ЛАБОРАТОРНАЯ РАБОТА 1. ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

**Исследование набора данных**

**Цели и задачи**

Цель лабораторной работы: изучение программных средств для организации рабочего места специалиста по интеллектуальному анализу данных.

Основные задачи:

* установка и настройка среды разработки Python;
* изучение принципов загрузки и очистки данных;
* получение навыков по предварительной обработке данных на языке

Python;

* изучение основных библиотек Python для работы с данными.

**Индивидуальное задание**

**Описание набора данных**

Данный набор данных является коллекцией пользовательских данных из социальных сетей. Он содержит информацию о отдельных пользователях, их паттернах использования и характеристиках.

Цель набора данных:

Целью этого набора данных является анализ поведения пользователей в социальных сетях, понимание их поведения, использования и выявление трендов. Этот набор данных может быть использован для построения моделей, предсказывающих вовлеченность пользователей, идентификации пользователей в сети и оптимизации маркетинга в социальных сетях.

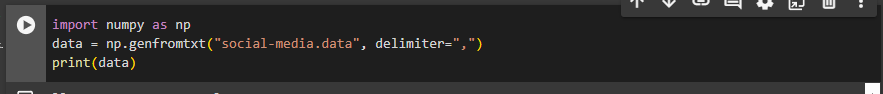
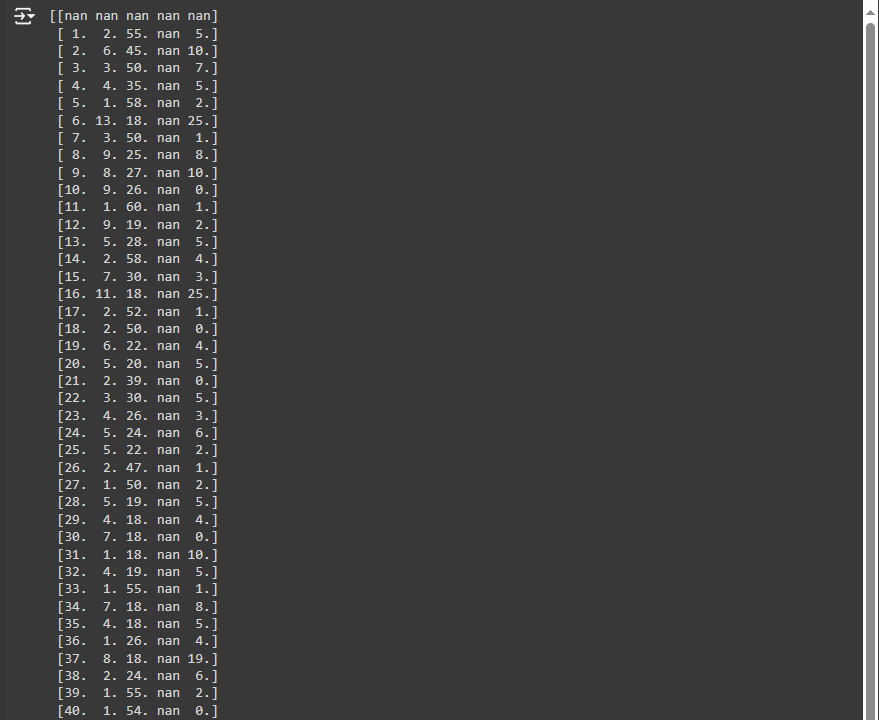
Описание каждого признака и его тип:

* UserId: Уникальный идентификатор для каждого пользователя в наборе данных (целое число)
* UsageDuration: Общее время, проведенное пользователем в социальных сетях, в часах (число)
* Age: Возраст пользователя в годах (число)
* Country: Страна проживания пользователя (категориальная переменная)
* TotalLikes: Общее количество лайков, поставленных пользователем за день (число)

**Форма набора данных**

Характеристики набора данных:

* Количество элементов (строк): 62
* Количество признаков (столбцов): 5
* Количество пропущенных значений: 62 (все значения в 4-м столбце равны NaN)

Сводные статистики:

1. Признак 1 (ID пользователя):

* Минимальное значение: 1
* Максимальное значение: 62

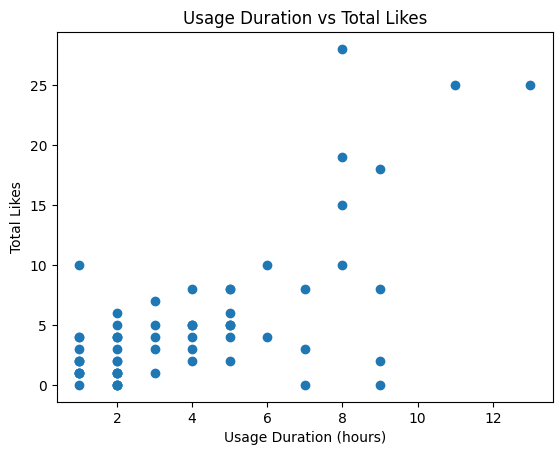
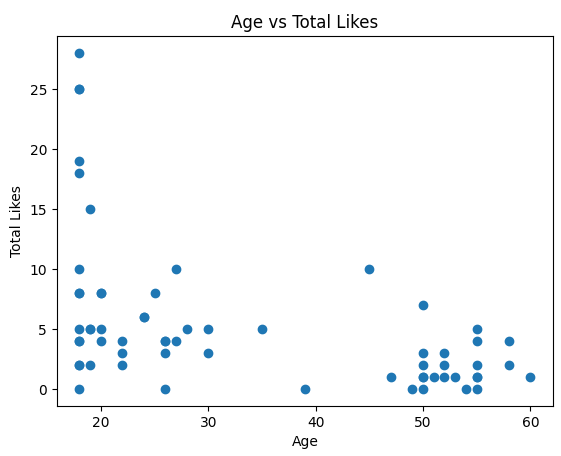
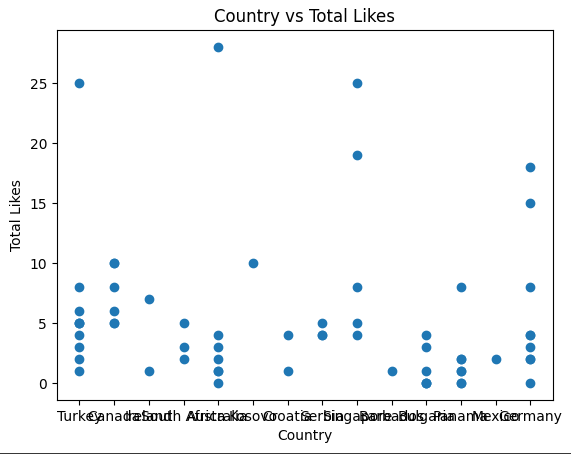
1. Признак 2 (Длительность использования):

* Среднее значение: 4,33
* Минимальное значение: 1
* Максимальное значение: 13

1. Признак 3 (Возраст):

* Среднее значение: 38
* Минимальное значение: 18
* Максимальное значение: 60

1. Признак 4 (Страна): (текстовое поле – поэтому расчеты по ней сделать невозможно)

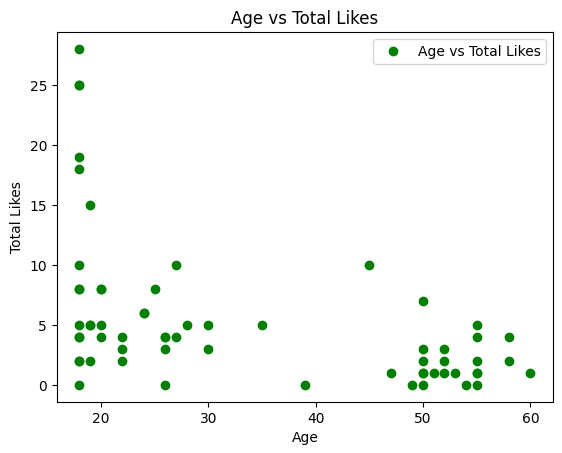
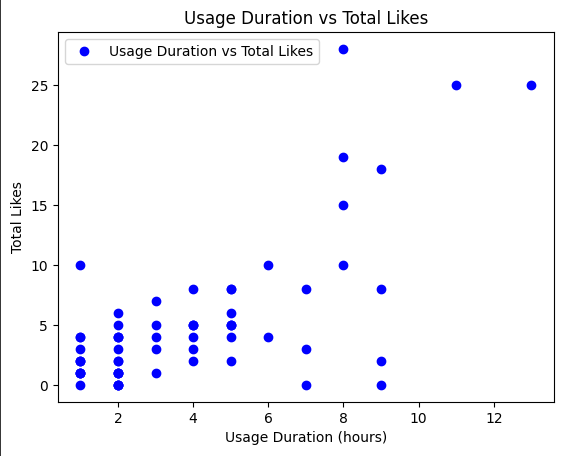
  

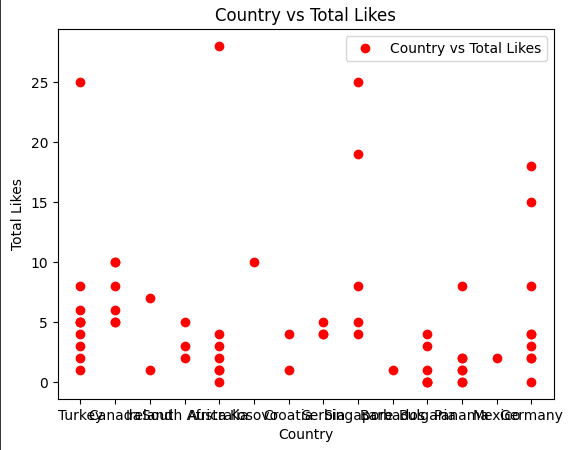
Предварительный анализ:

Основываясь на сводных статистиках, мы можем сделать некоторые наблюдения:

* Признак Длительность использования имеет смещенную распределение, с большинством значений между 1 и 5, что может указывать на то, что большинство пользователей используют систему в течение короткого периода времени.
* Признак Возраст имеет относительно нормальное распределение, с большинством значений между 20 и 50, что может указывать на то, что большинство пользователей находятся в возрасте от 20 до 50 лет.

**Графические представления**

****



**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?

**Экранные формы (консольный вывод)**

