

Generational Marketing Analysis

r/GenZ and r/Millennials

An Nguyen / 04.26.2024

Contents

01 Generational
Marketing Theory

02 Problem Statement

03 Data Collection

04 EDA & Visualization

05 Models Comparison

06 Conclusion

01. Generational marketing theory

emphasizes tailoring marketing strategies to the unique preferences of each generation to optimize marketing campaign effectiveness and customer engagement.



Millennials (1977-1995)

Prefer Facebook and Instagram, enjoy blogs and streaming services.

Invest in stocks and value-based purchases, focusing on sustainability and quality.

Look for authenticity, value personal expression, and prioritize work-life balance.



GenZ (1996-2012)

Drawn to TikTok and Twitch, favor short-form videos and live streams.

Budget-conscious, seek deals, and early adopters of cryptocurrencies.

Prefer interactive and immersive digital experiences influenced by social media trends.



02. Problem Statement

GenInsight Inc. is a marketing agency that wants to create effective digital marketing campaigns for Millennial and GenZ users on the popular social media platform Reddit.

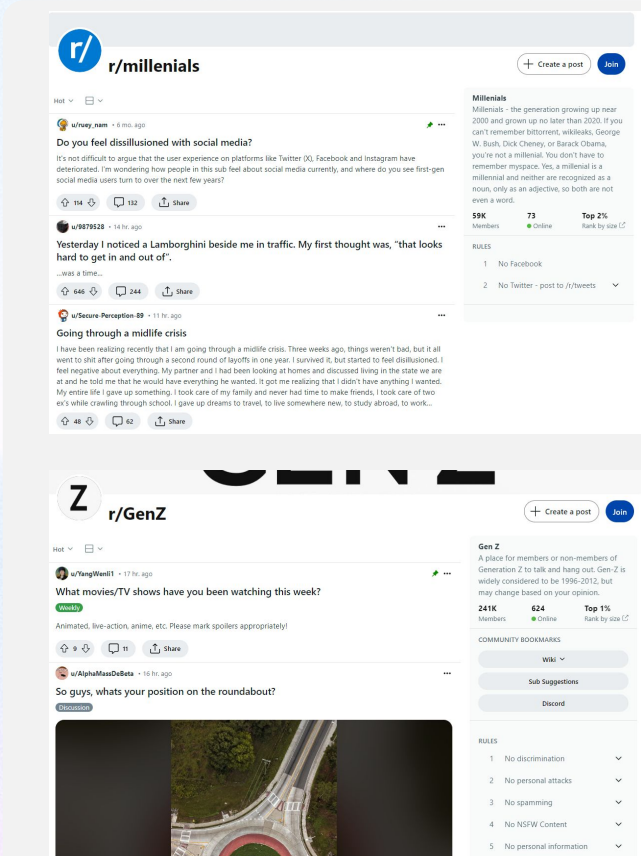
Manually sorting through vast amounts of user-generated content on platforms like Reddit is time-consuming and inefficient.

Objective

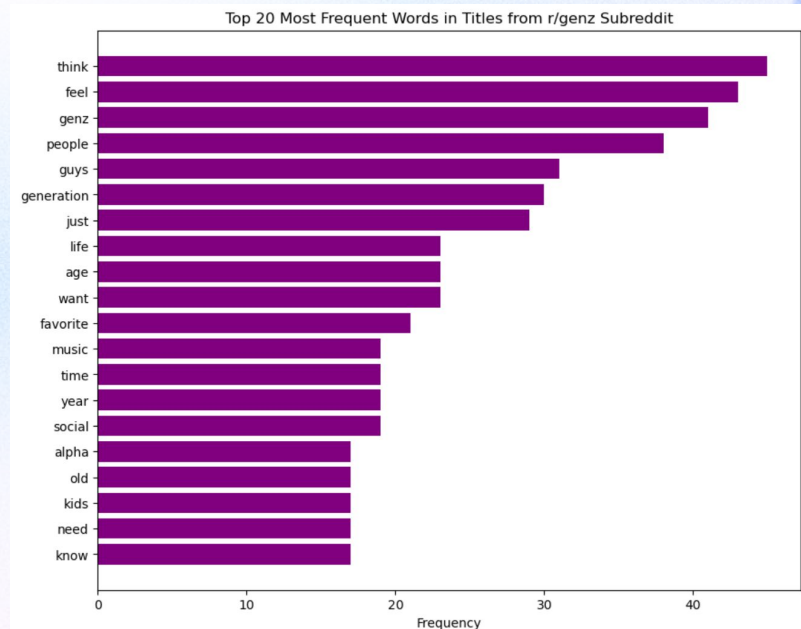
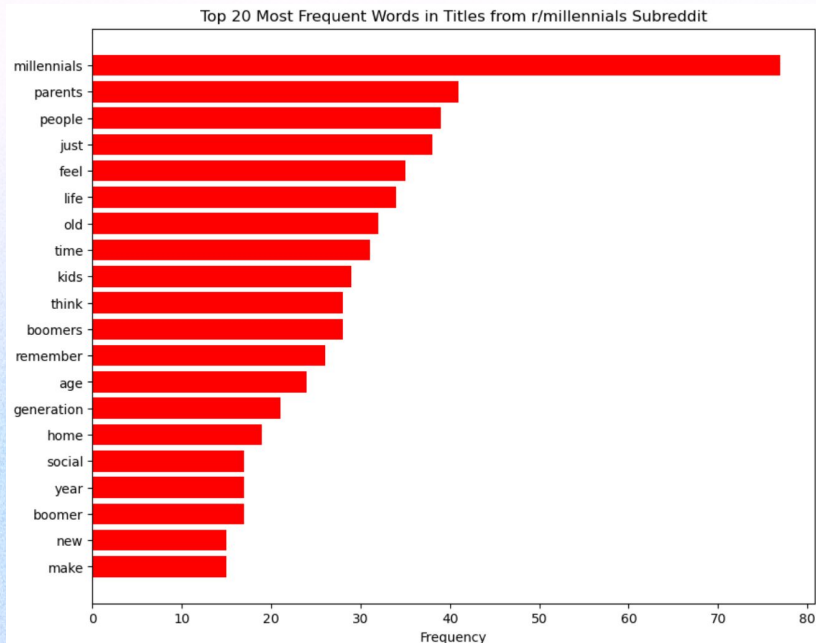
Develop a machine learning model to classify posts as 'millennials' or 'genZ' using post titles from the hottest 1000 posts this year from the r/millennials and r/genZ subreddits.

02. Data Collection

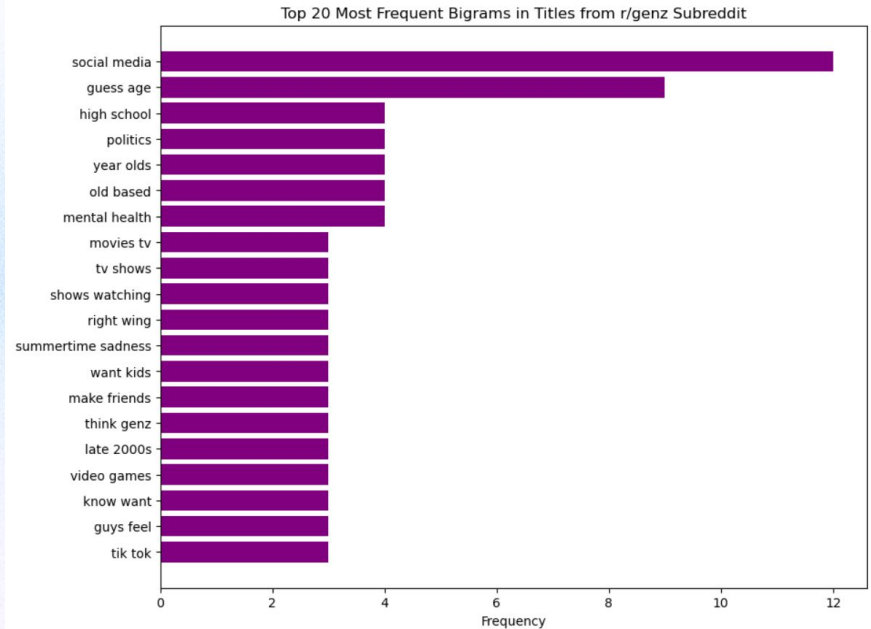
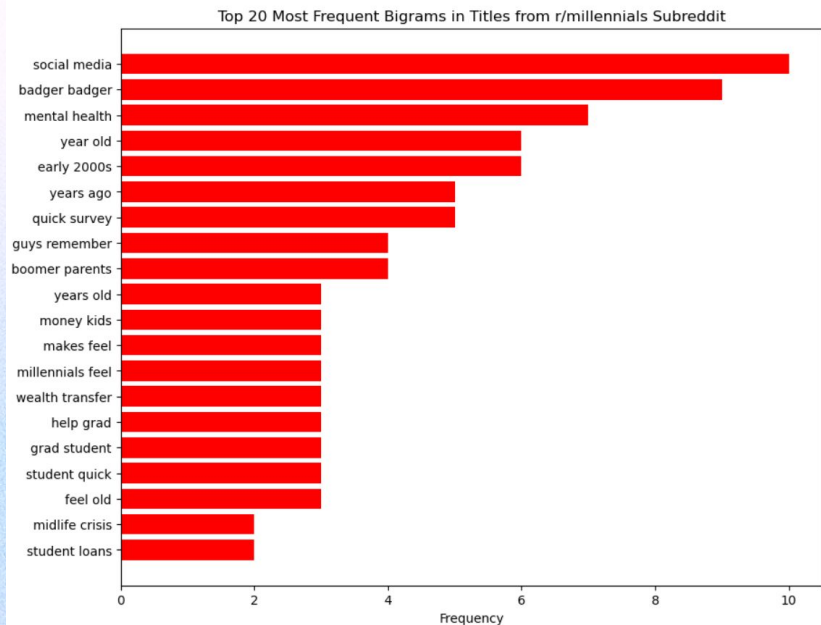
- **Collect** the top 1000 hottest posts this year from the subreddits r/millennials and r/genZ using Python Reddit API Wrapper (PRAW).
 - post ID, title, creation time, self-text (body of the post), number of comments, number of upvotes, and upvote ratio.
- **Clean** the data:
 - Convert datetime in utc format
 - Handle missing values in 'self-text'.
- **Create** new features:
 - word_count of each post title
 - segment_of_day
 - day_of_week
- **Tokenize** texts to transform text data into numerical formats for model training and analysis



03. EDA & Visualization: Top 20 unigrams



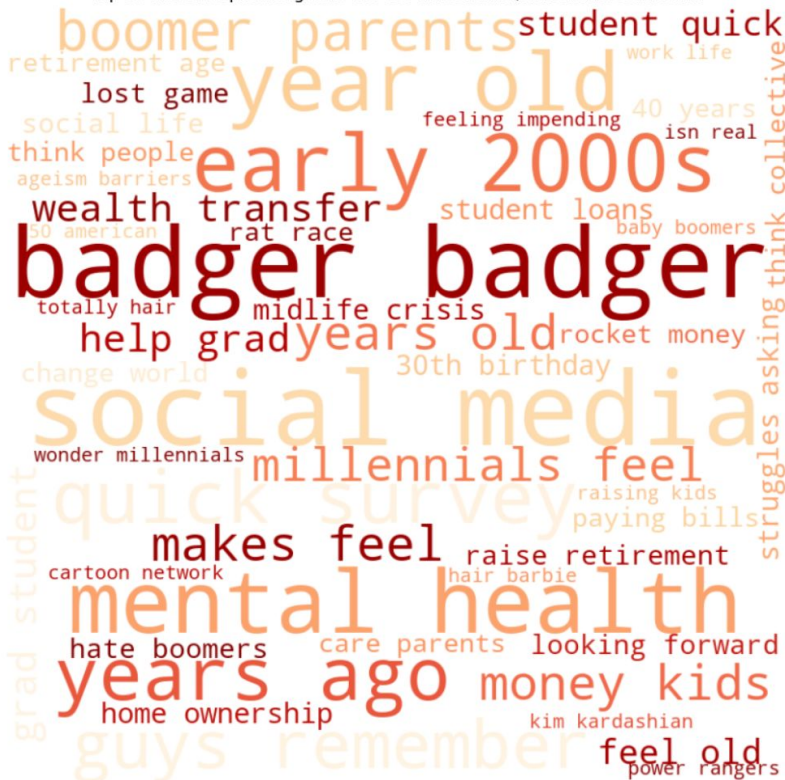
03. EDA & Visualization: Top 20 bigrams



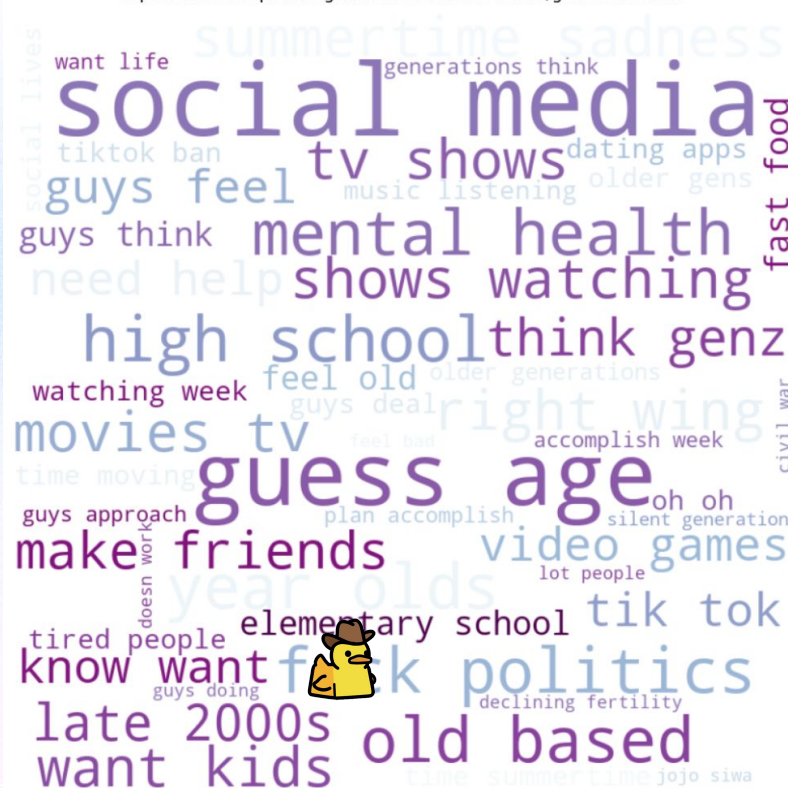


03. EDA & Visualization: Word Clouds from Top 50 Bigrams

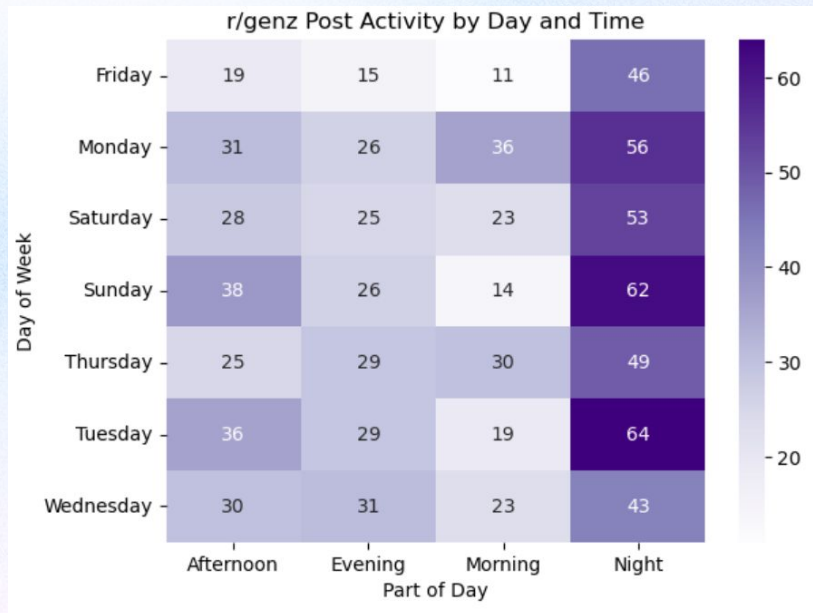
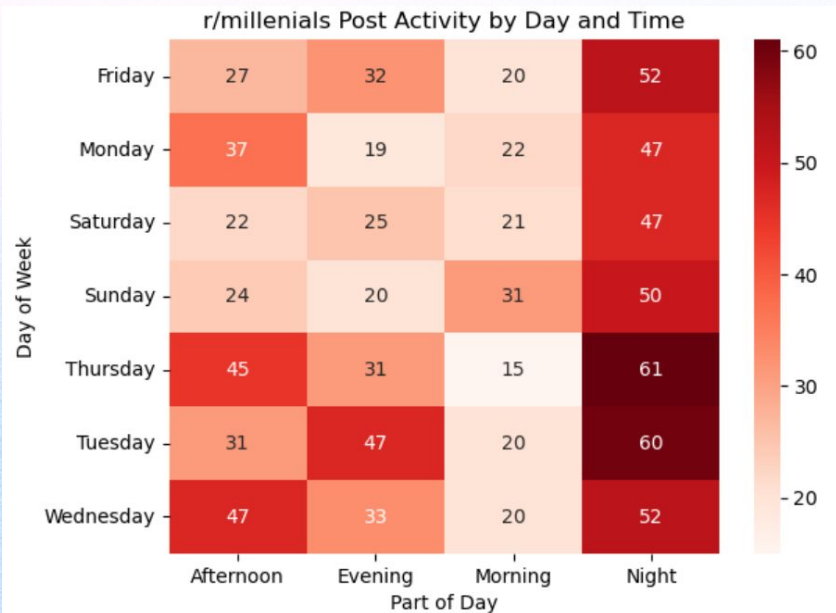
Top 50 Most Frequent Bigrams in Post Titles from r/millennials Subreddit



Top 50 Most Frequent Bigrams in Post Titles from r/genZ Subreddit




03. EDA & Visualization: Activities in r/millennials and r/genZ






05. Models Comparisons




Fine-tuned Binary Classification Models	Test Accuracy	Precision	Recall
Logistic Regression with CountVectorizer	63%	64%	56%
<i>Multinomial Naive Bayes with CountVectorizer</i>	66%	66%	63%
<i>Multinomial Naive Bayes with TfidfVectorizer</i>	66%	69%	55%
K-Nearest Neighbors with CountVectorizer	58%	59%	46%
RandomForest with CountVectorizer	64%	66%	54%
SVM with CountVectorizer	64%	62%	67%







06. Conclusion

- Both r/genZ and r/millennials are most active online at night and during the weekends.
 - Binary classification of posts:
 - Multinomial Naive Bayes model is the best performer with 66% test accuracy.
 - All models performed better than the baseline model,
 - *the test accuracy scores are not significant enough to clearly classify posts.*
- 

Recommendations

- **Collect more posts** from both r/millennials and r/genZ and similar subreddits.
 - **Stemmatize and lemmatize words** using libraries like NLTK and spaCy.
 - **Add additional features** to the models such as upvote ratio, segment of day, self_texts, etc.
 - **Analyze the misclassified posts** to understand where each model fails, and identify potential biases or similarities in languages and topics.
 - **GenInsight Inc. should re-evaluate the effectiveness of generational marketing strategy, especially for r/millennials and r/genZ redditors:**
 - *this tactic oversimplifies consumer behavior based on age, neglecting other significant factors like culture, socio-economic status, and technological adoption.*
 - *it does not account for the rapid changes in consumer preferences which transcends age.*
 - *it does not account for fuzzy generational boundaries.*
- 
- 

Thank you !