# BACKGROUND

- **Family Guy is an American adult animated sitcom created by Seth MacFarlane, which first aired in 1999**
- **Family Guy has become one of the most recognizable shows in the world due to its satirical humor which makes bold references to today's popular culture**
- **Recognizable by its distinctive style of cutaway jokes, flashbacks, and random comedy**
- **424 episode across 22 seasons have been released**

# OUR DATA

- Downloaded a dataset from Kaggle which includes 19 seasons of Family Guy, and contains all of the dialogue from each episode along with timestamps

# PROJECT MOTIVATION

- **Leverage NLP techniques learned in class to analyze humor within Family Guy**
  - **Since humor is very complex and subjective, our idea behind this is to gain a better understanding of what is generally considered "funny"**
- **Given our dataset has all of the episode dialogues, this gives us the opportunity to use the techniques learned in class and apply them to a real-world application**
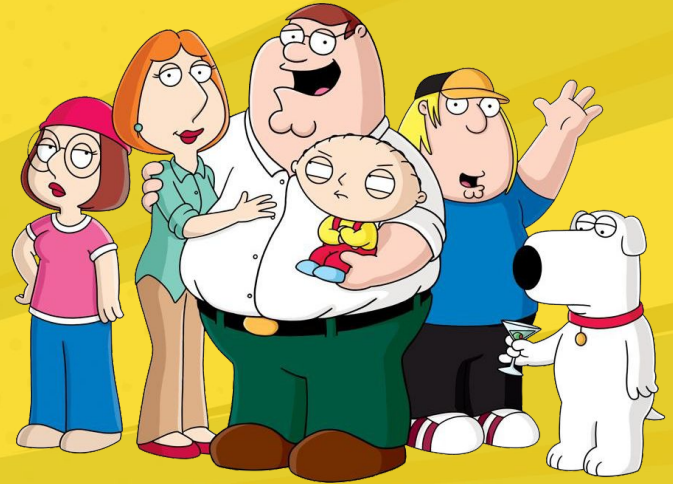
# DATA CLEANING

**Episodes average 20 - 27 minutes**

| Season | Episode | Time_Stamp | Dialogue |
|---|---|---|---|
| 1 | 1 | 00:00:07 | Mom, Dad, I found cigarettes in Greg's jacket. |
| 1 | 1 | 00:00:10 | Greg, were you smoking cigarettes? |
| 1 | 1 | 00:00:10 | No, Dad. |
| 1 | 1 | 00:00:12 | He's lying. There's no doubt about that. |
| 1 | 1 | 00:00:14 | Greg, I'm afraid your punishment will be four ... |

| Season | Episode | Dialogue |
|---|---|---|
| 1 | 1 | Mom, Dad, I found cigarettes in Greg's jacket.... |
| 1 | 2 | Look at Stewie. Isn't he adorable playing with... |
| 1 | 3 | It seems today that all you see is violence in... |
| 1 | 4 | When life is getting to you put some fresh in ... |
| 1 | 5 | Machiavelli! You've told me nothing I don't al... |
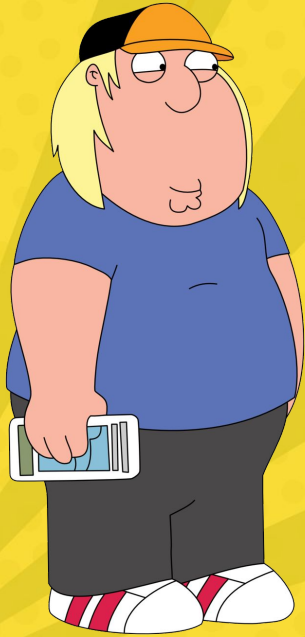
# OBJECTIVE

Analyze and rank episodes by humor using sentiment analysis

# METHODS

Keyword-based analysis and VADER sentiment analysis
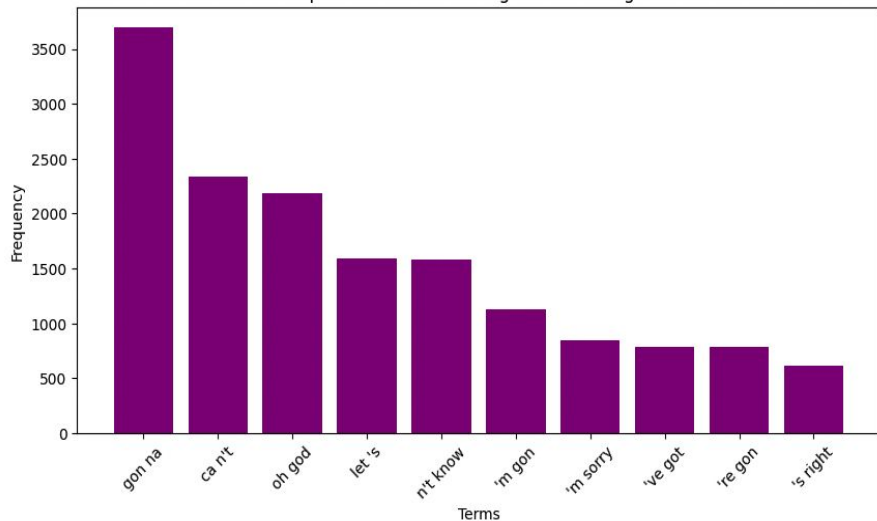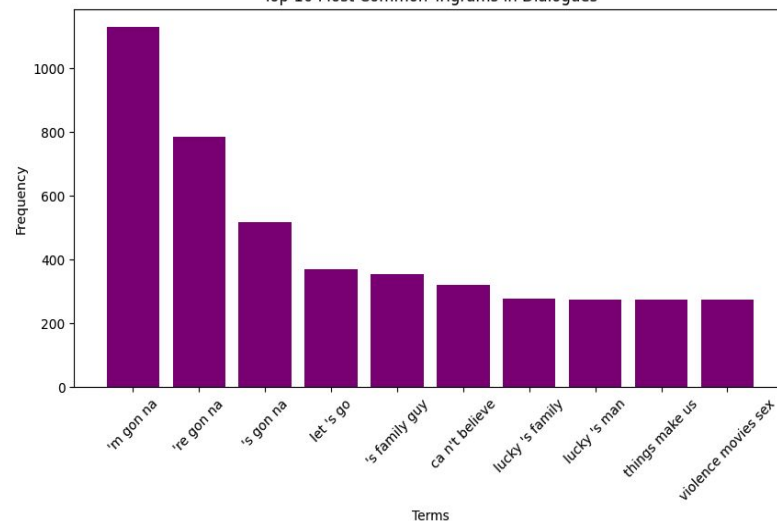
# KEYWORD SENTIMENT SCORING

Count based on predefined positive and negative keywords

# VADER SENTIMENT ANALYSIS

Utilizes pre-trained model for nuanced sentiment scoring.

| | Season | Episode | Sentiment_Score | Is_Funny |
|---|---|---|---|---|
| 0 | 1 | 1 | 2 | Funny |
| 1 | 1 | 2 | 0 | Not Funny |
| 2 | 1 | 3 | 0 | Not Funny |
| 3 | 1 | 4 | 5 | Funny |
| 4 | 1 | 5 | 2 | Funny |
| ... | ... | ... | ... | ... |
| 95 | 6 | 1 | 1 | Funny |
| 96 | 6 | 2 | 2 | Funny |
| 97 | 6 | 3 | 2 | Funny |
| 98 | 6 | 4 | 0 | Not Funny |
| 99 | 6 | 5 | 0 | Not Funny |

```python
import pandas as pd
from nltk.sentiment import SentimentIntensityAnalyzer

# NLTK's VADER SentimentIntensityAnalyzer
sia = SentimentIntensityAnalyzer()

# get sentiment score using VADER
def evaluate_sentiment_vader(dialogue):
    score = sia.polarity_scores(dialogue)
    return score['compound']  # Return the compound score

fm_combined_dialogue_only['VADER_Score'] = fm_combined_dialogue_only['Dialogue'].apply(evaluate_sentiment_vader)
```

```python
# group by season and episode
# calculate the mean sentiment score
episode_sentiment = fm_combined_dialogue_only.groupby(['Season', 'Episode']).agg({
    'VADER_Score': 'mean'
}).reset_index()



episode_sentiment.rename(columns={'VADER_Score': 'Average_Sentiment_Score'}, inplace=True)
```

# HOW MANY FUNNY EPISODES ARE THERE?



Distribution of Sentiment Scores



Proportion of Funny vs Not Funny Episodes

Not Funny

28.2%

71.8%

Funny
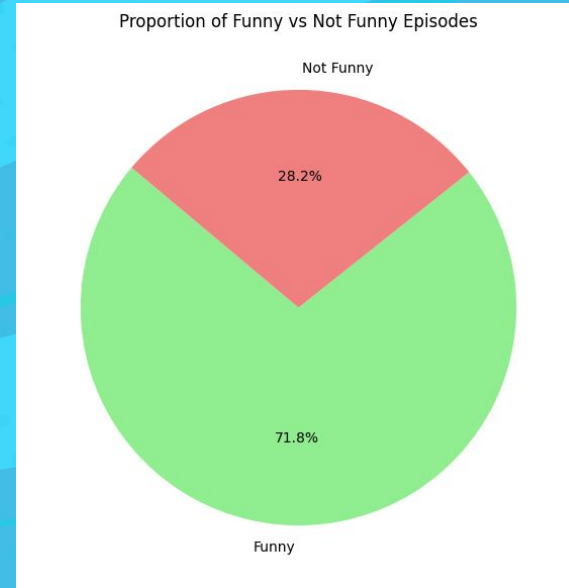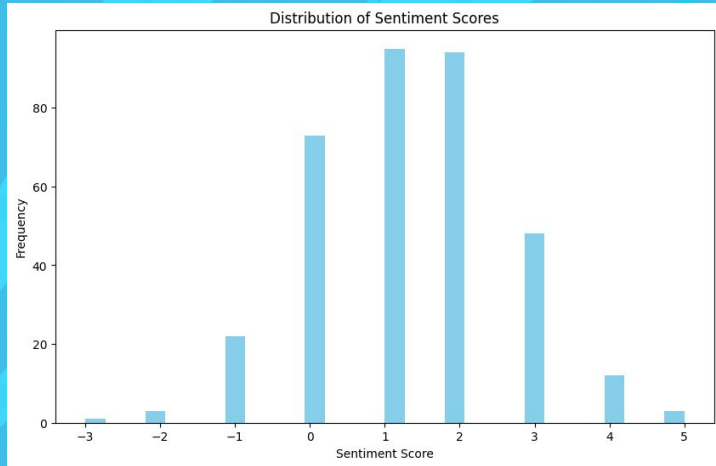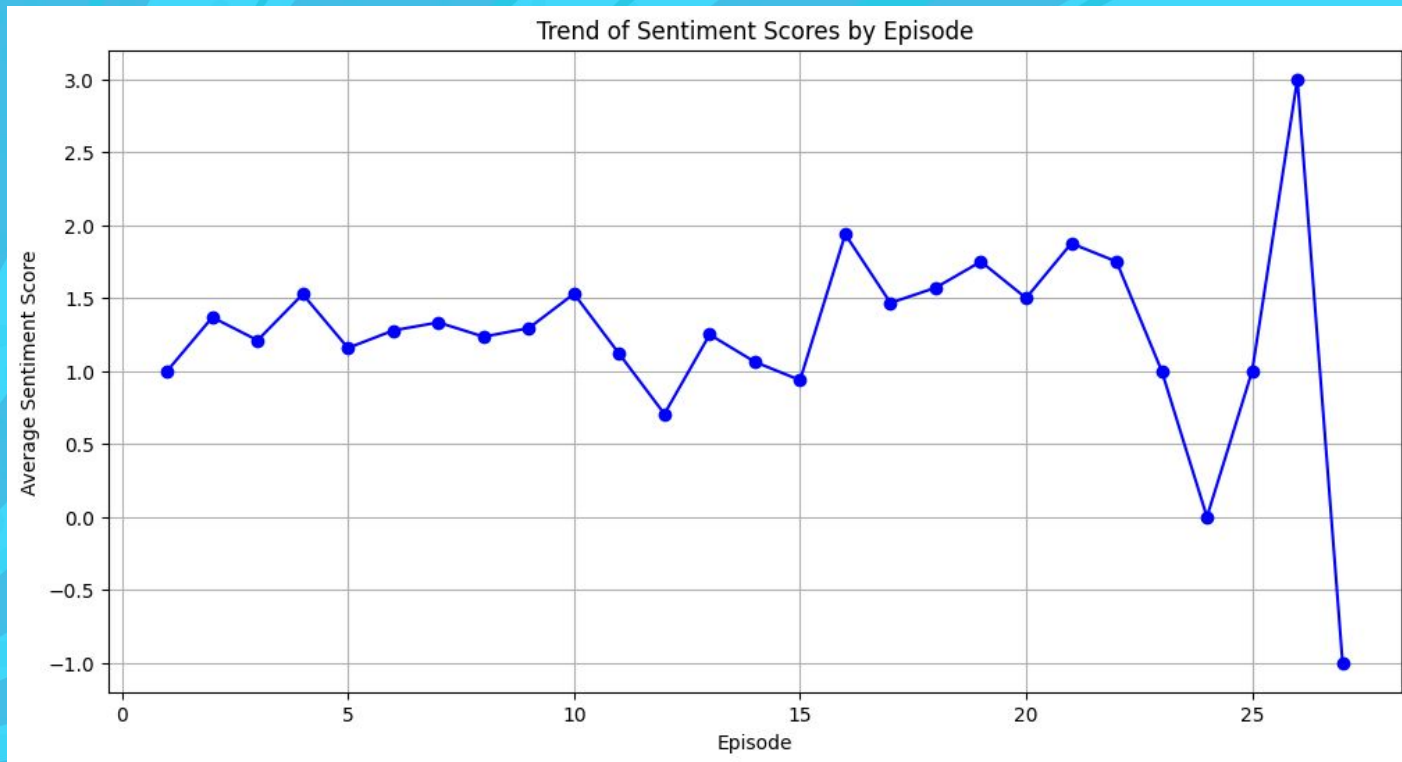
# Season 1

# RESULTS

| | Season | Episode | Average_Sentiment_Score | Humor_Rank |
|---|---|---|---|---|
| 132 | 8 | 10 | 1.0 | 1.0 |
| 140 | 8 | 18 | 1.0 | 1.0 |
| 46 | 3 | 19 | 1.0 | 1.0 |
| 40 | 3 | 13 | 1.0 | 1.0 |
| 250 | 14 | 5 | 1.0 | 1.0 |
| 35 | 3 | 8 | 1.0 | 1.0 |
| 120 | 7 | 14 | 1.0 | 1.0 |
| 32 | 3 | 5 | 1.0 | 1.0 |
| 103 | 6 | 9 | 1.0 | 1.0 |
| 29 | 3 | 2 | 1.0 | 1.0 |

# 2. CAN A JOKE BE DETERMINED BY ITS LENGTH?

# ANALYTICAL TECHNIQUES

## Feature extraction

- Sentence length and Word Count: essential metrics to gauge dialogue brevity and content density
- Vocabulary complexity: Evaluated to understand linguistic complexity

## Sentiment Analysis

- Tools Used: AFINN, BING, and TextBlob for sentiment and polarity scores
- Findings: Explored correlations between positive sentiments and increased humour

# LINGUISTIC PATTERNS

## Linguistic Analysis

- **POS Tagging: Identified common grammatical structures in humorous dialogues.**
- **NER: Highlighted frequent named entities involved in jokes**
  - **Indicating character or situational humour**

```
[44] nltk.download('maxent_ne_chunker')
     nltk.download('words')

     def get_named_entities(text):
         tokens = nltk.word_tokenize(text)
         tags = nltk.pos_tag(tokens)
         entities = nltk.chunk.ne_chunk(tags)
         return [entity for entity in entities if isinstance(entity, nltk.tree.Tree)]

     fm['Named_Entities'] = fm['Dialogue'].apply(get_named_entities)
```
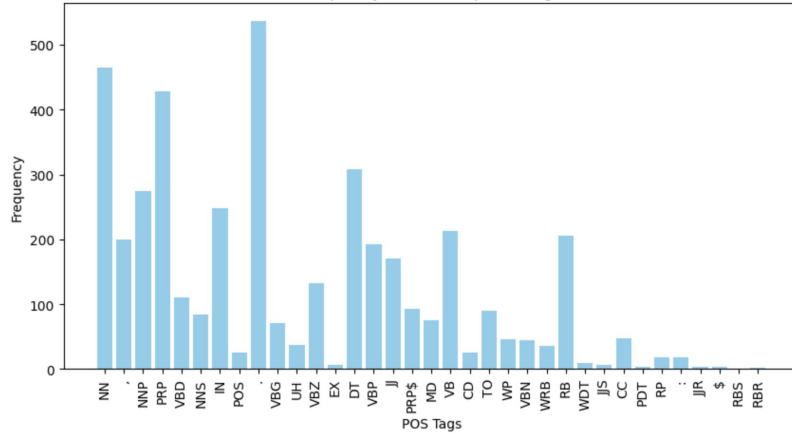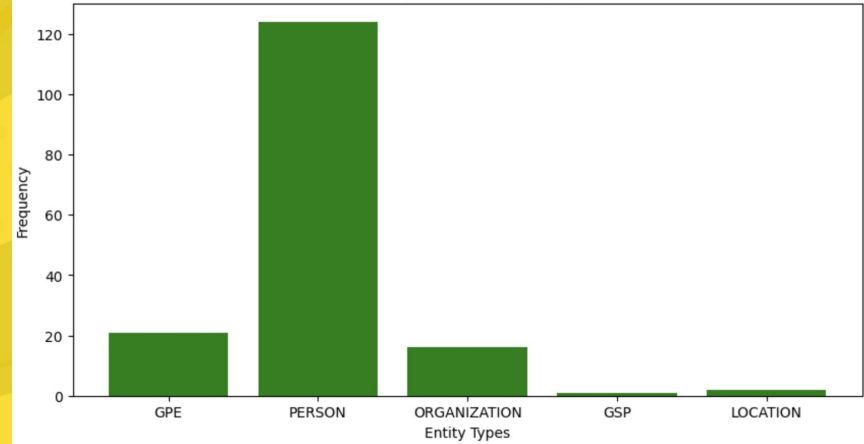
# VISUALIZING POS AND NER

# CLUSTERING

## METHOD

Used KMeans to group dialogues based on linguistic features

```
Examining dialogues in cluster 4
83                    That's just crazy enough to work.
116   Peter, I like you. But I need you to be more t...
135   Just not as much. So it might get a little com...
146   When she worries, she says, I told you so and:...
171   Screw this! I just came over to buy some firew...
180                      Lois, we just had dinner.
183                                        I just...
188                  It's just... It's not healthy.
212           I hate lying to Lois. It's just...
244   No. I just rented it. But they're gonna be tic...
283                        Just like the Kennedys.
285   The man I married would never think he could f...
297   Yeah, I know. I'm just making conversation. Co...
337                  You just want your toy back.
358   Okay, everybody, I feel really bad about what ...
424   Honey, it's not gonna go away just because you...
441                  Jeez. Did I just hit that ostrich?
488   You just have to remember that life has its li...
493   You just knocked out cable TV for the whole town!
506   I can't believe you just sold out your own dau...
```

## OUTCOME

Successfully identified clusters that significantly contained humor based on dialogue length and linguistic cues

```
       Season  Episode  Time_Stamp  \
35          1        1    00:02:26
63          1        1    00:03:48
65          1        1    00:03:49
94          1        1    00:05:08
100         1        1    00:05:30
...       ...      ...         ...
154938     19        5    00:05:46
155015     19        5    00:09:34
155132     19        5    00:15:58
155223     19        5    00:19:46
155229     19        5    00:20:08

                                                  Dialogue  \
35                                      Yeah, it's all right.
63                                             Right here.
65                                  All right. What do I win?
94                      I almost walked right into that one. God!
100                                      Mom, are you all right?
...                                                      ...
154938    All right, we'll find something else. Thank yo...
155015    I mean, this is the Mafia, not the Laugh Facto...
155132                              All right, I'll do it.
155223                   All right, nobody do anything stupid.
155229                           All right. So long, you two.

                                   NRC_Sentiment AFINN_Sentiment  \
35                                       Neutral        Positive
63                                       Neutral         Neutral
65                                       Neutral        Positive
94           anticipation,fear,joy,positive,trust        Positive
100                                      Neutral         Neutral
...                                          ...             ...
154938                                   Neutral        Positive
155015         fear,joy,negative,positive,surprise        Positive
155132         anger,disgust,fear,negative,sadness         Neutral
155223                                  negative        Negative
155229                              anticipation         Neutral
```

```
        AFINN_Sentiment_Score BING_Sentiment  Word_Count  Polarity  \
35                          2       Positive           4  0.285714
63                          0       Positive           2  0.285714
65                          4       Positive           6  0.542857
94                          1       Positive           8  0.357143
100                         0       Positive           5  0.285714
...                       ...            ...         ...       ...
154938                      2       Positive          10  0.272857
155015                      1       Positive          11  0.091071
155132                      0       Positive           5  0.285714
155223                     -2       Positive           6 -0.257143
155229                      0       Positive           6  0.117857

        Subjectivity                                     POS_Tags  \
35          0.535714  [(Yeah, UH), (,, ,), (it, PRP), ('s, VBZ), (al...
63          0.535714                  [(Right, RB), (here, RB), (., .)]
65          0.467857  [(All, DT), (right, NN), (., .), (What, WP), (...
94          0.535714  [(I, PRP), (almost, RB), (walked, VBD), (right...
100         0.535714  [(Mom, NN), (,, ,), (are, VBP), (you, PRP), (a...
...              ...                                            ...
154938      0.397857  [(All, DT), (right, JJ), (,, ,), (we, PRP), ('...
155015      0.441071  [(I, PRP), (mean, VBP), (,, ,), (this, DT), (i...
155132      0.535714  [(All, DT), (right, NN), (,, ,), (I, PRP), ('l...
155223      0.767857  [(All, DT), (right, NN), (,, ,), (nobody, NN)...
155229      0.467857  [(All, DT), (right, NN), (., .), (So, RB), (lo...

                    Named_Entities  Cluster
35                              []        3
63                              []        3
65                              []        3
94                              []        3
100                  [[(Mom, NN)]]        3
...                            ...      ...
154938                          []        3
155015   [[(Laugh, NNP), (Factory, NNP)]]   3
155132                          []        3
155223                          []        3
155229                          []        3
```
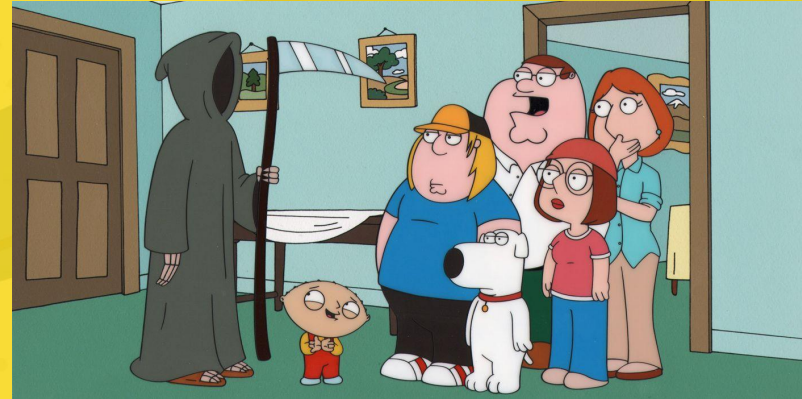
# IS A JOKE APPROPRIATE?

# FAMILY GUY RATING



-   The Movie Picture Association (MPAA) rating that has been assigned to Family Guy is: "language, some sexual content and drug use"
-   Early seasons are rated 12+
-   Season 10 and beyond are rated 14+

# IS A JOKE APPROPRIATE?

- **Profanity Analysis of classified jokes**
    - **Does a joke contain profanity?**
    - **How many of the classified jokes use profanity?**
- **Check Sentiment Scores**
- **Determine if a joke is inappropriate or appropriate**
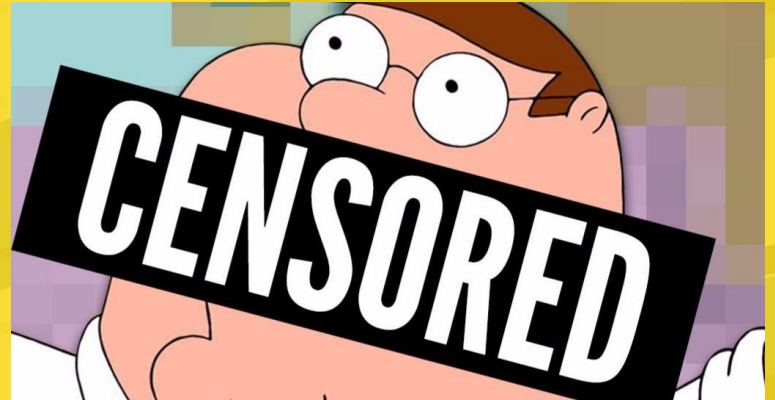
# PROFANITY ANALYSIS

- **Package: better_profanity (0.7.0)**
- **Steps:**
  - **Extract profane words**
  - **Identify words**
  - **Count the frequency**



```python
from better_profanity import profanity

profanity.load_censor_words()


def extract_profanity_words(dialogues):
    profane_words_list = []
    for dialogue in dialogues:
        words = word_tokenize(dialogue.lower())

        profane_words = [word for word in words if profanity.contains_profanity(word)]
        profane_words_list.extend(profane_words)
    return profane_words_list
```
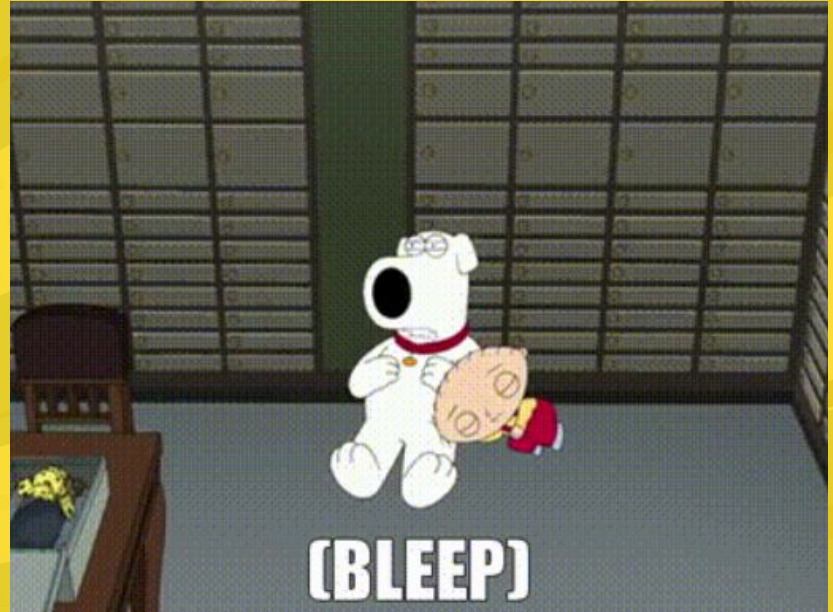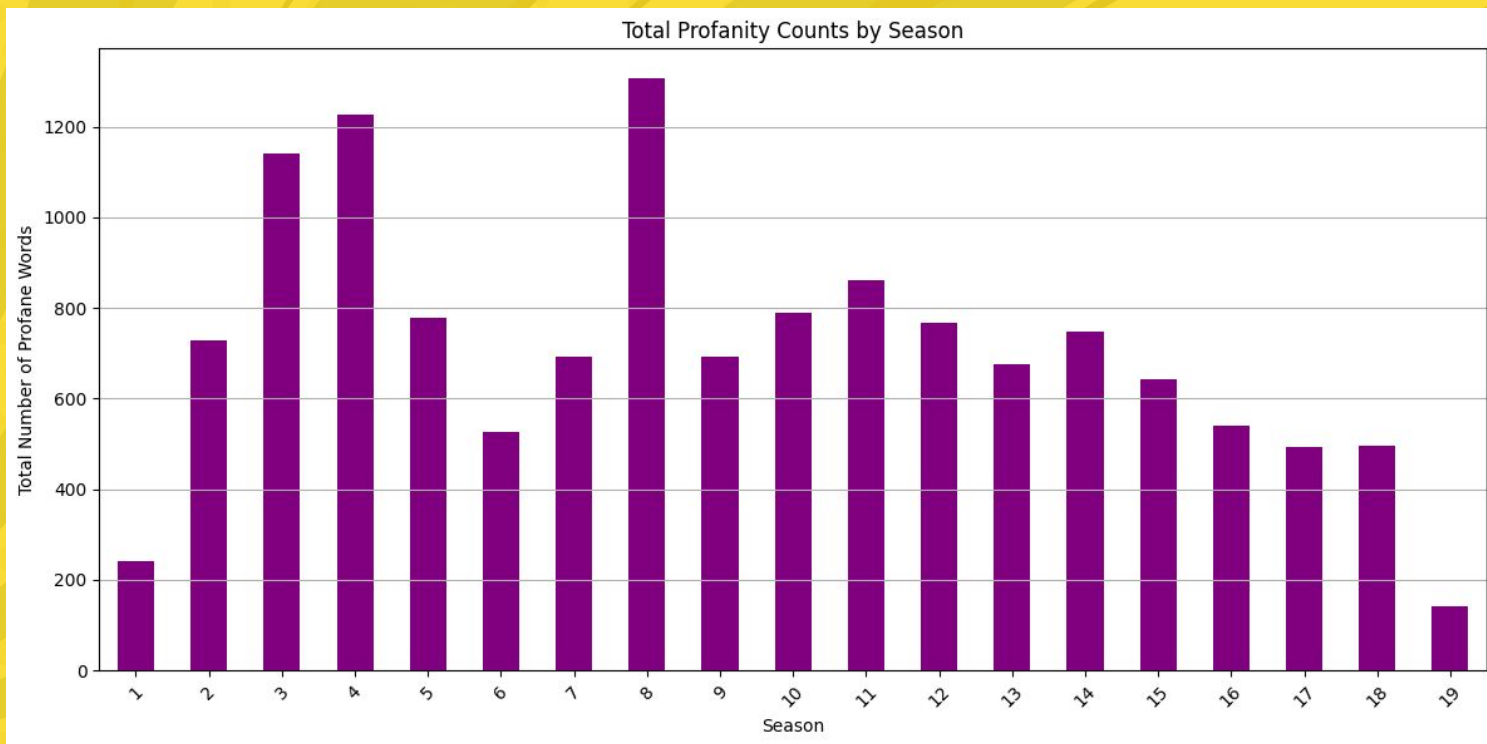
# *NOTABLE MENTIONS*

**Episode 1,1**
- **Stupid: 35**
- **Fat: 29**
- **Poop: 13**
- **F**king: 2**

# SEASONAL PROFANITY



Total Profanity Counts by Season

# LOOKING BEYOND PROFANITY

- **Using vader_lexicon and SentimentIntensityAnalyzer we analyzed the scores of each joke to determine if it was appropriate**
- **Positive, Neutral, and Negative**

```
     Season   Episode                                    Dialogue
83      1.0      1.0              That's just crazy enough to work.
116     1.0      1.0     Peter, I like you. But I need you to be more t...
135     1.0      1.0     Just not as much. So it might get a little com...
146     1.0      1.0     When she worries, she says, I told you so and:...
171     1.0      1.0     Screw this! I just came over to buy some firew...

     Appropriateness
83       Appropriate
116      Inappropriate
135      Appropriate
146      Appropriate
171      Inappropriate
```

```
Number of Appropriate dialogues: 12142
Number of Inappropriate dialogues: 989
```

# FUTURE WORK

- **Video and Audio Analysis**
  **Analyze videos and determine how does the inputted audio affect a joke**
- **Time Analysis**
  **Can we predict if a joke is more likely to be funny depending on when it appears in an episode?**
- **Edinburgh Fringe Joke detector**
- **Different Languages**

# VIDEO ANALYSIS

```
with quote as source:
    r.adjust_for_ambient_noise(source)
    audio=r.record(source)
```

```
result = r.recognize_google(audio)
```

```
  {'transcript': "do whatever you have to do Brian butt never let him know he's got talent yeah you're right I won't be so hard after all
I've certainly done worse I replaced Peter's I Can't Believe It's Not Butter with real butterfly",
   'confidence': 0.37292942},
```

```
'Do whatever you have to do, Brian, but never let him know he's got talent.Yeah, you're right.It won't be so hard.After all, I've certainly done worse.
I replaced Peter's I Can't Believe It's Not Butter with real butter.Buh... wuh... suh... muh-wuh...I... I can't... I can't believe...Muh... wuh...I do
n't know, Doctor.Looking back, I think it may have been real butter.Your husband murdered three children.'
```
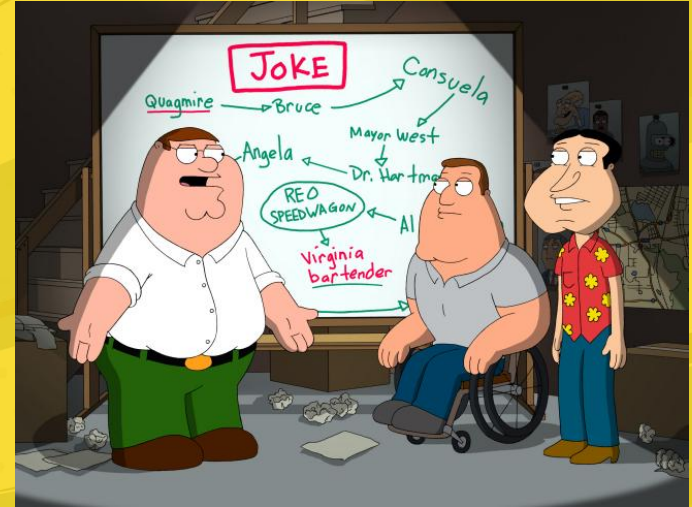
```
Accuracy: 0.16049382716049382
```

```
'Not Funny'
```

# AUDIO ANALYSIS

- **Audio Emotion Recognition**
  **We would like to analyze how does the background music, sound effects, and character's tone affect a joke delivery**
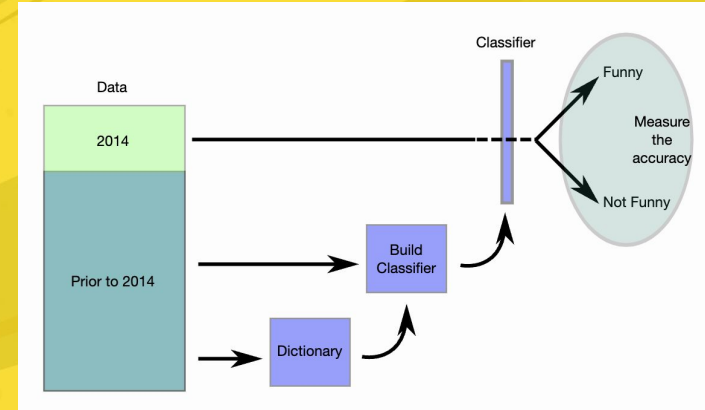
# TIME ANALYSIS

- **Given all of the information dialogue's have timestamps, can we determine when a joke is funniest?**

| Season | Episode | Time_Stamp | Dialogue |
|--------|---------|------------|----------|
| 1.0 | 1.0 | 00:04:33 | That's just crazy enough to work. |
| 1.0 | 1.0 | 00:06:14 | Peter, I like you. But I need you to be more t... |
| 1.0 | 1.0 | 00:07:20 | Just not as much. So it might get a little com... |

# EDINBURGH FRINGE

- **Use an open-source joke detection model inspired by the famous Edinburgh Fringe Festival**
- **Data includes the top ten jokes of every year since 2011**

THANK YOU