**Туториал**

**1. Описание данных:**

Набор данных содержит транзакции кредитных карт европейцев. Набор данных представляет транзакции за 2 дня, где есть 492 мошенника и 284807 транзакций.

**1.1. Цель исследования:**

изучение подготовки данных, их анализа, построение модели, сравнение модели на обученных данных и данных для тестирования.

**2. Первичный анализ данных:**

Набор не сбалансирован. Мошенничества составляют 0,172% от всех транзакций. Данные содержат только числовые значения, что является PCA трансформацией.

**2.1. Исследование признаков:**

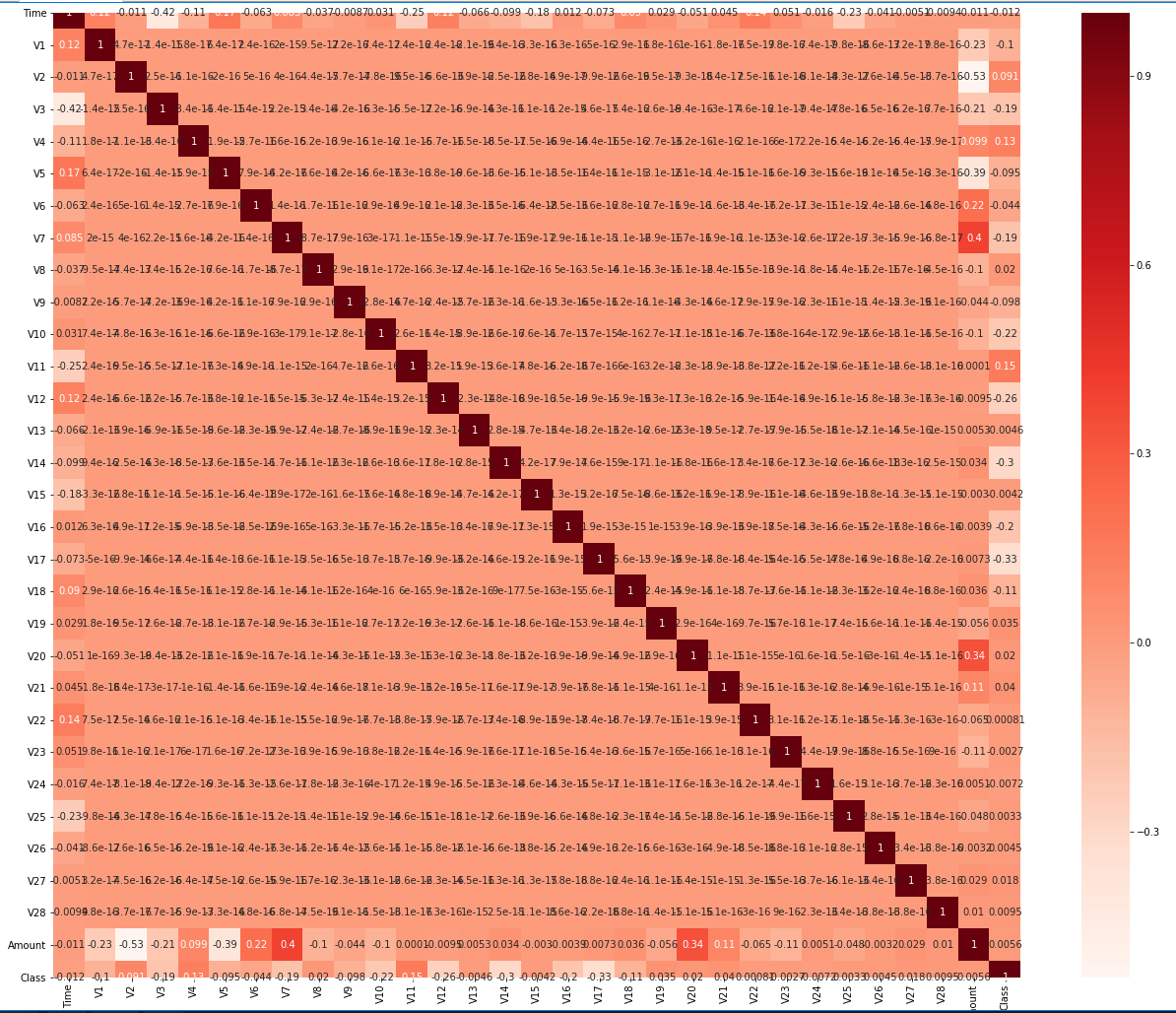
Мы не можем использовать оригинальные значения, свойства и фишки. Свойства называются **V1, V2, V3, …, V28**. Только **Time** и **Amount** не трансформированы. Time содержит время между первой транзакций и текущей транзакцией данных. Функция "Amount" - это сумма транзакции, эта функция может использоваться, например, для зависимого от затрат обучения. Class содержит 1 ,если это мошенничество и 0, если нет.

**2.2. Исследование целевой переменной:**

???

**2.3. Первичный визуальный анализ данных:**

Корреляция



**2.4. Найденные зависимости: (описание на основе корреляции)**

**Можно заметить, что некоторые квадраты более красные, чем другие. Так мы можем вывести некоторые зависимости.**

**Amount and V6 = 0.22**

**Amount and V7 = 0.4**

**Amount and V20 = 0.34**

**Amount and V21 = 0.11**

**V1 and Time = 0.12**

**V5 and Time = 0.17**

**V12 and Time = 0.12**

**V22 and Time = 0.14**

**Можно выявить, что все основные зависимости строятся на параметрах Amount and Time.**

**3.**