Analysis of Video Games reviews and Sales data

Introduction

Team

Under Prof. Jan-Willem van de Meent
Done by: Sumeet Dubey, Harshdeep Singh, Zhiguang Yu

Dataset

We are going to use a dataset of video game sales for our project. Below mentioned is the link to the dataset: https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings.
This dataset has fields like Name of the game, platform, year, publisher, along with sales values across North America, Europe, Japan and global sales as well. Also, we have user ratings and critic ratings available in the dataset.
Per our discussion the other day, we tried to find a supplementary dataset that would help in applying algorithmic analysis and produce some prediction models. We found this dataset that has game reviews by reviewers at Gamespot (www.gamespot.com). Though this does not contain every entry we have in our original dataset, it is of reasonable size to do analysis.

Data Mining Tasks to be performed

We will be performing basic exploratory analysis of the above data set in Milestone-1 and for Milestone-2 going to perform following algorithms:

- FP- Growth Algorithm to find Frequent item sets in games review dataset.
- Perform PCA on the games review dataset and plot the results to see if any conclusion can be drawn about the dataset.
- Implement Topic modelling to identify common topics amongst different documents or in this case game reviews.
- Perform Random forest and PCA Results on text data from Game Reviews Database.

Project Report – Milestone 1

Exploratory Analysis Over Dataset

Top companies in the gaming industry

- We calculated total Global sales for each publisher over all the games they have released and took the top 10 publishers.

Conclusion:
Nintendo seems to be a leader in sales with a good margin. All of the top companies are amongst the current big video game manufacturers, which was expected.

- We calculated the number of games released by each publisher and took the top 10 publishers with the maximum number of releases

Conclusion:
Interestingly, Nintendo has produced about half as many games as EA though it beats EA in global sales. Activision and Namco Bandai have released almost same number of games, but there is a considerable difference in their sales.
Also, note that both graphs contain the same set of companies in different order.

Popular Genres

Conclusion:
Action games have dominated gaming industry for a long time and have the maximum number of sales. Sales in Japan though for action games is the lowest compared to other places. But we can see a nice bump in Japan sales for Role-Playing games (in-fact the highest). We could say that the Japanese gaming industry prefers story-telling and role oriented games than genres like Action and Shooters

User/Critic Scores for top publishers

We calculated the mean User and Critic scores for each of our top 10 publishers.

Conclusion:
Nintendo seems to have the highest mean User/Critic rating which should explain it's top global sales. Also 7/10 have higher mean scores given by Users rather than Critics.

Distribution of Game Ratings:

ERSB ratings are given according to the content of the game to filter the users suitable for playing it. The popular ones are:

- E – Everyone
- E10+ - Everyone above 10
- T – Teen
- M - Mature

Conclusion:
Games rated E are the most made and most sold games. This is expected as many users of the gaming industry are kids and there are many genres that do not require unnecessary explicit content (sports, strategy, puzzles). Rated-M games have great sales compared to their strengh. Many of the Action and Role-playing games are often rated M and hence they are popular. Equally popular are Teen rated games, though their number is higher.

Note that only one game in our dataset has the AO rating (Adults Only). Interestinly it can be noted it the second graph that this game generated a profit of $1.95 million. It is the popular and critically aclaimed game by Take Two Interactive and Rockstar, Grand Theft Auto: San Andreas.

Above is a plot of sales per game for each rating. The bump for AO is because of GTA: San Andreas as we figured.

Sales Distributions for different Genres across different regions

Find top 5 publishers in the database based on global sales and then find out sales distributions for each genre across North America, Europe, Japan, Rest of the world(Others).

Conclusion
Top 5 publishers based on the global sales are: Electronic Arts, Ubisoft, Sony Computer Entertainment, Activision and Nintendo.
From the above plots, we can see that:

1. North America is the biggest market for games in Platform genre followed by Strategy Genre. With Electronic arts and Nintendo being the major publishers.
2. Nintendo is has done better sales in Japan (based on sales) as compared to the other regions.
3. Nintendo has sales for almost all its genres whereas other publishers are bit inconsistent across the genres and have sales for one or two specific genres.
4. But overall North America and Europe is the main hub for video games sales in the world.

Trends for Sales per year for top 5 publisher

For different regions plot trends for sales over the years for each of the top 5 publishers based on global sales. We have considered years as buckets of 5 years, tried to plot total sales for each of those buckets in North America, Europe, Japan and Rest of the world (Other).

Conclusion
Top 5 publishers based on the global sales are: Electronic Arts, Ubisoft, Sony Computer Entertainment, Activision and Nintendo.
From the plots, above we can see that:

1. There has been a sharp rise(increase) in sales for Nintendo for all regions in period 2005-2009.
2. Nintendo has been the oldest publisher in the world, accompanied by Activision whereas the other three publishers came into the market at later stages.
3. Sony Computer Entertainment had good sales in first 10 years but then it has seen continuous decline in sales across all regions.
4. We can also see that all the publishers had highest number of sales during the years 2005-2009 and sales fell drastically for each the publisher in all the regions for the corresponding period (2010-2014).
5. Also, third plot strengthens our claim in previous objective that Nintendo has performed exceptionally in Japan as compared to other publishers.

Most Selling Console Each Year:

Conclusion:
People from 1995 to 2015 prefer action games.

Conclusion:
PlayStation series is the most selling platform. From 2005 to 2010, the wii and X-BOX 360 take the place of most selling. However, from 2012 to 2015, the Playstation comes back to the most selling position.

Project Report – Milestone 2

Performing Data Mining Techniques

Topic Modeling with LDA:

We tried conducting topic modeling on our database of game reviews. We considered the GameSpot written reviews and ran it on different subsets of the dataset.

Preprocessing:

Our data organized by different consoles with a txt file for each game. We created a pandas data frame storing relevant information as columns. We used regex to find tags and store for each game its name, publisher, GameSpot score and written review.
Before conducting LDA, we vectorized the input into word vectors.

Observations:

- First we ran LDA on all PlayStation 3 games. We are able to identify some of the topics common to gaming genres that have been identified.
- There are a bunch of words like vehicle, racing suggesting a cluster of racing games. Yet another topic has words like character, story, enemies; suggesting action and adventure games.
- We then ran LDA specifically on good and bad rated games. First we ran it on all games that have a minimum score of 8/10.
- We are able to identify some interesting (and rather positive) topics with words such as fun, story, good, adventure, new. We also found topics specific to racing and soccer games, suggesting the good Fifa games and PES series.
- When we ran the tests on the low rated games (rating <= 3), we found some negative words and topics. Reviewers are identifying some bad features or poor design of the games by using words such as problems, crazy, unoriginal, bad.

Further work:

We aim to do a more comprehensive study of topic models under different conditions. We will be running it on some top companies and compare the results. Also we will extend topic modeling to user reviews (a larger dataset) and see how their features differ under various scenarios.

FP Growth and PCA Results on text data from Game Reviews Database

Objective

To find frequent item sets in the database, basically to find frequent words existing in the database that occur together. Main motive behind finding frequent item sets is that to see if we can figure out whether the frequent occurring words in the reviews tell something about the game and can we conclude about the genre, critic response to the game etc.
We ran PCA on the datasets to see if PCA tells us anything about how the data is spread across the different components and is there any conclusive relationship that can be drawn from PCA feature plotting or not.

Datasets: We have datasets for about 8000 games divided across 12 different platforms. But we choose 3 platforms namely PSN, PS2, PS3 and Xbox to perform the above exercise. The reason for choosing these three was basically the size constraint as we were not able to run FP-Growth for large data sets as our systems were not able to handle such large data. Total number of games covered were about 2400 games.

Results

FP-Growth: We tried to perform FP growth algorithm for finding Frequent Item sets and any pattern that could exist in the comments that could help us figure out some of the features of the game. After looking at the results across different datasets (mentioned above) we found that there is no clear conclusion that can be drawn from them.
There are instances where we can find words like (fun, like, game) words together and words like (playstation, game) etc but there is no clear pattern that can be seen here.
Maybe we need to do some more cleaning of the data or use some other kind Data mining techniques.

PCA-Results: PCA results for the above datasets, supports our findings about the FP growth.
If you see the appendix attached at the end, we have PCA results for each of the datasets. It shows that there is no specific pattern that can be seen there. We performed PCA with 20 components and when plotted them, we could not see any sudden drop in the variance caused by the components which we generally use to decide our components in PCA dimensionality reduction. And if we plot by taking top 2 components we can see that the data is mostly concentrated at (0,0) point in the space with no clear division and clustering. So, we can say that performing PCA on our dataset is not a useful task and we might need to use some other approach to see any kind of trends in the data.

Random forest and PCA Results on text data from Game Reviews Database

Objective

For the dataset, we previously reviewed sales across the world but there are so many other features including 'Platform','Genre','Publisher','Year_of_Release','Critic_Score','Global_Sales'. Now, we are trying to explore the weight of each feature in deciding the global sales. And based on the features, we try to find the possibility of a non-popular game to be a popular game. We define the popular game to be the one which sold more than one million dollars.

Datasets: We have only considered those records in the dataset where we have information for all the features. If any record has null value for any of the features, we have not considered it.

Results:

Multilayer perceptron neural network: We split the dataset into training and testing set. Then after training the model, we find that the correctness is 0.8274213836477987.

Random forest: After training the model, we find the correctness is 0.8510691823899371. So, we move forward with this algorithm. Then we come up with few results. We list the importance of features influencing the popularity. And found that the **Critic_Score , Year_of_Release, Publisher_Nintendo** are the top features deciding the popularity of games. Then we used the trained Random Forest model to predict the possibility of a non-popular game to be popular. We get a list of result showing the percentage. From this, we can see the potential popular games and better understand how the sales works. The results are available in the Appendix below.

APPENDIX

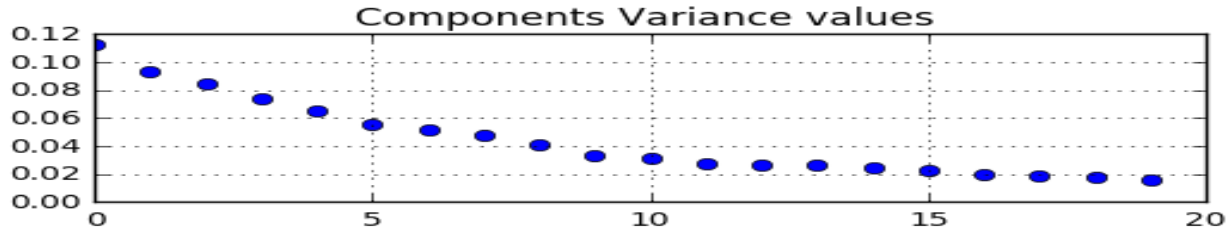
PSN - FP Growth results:

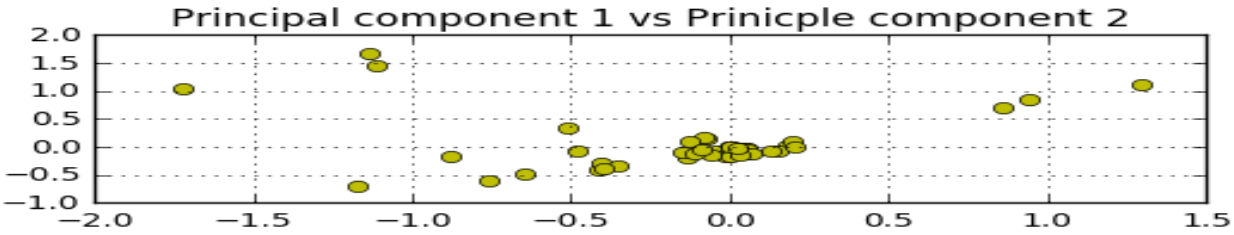
```
(['fun'], 36)
(['fun', 'game'], 35)
(['fun', 's'], 36)
(['playstation'], 36)
(['playstation', 't'], 35)
(['playstation', 't', 'game'], 34)
(['playstation', 't', 'game', 's'], 34)
(['playstation', 't', 's'], 35)
(['playstation', 'game'], 35)
(['playstation', 'game', 's'], 35)
(['playstation', 's'], 36)
(['play'], 45)
(['play', 'game'], 44)
(['play', 'game', 's'], 44)
(['play', 'like'], 40)
(['play', 'like', 'game'], 39)
(['play', 'like', 'game', 's'], 39)
(['play', 'like', 's'], 40)
(['just'], 42)
(['just', 's'], 42)
(['game'], 54)
(['game', 's'], 54)
(['make'], 36)
(['make', 'game'], 36)
(['make', 'game', 's'], 36)
(['make', 'like'], 34)
(['make', 'like', 'game'], 34)
(['make', 'like', 'game', 's'], 34)
(['make', 'like', 's'], 34)
(['make', 's'], 36)
(['level'], 36)
(['level', 't'], 35)
(['level', 't', 'game'], 34)
(['level', 't', 'game', 's'], 34)
(['level', 't', 's'], 35)
(['level', 'game'], 35)
(['level', 'game', 's'], 35)
(['level', 's'], 36)
(['like'], 50)
(['like', 'game'], 49)
(['like', 'game', 's'], 49)
(['like', 's'], 50)
(['time'], 42)
(['time', 't'], 41)
(['time', 't', 'game'], 40)
(['time', 't', 'game', 's'], 40)
(['time', 't', 's'], 41)
(['time', 'll', 's'], 36)
(['time', 'play'], 34)
(['time', 'play', 's'], 34)
(['time', 'like'], 39)
(['time', 'like', 't'], 38)
```

```
(['time', 'like', 't', 'game'], 37)
(['time', 'like', 't', 'game', 's'], 37)
(['time', 'like', 't', 's'], 38)
(['time', 'like', 'game'], 38)
```

PSN-PCA results:

```
{0: ['m', 'play', 'commando', 'sequences', 'enjoyable', 'explosion', 'solid',
'hour', 'don', '3', 'action', 'intense', 'shooter', 'enemies', 'big'], 1:
['enemies', '3', 'm', 'commando', 'crash', 'enjoyable', 'animated', 'story',
'solid', 'explosion', 'bug', 'overlay', 'wolf', 'shooter'], 2: ['enemies',
'don', 'm', 'play', 'sequences', 'big', 't', 'crash', 'experience',
'animated', 'explosion', 'action', 'overlay', 'problems', 'hour'], 3:
['character', 'characters', 'don', 'online', 'play', 'direction',
'sequences', 'choose', 'enjoyable', 't', '80s', 'action', 'bug', 'problems',
'hour'], 4: ['weapons', 'xbox', '3', 'online', '360', 'good', 'action',
'just', 'bug', 'negligible', 'game'], 5: ['weapons', 'enemies', 'characters',
'360', 'online', 'sequences', 'good', 'big', 'choose', 'way', 'modest',
'80s', 'action', 'bug', 'game'], 6: ['weapons', 'character', 'enemies',
'characters', '360', 'direction', 'sequences', 'good', 'big', 'choose',
'modest', '80s', 'action', 'pulling', 'bug'], 7: ['weapons', 'character',
'play', 'direction', 'big', 'way', 'experience', '80s', 'action', 'just',
'satisfying', 'levels', 'game'], 8: ['character', 'direction', 'way',
'pulling', 'right', 'levels'], 9: ['weapons', 'enemies', 'sequences', 'way',
'experience', 'story', 'action', 'just', 'wolf', 'satisfying', 'levels',
'game'], 10: ['weapons', 'enemies', '360', 'online', 'm', 'good', 'crash',
'way', '80s', 'explosion', 'bug', 'game'], 11: ['enemies', 'online',
'sequences', 'good', 'big', 'way', 'modest', '80s', 'bug', 'wolf', 'levels',
'game'], 12: ['wolf', 'story', 'shooter', 'bug', 'play'], 13: ['experience',
'story', 'bug', 'just', 'wolf', 'satisfying', 'shooter'], 14: ['online',
'play', 'way', 'experience', 'story', 'action', 'intense', 'just',
'negligible', 'bug', 'hour', 'shooter', 'levels', 'wolf'], 15: ['big',
'sequences', 'good', 'modest', 'experience', 'story', '80s', 'bug',
'satisfying', 'levels', 'game'], 16: ['weapons', 'online', 'play', 'story',
'bug', 'negligible', 'wolf', 'shooter'], 17: ['weapons', 'play', 'sequences',
'way', 'modest', 'story', '80s', 'action', 'bug', 'levels', 'game'], 18:
['weapons', 'enemies', 'sequences', 'way', 'modest', '80s', 'action',
'levels', 'game'], 19: ['weapons', 'direction', 'modest', '80s', 'bug',
'pulling', 'right']}
```





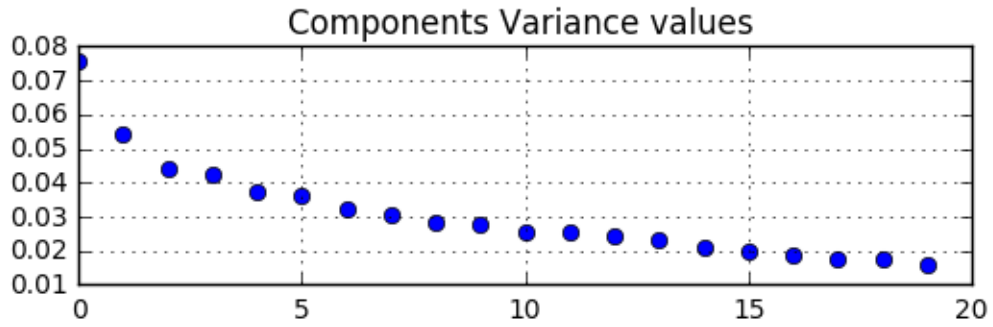
PS3 – FP Growth few important words

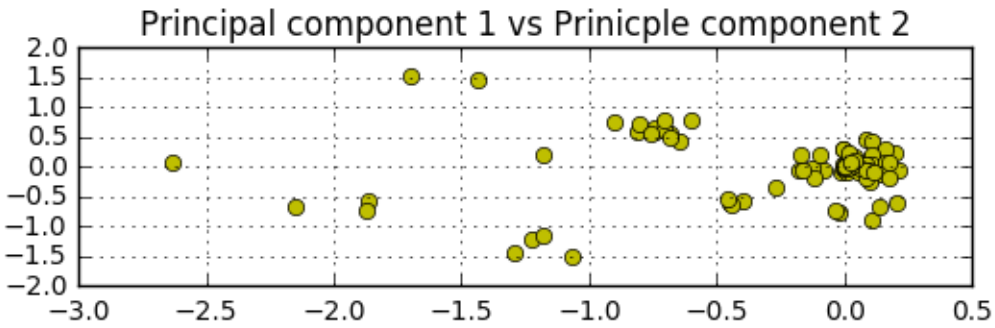
```
(['player', 'like', 'game'], 30)
(['fun', 'time', 'game'], 28)
(['quite', 'good'], 113)
(['quite', 'good', 'game'], 113)
(['bit', 'good', 'game'], 118)
```

(['bit', 'new', 'game'], 117

PS3 - Top 20 components:

```
{0: ['rarely', 'features', 'naruto', 'comes', 'problem', 'play', 'use',
'ninja', 'feel', 'lot', 'feels', 'ultimate', 'little', 'people', 'simply'],
1: ['rarely', 'story', 'comes', 'play', 'ninja', 'lot', 'ultimate', 'little',
'simply', 'highlight', 'action', 'storm', 'leaf', 'villagers', 's'], 2:
['great', 'locked', 'impossible', 'story', 'use', 'feel', 'feels', 'people',
'highlight', 'village', 'leaf', 'minigames', 'completely', 'arenas',
'doors'], 3: ['minimum', 'way', 'guard', 'story', 'characters', 'play',
'hours', 'mode', 'option', 'people', 'highlight', 'spend', 'village', 'long',
'10'], 4: ['locked', 'guard', 'story', 'perform', 'comes', 'problem', 'play',
'characters', 'feel', 'feels', 'jutsu', 'certain', 'option', 'highlight',
'fight'], 5: ['fighting', 'controls', 'minigames', 'completely', 'arenas',
'highlight', 'different', 'block', 'minutes', 'story', 'chakra', 'lands',
'battle', 'scrolls', 'items'], 6: ['mode', 'controls', 'ultimate', 'mission',
'minigames', 'arenas', 'people', 'highlight', 'camera', 'missions', 'option',
'story', 'great', 'guard', 'scrolls'], 7: ['way', 'story', 'comes',
'problem', 'characters', 'unlock', 'feel', 'certain', 'support', 'people',
'highlight', 'minutes', 'fight', 'spend', 'minigames'], 8: ['long',
'minigames', 'completely', 'arenas', 'highlight', 'guard', 'doors', 'locked',
's', 'story', 'roster', 'll', 'scrolls', 'rest', 't'], 9: ['combo',
'different', 'story', 'characters', 'problem', 'feel', 'feels', 'opponent',
'support', 'people', 'highlight', 'minutes', 'battle', 'spend',
'completely'], 10: ['combo', 'guard', 'impossible', 'story', 'play', 'feel',
'hours', 'challenges', 'feels', 'opponent', 'jutsu', 'option', 'highlight',
'include', 'minigames'], 11: ['opponent', 'certain', 'completely',
'minigames', 'arenas', 'people', 'highlight', 'combo', 'minutes', 'lengthy',
'scrolls', '30', 'comes', 'quiet', 've'], 12: ['locked', 'guard', 'features',
'story', 'feel', 'lot', 'feels', 'looking', 'people', 'highlight', 'place',
'long', 'minigames', 'completely', 'missions'], 13: ['minimum', 'guard',
'impossible', 'story', 'comes', 'problem', 'play', 'characters', 'feel',
'hours', 'feels', 'certain', 'option', 'people', 'highlight'], 14: ['rarely',
'job', 'does', 'free', 'great', 'camera', 'story', 'feel', 'feels',
'opponent', 'stops', 'people', 'battle', 'action', 'combatants'], 15:
['combo', 'guard', 'impossible', 'play', 'feel', 'lot', 'feels', 'opponent',
'option', 'people', 'highlight', 'place', 'minigames', 'completely',
'quiet'], 16: ['opponent', 'minigames', 'completely', 'mapped', 'arenas',
'people', 'combo', 'highlight', 'impossible', 'lengthy', 'story', 'chakra',
'end', 'scrolls', 'pad'], 17: ['minigames', '10', 'completely', 'like',
'arenas', 'highlight', 'way', 'camera', 'minimum', 'free', 'story',
'impossible', 'minutes', 'scrolls', '30'], 18: ['minimum', 'combo', 'way',
'camera', 'great', 'impossible', 'story', 'characters', 'feel', 'hours',
'feels', 'opponent', 'certain', 'people', 'highlight'], 19: ['minimum',
'combo', 'way', 'guard', 'impossible', 'story', 'charging', 'powerful',
'feel', 'hours', 'd', 'feels', 'opponent', 'certain', 'mapped']}
```

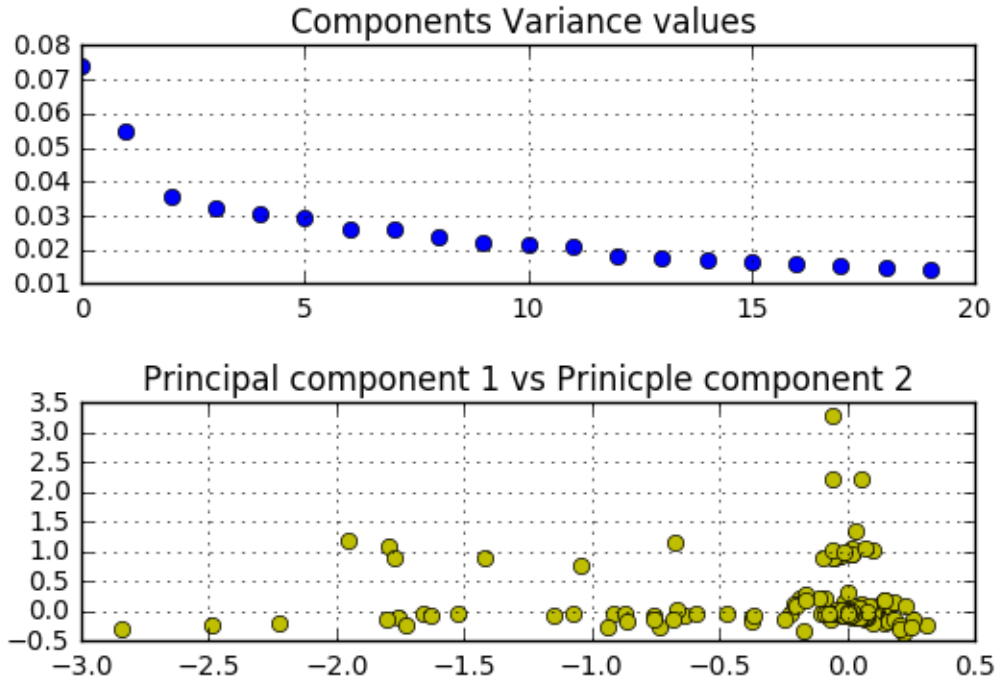




Xbox- PCA Results:

```
{0: ['zones', 'year', 'offense', 'new', 'legacy', 'thing', 'attributes',
'excellent', 'looks', 'fun', 'hoops', 'school', 'visuals', '2k6', 'lack'], 1:
['year', 'hand', 'mike', 'state', 'does', 've', 'clear', 'lets', 'quick',
'depending', 'legacy', 'fight', 'graphics', 'fun', 'sound'], 2: ['hand',
'means', 'lets', 'games', 'quickplay', 'depending', 'plays', 'accurate',
'shot', 'left', 'pretty', 'graphics', 'right', 'sound', 'execute'], 3:
['lets', 'simulation', 'legacy', 'plays', 'accurate', 'thing', 'open',
'attributes', 'playbooks', 'pretty', 'example', 'including', 'ncaa', 'looks',
'graphics'], 4: ['hand', 'lets', 'quickplay', 'plays', 'available', 'fight',
'easily', 'options', 'songs', 'different', 'player', 'lack', 'defense',
'quite', 'practice'], 5: ['hand', 'new', 'lets', 'legacy', 'plays',
'example', 'ncaa', 'graphics', 'fun', 'sound', 'visuals', 'player', '2k6',
'lack', 'college'], 6: ['want', 'ball', 'new', 'lets', '10', 'quickplay',
'plays', 'shot', 'open', 'traditional', 'ncaa', 'look', 'graphics', 'sound',
'big'], 7: ['zones', 'hand', 'new', '10', 'games', 'simulation', 'quickplay',
'plays', 'available', 'accurate', 'open', 'left', 'playbooks', 'traditional',
'ncaa'], 8: ['hand', 'means', 'lets', '10', 'quickplay', 'games', 'plays',
'accurate', 'left', 'traditional', 'ncaa', 'look', 'graphics', 'right',
'sound'], 9: ['new', 'ball', 'lets', '10', 'games', 'quickplay', 'plays',
'accurate', 'open', 'left', 'traditional', 'look', 'big', 'school',
'player'], 10: ['zones', 'hand', 'half', 'lets', 'games', 'quickplay',
'plays', 'accurate', 'shot', 'open', 'left', 'playbooks', 'pretty', 'ncaa',
```

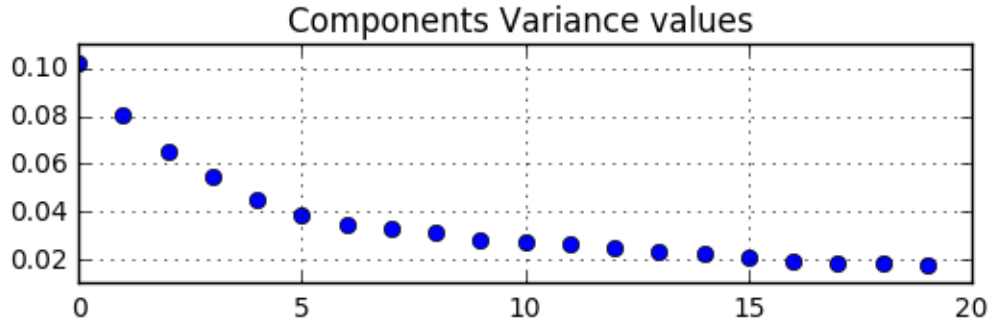
'ranging'], 11: ['hand', 'offense', 'new', 'lets', 'games', 'quickplay',
'plays', 'traditional', 'xbox', 'ncaa', 'look', 'graphics', 'sound',
'visuals', 'player'], 12: ['offense', 'include', 'new', 'lets', '10',
'games', 'quickplay', 'depending', 'plays', 'accurate', 'attributes',
'general', 'traditional', 'ncaa', 'graphics'], 13: ['zones', 'year', 'hand',
'lets', '10', 'games', 'plays', 'accurate', 'open', 'example', 'traditional',
'xbox', 'look', 'graphics', 'sound'], 14: ['zones', 'hand', 'include',
'want', 'new', '10', 'games', 'quickplay', 'shots', 'depending', 'plays',
'accurate', 'open', 'attributes', 'general'], 15: ['zones', 'hand', 'new',
'lets', '10', 'shots', 'plays', 'shot', 'open', 'general', 'clock',
'traditional', 'pretty', 'ncaa', 'graphics'], 16: ['hand', 'offense',
'include', 'want', 'lets', 'games', 'quickplay', 'depending', 'plays',
'accurate', 'shot', 'open', 'left', 'ncaa', 'including'], 17: ['year',
'hand', 'include', 'want', '10', 'games', 'quickplay', 'plays', 'accurate',
'traditional', 'example', 'presses', 'big', 'execute', 'school'], 18:
['zones', 'hand', 'games', 'plays', 'accurate', 'open', 'xbox', 'ncaa',
'look', 'effective', 'lack', 'bilas', 'invites', 'season', 'aspect'], 19:
['zones', 'year', 'hand', 'offense', 'new', '10', 'games', 'quickplay',
'depending', 'plays', 'accurate', 'shot', 'open', 'clock', 'example']}

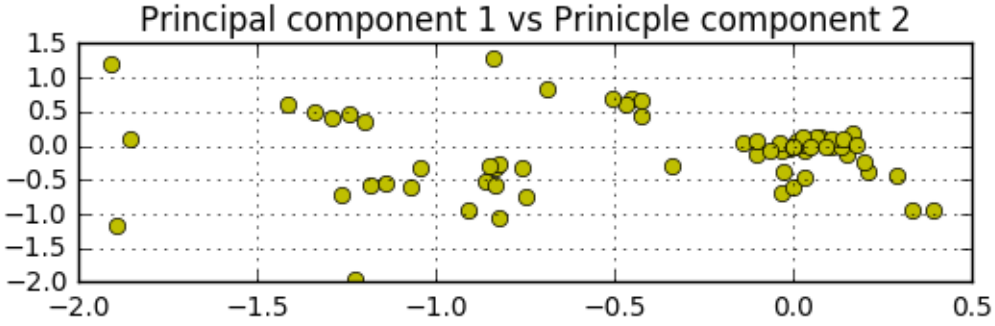


PS2 – PCA Results:

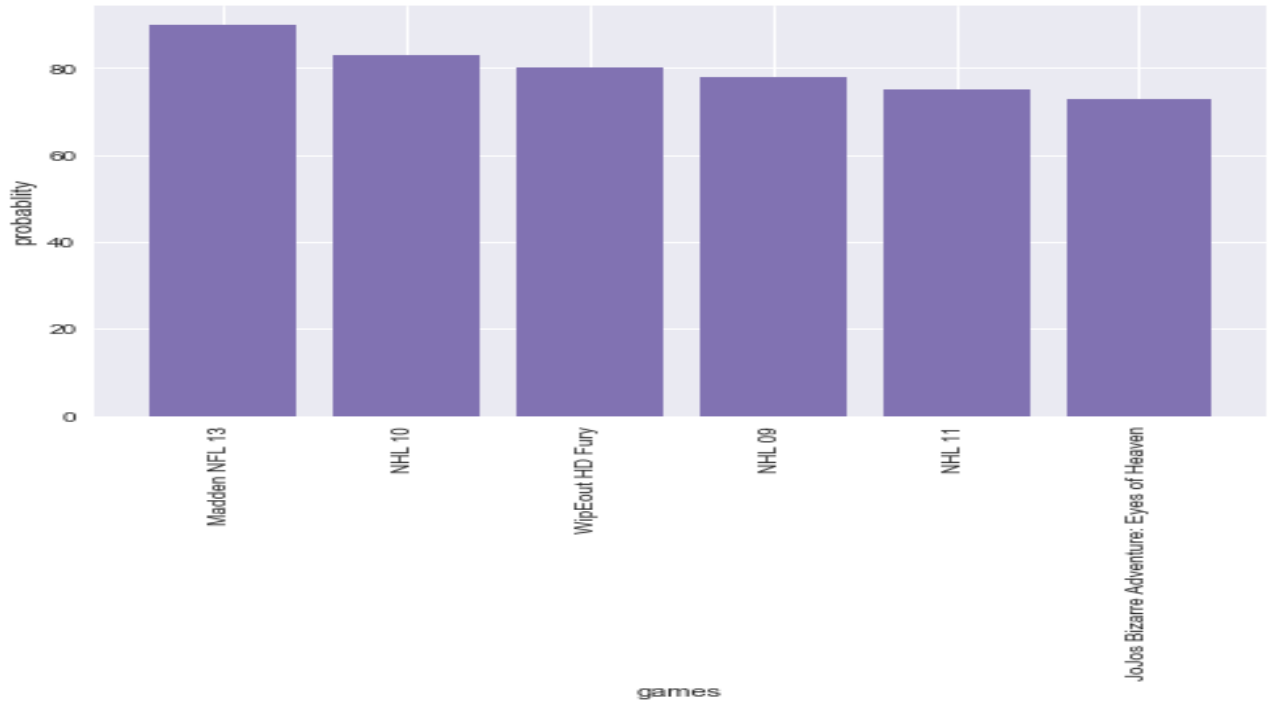
{0: ['game', 'just', 'weapons', 'basic', 'better', 's', 'whistler',
'fighting', 'games', 'basically', 'release', 'ii', 'blade', 'catchphrases',
'way'], 1: ['jump', 't', 'film', 's', 'ii', 'moving', 'blade', 'basically',
'used', 'whistler', 'vampires', 'based', 'better'], 2: ['pistol', 'techno',
'game', 'just', 'catchphrases', 'publishers', 'poor', 'weapons', 'enemies',
'movies', 'based', 'basic', 'really', 'games', 'movie'], 3: ['jump', 'skip',
'film', 's', 'ii', 'moving', 'game', 'catchphrases', 'used', 'weapons',
'movies', 'based', 'games', 'movie', 't'], 4: ['jump', 'little', 'used',

'movies', 'games', 'movie', 't'], 5: ['pistol', 'animations', 'run', 'enemy', 'using', 'enemies', 'movies', 'basic'], 6: ['like', 'lets', 'rage', 'fairly', 'foes', 'fighting', 'basic', 'unfortunately', 'slowly'], 7: ['locked', 'little', 'similar', 'run', 'vampires', 'movies', 'way', 'basic', 'doors', 'door'], 8: ['movies', 'usually', 'little', 'based', 'basic'], 9: ['pistol', 'kill', 'like', 'lets', 'rage', 'fairly', 'foes', 'usually', 'fighting', 'enemy', 'unfortunately', 'slowly'], 10: ['games'], 11: ['pistol', 'kill', 'like', 'little', 'foes', 'usually', 'enemy', 'movies', 'basic', 'unfortunately'], 12: ['like', 'stand', 'little', 'moving', 've', 'foes', 'movies', 'basic', 'start', 'unfortunately'], 13: ['pistol', 'kill', 'making', 'long', 'easy', 'usually', 'enemy', 'enemies', 'based', 'time'], 14: ['games', 'usually', 'based', 'basic'], 15: ['little', 'usually', 'movies', 'based', 'games', 'basic'], 16: ['kill', 'little', 'usually', 'enemy', 'movies', 'basic', 'games'], 17: ['like', 'foes', 'usually', 'movies', 'basic', 'based', 'unfortunately'], 18: ['little', 'usually', 'movies', 'based', 'games', 'basic'], 19: ['little', 'usually', 'movies', 'based', 'games', 'basic']}





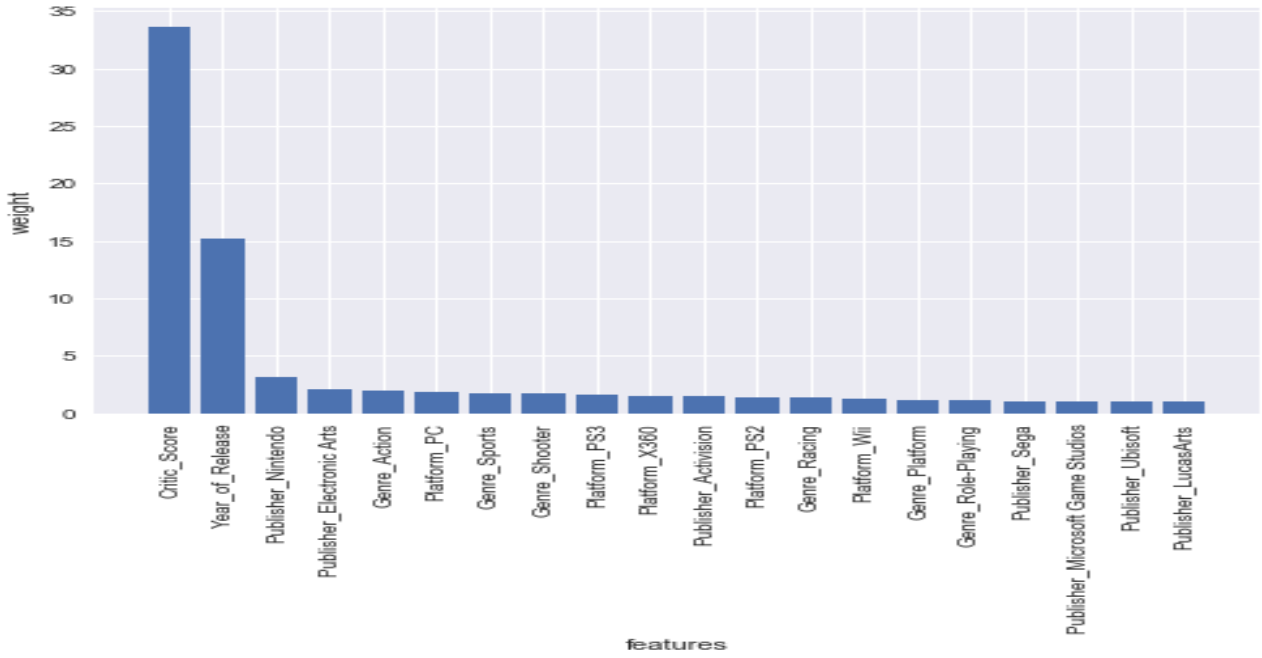
Top 6 games which are more likely to be popular



Probability score for each game:

0.32000000000000001: u'Swing Away Golf',
0.34999999999999998: u'Madden NFL 2003',
0.40000000000000002: u'Art Academy: Home Studio',
0.43666666666666665: u'Tiger Woods PGA Tour 2001',
0.45000000000000001: u'SSX 3',
0.48333333333333328: u'NHL 2K3',
0.5: u'Grand Theft Auto III',
0.53333333333333333: u'The Sims 4: City Living',
0.54000000000000004: u'The Sims 3: Ambitions',
0.55833333333333335: u'SSX On Tour',
0.59999999999999998: u'Rugby 08',
0.63333333333333341: u'The Sims: Medieval',
0.69999999999999996: u"JoJo's Bizarre Adventure: Eyes of Heaven",
0.73333333333333328: u'NHL 11',
0.75: u'NHL 2001',
0.78333333333333333: u'NHL 09',
0.80000000000000004: u'WipEout HD Fury',
0.83333333333333326: u'NHL 10',
0.90000000000000002: u'Madden NFL 13',

Importance of features in Random Forest model



Feature importance values:

('feature', u'Critic_Score', 'percentage', 33.601481571870941)
('feature', u'Year_of_Release', 'percentage', 15.169452501255209)
('feature', u'Publisher_Nintendo', 'percentage', 3.1916749623361182)
('feature', u'Publisher_Electronic Arts', 'percentage', 2.155176638336938)
('feature', u'Genre_Action', 'percentage', 1.9728119873928676)
('feature', u'Platform_PC', 'percentage', 1.8728756294784314)
('feature', u'Genre_Sports', 'percentage', 1.7751395422654046)
('feature', u'Genre_Shooter', 'percentage', 1.7484232651704283)
('feature', u'Platform_PS3', 'percentage', 1.6865055553680115)
('feature', u'Platform_X360', 'percentage', 1.5592908605859601)
('feature', u'Publisher_Activision', 'percentage', 1.5439420865532016)
('feature', u'Platform_PS2', 'percentage', 1.4415303397524417)
('feature', u'Genre_Racing', 'percentage', 1.3967067931376098)
('feature', u'Platform_Wii', 'percentage', 1.2238007734277945)
('feature', u'Genre_Platform', 'percentage', 1.2142413156033438)
('feature', u'Genre_Role-Playing', 'percentage', 1.1244790303369563)
('feature', u'Publisher_Sega', 'percentage', 1.0788184700808472)
('feature', u'Publisher_Microsoft Game Studios', 'percentage', 1.0774123442255348)
('feature', u'Publisher_Ubisoft', 'percentage', 1.0600226246794271)
('feature', u'Publisher_LucasArts', 'percentage', 1.0503466097838807)

Milestone Analysis graph for most selling platform