IS – 684 Web Mining Project

The Champions

Sentiment Analysis on Parents and Kids Game Reviews

By-
Jatin Bhowad
Meenal Sawant
Videet Malpe
Jagruti Divse

Table of contents

PROJECT OVERVIEW

- Common Sense Media rates movies, TV shows, books, and more so parents can feel good about the entertainment choices they make for their kids. They offer the largest, most trusted library of independent age-based ratings and reviews. Their timely parenting advice supports families as they navigate the challenges and possibilities of raising kids in the digital age.

- We chose to narrow down our web scraping and sentiment analysis to game reviews as we found them intriguing and they would provide us an interesting perspective of how games are perceived by parent and their kids.

PROJECT OBJECTIVE

- Scraping and processing relevant data from the exhaustive information available on the website and preparing it for opinion-based sentiment analysis of reviews.

- Building an Aspect based sentiment model to gain insights of parents and kids reviews by using widely known NLP libraries.

Data Collection

- For data collection we used Selenium and Beautifulsoup python library. **Selenium** was originally developed by Jason Huggins in 2004 as an internal tool at ThoughtWorks. Selenium is a portable framework for testing web applications. To interact with the website, we used google chrome based Selenium WebDriver. Selenium web browser accepts and sends commands to the browser. This is achieved through a browser-specific browser driver that sends commands and retrieves results to a browser.

- **Beautiful Soup** is a python library for pulling data out of HTML and XML files. It works to provide idiomatic ways to browse, check, and change the parser tree with your favorite parser. From isolating titles and links, to removing all the text from the html tags, to changing the HTML inside the document you are working with the Beautiful Soup documentation will give you a sense of the variety of items that the Beautiful Soup library will help with.

- As a part of the project our initial aim was to gather as much as data possible from the website. So, we collected data such as Game title, description, age suggestion, parents' ratings, kids rating, parents' reviews, kids review, game genre etc.

- We wrote multiple functions that will help us in the data collection process and also will give advantage of object-oriented programming such as code reuse. Below is the one of the functions written which will return a data frame of all the gathered information.

```python
def scrap_game_page(link):
    game_review_link=browser.find_element_by_partial_link_text(link.text.strip())
    game_review_link.click()
    time.sleep(5)
    html=browser.page_source
    soup=BeautifulSoup(html,'html.parser')
    game_desc_url = browser.current_url
    game_title=soup.find('h1').text.strip()
    age_data=soup.find('div',{'class':'csm-green-age'}).text.split(" ")[1].strip()
    parent_need_to_know=soup.find('div',{'class':'shutter-summary-pane panel-pane pane-entity-field pane-node-field-parents-need-
    game_description=story=soup.find('div',{'class':'field field-name-field-what-is-story field-type-text-long field-label-hidden
    talk_kid_about_this=soup.find('div',{'class':'field field-name-field-family-topics field-type-text-long field-label-hidden'})
    product_details=soup.find_all('ul',{'id':'review-product-details-list'})
    game_theme_info=product_details[0].find('li',{'class': 'themes'})
    if game_theme_info:
        game_theme = game_theme_info.text.split(":")[1].strip()
    else:
        game_theme = None
    game_developer_info=product_details[0].find('li',{'class': 'publishers'})
    if game_developer_info:
        game_developer=game_developer_info.text.split(":")[1].strip()
    else:
        game_developer=None
    parent_reviews = scrap_parent_reviews()
    browser.get(game_desc_url)
    time.sleep(5)
    kids_reviews = scrap_kids_reviews()
    browser.get("https://www.commonsensemedia.org/game-reviews?page={}".format(page_no))
    time.sleep(5)
    return [game_title,age_data,parent_need_to_know,game_description,talk_kid_about_this,game_theme,game_developer, parent_revie
```

Code Snippet for the game review scraper

- Also, we wrote one important function which is working as the main function in the script. It will go over each and every game page and collect the required data. Following is the code for it.

```python
scrapped_data=[]
page_no=0
while True:
    html=browser.page_source
    soup=BeautifulSoup(html,'html.parser')
    links_selected=soup.find_all('strong',{'class':'field-content'})
    for link in links_selected:
        scrapped_data.append(scrap_game_page(link))
    try:
        next_button = browser.find_element_by_xpath('//*[@id="content"]/div/div/div[2]/div/div[4]/div/div/div[2]/ul/li[11]/a')
        next_button.click()
        time.sleep(5)
        page_no += 1
    except:
        break
```

After collecting all the records final dataframe looked like this:

	Title	Age	Game Description	Parents need to know	Kids say about this game	Game theme	Developer	Parents Reviews	Kids Reviews
0	Astro's Playroom	6+	What parents need to know\n\n\nParents need to know that Astro's Playroom is a pack-in title wit...	In ASTRO'S PLAYROOM, players guide Astro through a series of lands, all of which tout the sellin...	Families can talk about progress. How have some products improved over time? How do those change...	Adventures, Robots	Sony Interactive Entertainment	None	None
1	Yakuza: Like a Dragon	18+	What parents need to know\n\n\nParents need to know that Yakuza: Like a Dragon is a role-playing...	YAKUZA: LIKE A DRAGON is both the latest sequel in Sega's long-running Yakuza series and a bit o...	Families can talk about the game's protagonist. Is he a good guy or bad guy? Is it possible to b...	Adventures	Sega of America	None	None
2	Assassin's Creed Valhalla	16+	What parents need to know\n\n\nParents need to know that Assassin's Creed Valhalla is a historic...	The next historic chapter in the epic Assassin's Creed saga begins with ASSASSIN'S CREED VALHALL...	Families can talk about historic representations in media. How can shows, movies, games, etc. se...	Magic and Fantasy, Adventures, History, Misfits and Underdogs, Monsters, Ghosts, and Vampires	UbiSoft	\nGreat game!\nReally good! A little bit of blood but nothing too serious. Fine if you are 12 o...	\nThis title contains: Violence & scarinessLanguageDrinking, Drugs & Smoking\n\nThis review...\n...
3	DIRT 5	10+	What parents need to know\n\n\nParents need to know that DIRT 5 is an off-road racing game for t...	While DIRT 5, like most racing games, doesn't have a story, it doesn't have a regular career mod...	Families can talk about advertising. Like in real racing, the tracks DIRT 5 are covered with ads...	Cars and Trucks	Deep Silver	None	None
4	Bright Memory	15+	What parents need to know\n\n\nParents need to know that Bright Memory is a downloadable first-p...	In BRIGHT MEMORY, a soldier named Shielia infiltrates a secret lab, only to accidentally create ...	Families can talk about violence in video games. Is the impact of the violence in Bright Memory ...	Adventures	Playism	None	None
5	Devil May Cry 5: Special Edition	18+	What parents need to know\n\n\nParents need to know that Devil May Cry 5: Special Edition is an ...	DEVIL MAY CRY 5: SPECIAL EDITION is a remastered version of 2019's action adventure game, Devil ...	Families can talk about violence in video games. Is the impact of the violence in Devil May Cry ...	Magic and Fantasy, Adventures, Monsters, Ghosts, and Vampires	Capcom	\nThe masterpiece returns and so does Virgil\nStill teen friendly. Fantasy violence with blood. ...	None
6	Marvel's Spider-Man: Miles Morales	13+	What parents need to know\n\n\nParents need to know that Marvel's Spider-Man: Miles Morales is a...	MARVEL'S SPIDER-MAN: MILES MORALES is the latest action adventure game to put players into the c...	Families can talk about violence in video games. Is the impact of the violence in Marvel's Spide...	Superheroes, Adventures, Friendship, Great Boy Role Models	Sony Interactive Entertainment	\nGood For Kids\nOk Come On it's Spiderman\nThis title contains: Positive MessagesPositive role ...	None
7	Floor 13: Deep State	15+	What parents need to know\n\n\nParents need to know that Floor 13: Deep State is a dystopian adv...	The folks at FLOOR 13: DEEP STATE would like to congratulate you on your new job as Director Gen...	Families can talk about conspiracy theories. How do conspiracy theories begin and what can peopl...	Activism, Adventures	Humble Games	None	None
8	Pikmin 3 Deluxe	10+	What parents need to know\n\n\nParents need to know that Pikmin 3 Deluxe is an adventure game ex...	PIKMIN 3 DELUXE is an enhanced Nintendo Switch version of the original Pikmin 3 for Wii U. Playe...	Families can talk about screen time. Pikmin 3 Deluxe is broken into individual days of explorati...	Adventures, Bugs, Space and Aliens	Nintendo	None	\nThe best thing since torn with your hands bread\nUNDERRATED it's a masterpiece it REALLY is i ...

DATA CLEANING

- Data cleaning is a process in which you go through all of the data in the data source and either remove or update information that is incomplete, incorrect, improperly formatted, duplicated, or irrelevant.
- Having clean data will ultimately increase overall productivity and allow for the highest quality information in your decision-making.
- For a better understanding of the result, we also performed some data cleaning. We cleared the incomplete records, none values, repetitive words which may skew the analysis.
- After cleaning all the records, we gathered 2101 records about different games which is our final cleaned data for analysis.

Data before cleaning:

```
'\nThis title contains: Violence & scarinessLanguageDrinking, Drugs & Smoking\n\nThis review...\n\nHelped me decide\n\n \nHad u
seful details\n\n \nRead my mind\n\n \n\nReport this review\xa0\n\n\n\nGood game\nIt is amazing what it can teach you and I r
ecommend it to everyone looking for a good game to play that is not to complicated but still enjoyable\nThis title contains: Po
sitive MessagesPositive role modelsEase of Play\n\nThis review...\n\nHelped me decide\n\n \nHad useful details\n\n \nRead my mi
nd\n\n \n\nReport this review\xa0\n\n\n\n'
```

Data after cleaning:

```
'contains violence scarinesslanguagedrinking drug smokingthis good amazing teach recommend everyone looking good compli
ill enjoyablethis contains positive messagespositive role modelsease '
```

EXPLORATORY DATA ANAYSIS

A. Total Count of Reviews per Genre.

Using Matplot library, we created a bar chart showing number of counts according to their respective genre.

```
In [34]: fig, axes = plt.subplots(figsize=(14, 4))
         review_dataframe.Genre.value_counts().plot.bar(title='Total Count of reviews per Genre')

Out[34]: <matplotlib.axes._subplots.AxesSubplot at 0x1e8fb925eb0>
```
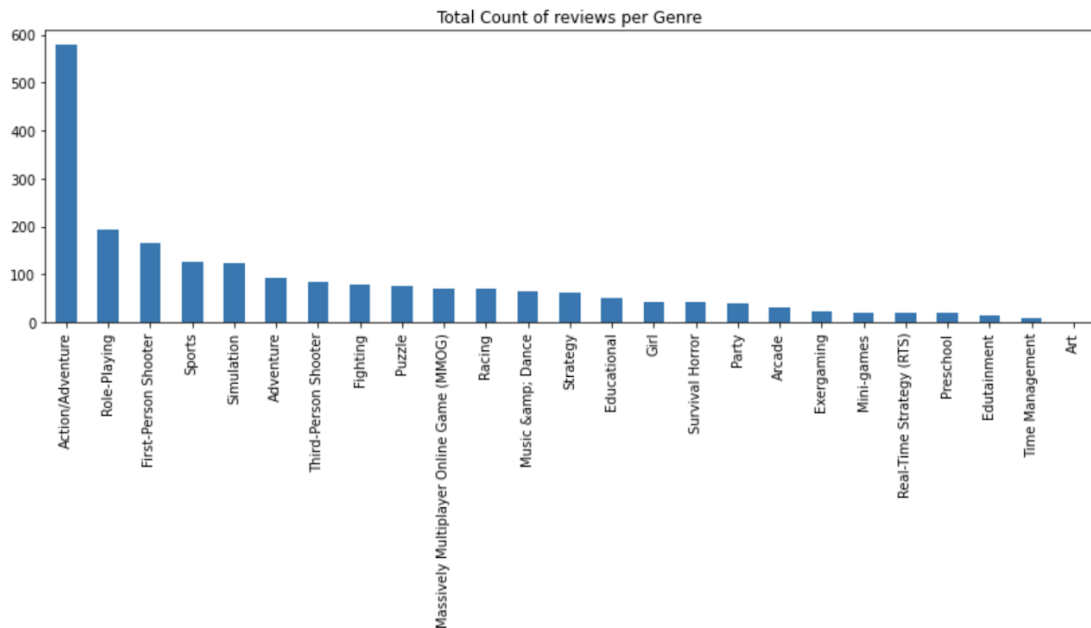

From the above, we can observe that Action/Adventure genre, role playing, and first-person shooter are amongst the most popular genres and hence, the most reviewed.

B. Top 10 Genres By Polarity for Parents and Kids

	Polarity	Subjectivity
Girl	0.275628	0.482803
Edutainment	0.245272	0.343483
Exergaming	0.234252	0.625075
Preschool	0.222617	0.299242
Sports	0.219415	0.534344
Real-Time Strategy (RTS)	0.214977	0.491454
Racing	0.198265	0.446700
Party	0.191938	0.427840
Educational	0.190513	0.350635
Role-Playing	0.185064	0.469998

Parents reviews

	Polarity	Subjectivity
Art	0.433333	0.383333
Arcade	0.355412	0.501491
Educational	0.302180	0.552543
Music & Dance	0.281200	0.504754
Puzzle	0.250920	0.479997
Massively Multiplayer Online Game (MMOG)	0.238193	0.402339
Girl	0.229002	0.401346
Role-Playing	0.206971	0.501101
Simulation	0.189117	0.495342
Real-Time Strategy (RTS)	0.185766	0.486118

Kids's reviews

C. Kid's and parent's Action genre Word Cloud:

Using Wordcloud library, we created a graphic description of most topics according to kids Action genre reviews, we can clearly determine that kids reviews included use of simple language terms like fun, good, love, story.

Kids Action WordCloud

Below is Wordcloud for parents reviews of action genres. We can conclude from these that parents' reviews are surrounding the topic of blood and violence that are mostly the theme of action genre. They have reviewed those games keeping these topics in mind and this shows in the Wordcloud image below.

Parent's Action WordCloud

SENTIMENT ANALYSIS

The process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.

For our data which is concerned with parent's and kid's review, **Aspect Based Sentiment Analysis(ABSA)** is the most suitable method for sentiment analysis and we have used the same for the entirety of our project.

Why **Aspect-based sentiment analysis:-**

- It is impossible to scan large amount of reviews manually, but **Aspect-based sentiment analysis model** is effective to analyze large amount of data to understand parents and kids review.
- This model used for information retrieval in order to perform sentiment analysis by identifying, extracting parents and kids reviews.
- Sentiment Analysis is the process of determining the tone of the text. In this case, we calculated polarity and subjectivity for sentiment analysis.
- We used TextBlob library to process textual data to calculate polarity and subjectivity of the text.
- Polarity lies in the range of (-1,1) where 1 means positive statement , -1 means a negative statement and 0 means neutral sentence.
- Subjectivity lies in the range of (0,1) to determine whether a sentence(reviews) expresses an opinion or not and to check whether the opinion is positive or negative.
- We used spacy model which is widely recognized as one of the powerful and advanced libraries in Natural Language Processing.

Details analysis of 3 genre :

1. After data cleaning, we processed our data to get tokens and used part of speech tagging that reads text in written in language.
2. After applying spaCy model, we were able to extract noun, adjective pairs and assigned noun as a topics and adjectives are their values.
3. Nouns are topics in text and adjectives are expression of their noun.
4. We extracted parents and kids reviews as per genre to get insights and to understand parents and kids perspective towards the game genre to analyze which game genre is most popular in parents and kids by calculating **sentiment** level.
5. As a visualization part, we are representing the data as per max count of topic and representing each topic's polarity and subjectivity.

Genre: Action/Adventure

Kids review:

Polarity Subjectivity and Topic Dictionary

	Polarity	Subjectivity	Appereances
violence	0.049874	0.611755	87.0
fun	0.179228	0.650835	55.0
role	0.359091	0.616818	44.0
ed	0.167898	0.583853	42.0
time	0.213549	0.596300	31.0

violence	fun	role
good	gorgeous	messagespositive
intense	adventurous	sexy
frequent	difficult	great
mild	outdated	good
awful	violent	positive

Parents review:

Polarity Subjectivity and Topic Dictionary

	Polarity	Subjectivity	Appereances
violence	0.163659	0.660536	55.0
year	0.141073	0.270393	38.0
fun	0.158272	0.532342	37.0
role	0.270248	0.550413	36.0
ed	0.036323	0.611835	32.0

violence	fun	role
mad	great	messagespositive
bloody	positive	valuepositive
terrible	intense	messagespositive
interesting	amazing	good
bad	best	positive

By looking at the above analysis of the Actions/Adventure genre, we can see that the most appeared word is 'violence'. We can infer that when it comes to Actions/Adventure parents and kids most likely talk about 'violence' part of the games. It is associated with adjectives like good, intense, awful. And both have neutral (slightly inclined towards positive) views about it. Other than that, they like to talk about other topics such as role, time.

Genre: Role-Playing

Kids review:

Polarity Subjectivity and Topic Dictionary

	Polarity	Subjectivity	Appereances
violence	0.195455	0.607008	21.0
ed	0.115794	0.499033	16.0
fun	0.186260	0.377686	15.0
e	0.118182	0.740422	15.0
rpg	0.355944	0.549650	14.0

violence	rpg	fun
much	older	much
easy	best	little
sexual	positive	vast
bloody	spectacular	favorite
harsh	boring	younger

Parents review:

Polarity Subjectivity and Topic Dictionary

	Polarity	Subjectivity	Appereances
violence	0.167127	0.713961	21.0
role	0.048182	0.552424	15.0
love	0.292727	0.464545	12.0
character	0.312121	0.400000	11.0
fun	0.113005	0.315657	10.0

violence	love	character
wonderful	golden	younger
old	minimal	male
intense	harmless	good
scary	educational	nice
fierce	entertaining	strong

Same as the Actions/Adventure genre, when it comes to Role-playing the most concerned word is 'violence'. In the role-playing genre, violence word is associated with adjectives such as much, easy, harsh, scary, intense. Even both see it with a neutral (slightly towards positivity) view. In the role playing genre, they also like to talk about role, fun, character, and love aspects of the game.

Genre: Sports

Kids review:

Polarity Subjectivity and Topic Dictionary

	Polarity	Subjectivity	Appereances
fun	0.249361	0.590341	18.0
violence	0.396419	0.648072	12.0
love	0.387500	0.677500	10.0
football	0.106250	0.720833	9.0
role	-0.236364	0.606061	9.0

football	love	fun
outrageous	good	super
educational	cool	cool
favourite	great	easy
great	kind	little
bad	educational	nice

Parents review:

Polarity Subjectivity and Topic Dictionary

	Polarity	Subjectivity	Appereances
football	0.144444	0.688889	13.0
fun	0.181667	0.592222	11.0
love	0.487500	0.682099	11.0
violence	0.188095	0.569048	8.0
madden	0.437500	0.375000	8.0

football	love	fun
stupid	oversaturated	realistic
great	much	bad
outrageous	sexy	beautiful
worst	extreme	complete
offensive	sure	easy

In the sports genre, by looking at analysis of kids' reviews they look at sports genre games as fun. Fun is the most appeared word in kids review while on other hand parents are mostly talking about football. Kids relate super, cool, easy adjectives with 'fun' topics. Parents finds 'football' as great, stupid, worst, offensive. Overall view of both are neutral (slightly towards positivity)

MODEL IMPROVEMENT

- After cleaning data and doing sentiment analysis on various genres we observed that the list of topics was flooded with words which don't add any value to our sentiment analysis and nor create any meaningful pattern to understand information.

  Here is one such example of a wordcloud created from parent's Action/Adventure reviews. We can observe that words such as game, review, title do not facilitate us in identifying meaningful pattern in the parent's reviews.
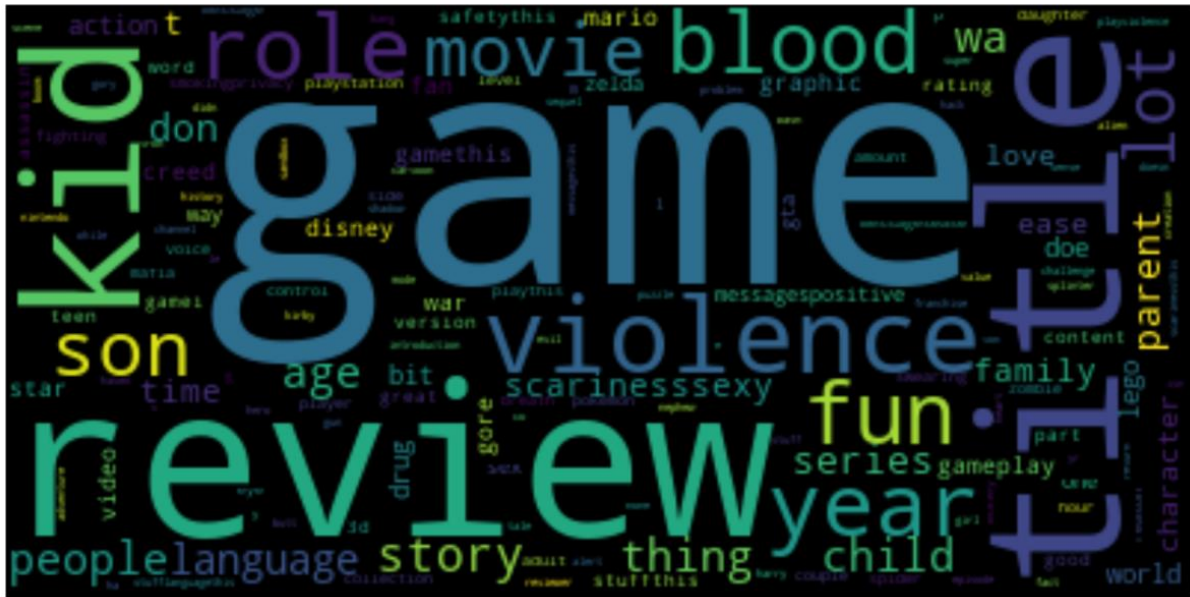

Parent's Word Cloud
Genre : Action/Adventure

Step taken to Improve our Model:

- After getting to know about this flaw, we went back to data cleaning and added these redundant words in our stopwords list.
- After doing this, We ran the sentiment model again and the list of new topics generated gave us a better information and useful context for review's analysis.

Below is the wordcloud after model improvement. We can observe that words like game, review, title were ignored and useful topics such a violence, blood were highlighted which gave us deeper understanding of parent's reviews.

Parent's Word Cloud After Improvement
Genre : Action/Adventure

CONCLUSION

- We can conclude that Action/Adventure genre, role playing, and first-person shooter are amongst the most popular genres and hence, the most reviewed ones.
- Based on polarity, parent's reviews were skewing towards positive for Girl, Edutainment & Exergaming and For Kid's, Art, Arcade & Educational genres reviews were positive.
- Parents and kids dig deep into all the aspects of the game and give their feedback on those aspects. The overall outlook of parents and children towards the games is largely neutral through the study of the reports.
- We observed that in regard to game content , both parents and kids are mostly satisfied with the game as the polarity is on the positive side but on the other hand with regards to quality/graphics of the game their sentiments is leaning on the negative side.
- By adding redundant words which aren't significant to reviews in the stop word list, we have significantly improved our model for better understanding of parent's and kid's reviews.