Министерство науки и высшего образования Российской Федерации  
Пензенский государственный университет  
Кафедра вычислительная техника

**ОТЧЕТ**

по лабораторной работе №6

по дисциплине «Программные средства кибертехнических систем»

на тему «Нормализация набора данных»

выполнили: ст -ты гр. 22ВВП1

Захаров А. С.

Беляев Д. И.

Демин М. С

Проверили:

Карамышева Н.С.

Зинкин С.А.

Пенза 2025

**Задание.**

1. Подготовить полученный набор данных в соответствии с этапами 3-й фазы методологии CRISP.

2. Описать использованные решения.

3. Организовать дополнительный сбор данных (если это необходимо и возможно).

4. Оформить отчёт, содержащий информацию о выполнении пунктов 1-3 задания.

Вариант 6. Задача кредитного скоринга (15-й столбец набора данных

{1 – кредит выдан, 2 – кредит не выдан}) по 14 признакам.

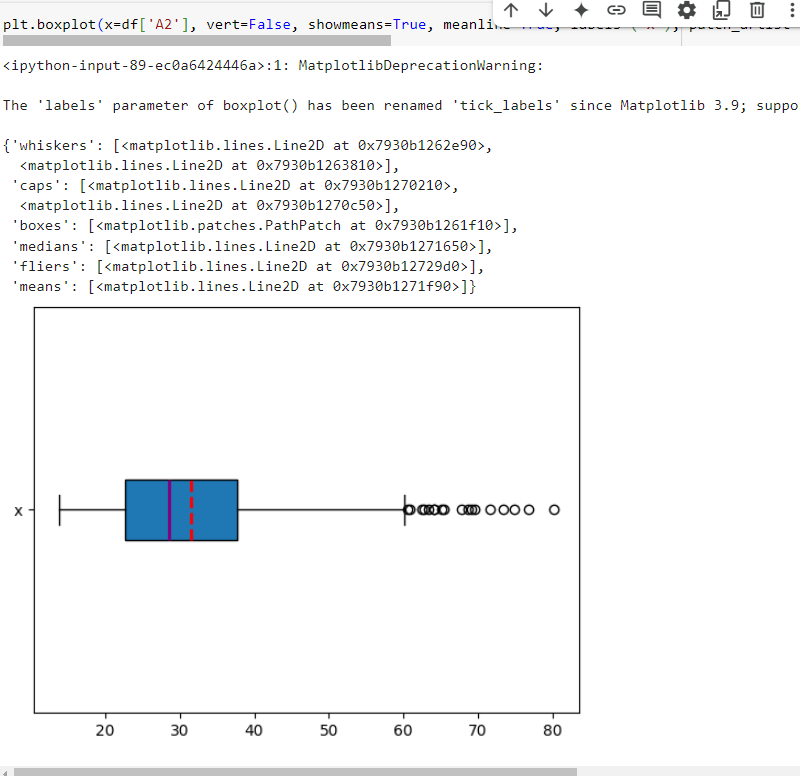
**Ход работы**

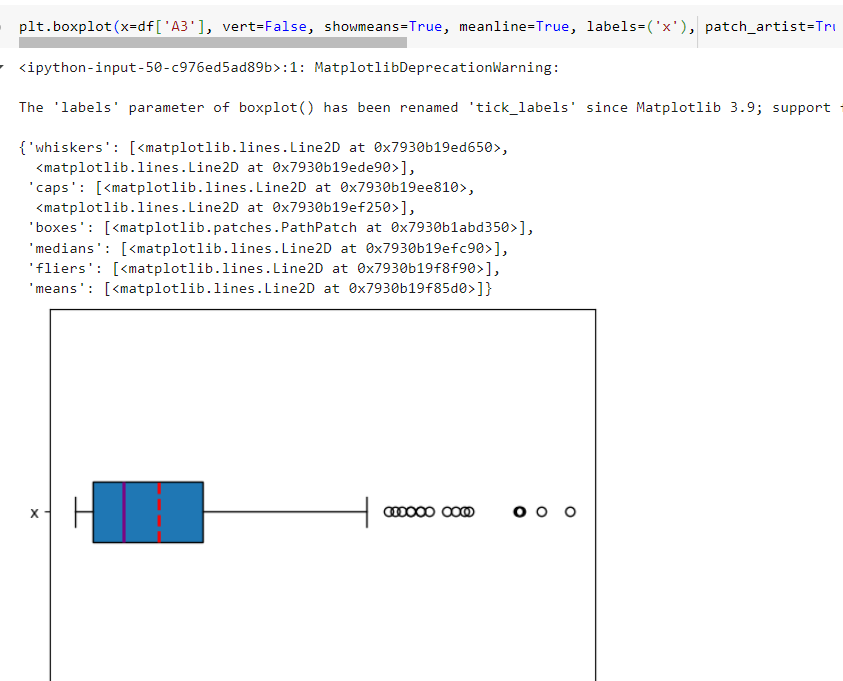
В ходе предыдущего анализа выявлены следующие особенности датасета:

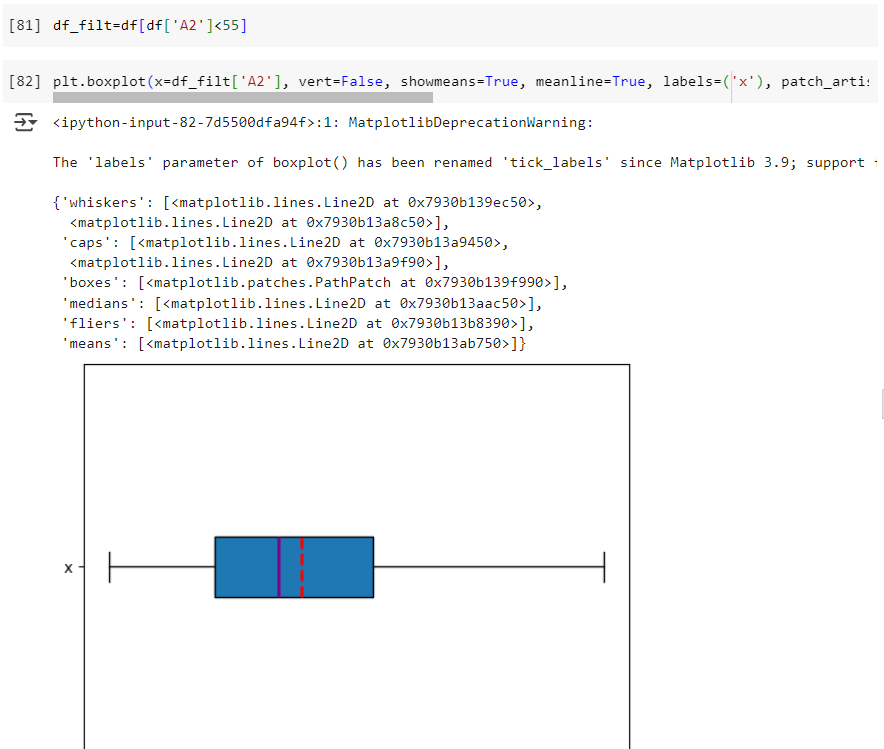
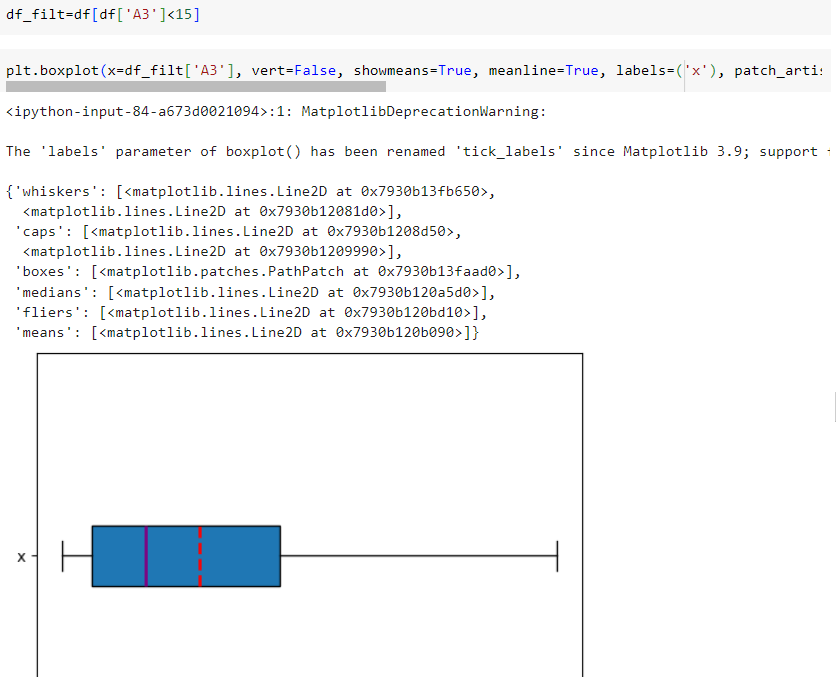
* Отсутствие дубликатов.
* Отсутствие пропусков.
* Присутствие выбросов во всех числовых признаках.
* Отсутствие категориальных признаков.

В ходе выполнения этого лабораторного задания нам нужно нормализовать значения.

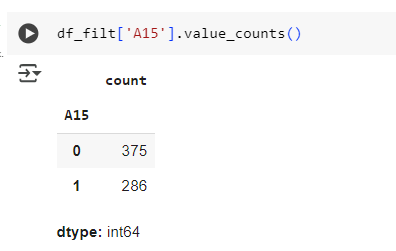
Избавимся от выбросов перед этим проанализировав их количество.  
Посмотрели выбросы в атрибутах А2 и А3



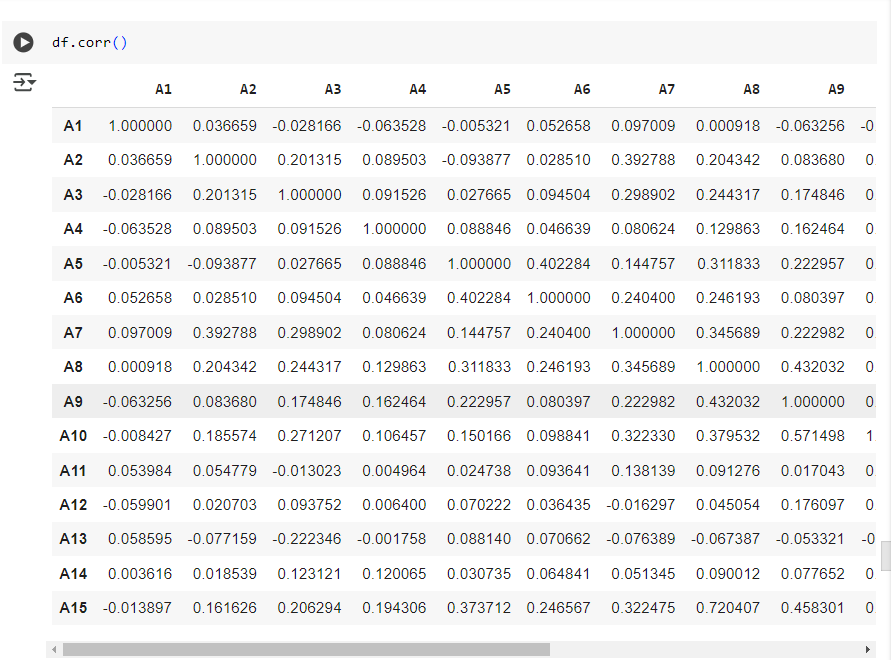


Отобразили признаки без выбросов путем фильтрации 

Оценив ситуацию по признакам А2, А3 принято решение удалить выбросы, так как неизвестно, нормальное ли такое значение для признака или нет.



Высоких корреляций не наблюдается



**Вывод**

Научились нормализировать датасеты.