# Report Milestone 2

## 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 aslo words like (playstation, game) etc but there is no clear pattern that can be seen here.

May be 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 kind of 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 there is not any 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. 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 an useful task and we might need to use some other approach to see any kind of trends in the data.

# 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)

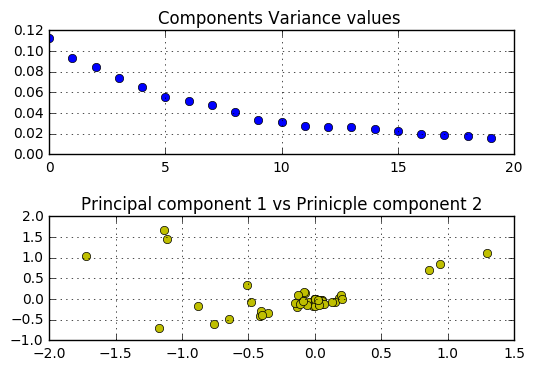
(['time', 'like', 't', 'game', 's'], 37)

(['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)

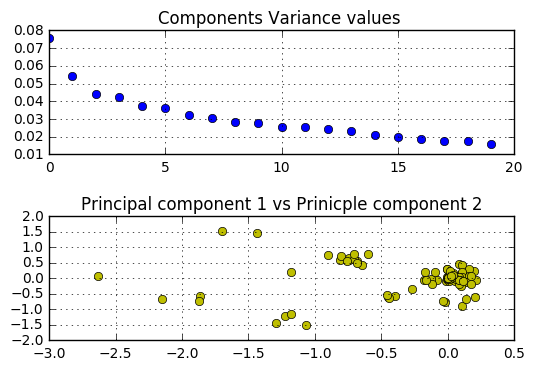
(['quite', 'good', 'game'], 113)

(['bit', 'good', 'game'], 118)

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

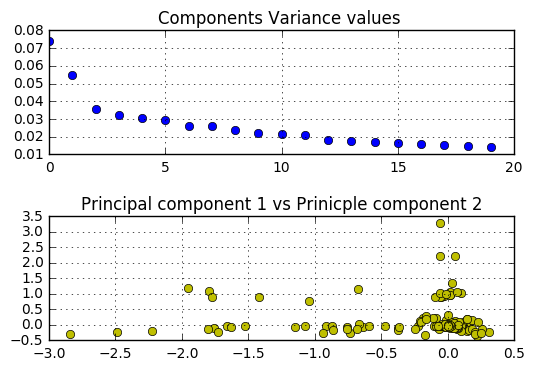
PS3 - Top 20 components:

{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']}



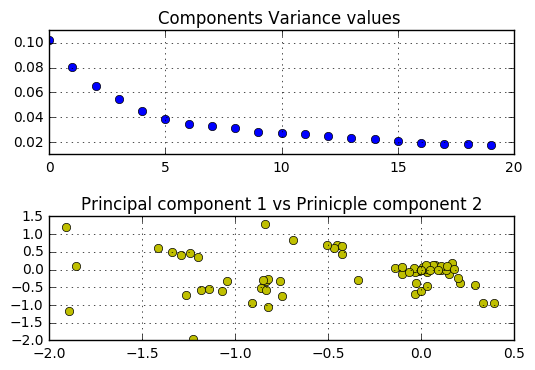
Xbox- PCA Results:

{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']}

****