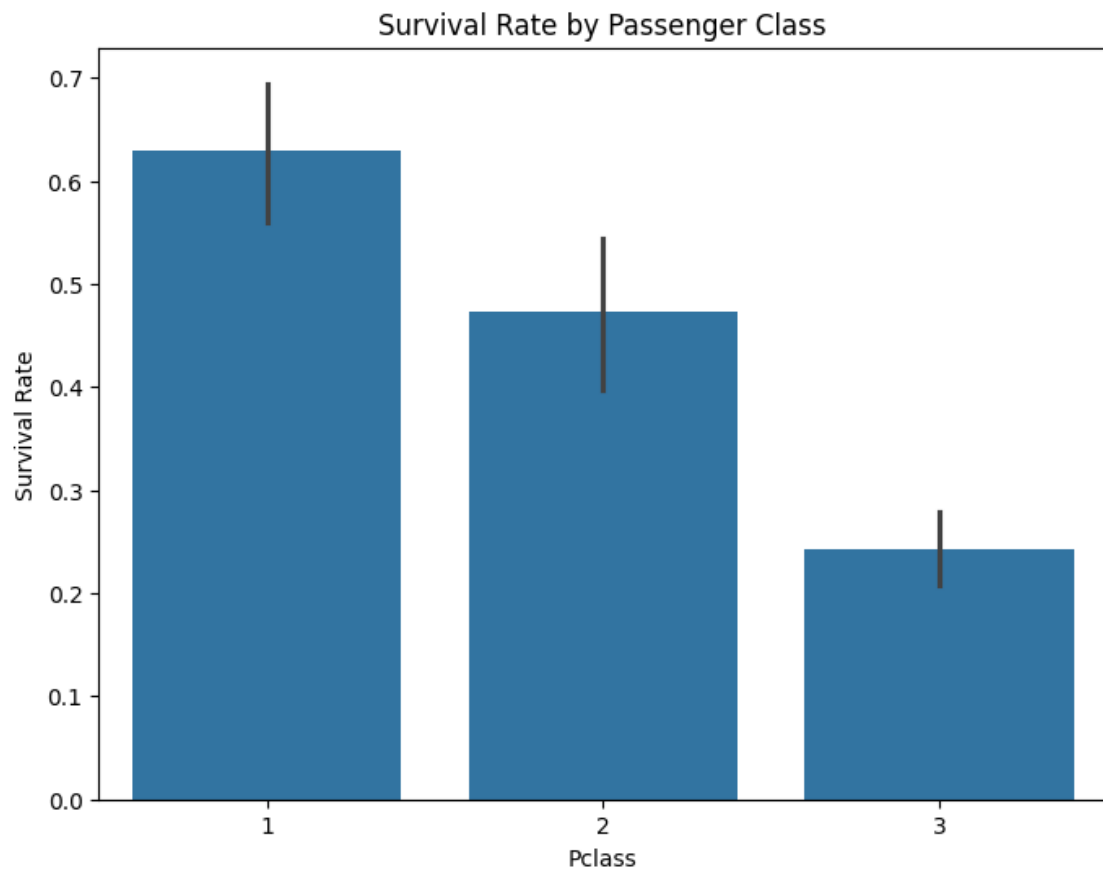
	PassengerId	Survived	Pclass	Age	SibSp \
count	891.000000	891.000000	891.000000	714.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
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

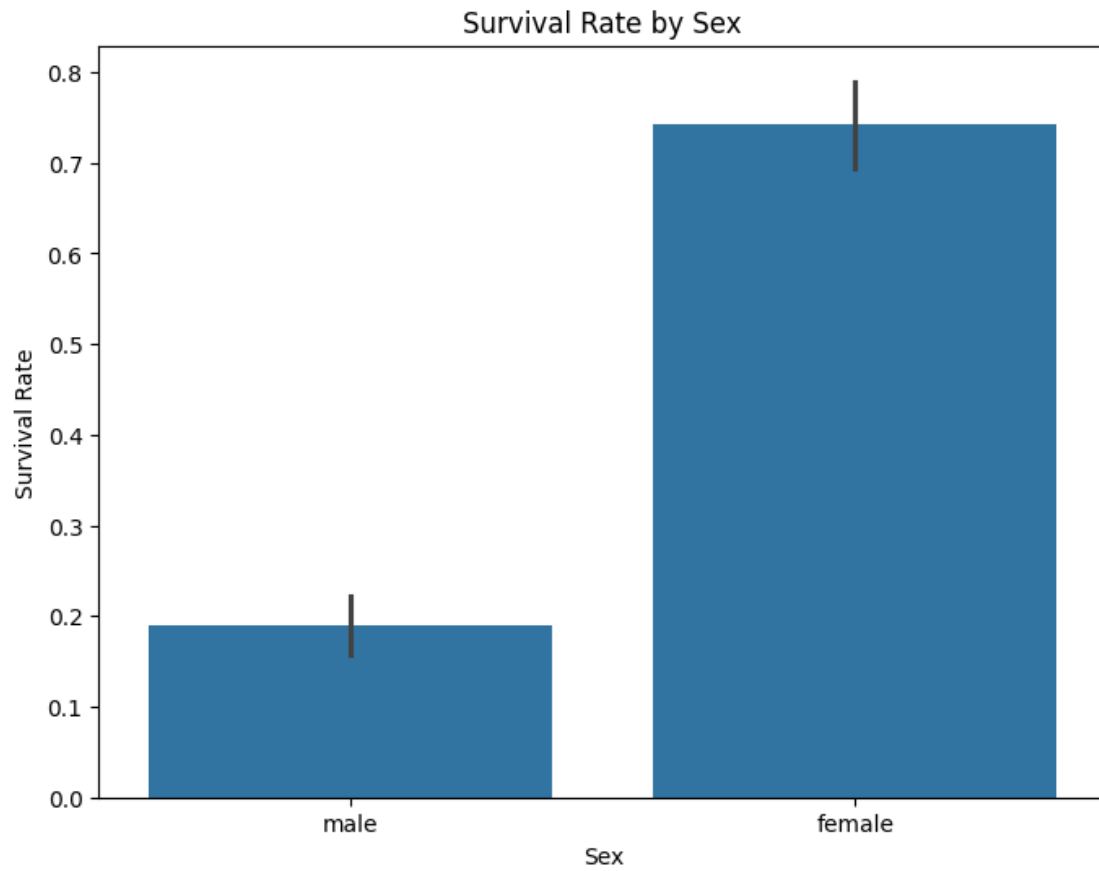
	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208

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

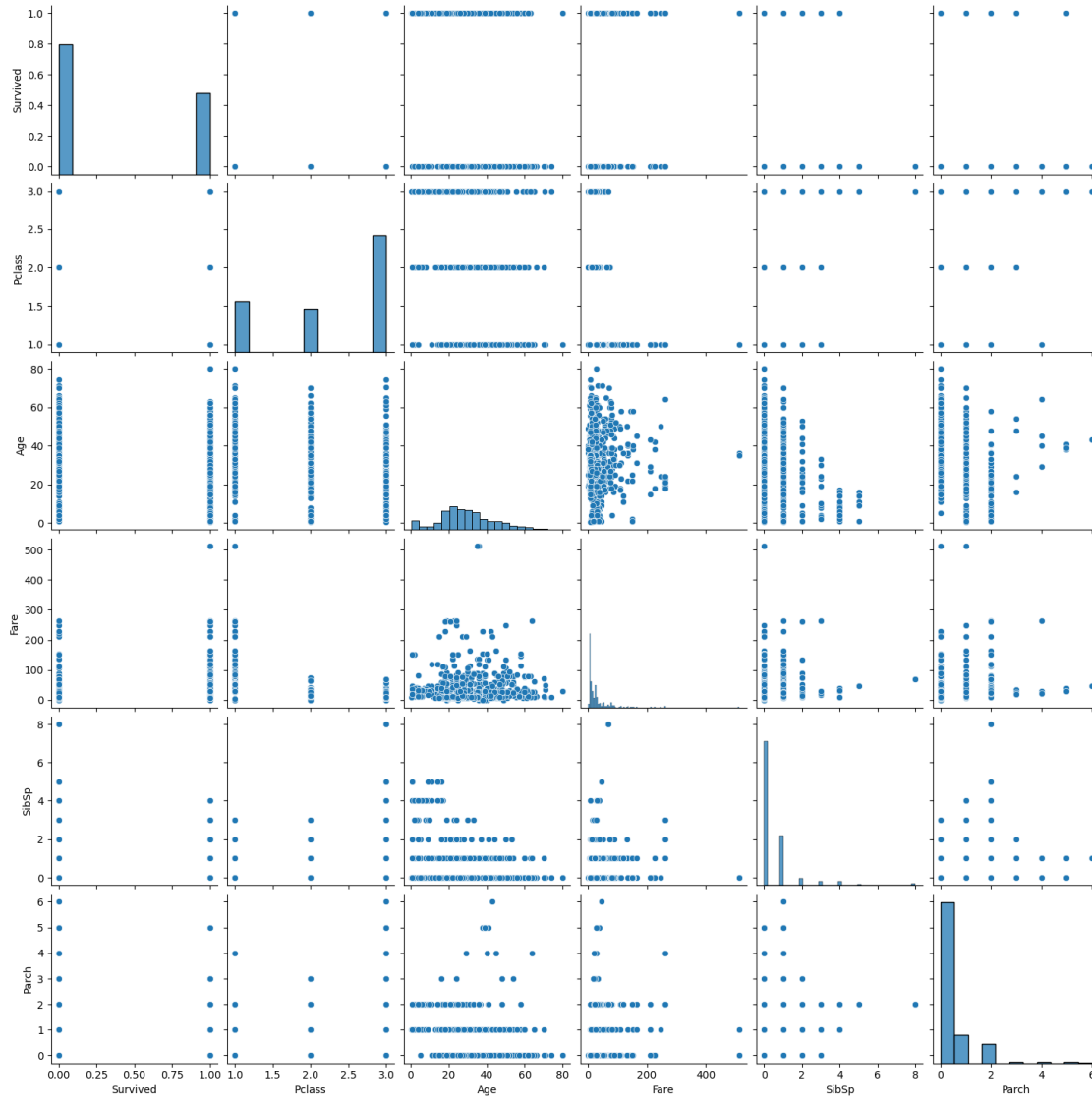








<Figure size 1200x1000 with 0 Axes>



ValueError

Traceback (most recent call last)

Cell In[4], line 61

```
58 plt.show()
```

```
60 plt.figure(figsize=(12, 8))
```

```
---> 61 sns.heatmap(data.corr(), annot=True, cmap='coolwarm', linewidths=0.5)
```

```
62 plt.title('Correlation Heatmap')
```

```
63 plt.show()
```

File

```
~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\frame
py:11022, in DataFrame.corr(self, method, min_periods, numeric_only)
11020 cols = data.columns
```



```

11021 idx = cols.copy()
> 11022 mat = data.to_numpy(dtype=float, na_value=np.nan, copy=False)
11024 if method == "pearson":
11025     correl = libalgos.nancorr(mat, minp=min_periods)

```

File

```

~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\frame
py:1981, in DataFrame.to_numpy(self, dtype, copy, na_value)
1979 if dtype is not None:
1980     dtype = np.dtype(dtype)
-> 1981 result = self._mgr.as_array(dtype=dtype, copy=copy, na_value=na_value)
1982 if result.dtype is not dtype:
1983     result = np.array(result, dtype=dtype, copy=False)

```

File

```

~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\internals\managers
py:1693, in BlockManager.as_array(self, dtype, copy, na_value)
1691     arr.flags.writeable = False
1692 else:
-> 1693     arr = self._interleave(dtype=dtype, na_value=na_value)
1694     # The underlying data was copied within _interleave, so no need
1695     # to further copy if copy=True or setting na_value
1697 if na_value is lib.no_default:

```

File

```

~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\internals\managers
py:1752, in BlockManager._interleave(self, dtype, na_value)
1750     else:
1751         arr = blk.get_values(dtype)
-> 1752     result[rl.indexer] = arr
1753     itemmask[rl.indexer] = 1
1755 if not itemmask.all():

```

ValueError: could not convert string to float: 'Braund, Mr. Owen Harris'

<Figure size 1200x800 with 0 Axes>

[]: