**Team Members**:

* Adrian Wood
* Tutors (David Pecot, Zubair Shaikh)
* students: (Mauvonte, Brandon, Tom)
* TA’s (Nick, Chris, Collin)
* Instructor (Dom)

**Topic:**

Anime has always been something I have enjoyed and have spent many years watching, talking and debating over anime, and this class has not allowed me to watch as much anime as I would like (and that a good thing in a lot of ways ha-ha). But this class has also shown me that there is data everywhere that needs to be analyzed. This project will let me lets me get my anime fix and do my group project (win-win to me). One thing I wanted to know was what was the best anime. One way I was thinking of finding this answer was to look at the rating of each anime, and the number of voters for each one. List them out by rating and number of votes. This would be the start, to see if there is any correlation between the number of users that voted and the average rating as well as weather the average rating is a good predictor of the number of users that voted for each show and lastly, if the length of the show would have any correlation to how the show would rate and how many users would be voting for each show based of of the data it was published.

**Steps**:

1. **Downloads** dataset.
2. **create a repository** for the project.
3. **clone** the data base/ **Look** **over** the data base.
4. Think about how to best do my testing with the data I have. Make it into workable database make it the **right size**, the **right information**, and **structure**, for the questions I have for this project.
5. **Remove the tv (##)** part of each entry on the data base under the episode’s column, and then also convert the **number of episodes** and **votes** column to **an integer** from a **variable.**
6. **Split** the data **column** to **a start date and end date**
7. **Clean data** format it into workable chunks.
8. **Create** the corresponding **charts**, including **bar graphs, line graphs. Scatter plots and box** and **whisker plots.**
9. **Formulate a summary** to be included in the **Jupiter Notebook**(s) or in the MD file.
10. Get everything all cleaned up and ready to **present**.

**\*\*\*The hypothesis’s test I intend to perform\*\*\***

-On average are **longer** shows (more episodes**) rated higher**? Or are shorter shows?

- Does **the number of voters** have any effect on the **rating** of the show?

(Do shows have a lot of votes and are the **positive** or **negative** **votes**)

-Is there a **sweet spot** for **the number of episodes, number of voters**, and **raking?**

**-Do older shows have a (xyz) better ratings and (xyz) more votes?**

- **Newer shows** have a (xyz %) **higher rating** then **older shows**, even if they have (xyz**) less voters**?

**Segregation of Duties:**

1. I(Adrian Wood),

must get it done myself and ask for help if I need it going forward

**\*\*\*Dataset(s):\*\*\***

<https://www.kaggle.com/datasets/srivnaman/top-100-anime-animelist?resource=download>

**\*\*\*Things I need to ask Dom About \*\*\***