WGU C964

Capstone

Video Game Recommendation System

Brad Pappan

002656581

12/24/2022

**Contents**

**Project recommendation (Section A)**............................................................................................ 3

1. Problem Summary........................................................................................................... 3

2. Application Benefits........................................................................................................ 3

3. Application Outline…...................................................................................................... 3

4. Data Description.............................................................................................................. 4

5. Objective and Hypotheses............................................................................................... 4

6. Methodology ................................................................................................................... 4

7. Funding Requirements .................................................................................................... 5

8. Stakeholder Impact ......................................................................................................... 5

9. Data Precautions.............................................................................................................. 5

10. Developer Expertise....................................................................................................... 5 **Technical Proposal (Section B)** ..................................................................................................... 6

1. Problem Statement........................................................................................................... 6

2. Customer Summary ........................................................................................................ 6

3. Existing System Analysis................................................................................................ 6

4. Data.................................................................................................................................. 6

5. Project Methodology........................................................................................................ 7

6. Project Outcomes............................................................................................................. 7

7. Implementation Plan........................................................................................................ 8

8. Evaluation Plan................................................................................................................ 8

9. Resources and Costs ........................................................................................................ 8

10. Timeline and Milestones................................................................................................ 9

**Post Implementation Report (Section D)**..................................................................................... 10

1. Project Purpose ............................................................................................................... 10

2. Datasets ........................................................................................................................... 10

3. Data Product Code........................................................................................................... 11

4. Hypothesis Verification................................................................................................... 13

5. Effective Visualization and Reporting ............................................................................ 13

6. Accuracy Analysis .......................................................................................................... 16

7. Application Testing.......................................................................................................... 16

8. Application Files.............................................................................................................. 17

9. User’s Guide..................................................................................................................... 18

10. Summation of Learning Experience .............................................................................. 18

**PROJECT RECOMMENDATION (SECTION A)**

**1. Problem Summary**

WildCard Games has noticed a trend that many users are only purchasing one game. Senior members of the company believe this is due to users not being able to find new games that may interest them. The main problem is that there are video games that receive little purchases that once users know exist, they will purchase. We have been tasked to come up with a machine learning solution to recommend different games to users to attempt to increase sales and user satisfaction.

**2. Application Benefits**

The video game recommendation system would allow users to engage with it to get recommendations on new games that they may want to buy based on critic’s reviews. This will allow sales to increase and increase user satisfaction with WildCard Games platform.

**3. Application Outline**

The application consists of a Jupyter Notebook that will be available to view the data and all information regarding the recommendation system. It will allow a user to interact with it and run the program to get recommendations at the bottom of the notebook. They will be able to see visuals regarding the data that is used to give them recommendations as well.

With the use of this Jupyter Notebook, the user can help train the model from it, which will continually improve the system. The model and system will be entirely done in python and use pandas, numpy, python-Levenshtein, fuzzywuzzy and scipy libraries.

**4. Data Description**

The data that will be used for the recommendation system is data collected from the Steam platform and can be found on Kaggle’s website. It consists of over 16000 games that have reviews for every game, and scores from those reviews attached.

The raw csv data can be found at the following URL.

[Video-Game-Recommendation-System/metacritic\_critic\_reviews.csv at master · bradpappan/Video-Game-Recommendation-System · GitHub](https://github.com/bradpappan/Video-Game-Recommendation-System/blob/master/metacritic_critic_reviews.csv)

**5. Objective and Hypotheses**

Will users that like similar games, continue to like games reviewed similarly in the future? We will test this by obtaining the data from users interacting with the system. The data will show whether the users that agree with certain games reviews or not, continue to like similarly reviewed games in the future.

Our objective is to create a system that will provide meaningful recommendations to users that will increase sales and the satisfaction of the user. This will solve the problem introduced by senior members of WildCard Games.

**6. Methodology**

The project will follow the SEMMA methodology. This methodology will provide a great structure for the machine learning application, which the data is a main focus for SEMMA.

The phases of SEMMA are:

1. Sample
2. Explore
3. Modify
4. Model
5. Assess

**7. Funding Requirements**

|  |  |  |
| --- | --- | --- |
| **Resource** | **Hourly Cost** | **Total Hours** |
| Planning and design | $110 | 10 |
| Development | $125 | 80 |
| Design review and meetings | $125 | 10 |
|  | **Total Cost** | $12,350 |

**8. Stakeholder Impact**

The recommendation system will provide a great tool for WildCard Games sales team to use to improve sales for the company.

**9. Data Precautions**

The data is available to anyone online and does not contain any sensitive data or pose any risk at all.

**10. Developer Expertise**

Machine learning engineers with at least 2 years of python experience will be recruited for this specific project.

**TECHNICAL PROPOSAL (SECTION B)**

**1. Problem Statement**

Design and develop a system that allows a user to receive a video game recommendation based on a review and provide feedback on that recommendation. This system will take in the user’s response for future recommendations.

**2. Customer Summary**

WildCard Games has noticed a trend that many users are only purchasing one game. Senior members of the company believe this is due to users not being able to find new games that may interest them. The main problem is that there are video games that receive little purchases that once users know exist, they will purchase. We have been tasked to come up with a machine learning solution to recommend different games to users based on the reviews to attempt to increase sales and user satisfaction. The system will be available to all users for the platform once it has been implemented into the platform in the future.

**3. Existing System Analysis**

WildCard Games currently has no way to provide meaningful video game recommendations to its users. All games are displayed and available for users to search through however they want. This system allows a user to interact with it to skip the need to search and provides meaningful recommendations quickly. The users will also be able to provide feedback on the recommendation.

**4. Data**

Video game reviews and scores are collected from 2011 to 2019 and is limited. The dataset is publicly available and was acquired from Kaggle.com. The data is in csv form and will be imported into the project and used to train the model. All the data will be cleaned and removed of any null values prior to use.

**5. Project Methodology**

Development will follow the SEMMA methodology.

• **Sample**: Will be identify independent and dependent variables in this phase. An independent variable will be questions that will be ask. The dependent variable will be the responses from the video game recommendation system.

• **Explore**: Univariate and multivariate are conducted in this phase. We will be using univariate to understand what part the factors play in the project. Multivariate will be used to understand the relationship between variables.

• **Modify**: Data collected in the explore phase is parsed and cleaned in this phase. Once that is done, it passed onto the model phase.

• **Model**: In this phase a tree-based model is created with the cleaned data.

• **Assess**: The model will be evaluated, and the data can now be tested.

**6. Project Outcomes**

The finished project will include the following:

* Interactive Jupyter Notebook, to include
  + Code used to create the system
  + Visual charts and graphs
  + Output from code
  + Ability to provide feedback to the system via UI
* Working video game recommendation system

**7. Implementation Plan**

* Get dataset
* Set up environment to build project
* Clean dataset
* Import dataset into PyCharm
* Verify dataset is imported correctly with correct data
* Create model
* Train model
* Evaluate
* If errors are present, correct them and re-evaluate
* Document work and create UI within Jupyter Notebook

**8. Evaluation Plan**

Validation of the system will be performed from the average responses received from user interaction. Users can input up to + 2 or down to - 2 from the recommendation provided, which will be used to calculate the average response.

**9. Resources and Costs**

* **Programming environment**
  + The project will be created on Windows 10 and will use Jupyter Notebooks. The programming language will be Python.
* **Human Resource Requirements**

|  |  |  |
| --- | --- | --- |
| **Resource** | **Hourly Cost** | **Total Hours** |
| Planning and design | $110 | 10 |
| Development | $125 | 80 |
| Design review and meetings | $125 | 10 |
|  | **Total Cost** | $12,350 |

* **Environment Costs**

|  |  |  |
| --- | --- | --- |
| **Resource** | **Description** | **Cost** |
| Laptops | 4X HP EliteBook 850 | $5,000 |
|  | **Total** | $5,000 |

**10. Timeline and Milestones**

|  |  |  |  |
| --- | --- | --- | --- |
| **Phase** | **Start** | **End** | **Tasks** |
| 1 | 11/01/2022 | 11/05/2022 | Requirement gathering |
| 2 | 11/05/2022 | 11/08/2022 | A technical proof of concept is presented and submitted for review |
| 3 | 11/08/2022 | 11/10/2022 | Design review |
| 4 | 11/10/2022 | 11/11/2022 | Environment setup |
| 5 | 11/11/2022 | 11/18/2022 | Clean and parse data. |
| 6 | 11/18/2022 | 11/28/2022 | Model is created with cleaned data. |
| 7 | 11/28/2022 | 12/1/2022 | Model is evaluated. |
| 8 | 12/1/2022 | 12/15/2022 | Deliverables completed |
| 9 | 12/15/2022 | 12/31/2022 | The project is delivered |

**POST IMPLEMENTATION REPORT (SECTION D)**

**1. Project Purpose**

The purpose of this project is to build a video game recommendation system that recommends games based on similar critic reviews. This system will take in responses from the user and adjust the game and review displayed in the future.

The finished product will consist of an interactive Jupyter notebook that allows the user to input responses to the recommended video games.

**2. Datasets**

The dataset for this project was acquired from Kaggle’s website. The data was cleaned of null values.

The dataset consists of 6 columns where it is shown below:

Text

Description automatically generated

Below is the code used to clean the dataset of null values:

Graphical user interface, website

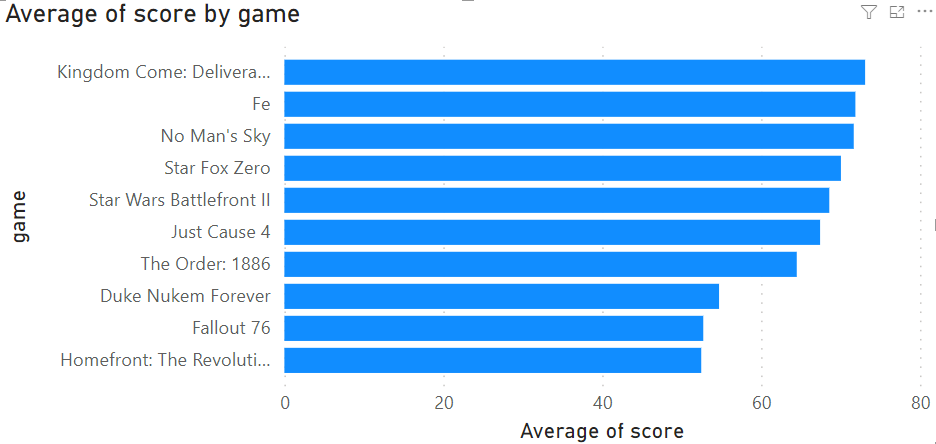
Description automatically generated

**3. Data Product Code**

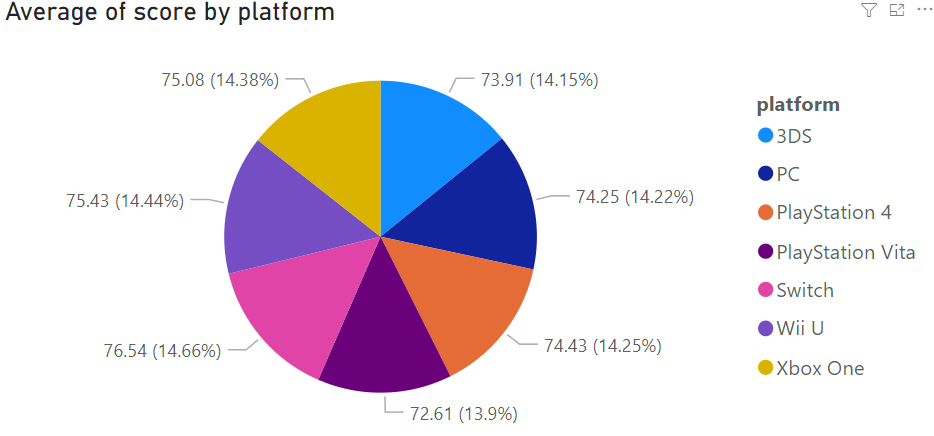
The data was saved to a csv format to allow easy access to the data. The data was saved within the project, then accessed via Panda’s library. The data was then cleaned of null values and was ready for descriptive and non-descriptive analysis. Code was not used to create the descriptive analysis and non-descriptive analysis; Microsoft Power Bi was a better choice. We decided that their product was able to create better charts, and more effective results from their Power Bi platform. We even decided to add an extra chart because of how effective this tool is.

The entire dataset was loaded into the platform and a descriptive analysis was created via their reports section. We went with average score by game in a bar graph, average score by platform in a pie chart, and count of reviews in a line graph. All of these show a great amount of information when it comes to what may be recommended to a user.

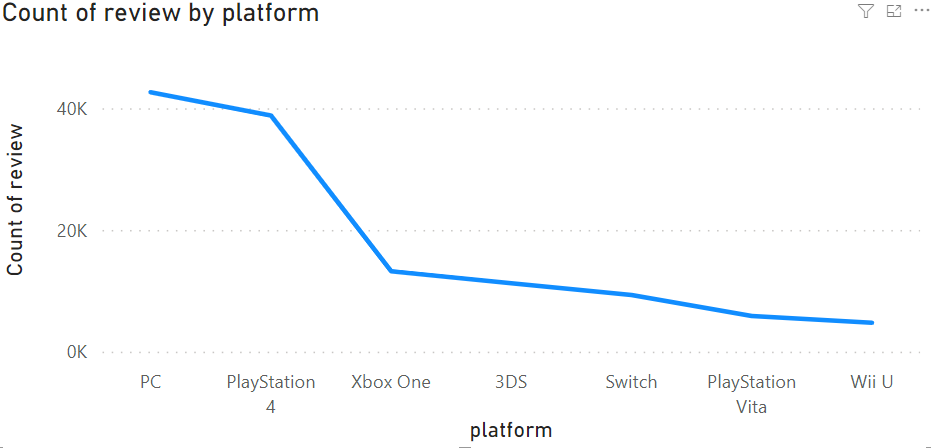
You can see top games by score,



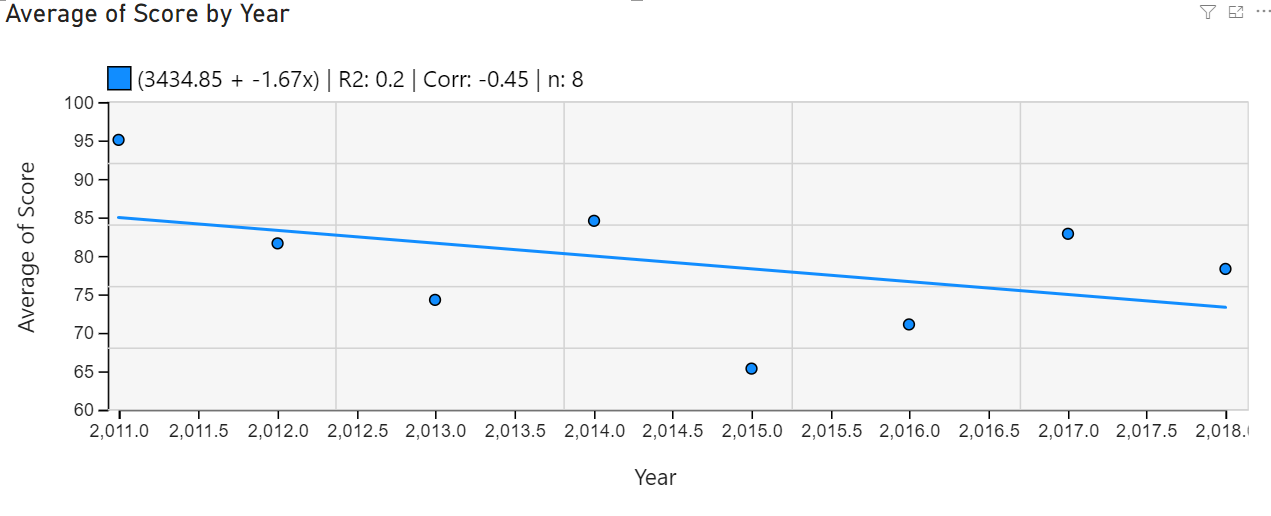
top average scores for platforms,



and as well what platform is primarily reviewed.



As well Power Bi created our non-descriptive analysis using an additional report function from Craydec Regression Chart. We inserted average scores from every year into the chart, and it clearly displays an ongoing trend for the scores. We can confidently say that average scores are trending downwards in the future.

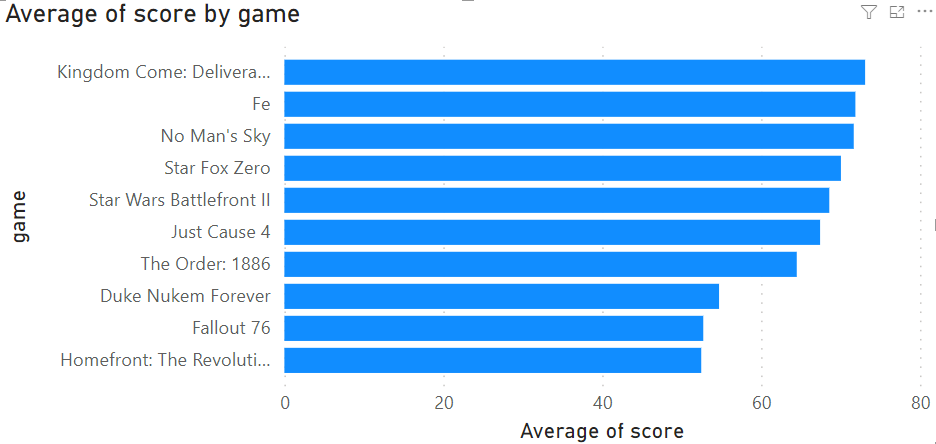


**4. Hypothesis Verification**

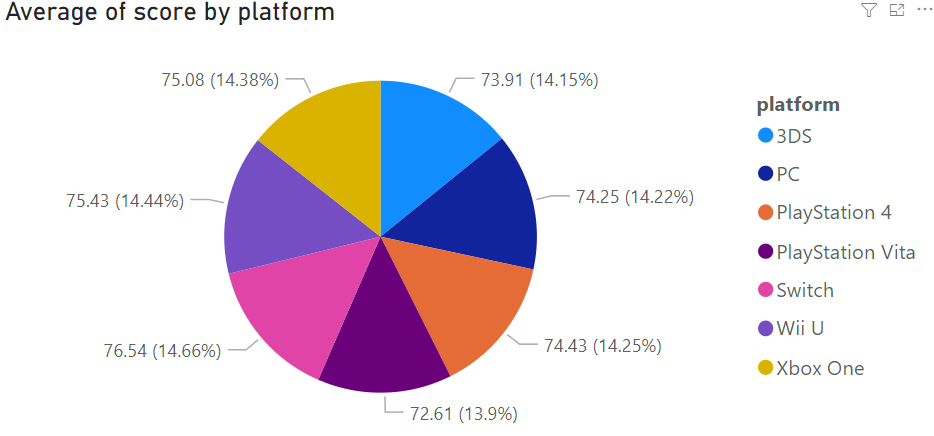
The hypothesis is as follows “Will users that like similar games, continue to like games reviewed similarly in the future?”. This is difficult to analyze before being implemented since the model needs to be trained by users of the platform to avoid any unnecessary bias added to the system. However, with the testing that we were able to do, we determined that the system does indeed maintain similar reviews so long as you continue to give it either a +1 or a +2 response and will adjust with the -1 and -2 responses. This system is entirely based on opinion on each individual person and is very hard to analyze its accuracy.

**5. Effective Visualization and Reporting**

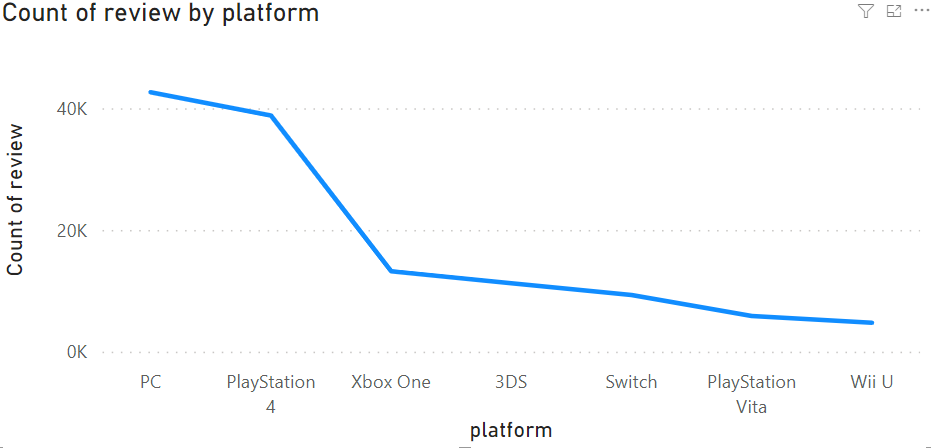
Below are the visualizations that were created from the cleaned dataset. We have a bar graph that displays the top 10 games by their average score.



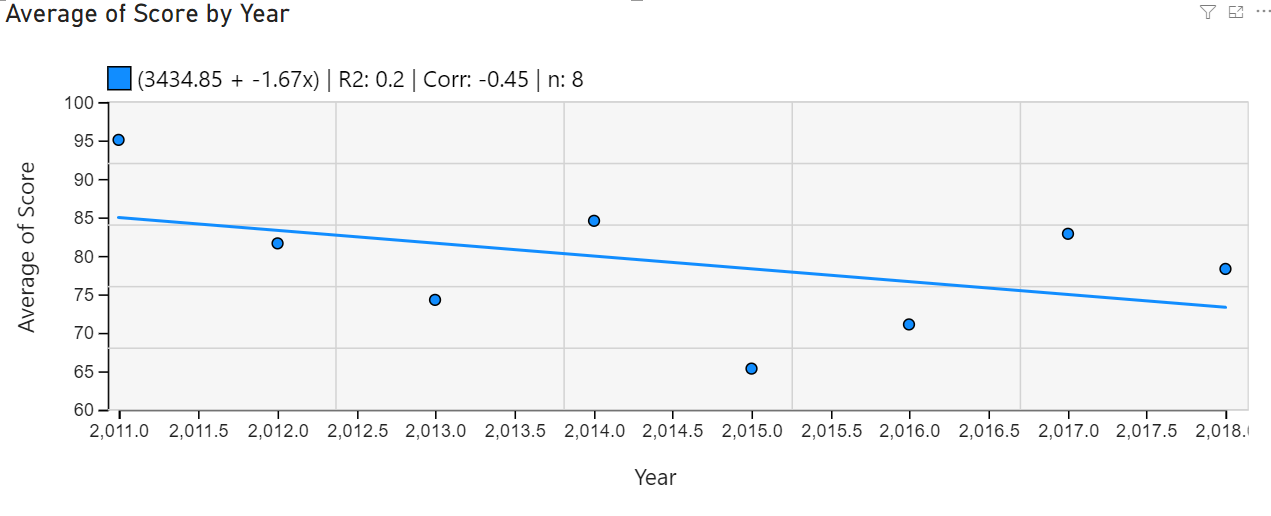
Then we have a chart showing average score by each platform in a pie chart.



Then we have a total count of reviews for each platform in a line graph.



And finally, we have a regression chart, showing a downwards trend in average score in the future.



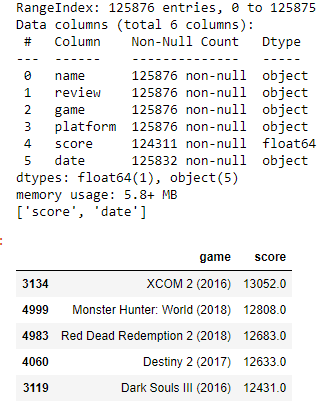
All of these are effective descriptions of the dataset that we used, and possibly additional data that will be entered in the future.

**6.** **Accuracy Analysis**

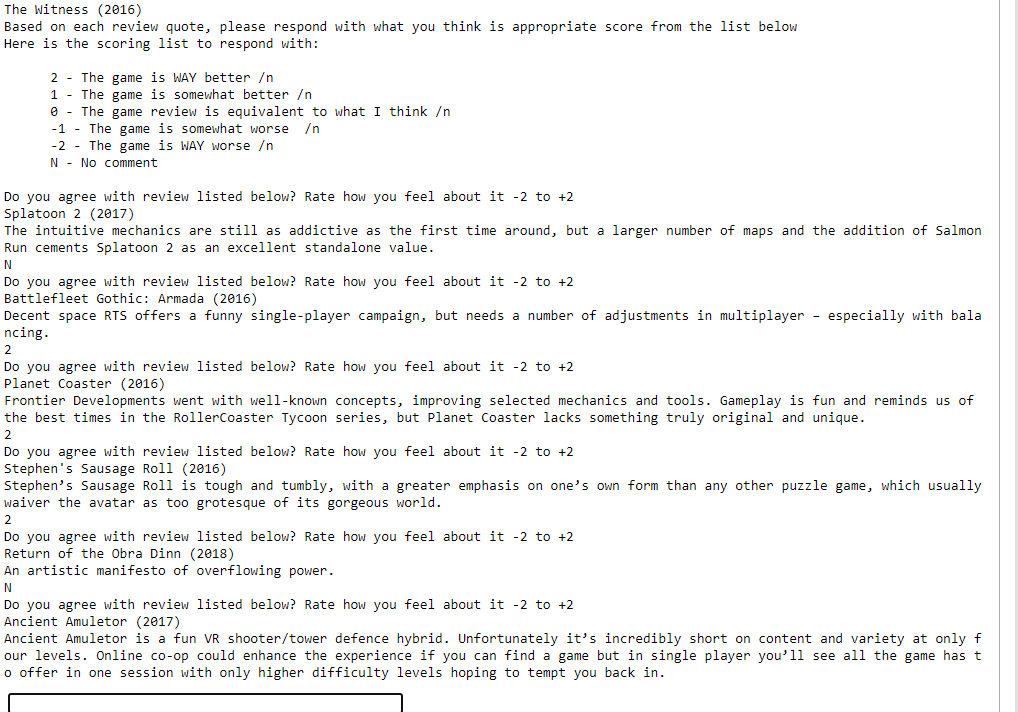
Our product is difficult to analyze accuracy due to the nature of the system. Every recommendation provided may have a different result to each user. Providing an accuracy analysis before giving the system time to train with users on WildCard Games would provide inaccurate results. However, with our testing we were able to determine that reviews do remain similar with a +2 or +1 score. They also will change with a -2 and -1 score.

**7. Application Testing**

We began with unit testing sections of our code. Below is a screenshot of output to determine that we have accurate data being pulled.



We also used dynamic and static code analysis throughout our development cycle. We also ran black box testing and usability testing with members of our quality assurance department. Below is a screenshot of one of those tests attempting to find relevant reviews from providing +2 or N as responses and getting similar reviews.



**8. Application Files**

The source code files were included in the submission. In the attached files will be:

* Video Game Recommendation System.ipynb file that contains my notebook
* Log.txt that will hold logs from the system
* Metacritic\_critic\_reviews.csv that is the dataset

**9. Users Guide**

**NOTE:** This guide is designed for windows 10 machines.

To access my Jupyter Notebook you will need to install the latest version of python installed, this can be found at <https://www.python.org/downloads/>. You will need to select the system you are currently on and using.

Next you need to install anaconda3 to your machine, this can be done by going to <https://www.anaconda.com/products/distribution>.

Next you will need to download the provided jupyter notebook files and place them in your user account directory in windows, for example mine is C:\Users\Brad Pappan.

Once that is completed, you can open jupyter notebook via the search menu. Run that and jupyter will open in a browser. You will see the folder titled C964 Notebook, open the folder

Click on Video Game Recommendation System and you can run the notebook from the play button at the top. On the very bottom of the web page you can enter your responses to the system.

**10. Summation of Learning Experience**

A lot of course from WGU contributed greatly to this project, like Intro to Artificial Intelligence, Project Management, and even Software II. I had to research and learn a lot more about machine learning and how to use jupyter notebooks to make this. This project helped me realize that there is a lot more that can be done in something this complicated, but its still fun to do it.