МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра «Информационные системы»

Дисциплина «Анализ многомерных данных»

**Лабораторная работа № 1**

Подготовка и оценка набора многомерных данных

Выполнил:

студент гр. ИИмд-21

Шевалдов С.С

Проверил:

Доцент, к.т.н., доцент

Романов А.А.

Ульяновск,   
2023 г.

**Цель работы**: подготовить набор данных

Задачи:

1. Найти открытый набор многомерных данных.
2. Выделить из всего множества атрибутов самые значимые для дальнейшего анализа.
3. Нормализовать данные атрибутов объектов.
4. Выбрать шкалу для преобразования качественных значений и представить полученный набор данных
5. Составить отчет со следующей структурой:
   1. Цель работы.
   2. Описание набора многомерных данных.
   3. Описание решения поставленных задач.
   4. Список использованных источников.
6. Оформить и защитить отчет.

**Описание набора многомерных данных:** набор данных содержит в себе информацию о характеристиках заёмщиков и статусах их заявок на кредиты, такие как:

1. Возраст.
2. Профессия.
3. Семейное положение.
4. Образование.
5. Наличие дефолтов в прошлом.
6. Баланс на банковском счету.
7. Статус наличия недвижимого имущества.
8. Статусе о наличие непогашенного кредита.
9. Информация о способе связи.
10. Статус одобрения заявки на кредит.

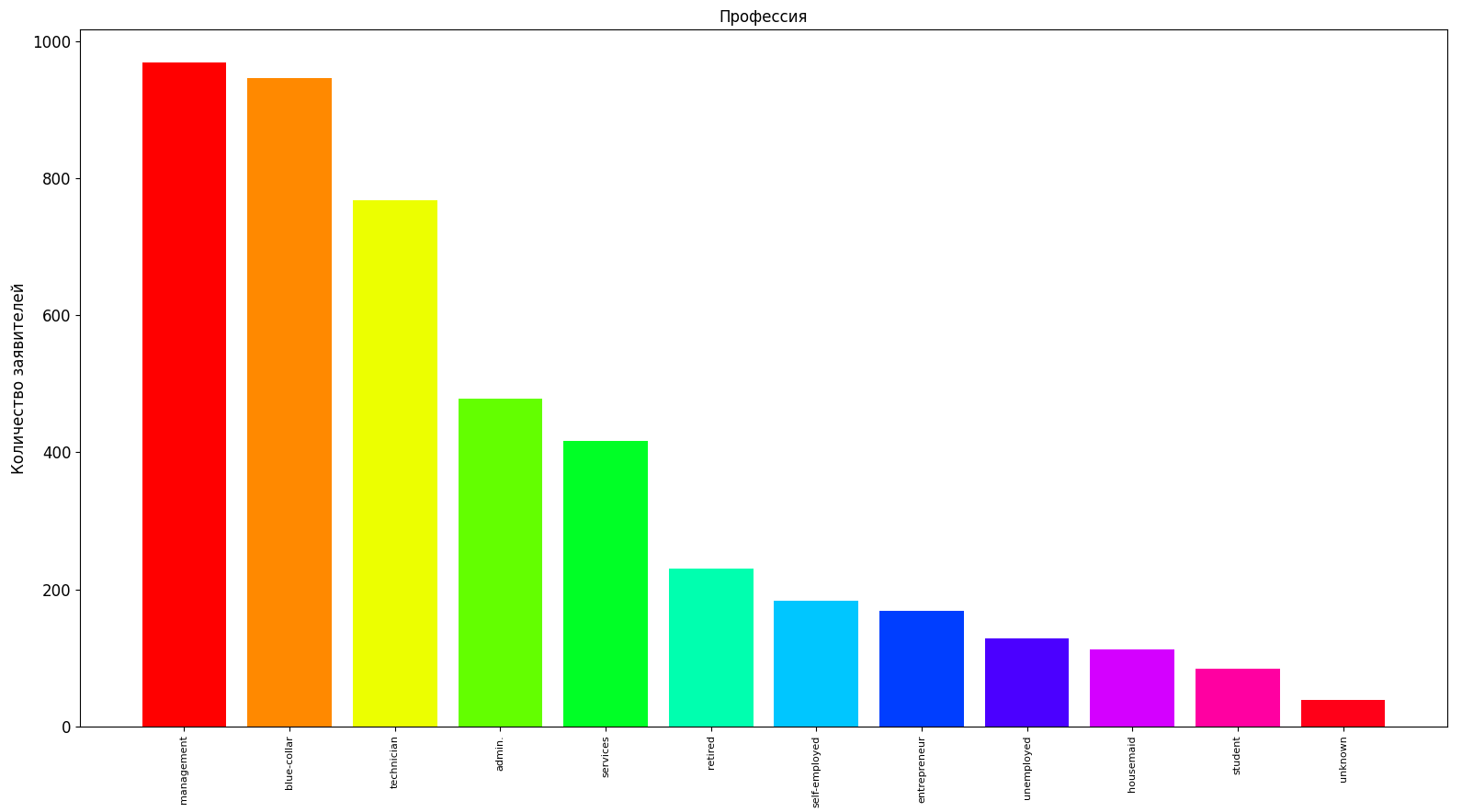
Для решения задачи анализа кредитных заявок были определены наиболее влияющие на решение о выдаче кредита характеристики заёмщиков. К ним относятся:

1. Наличие дефолтов в прошлом.
2. Возраст.
3. Статус о наличии непогашенного кредита.
4. Профессия.

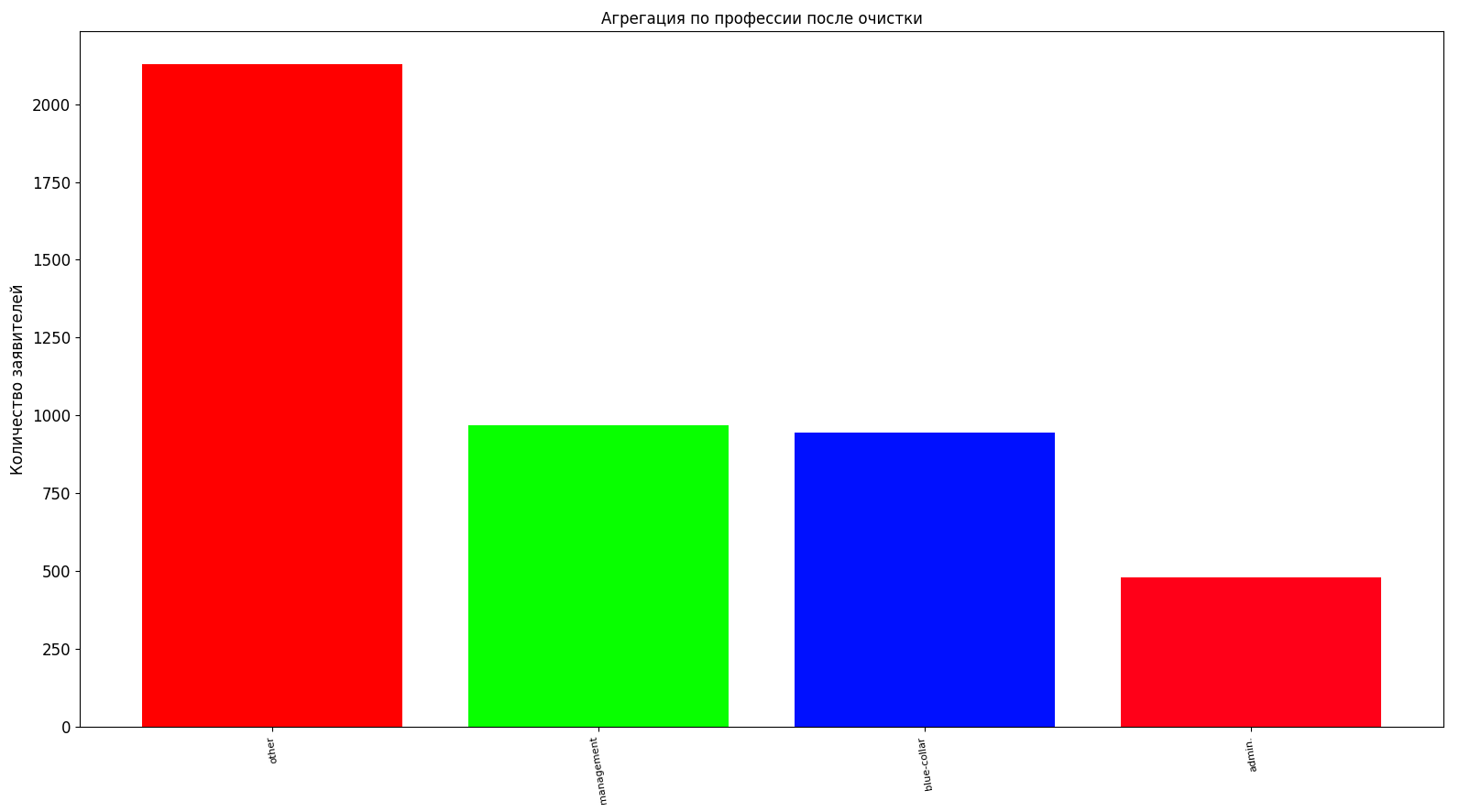
Порядок выполнения работы

В ходе лабораторной работы был проанализирован вышеописанный датасет на предмет наиболее встречающихся значений у заёмщиков по характеристикам профессия и возраст.

На рисунке ниже приведена статистика заёмщиков по профессиям:

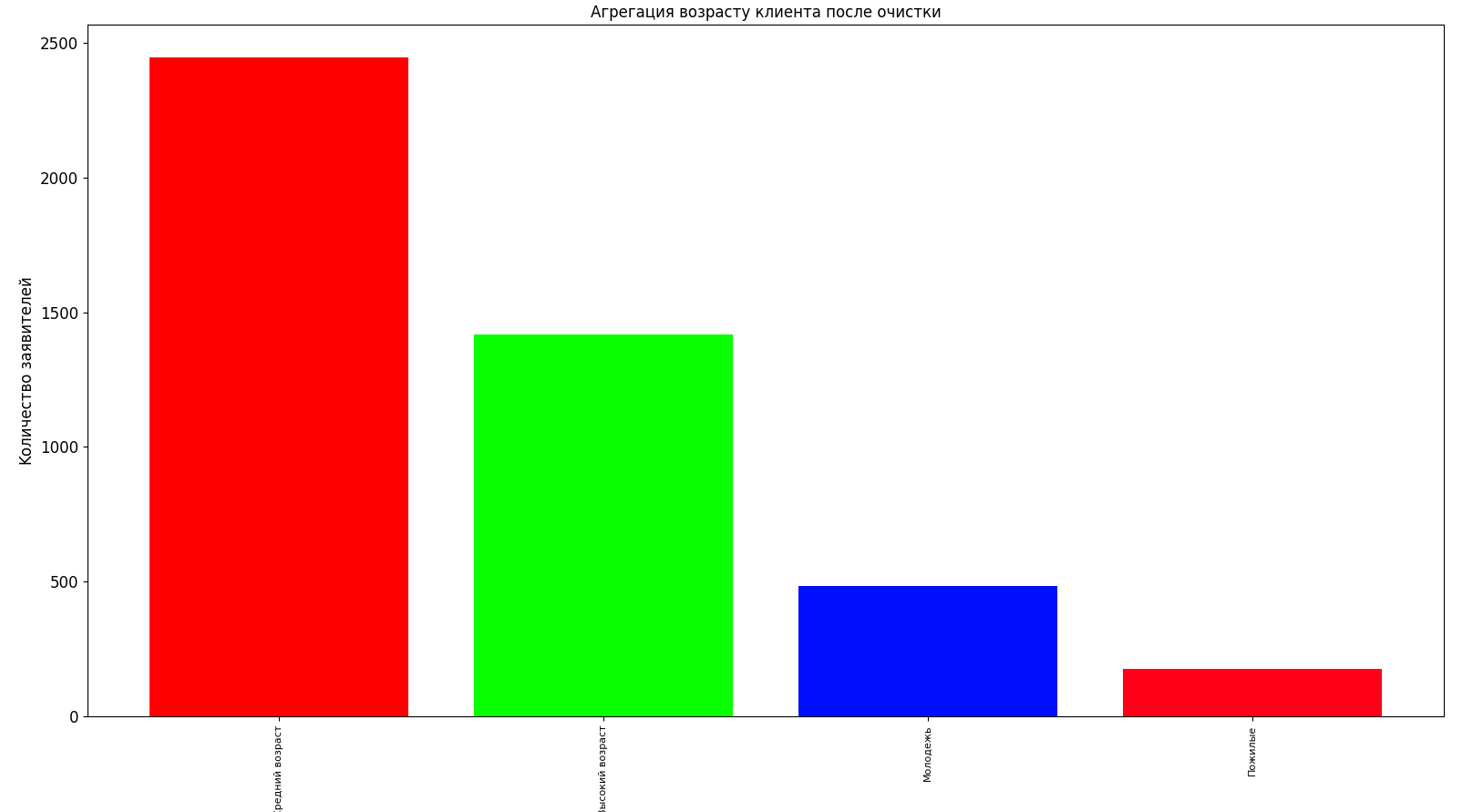


На рисунке ниже представлена статистика заёмщиков по профессиям, где наиболее редко встречающиеся профессии объединены в группу «другое»:



На рисунке ниже представлена статистика разбиения заёмщиков на группы по возрастам, где возраст человека принадлежит конкретной возрастной группе:

1. Группа «Молодежь» - возраст от 18 до 29 лет
2. Группа «Средний возраст» - возраст от 30 до 44 лет
3. Группа «Высокий возраст» - возраст от 45 до 59 лет
4. Группа «Пожилые» - возраст от 60 лет и выше.



Вывод

По результатам проведенного анализа можно сделать вывод, что больше всего заёмщиков по кредитам относятся к средней возрастной группе. Так же можно сделать вывод, что большинство заёмщиков работают в сфере управления.