

Behavior Insights from Text Analysis Using NaughtyDog Subreddit Data

Project By: Arnav Shrivastava (80224)

Introduction and Data Selection

In the realm of gaming, understanding player engagement and preferences is crucial for business insights. This project utilizes a corpus of textual data from the NaughtyDog subreddit, a community centered around discussing games developed by Naughty Dog, a renowned video game developer. The subreddit provides a rich dataset for analyzing gaming culture, fan opinions, and reactions to game narratives and gameplay mechanics. Naughty Dog's titles, like "The Last of Us" and "Uncharted," are known for their immersive storytelling and dynamic gameplay, making discussions around these games a valuable resource for insights into consumer behavior and preferences in the gaming industry.

Data Scraping Methodology

The code for scraping the data was taken from a premade [Github repository](#). Data was collected through a Python script using the requests library. The script targeted the NaughtyDog subreddit, extracting post details through Reddit's API. The headers parameter, set with a User-agent value, ensured compliance with Reddit's scraping policies.

The script performed the following functions:

- **URL Handling:** Constructed a URL to access the subreddit's JSON data, handling pagination through the after parameter.
- **Data Extraction:** Used requests.get to fetch data from the constructed URL and parsed it as JSON.
- **Fieldnames Specification:** Defined a comprehensive list of fieldnames corresponding to various attributes of a subreddit post, such as 'title', 'author', and 'num_comments'.
- **CSV Writing:** The data for each post was written into a CSV file named 'data.csv', using Python's csv library.

The script looped 200 times to aggregate a substantial amount of data, with pauses in between requests to adhere to Reddit's rate limit.

[The data, the code for the analysis along with code for data scraping can be found through this link](#)

Data Preprocessing

The dataset, sourced from the NaughtyDog subreddit and stored in a CSV format, primarily focused on the 'title' column containing post titles. The preprocessing involved several steps:

- **Combining and Cleaning Text:** The individual post titles were merged into a single text string and cleaned by removing punctuation and converting the text to lowercase.
- **Tokenization:** The cleaned text was then tokenized using NLTK's `word_tokenize` method, breaking it down into individual words.
- **Stemming:** Words were stemmed to their root forms using the PorterStemmer, helping in consolidating different forms of the same word.
- **Removing Stop Words:** Common stop words were removed to focus on more significant words, using NLTK's list of English stop words.

Analysis: Frequency Distribution and Word Cloud

The analysis comprised a frequency distribution plot and a word cloud:

- **Frequency Distribution:** This was visualized using matplotlib and seaborn libraries. The frequency of each word was determined, and the top 15 most frequent words were plotted, offering a clear view of dominant words in the discussions.
- **Word Cloud:** A word cloud was generated using the WordCloud library. This visualization represented the frequency of words, where larger and more central words indicated higher frequency. The word cloud provides an immediate visual representation of the most prominent themes in the NaughtyDog subreddit.

Interpretation of Results Based on Word Frequencies

The dataset's most frequent words, including 'last', 'us', 'part', 'naughty', 'uncharted', 'dog', 'ellie', and '2' (from "The Last of Us 2"), provide a wealth of insights into the NaughtyDog subreddit discussions:

- **Game Titles Dominance:** The prominence of titles like "The Last of Us" and "Uncharted" in discussions underscores the strong community engagement with these flagship games. This suggests that future titles in these series or similar genres are likely to be well-received.
- **Character Discussions:** The active discussion around characters such as 'ellie', 'joel', and 'abby' highlights the audience's deep connection to story-driven elements. This underscores the importance of character development and narrative depth in Naughty Dog's games.
- **Gameplay and Game Mechanics:** Interest in 'gameplay', 'remake', and 'playing' indicates a consumer base that values the quality of gameplay and is interested in new iterations or remakes of classic titles.
- **Comparative Discussions:** Discussions comparing different games or game versions ('vs', 'first', 'better') suggest a competitive and discerning player base, emphasizing the need for

continual innovation and improvement in game design.

- **Community Engagement:** The presence of terms like 'think', 'like', 'know', and questions marks an interactive community, indicating the importance of community management and engagement strategies for Naughty Dog.

Business Insights

The text analysis of the NaughtyDog subreddit not only reveals a community deeply engaged with Naughty Dog's games but also offers critical insights for business strategies:

- **Development Focus:** The high engagement with specific game titles and characters points to the potential success of sequels, spin-offs, or DLCs that expand on these elements. Naughty Dog can leverage this by focusing development efforts on these areas.
- **Narrative Importance:** The interest in narrative and characters underscores the need for strong storytelling in future projects. This could guide Naughty Dog's hiring and training practices, emphasizing skills in narrative development.
- **Gameplay Innovation:** The discussions around gameplay mechanics suggest a need for continual innovation. This could influence R&D investments in new game technologies and mechanics.
- **Market Research and Testing:** The comparative discussions highlight the community's discerning nature. Utilizing this subreddit for informal market research or beta testing feedback could provide valuable insights for game development.
- **Community Engagement Strategies:** The active subreddit community represents a powerful resource for direct marketing and feedback. Developing strategies to further engage this community could enhance customer loyalty and provide real-time feedback loops for game development.
- **Expansion into New Genres:** While current discussions focus heavily on existing titles, exploring interest in potential new genres or game types through community engagement could open avenues for diversification.
- **Remakes and Remasters:** Interest in remakes suggests a market for revisiting and updating older titles with modern technology, potentially tapping into nostalgia and reaching new audiences.

In summary, the NaughtyDog subreddit is a goldmine of player insights, directly applicable to business decisions in game development, marketing, and community engagement. By strategically analyzing and responding to these insights, Naughty Dog can enhance its market position and continue to deliver experiences that resonate deeply with their audience.

Conclusion

The text analysis of the NaughtyDog subreddit reveals a community deeply engaged with Naughty Dog's games, discussing various aspects from game narratives to gameplay mechanics. This insight is invaluable for understanding consumer behavior and preferences in the gaming industry, especially regarding narrative-driven games. Such analysis can inform game developers and marketers about

player preferences, guiding future game development and marketing strategies. In conclusion,, the NaughtyDog subreddit is a goldmine of player insights, directly applicable to business decisions in game development, marketing, and community engagement. By strategically analyzing and responding to these insights, Naughty Dog can enhance its market position and continue to deliver experiences that resonate deeply with their audience.