**Многопрофильная олимпиада «Путь к успеху»**
**Секция «Основы искусственного интеллекта»**

*Задание заключительного тура*

*Вступление*

*Кто мы? Мы команда BIiIT, которая любит собирать данные, создавать и предобрабатывать большущие датасеты, а также придумывать неожиданные гипотезы.*
*А кто ты? Говорят, ты Data Scientist, умеешь работать с большими данными, получать описательные статистики и круто их визуализировать. Но мы не знаем: junior, middle, senior ли ты. Вот и узнаем кто есть кто* 😉*!*

---

## *Тема задания: Анализируй визуализируя!*

---

Вам представлен датасет (файл: YT_comments_dataset.pqt) с комментариями одного из популярных русскоязычных каналов в YouTube (скачать файл: https://drive.google.com/file/d/1wzg7pXDi2dAi0FEOUWgeGR4qXEXmd5yJ/view).
Мы поскрабили все, что можно было, почистили данные и передаем эту ценную информацию тебе.
Что мы ждём от тебя:

1. Получить основные показатели описательной статистики.
2. Представить визуализацию содержимого датасета.
3. Определить: чей канал ты анализировал(ла) (Название канала)

**В качестве ответа мы ждём файл в форматах: py или ipynb.**
Кстати, комментарии к коду будут не лишними, заработаешь больше баллов!

**ОПИСАНИЕ ДАТАСЕТА**
**ДАТАСЕТ СОДЕРЖИТ СЛЕДУЮЩИЕ ПОЛЯ:**
- **ПОЛЕ “DATE” - СОДЕРЖИТ ДАТУ И ВРЕМЯ РАЗМЕЩЕНИЯ КАЖДОГО КОММЕНТАРИЯ;**
- **ПОЛЕ “TEXT” - СОДЕРЖИТ ТЕКСТ КОММЕНТАРИЯ;**
- **ПОЛЕ “AUTHOR” - СОДЕРЖИТ НИКНЕЙМ АВТОРА КОММЕНТАРИЯ;**
- **ПОЛЕ “AVATAR” - СОДЕРЖИТ URL АВАТАРА АВТОРА КОММЕНТАРИЯ;**
- **ПОЛЕ “VIDEO THEME” - СОДЕРЖИТ ТЕМУ (О КОМ ИЛИ О ЧЕМ) ВИДЕО, ПОД КОТОРЫМ ОСТАВЛЕН КОММЕНТАРИЙ.**

### Подробнее о задании

1. Получить следующие данные из датасета:
- количество записей в датасете;
- типы записей;
- количество уникальных комментаторов;
- вывести имя автора, аватар, количество комментариев данного автора;
- вывести 5 самых обсуждаемых героев выпусков на основании количества комментариев.

2. Получить частоту комментирования видео пользователя по дням недели и по годам:

| Год<br>День недели | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|
| Понедельник | | | | | |
| ... | | | | | |
| Воскресенье | | | | | |

3. Получить количество комментариев по годам:

| Год | Количество комментариев |
|---|---|
| 2017 | |
| 2018 | |
| 2019 | |
| 2020 | |
| 2021 | |

4. *С помощью анализа социальных сетей исследователи получили список самых популярных эмодзи.* Необходимо получить по каждому автору комментариев статистику использования данных эмодзи:

- определить эмоциональную окраску комментариев на всем канале;
- определить самых доброжелательных комментаторов (топ-10), самых негативных комментаторов (топ-10), самых эмоциональных комментаторов (топ-10).

| Изображение эмодзи | В Unicode | Эмоциональная окраска |
|---|---|---|
| 😀 | U+1F600 | Позитивный |
| 😁 | U+1F601 | Позитивный |
| 🤣 | U+1F923 | Позитивный |
| 😂 | U+1F602 | Позитивный |
| 😉 | U+1F609 | Позитивный |
| 😊 | U+1F60A | Позитивный |
| 😍 | U+1F60D | Позитивный |

| | | |
|---|---|---|
|  | U+1F618 | Позитивный |
|  | U+1F914 | Нейтральный |
|  | U+1F612 | Нейтральный |
|  | U+1F60E | Позитивный |
|  | U+1F62D | Негативный |
|  | U+1F48B | Позитивный |
|  | U+1F495 | Позитивный |
|  | U+2764 | Позитивный |
|  | U+2665 | Позитивный |
|  | U+1F44D | Позитивный |
|  | U+1F937 | Нейтральный |
|  | U+1F525 | Позитивный |
|  | U+1F44E | Негативный |
|  | U+1F644 | Негативный |
|  | U+1F922 | Негативный |
|  | U+1F620 | Негативный |
|  | U+1F92E | Негативный |
|  | U+1F615 | Негативный |
|  | U+1F928 | Негативный |
|  | U+1F47F | Негативный |
|  | U+2642 | Негативный |
|  | U+1F926 | Негативный |
|  | U+1F929 | Позитивный |

| | | |
|---|---|---|
| 🥰 | U+1F970 | Позитивный |
| 👏 | U+1F44F | Позитивный |

5. Визуализировать полученные данные:
   - частоту комментирования по дням неделям и годам (График 1);
   - количество комментариев по годам (График 2);
   - определить доминирующую эмоцию каждого автора из топ-10 самых активных комментаторов. Построить график, отображающий доминирующую эмоцию по каждому автору из топ-10 самых активных комментаторов. На графике вывести по каждому автору его аватар, вывести доминирующие эмоции в виде эмодзи (График 3).

Вывод графиков должен быть реализован с анимацией (построение графика, вывод дополнительной, уточняющей информации).

Все графики должны содержать легенды, подписи осей, название графика.

6. Написать в конце ноутбука название канала, чьи комментарии были в датасете.