# Study to Improve Number of Engagements Per Tweet: Fresh Tape Media

Jess Tillis

# Fresh Tape Media:
# Current State of Twitter Content Creation

- Sports broadcasters/Professional Athletes/Sports Teams/Sports Leagues contract with Fresh Tape Media (FTM) to create made-for-social content
- Currently, FTM “goes with their gut” or with client suggestions when deciding on the content of videos/GIFs.

***Business Question: What type of content sentiment generates the most views/engagements on Twitter?***

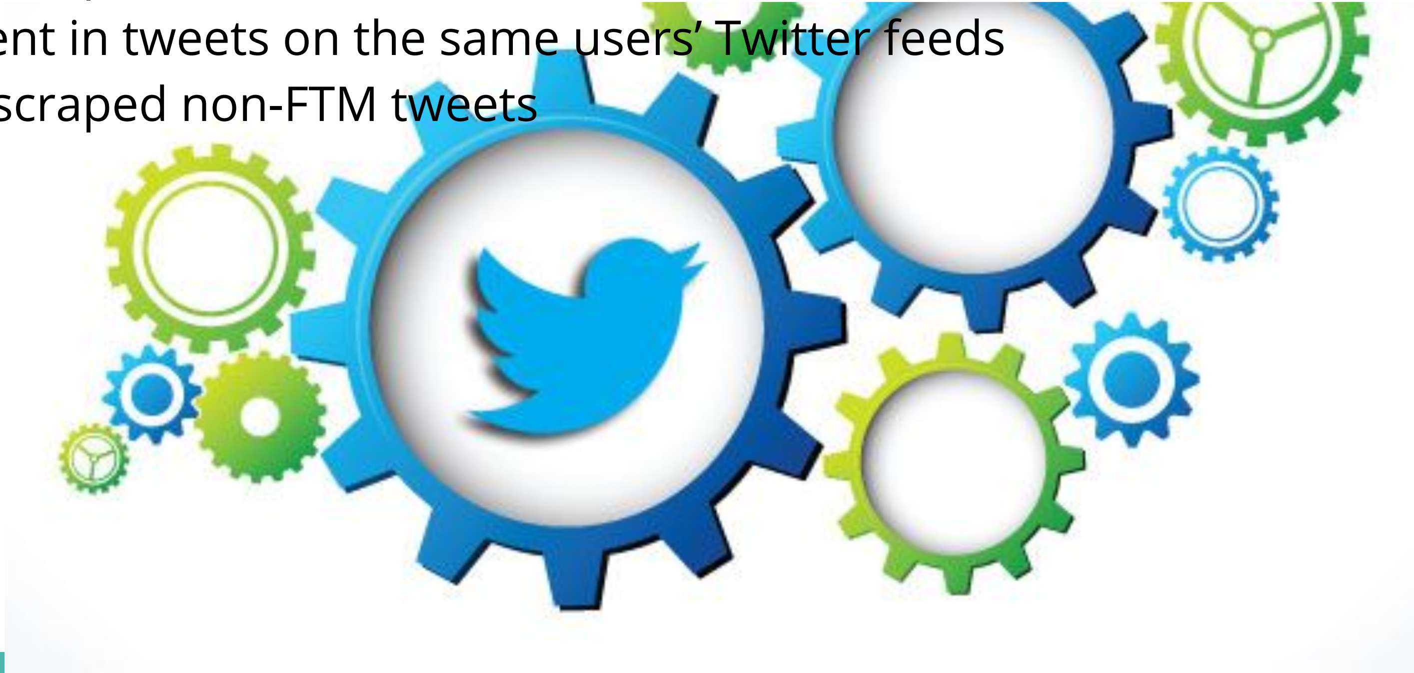
# Steps Taken To Answer the Business Question

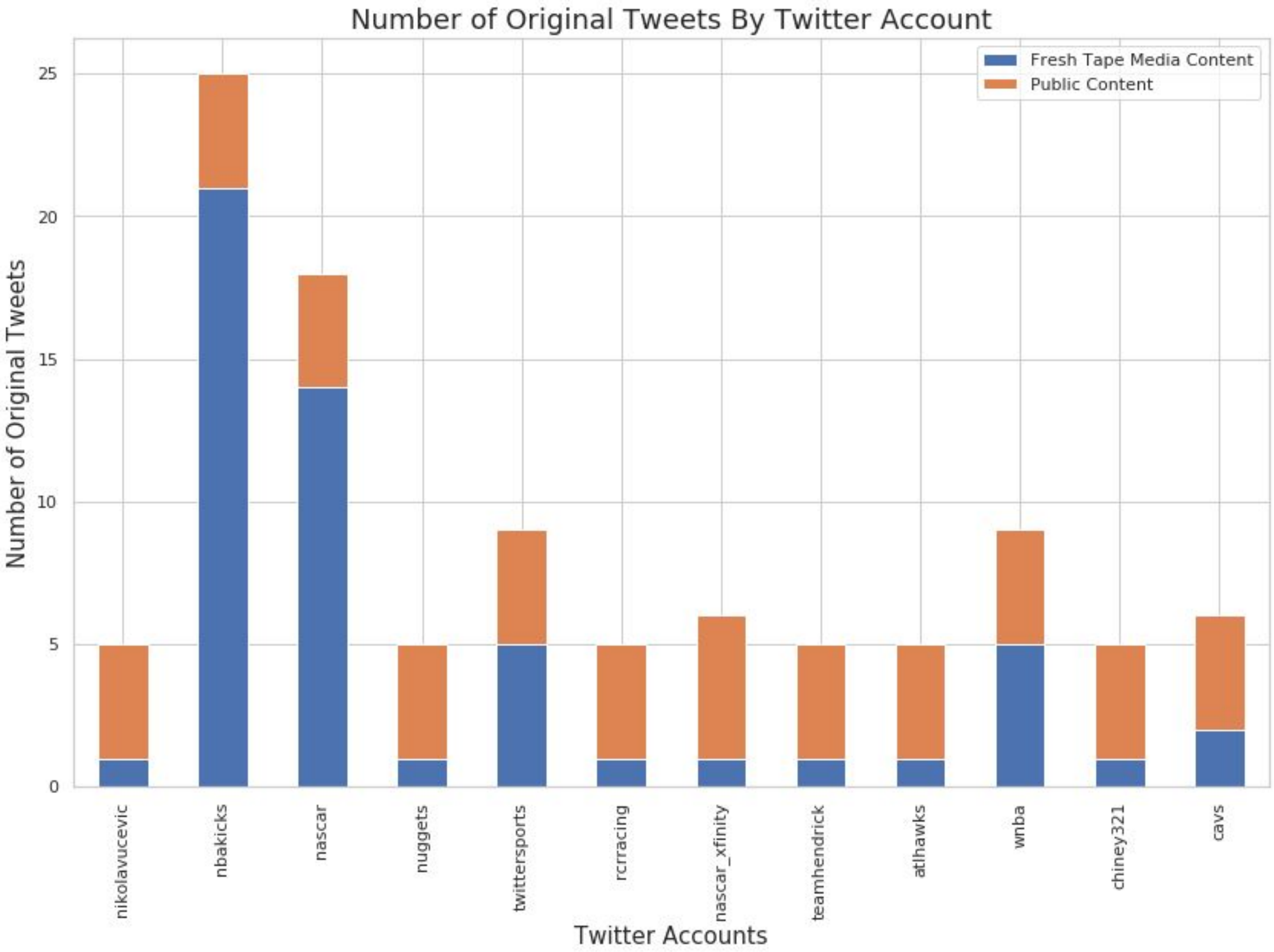
# Steps Taken To Answer the Business Question

1. Scraping the Twitter API and feature engineering
2. Deciding on the key metric to model/predict
3. Feature Engineering & Selecting the model with the most predictive power
4. Feature engineering to improve predictive accuracy
5. Using feature importance to pull out most valuable features when predicting the success of a tweet
6. Considering what's next

# Scraping the Twitter API and Feature Engineering

- FTM content in tweets on clients' Twitter feeds
- Replies to the scraped FTM content
- non-FTM content in tweets on the same users' Twitter feeds
- Replies to the scraped non-FTM tweets




Number of Original Tweets By Twitter Account
Fresh Tape Media Content
Public Content
Number of Original Tweets
0
5
10
15
20
25
nikolavucevic
nbakicks
nascar
nuggets
twittersports
rcrracing
nascar_xfinity
teamhendrick
atlhawks
wnba
chiney321
cavs
Twitter Accounts

# Deciding on the key metric to model/predict

Views vs. Likes

Likes vs. Percent Likes

Percent Likes as A Classification Problem

1.9%

Average Percent Likes for FTM Content

11%

Highest Percent Likes for FTM Content

2.4%

Average Percent Likes for non-FTM Content

13%

Highest Percent Likes for non-FTM Content

Percent likes < median … "Below Average"

median < Percent likes < mean … "Average"

Percent likes > mean … "Above Average"

# Feature Engineering & Selecting the model with the most predictive power

| Model | Predictive Power on Test Data |
|---|---|
| k-NN | 44.19% |
| Logistic (Lasso) Regression | 53.51% |
| Logistic (Ridge) Regression | 53.49% |
| Random Forest | 64.31% |

# Feature engineering to improve predictive/explanatory accuracy

1. Removing users
2. Removing pronouns
3. Removing more generic words like "good"/"great"
4. Normalizing the data
5. Limiting the number of features to 10

# The Results

Most important POSITIVE features of an above average tweet: 'need', '🔘', 'U', 'soo', '❤', 'luck', '🐍', '치치치치', '🔥'

Most important NEGATIVE features of an above average tweet: 'fan', '👎', 'not', '🏀', 'nbatwitter', 'jimmie', '👍', '🙌', 'look', 'race'

# Considering what’s next

- Paid access to Twitter API
- Turning the code/model into a website for Fresh Tape Media’s everyday use

# Questions?