

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

# Загружаем датасет Titanic (встроен в seaborn)
df = sns.load_dataset('titanic')

# Выводим первые 5 строк
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	

Далее: [New interactive sheet](#)

```
# Размер данных
print("Размер датасета:", df.shape)

# Общая информация
print("\nИнформация о столбцах:")
df.info()

# Пропущенные значения
print("\nПропущенные значения по столбцам:")
print(df.isnull().sum())

# Описательная статистика для числовых признаков
print("\nОписательная статистика:")
df.describe()
```

Размер датасета: (891, 15)

Информация о столбцах:

```
<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	category
9	who	891 non-null	object
10	adult_male	891 non-null	bool
11	deck	203 non-null	category
12	embark_town	889 non-null	object
13	alive	891 non-null	object
14	alone	891 non-null	bool

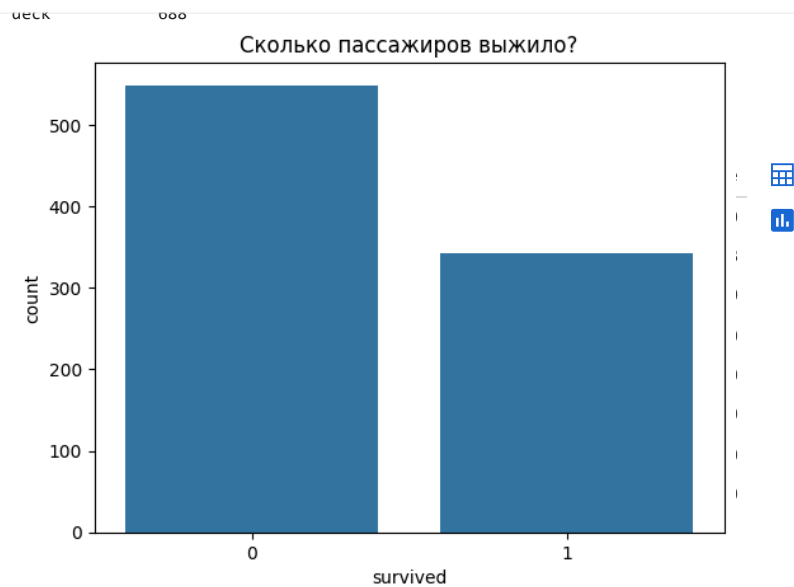
dtypes: bool(2), category(2), float64(2), int64(4), object(5)

memory usage: 80.7+ KB

Пропущенные значения по столбцам:

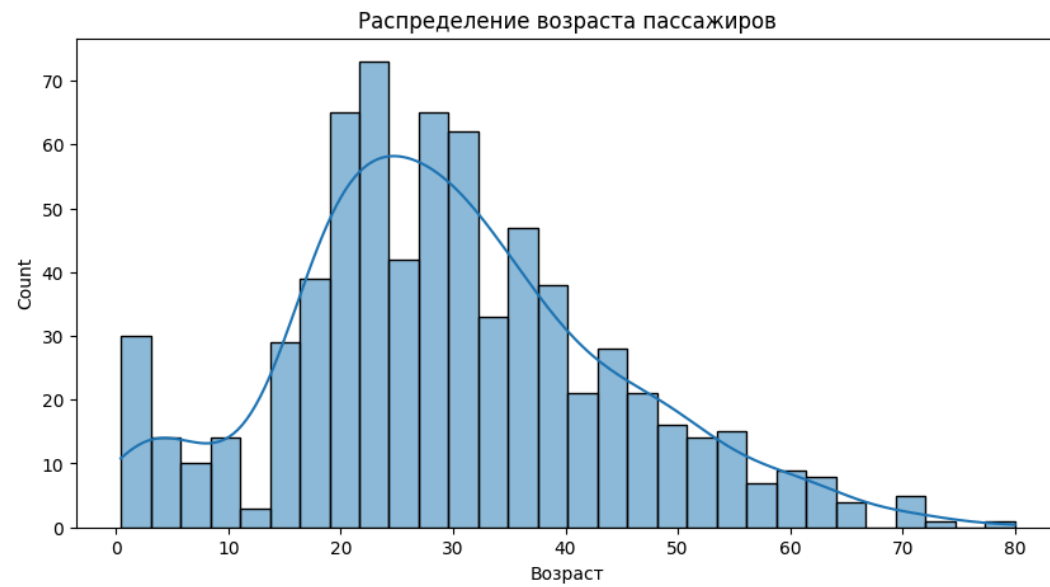
survived	0
pclass	0
sex	0
age	177
sibsp	0
parch	0
fare	0

```
sns.countplot(data=df, x='survived')
plt.title('Сколько пассажиров выжило?')
plt.show()
```

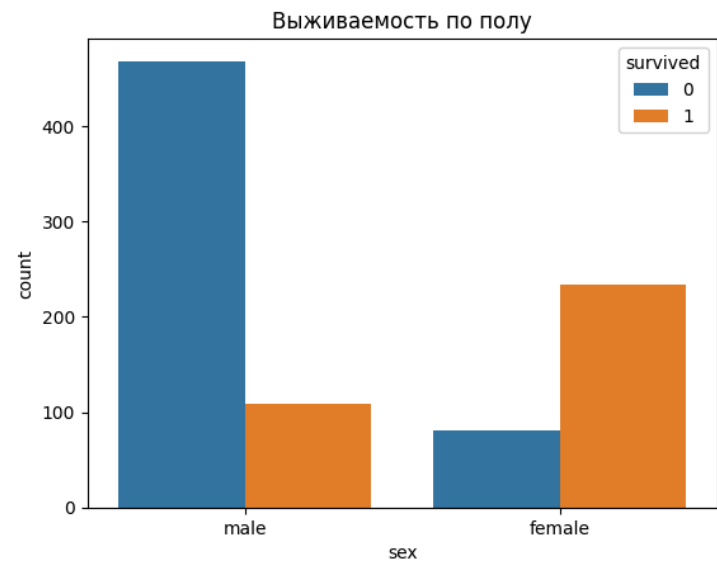


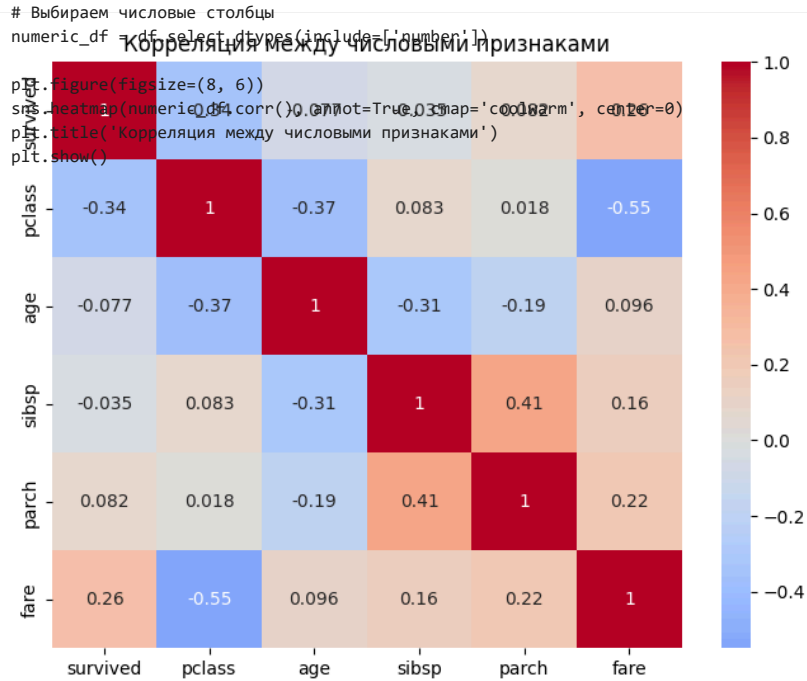
```
plt.figure(figsize=(10, 5))
sns.histplot(df['age'].dropna(), kde=True, bins=30)
plt.title('Распределение возраста пассажиров')
```

```
plt.xlabel('Возраст')  
plt.show()
```



```
sns.countplot(data=df, x='sex', hue='survived')  
plt.title('Выживаемость по полу')  
plt.show()
```





Уровень 1: Практика Построй график: выживаемость по классу каюты (pclass). Найди средний возраст выживших и не выживших. Сколько женщин выжило? Сколько мужчин?

Уровень 2: Исследование Создай новый признак: is_child — True, если возраст < 16, иначе False. Проанализируй выживаемость детей. Построй boxplot возраста по классу каюты (pclass). Есть ли выбросы?

Уровень 3: Творчество Выбери другой датасет из seaborn (например, tips, flights, iris) и проведи аналогичный EDA: загрузка, пропуски, 2–3 графика, выводы.