HELP DESK

Gen Z came to 'slay.' Their bosses don't know what that means.

(Emma Kumer/ The Washington Post; photos courtesy of Janvi Karla and Molly Foulkes)

'Milking' profits: just knowing how Gen Z and Millennials talk

Are they even speaking the same language?

SG-DSI-41, Project 3, Group 01, Members: Daryl Chia, Germaine Choo, Sharifah Nurulhuda, Tan Wei Chiong

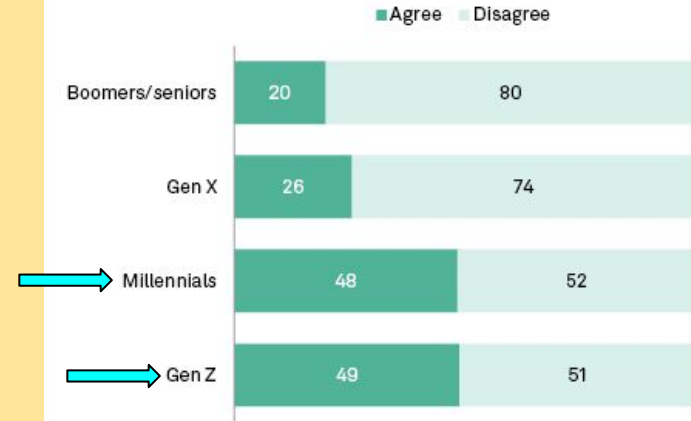# For US Gen Z adults and millennials, social media is a way of life

1 "... the survey data confirmed that 75% of those who spend five or more hours on social media were young Gen Z adults and millennials ..."

2 "Gen Z adults and millennials are much more likely than older generations to rely on social media for their news, sports highlights, short-form videos, TV/video content discovery and product/service information."

3

US: I often purchase products after seeing an ad or post about the product on social media, 2023
Percentage of age generation

■ Agree   ■ Disagree

Generation	Agree	Disagree
Boomers/seniors	20	80
Gen X	26	74
Millennials	48	52
Gen Z	49	51

Question: Please indicate if you agree or disagree with the following statements regarding your use of social media.
Base: Gen Z - 297, millennials - 671, Gen X - 336, baby boomers/seniors - 433.
Source: Kagan US first-quarter 2023 Consumer Insights survey.
Kagan, a media research group within the TMT offering of S&P Global Market Intelligence.
© 2023 S&P Global.

Gen Zs and Millennials consume social media content differently

Gen Zs

- Content that quickly cuts to the chase
- Use videos to learn about products - short-form video formatting & mobile-optimized content
- Focus on trust/connection with brand - sponsor influencers who know audience

Millennials

- Longer attention span - explore more creative longer-form content e.g. longer marketing videos
- Consume multiple different styles of multimedia content, from videos to podcasts
- Focus on products that make lives easier/ how it solves daily problems
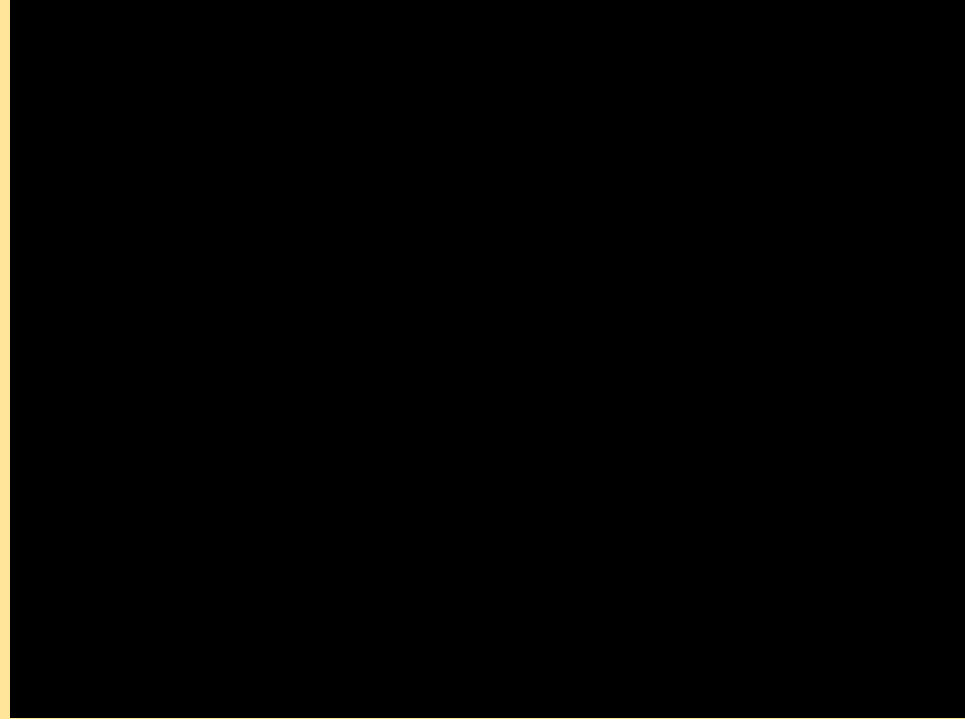
Target Audience

Companies that advertise on Instagram

Problem Statement

How might we make use of language patterns to differentiate between different groups of users (i.e. Millennials and Gen Z) on Instagram in order to maximise profits through targeted campaigning?

How ads currently work on Instagram

"*The way you engage* with other creators and followers tells us whose posts you'd like to see.

These signals can help your content reach the people who care about it most ..."
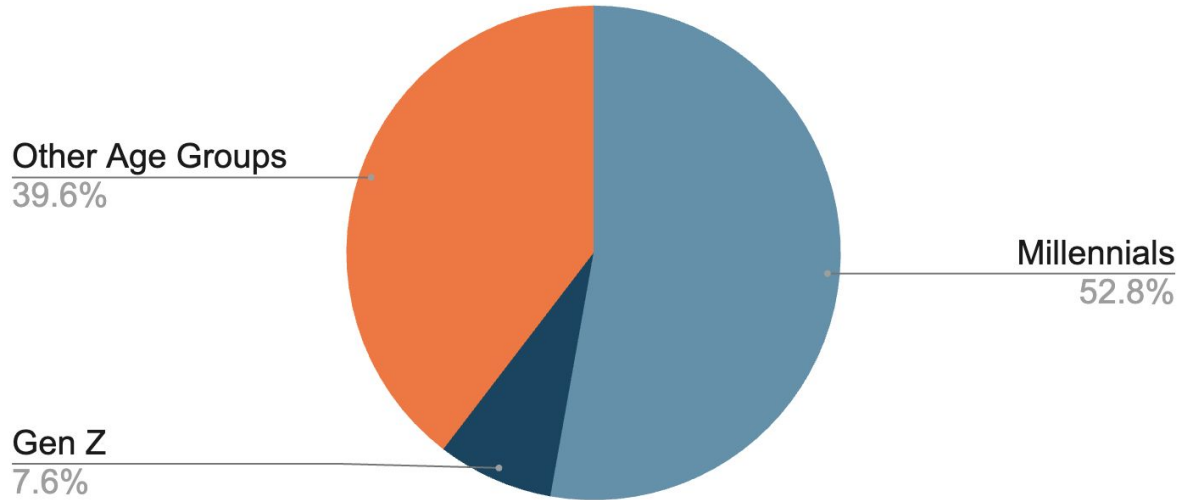
Taken from Instagram Website

You as the advertiser

Your target audience for this product: Gen Z

Purchase data on your website:

Purchase Volume

Other Age Groups
39.6%

Millennials
52.8%

Gen Z
7.6%

EDA and Cleaning

```c
#define ; rn
#define false cap //use instead of 0 or null or EOF when appropriate
#define ! no
#define ? sussin
#define long fr
#define main chief
#define & bouta
#define = finna
#define * mf
#define : bussin
#define == be
#define break yikes
#define -- vibin
#define return deadass
#define ++ ongod
#define } tho

#include <stdio.h>

int chief() {
    fr fr int c rn
    scanf("%d", bouta c) rn
    fr fr int mf p rn
    c>3 sussin p finna bouta c[0] bussin p finna null rn
    while(no cap) {
        if(p be cap)
            yikes rn
        printf("%x, %d\n",p , c) rn
        p vibin rn //carefull don't use mf p or it segfaults fr fr
    tho
    deadass 0 rn
tho
```

**Coding with Gen Z
(Source: r/ProgrammingHorror)**

EDA - Description of Scraped Data Sets

Subreddit	Description
r/Millennials	Before EDA: - 48 145 comments After EDA and undersampling: - 36 941 comments
r/GenZ	Before EDA: - 38 065 comments After EDA: - 36 941 comments

EDA - Description of Scraped Data Sets

Considerations	Rationale
'Top' filter used when scraping each subreddit	Robust to time and individual search history
Only 'comments' were used in final dataset	Contains a large number of interactions between users of the respective subreddit which reveals actual language used between generations
Definitions of Gen Z and Millennials	Millennials: Born 1981 – 1996 Gen Z: Born 1997 – 2010

Structure and general sequence of EDA

Cleaning:

- Initial dropping of rows
- Identification and removal of irrelevant text

EDA:

- Comparison of mean word count
- N-gram analysis

EDA using Regular Expression (RegEx)

Removed Comments

Bot Comments
"I am a bot"

Comments that had already been Deleted, and Null Values
"[deleted], [removed]"

Replaced with ' '

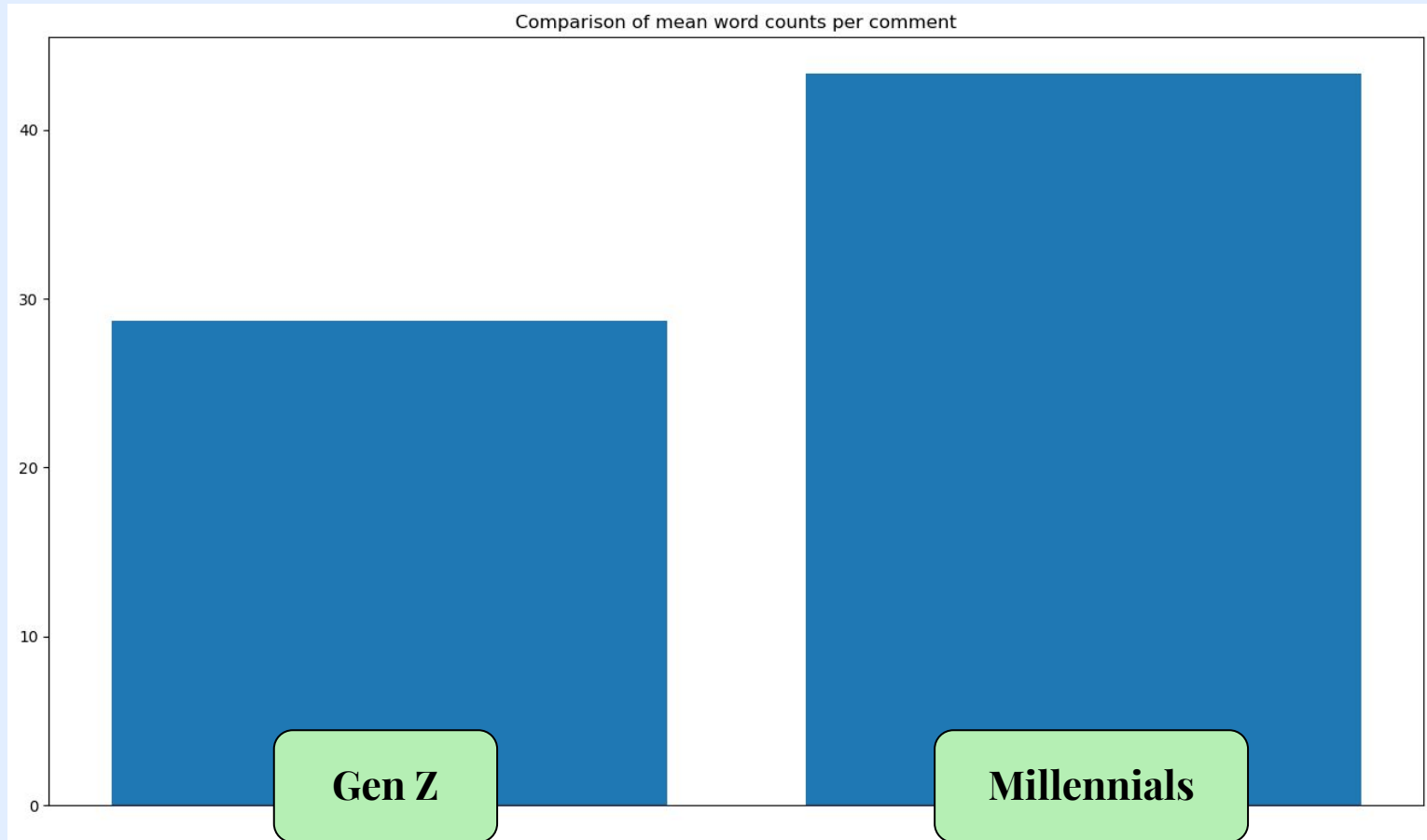**Whitespaces, Escape Characters, Numbers, Punctuation**
"\n, \t, \r"

Subreddits, Username References
"r/…, u/…"

Website URLs, Gif Images, HTML, Emoji Hexadecimal Codes
"https://, \u2ood, "

EDA Insights : Average number of words used per comment

Comparison of mean word counts per comment

Gen Z

Millennials

EDA Insights : General themes

Common themes across both groups

- Talking about other generations
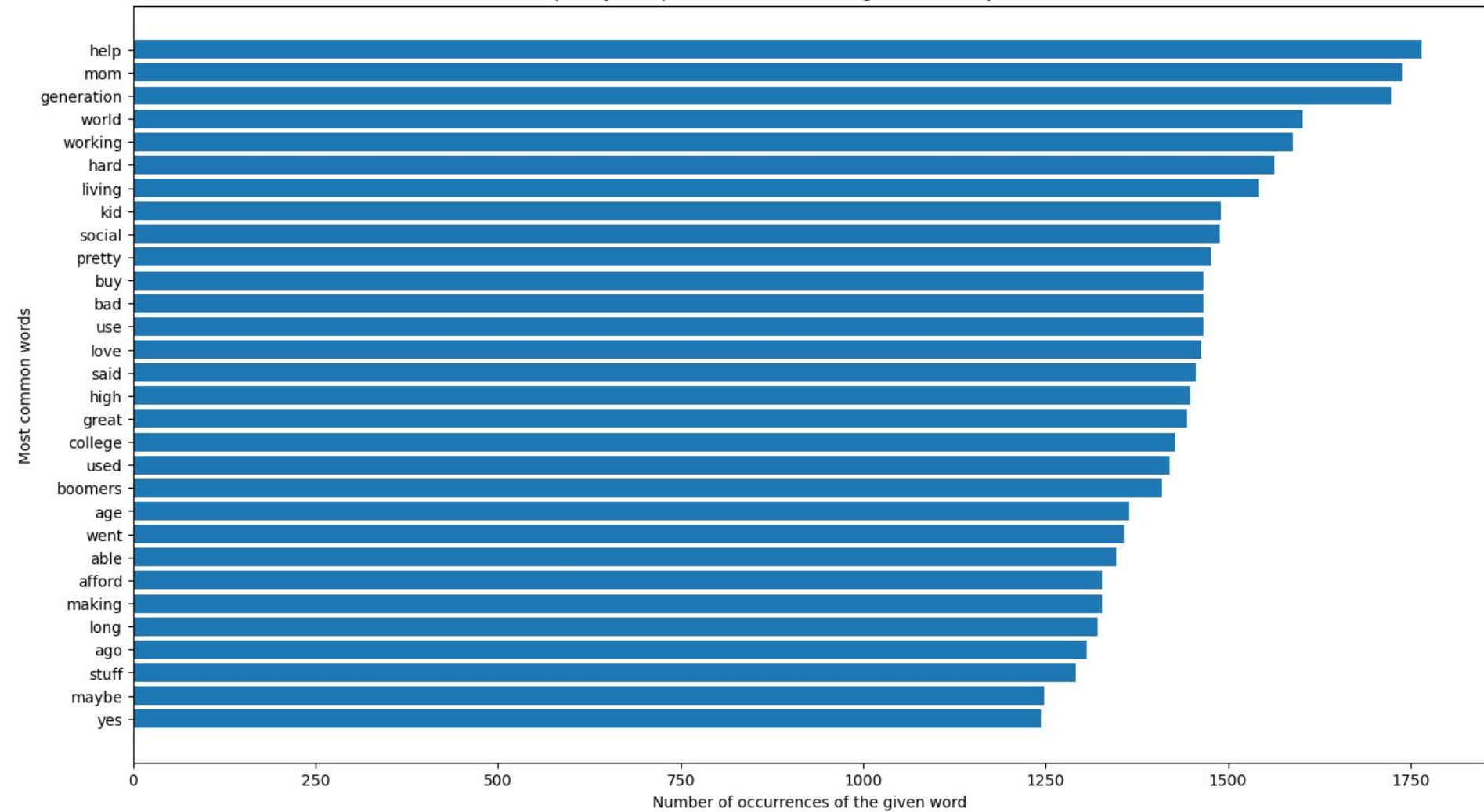- Social Media
- Not wanting to have kids

r/Millennials

- Mental health
- Housing and financial difficulties
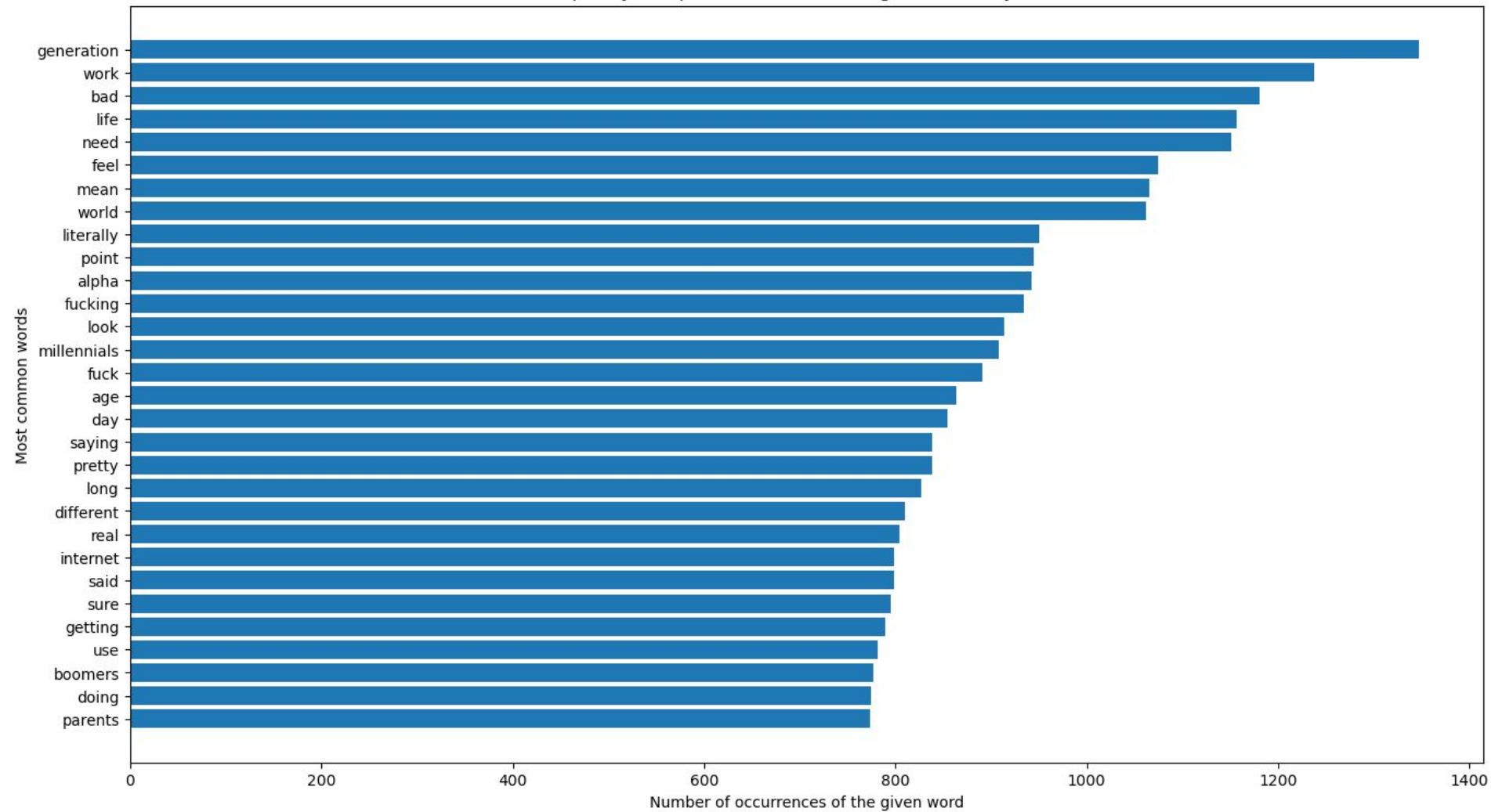- Family life
- Work–life balance

r/GenZ

- Schooling life
- Gen Alpha
- Self-focused
- Gaming

Frequency of top 30 most common 1-grams used by Millennials

Frequency of top 30 most common 1-grams used by Gen Z
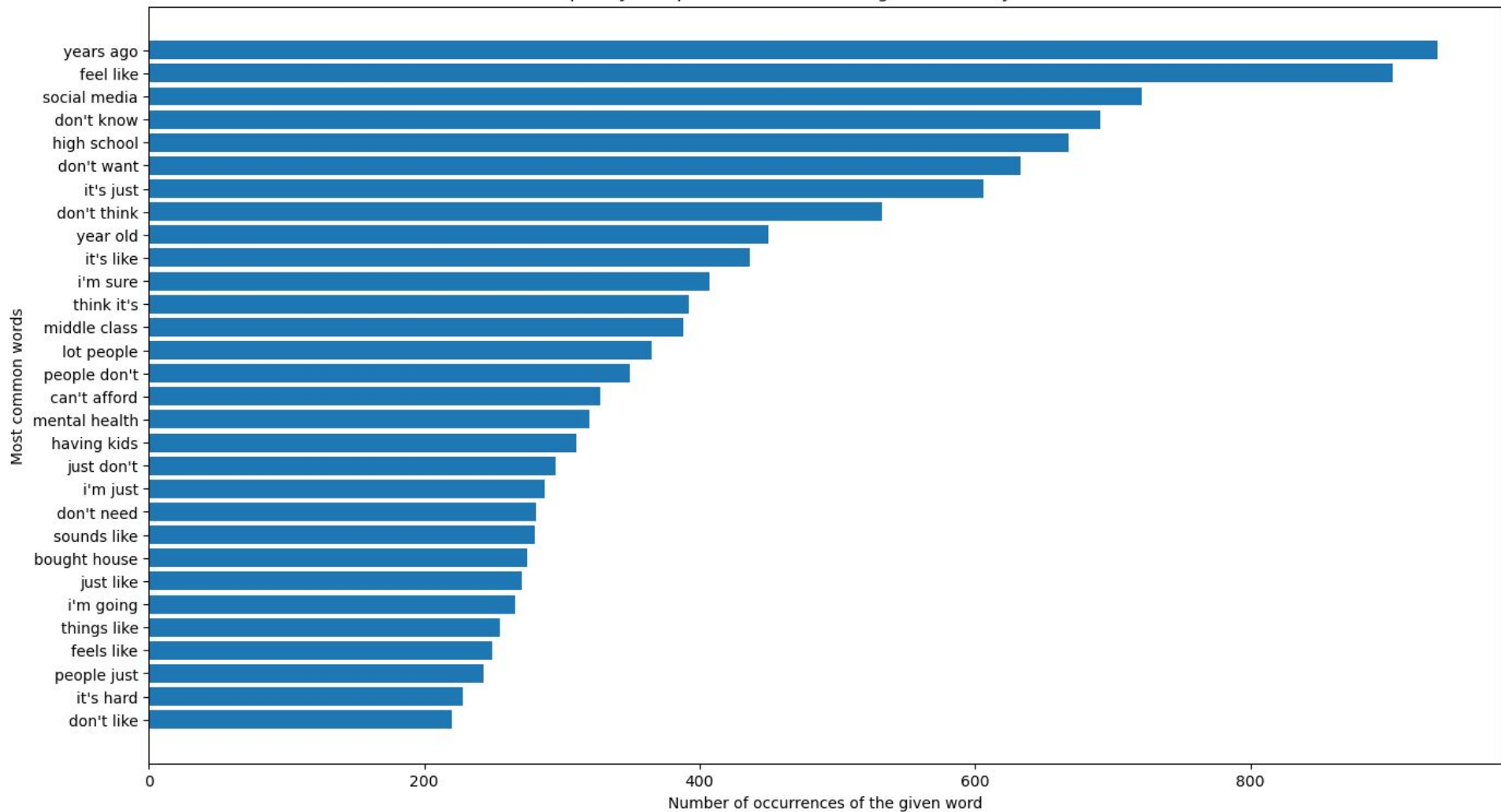
EDA Insights : top 10 phrases from 1-gram analysis

r/Millennials

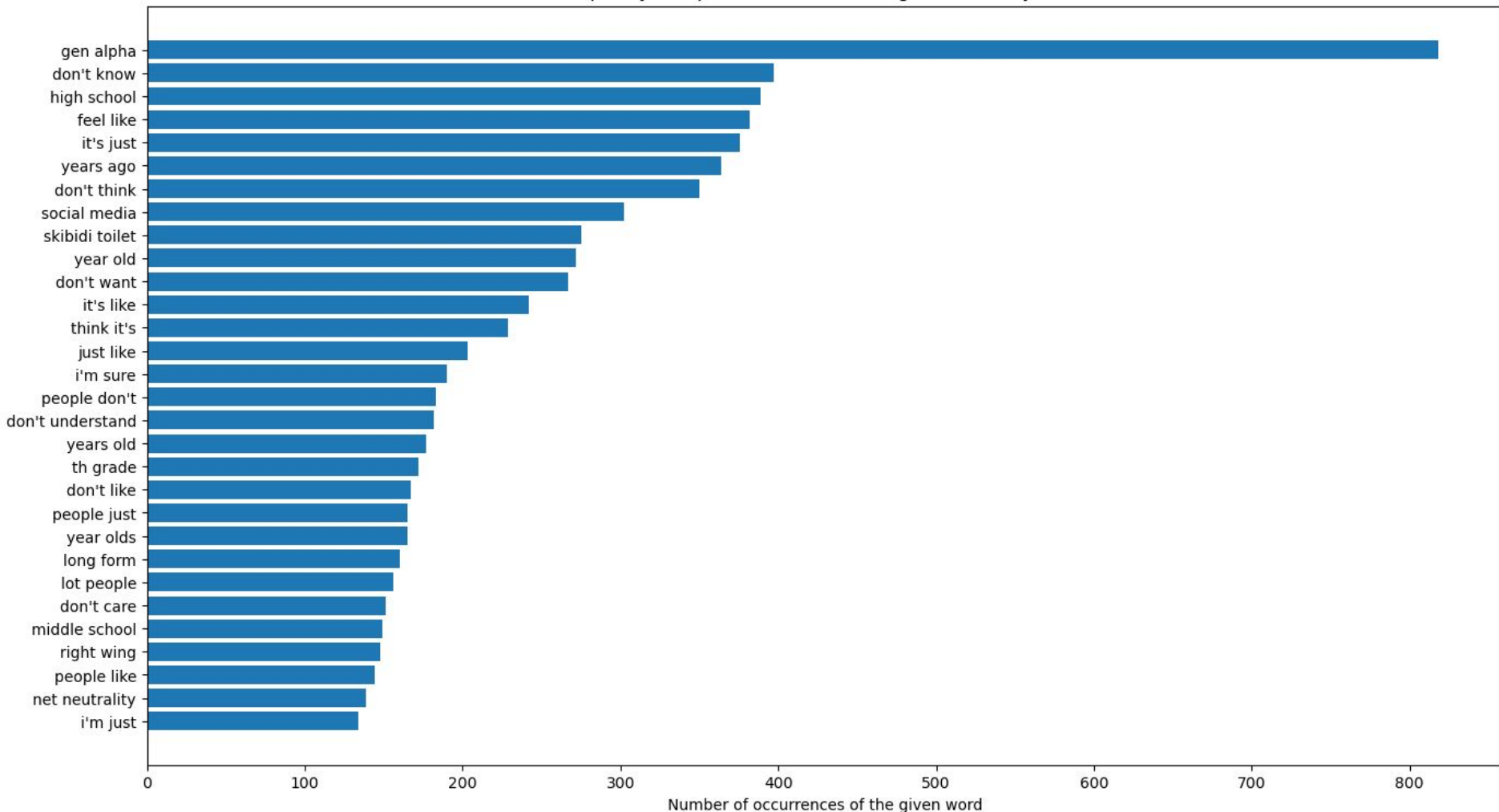|	
1. help	6. hard
2. mom	7. **living**
3. **generation**	8. kid
4. **world**	9. social
5. **working**	10. pretty

r/GenZ

1. **generation**	6. feel
2. **work**	7. mean
3. bad	8. **world**
4. **life**	9. literally
5. need	10. point

Frequency of top 30 most common 2-grams used by Millennials

Most common words	Number of occurrences of the given word
years ago	
feel like	
social media	
don't know	
high school	
don't want	
it's just	
don't think	
year old	
it's like	
i'm sure	
think it's	
middle class	
lot people	
people don't	
can't afford	
mental health	
having kids	
just don't	
i'm just	
don't need	
sounds like	
bought house	
just like	
i'm going	
things like	
feels like	
people just	
it's hard	
don't like	

Frequency of top 30 most common 2-grams used by Gen Z

EDA Insights : top 10 phrases from 2-gram analysis

r/Millennials

1. years ago
2. feel like
3. social media
4. don't know
5. high school
6. don't want
7. it's just
8. don't think
9. year old
10. it's like

r/GenZ

1. gen alpha
2. don't know
3. high school
4. feel like
5. it's just
6. years ago
7. don't think
8. social media
9. skibidi toilet
10. year old

EDA Insights : top 10 phrases from 2-gram analysis

r/Millennials

1. years ago
2. feel like
3. social media
4. ~~don't know~~
5. high school
6. ~~don't want~~
7. ~~it's just~~
8. ~~don't think~~
9. year old
10. ~~it's like~~

r/GenZ

1. gen alpha
2. don't know
3. high school
4. feel like
5. ~~it's just~~
6. years ago
7. ~~don't think~~
8. social media
9. skibidi toilet
10. year old

EDA Insights : top 10 phrases from 2-gram analysis (noise removed)

r/Millennials

1. years ago
2. feel like
3. social media
4. high school
5. year old
6. **middle class**
7. **can't afford**
8. **mental health**
9. **having kids**
10. **bought house**

r/GenZ

1. gen alpha
2. high school
3. years ago
4. social media
5. skibidi toilet
6. year old
7. **middle school**
8. **right wing**
9. **people like**
10. **net neutrality**

Frequency of top 30 most common 3-grams used by Millennials
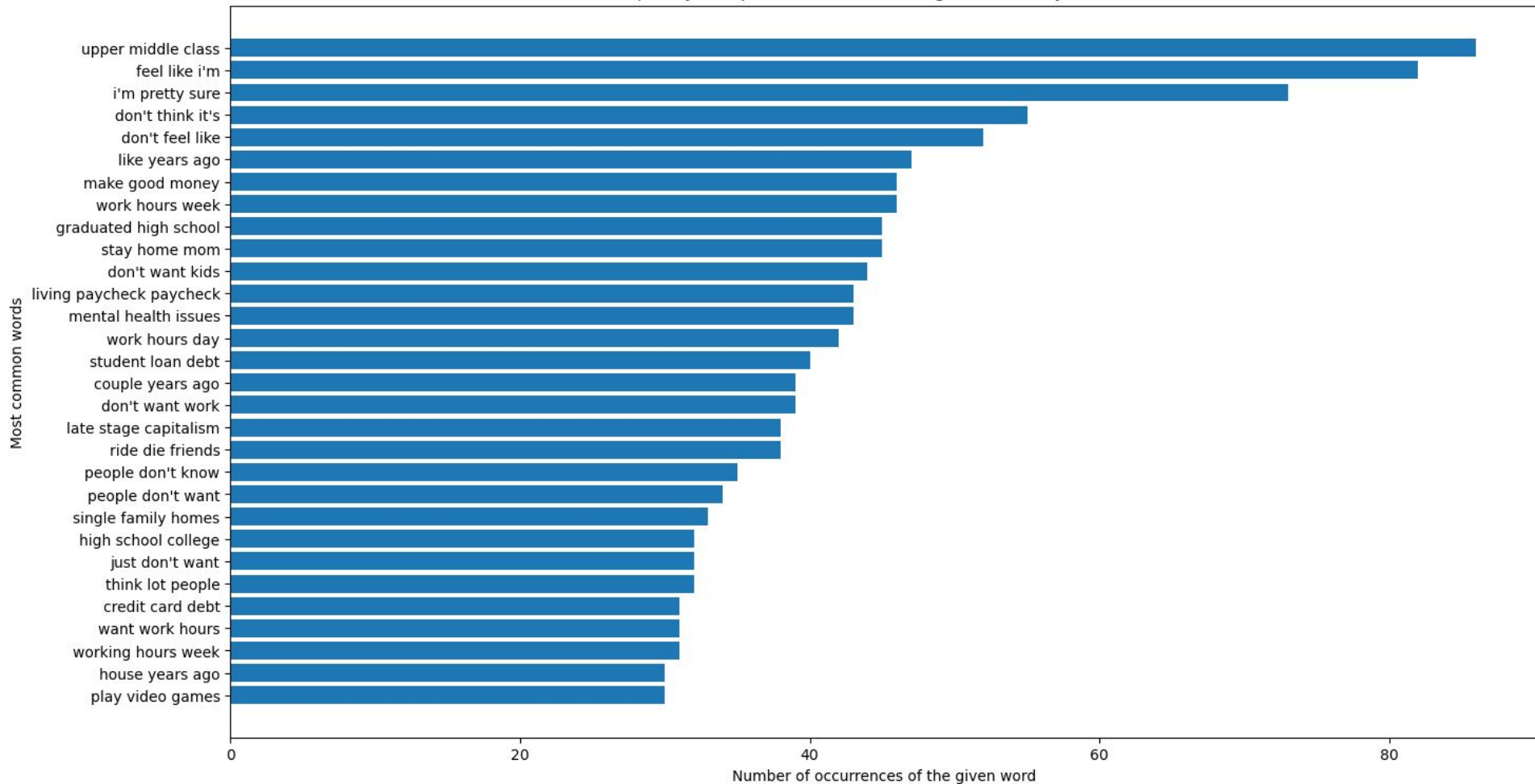
Frequency of top 30 most common 3-grams used by Gen Z
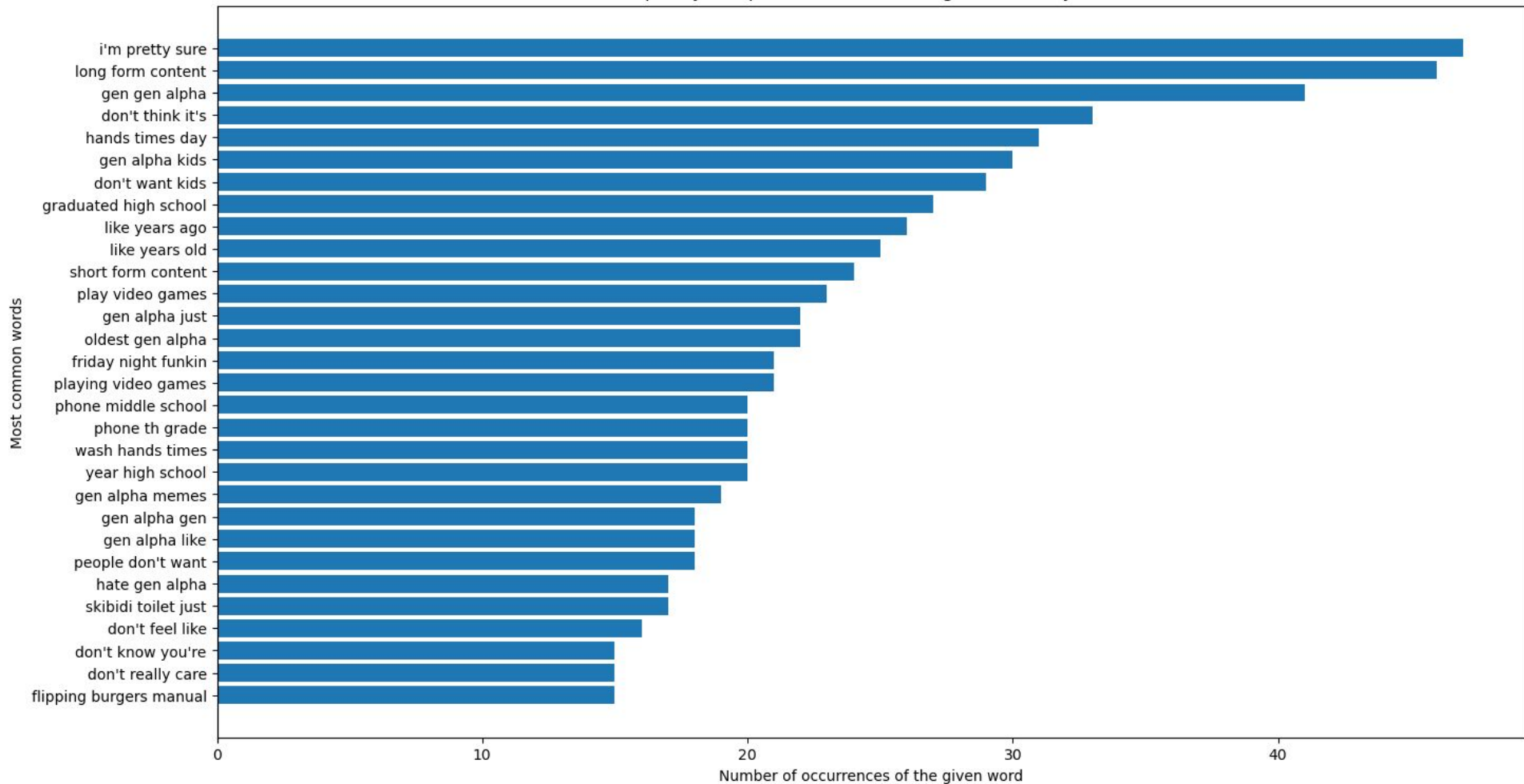
EDA Insights : Insights from 3-gram analysis (noise removed)

r/Millennials

upper middle class

feel like i'm

like years ago

make good money

work hours week

graduated high school

stay home mom

don't want kids

living paycheck (to) paycheck

r/GenZ

long form content

hands times day

gen alpha kids

don't want kids

graduated high school

like years ago

short form content

play video games

friday night funkin

EDA Insights : General themes (recap)

Common themes across both groups

- Talking about other generations
- Social Media
- Not wanting to have kids

r/Millennials

- Mental health
- Housing and financial difficulties
- Family life
- Work-life balance

r/GenZ

- Schooling life
- Gen Alpha
- Self-focused
- Gaming

https://pbs.twimg.com/media/EveFNIUXYAE8yMV.png

Modelling

Our Implemented Models

Model	Description
Logistic Regression	- Using probabilities to classify the predicted binary response with reference to a threshold probability e.g 0.5 mid-point - Relatively faster run time
Bernoulli Naive Bayes	- Classification model relying on Bayes Theorem for binary data
Multinomial Naive Bayes	- Classification model relying on Bayes Theorem for count data

Metric Selected: Accuracy

Accuracy:

$$\frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

Precision:

$$\frac{\text{True Positives}}{\text{True Positives + False Positives}}$$

Recall:

$$\frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

F1-Score:

$$\frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision + Recall}}$$

Rationale on score metric:

- Our goal is to accurately identify if a comment was written by a Gen Z or a Millennial – equal emphasis on accurate prediction for both classes
- Other metrics such as recall/precision are not prioritised
- Selected data set is also relatively balanced between the 2 classes
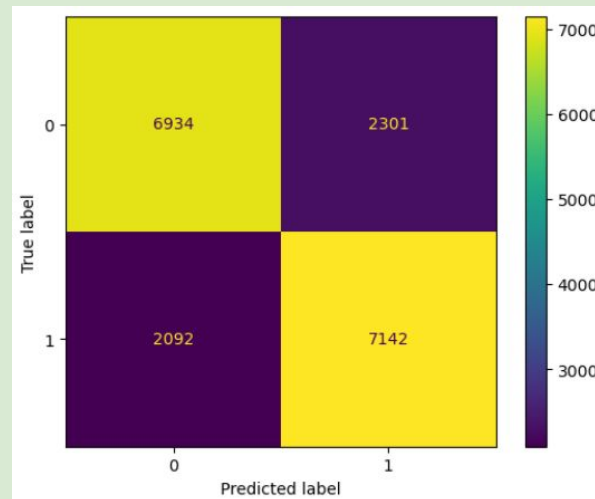
Confusion Matrix for Logistic Regression Model

Model Scores

Model	Training Accuracy	Testing Accuracy	Precision	Recall	F1-score
Logistic Regression	**0.85**	**0.76**	0.76	0.77	0.76
Bernoulli Naive Bayes	0.73	0.69	0.64	0.88	0.74
Multinomial Naive Bayes	0.84	0.76	0.77	0.75	0.76

Selected Model: Logistic Regression

Authentic Data Testing

Let's Test This.

Comments	Model Prediction	Actual
I was skeptical, but took a chance on the Black Friday deal. Haven't used another cup since it arrived. Everything I drink from it tastes immaculate. My 3rd and 4th cups are in the mail. The Hydration+ and Tea Infusers are great too. Couldn't recommend more highly.		

Let's Test This.

Comments	Model Prediction	Actual
I was skeptical, but took a chance on the Black Friday deal. Haven't used another cup since it arrived. Everything I drink from it tastes immaculate. My 3rd and 4th cups are in the mail. The Hydration+ and Tea Infusers are great too. Couldn't recommend more highly.	Millennial	Millennial

Let's Test This.

Comments	Model Prediction	Actual
What kids? I can't afford to have kids.		

Let's Test This.

Comments	Model Prediction	Actual
What kids? I can't afford to have kids.	Gen Z	Gen Z

To Conclude: "Milking" Profits from Generations

- **Understanding customer base/ demographic for product: Gen Z/ Millennial?**
- **Shaping new strategies to better target preferred demographic:**
  - Framing advertisements to address top themes in each generation
  - Customising ads' type (e.g. presenting more reels to reach more Gen Zs)
- **Targeted ads to better resonate/ connect with specific audience in increasing interaction and sales with product**

Future work

Future Work

- Classify more generations e.g. Gen X, Boomers; to extend models across more social media platforms e.g. Facebook, increasing advertisement outreach to a larger audience

US Social Network User Penetration, by Platform, 2023

% of social network users, by generation

	Gen Z	Millennials	Gen X	Baby boomers
Snapchat	83.8%	43.7%	16.8%	7.5%
TikTok	82.2%	49.9%	23.5%	18.8%
Instagram	78.2%	75.3%	47.7%	27.7%
Facebook	55.6%	86.1%	87.8%	92.2%
Pinterest	39.2%	41.6%	39.7%	35.3%
Twitter	32.2%	26.7%	21.0%	15.3%
LinkedIn	32.1%	40.5%	41.9%	17.5%
Reddit	24.7%	31.1%	18.8%	9.5%

Future Work

- Exploration of other classification models in improving accuracy score e.g. K-nearest-neighbours
- Identification of slangs/ memes/ emojis specific to generation - incorporate into ads for better product-consumer connection
- Collaboration with social media platforms for live-feedback - feeding generation specific ads after classification

Let's Chat.