ЛАБОРАТОРНАЯ РАБОТА 2. ВИЗУАЛИЗАЦИЯ ДАННЫХ

**Исследование набора данных**

**Цели и задачи**

Цель лабораторной работы: изучение программных средств для визуализации наборов данных.

Основные задачи:

* установка и настройка matplotlib, seaborn;
* изучение основных типов графиков библиотеки matplotlib;
* изучение основных типов графиков библиотеки seaborn;
* получение навыков анализа данных по визуальным представлениям данных.

**Индивидуальное задание**

**Описание набора данных**

Данный набор данных является коллекцией пользовательских данных из банка. Он содержит информацию о отдельных пользователях, и их персональными данными.

Цель набора данных:

Основной целью этого набора данных является предсказание, покинет ли клиент банк в ближайшее время. Это может помочь банку разработать стратегии удержания клиентов.

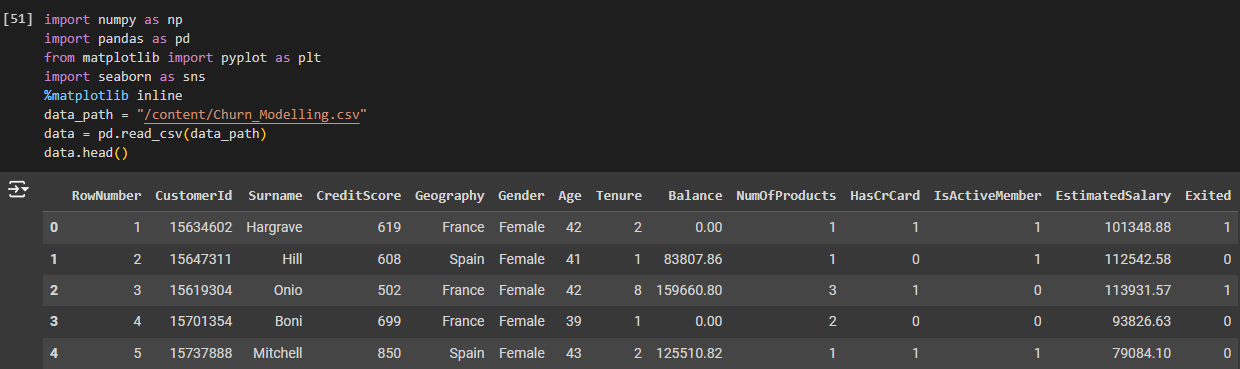
Описание каждого признака и его тип:

* RowNumber (Тип: Целое число): Номер строки в наборе данных.
* CustomerId (Тип: Целое число): Уникальный идентификатор клиента.
* Surname (Тип: Категориальный): Фамилия клиента.
* CreditScore (Тип: Целое число): Кредитный рейтинг клиента, который отражает его кредитоспособность.
* Geography (Тип: Категориальный): Географическое положение клиента (Франция, Испания, Германия).
* Gender (Тип: Категориальный): Пол клиента (Мужской, Женский).
* Age (Тип: Целое число): Возраст клиента.
* Tenure (Тип: Целое число): Срок пребывания клиента в банке в годах.
* Balance (Тип: Вещественное число): Баланс счета клиента.
* NumOfProducts (Тип: Целое число): Количество продуктов (услуг), которыми пользуется клиент.
* HasCrCard (Тип: Булево): Имеет ли клиент кредитную карту этого банка (1 = Да, 0 = Нет).
* IsActiveMember (Тип: Булево): Является ли клиент активным членом банка (1 = Да, 0 = Нет).
* EstimatedSalary (Тип: Вещественное число): Оценка годового заработка клиента.
* Exited (Тип: Булево): Покинул ли клиент банк (1 = Да, 0 = Нет). Это целевая переменная.

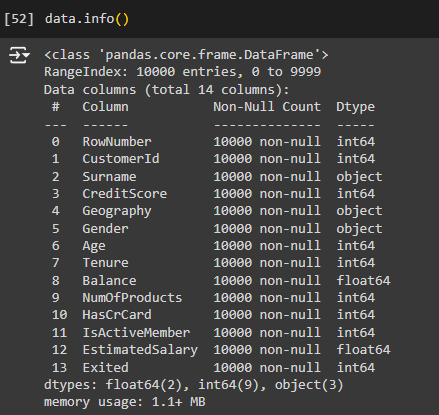
**Форма набора данных**

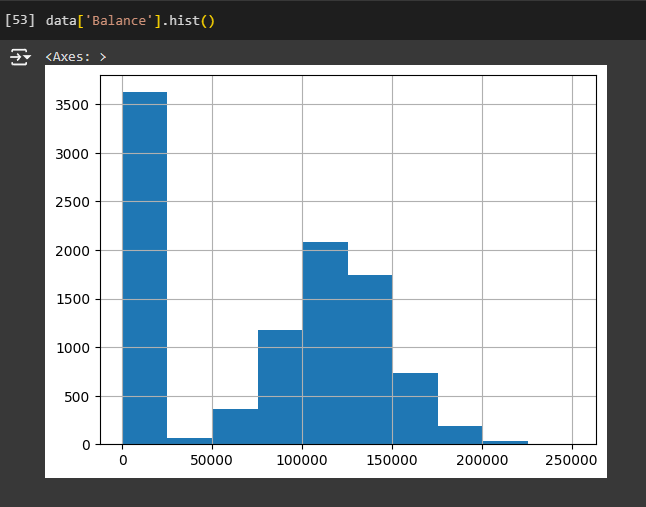
Характеристики набора данных:

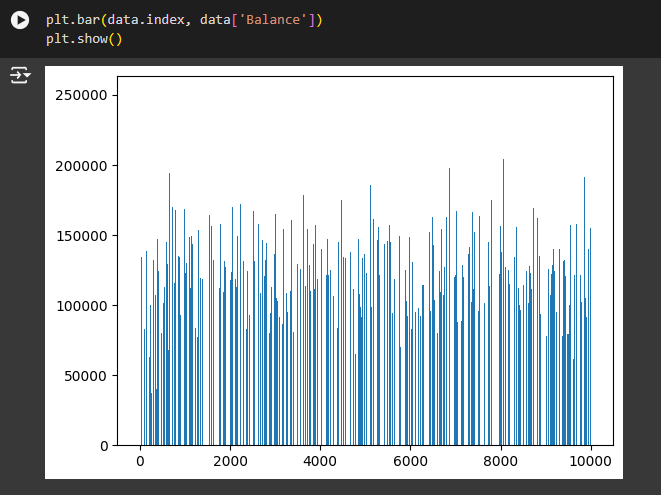
* Количество элементов (строк): 10000
* Количество признаков (столбцов): 14
* Типы данных: Float (2 шт.), int(9 шт.), object(3 шт.)

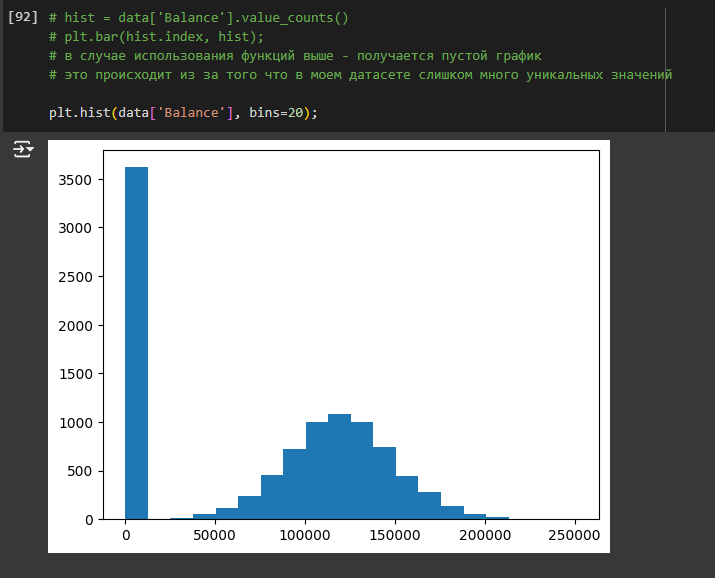


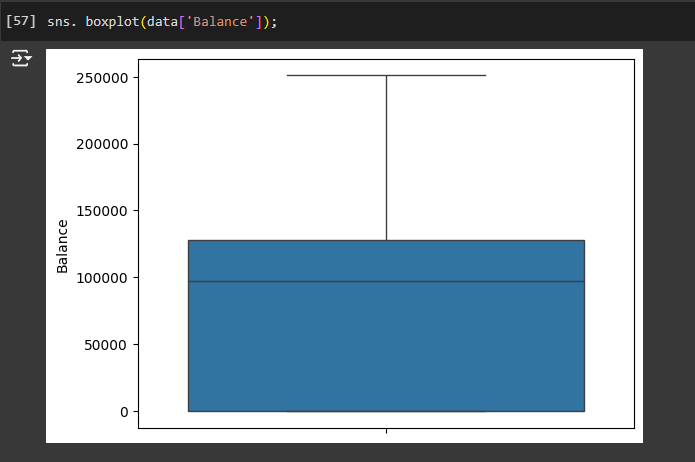
Детальные характеристики признаков:



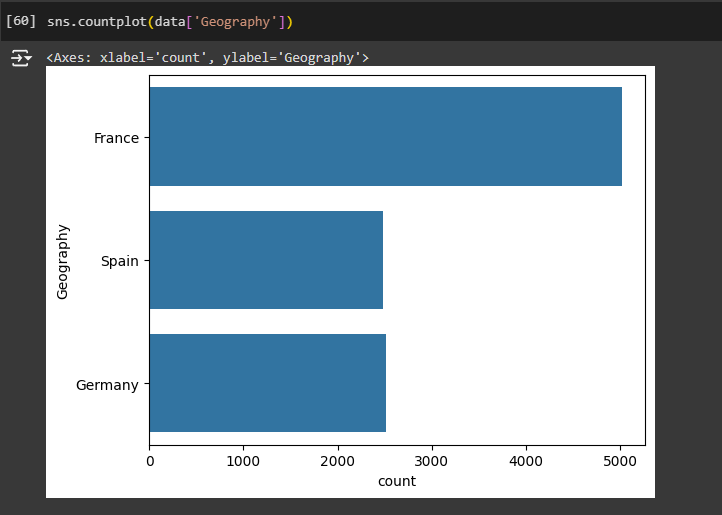


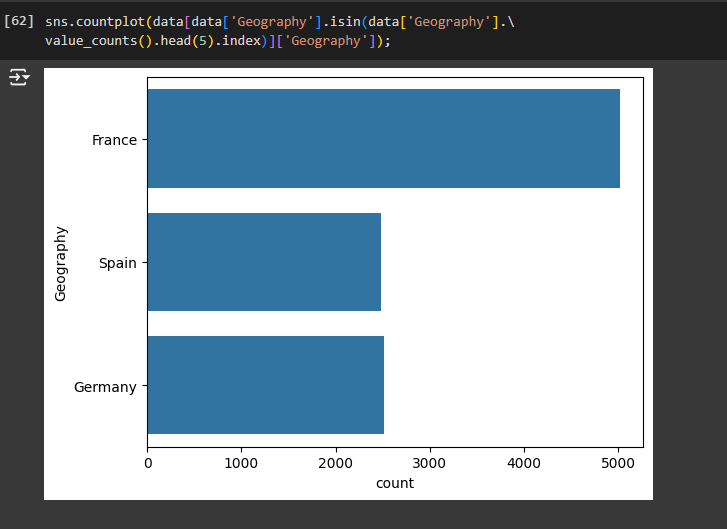


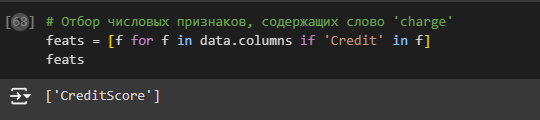


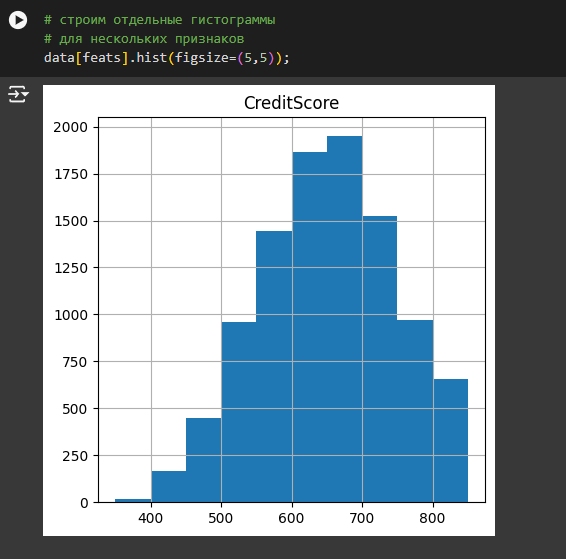


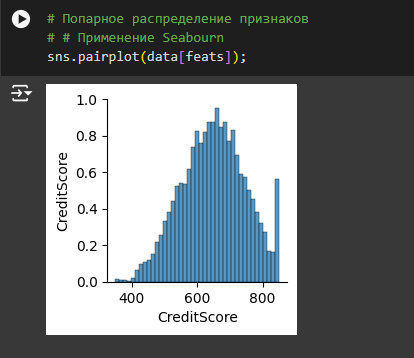


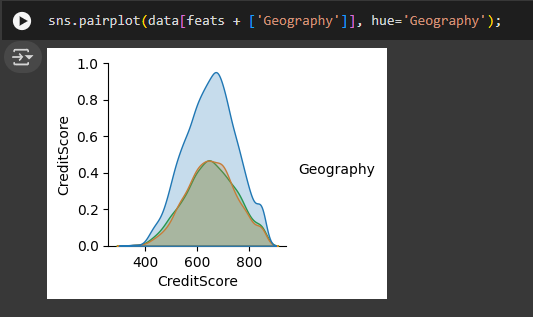


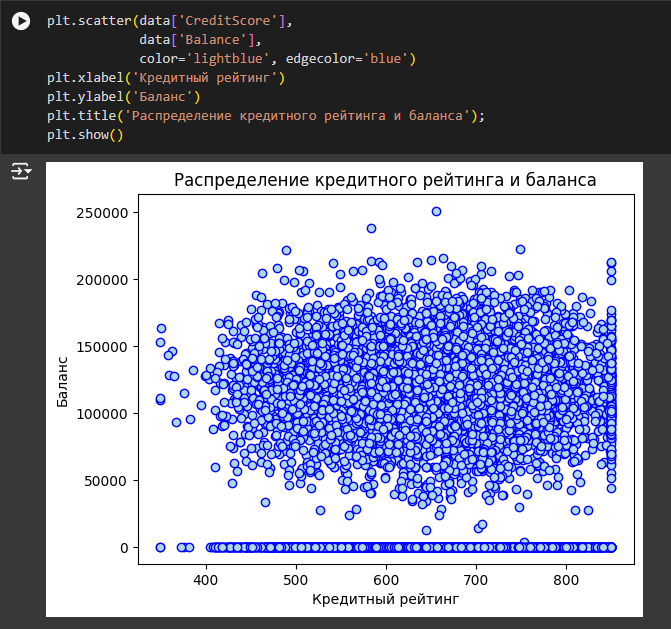


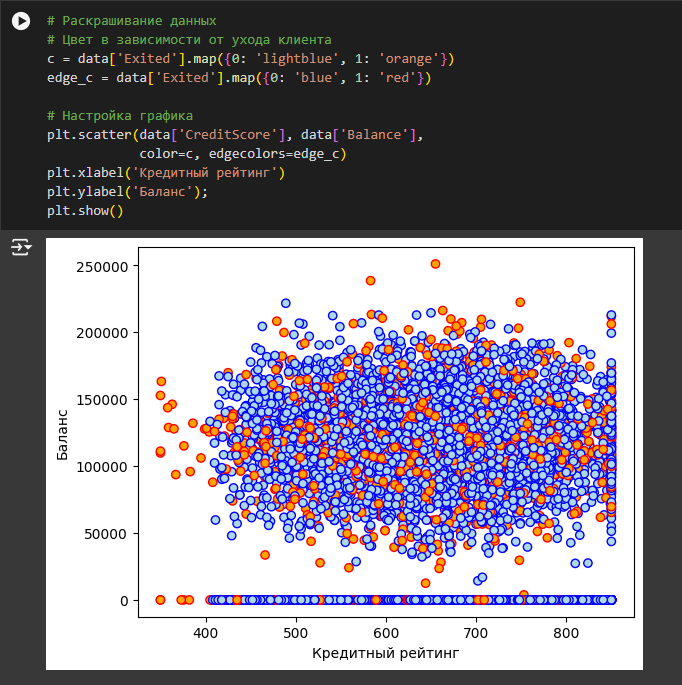


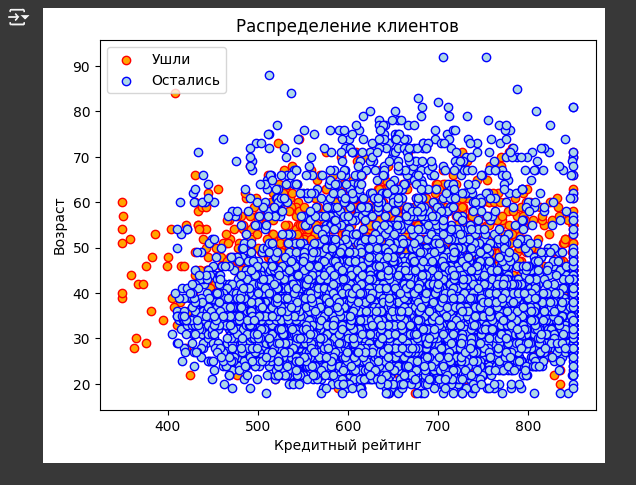


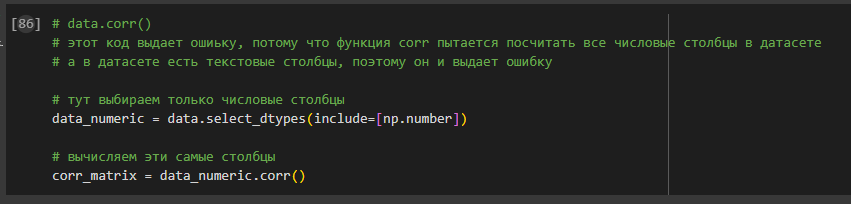


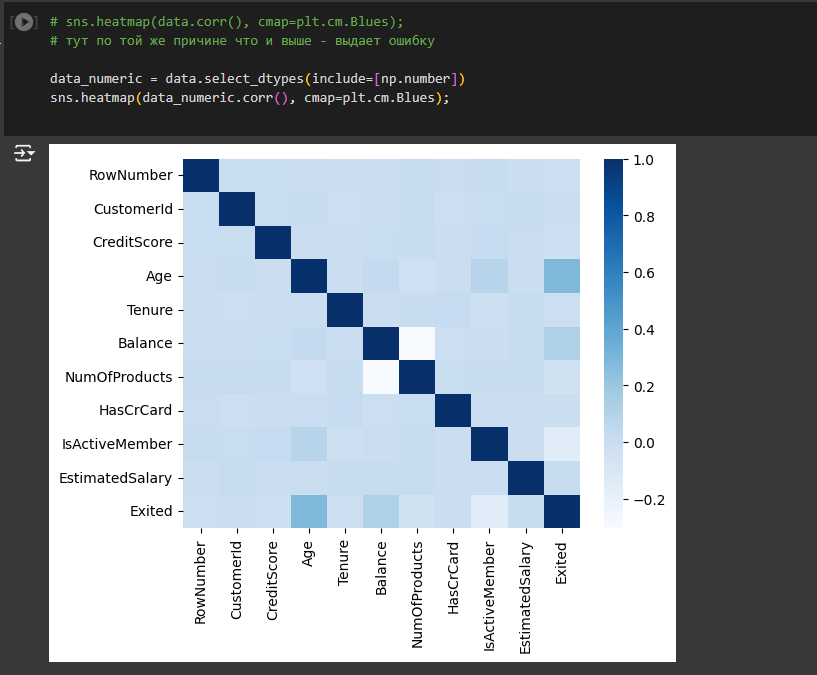










**Контрольные вопросы**

1. Какие инструментальные средства используются для организации рабочего места специалиста Data Science?
2. Какие библиотеки Python используются для работы в области машинного обучения? Дайте краткую характеристику каждой библиотеке.
3. Почему при реализации систем машинного обучения широкое распространение получили библиотеки Python?