

Аналитический отчет по датасету Titanic

Автор: Аскар Подогов

Проект: Exploratory Data Analysis (EDA)

В данном отчете представлены результаты разведочного анализа данных о пассажирах Титаника. Цель исследования — выделить факторы, которые сильнее всего влияли на вероятность выживания.

1. Описание данных

Для анализа использовался классический датасет Titanic. Он содержит информацию о пассажирах: пол, возраст, класс билета, стоимость билета, порт посадки, наличие родственников на борту и признак выживаемости. Всего в выборке 891 наблюдение и 12 признаков.

2. Демография пассажиров

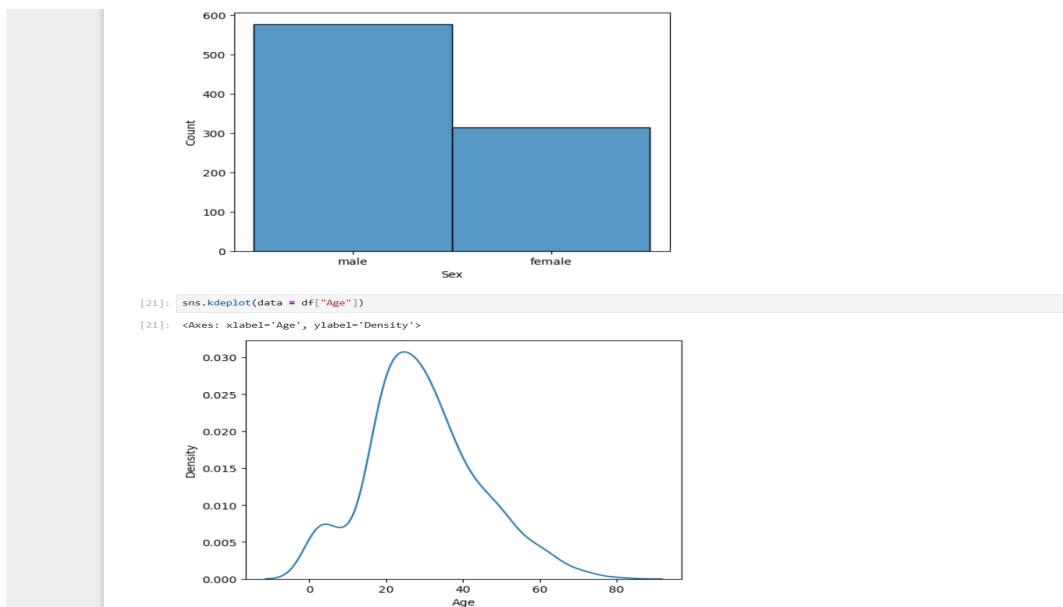


Рис. 1. Распределение пассажиров по полу и плотность распределения возраста.

На первом графике видно, что большинство пассажиров составляли мужчины. Плотность распределения возраста показывает, что основная масса пассажиров — взрослые люди 20–40 лет, но присутствуют и дети, и пожилые люди.

3. Класс билета и социальный статус



Рис. 2. Распределение классов билета и выживаемости по классам.

Анализ распределения классов показывает, что больше всего пассажиров путешествовали в 3-м классе. При этом выживаемость существенно выше у пассажиров 1-го класса по сравнению с 3-м.

4. Порт посадки (Embarked)

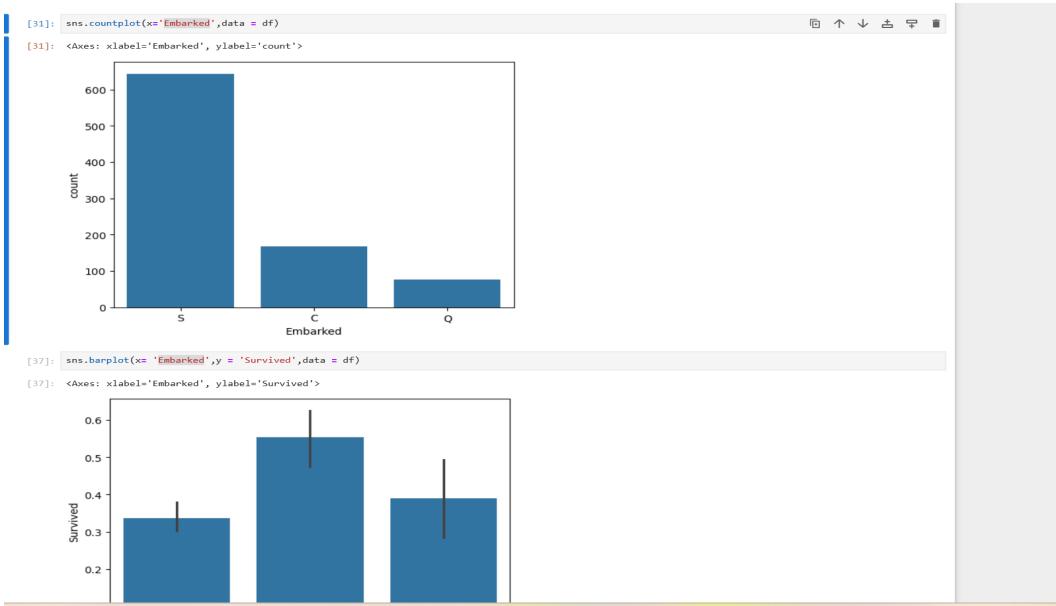


Рис. 3. Распределение пассажиров по порту посадки и выживаемость в разрезе портов.

Большинство пассажиров сели в порту Southampton (S). Пассажиры из Cherbourg (C) чаще покупали более дорогие билеты и относились к более высокому классу, в то время как из Queenstown (Q) в основном отправлялись пассажиры 3-го класса.

4. Порт посадки (продолжение)



Рис. 4. Стоимость билета по портам посадки (boxplot).

Boxplot стоимости билета по портам посадки подтверждает, что в порту Cherbourg медианная цена билета заметно выше, а разброс значений шире — там чаще встречаются дорогие билеты 1-2 класса.

5. Влияние класса и семейного статуса на выживаемость

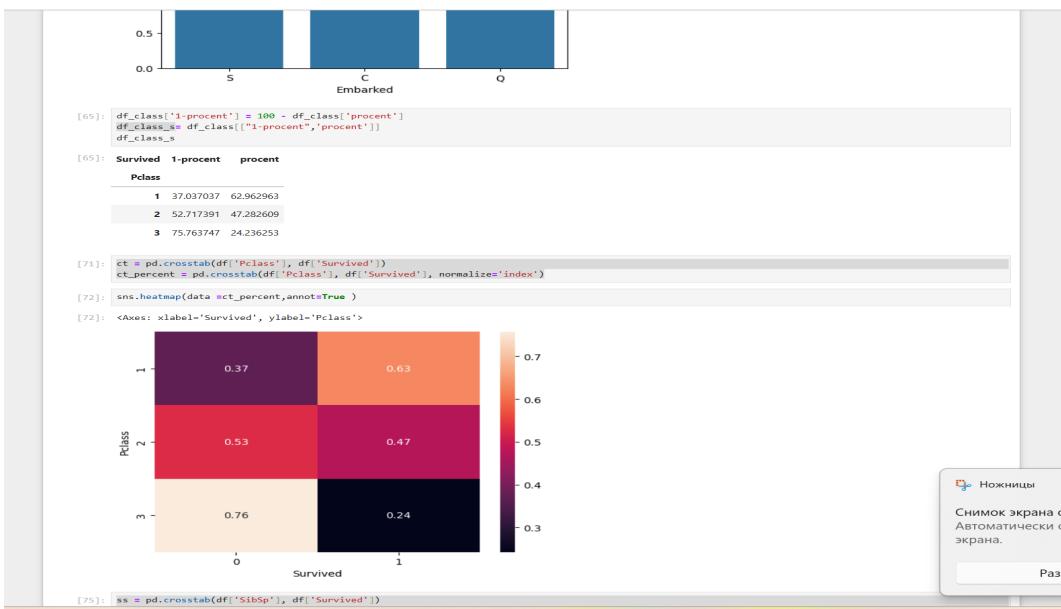


Рис. 5. Тепловая карта выживаемости по классам билета.

Тепловая карта по классам билета показывает, что доля выживших в 1-м классе существенно выше, чем во 2-м и особенно в 3-м. Это подтверждает сильное влияние социального статуса и условий размещения на вероятность спасения.

5. Влияние класса и семейного статуса на выживаемость (продолжение)

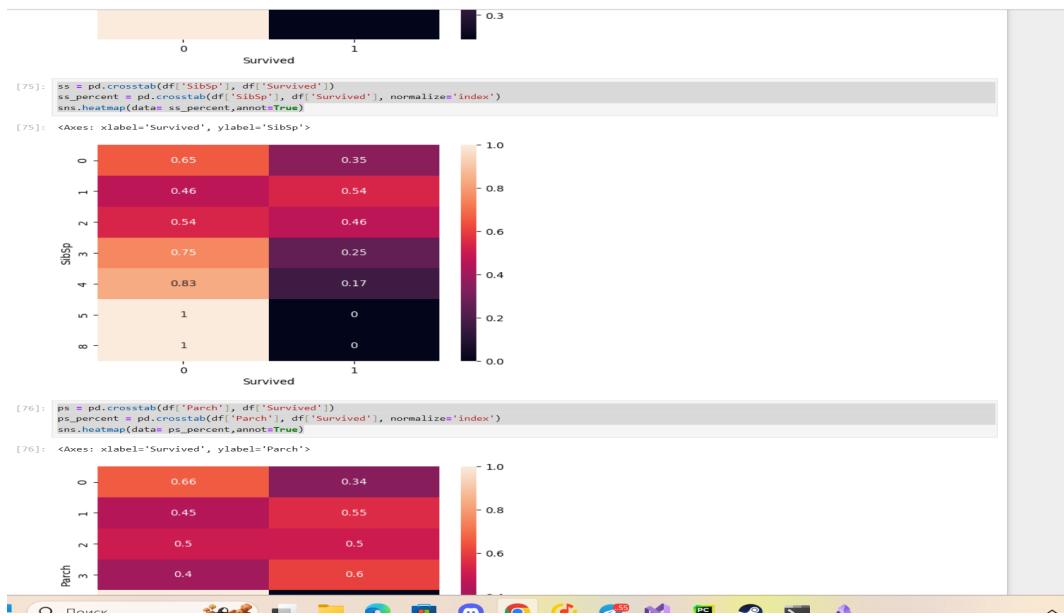


Рис. 6. Тепловые карты выживаемости в зависимости от количества родственников (SibSp, Parch).

Тепловые карты для признаков SibSp и Parch показывают, что наибольшие шансы на выживание были у пассажиров, путешествовавших с 1-2 родственниками. Большие семьи (3 и более родственников) выживали заметно хуже, так же как и полностью одинокие пассажиры.

6. Взаимосвязи между признаками (pairplot)

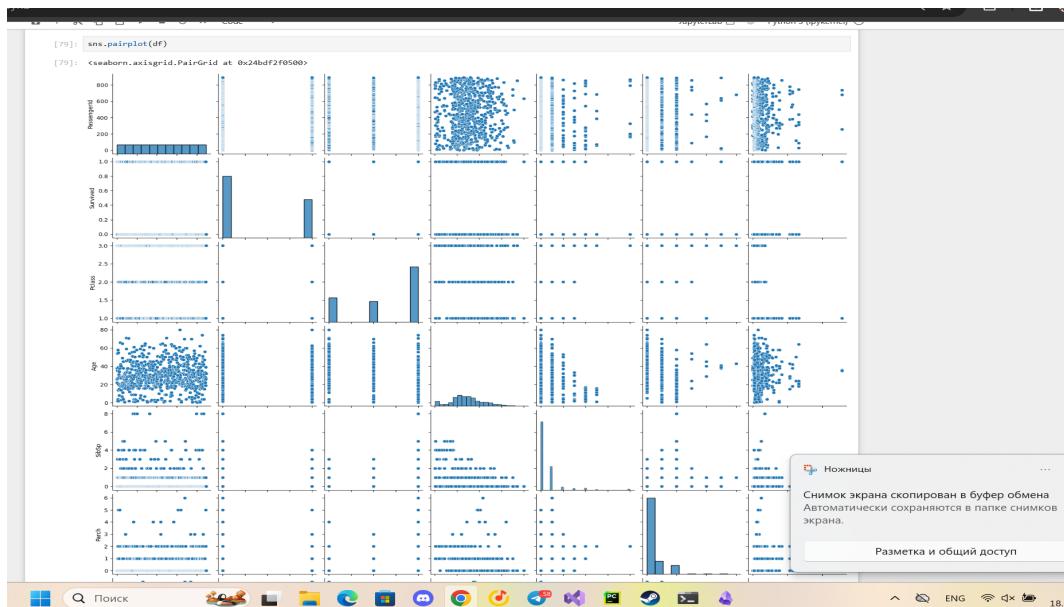


Рис. 7. Pairplot по основным числовым признакам датасета.

Pairplot позволяет визуально оценить взаимосвязи между числовыми признаками. Видно, что стоимость билета и класс билета сильно связаны между собой, а также что выжившие чаще сосредоточены в области более высоких тарифов и младших номеров классов.

7. Итоговые выводы

На основе проведенного анализа можно выделить следующие ключевые факторы, влияющие на выживаемость пассажиров Титаника:

- 1) Класс билета (Pclass). Пассажиры 1-го класса выживали значительно чаще пассажиров 3-го класса. Класс билета отражает уровень комфорта и близость к палубам, откуда осуществлялась эвакуация.
- 2) Стоимость билета (Fare). Более дорогие билеты связаны с более высокой вероятностью выживания. Этот признак тесно коррелирует с классом билета и косвенно характеризует социальный статус пассажира.
- 3) Порт посадки (Embarked). Пассажиры, севшие в Cherbourg, чаще относились к более обеспеченным слоям и имели лучшие условия размещения, что положительно влияло на их шансы.
- 4) Наличие семьи на борту (SibSp, Parch). Небольшие семьи (1-2 родственника) имели более высокие шансы на спасение, в то время как большие семьи и одиночные пассажиры выживали реже.

В целом результаты подтверждают, что выживаемость на Титанике во многом определялась социальным статусом пассажира: класс билета, стоимость билета и условия размещения играли ключевую роль. Дополнительное влияние оказывали структура семьи и порт посадки.