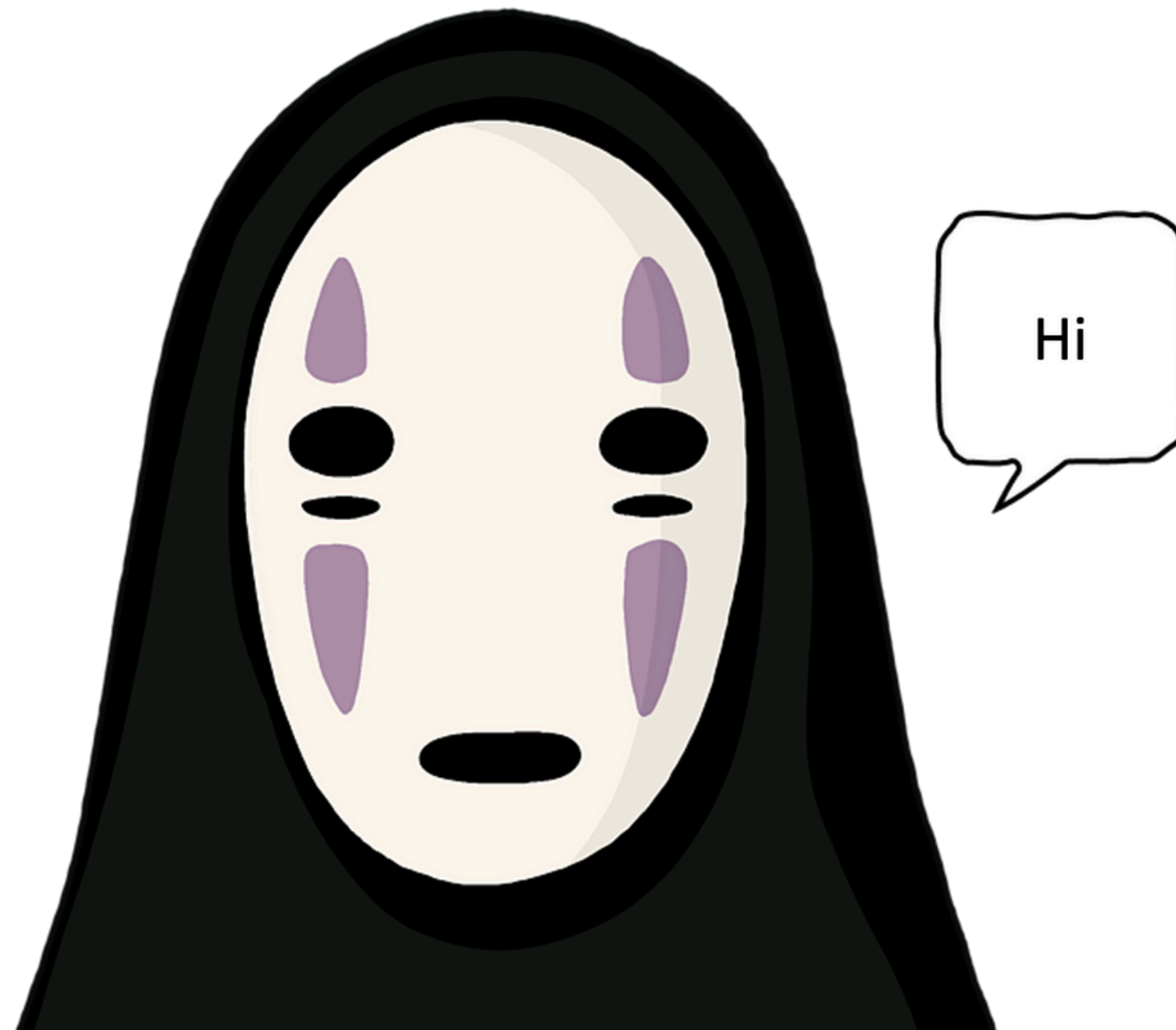


# DATA 400 PROJECT

Faye Le & Dylan Nguyen





# Predicting Anime Score Based on Common Media Metadata Fields

# DATA COLLECTING



Source: [myanimelist.net/topanime](http://myanimelist.net/topanime)

Scrape data using Selenium and BeautifulSoup

- Selenium: Rank, Title, URL, Score - 10,000 data
- BeautifulSoup: Genre, Episode, Premiere, Studio, Source - 4,000 data

Limitation:

- Missing value = Not found, regression model = 0
- Mostly categorical variables
- Time consuming

|      | Rank | Title   | URL   | Score | Genres  | Episodes | Premiered                  | Studio         | Type       | Source           |
|------|------|---|---|-------|---|----------|----------------------------|----------------|------------|------------------|
| 0    | 1    | Sousou no Frieren                                 | https://myanimelist.net/anime/52991/Sousou_no_... | 9.32  | Adventure, Drama, Fantasy, Shounen                | 28       | Fall 2023                  | Madhouse       | TV         | Manga            |
| 1    | 2    | One Piece Fan Letter                              | https://myanimelist.net/anime/60022/One_Piece_... | 9.17  | Action, Adventure, Fantasy, Shounen               | 1        | Premiered season not found | Toei Animation | TV Special | Light novel      |
| 2    | 3    | Fullmetal Alchemist: Brotherhood                  | https://myanimelist.net/anime/5114/Fullmetal_A... | 9.09  | Action, Adventure, Drama, Fantasy, Military, S... | 64       | Spring 2009                | Bones          | TV         | Manga            |
| 3    | 4    | Steins;Gate                                       | https://myanimelist.net/anime/9253/Steins_Gate    | 9.07  | Drama, Sci-Fi, Suspense, Psychological, Time T... | 24       | Spring 2011                | White Fox      | TV         | Source not found |
| 4    | 5    | Shingeki no Kyojin Season 3 Part 2                | https://myanimelist.net/anime/38524/Shingeki_n... | 9.05  | Action, Drama, Suspense, Gore, Military, Survi... | 10       | Spring 2019                | Wit Studio     | TV         | Manga            |
| ...  | ...  | ...   | ...   | ...   | ...   | ...      | ...                        | ...            | ...        | ...              |
| 3843 | 3996 | Gotou ni Naritai.                                 | https://myanimelist.net/anime/8120/Gotou_ni_Na... | 7.07  | Drama, Slice of Life, School                      | 1        | Premiered season not found | add some       | Movie      | Source not found |
| 3844 | 3997 | Hanyou no Yashahime: Sengoku Otogizoushi - Ni ... | https://myanimelist.net/anime/48505/Hanyou_no_... | 7.07  | Action, Adventure, Fantasy, Historical, Mythol... | 24       | Fall 2021                  | Sunrise        | TV         | Source not found |
| 3845 | 3998 | Hello! Sandybell                                  | https://myanimelist.net/anime/2799/Hello_Sandy... | 7.07  | Adventure, Romance                                | 47       | Spring 1981                | Toei Animation | TV         | Source not found |
| 3846 | 3999 | Higepiyo  | https://myanimelist.net/anime/5983/Higepiyo       | 7.07  | Comedy  | 39       | Spring 2009                | Kinema Citrus  | TV         | Source not found |
| 3847 | 4000 | Highschool of the Dead                            | https://myanimelist.net/anime/8074/Highschool_... | 7.07  | Action, Horror, Supernatural, Suspense, Ecchi,... | 12       | Summer 2010                | Madhouse       | TV         | Manga            |

3848 rows × 10 columns

# EDA



## Distribution Analysis

- Score Distribution
- Episodes Distribution

## Categorical Variable Exploration

- Genre Analysis: Plot a bar chart of genre frequencies
- Studio Analysis: Plot the number of anime per studio and their average scores
- Source Analysis: Analyze which sources tend to have higher average scores
- Premiere Analysis: Analyze trends over years and seasons

# EDA



## Correlation Analysis

- Numerical Correlations ('Score' and 'Episodes')
- Categorical Associations ('Score' and other variables)

## Text Analysis on Titles

- Analyze if there's any correlation between title length and 'Score'

## Temporal Analysis

- Plot the count of anime releases per year --> Analyze average scores per year to identify if the quality is improving or declining over time

