Blake Calvin and Ryan Moline

1. Introduction

The motivation for this project is that both members of the team are interested in video games and thought that this was a useful project. We also believe that the real life application makes sense for some online game stores to recommend to users future game purchases. There's so many video games out on the market whether they be AAA companies releasing high quality games annually, or indie games that are overlooked, but also just as enjoyable. So a recommender system to find games similar to a specific game sounds like a fun and helpful project to develop.

Some thoughts or questions we had for this project were:

- What dataset should we use that provides information we need?
- How should we organize the data to only provide relevant features?
- Should we create one large recommender or multiple smaller ones specific to certain qualities in games?

These questions drove us to work hard and to give our project plenty of thought before, and during our development process.

A challenge we came across was that there were duplicates for games because they were able to be played on multiple platforms. At first, we were not sure what to do, or if we should remove some of the duplicates until we came to the realization that having the input take the game and the platform would improve results and make the project better. This means that it would give recommendations based on the game criteria, but also ensure it's available for that console.

To summarize our results, we were able to create a program that takes a game title and platform and generates a list of top 5 game recommendations for the user.

```
----------- Top Recommendations for Super Mario Maker on the WiiU ------------
Recommendation 1: Mario Kart 8
Recommendation 2: New Super Mario Bros. U
Recommendation 3: Super Smash Bros. for Wii U and 3DS
Recommendation 4: Splatoon
Recommendation 5: Nintendo Land
```

We think that this was a very successful project as our recommendations seem fairly accurate when it comes to genre. As you can see above when we look for a recommendation for Super Mario Maker played on the WiiU, it recommends many Nintendo games that are fairly similar in genre, ratings, and style.

2. Data Mining Task

The input dataset we chose to work with is a CSV file where each row belongs to a specific video game with columns as features which include a game title, platform, year of release, genre, publisher, NA sales, EU sales, JP sales, other sales, global sales, critic score, user count, developer, and a rating. There are many duplicates of games because the platform column has values which are gaming platforms, like what was hinted at in the introduction and what we will get to in more detail later. However, to briefly summarize, one platform could share game titles with other platforms. Below are a few example columns of our dataset we are working with

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Cou	Developer	Rating
2	Wii Sports	Wii	2006	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53	76	51	8	322	Nintendo	E
3	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24						
4	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52	82	73	8.3	709	Nintendo	E
5	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77	80	73	8	192	Nintendo	E
6	Pokemon Red/Pokemon Blue	GB	1996	Role-Playi	Nintendo	11.27	8.89	10.22	1	31.37						
7	Tetris	GB	1989	Puzzle	Nintendo	23.2	2.26	4.22	0.58	30.26						
8	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.28	9.14	6.5	2.88	29.8	89	65	8.5	431	Nintendo	E
9	Wii Play	Wii	2006	Misc	Nintendo	13.96	9.18	2.93	2.84	28.92	58	41	6.6	129	Nintendo	E
10	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.44	6.94	4.7	2.24	28.32	87	80	8.4	594	Nintendo	E
11	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	0.63	0.28	0.47	28.31						
12	Nintendogs	DS	2005	Simulatio	Nintendo	9.05	10.95	1.93	2.74	24.67						
13	Mario Kart DS	DS	2005	Racing	Nintendo	9.71	7.47	4.13	1.9	23.21	91	64	8.6	464	Nintendo	E
14	Pokemon Gold/Pokemon Silv	GB	1999	Role-Playi	Nintendo	9	6.18	7.2	0.71	23.1						
15	Wii Fit	Wii	2007	Sports	Nintendo	8.92	8.03	3.6	2.15	22.7	80	63	7.7	146	Nintendo	E

At this point, some of the questions we were now faced with were:

- What are the features in this dataset that are going to be helpful for generating a game recommendation
- How can we create recommendations that are relevant to those features we feel are necessary? What ways can we do this that are more advanced than just returning games with the same genre or of similar simplicity?
- How can we effectively handle all the duplicate titles? If we create recommendations for a title with duplicates in the dataset, wouldn't it just give back the same title but for a different platform?

The output data that is used in the program would be five game recommendations that are similar to a specific game title. From there, we print to the console in a listed order, the games that are recommended in order from most similar to least similar. Again, here is the same example of a sample recommendation for Super Mario Maker on the WiiU which puts this idea into a picture.

```
----------- Top Recommendations for Super Mario Maker on the WiiU ------------
Recommendation 1: Mario Kart 8
Recommendation 2: New Super Mario Bros. U
Recommendation 3: Super Smash Bros. for Wii U and 3DS
Recommendation 4: Splatoon
Recommendation 5: Nintendo Land
```

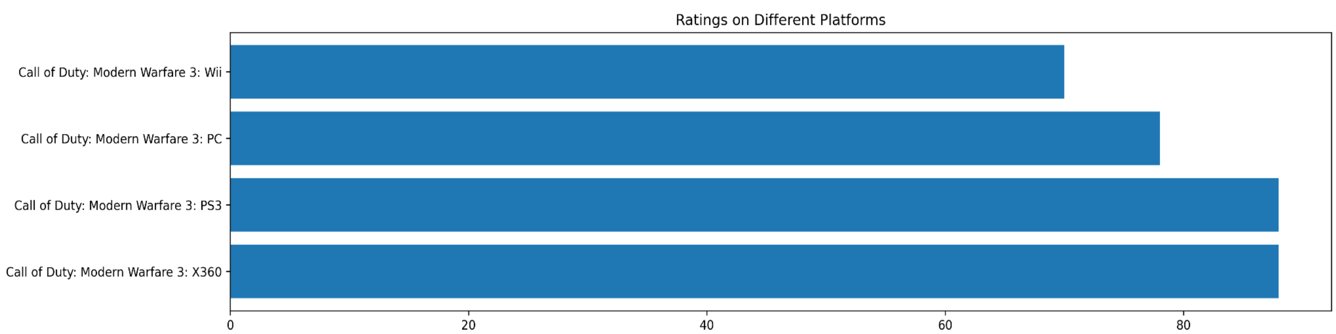
The main challenges were building the recommender system and also dealing with the duplicates of game titles. However, we have set out to create a program that, given an extensive dataset of games, determines the recommendations that are developed with strategies we learned in class.

3. Technical Approach

The overall pseudo code for our program will be the following:

1) Read the dataset
    a) Filter the dataset
2) Create new specific dataframes
3) Create a recommendation column for help later in creating recommendations, this column will have features that are important.
4) Fit the data
5) Create transform and cosine similarity matrices for each data frame
6) Create a function that…
    a) Asks for a game title, and platform for the game
    b) Sorts and filters the cosine similarity matrix to find the top five best fit games considering the input criteria
    c) Print them in the console to see the results

1. First we read in the CSV file containing all the games and save them in a pandas data frame. We remove some of the columns that we feel are not necessary like "Release Year" and "JP_Scores" and keep the important ones such as "Genre", "Global Sales", and some scores/ratings. As mentioned earlier, there are several duplicates of games in the dataset where the difference is the platform so we can use the platform column to differentiate recommendations.



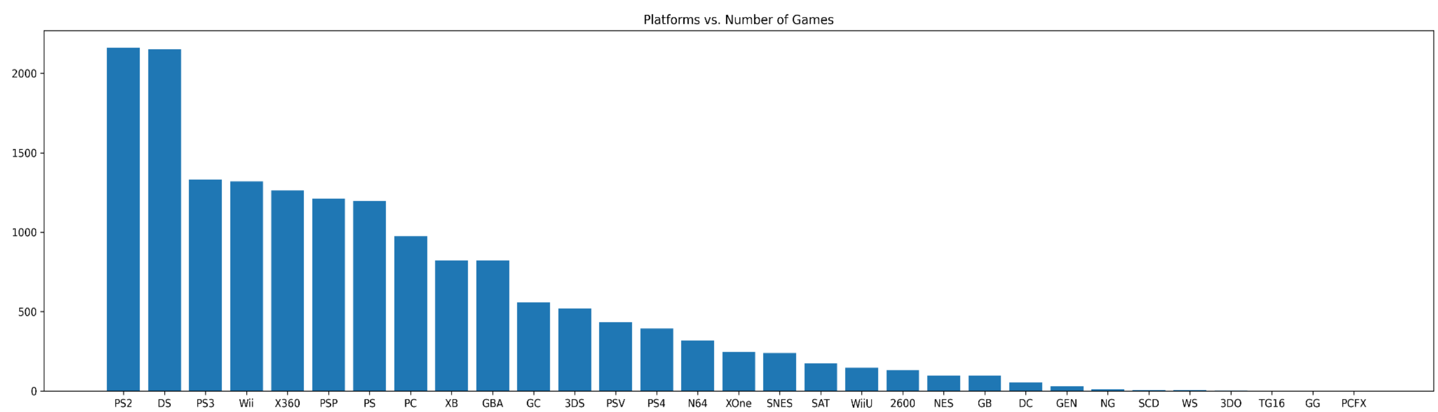
As the picture above describes, Call of Duty: Modern Warfare 3 is released on Wii, PC, PS3, and X360. And not only that, but there also is a difference in global ratings depending on what platform the game is played on. This could be because of many factors, but the two we think make the most sense are

- The demographic of Wii owners might not like the style of games like Call of Duty since Nintendo makes Wii and Nintendo normally makes more family friendly games.
- Some platforms might not run a specific game as well as other platforms, for example X360's hardware might more optimized for running a game like Call of Duty

2. Because of this, we decided we will create multiple different data frames, where each data frame is specific to one platform, so only PC games will be listed in the PCData dataframe, and the same applies for other platforms like 3DS, WiiU, etc.

	Name	Platform	Genre	Publisher	Global_Sales	Critic_Score	User_Score	Rating
969	Pony Friends 2	PC	Simulation	Eidos Interactive	0.010000	nan	nan	nan
970	Metal Gear Solid V: Ground Zeroes	PC	Action	Konami Digital Entertainment	0.010000	80.000000	7.600000	M
971	Breach	PC	Shooter	Destineer	0.010000	61.000000	5.800000	T
972	STORM: Frontline Nation	PC	Strategy	Unknown	0.010000	60.000000	7.200000	E10+
973	15 Days	PC	Adventure	DTP Entertainment	0.010000	63.000000	5.800000	nan

Above shows that we were able to separate games based on their platform, and so the diagram is an example of what some of the rows look like for the PC platform dataframe. We chose to only work with the following platforms: PC, PS4, XOne, WiiU, 3DS, and N64. This is because all these platforms (except N64) were new for the time this dataset was released (2016), so less knowledge on the games they contain would be out there in the world, so we wanted to focus on them to help people get an idea of some of the games that are similar to games like but don't know what game to play next. This picture shows that fact in a visual.



N64 isn't a new platform however, and is in our project simply because we both enjoyed that console in our youth and wanted to have fun adding that in as well.

3. We then add a new column to all the filtered data frames that is used for recommendations. This new column involves features from each game we thought was necessary to factor in when creating recommendations.

4. We fit the data now into a Vectorizor and call the fit_transform function on the recommended column we created.

5. Afterwards, we create the transform matrix for each platform data frame using the recommended column we created. After the transform matrix is made, we use it to create a cosine similarity matrix to actually have a numeric similarity value for all the games in each dataframe.

6. Next, we create a function that handles all the recommendation logic. This function takes in a game title, a cosine similarity matrix, and a platform. For example, if you wanted recommendations for Borderlands on the PC, you would enter “Borderlands”, PCcsMatrix, and “PC” into the input parameters for our GameRecommendation function. In this function we sort out and filter the top scores for different titles of games compared to the game that is passed in to the function. We then loop through the top five scores and print the name and the order number.

After all of that, we should have our desired output we are looking for. This will be done by making good use of pandas dataframes and the scikitlearn functions that help with generating everything we need in our steps listed above. This helped us a lot with dealing with this data mining task because it answered a lot of the questions we had and gave optimal solutions to satisfy our specification and requirements.

4. Evaluation Methodology

As mentioned earlier, all the columns in the dataset were:

Name,Platform,Year_of_Release,Genre,Publisher,NA_Sales,EU_Sales,JP_Sales,Other_Sales,Global_Sales,Critic_Score,Critic_Count,User_Score,User_Count,Developer,Rating

We got this dataset from Kaggle and we decided on this dataset because we felt that it would be a fun idea to recommend video games and this dataset was not too big to easily test, and big enough that it tested our ability to work with large datasets. At first we thought the duplicates on different platforms was going to be an issue, however, it ended up improving our project by adding another layer of work that made us check the user’s platform before making a recommendation. After we pruned the data, we only used the columns:

Name, Platform, Genre, Publisher, Global_Sales, Critic_Score, User_Score, and Rating.

We decided to trim the dataset down because we believed that a lot of the other pieces of data were not things that we felt were necessary for making the correct recommendation. Although information such as North America Sales vs Europe Sales could have a slight impact on game recommendations, we believed this to not be important for our project goal that we had in mind.

5. Results and Discussion

```
---------- Top Recommendations for The Sims: Unleashed on PC -----------
Recommendation 1: The Sims 3
Recommendation 2: World of Warcraft
Recommendation 3: Diablo III
Recommendation 4: Microsoft Flight Simulator
Recommendation 5: StarCraft II: Wings of Liberty
------------------------
---------- Top Recommendations for Call of Duty: Ghosts on the XboxOne -----------
Recommendation 1: Call of Duty: Black Ops 3
Recommendation 2: Grand Theft Auto V
Recommendation 3: Call of Duty: Advanced Warfare
Recommendation 4: Halo 5: Guardians
Recommendation 5: Fallout 4
------------------------
---------- Top Recommendations for Call of Duty: Ghosts on the Playstation4 -----------
Recommendation 1: Call of Duty: Black Ops 3
Recommendation 2: Grand Theft Auto V
Recommendation 3: FIFA 16
Recommendation 4: Star Wars Battlefront (2015)
Recommendation 5: Call of Duty: Advanced Warfare
------------------------
---------- Top Recommendations for Super Mario Maker on the WiiU -----------
Recommendation 1: Mario Kart 8
Recommendation 2: New Super Mario Bros. U
Recommendation 3: Super Smash Bros. for Wii U and 3DS
Recommendation 4: Splatoon
Recommendation 5: Nintendo Land
------------------------
---------- Top Recommendations for Pokemon Omega Ruby/Pokemon Alpha Sapphire on the 3DS -----------
Recommendation 1: Pokemon X/Pokemon Y
Recommendation 2: Mario Kart 7
Recommendation 3: Pokemon Omega Ruby/Pokemon Alpha Sapphire
Recommendation 4: Super Mario 3D Land
Recommendation 5: New Super Mario Bros. 2
------------------------
---------- Top Recommendations for Donkey Kong 64 on the N64 -----------
Recommendation 1: Super Mario 64
Recommendation 2: Mario Kart 64
Recommendation 3: GoldenEye 007
Recommendation 4: The Legend of Zelda: Ocarina of Time
Recommendation 5: Super Smash Bros.
------------------------
```

This is what our results look like. We called our GameRecommendation function on five games for six different platforms. The reason being for only five games instead of six, is to show the interesting fact mentioned earlier that not all games are rated the same or perform the same success depending on what platform it's played on. So For XOne and PS4, we ran the same title "Call of Duty: Ghosts" in the recommender algorithm and we got different recommendations. We can see that they both share a top recommendation spot, but the rest are shuffled in different order or have completely different games in them, like how Advanced Warfare is the third recommendation for XOne but only fifth for PS4. This is super interesting to us because we never really thought about this up until this project that some games might not be recommended as highly depending on the platform you play it on.

Lastly, something else to note is that each game in the recommendations is specific for the platform they are being recommended with. What we mean by that is there's no Wii games in the

Xbox recommendations or vice versa, all N64 game recommendations are N64 games, all 3DS game recommendations are 3DS games, etc.

6. Lessons Learned

We've both worked with pandas in the past, but this project also helped us get more used to working with pandas dataframes. We also were able to implement a lambda function in our generation for the Recommended column which helped us learn more about the use of lambda functions and how it can make code much cleaner and more concise. We also learned how to use scikitlearn and how it can be used to make machine learning programs. Neither of us really had much machine learning experience so we learned how useful a recommender application can be and how a simple recommender can be created without too much experience or extensive knowledge.

One thing we valued highly in the lessons we learned is to work on a project that you are honestly interested in because we started out with a different project in mind and pivoted because we were finding it difficult to be motivated to work on. Once we found a project we were passionate about it became much easier to create the solution and gave us more passion to create a quality program. We think this can really be a life lesson for as we are graduating and starting our careers, to apply for jobs that we truly find interesting, and isn't just a big paycheck. It will make our quality of life much higher in the future and is something we will remember as we are leaving our college chapter in life.

In hindsight, and if we had more time, implementing some of the other variables in the dataset would have been fun and might have made our recommender more accurate. Things such as sales in different regions to give a better recommendation if the user is able to provide the region they are from. Also maybe having some recommendations based on developers because we mostly use genre and overall ratings because we felt that was a good start.

7. Acknowledgements

We would like to thank the professor and all the TA's for this semester and helping us learn more about data mining!