**Team Members**:

* Adrian Wood (myself)
* Tutors (David Pecot, Zubair Shaikh, Asiha Braxton-Garvin, Kyle Goode)
* students: (Mauvonte.R, Brandon.G, Tom.L, Eric.B , Scott.N Akihil.B, Sung.A Amanda.N)

not all of them helped me with the project directly but they all have been very supportive and helped me learn a lot for me to be able to try and do this project alone.

* TA’s (Nick, Chris, Collin,SlackBot/AskBCS Learning Assistant)
* 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 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 dataset/ **Look** **over** the dataset.
4. Think about how to best do my testing with the data I have. Make it into a workable dataset 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 dataset 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 aired 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**.

**Extras**

- as a part of the data cleaning process will be splitting the ‘Aired’ column into two new columns titled “date aired” and “series finally”

-This will just be done so I can do more analysis on this dataset later as well as do any other testing if there is time before the end of class

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

Hypothesis:

**Shows with more episodes are rated higher than shows with fewer episodes.**

I plan to test this hypothesis using a ANOVA test.

What is ANOVA (ANalysis Of VAriance)

* **ANOVA test used to compare the means of more than 2 groups (t-test or**[***Z*-test**](https://www.reneshbedre.com/blog/z-test-in-python.html)**can be used to compare 2 groups)**
* **Groups mean differences inferred by analyzing variances**
* **ANOVA uses variance-based *F* test to check the group mean equality. Sometimes, ANOVA *F* test is also called omnibus test as it tests non-specific null hypothesis i.e. all group means are equal**
* **Main types: One-way (one factor) and two-way (two factors) ANOVA (factor is an independent variable)**

**Will want to talk with dom on how to make sure I do this correctly**

**I have a few youtube videos that I have been watching wanted to get some insight on these resources**

**Web resources**

* [ANOVA using Python (with examples) (reneshbedre.com)](https://www.reneshbedre.com/blog/anova.html)
* [Four Ways to Conduct One-Way ANOVA with Python - Erik Marsja](https://www.marsja.se/four-ways-to-conduct-one-way-anovas-using-python/)
* [One-way ANOVA with Python (pythonfordatascience.org)](https://www.pythonfordatascience.org/anova-python/)
* [Statistics in Python — Using ANOVA for Feature Selection | by Wei-Meng Lee | Towards Data Science](https://towardsdatascience.com/statistics-in-python-using-anova-for-feature-selection-b4dc876ef4f0)

**Youtube videos**

* [(76) Python for Data Analysis: ANOVA - YouTube](https://www.youtube.com/watch?v=EWYzeZbchR0)
* [(76) How To Know Which Statistical Test To Use For Hypothesis Testing - YouTube](https://www.youtube.com/watch?v=ChLO7wwt7h0)
* [(76) Hypothesis Testing - Anova & Chi Square Test of Independence using Python - YouTube](https://www.youtube.com/watch?v=X3nRFKJQdRs)
* [(76) How to perform Analysis of Variance (ANOVA) | One-way and Two-way ANOVA with Python - YouTube](https://www.youtube.com/watch?v=AhZ-hllEVxs)

**3 categories.**

**-Votes,**

**-Score,**

**-and number of episodes**

**visualizations**

Using these categories I will visualize the data to help me look at trends in the following ways

**1** -Number of Episodes (X) Vs. Score (Y) graph 1

(Main test 1)

This is just how to take a surface level look at the data to see if more episodes equal a higher rating.

**2** -number of episodes (X) VS. Votes (Y)

(Supporting graph 1)

based on the number of votes casted if there is some overlap from the first graph

**3** -Votes(X) VS. Score (Y)

(Tertiary graph 1)

this last one will just be a precaution test to double check the results of the first two

**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 \*\*\***

I think this should be all set to go I have taken some time to get caught up with the class and chipped away at this for some time now I hope I can just knock this out and keep moving forward to project 4 and hopefully the end of the class with a pass (no matter the score) and a certificate of completion