|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ | | | | | | | | | | | | | | | | | | | | |
| Федеральное государственное автономное образовательное  учреждение высшего образования | | | | | | | | | | | | | | | | | | | | |
| **Дальневосточный федеральный университет** | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
| **ШКОЛА ЕСТЕСТВЕННЫХ НАУК** | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
| **Кафедра информационной безопасности** | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
| **О Т Ч Е Т** | | | | | | | | | | | | | | | | | | | | |
| о прохождении учебной практики (учебно-лабораторного практикума) | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | |  |  | Выполнил студент  гр. С8117-10.05.01ммзи | | | | | | | | |
|  |  |  | | | | | | Шульга М.И. | | |
|  | | | | | | | | | |  |  | (подпись) | | | | | |  | | |
|  | | | | | | | | | | | | | | | | | | | | |
| Отчет защищен с оценкой | | | | | | | | | |  |  | Руководитель практики | | | | | | |  | |
|  | | | | | | | | | |  |  | Старший преподаватель кафедры информационной безопасности ШЕН | | | | | | | | |
|  | | | |  | С.С. Зотов | | | | |  |  |  | | | | |  | С.С. Зотов | | |
| (подпись) | | | |  | (И.О. Фамилия) | | | | |  |  | (подпись) | | | | |  | (И.О. Фамилия) | | |
| « | 26 | » | июня | | | | | | 2021 г. |  |  |  | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
| Регистрационный № | | | | | | | |  | |  |  | Практика пройдена в срок | | | | | | | | |
| « | 26 | » | июня | | | | | | 2021 г. |  |  | с | « | 22 | » | февраля | | | | 2021 г. |
|  | | | | | | | | | |  |  | по | « | 26 | » | июня | | | | 2021 г. |
|  | | | | | |  | Е.В. Третьяк | | |  |  | на предприятии | | | | | | | | |
| (подпись) | | | | | |  | (И.О. Фамилия) | | |  |  |
|  | | | | | | | | | |  |  | Кафедра информационной | | | | | | | | |
|  | | | | | | | | | |  |  | безопасности ШЕН ДВФУ | | | | | | | | |
|  | | | | | | | | | |  |  |  | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | |
| г. Владивосток | | | | | | | | | | | | | | | | | | | | |
| 2021 | | | | | | | | | | | | | | | | | | | | |

# Оглавление

[Оглавлние 2](#_Toc75226631)

[Разработка программы для детектирования аномалий с помощью машинного обучения 5](#_Toc75226632)

[Изменения в парсере ивентов 5](#_Toc75226633)

[Обработка входных данных 6](#_Toc75226634)

[Кластеризация алгоритмом KMeans 9](#_Toc75226635)

[Python скрипт детектирующий аномалии в ивентах безопасности Windows 11](#_Toc75226636)

[Список литературы 13](#_Toc75226637)

**Задание на практику**

- Проведение исследования в области защиты от взлома программных средств.

- Написание отчета по практике о проделанной работе.

**Введение**

Учебная практика (учебно-лабораторный практикум) проходил на кафедре информационной безопасности ШЕН ДВФУ в период с 22 февраля 2021 года по 26 июня 2021 года.

Целью прохождения практики является приобретение практических и теоретических навыков по специальности, а также навыков оформления проведенного исследования в отчетной форме.

Задачи практики:

1. Ознакомиться с существующими подходами кластеризации.
2. Применить полученные знания в обнаружении аномальных ивентов безопасности Windows.
3. На основе полученных знаний написать отчет по практике о проделанной работе.

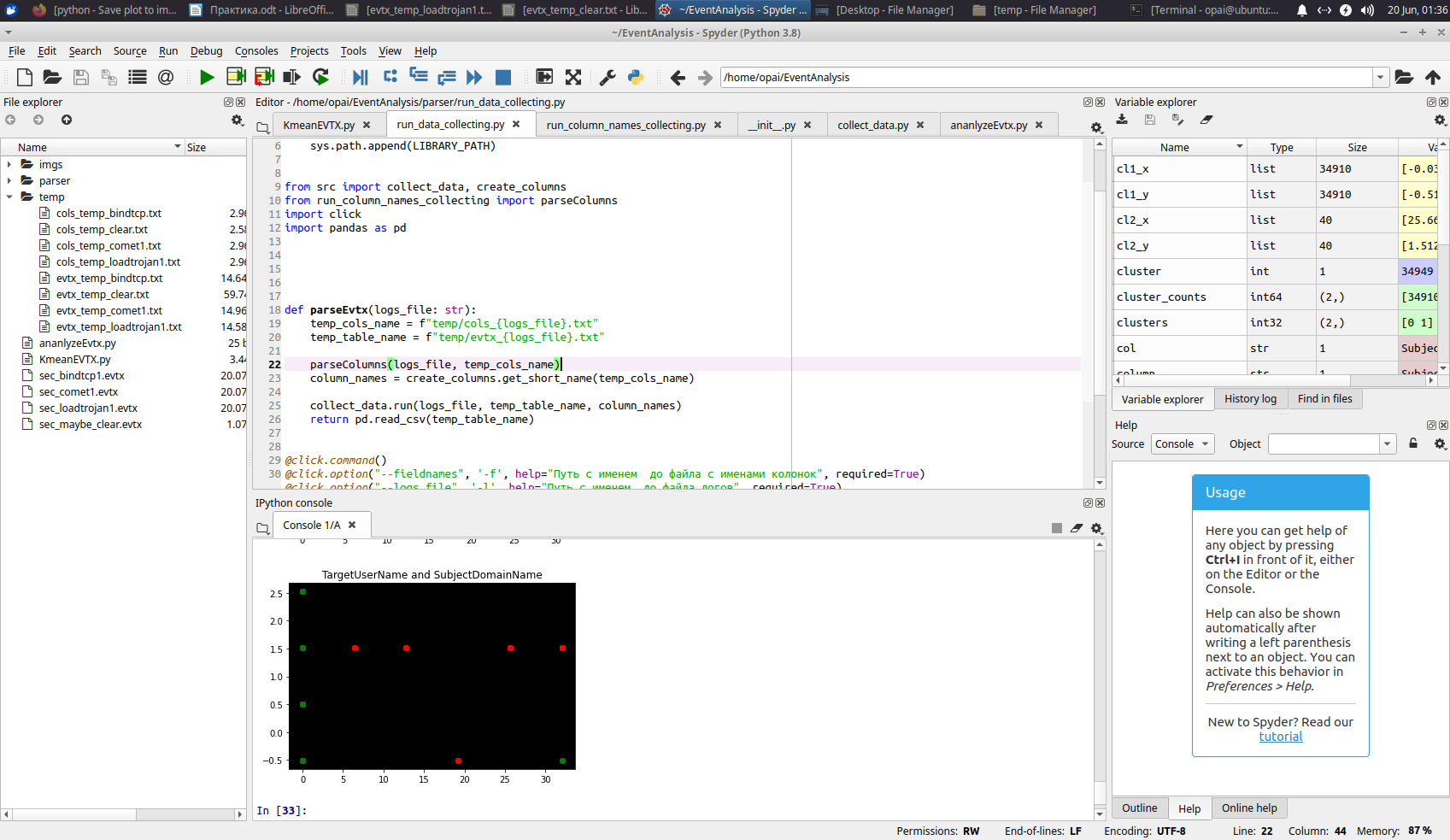
# Разработка программы для детектирования аномалий с помощью машинного обучения

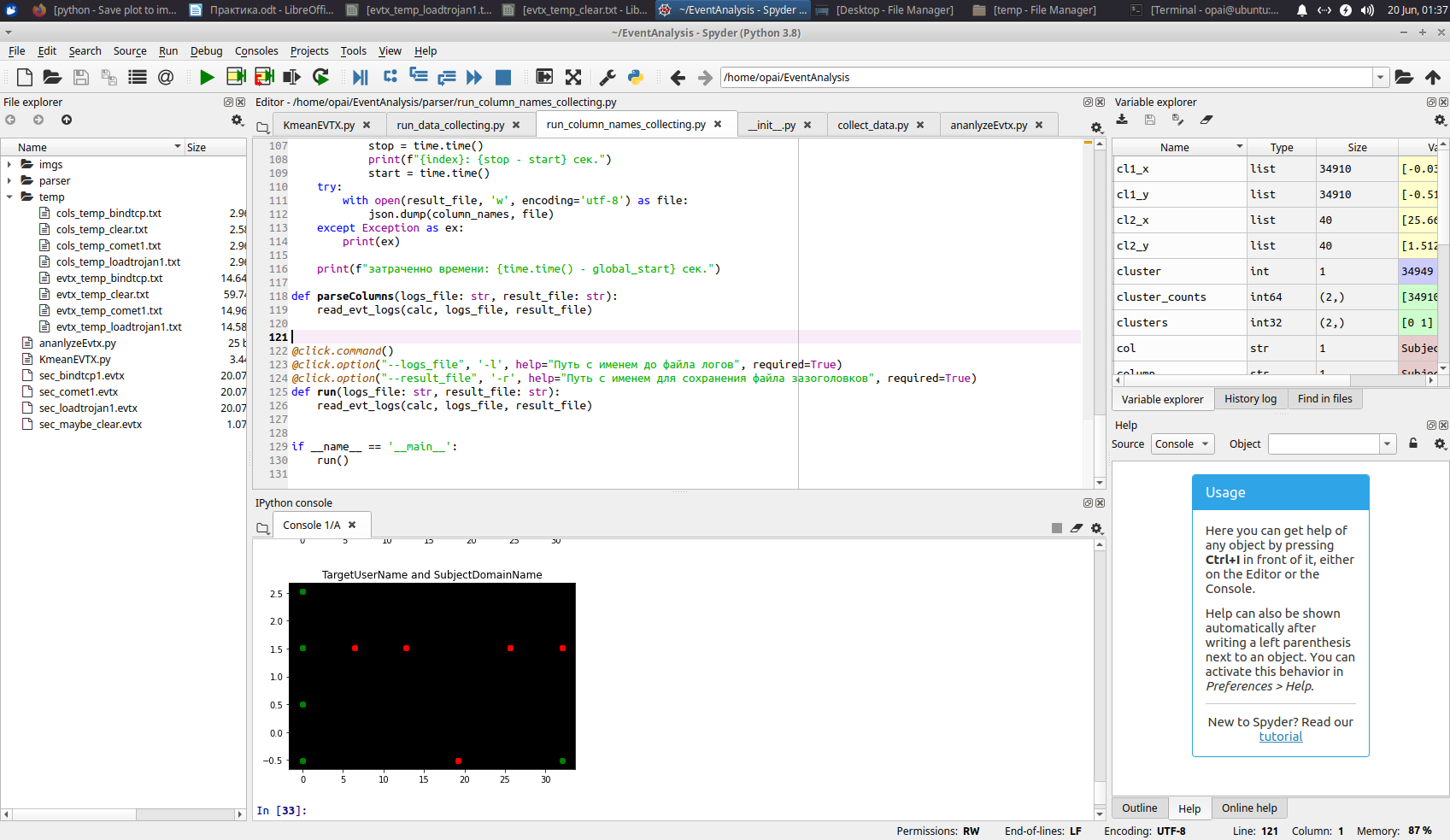
Для выполнения кластеризации были собраны данные с журналов событий Windows 7. Операционная система преднамеренно подвергалась атакам, для того чтобы собрать информацию о подозрительной активности. Для формирования подозрительной активности использовались: DarkComet, созданные с помощью msfvenom бекдоры и различные троянские программы

Далее будет описан процесс обработки полученных данных, их кластеризации, а также создании скрипта детектирования подозрительной активности в файлах событий Windows 7.

## Изменения в парсере ивентов

Для парсинга evtx логов использовался парсер winlog\_parser-master с небольшими правками. В парсер были добавлены функции, позволяющие вызывать парсер “на лету”, для того, чтобы упростить интерфейс полученного детектора.

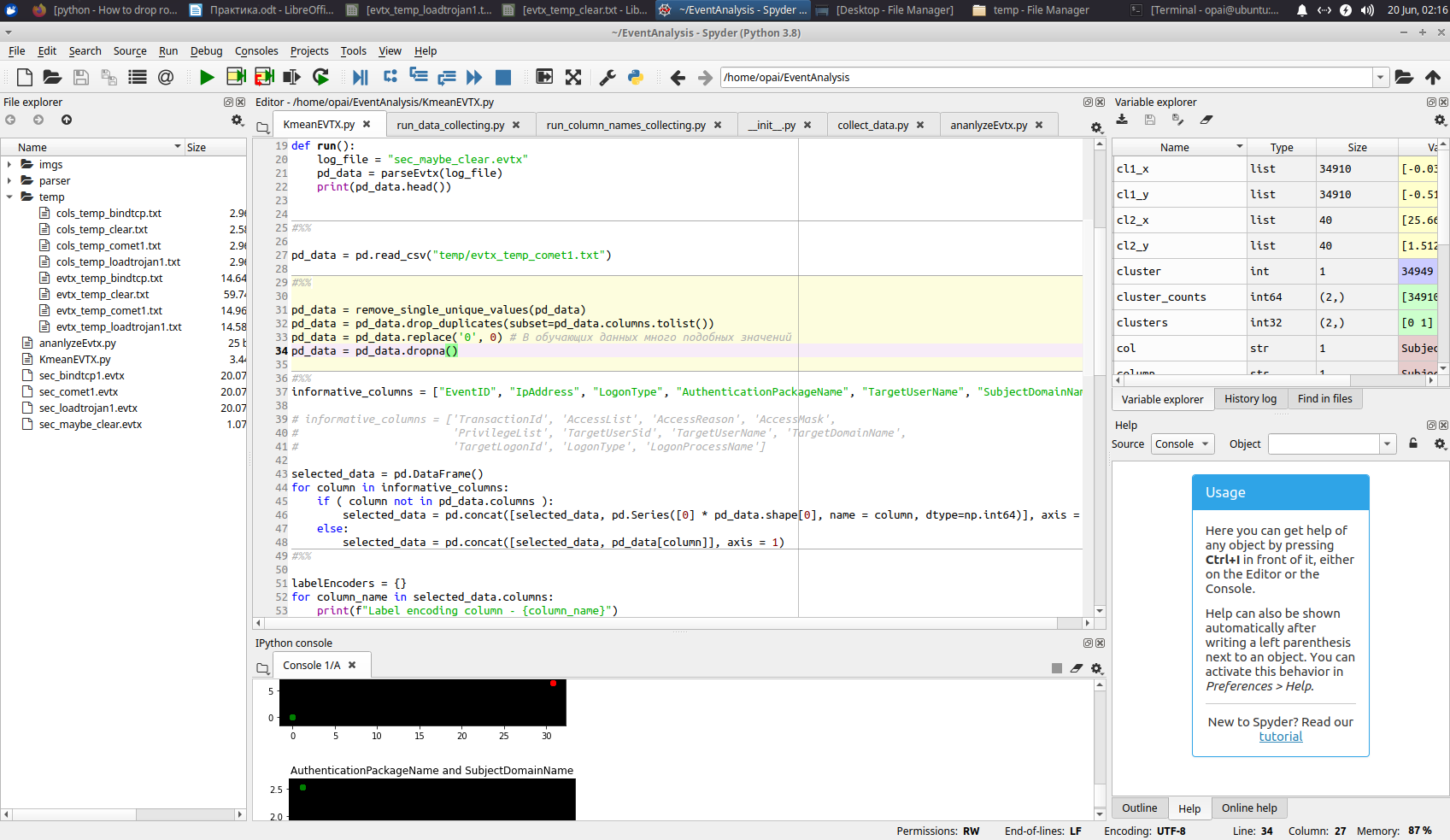
*Рисунок 1. Дополнительная функция в run\_data\_collecting*

**

*Рисунок 2. Дополнительная функция в run\_column\_names\_collecting*

## Обработка входных данных

В первую очередь датасет очищается от дубликатов, и Nan значений. Также, в датасете заменяются “0” типа str на 0 типа int, так как значения одинаковые, но если оставить их в таком виде, то в дальнейшем преобразовании данных получится два различных значения.

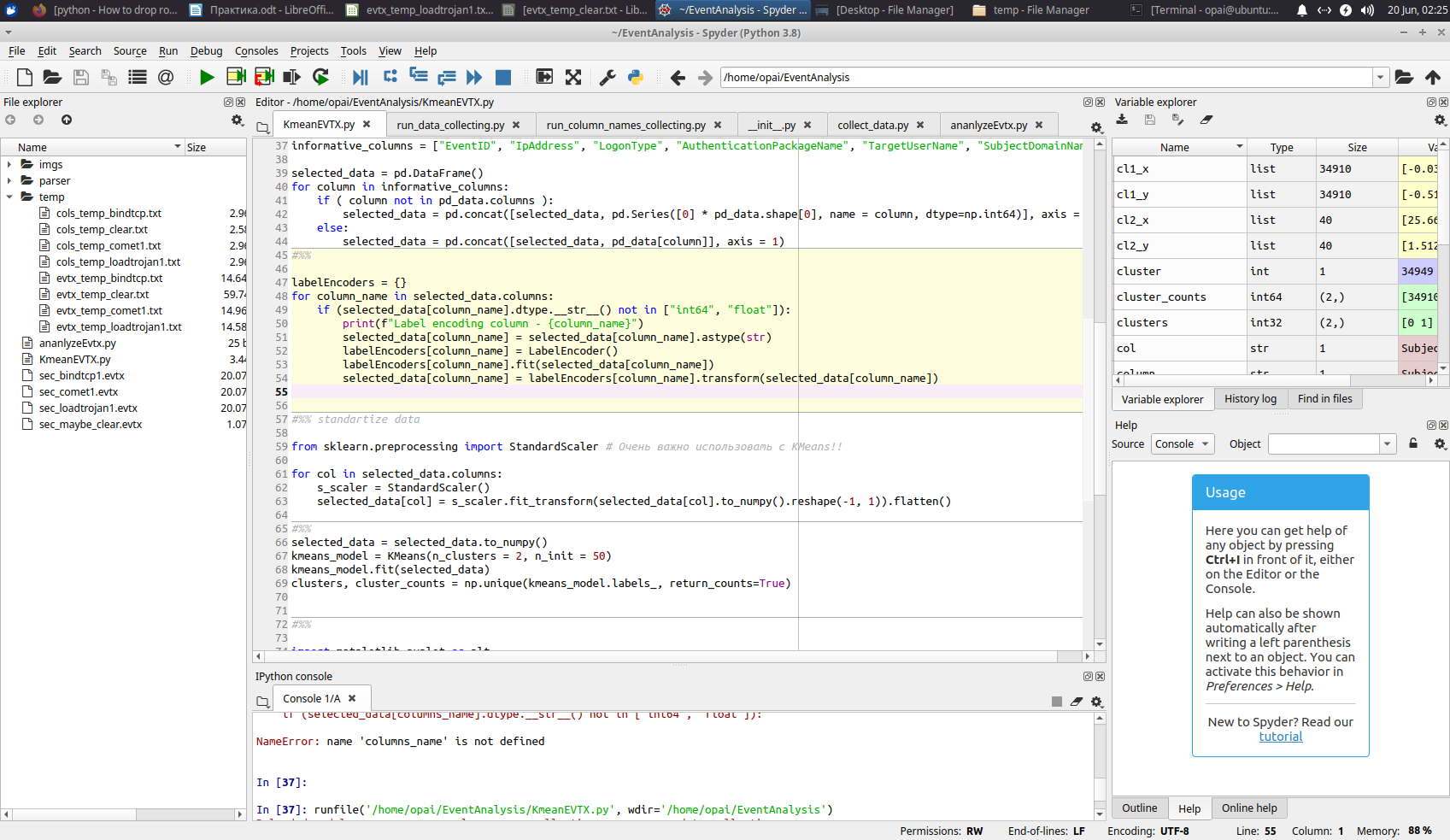
Рисунок 3. Очистка датасета от дубликатов, nan значений.

Для выполнения кластеризации методом перебора были выбраны следующие параметры security ивентов:

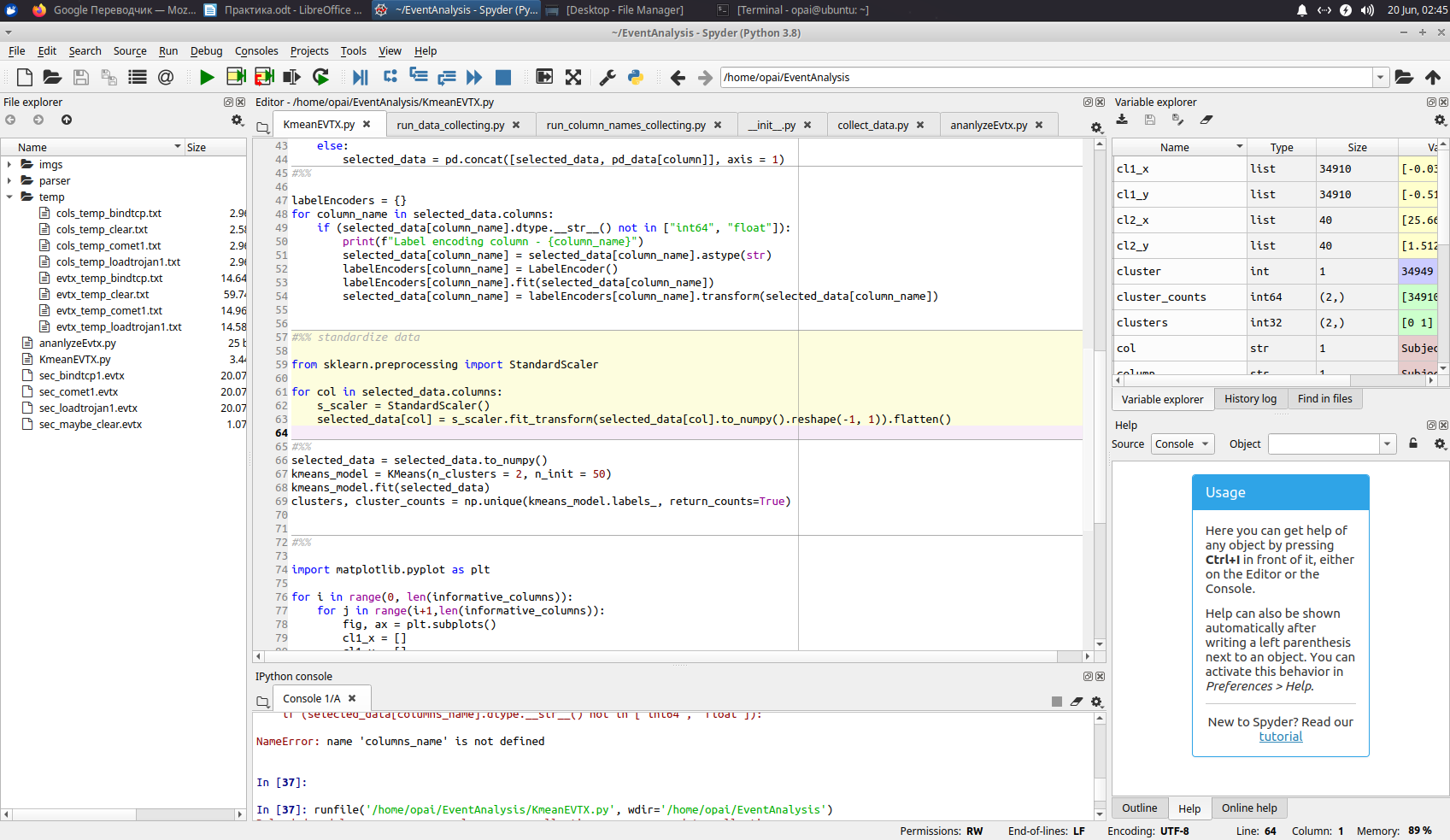
1. EventID
2. IpAddress
3. LogonType
4. AuthenticationPackageName
5. TargetUserName
6. SubjectDomainName

Такой набор выбранных параметров объясняется хорошим результатом кластеризации при небольшом количестве параметров.

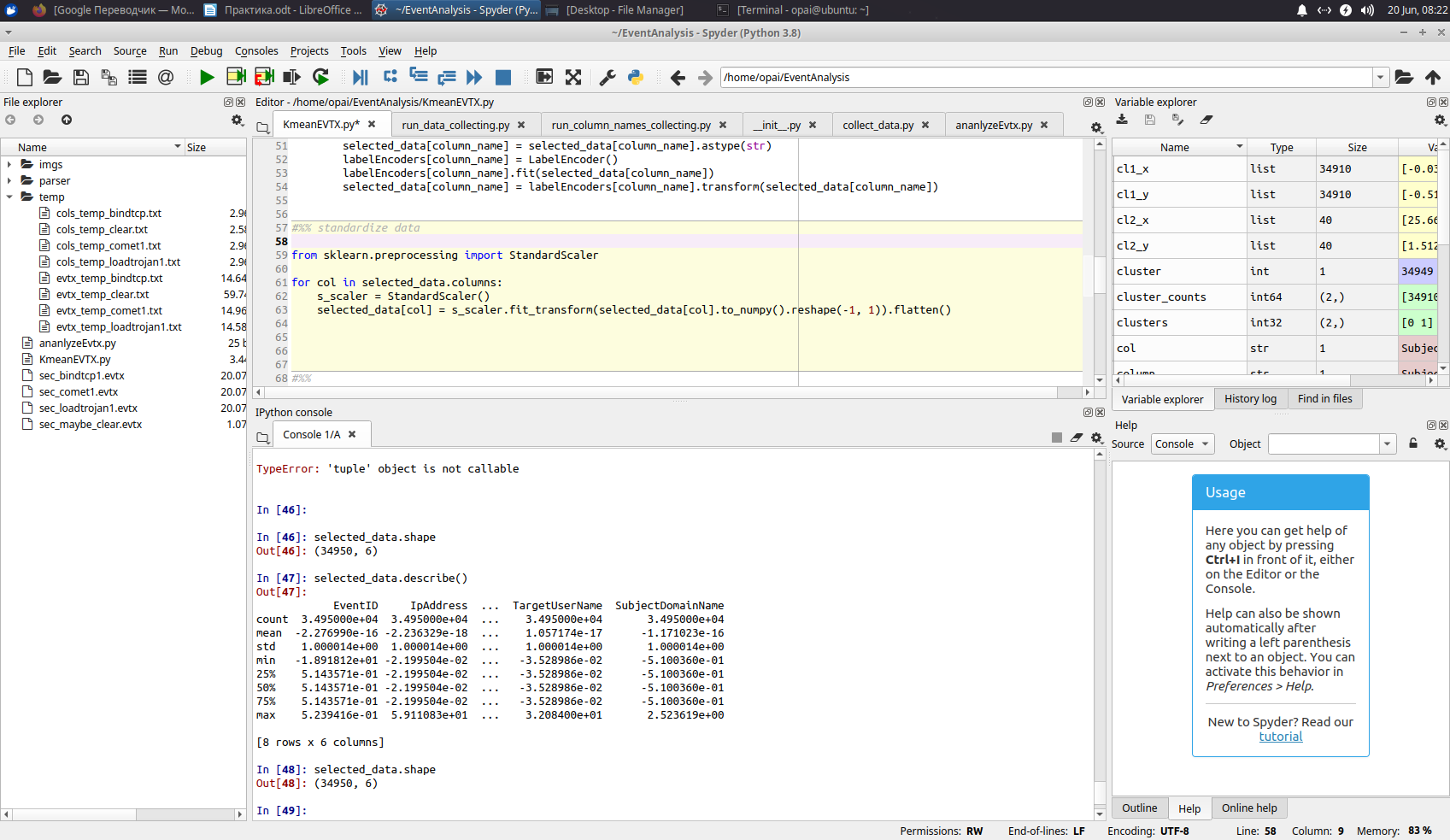
После этого столбцы, содержащие нечисловые значения преобразуются в числовые.

*Рисунок 4. Преобразование нечисловых значений*

Одним из ключевых моментов обработки входных данных является cтандартизация значений в рамках одного параметра путем удаления среднего и масштабирования до единичной дисперсии. В работе [1] показана эффективность используемой стандартизации. Так как, алгоритм Kmeans основан на применении геометрических мер (например, Евклидово расстояние), то до признаков с большей разницей между максимальных и минимальных элементов будут неявно оказывать большее влияние на кластеризацию в отличие от признаков с меньшим диапазоном.

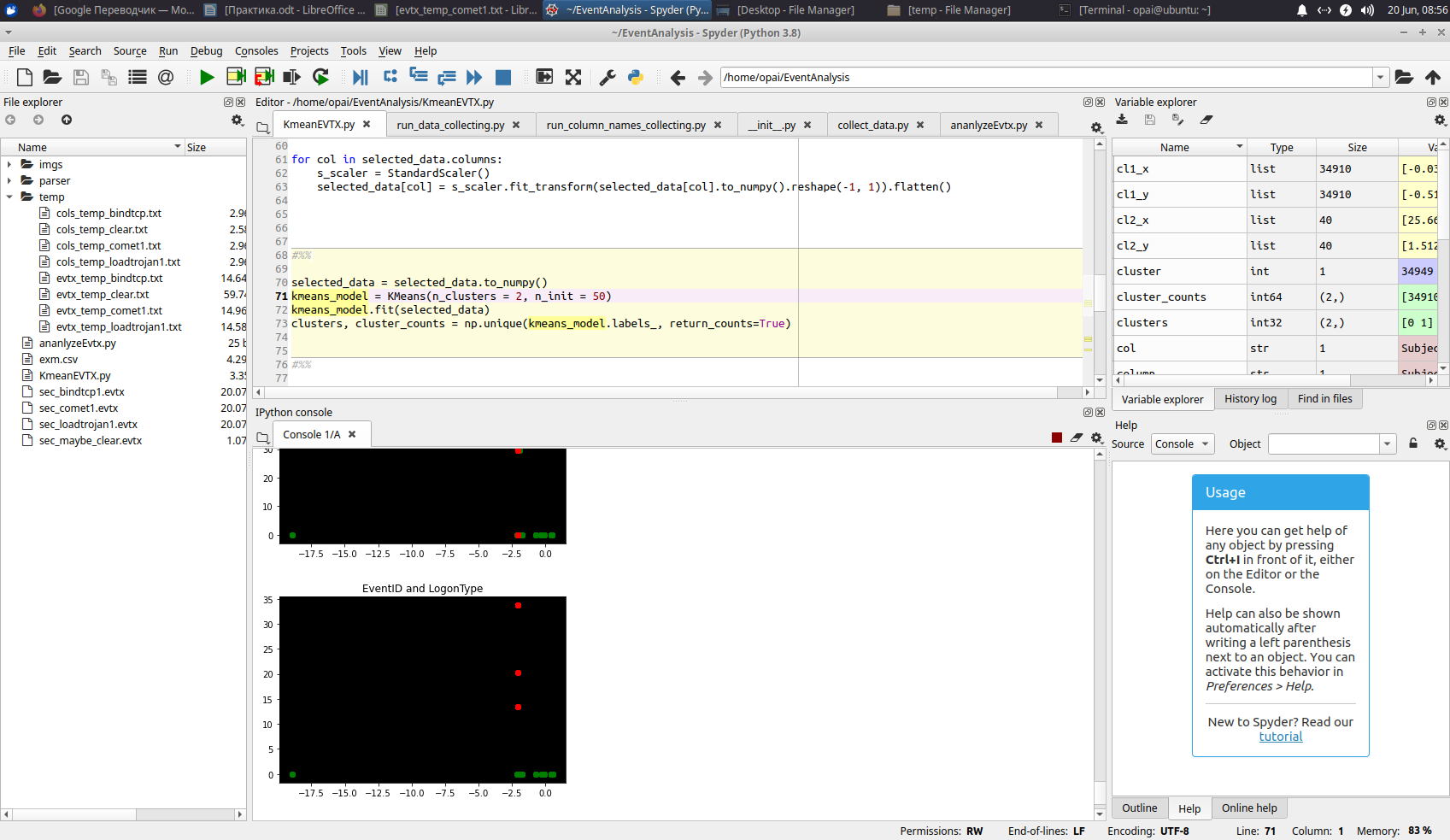
*Рисунок 5. Стандартизация входных данных*

Результат обработки данных изображен на рисунке 6.

*Рисунок 6. Статистические характеристики и размер выборки*

## Кластеризация алгоритмом KMeans

Код выполненения кластеризации представлен на Рисунке 7.

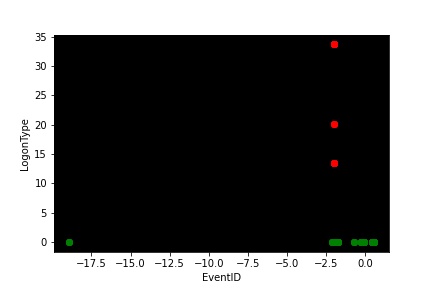
*Рисунок 7. Код выполнения кластеризации*

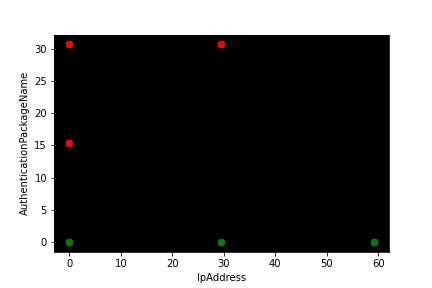
В таблице 1 представлена информация о полученных размерах кластеров для различных входных данных.

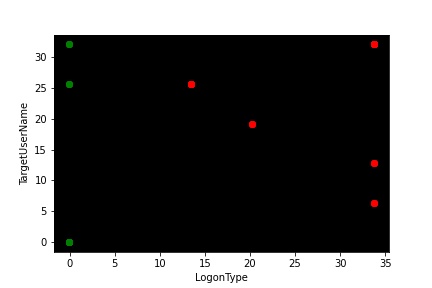
|  |  |  |  |
| --- | --- | --- | --- |
| **Вредоносное ПО** | **Количество ивентов** | **Кластер 1** | **Кластер 2** |
| Dark Comet | 34950 | 34910 | 40 |
| Bind TCP | 32284 | 32193 | 91 |
| LoadTrojan1 | 31722 | 31621 | 101 |
| MailTrojan1 | 31703 | 31602 | 101 |

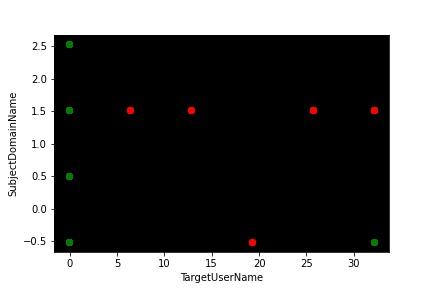
*Таблица 1. Размер получившихся кластеров для различных логов*

Графики по некоторым парам признаков представлены на Рисунках 8 – 11.

*Рисунок 8*

*Рисунок 9*

*Рисунок 10*



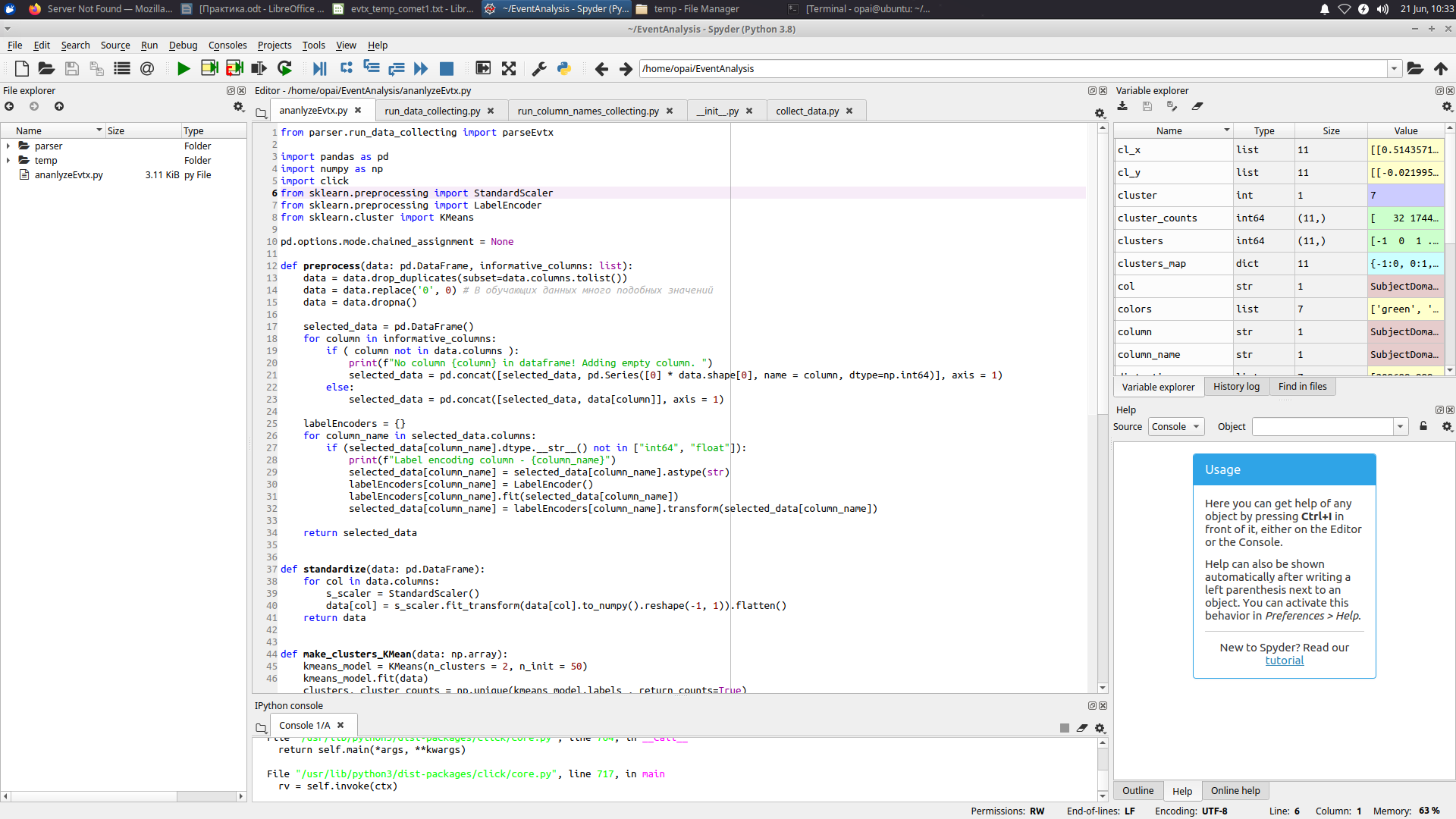
*Рисунок 11*

## Python скрипт детектирующий аномалии в ивентах безопасности Windows

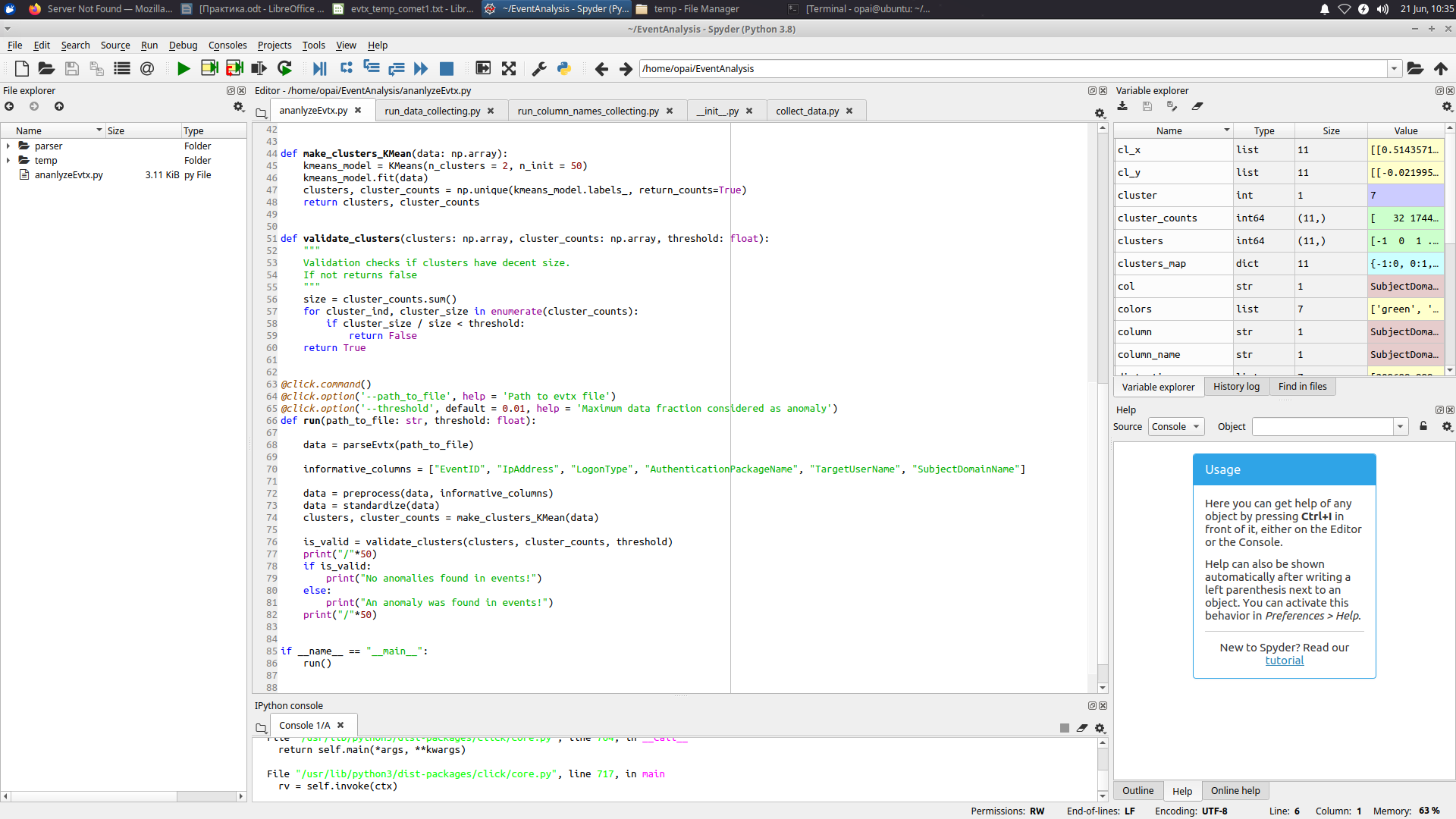
Детектирование аномалий выполняется при помощи кластеризации на два кластера, с помощью алгоритма KMeans. Если размер одного из кластеров оказывается меньше порогового значения, то данный кластер рассматривается как аномальный.

Такой подход объясняется тем, что количество нормальных ивентов безопасности зачастую сильно превышает число аномальных ивентов. Также, алгоритм кластеризации может выполнить кластеризацию «чистого» трафика на несколько кластеров, но по соотношению размеров кластеров можно определить, что данный трафик не аномальный.

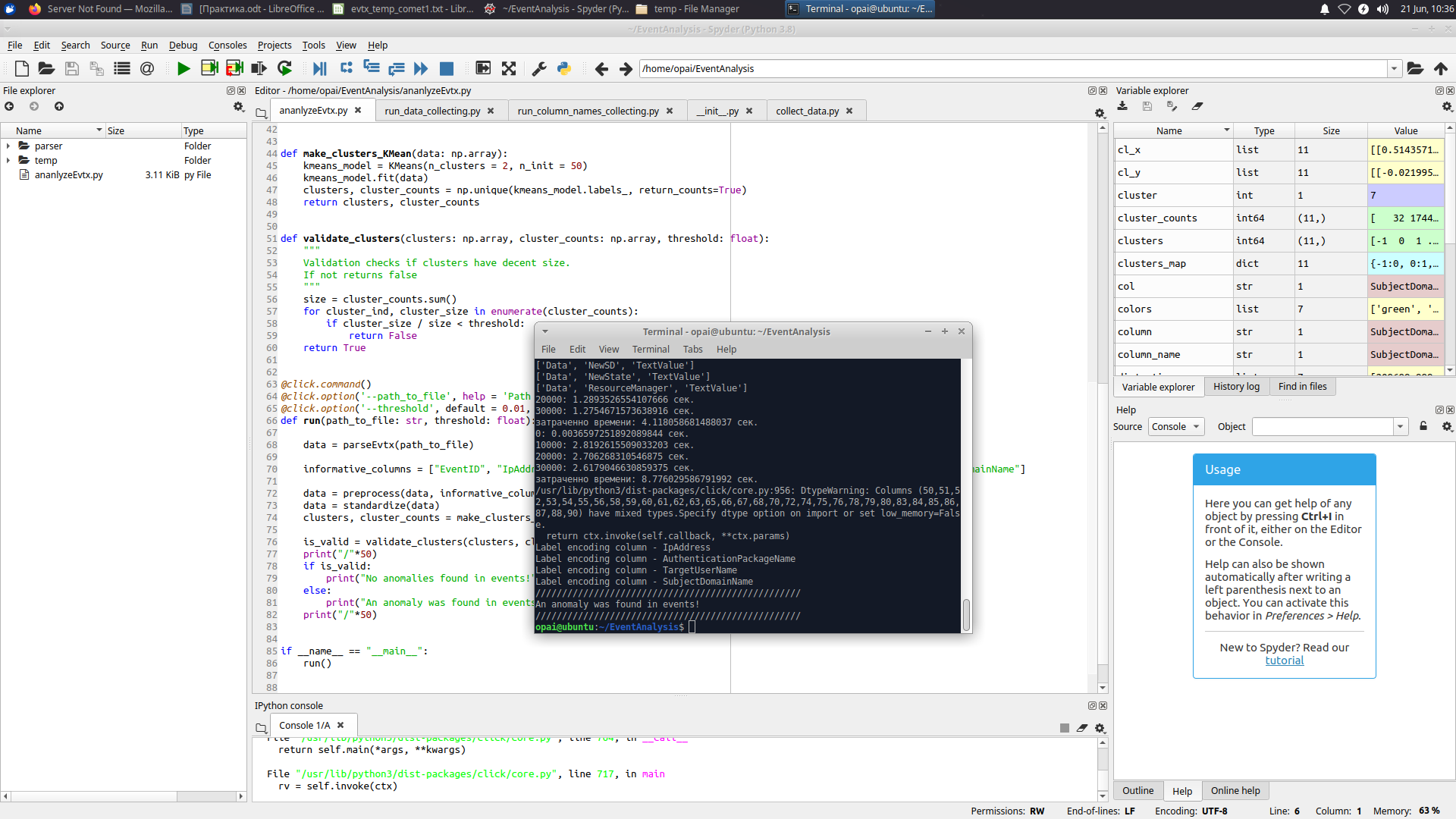
Код и пример использования скрипта показан на рисунках 12 - 14.



*Рисунок 12*



*Рисунок 13*



*Рисунок 14*

# Список литературы

1. Mohamad, Ismail & Usman, Dauda. (2013). Standardization and Its Effects on K-Means Clustering Algorithm. Research Journal of Applied Sciences, Engineering and Technology. 6. 3299-3303. 10.19026/rjaset.6.3638.