

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ – ПРОЦЕССОВ
УПРАВЛЕНИЯ

ОТЧЕТ
по лабораторной работе №2
по дисциплине «Алгоритмы и структуры данных»
на тему «Генерация данных»
Вариант 3 (просмотр рекламы)

Выполнила
студентка 2 курса
группы 21-Б15.ПУ
Павлова Ксения Андреевна

Преподаватель
Дик Александр Геннадьевич

Санкт-Петербург
2022

СОДЕРЖАНИЕ

Цель.....	3
Задачи.....	3
Блок-схемы общего алгоритма и его функций	5
Примечания к блок-схемам	6
Рекомендации программиста.....	7
Контрольный пример	8
Вывод	9
Список литературы	10

Цель: написать программу генерации набора данных с заданными условиями.

Задачи:

А. Сгенерировать датасет в котором будут следующие наборы свойств:

1. Пользователь - example@mail.ru
2. IP адрес - 10.10.10.12
3. Платформы - youtube.com
4. Дата просмотра - 12.12.2022
5. Кол-во рекламы - 2 раза
6. Время просмотра рекламы - 1:03 минут
7. Вид рекламы - кроссовки Nike

В. Дополнительная информация по каждому свойству:

1. Пользователь - свободный вариант генерации данных.
2. IP адрес - только глобальный IP.
3. Платформы – «Словарь» по социальной сети и по видеохостингам
4. Дата просмотра - определенный вариант генерации данных.
5. Кол-во рекламы - определенный вариант генерации данных.
6. Время просмотра рекламы - зависит от количества рекламы и высчитывается по определенной формуле, где возможно регулировать среднее значения продолжительности рекламы.
7. Вид рекламы – «Словарь» по которому будет генерироваться разнообразия товаров по категориям. Учитывать сезонные товаров по дате

С. Ограничения датасета:

1. Всего строк в датасете – минимум 50 000.
2. Пользователь – по email.
3. IP адрес – только глобальный IP.
4. Платформы – «Словарь» должен состоять минимум из 50 платформ.
5. Дата просмотра – должен удовлетворять сезону, который заранее обсуждается.
6. Кол-во рекламы – Имеет случайный диапазон от 1 до 100 раз.
7. Время просмотра рекламы = кол-во рекламы*Коэффициент.
Коэффициент имеет, или случайную генерацию от 30 секунд до 120 секунд, или определенную константу равную величине, которую попросит заказчик.
8. Вид рекламы – «Словарь» должен состоять минимум из 50 видов, которые учитывают сезон, диапазон которого указан в дате.

Блок-схемы общего алгоритма и его функций.

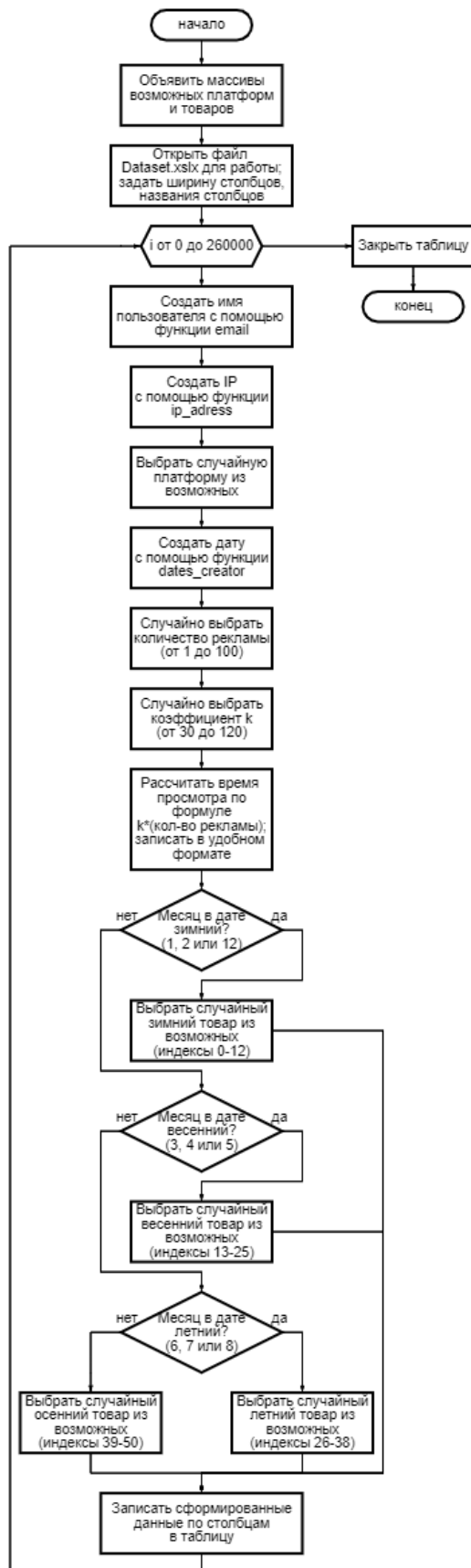


Рис. 1. Блок-схема общего алгоритма

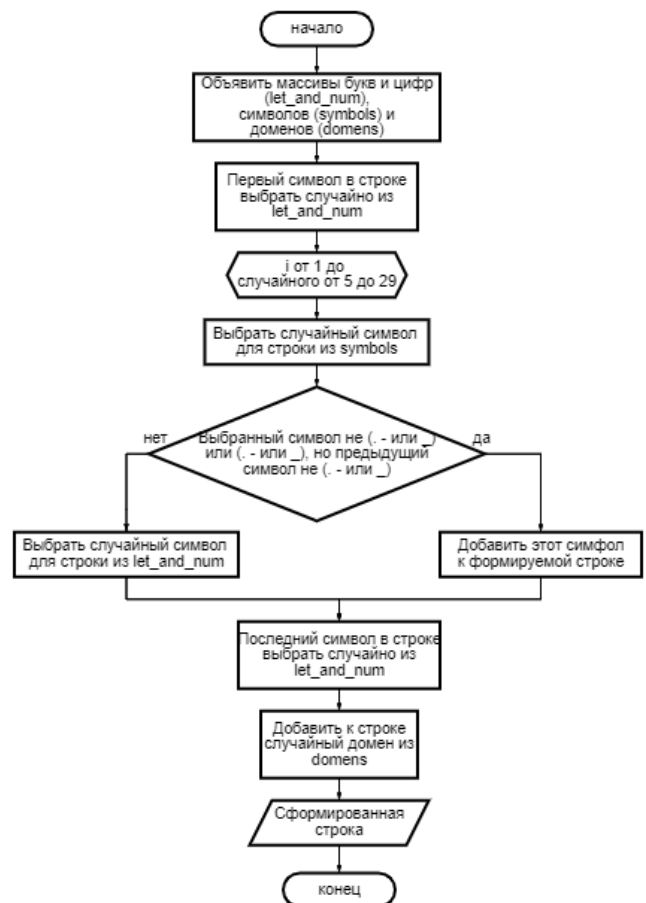


Рис. 2. Блок-схема функции email

Примечания к блок-схемам.

Общий алгоритм (рис. 1)

Работа с файлами формата `xlsx` происходит с помощью библиотеки `xlsxwriter`.

Случайный выбор производится с помощью библиотеки `random`.

Функция `ip_adress` формирует строку в формате `XXX.XXX.XXX.XXX`, где `XXX` – число от 1 до 255 (без незначащих нулей).

Случайная дата формируется с помощью библиотеки `time`.

Блок-схема функции формирования имени пользователя представлена на рис.2. Имя пользователя формируется по следующим правилам:

Имя почтового ящика (логин) — это уникальная комбинация букв, цифр и специальных символов, расположенная перед доменом:

`mail.ru`, `list.ru`, `bk.ru`, `internet.ru` или `inbox.ru`.

Имя состоит из 5-31 символов. В имени можно использовать:

латинские буквы: `a-z`, `A-Z`;

цифры: `0-9`;

точку, подчеркивание или дефис: `._-`. Эти символы не могут идти друг за другом или стоять в начале или конце имени.

Рекомендации программиста.

Для работы с кодом необходимо приложение PyCharm и установленные библиотеки random, time, xlswriter.

Контрольный пример.

В данном разделе представлен пример работы данной программы.

На рис. 3 представлена часть таблицы - результат работы программы.

Пользователь	IP адрес	Платформа	Дата просмотра	Кол-во рекламы	Время просмотра рекламы	Вид рекламы
Y5vMMUjR_QEFq@bk.ru	147.67.88.185	last.fm	14/02/2021	11 раз	19:4 минут	Коньки Jackson
B0KCAQcgnX0qaU1xA2078zEbg@list.ru	103.68.183.78	videos.kaltura.com	02/04/2022	26 раз	46:22 минут	Костюм Fila
l8h1789a_iBAk86Zd@inbox.ru	173.111.31.189	snapchat.com	10/08/2021	53 раз	75:5 минут	Купальник ПАЧЕ
b83FQZvT0wqG5wT.Z@list.ru	154.77.238.137	twitter.com	26/01/2021	4 раз	7:40 минут	Коньки Wifa
813cyT@bk.ru	212.94.232.148	likee.video	25/10/2021	21 раз	23:6 минут	Плед Hermes
kn0grnNDRwe@bk.ru	179.254.107.126	panopto.com	01/10/2021	4 раз	7:4 минут	Плед Hermes
AbhWNvQSSa7bEC6K@internet.ru	107.174.174.250	teamo.ru	09/04/2021	39 раз	33:9 минут	Ветровка Trespass
596HaDwIXKyMTHgQ319PLF@bk.ru	212.82.52.250	likee.video	14/02/2021	43 раз	83:51 минут	Коньки Jackson
q2YnVsdOWJ3Xnzd@list.ru	110.22.46.129	liveinternet.ru	18/05/2022	99 раз	133:39 минут	Костюм Reebok
OKwT2t@mail.ru	85.124.60.74	instagram.com	03/05/2021	26 раз	43:46 минут	Костюм Under armour
oI7m1QrWtaWb@internet.ru	75.200.193.137	tiktok.com	15/12/2021	46 раз	41:24 минут	Угги Teva
TQ_TUz8BSI.SBVC25B.SW@internet.ru	239.41.94.179	subscribe.ru	04/05/2022	39 раз	59:48 минут	Костюм New balance
3yhVh3YP_JWA@mail.ru	234.206.35.195	twitich.tv	22/03/2021	92 раз	65:56 минут	Ветровка Trespass
L7xAdY6Ob2G@bk.ru	248.3.159.52	badoo.com	14/11/2021	65 раз	61:45 минут	Зонт Fulton
NWleYHN6R7t@inbox.ru	119.74.104.171	dacast.com	09/06/2021	84 раз	135:48 минут	Очки Police
T5rvoGnPaFW-QVc7VKcY2cWSQYF@internet.ru	125.113.44.155	jwplayer.ru	18/04/2022	37 раз	70:18 минут	Ветровка Helly Hansen
E32zaEZq7A@list.ru	153.77.56.111	videos.kaltura.com	25/02/2021	91 раз	165:19 минут	Пуховик Marmot
exwVx-dnA2w_F_Z@bk.ru	17.104.34.74	scipeople.ru	04/11/2021	47 раз	33:41 минут	Зонт Blunt
Dje8Uw74hbhq5GnLUUJH@mail.ru	233.247.90.138	blogger.com	22/02/2022	15 раз	28:30 минут	Валени Russy Valenki
e5WysyT@list.ru	52.137.137.247	blogger.com	24/04/2022	42 раз	46:12 минут	Ветровка Under Armour
4t1muwEt1rWZ1Yhlyy@bk.ru	128.73.121.229	vidyard.com	27/04/2022	34 раз	42:30 минут	Костюм Fila
JE6nJPY7dPq75sxHn2o8@inbox.ru	144.97.150.132	wechat.com	23/05/2021	17 раз	18:59 минут	Костюм Under armour
D5saFOVbfr4qRt06UeXlek@list.ru	122.90.95.208	twitter.com	22/09/2021	86 раз	67:22 минут	Плед Кристин Диор
6c7Rb-7VCK-jp00e@internet.ru	148.91.49.21	rutube.ru	29/04/2022	29 раз	45:26 минут	Костюм Putin team
BehbOJXmPFADMKu@list.ru	46.40.216.210	nportal.ru	11/04/2022	86 раз	101:46 минут	Костюм Putin team
MCzITux@inbox.ru	198.244.169.31	agrobook.ru	26/01/2022	63 раз	123:54 минут	Коньки Wifa
LZa-rtDMVtVtKMShezjvl@mail.ru	19.121.32.3	youtube.com	03/01/2022	8 раз	11:4 минут	Пуховик BASK
9NaQoYU8By_PovGfOs.vWrdhVoIx9X@list.ru	121.110.154.143	facebook.com	19/04/2021	48 раз	93:36 минут	Костюм Under armour

Рис. 3. Вывод созданного датасета

Вывод: в ходе выполнения данной работы была написана программа генерации данных для датасета просмотра рекламы по заданным условиям.

Список литературы.

1. Introducing Python. Modern computing in simple packages / Bill Lobanovic.
2. <https://help.mail.ru/mail/account/signup/choose>
3. Ссылка на код и сформированную таблицу
<https://github.com/ksenkap/dataset>