Social Network Analysis on Formula 1

Members:
Vennela Muvva
Krishndutt Shirodkar
Bhagyashri Patil
Priyanshu Singh
Shayne Mogilski
Chi Hsu

What is Formula 1?

- Formula 1 is one of the most popular and technologically advanced sports in the world. It is the pinnacle of motorsport/automobile technology.

- It is the biggest and highest degree of international racing for single-seater formula racing cars.

- .A Grand Prix race week comprises a practice day, qualifying race day, and final race day. Together all these races are called the F1 season, governed and sanctioned by the Fédération Internationale de l'Automobile (FIA).

- In the race, FIA evaluates the performance of the driver and the constructor.

- At the end of a season, the FIA combines the scores made by each and awards two annual World Championships: F1 Driver Championship and Constructor Championship.

Why Formula 1?

- This generationally established motorsport predominantly taking place in Europe is picking up popularity in other huge markets since its takeover by a US media/entertainment company, Liberty Media.

- F1 as an organization in itself has pushed for increasing its presence in potential markets like the US through the introduction of a Netflix series 'Drive To Survive' and ever-increasing popularity on social media channels

- Teams have accepted this new influx of fans with open-arms and have moved on from 'closed-door elite racing club'

Common social media platforms used in F1

- Instagram
- Twitter
- TikTok
- Reddit

Our focus points for data collection included:
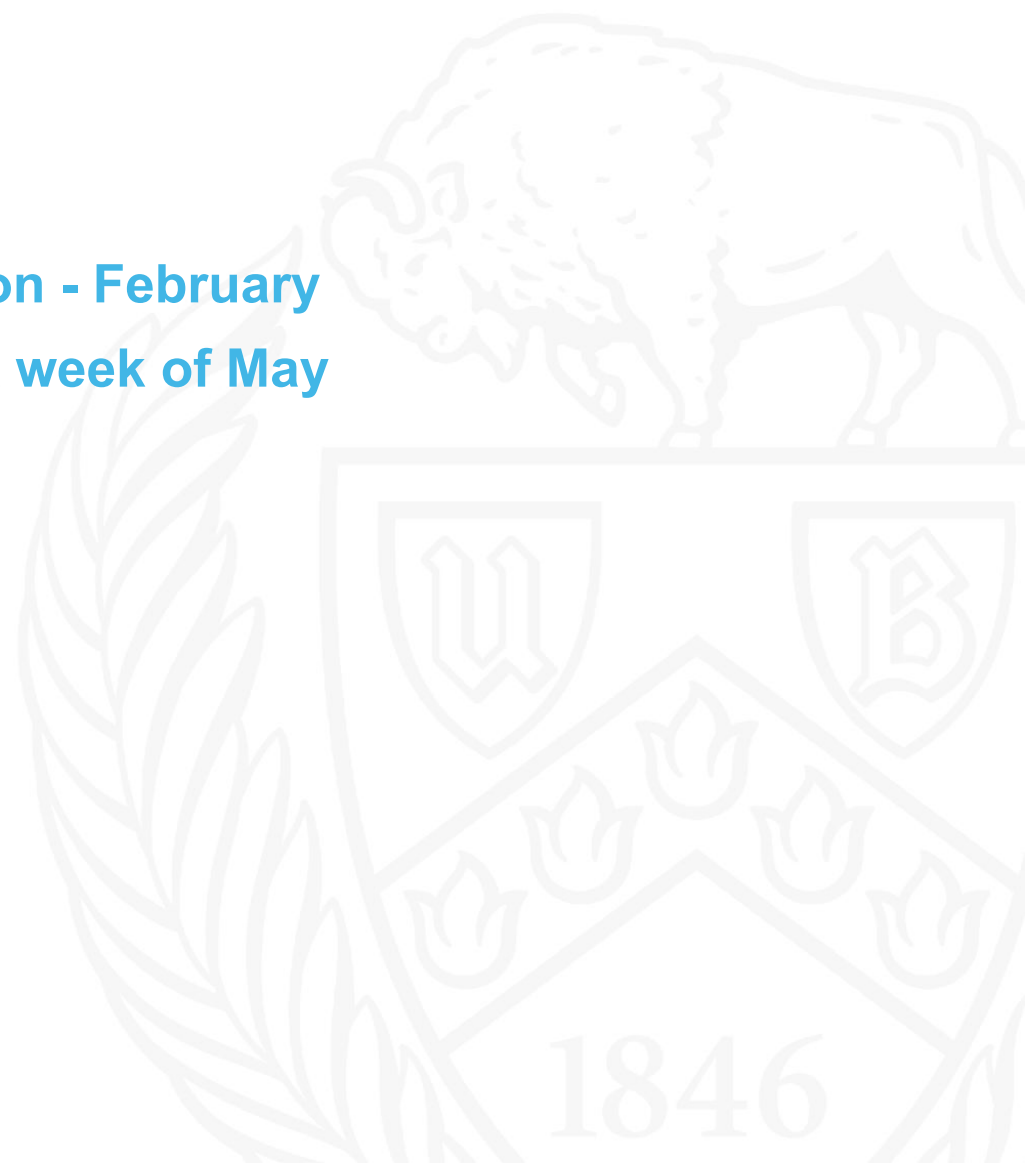
- Twitter
- Reddit

Data Collection Period:

- Preliminary Data collected for exploration: **Launch Season - February**
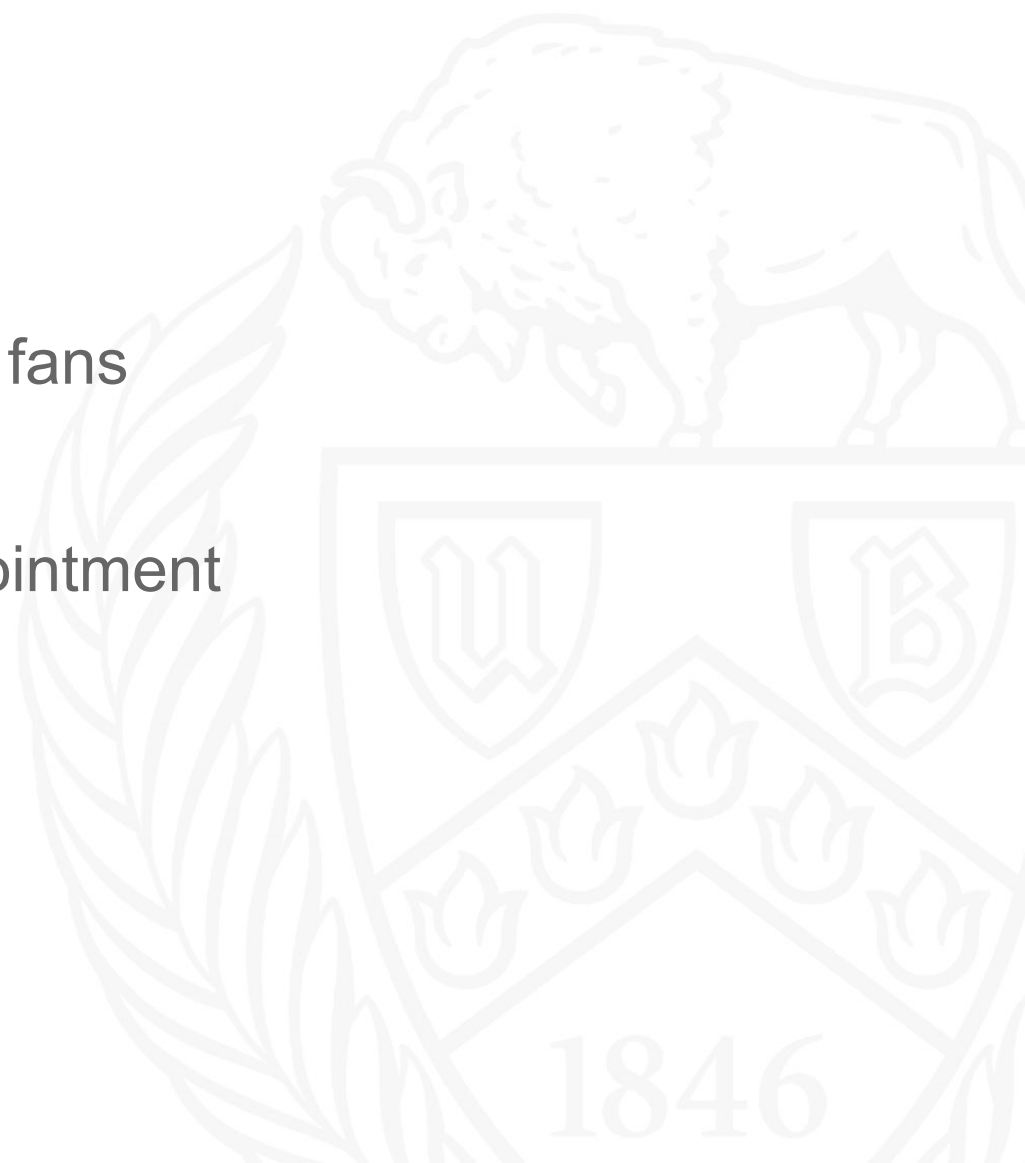- Data Collected for Analysis: **3rd week of March until 1st week of May**

Hashtags/Threads:

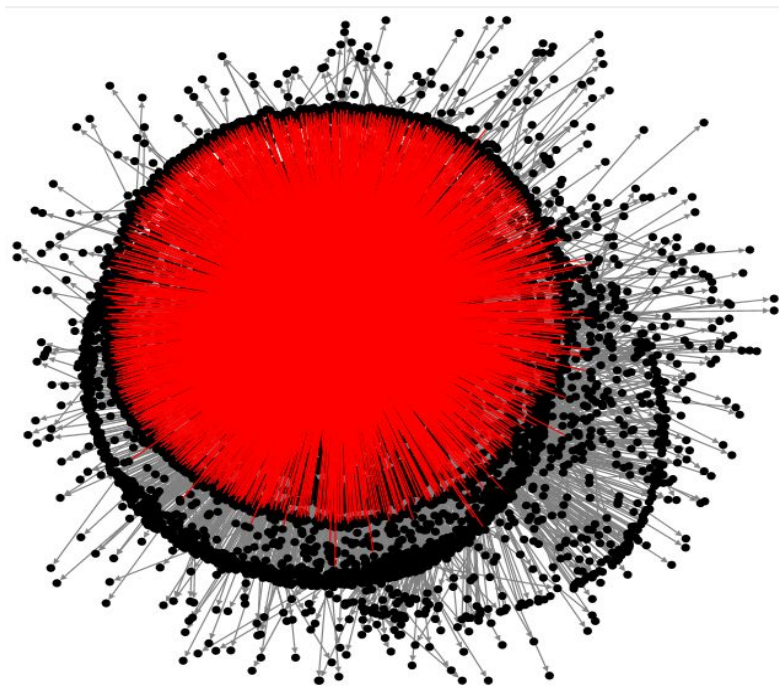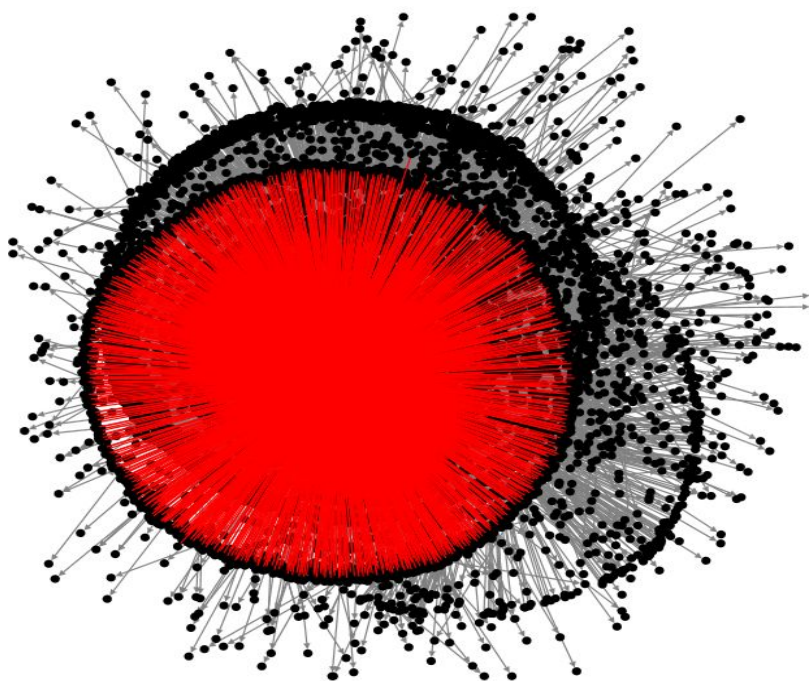- Formula 1
- AstonMartinF1

Premise/Objective:

- Track changes in opinion/behavior patterns of fans across different social media platforms

- Track apparent changes in behaviour/ disappointment of Aston Martin fans across different platforms
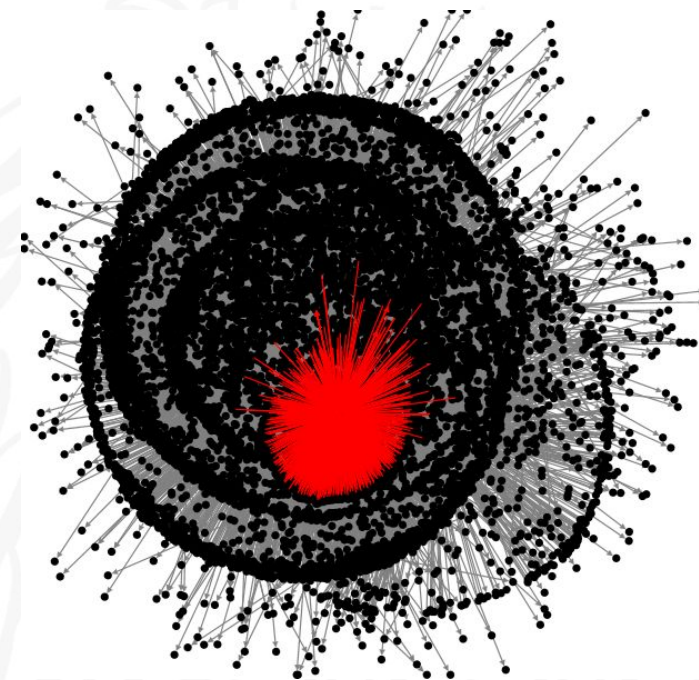
Top 3 Influential Nodes

@F1

@vegas

@mgmresorts

Formula 1 Network Metrics:

Nodes/Accounts	In Degree	Betweenness Centrality	Closeness Centrality
@F1	7730	69652278.71	0.886
@vegas	3514	7391549.99	0.595
@mgmresorts	1012	292580.90	0.502

Generalized Metrics Inference:

- A few nodes act as initiators of conversations in the network indicated by the relatively low out-degree and high in-degree scores

- These nodes set the topic of the conversation and the less influential nodes follow the tone set by these influential nodes through comments, retweets, and mentions

- Any ingenious opinions are subdued by the overall tone set

- This induces a sort of imaginary restriction of the topic

WordCloud

- Word Cloud or Tag Clouds is a visualization technique for texts for visualizing the tags or keywords from the websites.

- These keywords are single words that depict the context of the webpage the word cloud is being made from.

- Each word in this cloud has a variable font size and color tone which helps to determine words of prominence.

Libraries utilised in Python

Numpy
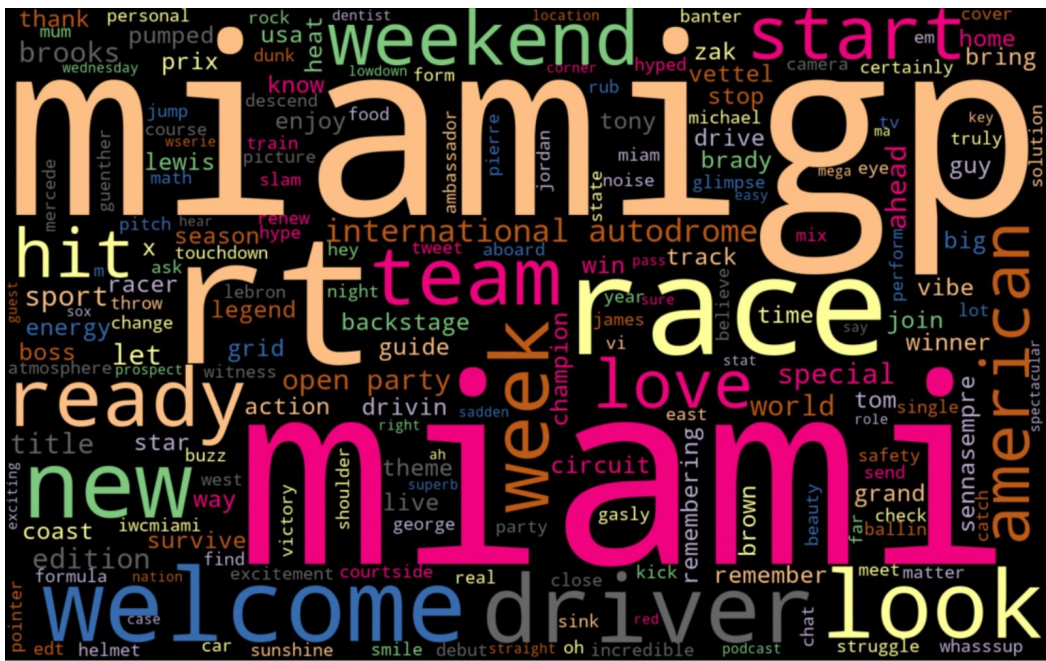**Pandas**
Spacy
Matplotlib
WordCloud

Formula 1

Twitter word-cloud:

Reddit word-cloud:

Formula 1 word-cloud inference:

- Users on Twitter generally talk about the event and the celebrities joining the event, Miami GP, which was a new addition to the 2022 race calendar and a new US race location

- Users on Reddit generally talk about the actual race, teams, drivers, cars, and other aspects pertaining to racing rather than commercial aspects of the event

What is N-grams?

N-grams is a sequence of n items from a given sample of text or speech collected from a text or speech corpus.

We have used the following python libraries to implement n-grams:

1. Pandas
2. String
3. Clean
4. nltk,spacy,wordcloud, ngrams

N-grams for Twitter Formula1 data

n=2

```
[(('miamigp', 'f1'), 29),
 (('miamigp', 'f1miami'), 18),
 (('miami', 'international'), 16),
 (('international', 'autodrome'), 16),
 (('f1miami', 'miamigp'), 16),
 (('race', 'week'), 16),
 (('opening', 'party'), 12),
 (('welcome', 'miami'), 12),
 (('landonorris', 'georgerussell63'), 8),
 (('allnew', 'miami'), 8),
 (('miami', 'vibe'), 8),
 (('miami', 'welcome'), 8),
 (('f1miami', 'miami'), 8),
 (('pumped', 'miamigp'), 8),
 (('love', 'american'), 8),
 (('grand', 'prix'), 8),
 (('world', 'title'), 8),
 (('tom', 'brady'), 8),
 (('miami', 'ready'), 8),
 (('autodrome', 'miamigp'), 8)]
```

n=3

```
[(('miami', 'international', 'autodrome'), 16),
 (('international', 'autodrome', 'miamigp'), 8),
 (('welcome', 'race', 'week'), 8),
 (('miamigp', 'f1', 'f1miami'), 6),
 (('miamigp', 'f1miami', 'miami'), 6),
 (('certainly', 'enjoyed', 'miami'), 4),
 (('enjoyed', 'miami', 'far'), 4),
 (('look', 'getting', 'miamigp'), 4),
 (('know', 'struggle', 'math'), 4),
 (('struggle', 'math', 'landonorris'), 4),
 (('math', 'landonorris', 'georgerussell63'), 4),
 (('landonorris', 'georgerussell63', 'enjoyed'), 4),
 (('georgerussell63', 'enjoyed', 'backstage'), 4),
 (('enjoyed', 'backstage', 'banter'), 4),
 (('backstage', 'banter', 'miam'), 4),
 (('banter', 'miam', 'first'), 4),
 (('miam', 'first', 'glimpse'), 4),
 (('first', 'glimpse', 'action'), 4),
 (('glimpse', 'action', 'around'), 4),
 (('action', 'around', 'miami'), 4),
 (('around', 'miami', 'safety'), 4),
 (('miami', 'safety', 'car'), 4),
 (('safety', 'car', 'hit'), 4),
 (('car', 'hit', 'track'), 4),
 (('hit', 'track', 'first'), 4)]
```
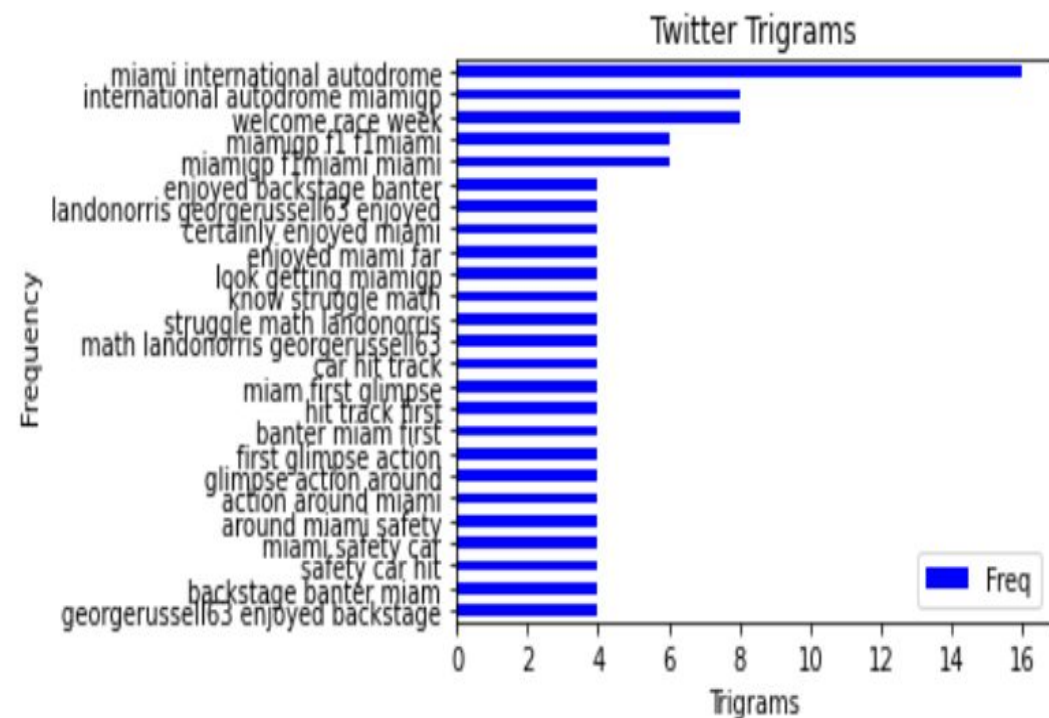
Yake model

```
[('weekend miamigp driver', 1.7652876687240923e-05),
 ('ready rock miami', 1.752315582783526e-05),
 ('week miami excited', 1.723647037674567e-05),
 ('guy miamigp week', 1.7204419975468976e-05),
 ('miami guide weekend', 1.671983583570523e-05),
 ('allnew miami international', 1.6441886260511323e-05),
 ('heat miami ready', 1.5608305630228093e-05),
 ('hitting miami race', 1.5489848764569187e-05),
 ('miami opening party', 1.5198383041207984e-05),
 ('special weekend miami', 1.4966025041833326e-05),
 ('ahead miami weekend', 1.4966025041833326e-05),
 ('race week miamigp', 1.0662738035226531e-05),
 ('miami beauty miamigp', 9.686626898299729e-06),
 ('miami race week', 9.353921173593649e-06),
 ('race week miami', 9.353921173593647e-06),
 ('miami track miamigp', 8.482665461244785e-06),
 ('miami guy miamigp', 8.482665461244785e-06),
 ('international autodrome miamigp', 7.3537679434664755e-06),
 ('miami weekend miamigp', 6.395007378855243e-06),
 ('miami international autodrome', 3.2254300489517883e-06)]
```
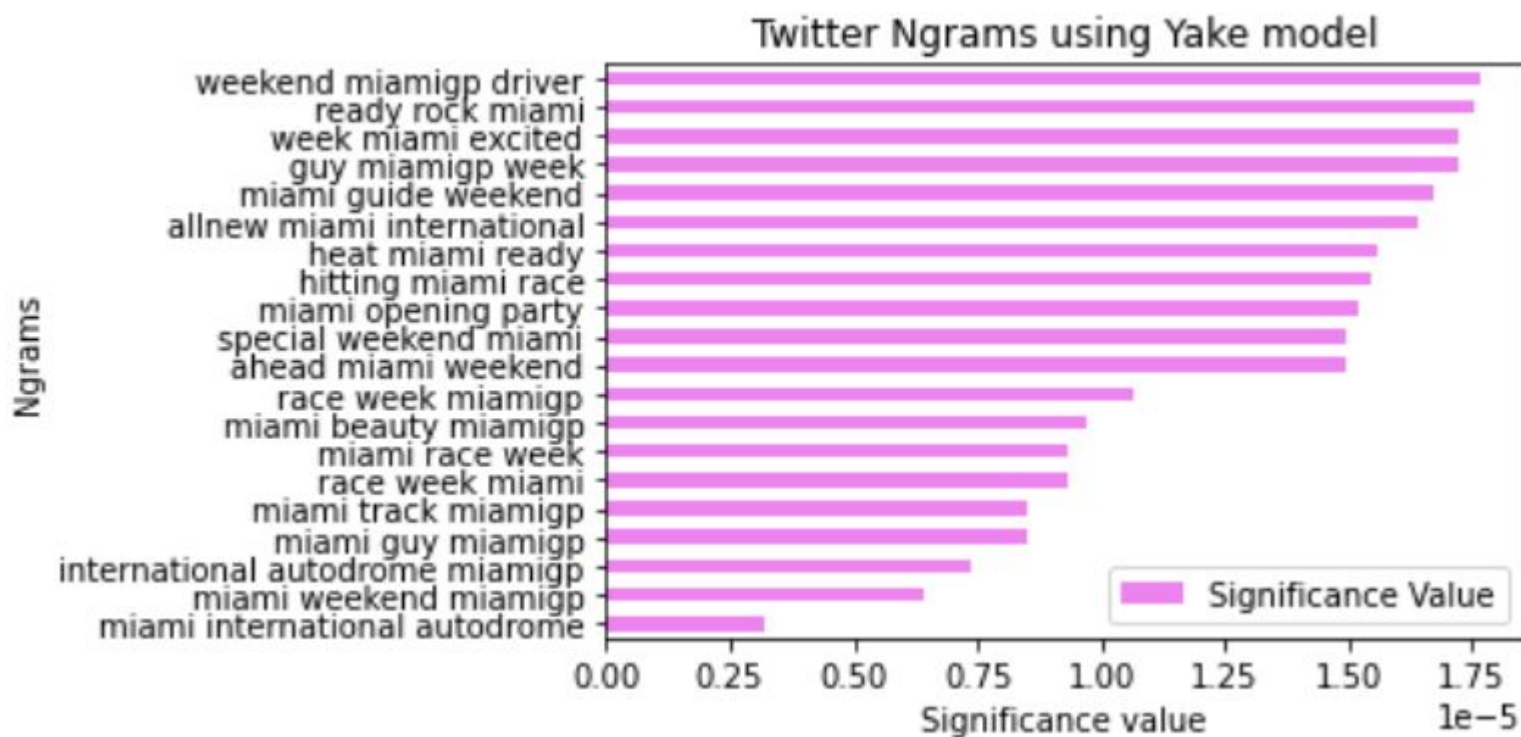
N-grams visualization for twitter

n=2
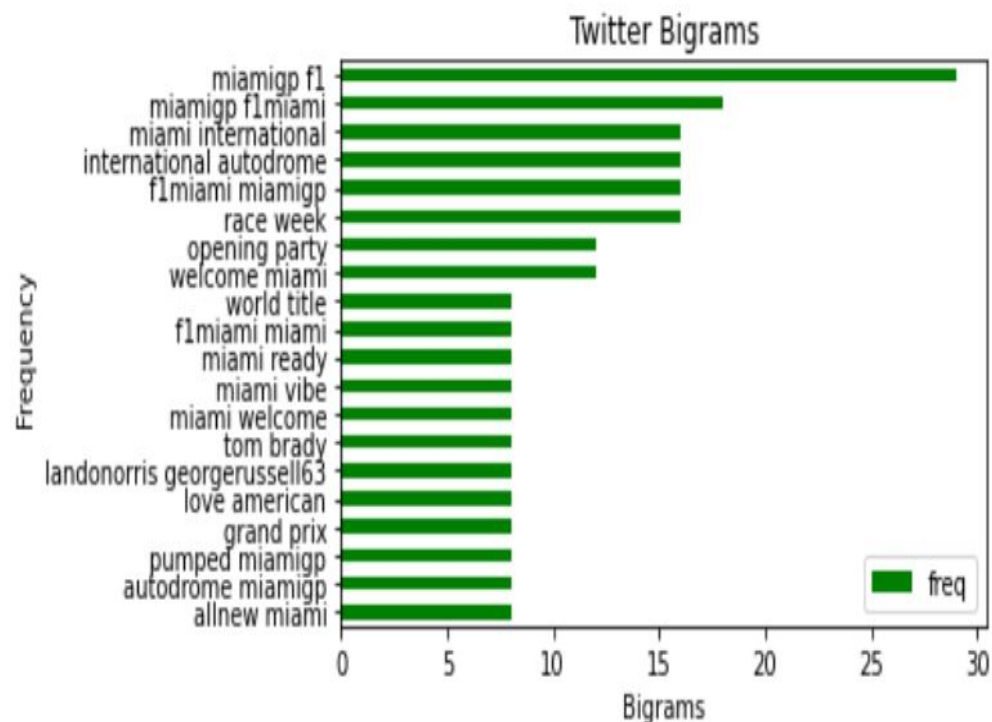
n=3

TWITTER N-GRAMS USING YAKE MODEL

N-grams for Reddit Formula1 data

n=2

```
[(('helmet', 'design'), 16),
 (('red', 'bull'), 14),
 (('grand', 'prix'), 13),
 (('original', 're'), 13),
 (('aston', 'martin'), 13),
 (('let', 'u'), 11),
 (('anyone', 'know'), 10),
 (('last', 'year'), 9),
 (('ever', 'seen'), 9),
 (('indian', 'flag'), 9),
 (('de', 'vries'), 9),
 (('last', 'season'), 8),
 (('really', 'cool'), 8),
 (('miami', 'vice'), 8),
 (('f1', 'fan'), 7),
 (('palm', 'tree'), 7),
 (('spec', 'series'), 6),
 (('social', 'medium'), 6),
 (('formula', 'one'), 6),
 (('missed', 'apex'), 6)]
```

n=3

```
[(('helmet', 'design', 'ever'), 4),
 (('direct', 'link', 'video'), 3),
 (('link', 'video', 'hq'), 3),
 (('original', 're', 'photo'), 3),
 (('mick', 'schumachers', 'future'), 3),
 (('side', 'pod', 'merc'), 3),
 (('design', 'ever', 'seen'), 3),
 (('helmet', 'ever', 'seen'), 3),
 (('happy', 'collaboration', 'mercedes'), 3),
 (('rule', 'driver', 'allowed'), 2)]
```
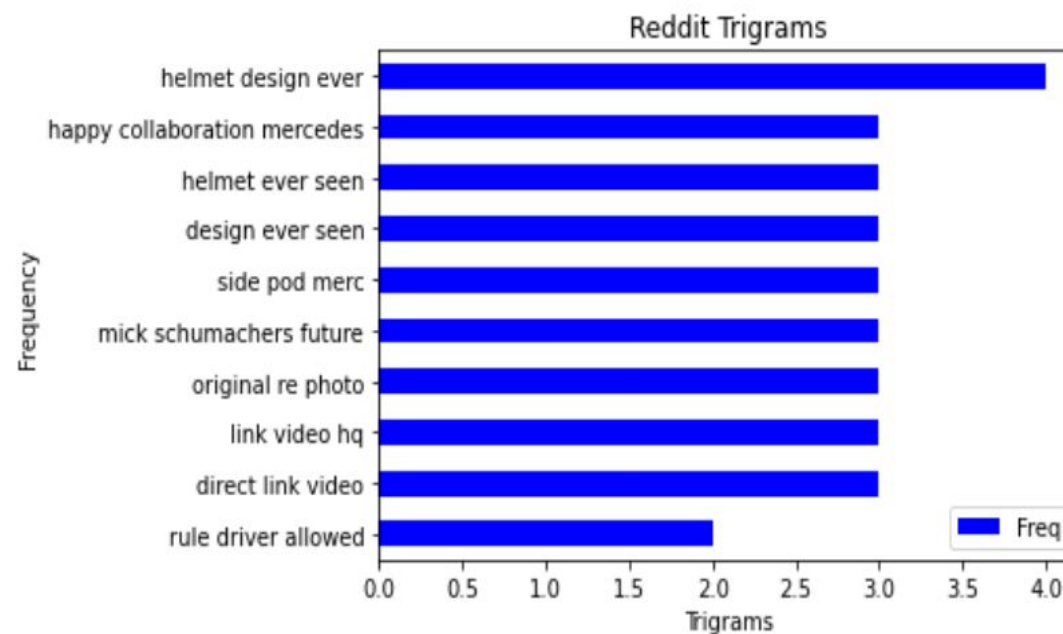
Yake model

```
[('miami helmet basketball', 4.001184070266171e-05),
 ('thought indian flag', 4.00034994764919e-05),
 ('world formula full', 3.845276221961887e-05),
 ('happy collaboration mercedes', 3.708233974517143e-05),
 ('kind cool helmet', 3.601610226614899e-05),
 ('helmet love lando', 3.575006127902042e-05),
 ('love lando helmet', 3.575006127902042e-05),
 ('racing helmet design', 3.552787912700982e-05),
 ('mick schumachers future', 3.439455612963848e-05),
 ('great helmet design', 3.384854790991647e-05),
 ('helmet helmet weekend', 3.298092871500331e-05),
 ('cool helmet good', 3.27524189829258e-05),
 ('miami grand prix', 3.274359252313914e-05),
 ('love love helmet', 3.264379941440926e-05),
 ('helmet design time', 3.010195436966762e-05),
 ('helmet miami helmet', 2.9261347563859698e-05),
 ('love helmet helmet', 2.8198184782229723e-05),
 ('miami flag bottas', 2.7318308165673354e-05),
 ('kind cool kind', 2.364253267779779e-05),
 ('helmet design miami', 1.5829891285085457e-05)]
```

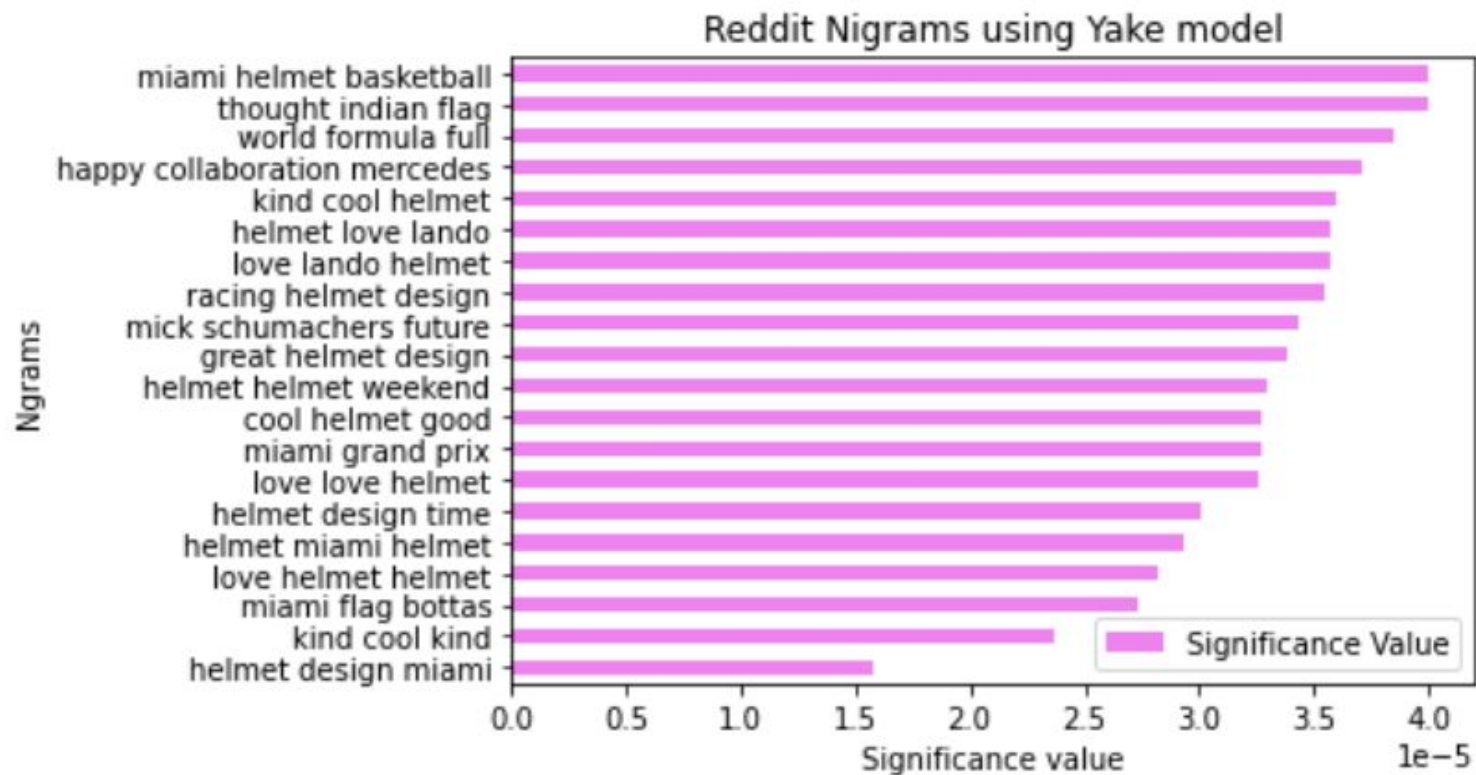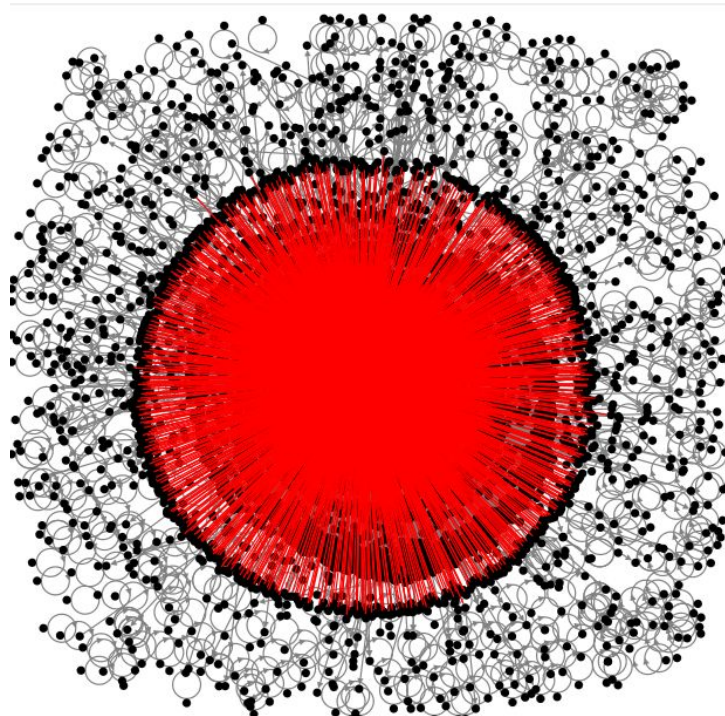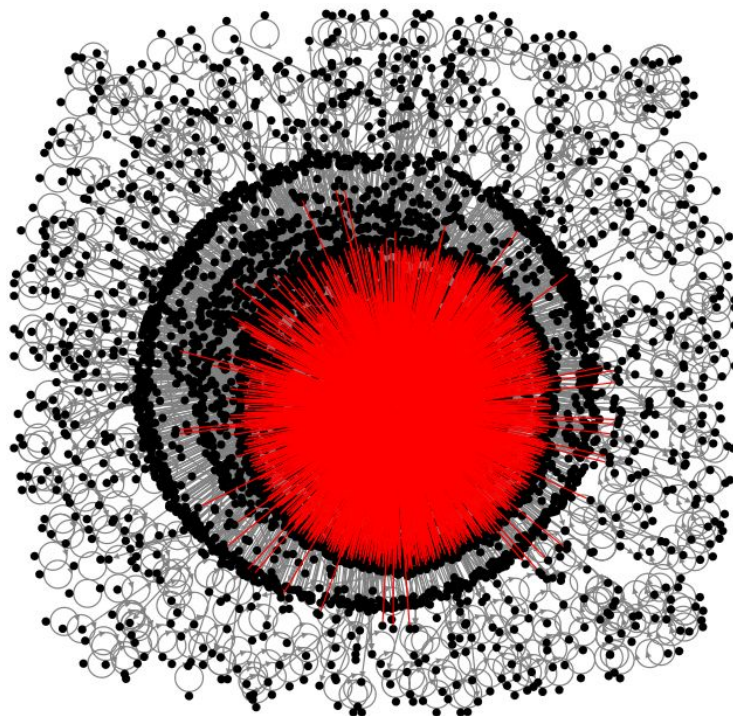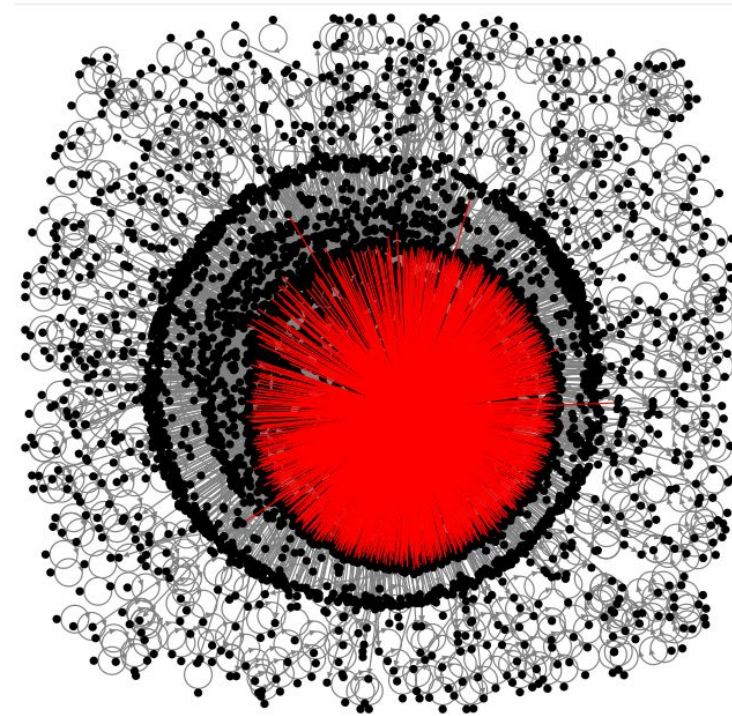# N-GRAMS VISUALIZATION FOR REDDIT

n=2
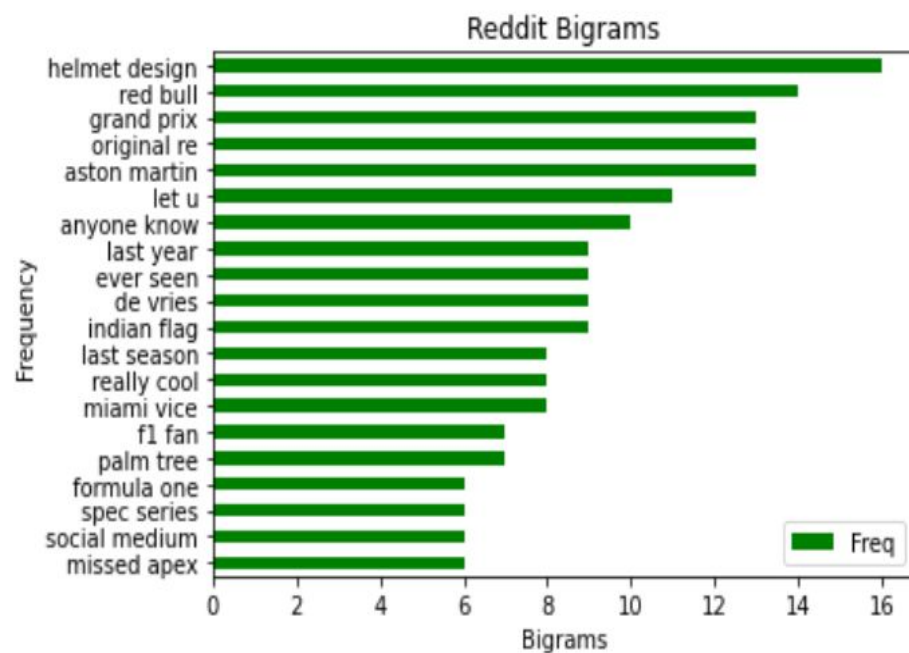
n=3

REDDIT N-GRAMS USING YAKE MODEL

Top 3 Influential Nodes

@AstonMartinF1

@Hulkhulkenberg

@lance_stroll

Aston Martin Network Metrics:

| Nodes/Accounts	In-Degree	Betweenness Centrality	Closeness Centrality
@AstonMartinF1	3708	18892148.53	0.666
@Hulkhulkenberg	2504	6962487.64	0.550
@lance_stroll	1725	4106544.56	0.499

Aston Martin F1

Twitter word-cloud:
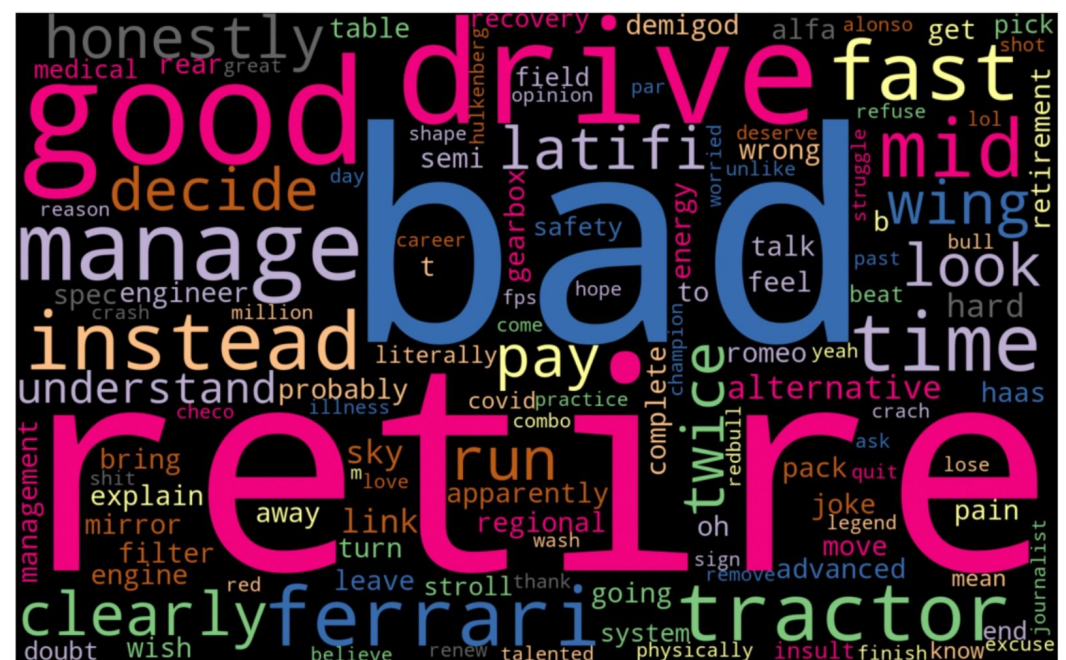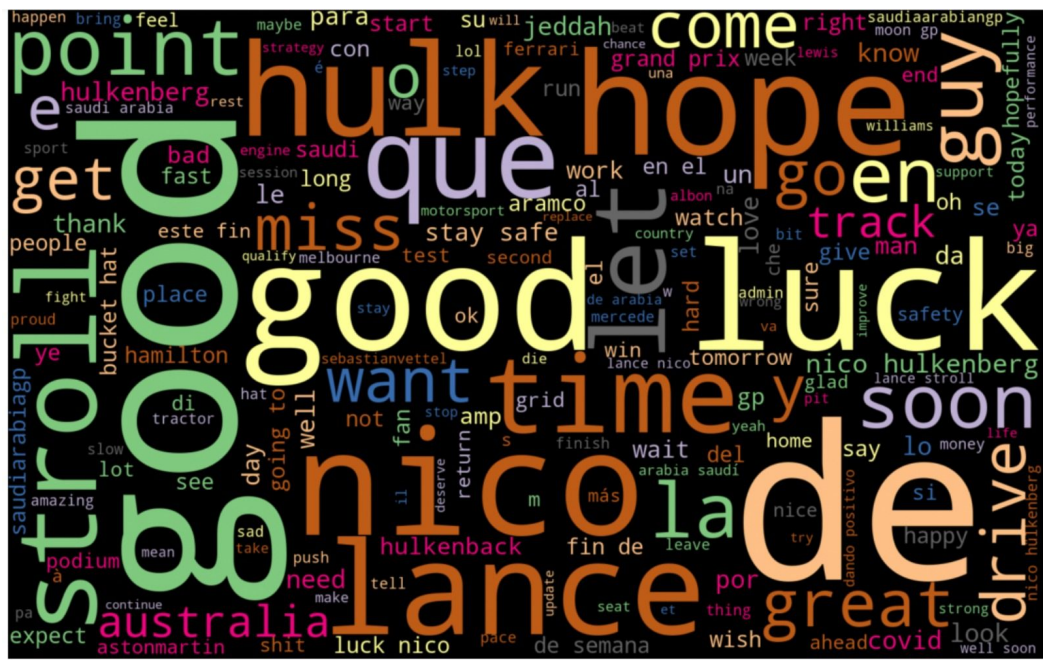
Reddit word-cloud:

Aston Martin word-cloud inference

- Users on Twitter express positive sentiment/follow the sentiment as per the original post

- Reddit users voice their unfiltered opinions and we can trace the apparent disappointment.
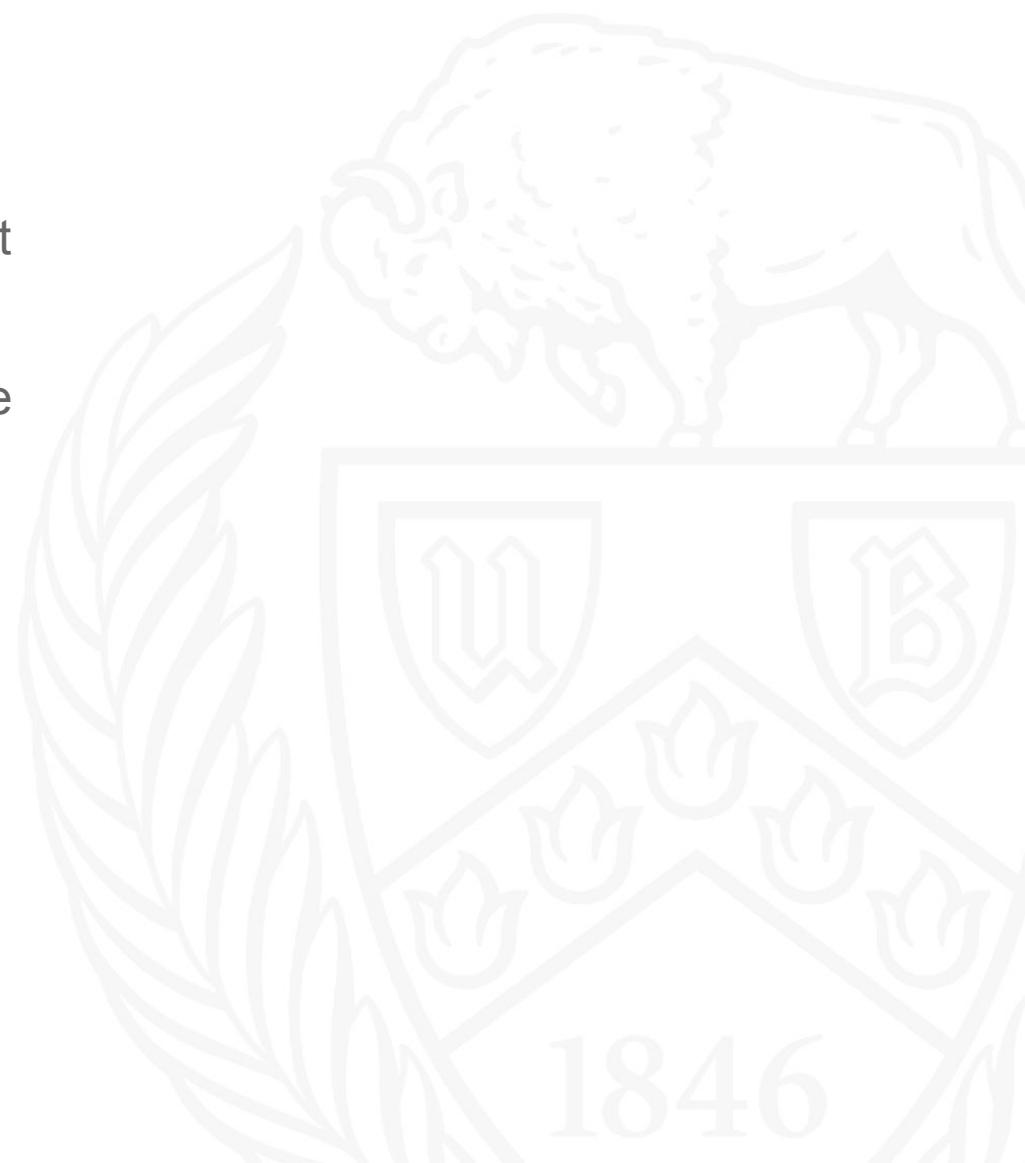
Conclusion:

- Each platform attracts different demographic based on privacy features offered to its users

- Formula 1 as an organization can use both platforms for specific surveying and commercialization purposes

- Twitter could be used by participating brands to improve brand visibility

- Reddit could be used as a platform to conduct surveys to gauge true unbiased opinions of individuals brought in by the anonymity feature.