UNVCOVERING TEEN INTERESTS:

AN UNSUPERVISED NLP EXPLORATION

By: Jason Palumbo

MOTIVATION

"To curb the small but present semantic divide between young and old, I wanted to explore YouTube video transcript data in order to learn more about the type of interests, words, and phrases teens are using."
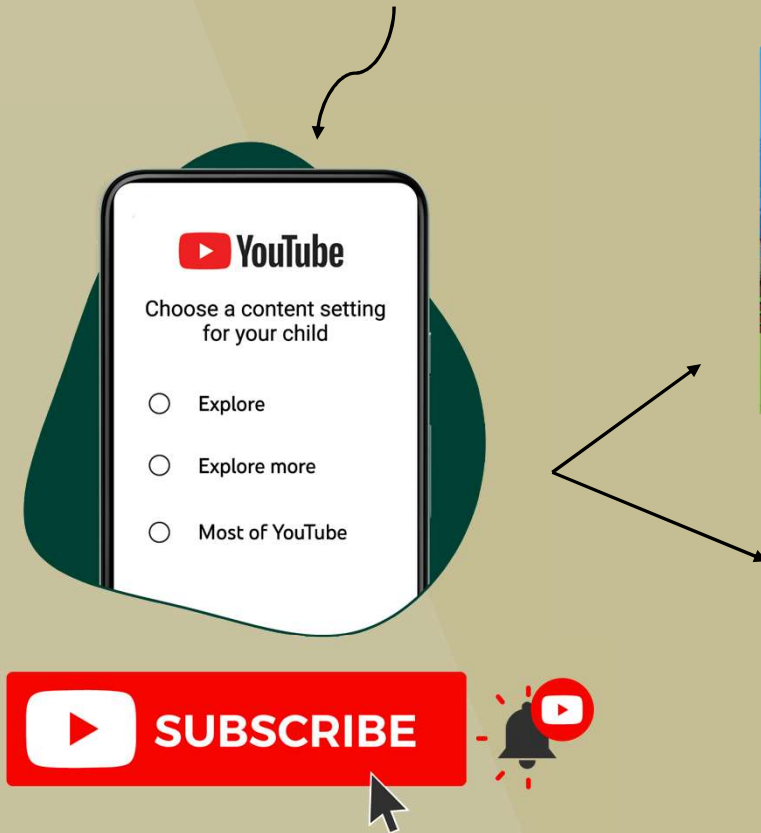
TARGETED DATA

PEW DIE PIE

"MINECRAFT"
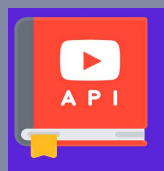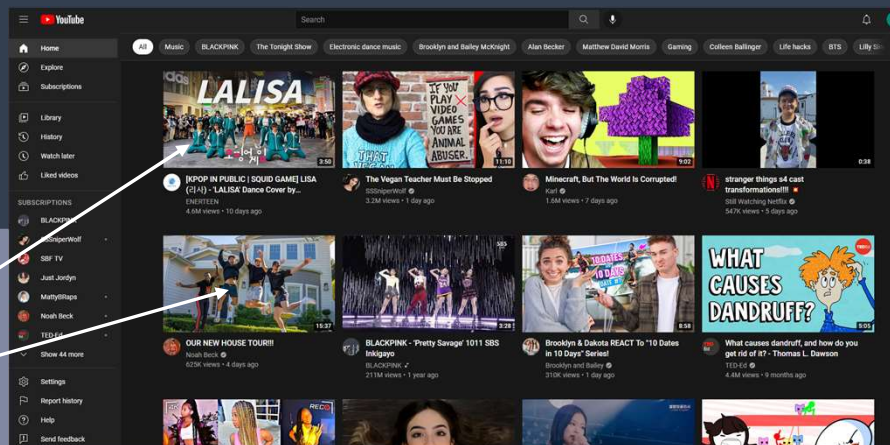
LILLY SINGH

"MAKEUP"
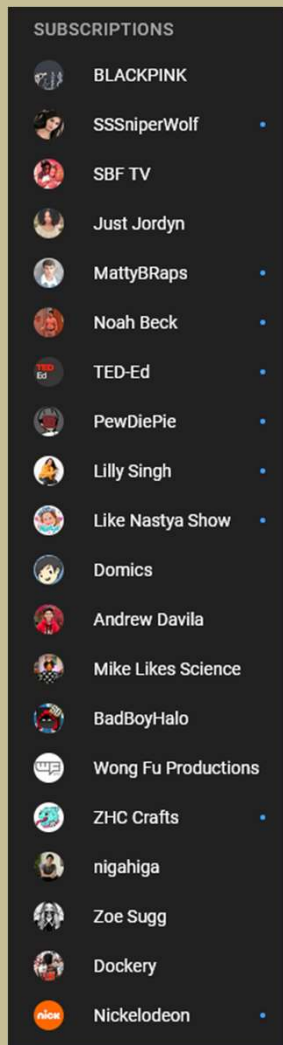
PIPELINE

URLs

COLLECT DATA	CLEAN DATA	PREPROCESS TEXT	TOPIC MODELING	SENTIMENT ANALYSIS
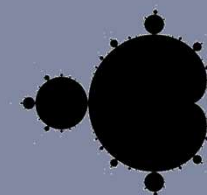
ONE ROW OF DATA CONSISTS OF A SINGLE VIDEO'S CLEANED TEXT DATA AND VIDEO METADATA

SUBSCRIPTIONS
- BLACKPINK
- SSSniperWolf
- SBF TV
- Just Jordyn
- MattyBRaps
- Noah Beck
- TED-Ed
- PewDiePie
- Lilly Singh
- Like Nastya Show
- Domics
- Andrew Davila
- Mike Likes Science
- BadBoyHalo
- Wong Fu Productions
- ZHC Crafts
- nigahiga
- Zoe Sugg
- Dockery
- Nickelodeon

- Only 1500 videos

- One source: YouTube

- Need for more breadth of subscriptions

RESULTS

DOCUMENT-TERM MATRIX	TfidfVectorizer
MODEL	Non-Negative Matrix Factorization ("NMF")
TOPICS RETURNED	15

TOPICS

- Video Reviews
- Hair Care
- Prizes
- Food
- Minecraft
- Games
- School
- Educational
- Pregnancy
- Dance/Music
- Halloween
- Clothes
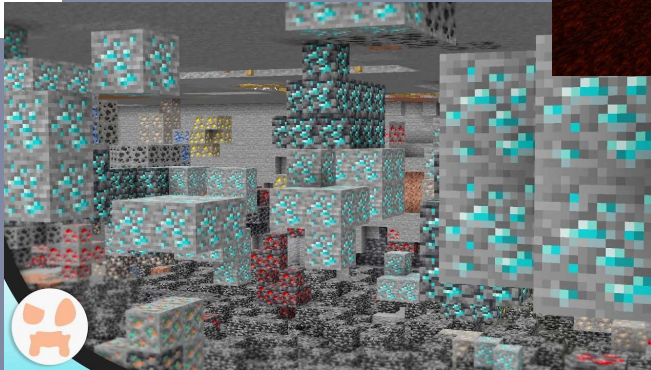- Challenges
- Beauty Care
- Christmas

NEW WORDS/CHARACTERS

MINECRAFT – MR.BEAST

"Creeper"

"Diamond"

"Nether"

"Chandler,"
"Mr. Beast,"
"Chris"

TOPIC EDA

Video Count/Topic

1. Beauty Care

2. Food

3. Minecraft

Avg. Comment Count/Topic

1. Challenges

2. Dance/Music

3. Beauty Care

Avg. Like/Dislike Ratio

1. Pregnancy

2. Educational

3. Beauty Care

Avg. Views per Topic

1. Prizes

2. Christmas

3. Dance/Music

SENTIMENT ANLAYSIS

Average 'text_polarity' per Topic
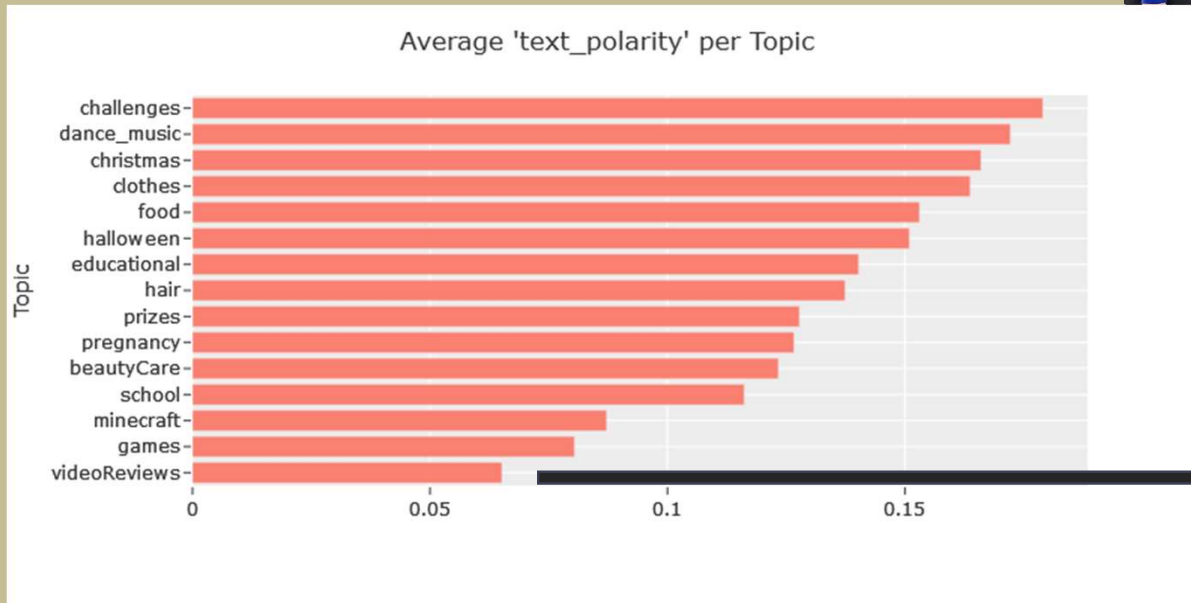
Case of the 'Karen's' and the 'SSSniperWolf' channel

CONCLUSION

- Latent topics exist

- Potential for Business Use Case

"This type of information could be immensely helpful to parents, educators, therapists, and psychologists who want to better understand and relate to the teens in their lives."

FUTURE WORK

- More Data/Diverse Data

- Deep Learning and Image Recognition

- Different Sources

THANK YOU

Jason Palumbo

GitHub: giasonep

APPENDIX