

## Decision Tree - SpaceShip Titanic

Ежова Елена, группа 5030102/10401

В данном коде используется метод **решающих деревьев** (Decision Trees) для классификации данных о пассажирах космического корабля.

### **Принцип метода решающих деревьев (Decision Trees)**

Решающее дерево — это метод машинного обучения, который использует дерево решений для моделирования данных и принятия решений. Каждый внутренний узел дерева представляет собой тест на определенный атрибут (например, возраст, планета), каждая ветвь — результат этого теста, а каждый лист — итоговый класс или прогнозируемое значение.

### **Основные этапы метода решающих деревьев:**

#### *1. Создание дерева решений:*

- На первом этапе алгоритм выбирает атрибут (или комбинацию атрибутов), который будет разделять набор данных на подгруппы наилучшим образом.
- Для каждого разделения алгоритм задает правило, используя наиболее значимые атрибуты и делит набор данных на группы.
- Этот процесс продолжается до тех пор, пока не будет достигнут критерий завершения (например, глубина дерева или количество объектов в группе меньше заданного значения).

#### *2. Прогнозирование:*

- Когда новые данные поступают на вход, алгоритм начинает от корня дерева и следует по ветвям вниз, делая выбор на основе каждого разделения. Каждый следующий узел дерева соответствует решению о том, какой путь продолжить (например, назначить ли пассажира на транспортировку или нет).
- Дерево продолжается до достижения листа, который определяет конечное решение для объекта.

## Применение метода в коде:

В данном коде решающее дерево используется для предсказания, транспортированы ли пассажиры на космическом корабле. Процесс начинается с подготовки данных: заполнение пропусков, применение `pd.get_dummies` для категориальных переменных и деление данных на обучающие и тестовые выборки. Затем создаётся модель решающего дерева (`DecisionTreeClassifier`), которая обучается на тренировочных данных и тестируется на тестовых данных. Результаты предсказаний оцениваются с помощью метрик `accuracy_score`, `classification_report` и визуализации `confusion_matrix`. Важность признаков оценивается с помощью `feature_importances_`

## Результаты:

График распределения пассажиров с разных планет:

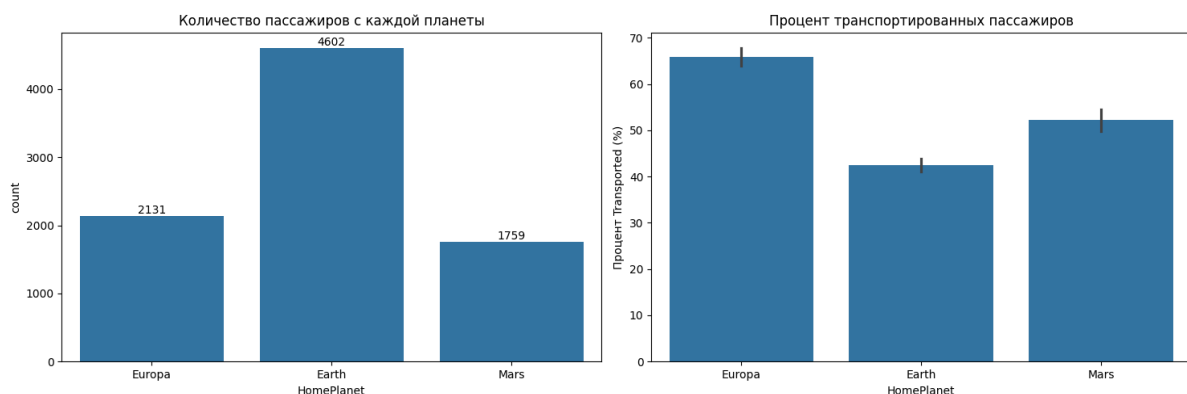
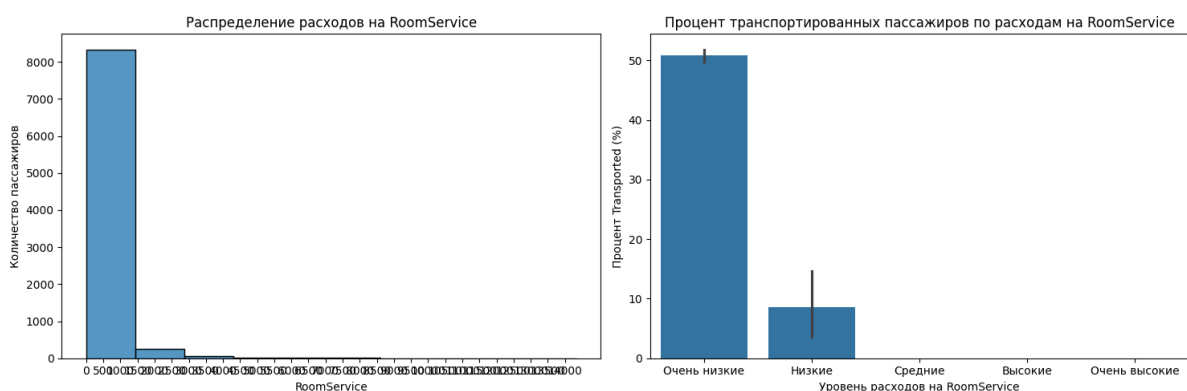
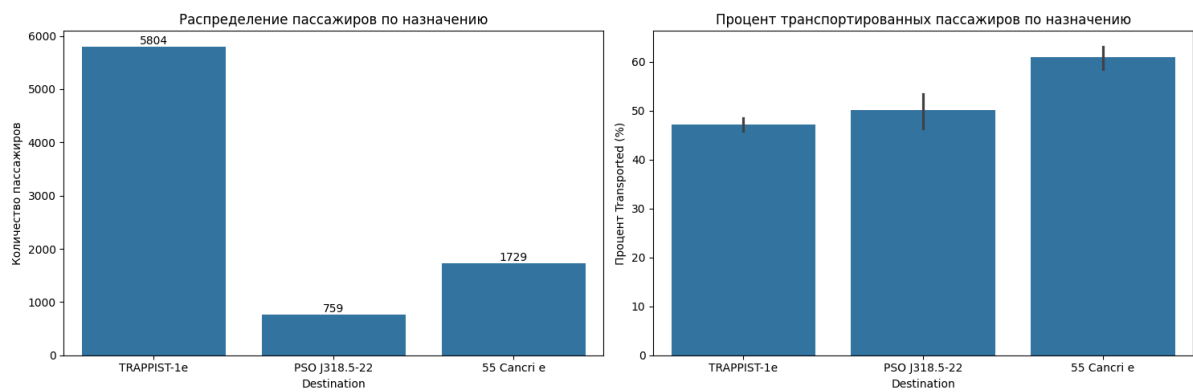


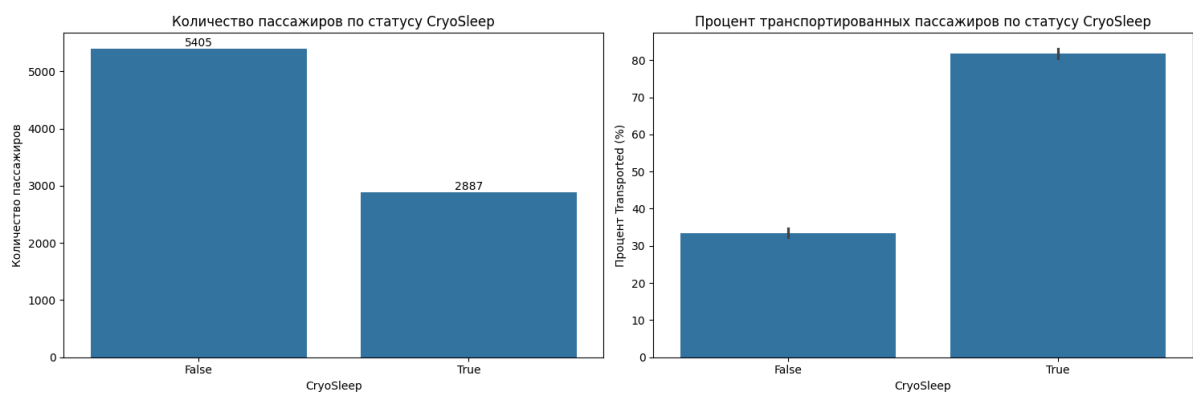
График распределения пассажиров по расходам на RoomService:



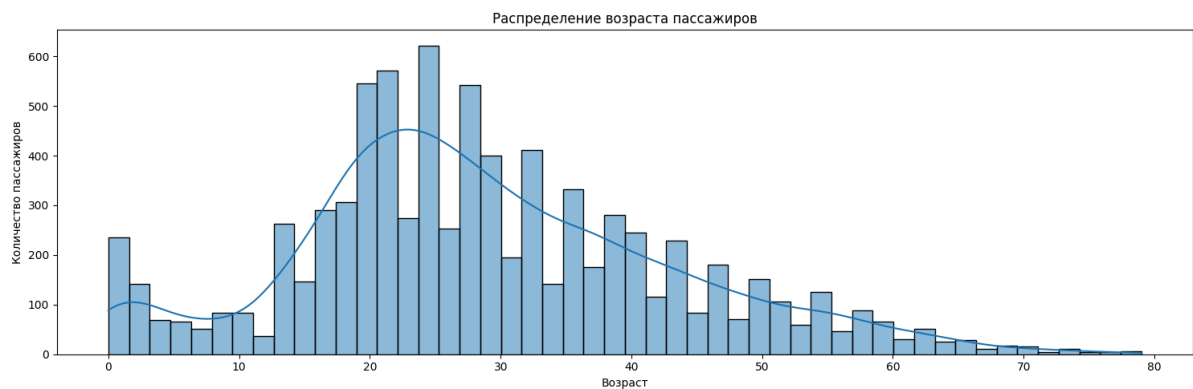
## График распределения пассажиров по назначению (Destination):



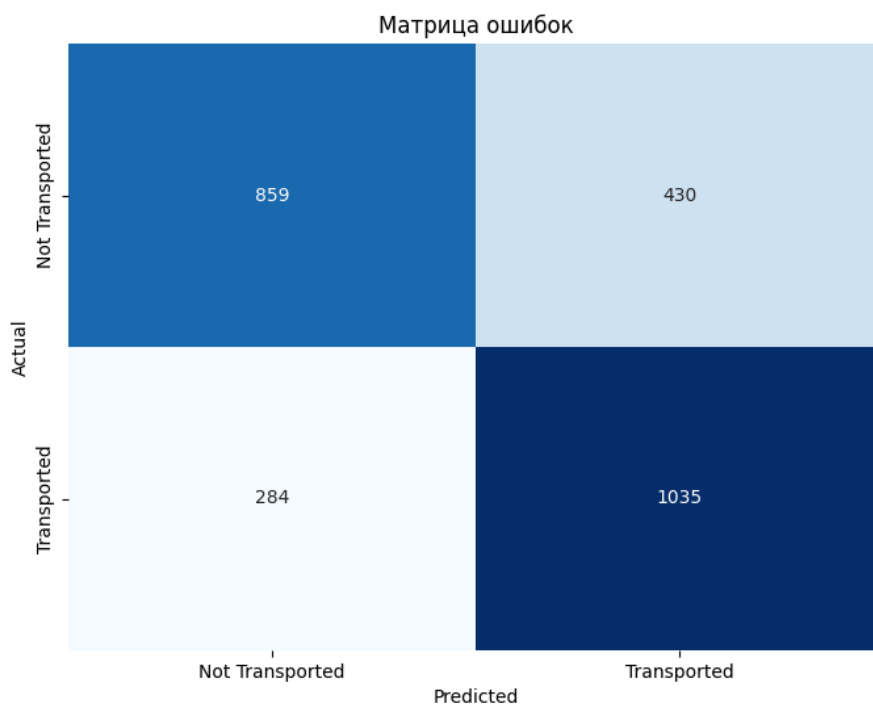
## График количества пассажиров в зависимости от статуса CryoSleep:



## Гистограмма и плотность распределения возраста:



## Матрица ошибок:



## Гистограмма важности признаков:

