Санкт-Петербургский государственный университет

Факультет прикладной математики – процессов управления

Лабораторная работа №1

**Отчет**

по дисциплине «Алгоритмы и структуры данных»

**Генерация синтетических данных**

Вариант 3

Автор работы: Дацык Р.В.

Группа: 22.Б15-пу

Преподаватель: Дик А.Г.

Санкт-Петербург, 2023

**Оглавление**

[**1.** **Цель работы** 3](#_Toc148043577)

[**2.** **Задачи** 3](#_Toc148043578)

[**3.** **Введение** 3](#_Toc148043579)

[**4.** **Теоретическая часть** 3](#_Toc148043580)

[**5.** **Алгоритм метода** 3](#_Toc148043581)

[**6.** **Опsисание программы** 5](#_Toc148043582)

[**7.** **Рекомендации пользователю** 5](#_Toc148043583)

[**8.** **Рекомендации программисту** 6](#_Toc148043584)

[**9.** **Контрольный пример** 6](#_Toc148043585)

[**10.** **Вывод** 6](#_Toc148043586)

[**11.** **Список использованной литературы** 7](#_Toc148043587)

# **Цель работы**

Разработать программу, реализующую генерацию синтетических данных о просмотрах и количестве сезонной рекламы на различных платформах.

# **Задачи**

1. Изучить методы генерации данных.
2. Проанализировать их достоинства и недостатки.
3. Выбрать наилучший метод генерации.
4. Реализовать выбранный алгоритм.
5. Протестировать программу.
6. Провести контрольный эксперимент

# **Введение**

В современном мире анализ данных и машинное обучение играют ключевую роль в принятии информированных решений в различных областях. Для успешного проведения исследований и разработки алгоритмов необходимы качественные и разнообразные датасеты. В данном отчете представлен проект, целью которого является создание программы для генерации синтетического датасета с данными о просмотрах рекламы.

# **Теоретическая часть**

Генерация персональных данных также имеет важные особенности. Во-первых, в зависимости от определенного сезона, должна показываться ориентированная реклама, значит для каждого сезона должен быть свой словарь с видами рекламы. Во-вторых, время просмотра высчитывается по определенной формуле, которую предоставляет заказчик.

# **Алгоритм метода**

1. Запуск последовательной генерации данных
2. Генерация уникального почтового адреса
3. Генерация уникального ip-адреса
4. Случайный выбор платформы
5. Генерация даты просмотра
6. Вычисление формулы времени просмотра рекламы
7. Все полученные данные записываются в таблицу.
8. Созданная таблица сохраняется в Excel-файл.

На рисунке 4.1 представлена блок-схема алгоритма.

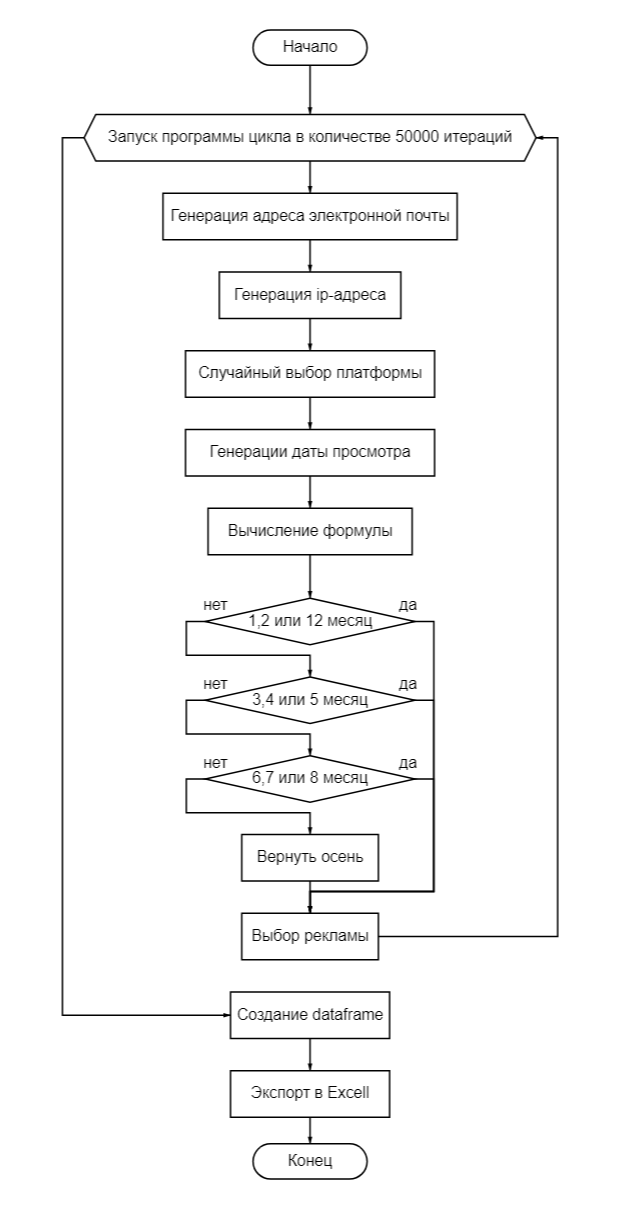


Рисунок 4*.*1 Блок-схема программы.

# **Описание программы**

Алгоритм реализован на языке python 3.10 с использованием следующих пакетов: pandas, random, datetime.

В программе используются 7 функций: 6 связанных с генерацией данных и одна для обработки интерфейса. В таблице 5.1 представлено описание всех функций.

Таблица 5.1. Описание функций

|  |  |  |  |
| --- | --- | --- | --- |
| Имя функции | Аргумент | Возвращаемое значение | Условия формирования данных |
| generate\_ip | None | От 8 до 12 цифр, составляющих ip-адрес | Цифры генерируются в формате «192.104.XXX.XXX». Выбираются случайным образом. |
| generate\_date | None | Дата просмотра | Дата просмотра рекламы, в зависимости от выбранного сезона |

# 

# **Рекомендации пользователю**

Для успешного запуска программы необходимо устройство с операционной системой Linux, macOS или Windows, а также среда разработки, поддерживающей запуск python 3.10

Перед запуском программы следует выбрать количество рекламы и коэффициент в зависимости от требований заказчика.

Для запуска генерации достаточно запустить файл с кодом. Как только таблица будет создана, появится уведомление об успешном завершении. Файл с данными будет сохранен в директорию, где находится код программы.

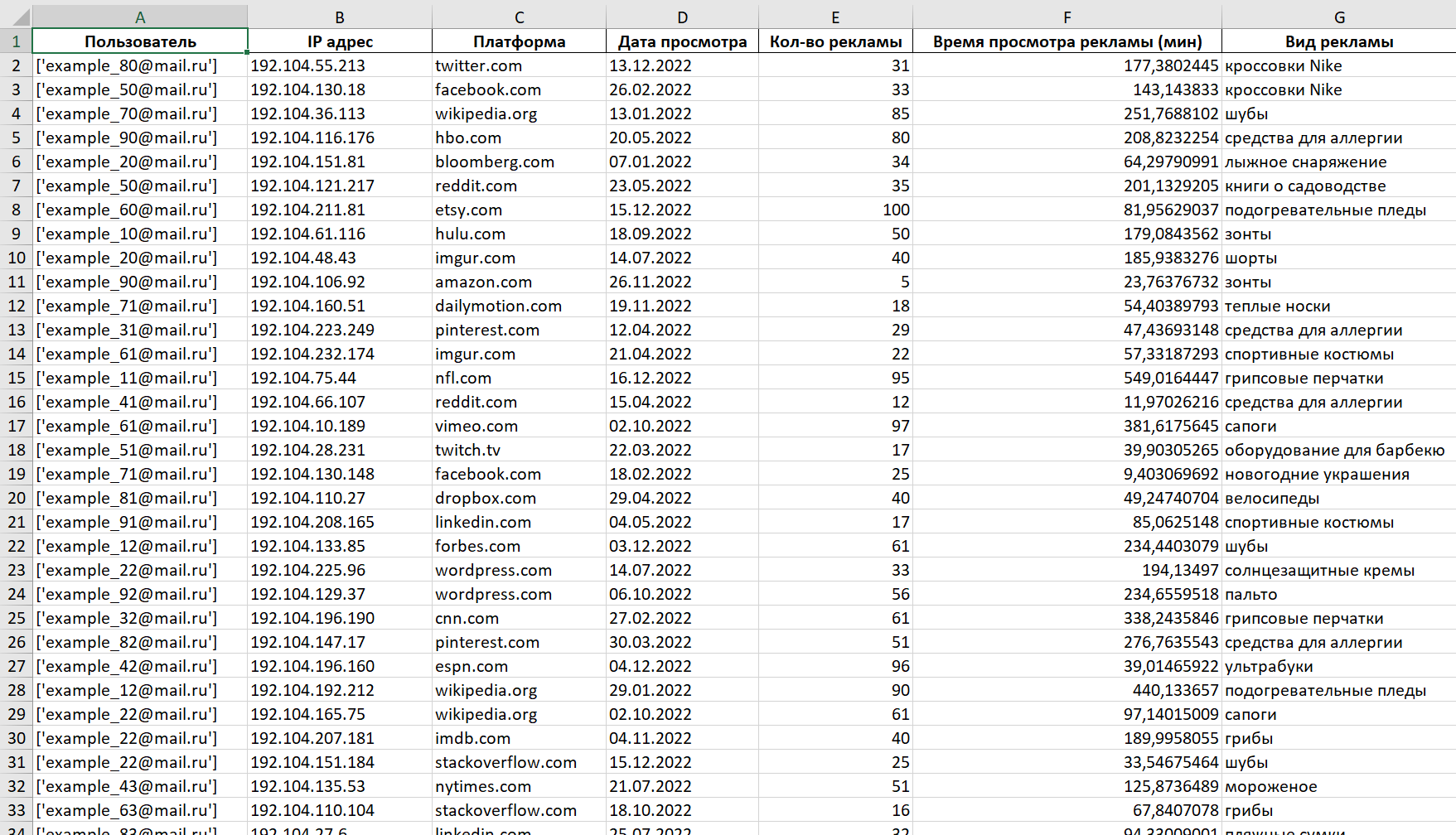
# **Рекомендации программисту**

Для запуска программы необходима 64-битная операционная система Windows, Linux или macOS. Для работы с кодом необходима среда разработки, совместимая с python 3.10 и библиотеки pandas, random, datetime.

Исходный код программы доступны по ссылке:

<https://github.com/CapTopGrade/Algorithms/blob/main/Alg1.py>

# **Контрольный пример**

Пример полученной таблицы изображен на следующем рисунке. Рисунок 8.1 Пример полученной таблицы

# **Вывод**

В результате выполнения данной работы разработан алгоритм формирования датасета с синтетическими данными о просмотрах и количестве сезонной рекламы на различных платформах.

# **Список использованной литературы**

[1] Документация библиотеки Pandas: https://pandas.pydata.org/docs/

[2] Документация библиотеки Random: https://docs.python.org/3/library/random.html