

# Итоговый проект по курсу «Инженер данных»

Лисниченко Марина

2022

# Структура

- Постановка задачи
- Решение
- Представление обработанных данных
- Визуализация данных
- Выводы

# Постановка задачі

- Сформировать витрину на основе логов web-сайта
- Пример строки:

```
'54.36.149.41 - - [22/Jan/2019:03:56:14 +0330] "GET  
/filter/27|13%20%D9%85%DA%AF%D8%A7%D9%BE%DB%8C%DA%A  
9%D8%B3%D9%84,27| %DA%A9%D9%85%D8%AA%D8%B1%20%D8%  
A7%D8%B2%205%20%D9%85%DA%AF%D8%A7%D9%BE%DB%8C%D  
A%A9%D8%B3%D9%84,p53 HTTP/1.1" 200 30577 "-" "Mozilla/5.0  
(compatible; AhrefsBot/6.1; +http://ahrefs.com/robot/)" "-" \n'
```

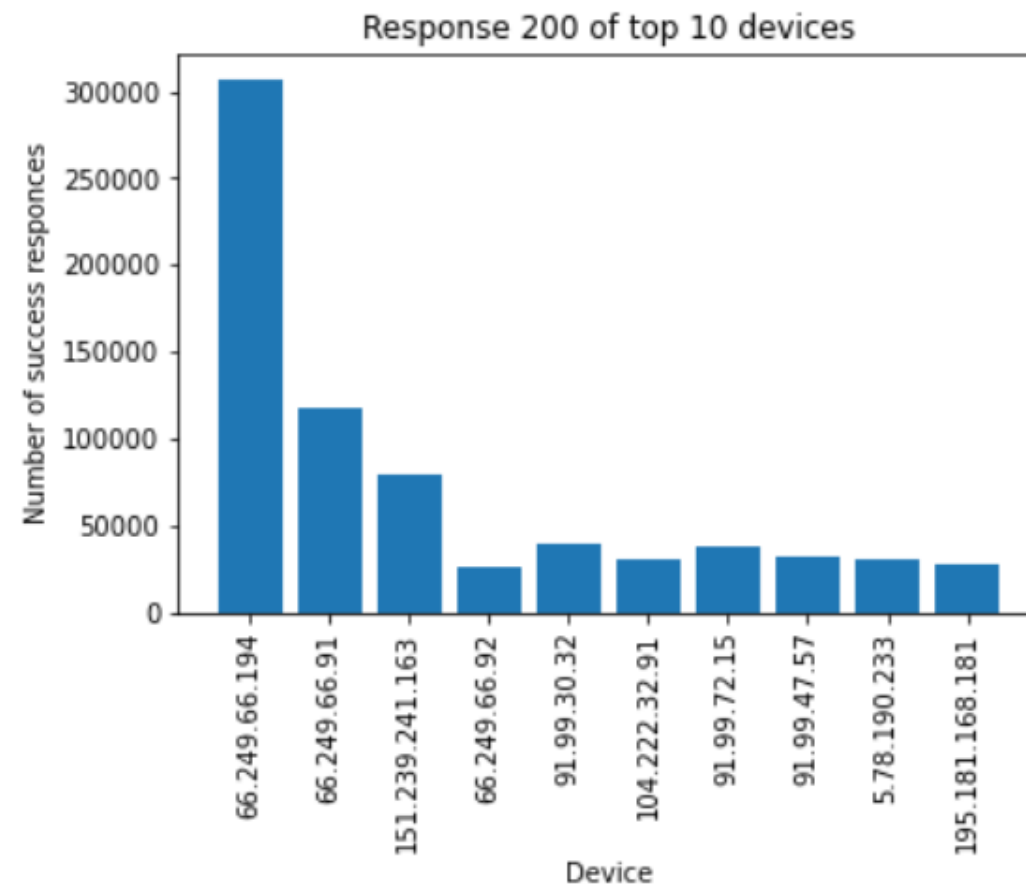
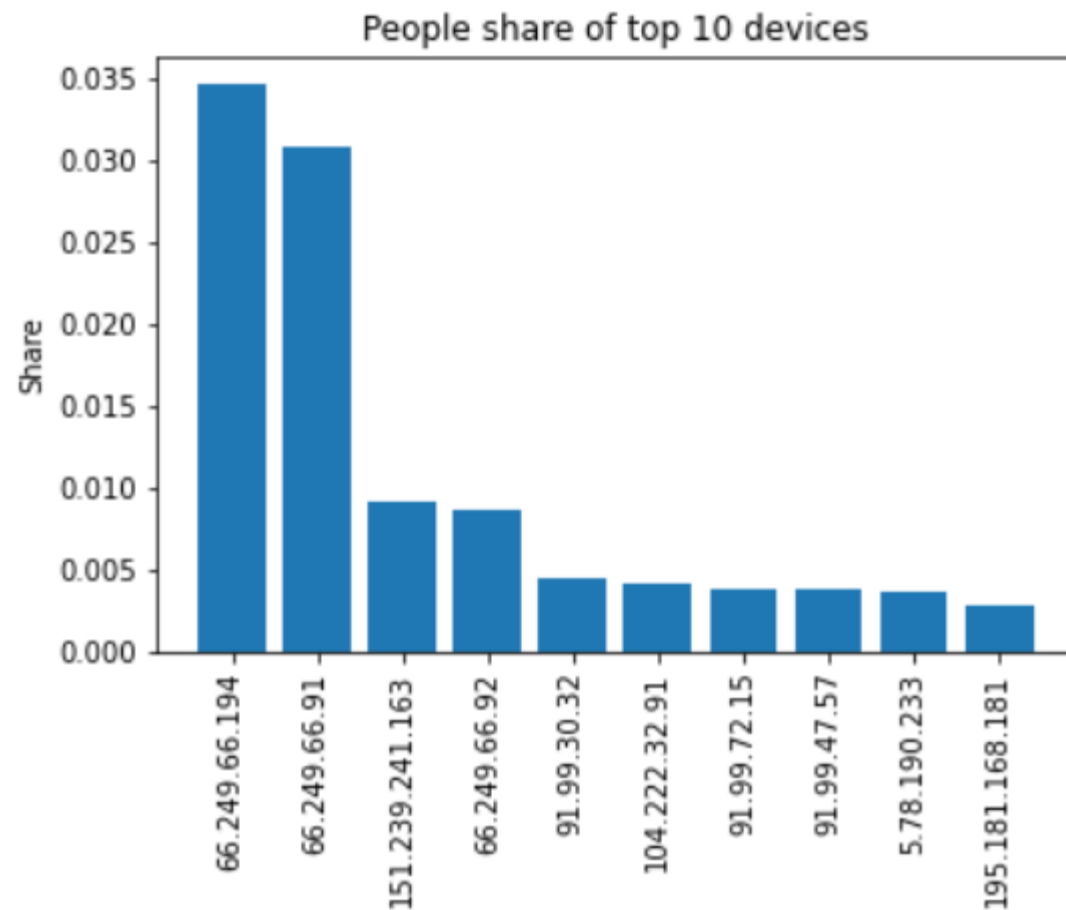
# Решение

- Разбить строку на **имя устройства, совершенное действие, ответ сервера**.
- Подсчитать статистику числа **пользователей**, числа **запросов**, числа **успешных запросов** и числа **запросов с ответом отличным от 200**.
- Составить объект типа dict и перевести в pandas Dataframe с соответствующими колонками.
- Визуализировать данные.

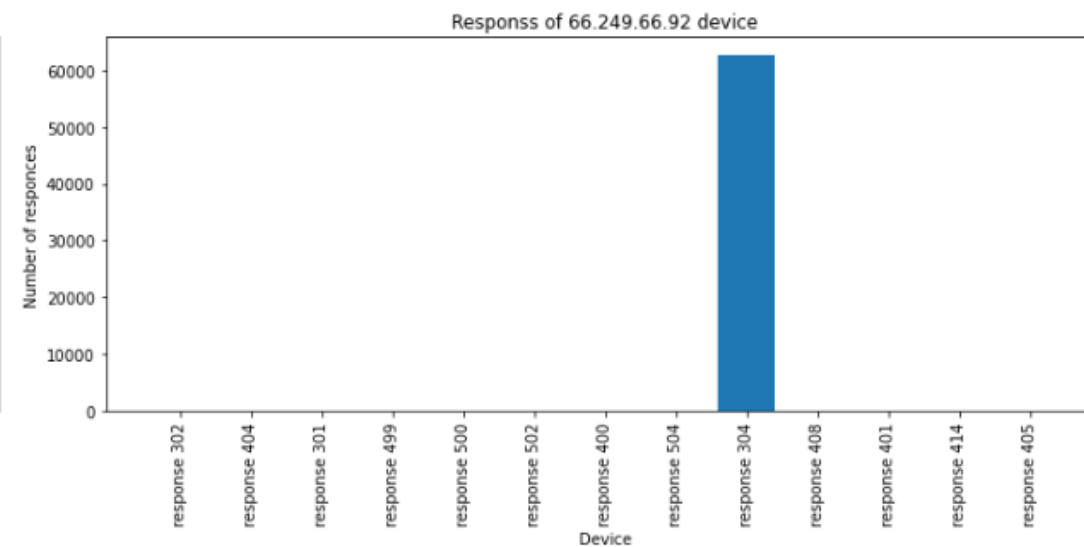
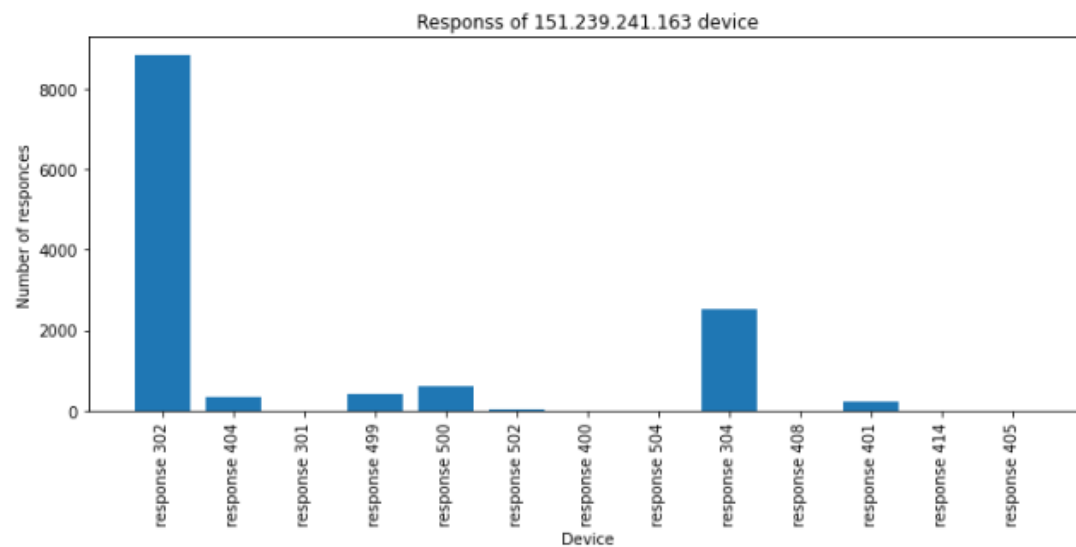
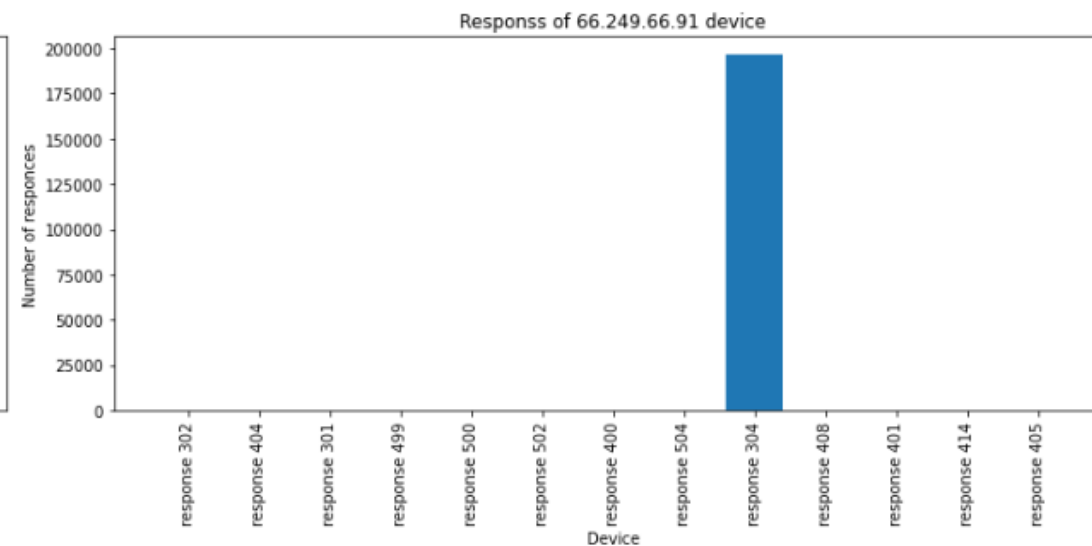
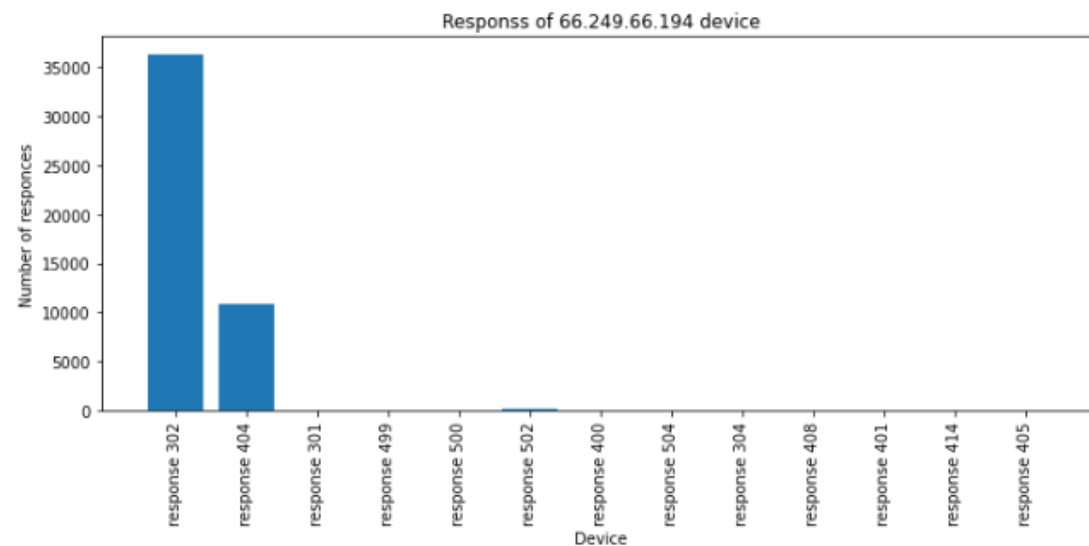
# Представление обработанных данных

[illegible]

# Визуализация данных



# Визуализация данных



# Выводы

В ходе работы было проведено чтение необработанных данных, их структурирование и обработка. После этого данные были представлены в виде таблицы с полями, предоставляющими статистическую информацию об устройстве. На последнем шаге статистические данные были визуализированы.