April 5, 2021

Mr. Lee A. Sun  
My AnimeRoll  
948 Leisure Lane  
Los Angeles, CA, 90013

### Letter of Transmittal – Anime Recommendation Application (Metis)

Dear Mr. Sun,

It was a pleasure speaking with you on April 1, 2021 regarding the creation of an anime recommendation application. One of the big rocks that you mentioned your company is currently facing is maintaining viewership. The company’s metrics surrounding attracting new viewers are excellent, however, My AnimeRoll has an opportunity when it comes to answering the “Now what?” moment that your viewers ask themselves when they reach the end of their favorite anime. Together, we can eliminate the stress of discovering a new show after the “Now what?” moment by offering customers the opportunity to follow-up a show they are enjoying with another show that your service provides that they might also enjoy. Answering the “Now what?” moment will allow you to not only continue to attract new viewers - as is your strength - but also take advantage of the opportunity to entice them to continue using My AnimeRoll exclusively for their anime watching needs.

Pursuant to this idea, I have designed Metis, a stand-alone anime recommendation application that houses an alphabetized list of all the titles that your company offers. Users will be able to enter the name of a show they enjoy or just finished watching and with the click of a button have 5 recommendations pop up for them. Metis does not require any specialized equipment or expensive software, keeping costs low. The project as it is laid out below will take 120 hours to create with an expected completion date of April 30, 2021. The resultant cost is only $6,000 to help bring My AnimeRoll to the forefront of anime streaming services.

With my 8 years working in a multi-level retail management environment, and 2 years of experience working with software, I can both build the software, and communicate with all levels of leadership and associates that will be part of the project team from both of our organizations, both technical and non-technical.

I look forward to working with you for the duration of this project.

Respectfully,

Bethany Carpio

|  |
| --- |
|  |
| Anime Recommendation Application (Metis) |
|  |

|  |
| --- |
| Bethany Carpio  4-9-2021  [Version 1.0] |

Contents

[Letter of Transmittal – Anime Recommendation Application (Metis) 1](#_Toc70017148)

[Section A 6](#_Toc70017149)

[Letter of Transmittal 6](#_Toc70017150)

[Project Recommendation 6](#_Toc70017151)

[Problem Summary 6](#_Toc70017152)

[Application Benefits 6](#_Toc70017153)

[Application Description 6](#_Toc70017154)

[Data Description 6](#_Toc70017155)

[Objective and Hypothesis 7](#_Toc70017156)

[Methodology 7](#_Toc70017157)

[Funding Requirements 7](#_Toc70017158)

[Stakeholder Impact 8](#_Toc70017159)

[Data Precautions 8](#_Toc70017160)

[Developer Experience 8](#_Toc70017161)

[Section B 9](#_Toc70017162)

[Problem Statement 9](#_Toc70017163)

[Customer Summary 9](#_Toc70017164)

[Existing System Analysis 9](#_Toc70017165)

[Data 9](#_Toc70017166)

[Project Methodology 10](#_Toc70017167)

[Project Outcomes 11](#_Toc70017168)

[Project Deliverables 11](#_Toc70017169)

[Product Deliverables 11](#_Toc70017170)

[Implementation Plan 11](#_Toc70017171)

[Environment Set-Up 11](#_Toc70017172)

[Application Testing 11](#_Toc70017173)

[Deployment 12](#_Toc70017174)

[Evaluation Plan 12](#_Toc70017175)

[Resources and Costs 13](#_Toc70017176)

[Programming Environment 13](#_Toc70017177)

[Environment Costs 13](#_Toc70017178)

[Human Resource Requirements 13](#_Toc70017179)

[Timeline and Milestones 14](#_Toc70017180)

[Section C 15](#_Toc70017181)

[Data Methods 15](#_Toc70017182)

[Data SetS 15](#_Toc70017183)

[Analytics 15](#_Toc70017184)

[Data Cleaning 16](#_Toc70017185)

[DatA Visualization 16](#_Toc70017186)

[Real-Time Queries 16](#_Toc70017187)

[Adaptive Element 16](#_Toc70017188)

[Outcome Accuracy 16](#_Toc70017189)

[Security Measures 17](#_Toc70017190)

[Product Health Monitoring 17](#_Toc70017191)

[DashBoard 17](#_Toc70017192)

[Section D 17](#_Toc70017193)

[Project purpose 17](#_Toc70017194)

[Dataset 17](#_Toc70017195)

[Data Product Code 18](#_Toc70017196)

[Hypothesis Verification 18](#_Toc70017197)

[Effective Visualizations and Reporting 18](#_Toc70017198)

[Accuracy Analysis 19](#_Toc70017199)

[Application Testing 19](#_Toc70017200)

[Application Files 19](#_Toc70017201)

[User’s Guide 20](#_Toc70017202)

[Summation of Learning Experience 21](#_Toc70017203)

[Section E 22](#_Toc70017204)

[Appendix A 23](#_Toc70017205)

[Score Frequency Analysis Example 23](#_Toc70017206)

[Studio Frequency Analysis 23](#_Toc70017207)

[Main Application Screen 24](#_Toc70017208)

[Recommendation Screen 24](#_Toc70017209)

[Correlation Matrix 24](#_Toc70017210)

[Genre Bar graph 25](#_Toc70017211)

[Ratings Bar Graph 25](#_Toc70017212)

[Appendix B 26](#_Toc70017213)

[Original Columns 26](#_Toc70017214)

[Final Columns 26](#_Toc70017215)

[Duration Outliers 27](#_Toc70017216)

# Section A

## Letter of Transmittal

See above ([Letter of Transmittal – Anime Recommendation Application (Metis)](#_Letter_of_Transmittal))

## Project Recommendation

### Problem Summary

My AnimeRoll is struggling to maintain their viewership despite having strong metrics surrounding new viewers. Currently, when a customer finishes watching a show, there is no way for them to get suggestions on what to watch next. Metis will provide the ability for a customer to request recommendations of shows that are like a title of their choosing. Metis will be designed and executed as a standalone computer application, with the intent that My AnimeRoll will incorporate it into their structure as they see fit post-delivery. Due to data information security concerns, Metis will not be making recommendations based on other user input, nor will it be storing any new information from users.

### Application Benefits

Metis will provide a way for My AnimeRoll to answer the “Now what?” question that users often ask after completing an anime. By answering this question with recommendations on what to watch next, Metis will help My AnimeRoll maintain their viewers’ attention, thereby increasing their profit margin.

### Application Description

The application will include an alphabetized list of all 4,783 shows (rated PG-13 and below) available in the My AnimeRoll archives. Users will be able to scroll down to their chosen title or use the convenient filter bar to narrow down the list so that it is more manageable. Once the customer has selected a show by clicking on it, they will need to hit the “Recommend” button and a box will pop up with the Top 5 recommendations for them based upon that show.

Metis will have 4 additional buttons to those described above. One will be to exit the application. The other 3 will allow the program user to view the graphs that will be created for this project: a correlation graph of all elements that are not related to genre, a bar graph of anime count by genre, and a similar bar graph showing anime count by rating.

The intent is to create as clean and streamlined of an application as possible to optimize the user experience and allow them to return to watching shows as quickly as possible.

### Data Description

The dataset will come from Kaggle (Račinský 2018) since this list incorporates most of the shows that My AnimeRoll uses on their platform. There will be several columns that Metis will use to make its predictions after preprocessing the data. Most of the columns will hold binary values relating to a title’s genres. There will also be two categorical columns, type, and rating, that have numbers that are associated with other values. (i.e., a rating of 0 indicates that a show is PG-13 - Teens 13 or older) The title will be the only non-numeric column, since it has alphanumeric values, however towards the end of the process that will be converted to numerical values as well for use by the machine learning model.

There are several anomalies in the dataset that will be handled during preprocessing. First, the genre category occupies one column even in instances where multiple genres apply to one title. To ensure the data is consistent, each genre will be given a separate column. For each title, the applicable genres will be marked with a 1 while the remaining genres will be marked with a zero. Additionally, there are also several outliers when it comes to scoring. Due to My AnimeRoll’s high standard of excellence for shows, the outliers at the lower end of the spectrum will be removed. Finally, to ensure the application and My AnimeRoll remain family friendly, all shows above the PG-13 rating will be removed as well.

The largest limitation to the data is that the dataset has not been updated since 2018 despite new titles having been released since then. This won’t be an issue for My AnimeRoll at this time as, per our conversation, the platform focuses mostly on older shows. Adding new shows is time consuming and must be done manually. The ability to add shows in the future can be discussed, however it is currently outside the scope of Metis.

### Objective and Hypothesis

The main objective of Metis is to allow users to choose an anime they recently watched and enjoyed, and return recommendations of similar shows for them to watch next. If the user chooses a valid title, then five recommendations should be returned to them. The user should not be able to submit an invalid title to the application, nor should they be able to select more than one title at a time. If the user selects one of the three buttons related to graphs, then the corresponding graph should open in a separate window and populate appropriately.

### Methodology

The creation of Metis will follow the Agile methodology. Agile has 5 steps: Plan, Design and Develop, Testing and Quality Assurance, Deliver, and Assess. These steps are completed in sprints of 2 weeks to create products rapidly and efficiently. The Agile methodology is appropriate for this project given the short turnaround time, and the potential for adding new features later.

The planning stage is taking place here with the proposal and conversations that My AnimeRoll has had with the Metis team surrounding the needs and requirements of their business. The design and develop phase will take place next with the creation of the user interface (UI) mock-ups, planning the flow of the application and choosing an applicable machine learning model. Development will include the actual creation of the machine learning model in Jupyter Notebook, the creation of the UI in QT Designer, and the finalization of the program including making the UI functional via PyCharm.

Following this stage will be extensive testing. The Metis team will ensure all the buttons and features work as expected, that there are no spelling errors, and generally polishing the application. The application will then be delivered to My AnimeRoll for deployment to their users. Finally, the Metis team would appreciate any feedback that My AnimeRoll has to offer in the post-implementation review, including ideas of what should come next for the application.

### Funding Requirements

As stated in the “Letter of Transmittal” above (p.1), the total funding requirements for this project are $6,000. The project does not require any specialized software, or hardware, nor will it require specialized training to use. Our goal at Metis is to provide a simple easy to use product that will provide a great return on investment to My AnimeRoll.

### Stakeholder Impact

As the primary stakeholder in the project, My AnimeRoll should expect to see an increase in returning customer traffic due to the practicality of the Metis application. They should also see an increase in their customer satisfaction as they discover new shows to watch that are like titles that they already enjoy.

### Data Precautions

The data that will be used for Metis will have no sensitive or protected Personal Identifiable Information (PII); no payments will be taken, no usernames will be required, and all shows that rate above PG-13 will be removed.

Though there is no legal requirement to follow the MPAA movie rating system, given the ethical concerns with children watching shows that are potentially damaging to their psyches, it is important to the Metis team that we ensure that all shows are age appropriate. There have been several instances where legal action was taken against persons who have not abided by the ratings in a public setting. In Ohio State vs Kearns, a substitute Spanish teacher was jailed for showing a rated R movie in their classroom. There were several instances in California of parents raising concerns over the showing of content they felt was inappropriate for their children in classrooms. Additionally, YouTube has also had to enforce stricter controls over channels that they identify as “kids TV” to ensure that children are not getting suggestions for shows that are not age appropriate. The best route for Metis at this time to make these same assurances will be to simply not include them in the application.

### Developer Experience

My AnimeRoll is a strong player in anime streaming services, that has a phenomenal model for attracting new viewers. With 8 years working in a multi-level retail management environment, and 2 years of experience working with software, I can both build the Metis software, and communicate with all levels of leadership and associates that will be part of the project team from both of our organizations, both technical and non-technical. Together, My AnimeRoll and Metis will be able to improve My AnimeRoll’s viewer retention and become an industry leader.

# Section B

## Problem Statement

Metis will be designed to aid My AnimeRoll in maintaining their viewership by recommending shows that are like one that they select. Currently, My AnimeRoll does not offer this ability with their platform, so Metis will be used alongside it as a standalone application for Windows 10. The selections for shows will be made using an unsupervised item-based K-Nearest Neighbors collaborative filtering model. The model will focus on similarities between movies, rather than similarities between users in the interest of maintaining user privacy.

## Customer Summary

The customer base for My AnimeRoll and Metis are like those of their competitor, CrunchyRoll another anime streaming service. 75% of the customers are under the age of 35 with a median age of 18. (Smith 2021) Due to the repeated quarantines that have occurred due to COVID-19, My AnimeRoll’s viewer range may be broader now than it was originally as people have hunted for ways to spend their time at home.

Additionally, 90% of the users would consider themselves gamers (Smith 2021), which would imply that minimal training should be needed for them to comfortably use Metis. The application will be designed to run in Windows 10 and feature a streamlined UI with minimal interaction needed from the customer once it is installed. As such, no special skill sets should be necessary.

## Existing System Analysis

My AnimeRoll is a computer-based application for streaming anime. The My AnimeRoll information is stored in a database, with basic information about the shows regularly exported to a .csv file for analytical purposes. It is this .csv file that Metis will be based upon. Metis will run in the same environment as My AnimeRoll, Windows 10., and should be able to integrate smoothly.

## Data

The data for this project will be from the aforementioned .csv file, pulled from Kaggle (Račinský 2018). The original file available in the database has roughly 14,000 entries while the cleaned file has 6668. There is no mention in the dataset of the cleaning that was done to the anime file specifically. There will also be some more cleaning that will take place to help with the functionality of Metis.

Many of the columns will be dropped because they will not be instrumental in making a recommendation. In total, 8 of the original 32 columns will be left untouched. The genre column currently represents multiple genres in each cell. As mentioned above, each genre will be given its own column to allow for more accurate comparisons between titles. This will change the genre column from 1 to 41 columns. For each title, applicable genres will be indicated with a 1 while the rest will be filled with 0.

Additionally, there are several outliers when it comes to scoring. Due to My AnimeRoll’s high standard of excellence for shows, the outliers at the lower end of the spectrum will be removed. Finally, to ensure the application and My AnimeRoll remain family friendly, all shows above the PG-13 rating will be removed as well.

A Jupyter notebook will be available with the application so that all changes can be reviewed and repeated if necessary.

## Project Methodology

Metis will follow the Agile methodology for its creation. Currently, Metis and My AnimeRoll are in the planning stage. Once we have finalized the needs and requirements to ensure the application is as useful as possible for My AnimeRoll, the design phase can begin. In this stage, the UI will be designed, the flow of the application will be laid out, and an applicable machine learning model will be chosen. Additionally, the Metis team will install the necessary software to our systems for the creation of the application.

Third, Metis will be created. The machine learning model and cleaning of the data will take place in Jupyter notebooks. Several plots will be used to determine the usefulness – or lack thereof – of the features available in the original cleaned dataset. The reasoning behind keeping or dropping a column should be indicated within the notebook as well. Concurrently, the UI for the application will be built using QT Designer. After both components are completed, the model and UI will be brought together in PyCharm where the final functionality will be built.

Following the third stage will be copious amounts of testing. Each function will be tested individually prior to ensuring it functions with the rest of the application. Once each feature is working properly, integration testing will begin to ensure all the features work properly together. This part of the process should be quick compared to the unit testing since most of the features are independent from each other. The application will be tested to ensure it will function properly with My AnimeRoll’s current system, and then My AnimeRoll will be able to complete their acceptance testing.

Finally, after the application passes My AnimeRoll’s acceptance testing, it will be officially delivered. Metis will assess the success of the project and will welcome any input that My AnimeRoll has to offer for the future.

## Project Outcomes

### Project Deliverables

1. A Jupyter Notebook: This will be where the model used in the application is created. It will also hold notes and plots relating to changes made to the dataset.
2. Two .csv files: The first is the file from the Kaggle dataset. The second will be created from the Jupyter notebooks file. It will contain the scrubbed data prior to the final transformation for the model.
3. Pickle file: This will house the model created in the Jupyter Notebook. It will be imported into the application itself.
4. QtDesigner Files: These will house the UI files for the application. There will be two .ui files here, one for each screen that will eventually be incorporated into the main.py file in PyCharm. The original mock-ups will take place on a whiteboard and therefore will be unavailable.

### Product Deliverables

1. PyCharm file: This will be where all the pieces come together into a cohesive, useable application.
   1. 3 visualizations of the final dataset: A correlation graph to show the relationship between the non-genre related data. A bar graph by genre to show the relationship between the anime titles and their genres. (i.e., which are the most popular) The third and final visualization will also be a bar graph of the shows by rating.
   2. User Dashboard: The main .ui file which allows for the user to click on any of the buttons and display the appropriate graph, recommendation, or exit the program.

## Implementation Plan

### Environment Set-Up

My AnimeRoll will want to ensure that they have all the necessary libraries installed so that they can run the Jupyter Notebook and PyCharm File smoothly. Metis plans to accomplish this via Anaconda 3 and recommends that My AnimeRoll do the same. By installing Anaconda, all the libraries that Metis uses should be available. The program will use Python 3.8 and PyQt5 as well as NumPy, Pandas, SciKit-Learn, MatPlotLib, Seaborn, SciPy, and Pickle. By this stage, the UI will already be designed in QT Designer, so the My AnimeRoll team should not need to install it if they have PyQt5. To run the application, Jupyter Notebook will not be required, however, to see how the model was created My AnimeRoll will want Jupyter Notebook installed.

### Application Testing

The testing of the application will be mostly complete by this point and will only need to involve installing the program with the file structure intact, and then clicking each button to ensure that the expected output still occurs.

### Deployment

Given the lack of dependencies outside of the application (i.e., no internet or access to other applications needed), it should be possible to deploy quickly even as a phased roll-out. During the Application Testing phase of the Implementation Plan, they will want to begin letting users know of the upcoming recommendation features and make the User Manual/Screenshots available.

Metis should first roll-out to users with the highest view counts as a “Loyalty Reward”, these users are less likely to leave the platform if there are issues and should also appreciate being able to be Beta testers. Once any major issues that these users uncover are handled, or two weeks pass with no major issues, the roll-out should continue to all other users. Any bug reporting will be handled through the My AnimeRoll team.

## Evaluation Plan

The evaluation process will consist of two phases, one to be completed by Metis and one to be completed and tracked by My AnimeRoll. Due to the application’s design, there should be minimal security or liability concerns. The application will not include any PII, or regulatory standards that need to be met because of the lack of content that is above a PG-13 rating.

Metis will include a section at the bottom of the Jupyter notebook where the model will be tested, and the accuracy determined. After contemplating several different ideas, it has been decided that accuracy will be tested using the kneighbors\_graph functionality from SciKit-Learn. By determining the shortest distance between all elements in the DataFrame, a baseline can be established for best performance of the model. Each distance returned from the kneighbors model will be measured against this distance to determine the accuracy of the recommendation. The closer it is to the minimum value, the higher the accuracy.

My AnimeRoll will judge the efficiency of the application by checking their viewer counts. Given how their system is established, they should also be able to see which shows are recommended to viewers, and whether the users go on to watch any of the recommended titles. If My AnimeRoll sees the expected increase in returning viewers, as well as viewers using the Metis software and going on to watch the recommended shows, the application will be considered successful. This is outside of the scope of Metis and must be handled by My AnimeRoll.

## Resources and Costs

### Programming Environment

Hardware and Software Complete List:

* MSI GS65 Stealth (and power)
* External Mouse
* Internet access
* Windows 10
* Anaconda 3
  + Jupyter Notebook
  + PyCharm Community Edition
  + Qt Designer
  + Python 3.8
  + PyQt5
  + Several libraries: NumPy, MatPlotLib, SciKit-Learn, Pandas, SciPy, Pickle
* Kaggle Dataset

This project will be created on an MSI GS65 Stealth Laptop running Windows 10. The computers that will run this application do NOT need to have the same specifications, nor will they need much space for the application to run. The creation of Metis will require at least 5 free Gigabytes of space for the software that will be used to build it, as well as the dataset from Kaggle.

### Environment Costs

Due to the small development team, the community edition for all software should be able to be used which will result in no licensure costs. Metis will not require any specialized software or environments to run, thereby incurring no additional costs outside of the Human Resource Requirements.

### Human Resource Requirements

The time and cost to develop the application as outlined above will be $6,000 for 120 hours of labor. The delivery date will be April 30.

### Timeline and Milestones

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start DAte | End Date | Milestone | Resources | Dependencies |
| 4/01/2021 | 4/30/2021 | Project Overall Timeline |  |  |
| 4/01/2021 | 4/03/2021 | Research |  |  |
| 4/03/2021 | 4/07/2021 | Project Approval | Task 1 Approval Form |  |
| 4/07/2021 | 4/10/2021 | Machine Learning Udemy Course |  |  |
| 4/11/2021 | 4/16/2021 | Machine Learning Creation and Data Cleaning | Kaggle Dataset, Anaconda 3 |  |
| 4/11/2021 | 4/16/2021 | UI Design | Qt Designer |  |
| 4/16/2021 | 4/18/2021 | Application Creation | Anaconda 3, Qt Designer | Machine Learning Model, UI Files |
| 4/19/2021 | 4/22/2021 | Documentation | Access to all Previous Resources, Microsoft Word | Completed Application |
| 4/23/2021 | 4/30/2021 | Project Submission and Revisions |  | Completed Application, Completed Documentation |

# Section C

## Data Methods

One of the main descriptive methods used to determine whether a row should be eliminated from the data set is Frequency Distribution. (see [Score Frequency Analysis Example](#_Statistic_Analysis_Example)) Data that fell below 3 standard deviations was eliminated allowing the data to be in a balanced bell curve. Though the visualization was not created for each instance, there were several instances where it was used. The [Score Frequency Analysis Example](#_Statistic_Analysis_Example) below was used to determine that the scores were out of balance. [Studio Statistical Analysis](#_Studio_Statistical_Analysis) is another example of frequency distribution without the graphical visualization. Ultimately the Scores were deemed essential, and the outlier rows were dropped. While the studio was deemed non-essential, and the whole column was dropped. Additionally, I used correlation graphs to eliminate columns that were too like each other to lessen the number of features that would have to be accounted for in the KNN model. (see [Correlation Matrix](#_Correlation_Matrix))

The non-descriptive method that was used to create the model is an unsupervised K Nearest Neighbors Model which found the associations between all columns, and the “distances” between them. Because there is no “right” answer when it comes to recommendations, an unsupervised model had to be chosen. By using a model that looked at the Nearest Neighbors, Metis can choose titles that are like each other, primarily based upon their genres.

## Data SetS

The dataset used in this project was available on Kaggle, called the My AnimeList Dataset by Matěj Račinský. Račinský compiled the information from the My AnimeList website, which is publicly available. Metis begins with the anime\_cleaned.csv file, and, after running the Jupyter Notebook [anime-recommendation-KNN-Model.ipynb](anime-recommendation-application/anime-recommendation-KNN-Model.ipynb) file, returns [anime\_list\_scrubbed.csv](anime-recommendation-application/MyAnimeList%20Dataset/anime_list_scrubbed.csv) which is further used in the notebook as well as the PyCharm application. Both datasets can be found in the “MyAnimeList Dataset” folder.

There was one main disadvantage to this dataset that required an entire restructuring of the project before it began. Originally, two files from the dataset were going to be used, anime\_cleaned.csv and animelists\_cleaned.csv. Unfortunately, the animelists\_cleaned.csv file is upwards of 2G on its own. After cleaning and eliminating more than half of the columns, it remained above 1G and there were several operations that simply took too long to run. This resulted in using only the anime\_cleaned.csv file rather than both as originally planned. While the recommendations are still fairly accurate, the approach had to be entirely different.

## Analytics

As mentioned above, this application forms decisions on recommendations using an unsupervised KNN model. Determining the accuracy was difficult due to the unsupervised nature, there are no “wrong” answers when giving movie recommendations. To give some idea of how accurate the predictions are, however, I found that applying the kneighbors\_graph to the fit and transformed model returned a list of distances between all the neighbors for each point. By removing all zeros (instances where an item matched with itself), I found the minimum value for the distances, and used it as a baseline target for all other distances. The accuracy is then determined by the percentage of deviation from this point.

The accuracy measure is found at the bottom of the Jupyter Notebook [anime-recommendation-KNN-Model.ipynb](anime-recommendation-application/anime-recommendation-KNN-Model.ipynb) file.

## Data Cleaning

The data cleaning all took place in Jupyter Notebooks, [anime-recommendation-KNN-Model.ipynb](anime-recommendation-application/anime-recommendation-KNN-Model.ipynb), and is available for review. By clicking Kernel -> Restart and Run All, the data can be cleaned again if needed and all the necessary files will be output. A small amount of cleaning also takes place in [main.py](anime-recommendation-application/main.py) in the prepData() function.

## DatA Visualization

The three visualizations used on the user dashboard are [Correlation Matrix](#_Correlation_Matrix), [Genre Bar Graph](#_Genre_Bargraph) and [Ratings Bar Graph](#_Ratings_BarGraph). Each has its own button on the dashboard and will pop-up as shown in the linked images.

## Real-Time Queries

Due the additional data preparation that takes place when the application is opened, it does take a couple seconds to load up. Once everything is loaded, users can complete real-time queries in the main dashboard using the filter feature and by clicking their chosen title and then “Recommendations” the results appear in a second or less. (see [Main Application Screen](#_MaIn_Application_Screen))

## Adaptive Element

Adaptive Elements do not apply for Metis; no user information or feedback is received or stored.

## Outcome Accuracy

As mentioned in the [Analytics](#_Analytics) section above, determining the accuracy was difficult due to the unsupervised nature, there are no “wrong” answers when giving movie recommendations. To give some idea of how accurate the predictions are, however, I found that applying the kneighbors\_graph to the fit and transformed model returned a list of distances between all the neighbors for each point. By removing all zeros (instances where an item matched with itself), I found the minimum value for the distances, and used it as a baseline target for all other distances. The accuracy is then determined by the percentage of deviation from this point.

The way the model generates the 5 recommendations ensures that there will always be 5, even if the accuracy is lower, so no additional precautions needed to be taken to ensure the output would be valid.

## Security Measures

No security measures were needed for Metis. No PII or other sensitive data is present in the application. Additionally, all shows rated above PG-13 are removed so there should be neither ethical nor legal concerns from children viewing shows that are damaging to their psyche. See [Data Precautions](#_Data_Precautions) for more details.

## Product Health Monitoring

No changes are intended to be made to the Metis application. As such, minimal Product Health monitoring is necessary. A title will always be selected when opening the application, even when a filter is applied, so no errors should occur from the model receiving an empty data point.

Two files are loaded in however, the [anime\_list\_scrubbed.csv](anime-recommendation-application/MyAnimeList%20Dataset/anime_list_scrubbed.csv) and [movie\_info\_model.pkl](anime-recommendation-application/movie_info_model.pkl). Both paths will be verified before loading, if there are any issues with the path, an error will be thrown to the console and the application will be halted.

## DashBoard

Metis is a standalone application with one main dashboard ([Main Application Screen](#_MaIn_Application_Screen)), 3 buttons for graphs ([Correlation Matrix](#_Correlation_Matrix), [Genre Bar Graph](#_Genre_Bargraph), [Ratings Bar Graph](#_Ratings_BarGraph)) , 1 button to get recommendations ([Recommendation Screen](#_Recommendation_Screen)) and an exit button. Each button, aside from Exit, creates a pop-up window with the information the user requested.

# Section D

## Project purpose

Metis was created to provide 5 recommendations of anime titles to users based on their selection in the alphabetized window of the main screen. The application also provides 3 visualizations of the data contained within that may be useful to My AnimeRoll, the fictitious client for the software. My AnimeRoll is using the software to increase their viewer retention by providing them with these recommendations, which is something they did not offer originally in their anime streaming service.

## Dataset

The dataset was cleaned extensively, and a record of all the changes made and the reasoning are available via the [Jupyter Notebook](anime-recommendation-application/anime-recommendation-KNN-Model.ipynb) that accompanies this project. Approximately 25 columns were removed from the [original](anime-recommendation-application/MyAnimeList%20Dataset/anime_cleaned.csv) dataset for various reasons including: similarities to other columns, unnecessary to the model (ex. opening\_them and closing\_theme), columns with more than 50% null values, the information was sparse as with the studio data. The decision was then made to expand the genre column into its component parts to allow for better analyzation from the model, leaving the dataset with 47 columns upon completion. (The anime\_id column is used several times throughout the notebook for analyzation purposes but is dropped when the data undergoes its final transformation) See the [initial](#_Original_Columns) and [final](#_Final_Columns) column lists below.

Several changes were made to rows as well; many were dropped for NaN values, some were dropped for being outliers (See [code sample](#_Duration_Outliers) below), and several were dropped for being above a PG-13 rating. They type and rating were both converted to numerical values. Types: {0: “TV”, 1: “OVA”, 2: “Special”, 3: “ONA”, 4: “Movie”}. Ratings: {0: “PG-13 - Teens 13 or older”, 1: “PG – Children”, 2: “G - All Ages”} Finally, the entire file was sorted alphabetically so that it would appear properly in the main window.

## Data Product Code

There were two sets of code used for this project, one was a [Jupyter Notebook](anime-recommendation-application/anime-recommendation-KNN-Model.ipynb) and one was created in [PyCharm](anime-recommendation-application/main.py). The Jupyter Notebook was used for the data analysis and cleaning of the dataset due to its ease of use for visualizations and data manipulation. Most of the descriptive and predictive outputs were created in Jupyter Notebook prior to being imported into PyCharm. At this point, the data cleaning still takes place in Jupyter Notebook, however the file should be able to be converted to a .py file if desired.

PyCharm was used for easier integration with Qt Designer, the application used for creating the UI. I also had more experience with PyCharm, and JetBrains applications, making it more comfortable for me to use.

Between these IDE’s, the code that was created is fully functional, and can be reviewed in the <anime-recommendation-application> folder.

## Hypothesis Verification

There were 2 hypotheses made in Section A, [Objective and Hypothesis](#_Objective_and_Hypothesis). Both were proven via Metis’ creation. First, if the user chooses a valid title, then 5 recommendations should be returned to them. The user should not be able to submit an invalid title to the application, nor should they be able to select more than one title at a time. These are all accomplished by the application. The first is handled by the Recommendations button. The conditions are fulfilled by not allowing the user to enter titles outside of the filter box and automatically selecting the first option on a list when it populates. The average accuracy of the recommendation is 83%.

Second, if the user selects one of the three buttons related to graphs, then the corresponding graph should open in a separate window and populate appropriately. This is also demonstrable when running the application, each of the graphs populates according to the button that the user presses as expected.

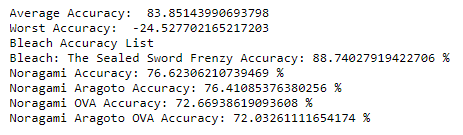
## Effective Visualizations and Reporting

Several different visualizations were used throughout the creation of the application and the Jupyter Notebook. Three are included in the application itself because they are the most likely to be valuable resources for the user. The [Correlation Matrix](#_Correlation_Matrix) shows the relationship between all non-genre related elements. The [Genre Bar Graph](#_Genre_Bargraph) shows the relationship between all the titles and their respective genres. This allows the user to see which genres are likely to get better recommendations. Finally, the [Ratings Bar Graph](#_Ratings_BarGraph) shows all the titles by their ratings (G, PG and PG-13). This one was included as verification that there are no ratings above PG-13 included in the application.

Jupyter Notebook has quite a few more data visualizations to aid in making decisions regarding which data to keep and which to drop. Each visualization and decision that was made in the Jupyter Notebook file has been detailed for review if necessary.

## Accuracy Analysis

As mentioned in the [Analytics](#_Analytics) and [Outcome Accuracy](#_Outcome_Accuracy) sections above, determining the accuracy was difficult due to the unsupervised nature of the model. Ultimately, the accuracy was determined by finding the minimum distance between all the points and using it as a target. The accuracy of each recommendation can then be calculated based on its percentage of deviation from this point. The average accuracy of the model is 83%, while the worst accuracy is -24%. As expected, the first of the 5 recommendations were always the most accurate.



## Application Testing

There were numerous levels of testing involved in this project, mainly once the application as created in PyCharm. Unit testing took place as each function was completed to ensure that it was behaving as expected. Once this was established, the function was integrated with the rest of the application and tested to ensure it was still functioning as well. Regression testing took place in several places as well, changes were made to the Jupyter file, and so the PyCharm file needed to be validated as well. In some cases, remediation was needed due to the changes in one or both files. For example, a feature was deleted from the Jupyter Notebook file at the last minute, which threw off the alignment needed for the PyCharm file. It is now addressed in the prepData() function, however initially it did stop the program from running.

## Application Files

Several files are involved with Metis. They should be left in the hierarchy in which they are received to ensure everything works properly. The files will be zipped as is for submission.

1. Capstone\_ProjectFolder: this is the folder that houses the application, as well as this document. It has no dependencies
   1. anime-recommendation-application: The application folder, 00469842 Carpio C951 Task 2.docx is dependent upon this for many of the links to function.
      1. .idea: Houses JetBrains project files.
      2. .ipynb\_checkpoints: Houses save points from the Jupyter Notebook
      3. MyAnimeList Dataset: Houses the .csv files that the programs need to run
         1. Anime\_cleaned.csv: Original dataset downloaded from Kaggle
         2. Anime\_list\_scrubbed.csv: Dataset post pre-processing in Jupyter Notebooks. Can be regenerated from JN if needed.
      4. UI Files: Has the initial UI files before they were incorporated into main.py. These have no interdependencies. The .ui files were both created in Qt Designer and converted to .py files in Anaconda. The .py files were copied and pasted into main.py.
         1. main\_window.py
         2. main\_window.ui
         3. recommendation\_window.py
         4. recommendation\_window.ui
      5. Anime-recommendation-KNN-Model.ipynb: The Jupyter Notebook file used for data analysis. Dependent on MyAnimeList Dataset
      6. main.py: main application file dependent upon MyAnimeList Dataset file and movie\_info\_model.pkl
      7. movie\_info\_model.pkl: Pickle file with the KNN model, can be regenerate from anime-recommendation-KNN-Model.ipynb if needed.
   2. 00469842 Carpio C951 Task 2.docx: The task write-up document. It has several dependencies so that the links will function properly

## User’s Guide

Metis, the anime recommendation application, is quite easy to use once everything is set up. The first step is to ensure that the computer that will be running the program has Anaconda 3 installed. (<https://www.anaconda.com/products/individual>) The user will need to download and unzip the project file, leaving the structure intact. The Capstone\_ProjectFolder can be nested into other folders, however its internal structure cannot change.

To view the. ipynb file, open Jupyter Notebook from the Anaconda Navigator. This will open a window in the user’s browser showing their file structure. Navigate to where the file is saved and click on it to open. To run the file, click Kernel->Restart and Run All. The file should be up to date and ready to explore. When the user is done viewing the file, they should click File->Close and Halt. Jupyter Notebook will notify them when it’s safe to close the window. To exit Jupyter Notebooks, click Quit on the main screen and then close the window.

To run the application, open PyCharm Community from the Anaconda Navigator. In PyCharm, go to File -> Open and then navigate to the Capstone\_ProjectFolder. Press Alt+Shift+F10 to run the application. To use the filter, type in all or part of the title the user is looking for and then click filter. Highlight, by clicking, the chosen title and then press “Recommendations.” This should result in a pop-up window with 5 recommendations appearing. Press the X to close the window.

To view the graphs, click any of the three upper buttons at the bottom of the screen and the associated graph should appear. As before, click the X to close the window.

To exit the application, click Exit or the X.

## Summation of Learning Experience

Outside of the classes I took through WGU, I have had no real-world prior experience with programming outside of a little bit of freelance web designing. Unfortunately, web design was of little help to me given how I decided to approach this task. I did base part of the design and implementation of the UI on the C195: Software II Advanced Java Concepts coursework. The differences between Qt Designer and SceneBuilder meant that I had to make quite a few changes to my expectations, but I did feel that I had some idea of how to approach the problem initially. The best skill I learned throughout the program that was of use to me in the Capstone was how to research things for myself. Though I picked up quite a bit of useful information from the degree program, being able to confidently use available resources to find answers to the inevitable questions that arise in any industry will be an invaluable asset.

I purchased the Udemy Course “Complete Machine Learning and Data Science Bootcamp” which was extremely helpful in understanding Jupyter Notebooks and Machine Learning. Prior to viewing the course, I did not feel that I had a solid grounding in machine learning. I also sought quite a bit of help from non-computer sources. My family was instrumental as emotional support, grading my papers, and with building my skill in explaining technical problems in a more simplified manner for people that are not trained in computer science. When I was stuck on a concept or when something in this project wasn’t working properly; breaking it down into smaller parts and explaining it to someone else often helped identify where the problem was.

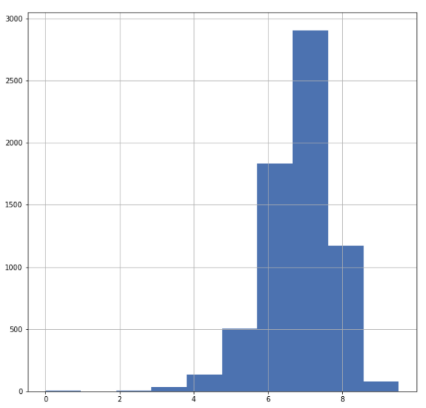
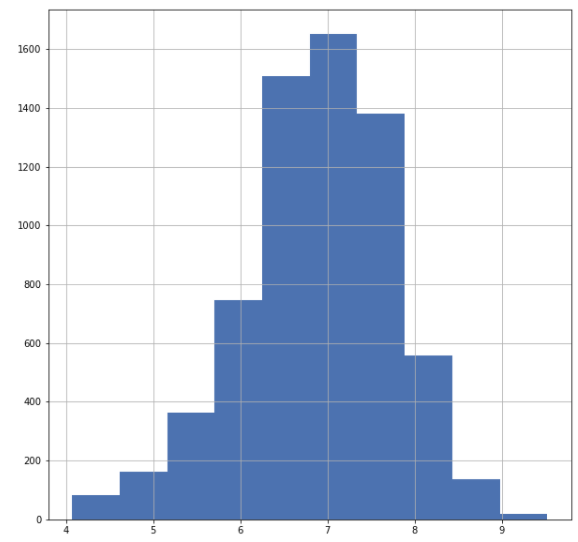
As much as I have honestly *loathed* parts of this experience when they weren’t working properly - when I had to rebuild the model from scratch for the 10th time because I forgot something important, or when I saw something online that completely contradicted how I was doing things even though my way worked - the excitement over finishing the project and seeing everything come together the way I had envisioned it made all the frustration and loathing worth it. I am looking forward to continuing my journey of probable frustration and hopeful reward in the future as I begin job hunting. And who knows, maybe a master’s degree in Computer Science is in my future some day!

# Section E

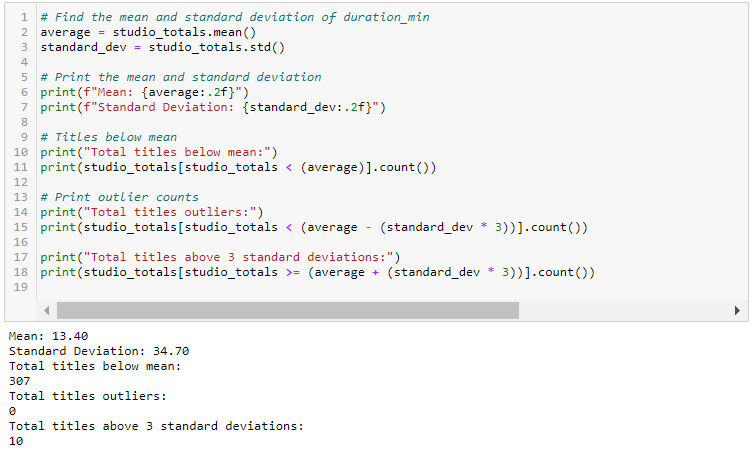
* Matěj Račinský. (2018). MyAnimeList Dataset [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/45582>
* Smith, C. (2021, March 23). *CrunchyRoll Facts and Statistics (2021) | By the Numbers*. DMR. https://expandedramblings.com/index.php/crunchyroll-facts-statistics/.

# Appendix A

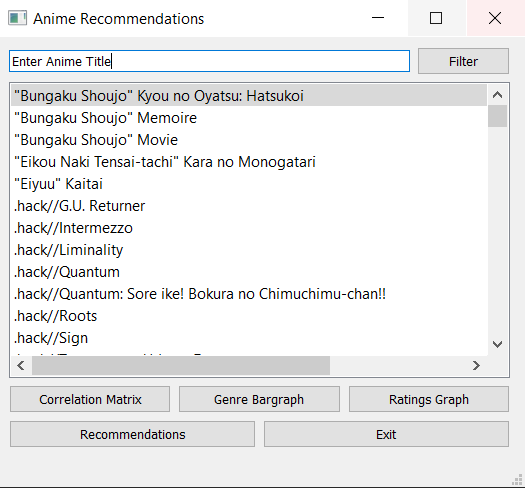
## Score Frequency Analysis Example

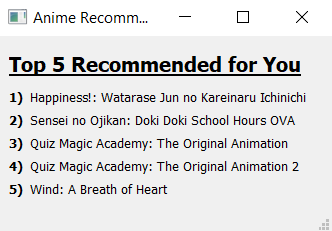
## Studio Frequency Analysis



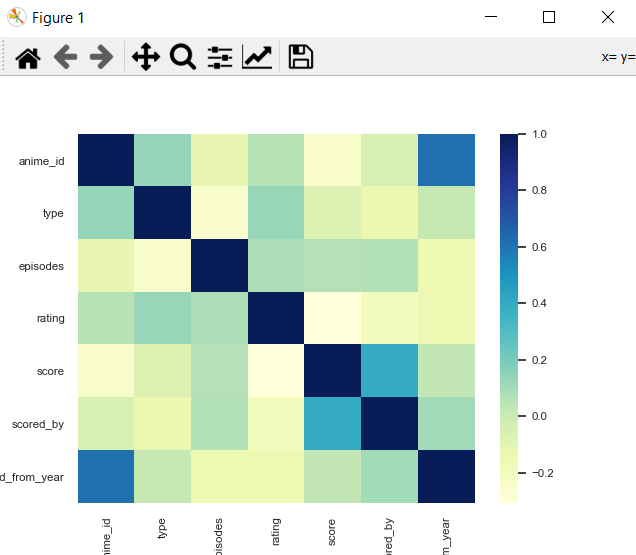
## Main Application Screen



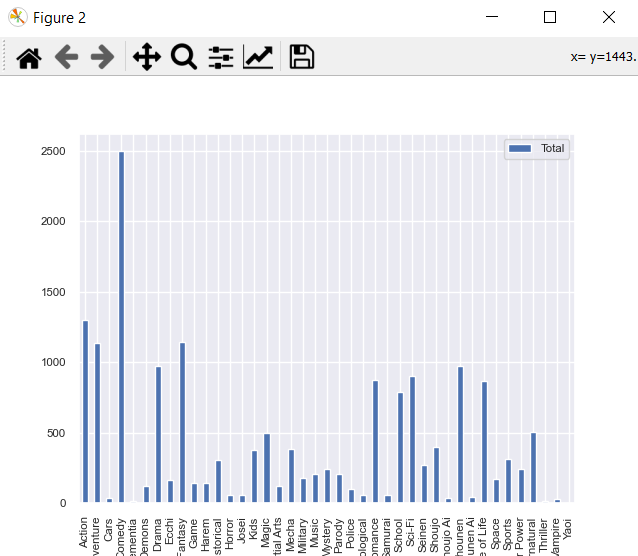
## Recommendation Screen



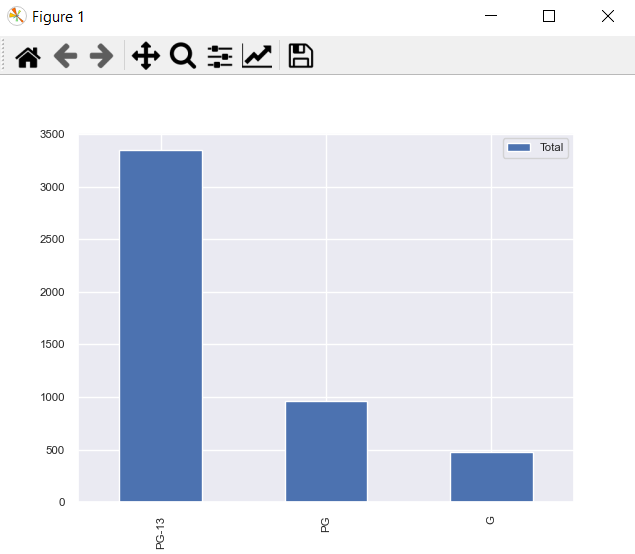
## Correlation Matrix



## Genre Bar graph

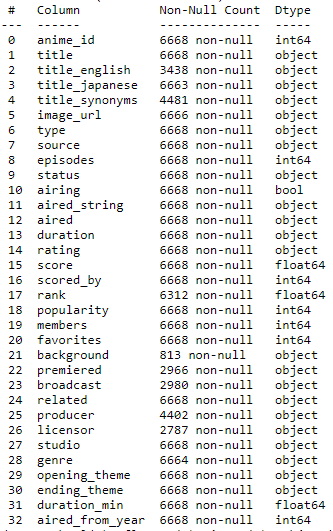


## Ratings Bar Graph

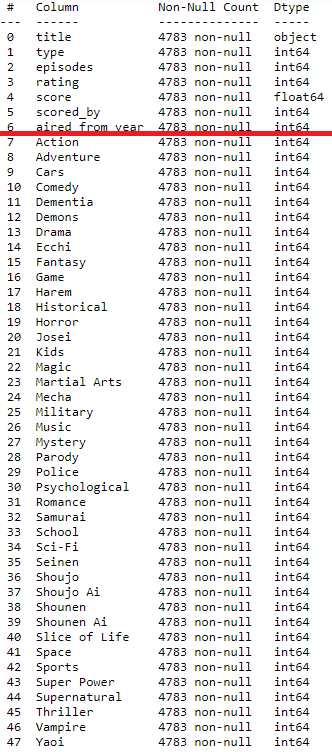


# Appendix B

## Original Columns



## Final Columns

The red line denotes the point between the remaining original columns, and the genres.

## Duration Outliers

