April 17, 2023

C964: Computer Science Capstone – GameScores, INC. Ratings Prediction

Polefko, Alexander - 001405572

Part A: Project Proposal for Business Executives

Letter of Transmittal

Polefko, Alexander
Founder and Machine Learning Engineer
GameScores, INC.
April 17, 2023
a.polefk@wgu.edu

Letter of Transmittal: Ratings Prediction

Dear Mr. Bowser,

I hope this finds you well. I'm Alex, the Machine Learning Engineer at GameScores, INC., and below is the proposal for a Ratings Prediction data product to assist Nintendo with marketing, press, and game rating/scores.

There are a huge number of choices for gamers in 2023. Gamers frequently will "wish list" games that are already out, that were just released, or that will come out in the future. How does a gamers wish list correlate to the rating of a game?

This data product, Ratings Prediction by GameScores, INC., will assist in predicting the rating of games based on wish list data from 1983-2023. You and your content marketing team will be able to filter by the major players in the industry.

The Ratings Prediction data product will directly benefit Nintendo and partners by helping predict the power of a gamer adding a game to their wish list and what that meant to the games rating. The current plan for the data product is to use intense machine learning through a linear regression model.

GameScores, INC. has completed the total cost estimate at $51,520.

I'll be the lead machine learning engineer on this project. I helped found GameScores, INC. in 2019 and I earned my bachelor's degree in computer science from Western Governor's University in 2017. Since I founded the company, I've personally worked on machine learning solutions for 11 companies on the Fortune 500 (all in the top 50).

My team and I are honored to be working with a company as prolific as Nintendo. We greatly look forward to our partnership and the benefits our machine learning data product, Ratings Prediction, can provide to you and your team.

Very Respectfully,

Alexander Polefko
Founder and Machine Learning Engineer

Project Recommendation

Please view the subheadings below for more detail on the Ratings Prediction project.

Problem Summary

- The Ratings Prediction data product will help Nintendo assess and predict the connection between games put onto a wish list and their respective rating.
- In 2023, there are more new games than ever, as well as all the games that came before. Gamers add titles to a wish list and this data product can help predict the rating of games by developer by the number of times a game is on wish list.
- For Nintendo, this means they can evaluate their strategy with content marketing and product to boost wish list features. For example, including a discount for gamers if the game is on their wish list.
- GameScores, INC. will deliver a dashboard that allows filtering by Nintendo, Sony, or All games from a data set that spans 1980-2023. The Ratings Prediction data product will enable Nintendo to make informed decisions around existing and upcoming games and wish list behavior.

Application Benefits

- This data product meets a huge need of Nintendo, which is continued differentiation and time commitment from players. There are many games on the market and more to come, so using this machine learning model to assist in new strategies will be key.
- Nintendo will greatly benefit from this Ratings Prediction data product by enabling feature improvements to the wish list and examining existing games in their catalog. For example, Nintendo could see the results of the ML model's prediction and confirm that more wish lists will indict a higher rating and set Objective Key Results (OKRs) to reach that wish list metric.

Application Description

- The Ratings Prediction data product will use the 1980-2023 data set and parse out the rating and wish list count per game.
- A linear regression model will then be applied and viewable in a user-friendly dashboard.
- The model will be trained and tested.
- Users on content marketing or product teams at Nintendo can then use the dashboard to see the model's prediction for best fit on wish list count to rating for games (filtered by Nintendo, Sony, or All games).

Data Description

- This data set was obtained from Kaggle.com and includes games from 1980-2023.
- The data set includes quantitative data, such as the ratings and wish list count per game. Those are the primary variables used. This set is a CSV file an, using Python with the Pandas library, will be turned into a "dataframe" for manipulation by a machine learning model.
- The independent variable is the wish list count. The dependent variable is the rating. The supervised, linear regression model will be used to predict the rating and wish list maximization.
- The data set does include some games that have been added to a wish list much more than the mean. This is the primary anomaly, otherwise the data set has high usability.

Objectives and Hypothesis

- Please see the table below for objectives.
- Hypothesis: The more times a game is added to a wish list, the higher the rating will be, and machine learning will enable a predictable rating based on the count required on the wish list.
- Given the data set and using a linear regression model with "R^2 Score", the desired accuracy is .1 out of 1.

Objectives	Success Criteria
Accuracy	The prediction accuracy of the linear regression model will be .1 out of 1.
Escaped Defects	There will be no more than 8 at a Critical priority

Methodology

- GameScores, INC. operates with the agile methodology. We've found on the many projects we've done that being able to quickly prototype, test, and iterate within a sprint format is extremely effective for machine learning data products.
- This methodology allows GameScores, INC. to get valuable feedback from testers and stakeholders, then implement in the very next sprint.
- In general, our sprints follow this phased structure:
    - In the first sprint, or Phase 1, our squad will prepare the data and environments. This also includes confirm requirements and acceptance criteria.
    - In Phase 2, our squad moves onto confirm the linear regression model as well as manipulating the data set. Typically, this is where model training starts.
    - During Phase 3, the team will test the model and build the GUI. This is also a key sprint to get feedback in case a pivot is needed.

- Finally, Phase 4 involves correcting defects and delivering the final product. The product owner and I (Machine Learning Engineer) will also produce key documentation.

Funding Requirements

- The total estimate is $51,520 for all personnel and tools. The tools are already apart of GameScores, INC. and all personnel will be working exclusively on this project for four (4) sprints, or eight (8) weeks.
- The squad consists of a Product Owner, a Scrum Manager, a Machine Learning Engineer (myself), and a Junior Developer. Their hourly rates are listed in Section B.

Data Precautions

- The primary data set is accessible from Kaggle.com and has been verified. No sensitive data or any handling is required for this project. The data product itself, Ratings Prediction, will be delivered directly to Nintendo.

Developer's Expertise

Please see education and experience below:

- 2021-2023: Continued creating machine learning solutions for companies like Amazon and their games division.
- 2020: Created a classification machine learning data product for Supergiant Games.
- 2019: founded GameScores, INC. in 2019.
- 2017: bachelor's degree in computer science from Western Governor's University.

The capstone of my degree was in machine learning and helped me start my business two years after that. Prior to my time at WGU, I was a self-taught web developer at Salesforce.

Part B: Project Proposal

Please see the formal Project Proposal below. My primary contact is apolefk@wgu.edu for any questions, comments, or concerns.

Problem Statement

- In 2023, there are more new games than ever, as well as all the games that came before. Gamers add titles to a wish list and this data product can help predict the rating of games by developer by the number of times a game is on a wish list.

Customer Summary

- Nintendo is household name and a king in the video game industry. GameScores, INC. can partner with Nintendo to create a machine learning data product to help navigate the enormous market and help create solutions for gamers to choose Nintendo more and more.
- The Ratings Predication application can close the gap for Nintendo by helping predict the ratings from games wish listed. This is valuable because of the number of games already released and on wish lists as well as predict the wish list and rating changes for future games. This tool will allow Nintendo to examine their competitors progress as well.

Existing System Analysis

- Nintendo currently has no machine learning model for predicting the rating with wish list data. This is a huge gap GameScores, INC. can fill with the upcoming data product, Ratings Prediction.

Data

- The raw data set can be found here: https://www.kaggle.com/datasets/arnabchaki/popular-video-games-1980-2023.
- The data will be downloaded and accessed via CSV file in Excel. Then using Python and "pandas" it will be formed into a "dataframe" for access by the machine learning model.
    - Data will be accessed via the "dataframe".

    - The data set will then be manipulated through if-else statements filtering by the "dataframe" column containing team or developer name.
- Any duplicates or blanks in the data set will be eliminated via Excel. Outliers will be accepted in this case because a game added a huge number of times to wish lists is critical information.

Project Methodology

- GameScores, INC. operates with the agile methodology. The ability to quickly prototype, test, and iterate within a sprint format is extremely effective for machine learning data products.
- This methodology allows GameScores, INC. to get valuable feedback from testing and stakeholders, then implement in the very next sprint.
- In general, our sprints follow this phased structure:
    a. In the first sprint, or Phase 1, our squad will prepare the data and environments. This also includes confirm requirements and acceptance criteria.
    b. In Phase 2, our squad moves onto confirm the linear regression model as well as manipulating the data set. Typically, this is where model training starts.
    c. During Phase 3, the team will test the model and build the GUI. This is also a key sprint to get feedback in case a pivot is needed.
    d. Finally, Phase 4 involves correcting defects and delivering the final product. The product owner and I (Machine Learning Engineer) will also produce key documentation.
- The deployment of this application will be a hyperlink accessible for Nintendo stakeholders.

Project Outcomes

This squad will produce three key deliverables:

- Dashboard: This is a user-friendly interface accessing the machine learning model trained and tested on the 1980-2023 games data set. The data product will be accessible for stakeholders at Nintendo.
- User Guide: A comprehensive user guide will be created for enabling the data products use internally. In the post-implementation documentation when the product is complete, there will also be use cases or examples.
- Trained Model: The linear regression model will be trained and tested on the data set and accessible via the dashboard or visible as Python code.

Implementation Plan

- The overall strategy for Ratings Prediction is having an accessible link for Nintendo Stakeholders.
- Please see the sprint development timeline below with milestones.
- Each sprint represents a key phase with four (4) phases total.
- GameScores, INC. already has the personnel and data set ready, so no other dependencies at this time.
- Distribution will occur when I send the completed data product and documentation.

Sprint (two weeks per)	Start	End	Milestones
1	5/1/2023	5/12/2023	New repository created – Jr. Developer Data set parsed, cleaned, and evaluated – ML Engineer and Jr. Developer
2	5/15/2023	5/26/2023	Machine learning algorithm confirmed – ML Engineer and Jr. Developer Implementation and manipulation of the data set – ML Engineer Training conducted and evaluated – ML Engineer
3	5/29/2023	6/9/2023	Testing conducted and evaluated – ML Engineer GUI Built and published – Jr. Developer User Testing –Product Owner
4	6/12/2023	6/23/2023	Defects and feedback addressed from Nintendo – Product Owner, Scrum Manager, ML Engineer, and Jr. Developer Full documentation and User Guide – ML Engineer and Product Owner Delivery of finished data product to Nintendo – ML Engineer and Product Owner

Evaluation Plan

- After each phase, the Product Owner and myself (Machine Learning Engineer) will review Acceptance Criteria and Acceptance Tests. These will then be a part of a System Demonstration at GameScores, INC. to gather internal feedback before going to full release. Please see the key components per phase:
    a. Phase 1 – Data set parsed, cleaned, and evaluated.
        i. Verification primarily by ML Engineer via Code Review.

b. Phase 2 – Training conducted and evaluated.
    i. Verification primarily by ML Engineer via Code Review.
c. Phase 3 – Testing conducted and evaluated.
    i. Verification primarily by ML Engineer via Code Review.
d. Phase 4 – Delivery with documentation.
    i. Verification primarily by Product Owner.

- The validation method will be focused on the objectives and their success rate as well as the core accuracy metric of the machine learning model (linear regression). This will primarily be conducted by the Machine Learning Engineer and the Product Owner.

Resources and Costs

All employees listed will be out 40 hours a week (eight (8) hours per day, five (5) days a week). Sprints last two (2) weeks and there will be a total of four (4) sprints.

Resource	Description	Cost (8hrs, 5 days per week)
Product Owner	Hourly Rate = $39	$12,480
Scrum Manager	Hourly Rate = $42	$13,440
Machine Learning Engineer	Hourly Rate = $50	$16,000
Junior Developer	Hourly Rate = $30	$9,600
Laptops	Provided by GameScores, INC.	$0
Code Repository and GUI	Included within the package GameScores, INC. uses	$0
	Total	$51,520

Timeline and Milestones

Please see the table below for the sprint timeline.

Key Information: All milestones are due on the end date of the sprint by close-of-business (COB).

Sprint (two weeks per)	Start	End	Milestones
1	5/1/2023	5/12/2023	New repository created – Jr. Developer Data set parsed, cleaned, and evaluated – ML Engineer and Jr. Developer
2	5/15/2023	5/26/2023	Machine learning algorithm confirmed – ML Engineer and Jr. Developer Implementation and manipulation of the data set – ML Engineer

			Training conducted and evaluated – ML Engineer
3	5/29/2023	6/9/2023	Testing conducted and evaluated – ML Engineer GUI Built and published – Jr. Developer User Testing –Product Owner
4	6/12/2023	6/23/2023	Defects and feedback addressed from Nintendo – Product Owner, Scrum Manager, ML Engineer, and Jr. Developer Full documentation and User Guide – ML Engineer and Product Owner Delivery of finished data product to Nintendo – ML Engineer and Product Owner

Part C: Application

Please see the submission link for the Ratings Prediction application by GameScores, INC. or access the link below:

- https://mybinder.org/v2/gh/apolefko/CapstoneWGU/HEAD

Part D: Post-implementation Report

The Ratings Prediction data product is complete readily accessible. Please see the link in Section C or further down in this section.

A Business (or Organization) Vision

- In 2023, there are more new games than ever, as well as all the games that came before. Gamers add titles to a wish list and this data product can help predict the rating of games by developer by the number of times a game is on wish list.
- For Nintendo, this means they can evaluate their strategy with content marketing and product to boost wish list features. For example, including a discount for gamers if the game is on their wish list.
- Another example would be a sales representative using this to tool to determine the wish list and rating prediction on the dashboard then accessing further internal data to determine why a certain title didn't fit the model (or how to get it more wish listed to increase the rating).

Datasets

- The raw data is a data set from Kaggle.com focusing on games from 1980-2023.
    - This includes fields like Title, Team, Wish list count, Rating, and more.
- The data was processed by me to exclude many irrelevant fields and converting the Wish list count value from an abbreviated value to a true integer.
- For example, some games had 4.2K but the K was hard coded in the cell. I converted all those values to true integers so it would work with the Ratings Predication supervised, linear regression model.

- The core fields used were “Team”, “Wishlist”, and “Rating”.
- The main issue was if there were blanks or non-integers in those columns. I was able to solve that and make it accessible for the algorithm.

Raw Data

- There were multiple duplicates and blanks in the Team Column that were removed.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Title	Release Date	Team	Rating	Times Listed	Number of R	Genres	Summary	Reviews	Plays	Playing	Backlogs	Wishlist
2	0	Elden Ring	25-Feb-22	['Bandai Nar	4.5	3.9K	3.9K	['Adventure',	Elden Ring is	["The first pl	17K	3.8K	4.6K	4.8K
3	1	Hades	10-Dec-19	['Supergiant	4.3	2.9K	2.9K	['Adventure',	A rogue-lite	['convinced t	21K	3.2K	6.3K	3.6K
4	2	The Legend c	3-Mar-17	['Nintendo',	4.4	4.3K	4.3K	['Adventure',	The Legend c	['This game	30K	2.5K	5K	2.6K
5	3	Undertale	15-Sep-15	['tobyfox', '8	4.2	3.5K	3.5K	['Adventure',	A small child	['soundtrack	28K	679	4.9K	1.8K
6	4	Hollow Knigh	24-Feb-17	['Team Cher	4.4	3K	3K	['Adventure',	A 2D metroic	["this games	21K	2.4K	8.3K	2.3K
7	5	Minecraft	18-Nov-11	['Mojang Stu	4.3	2.3K	2.3K	['Adventure',	Minecraft fo	['Minecraft i	33K	1.8K	1.1K	230
8	6	Omori	25-Dec-20	['OMOCAT',	4.2	1.6K	1.6K	['Adventure',	A turn-based	["The best ga	7.2K	1.1K	4.5K	3.8K
9	7	Metroid Drea	7-Oct-21	['Nintendo',	4.3	2.1K	2.1K	['Adventure',	Join intergal	['Have only b	9.2K	759	3.4K	3.3K
10	8	Among Us	15-Jun-18	['InnerSloth']	3	867	867	['Indie', 'Stra	Join your cre	["it's a solid	25K	470	776	126
11	9	NieR: Autom	23-Feb-17	['PlatinumGa	4.3	2.9K	2.9K	['Brawler', 'F	NieR: Autom	['Holy shit', '	18K	1.1K	6.2K	3.6K
12	10	Persona 5 Rc	31-Oct-19	['Atlus USA',	4.4	2.7K	2.7K	['Adventure',	An enhanced	['Verdadeiro	12K	2.3K	5.1K	3K
13	11	Stray	19-Jul-22	['BlueTwelve	3.7	1.5K	1.5K	['Adventure',	Lost, alone, a	['Press B to f	7.7K	801	2.5K	3.4K
14	12	God of War	20-Apr-18	['Sony Intera	4.2	2.9K	2.9K	['Adventure',	God of War i	['freya te vej	21K	1.1K	4.8K	2.6K
15	13	Portal 2	18-Apr-11	['Valve', 'Ele	4.4	2.9K	2.9K	['Adventure',	Sequel to the	['This is my f	29K	471	3.9K	1.2K
16	14	Bloodborne	24-Mar-15	['FromSoftw	4.5	3.4K	3.4K	['Adventure',	An action RP	["I'm not tryi	17K	1.1K	5.6K	3.3K
17	15	Celeste	25-Jan-18	['Extremely (	4.2	2.8K	2.8K	['Adventure',	Help Madelir	['video game	20K	1.2K	5.9K	2K

Processed Data

- Observe the corrected values in “Wishlist”.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
		Title	Release Date	Team	Rating	Times Listed	Number of R	Genres	Summary	Reviews	Plays	Playing	Backlogs	Wishlist
	0	Elden Ring	25-Feb-22	['Bandai Nar	4.5	3.9K	3.9K	['Adventure',	Elden Ring is	["The first pl	17K	3.8K	4.6K	4800
	1	Hades	10-Dec-19	['Supergiant	4.3	2.9K	2.9K	['Adventure',	A rogue-lite	['convinced t	21K	3.2K	6.3K	3600
	2	The Legend c	3-Mar-17	['Nintendo',	4.4	4.3K	4.3K	['Adventure',	The Legend c	['This game	30K	2.5K	5K	2600
	3	Undertale	15-Sep-15	['tobyfox', '8	4.2	3.5K	3.5K	['Adventure',	A small child	['soundtrack	28K	679	4.9K	1800
	4	Hollow Knigh	24-Feb-17	['Team Cher	4.4	3K	3K	['Adventure',	A 2D metroic	["this games	21K	2.4K	8.3K	2300
	5	Minecraft	18-Nov-11	['Mojang Stu	4.3	2.3K	2.3K	['Adventure',	Minecraft fo	['Minecraft i	33K	1.8K	1.1K	230
	6	Omori	25-Dec-20	['OMOCAT',	4.2	1.6K	1.6K	['Adventure',	A turn-based	["The best ga	7.2K	1.1K	4.5K	3800
	7	Metroid Drea	7-Oct-21	['Nintendo',	4.3	2.1K	2.1K	['Adventure',	Join intergal	['Have only b	9.2K	759	3.4K	3300
	8	Among Us	15-Jun-18	['InnerSloth']	3	867	867	['Indie', 'Stra	Join your cre	["it's a solid	25K	470	776	126
	9	NieR: Autom	23-Feb-17	['PlatinumGa	4.3	2.9K	2.9K	['Brawler', 'F	NieR: Autom	['Holy shit', '	18K	1.1K	6.2K	3600
	10	Persona 5 Rc	31-Oct-19	['Atlus USA',	4.4	2.7K	2.7K	['Adventure',	An enhanced	['Verdadeiro	12K	2.3K	5.1K	300
	11	Stray	19-Jul-22	['BlueTwelve	3.7	1.5K	1.5K	['Adventure',	Lost, alone, a	['Press B to f	7.7K	801	2.5K	3400
	12	God of War	20-Apr-18	['Sony Intera	4.2	2.9K	2.9K	['Adventure',	God of War i	['freya te vej	21K	1.1K	4.8K	2600
	13	Portal 2	18-Apr-11	['Valve', 'Ele	4.4	2.9K	2.9K	['Adventure',	Sequel to the	['This is my f	29K	471	3.9K	1200
	14	Bloodborne	24-Mar-15	['FromSoftw	4.5	3.4K	3.4K	['Adventure',	An action RP	["I'm not tryi	17K	1.1K	5.6K	3300
	15	Celeste	25-Jan-18	['Extremely (	4.2	2.8K	2.8K	['Adventure',	Help Madelir	['video game	20K	1.2K	5.9K	200

- The data set is labeled “gameNoDup.CSV” in the Binder link.

Data Product Code

- The raw data had some critical errors that had to be processed as seen in the section above this. The key components were non-integers in the integer fields or blanks in the Team column to affect the filter for Nintendo, Sony, or All games.
- The descriptive methods were also used as visualizations:

    - Scatter Matrix for the All Games dataset.
    - Histograms for both Sony and Nintendo filters.
    - Scatter Plot focusing on the wish list and ratings variables.
- The non-descriptive method was a linear regression algorithm that was trained and tested, then visible with the Scatter Plot and prediction plot.
    - The non-descriptive section used the "train_test_split" method to separate the data and make initial predictions.
    - These methods worked here due to the nature of the data set enabling a supervised model and that these were integer values. Regression was most appropriate while classification would have been inappropriate.
    - The method was trained on a 20% section of the "dataframe" and tested using the method above's split as well.
- Please see the dashboard and code here: https://mybinder.org/v2/gh/apolefko/CapstoneWGU/HEAD.
- The analysis really enabled the split by filter of Nintendo, Sony, and All games. The broad view of all data and games with the Scatter Matrix is particularly useful to start out.

Objective (or Hypothesis) Verification

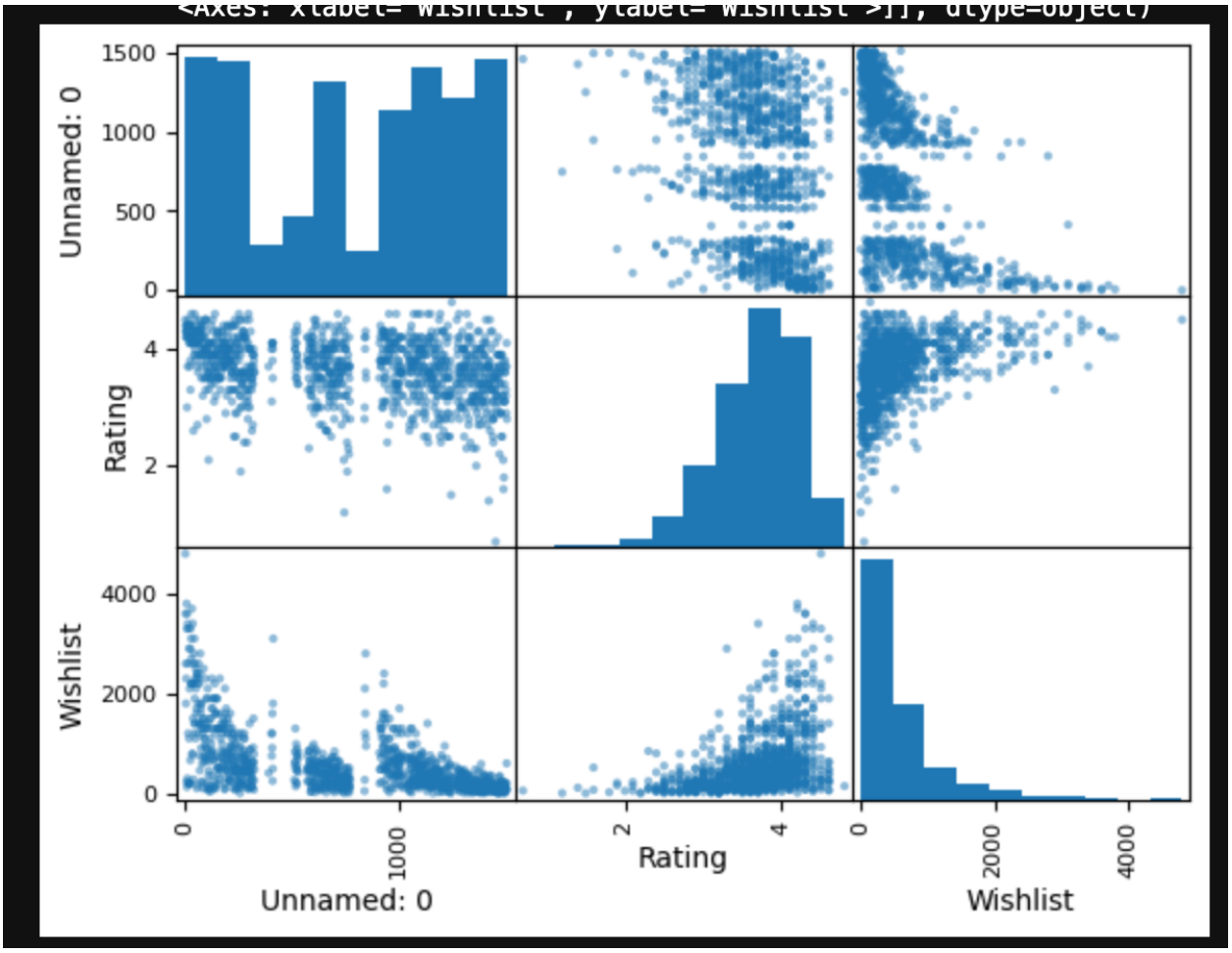
- The linear regression model was able to predict the wish list and rating trend above the .1 out of 1 goal with the "R^2 Score" as the accuracy metric.
    - This is the case for all three filters (Nintendo, Sony, and All games). Please see screenshots in the Accuracy Analysis section.
- No defects escaped testing at Critical or higher priority.
- The objectives and hypothesis were met due to effective planning and a well-trained and tested machine learning model (linear regression).

Effective Visualization and Reporting

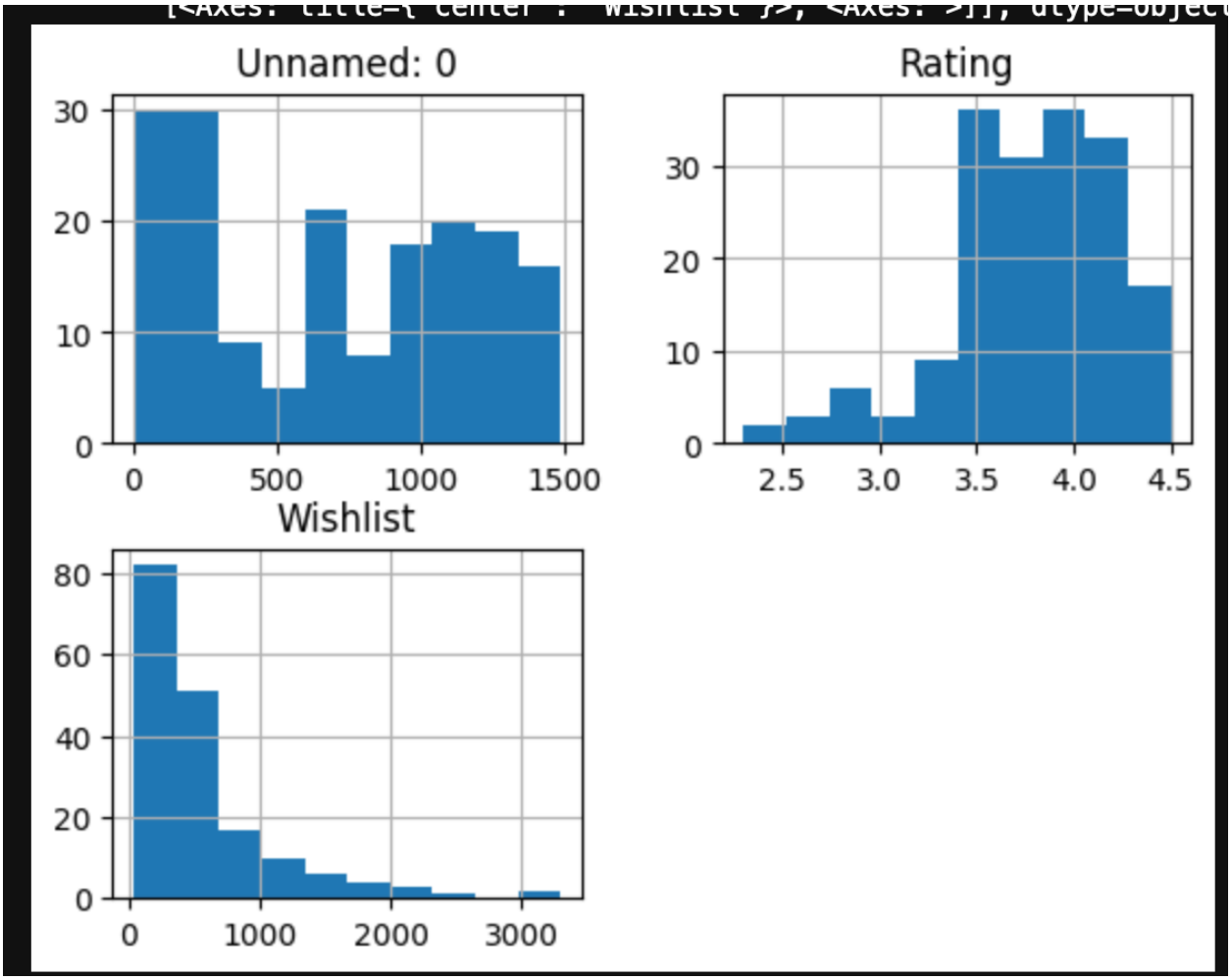
- The descriptive methods were key in initially exploring the data after it was processed.
    - The lower number of games by Team for Nintendo and Sony helped create an uncluttered visual to start before the non-descriptive method was implemented.

- Analysis with the visualizations showed a correlation to higher wish listed items and ratings, which enabled the predicating model later. This is present in the Histograms, the Scatter Matrix, and the Scatter Plot.
- In summary, the data shows a few outliers that are more highly wish listed, but the correlation stands for most games. The linear regression shows the predictive, supervised linear regression plot as well.

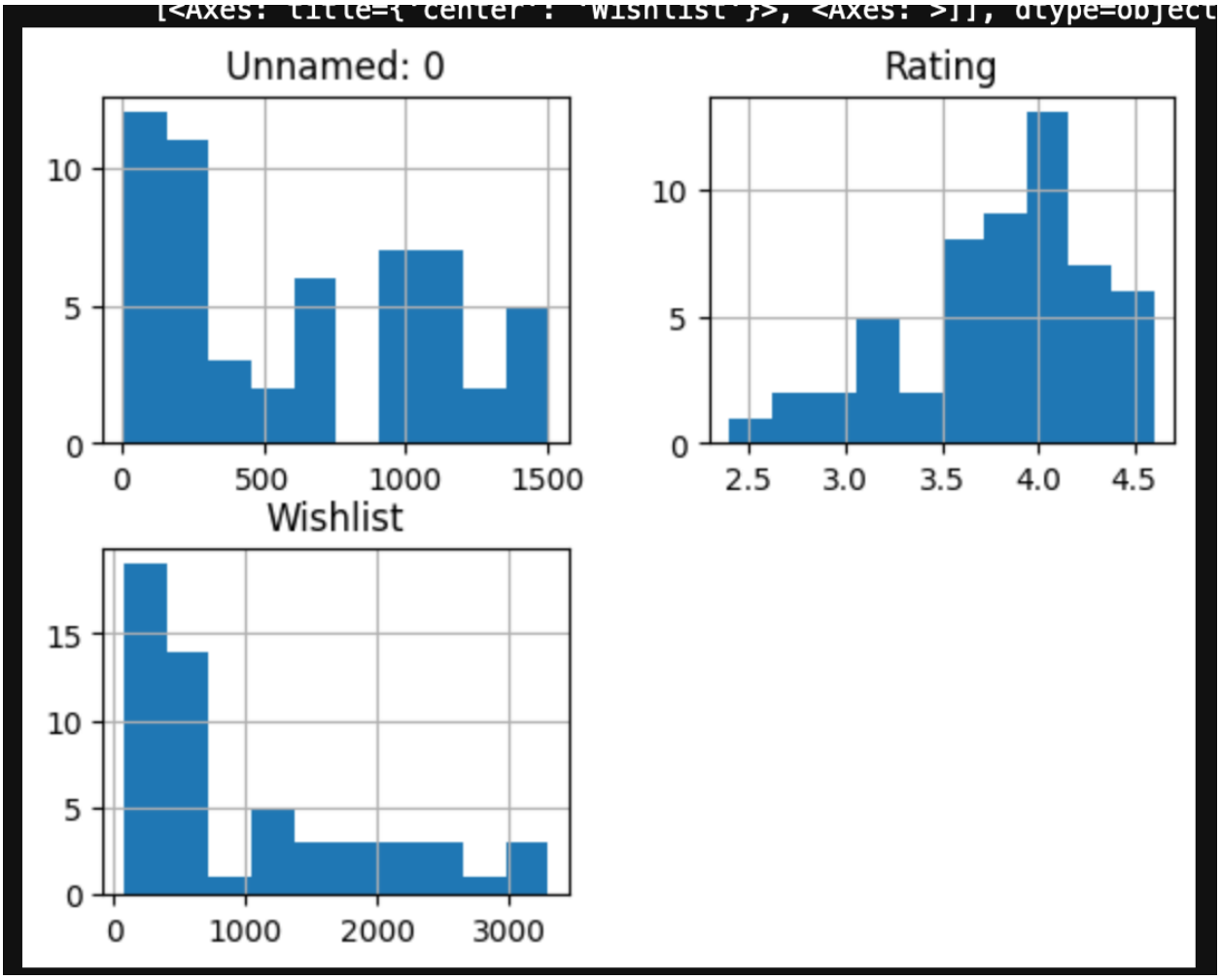
Scatter Matrix



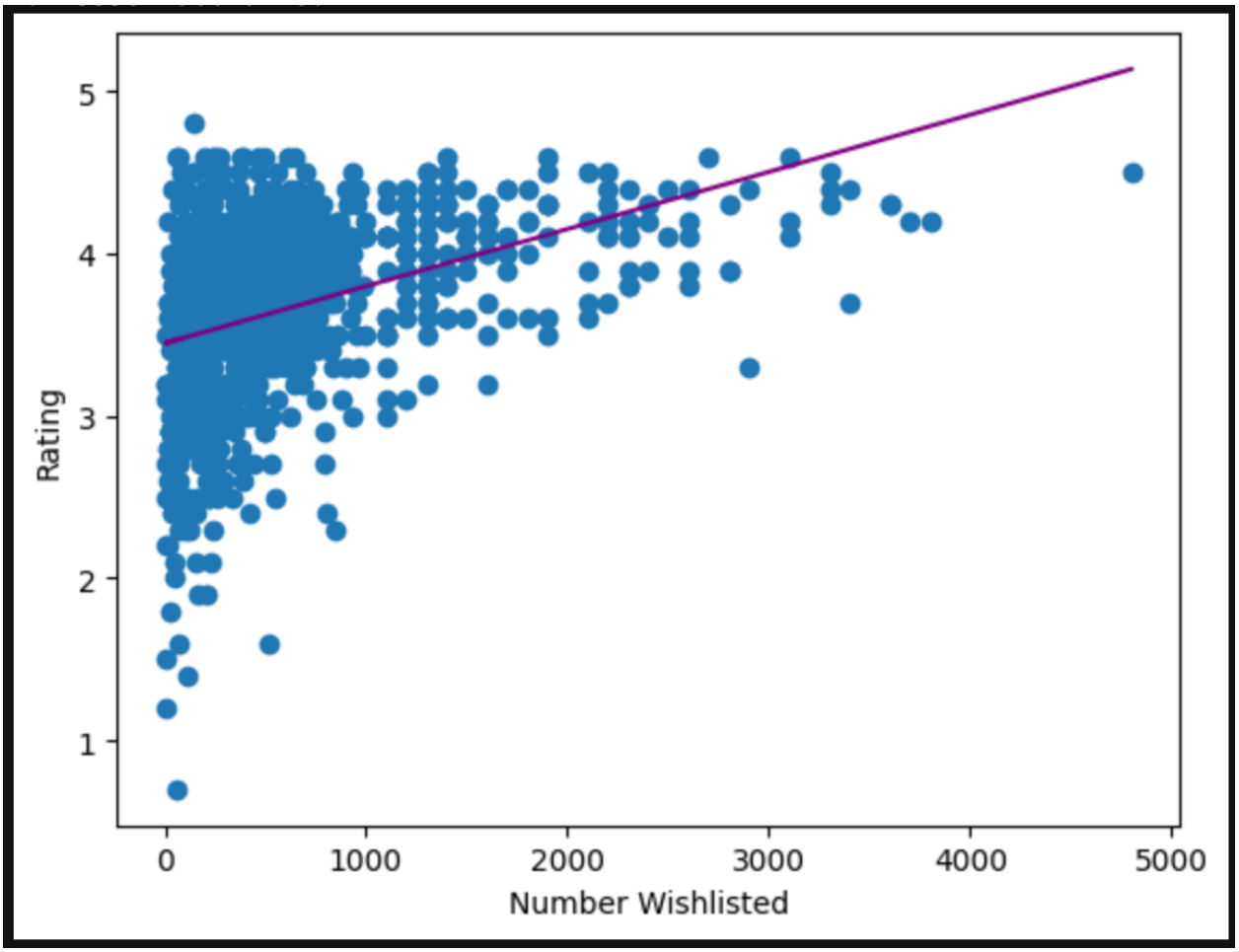
Histogram (Nintendo)



Histogram (Sony)



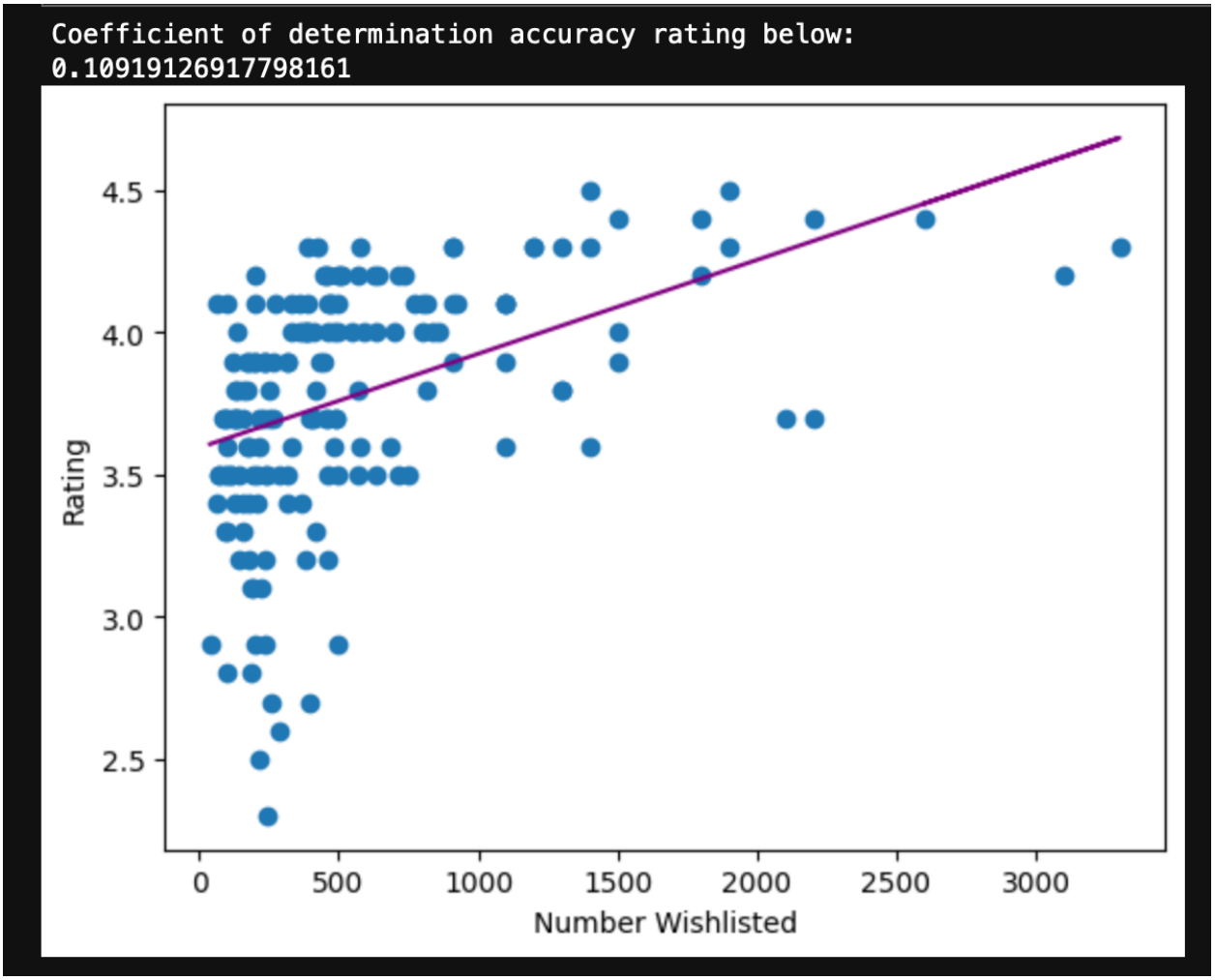
Scatter Plot (tied to the linear regression (non-descriptive))



Accuracy Analysis

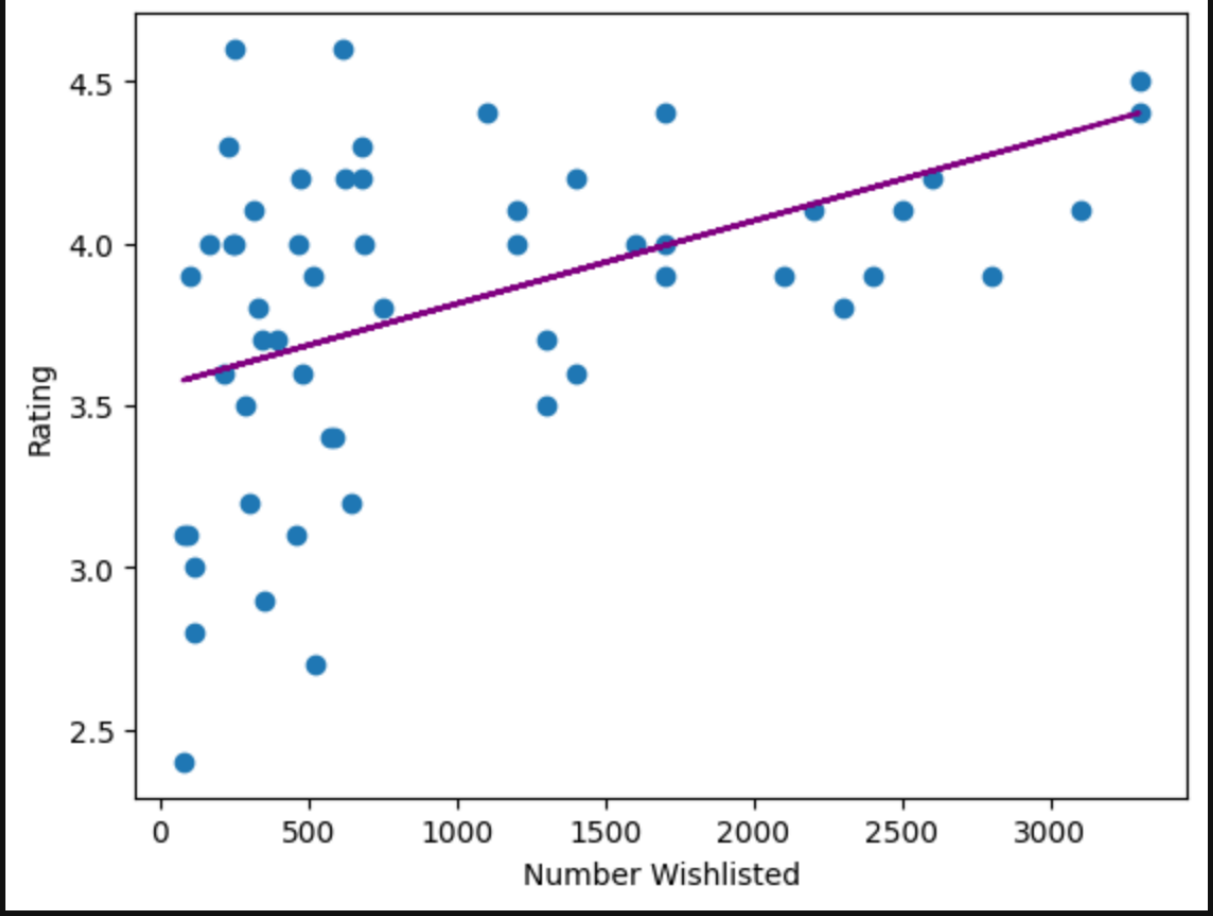
- The accuracy metric used from scikit-learn was the R^2 Score. This has a top score of 1.0 and is also called the coefficient of determination. This is implemented in the code after the training set and during training.
- The machine learning model for Ratings Prediction is inaccurate as all three options (Nintendo, Sony, or All) produce results just above .1 out of 1. This does meet the objectives outlined in this data product documentation, but accuracy can and should be improved with more training.

Nintendo



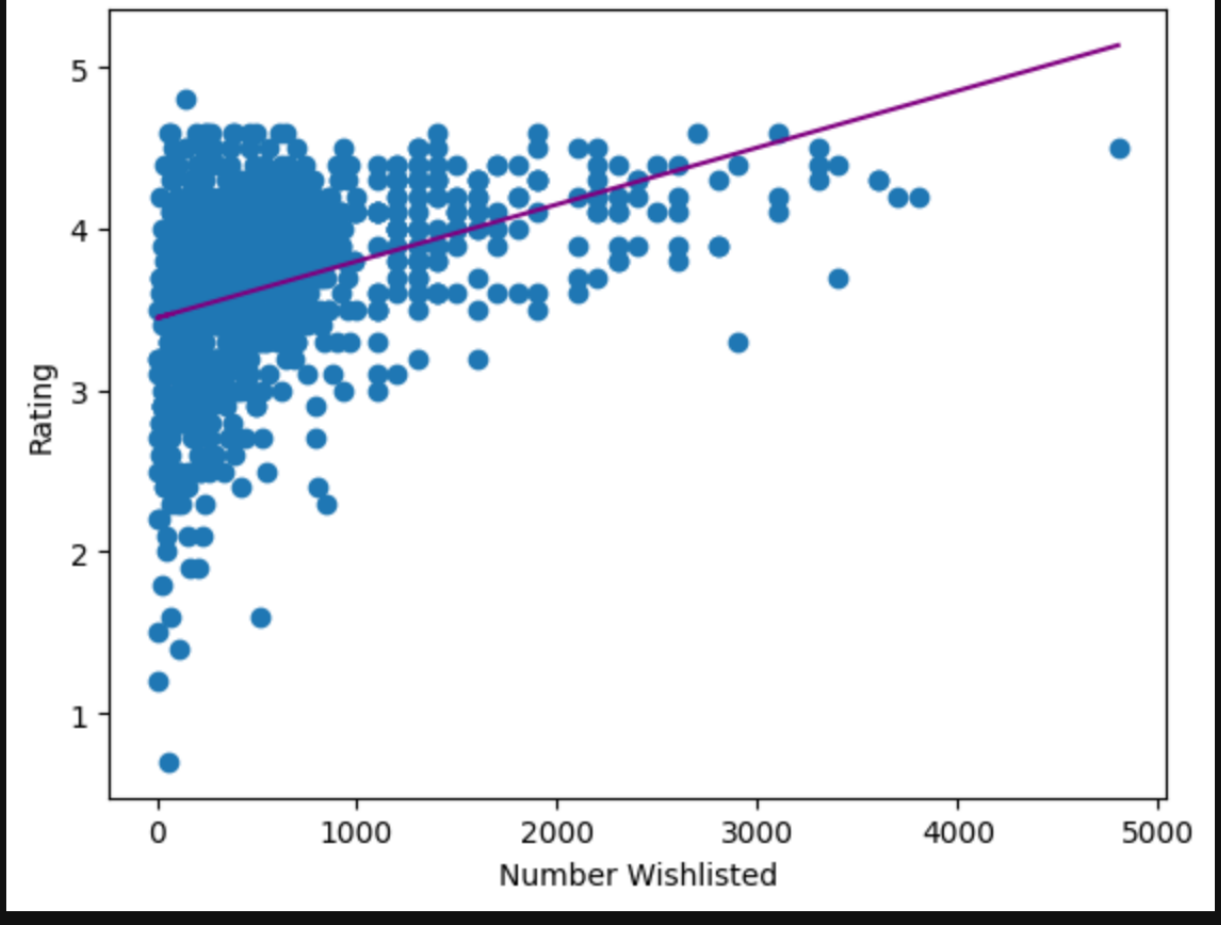
Sony

```
Coefficient of determination accuracy rating below:
0.1525279694945315
```



All

```
Coefficient of determination accuracy rating below:
0.13358219864912801
```



- Please see the screenshots above and the example below of the non-descriptive linear regression method along with the R^2 Score accuracy metric from scikit-learn.
    - The accuracy metric for Nintendo is at .109 which meets the objectives of this data product but could be improved greatly in prediction since R^2 Score is X out of 1.

```
X_wishlist_train, X_wishlist_test, y_ratings_train, y_ratings_test = train_test_split(independent_var, dependent_var,
                                                                                      test_size=0.2, random_state=20)
primaryModel = LinearRegression()

# Training
primaryModel.fit(X_wishlist_train, y_ratings_train)
y_prediction_training = primaryModel.predict(X_wishlist_train)
# accuracyRatingsTraining = r2_score(y_ratings_train, y_prediction_training)


# Testing
y_prediction_testing = primaryModel.predict(X_wishlist_test)
accuracyRatings = r2_score(y_ratings_test, y_prediction_testing)
print("Coefficient of determination accuracy rating below:")
print(accuracyRatings)


# Display
y_prediction = primaryModel.predict(independent_var)
pyplot.scatter(independent_var, dependent_var)
pyplot.plot(independent_var, y_prediction, color="purple")
pyplot.xlabel("Number Wishlisted")
pyplot.ylabel("Rating")
pyplot.show()
```

Application Testing

- The Ratings Prediction data product was tested using the “train_test_split” method. The test size was .2 or 20% and random state was 20.
- The “R^2 Score” accuracy metric was then applied which appears in the dashboard as “coefficient of determination”. The accuracy metric hit above .1 for the three filters (Nintendo, Sony, or All) so no modification was made.
    - In the future, accuracy can and should be increased with further training.

Application Files

- The binder link is as follows: https://mybinder.org/v2/gh/apolefko/CapstoneWGU/HEAD.
- The primary files are in the hosted Binder link.
    - gamesNoDup.csv
    - GUI_Game_Ratings.ipynb
    - requirements.txt
- All libraries and dependencies are in the “requirements.txt” file and will automatically load once the Binder loads.
- The “GUI_Game_Ratings.ipynb” file is the dashboard and contains all code.
- The submission will include this document and the hyperlink listed above (also found in the User Guide).

User Guide

Please see the user guide below. Generally, once the Binder link is clicked and the notebook loads, a user (example: Content Marketer) can click GUI_Game_Ratings and interact with the dashboards.

1. Open your preferred browser (confirmed on Firefox and Chrome).
2. Paste this link https://mybinder.org/v2/gh/apolefko/CapstoneWGU/HEAD and go to the site.
3. Allow the Binder to load (see Fig. D-1).
4. Once the Binder loads, the notebook will appear. This is the dashboard environment for the Ratings Prediction.
5. From the left menu, click GUI_Game_Ratings (see Fig. D-2).
6. The dashboard is now accessible.
7. Each "cell" can be "run" with the play button in the top bar OR the double arrow icon enables all cells (see Fig. D-2).
8. If the double arrow icon is clicked, this will load the code and requirements (dependencies).
9. Once the cells are loaded, the dashboard will now be on a text input box. Follow the instructions to input "Nintendo", "Sony", or "All" to receive an updated linear regression (see Fig. D-3).
10. The accuracy rating using R^2 score will also appear.
11. To see a different model based on Nintendo, Sony, or all Games, simply click the cell with "devrequest" and then the play button at the top bar.

Fig. D-1

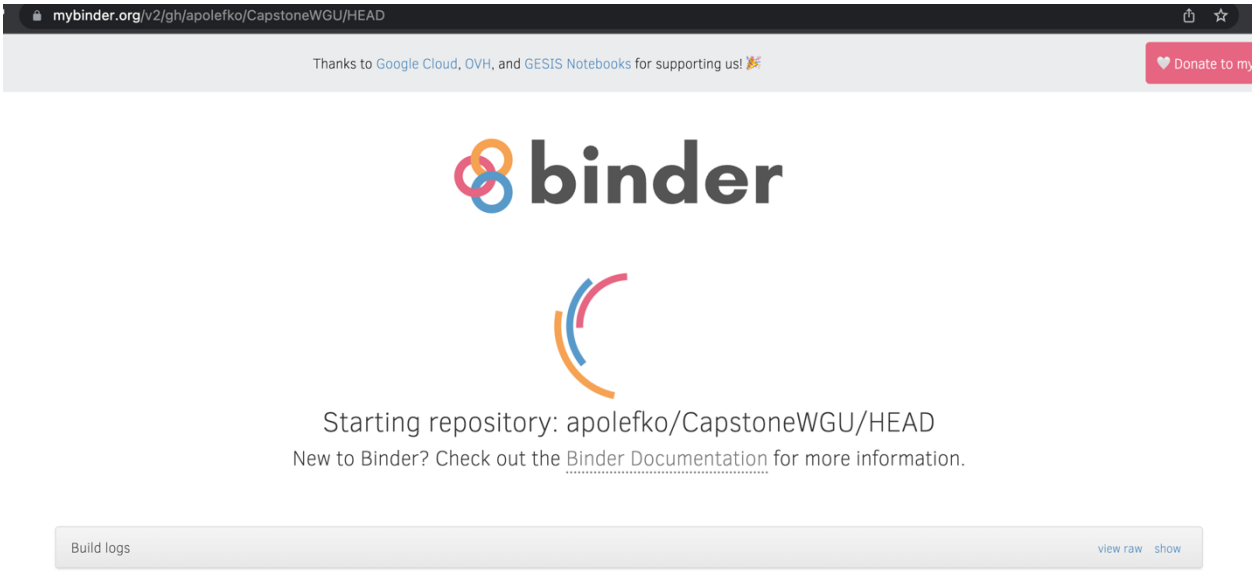


Fig. D-2

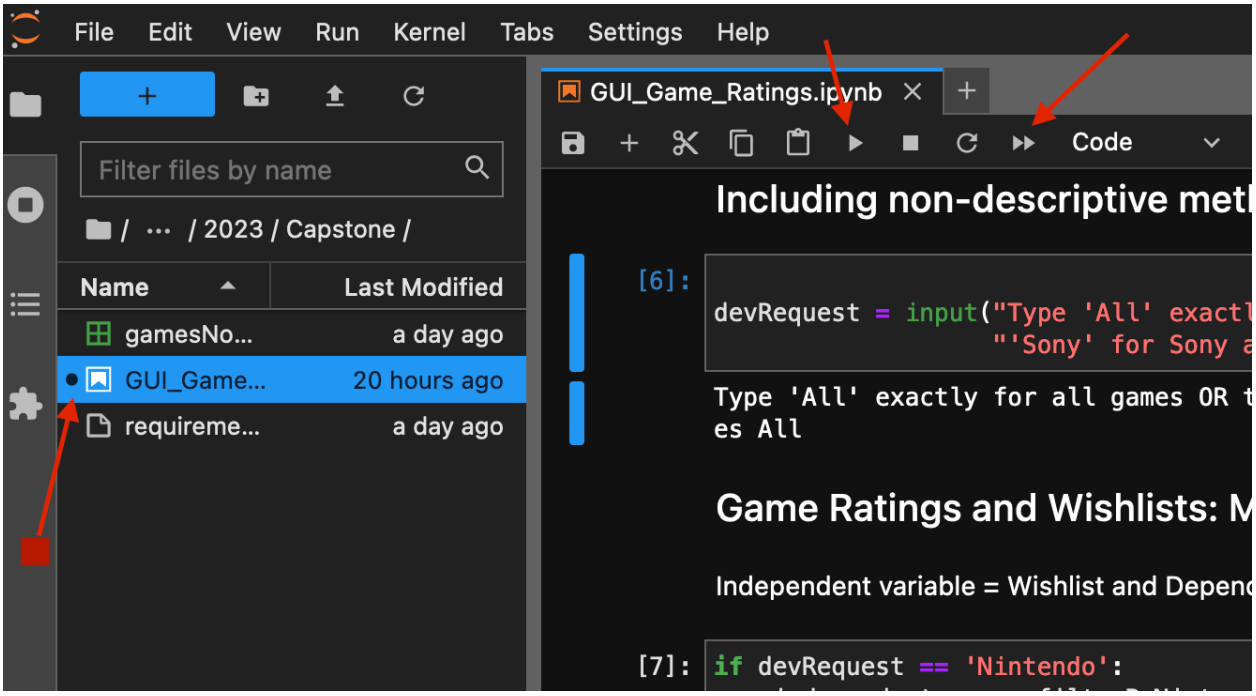
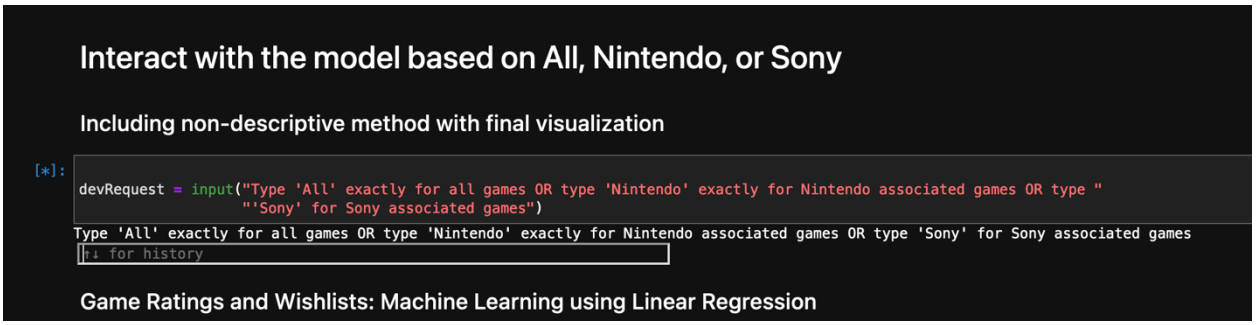


Fig. D-3



Summation of Learning Experience

- My Intro to Artificial Intelligence course at WGU was critical in helping me complete this data product. Additionally, my day job in the Product field at a healthcare company really helped when planning and designing this data product, Ratings Prediction.
- The documentation from scikit-learn, https://scikit-learn.org/stable/index.html, was extremely helpful as well as the pandas documentation, https://pandas.pydata.org/docs/index.html.
- This data product will greatly help my future career endeavors and enabled my understanding and my love of data. With machine learning, I was able to make this data product and tie it to another

passion of mine, videos games. I always want to discover, learn, and explore and this project pushed me to the next level, just like Nintendo!