

**Оглавление**

1.	Постановка задачи.....	2
2.	Описание решения .....	3
3.	Анализ результатов .....	5
4.	Выводы .....	7

## **1. Постановка задачи**

В рамках лекций по практике была изучена библиотека для сбора данных BeautifulSoup, созданная для анализа документов HTML.

Также был изучен Pandas для записи и анализа данных. Были показаны основы работы с таблицами и примеры разведывательного анализа. Также были рассмотрены библиотеки для визуализации данных, такие, как seaborn и matplotlib.

Были показаны примеры разработки веб-приложений.

Для выполнения домашнего задания требовалось разработать программу с использованием веб-данных и веб-приложений. Это мог быть веб-сервис на удобном фреймворке для веб-разработки, бот для соцсети, как VK или Telegram или программа для сбора интересных данных, их обработка, анализ и визуализация.

Требовалось использования API и BeautifulSoup. Нужно было как-то поработать с данными.

В коде программы приветствовались объектно-ориентированный стиль, статические методы, комментарии к строчкам кода, создание исключения для ловли ошибок, и проверка стиля написанного приложения на PEP8.

## 2. Описание решения

### 2.1. Данные

При помощи библиотеки Requests был отправлен HTTP get-запрос на сайт и библиотекой BeautifulSoup были собраны данные об аниме (японская анимация).

Далее из этих данных был создан Pandas Dataframe со следующими столбцами:

"anime\_name" - название аниме,  
"anime\_year" - год выхода аниме,  
"anime\_genre" - жанры аниме,  
"anime\_views" - количество просмотров,  
"number\_of\_episodes" - количество эпизодов,  
"anime\_url" - ссылка для просмотра,  
"anime\_comments" - количество комментариев,  
"anime\_description" - небольшое описание сюжета,  
"image" - ссылка на картинку,  
"ongoing" - индикатор выхода новых серий.

Также в каждом столбце данные были проредактированы для удобства чтения и обработки и приведены к нужным типам данных.

Все данные записывались в файл anime.csv для удобного использования.

### 2.2. Используемые методы

Был проведен анализ и визуализация собранных данных. При помощи библиотек Pandas и Seaborn были построены различные графики и гистограммы, были выведены некоторые факты и зависимости по данным. Была построена корреляция.

Далее при помощи библиотеки pyTelegramBotAPI был написан телеграмм бот.

При помощи кнопки 'Start' бот активировался.

В функционале бота присутствовали три раздела:

'Genre' – при нажатии данной кнопки появлялась возможность выбрать лучшие аниме по жанру. После выбора жанра пользователь получал информацию по самым просматриваемым аниме жанра. Данные брались из anime.csv.

'Name' – кнопка, отвечающая за поиск информации по аниме. После ее нажатия требовалось ввести название интересующего аниме. Далее, если в столбце 'anime\_name' файла anime.csv присутствовало нужное аниме, то пользователь получал полную информацию про него.

‘/Game’ – после нажатия этой кнопки пользователь попадал в режим игры-викторины. При помощи метода ‘sample’ библиотеки pandas выбиралось четыре аниме для вопроса, на экран выводилось описание одного из них, это аниме, соответствующие правильному ответу, которое выбиралось при помощи библиотеки ‘random’.

Пользователю выводилось описание одного из аниме из столбца ‘anime\_description’ и четыре варианта ответа названий.

### **2.3. Код**

Репозиторий:

<https://github.com/vladislareon/animedata>

### 3. Анализ результатов

В файле ‘anime\_analysis.ipynb’ при помощи библиотеки ‘seaborn’ были визуализированы некоторые зависимости и факты по аниме данным из anime.csv.

Было выяснено, что индустрия аниме сильно развивается. С 1983 года количество аниме, вышедших за год, только растет. (Рисунок 1)

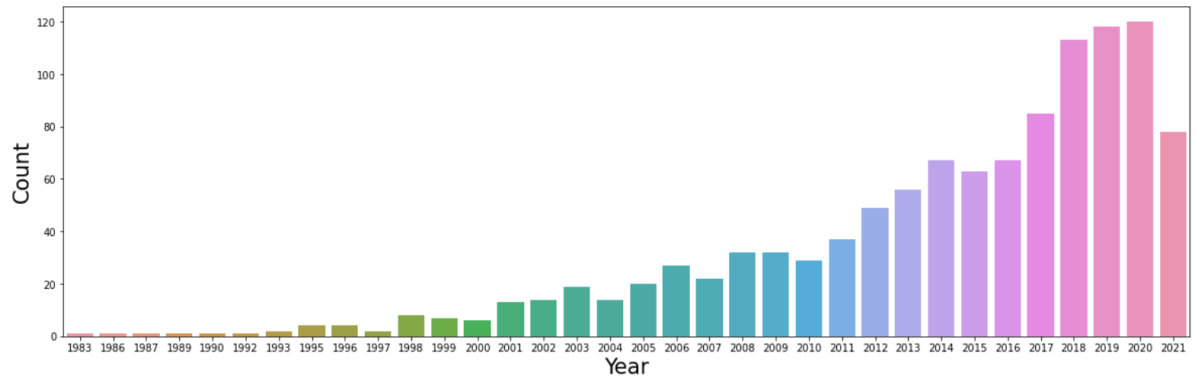


Рисунок 1

По средним значениям самые популярные аниме выходили в 2002, 1999, 2017 и 2020. (Рисунок 2)

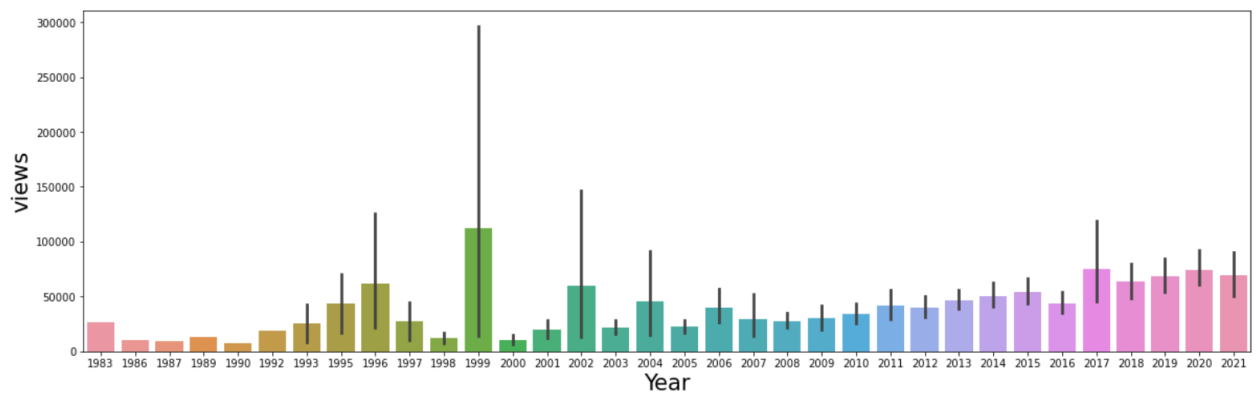


Рисунок 2

Самыми популярными оказались аниме со средним количеством серий от 170 до 300. (Рисунок 3)

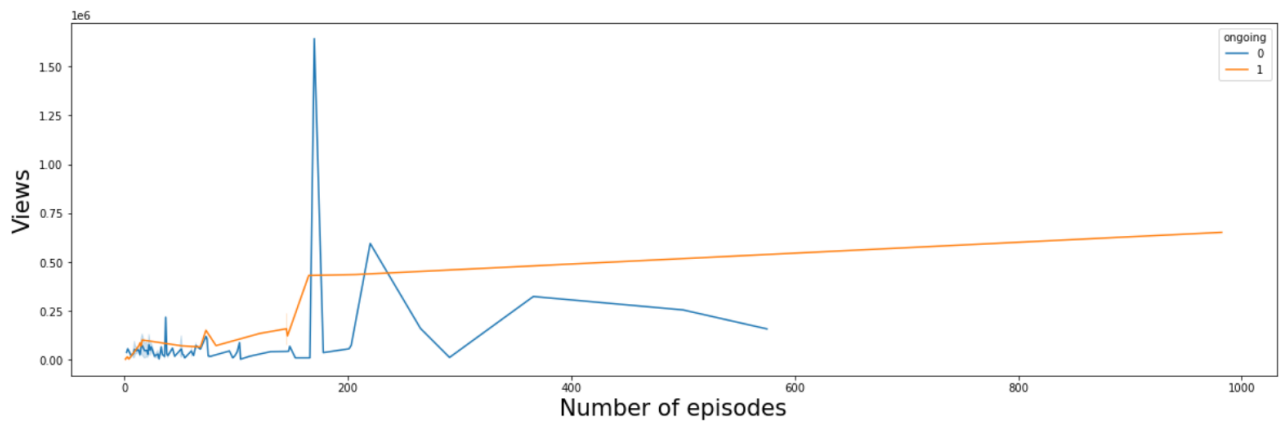


Рисунок 3

Больше всего аниме соответствуют жанру 'Комедия'.

Была замечена линейная зависимость количества серий у аниме, которые уже закончились и годом выхода. В современном мире стало актуальным выпускать аниме с небольшим количеством серий.

#### **4. Выводы**

В ходе практики были получены знания по сбору, анализу и использованию данных, опыт работы с такими библиотеками, как 'Pandas' и 'Seaborn'.

Получены навыки по написанию WEB-сервисов, web-scraper`ов и телеграмм ботов.

Был получен полноценный продукт, который состоит из множества частей, таких, как сбор данных, их визуализация и использование для работы с пользователями в backend`е.

В продолжении практики и улучшении полученного приложения можно добавить функцию рекомендации аниме по действиям пользователя при помощи методов машинного обучения. Так же возможность оставлять пользователями комментариев для использования их, как признаков в обучении модели.