ГУАП

КАФЕДРА № 41

ОТЧЕТ   
ЗАЩИЩЕН С ОЦЕНКОЙ

ПРЕПОДАВАТЕЛЬ

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ассистент |  |  |  | В.В. Боженко |
| должность, уч. степень, звание |  | подпись, дата |  | инициалы, фамилия |

|  |
| --- |
| ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ №2 |
| ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ДАННЫХ |
| по курсу: ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ |
|  |
|  |

РАБОТУ ВЫПОЛНИЛ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| СТУДЕНТ ГР. № | 4917 |  |  |  | В.С. Тихонов |
|  |  |  | подпись, дата |  | инициалы, фамилия |

Санкт-Петербург

2022

**Цель работы:** осуществить предварительную обработку данных csv-файла, выявить и устранить проблемы этих данных.

Вариант 21. Ссылка на репозиторий: https://github.com/Deiwas/analysis\_lab2

1. Получен набор данных от преподавателя для проведения анализа согласно цели выполняемой работы (рис. 1).

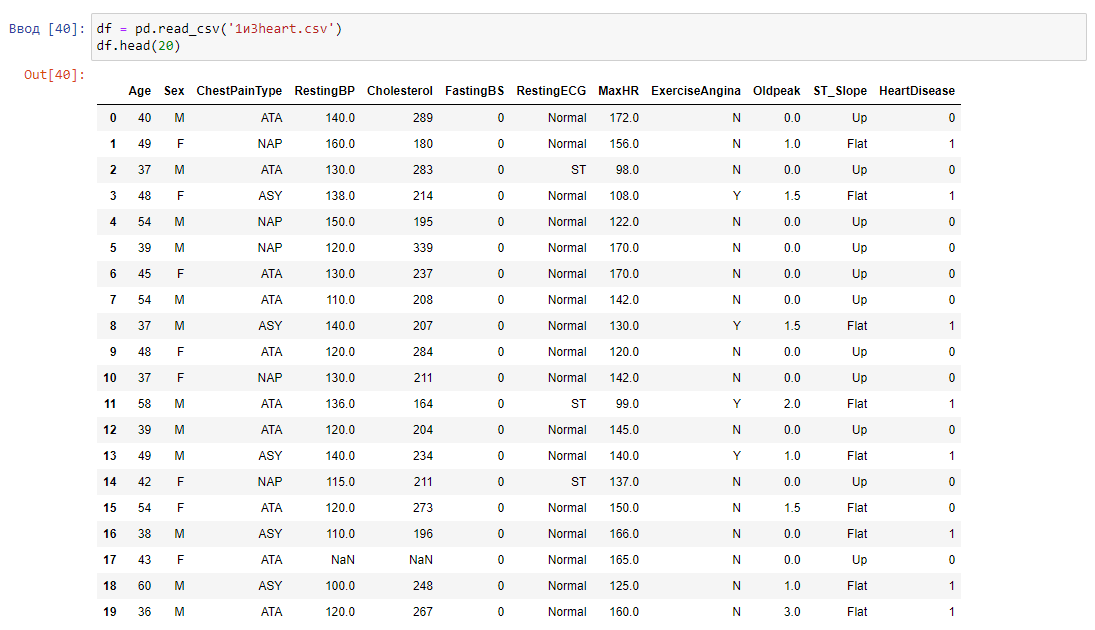


Рисунок 1 – Загруженный датафрейм

1. Проведена предварительная обработка данных. В процессе ее выполнения удалены дубликаты (рис. 2), пустые строки (рис. 3), проведена конвертация данных (рис. 4), а также проведена реиндексация датасета (рис. 5).

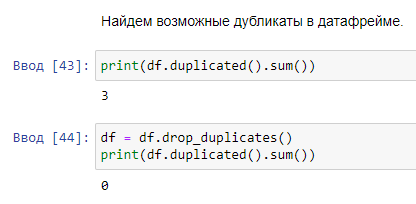


Рисунок 2 – Удаление дубликатов

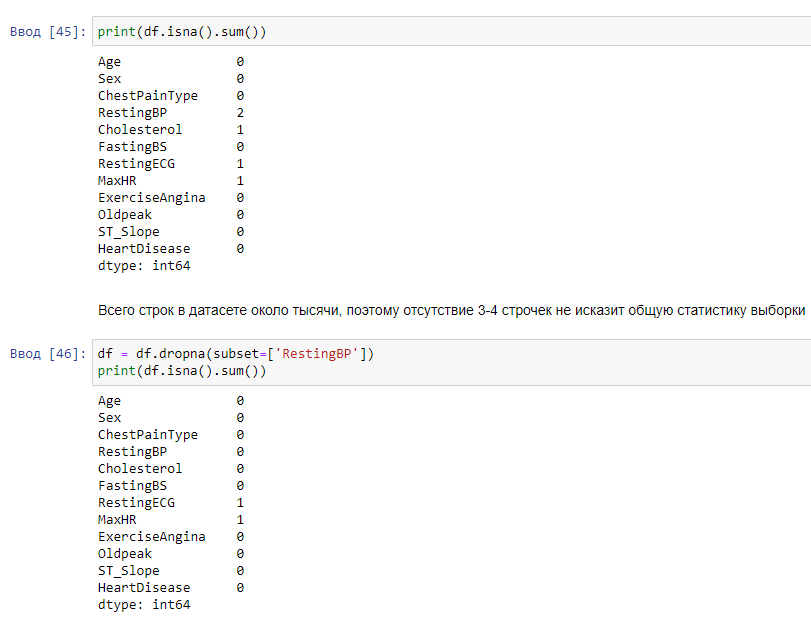


Рисунок 3 – Удаление пустых строк

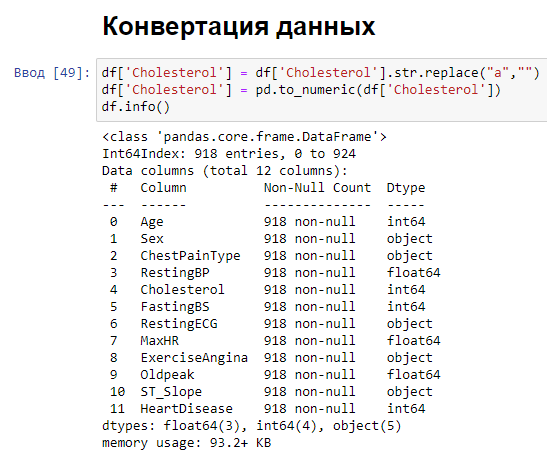


Рисунок 4 – Конвертация столбца «Cholesterol»

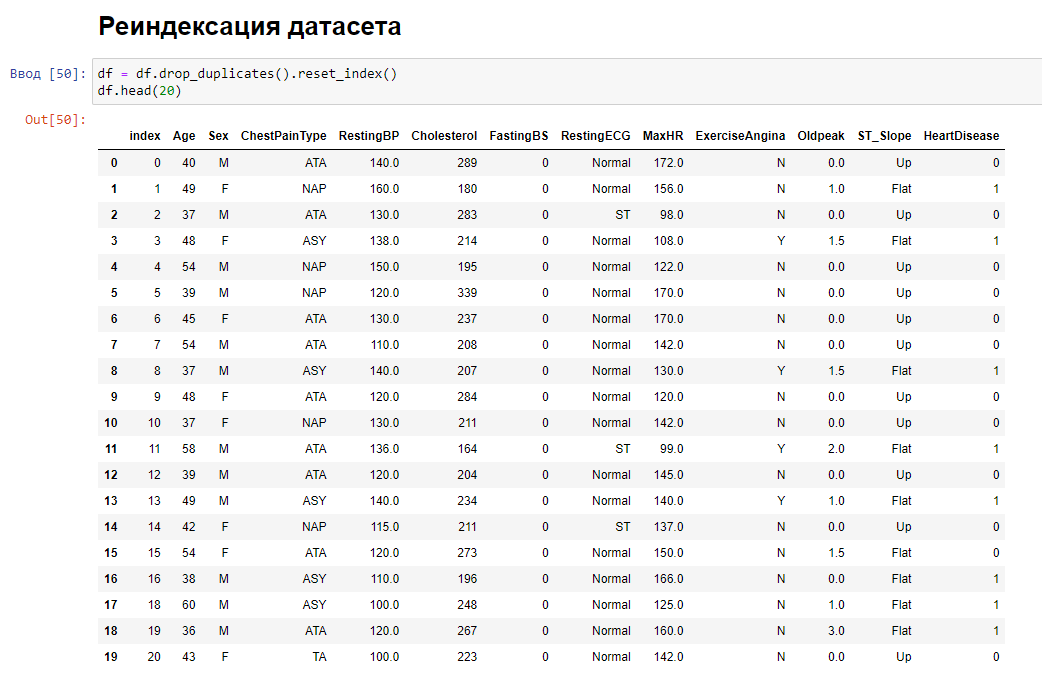


Рисунок 5 – Реиндексация датасета

1. Была проведена стандартизация и нормализация данных датафрейма с помощью класса StandardScaler().

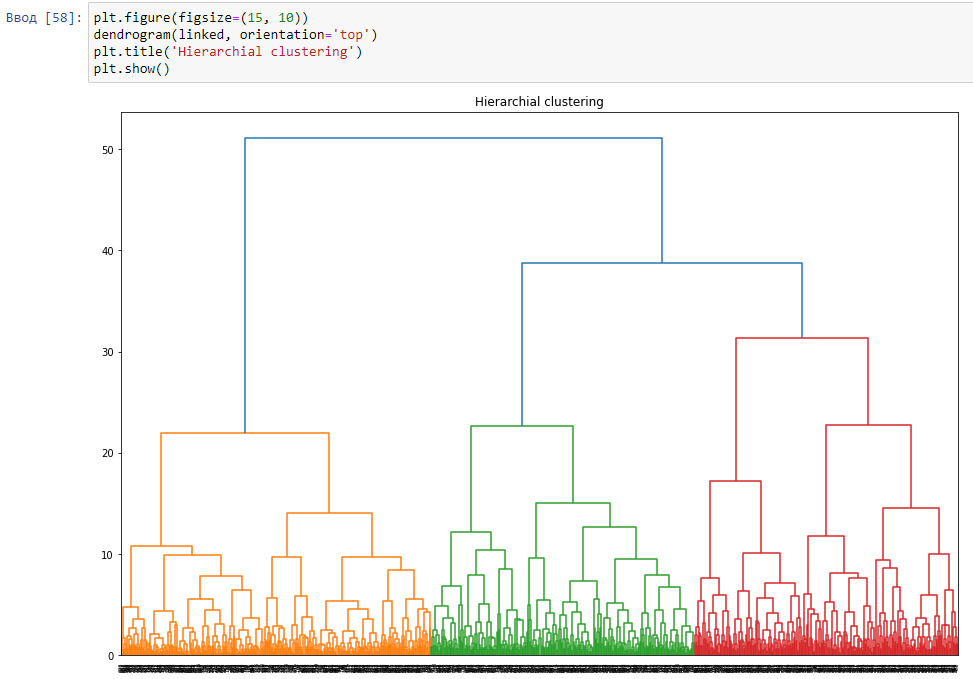


Рисунок 6 – Кластеризация датафрейма

1. Предположим, что число кластеров равно 3. Определим признаки, оказавшие наибольшее влияние на выделение кластеров (рис. 7).

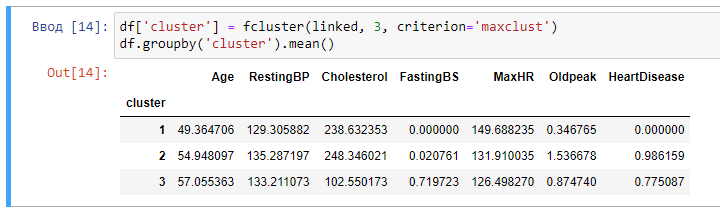


Рисунок 7 – Таблица с средними значениями столбцов по кластерам

Вывод: в ходе выполнения лабораторной работы были изучены алгоритмы и методы кластерного анализа. Над набором данных была проведена кластеризация иерархическим агломеративным методом. Было выявлено 3 кластера - "Здоровые люди", "Люди с высоки уровнем депрессии ST", "Люди с диабетом/пониженным холестерином".

Список источников:

1. Методические указания к лабораторной работе №3: https://pro.guap.ru/get-task/d67f94c9d2eb770cb301dc2759523ebd