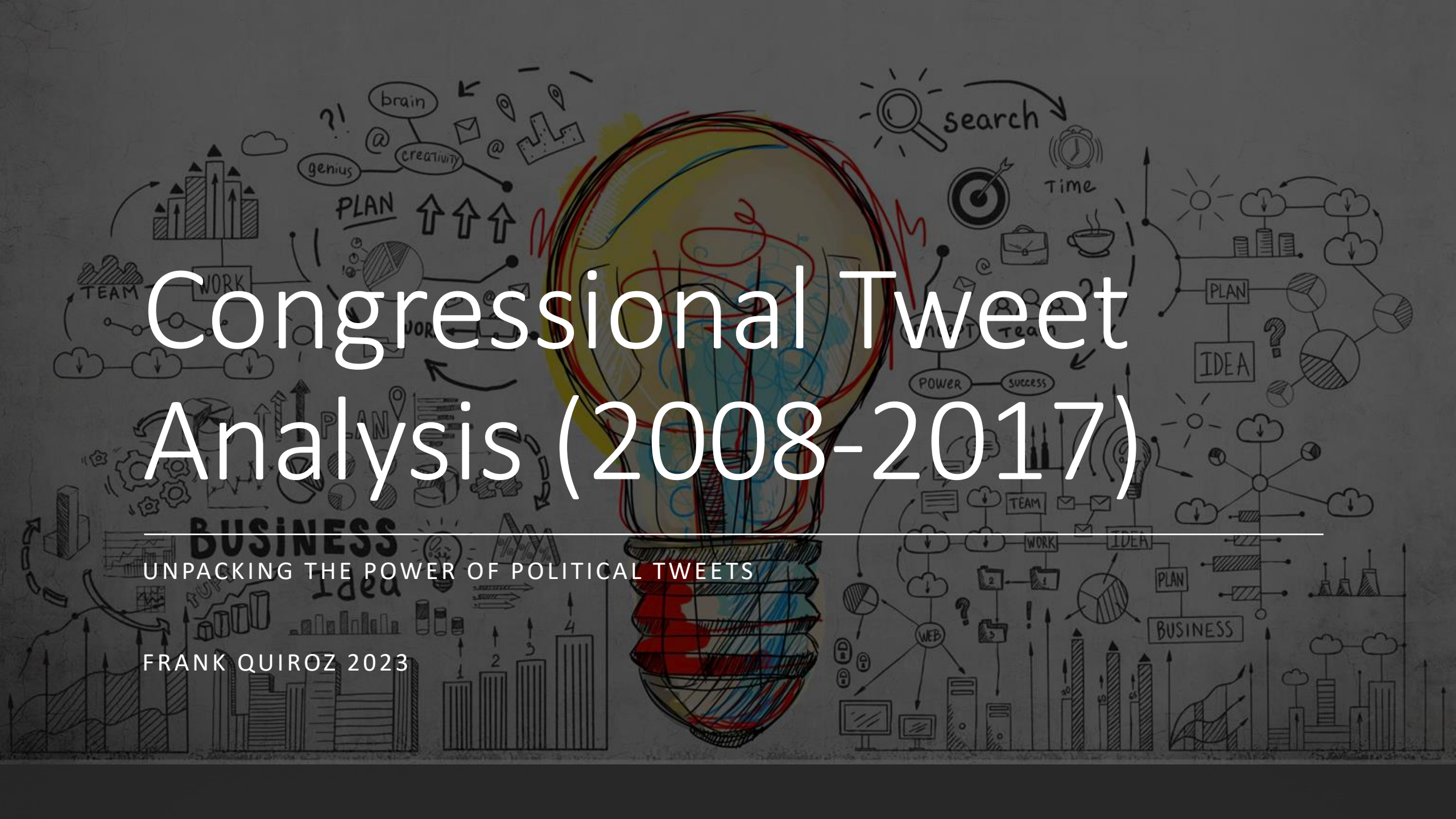


Congressional Tweet Analysis (2008-2017)

UNPACKING THE POWER OF POLITICAL TWEETS

FRANK QUIROZ 2023

Contents

Questions to Answer/Hypothesis/Approach

Technical Challenges

Entity Relationship Discuss Technical Challenges (ERD)

Initial Findings

Deeper Analysis

Hypothesis Results

Questions to Answer

At what specific hours of the day do tweets receive the most likes (favorites)?

At what specific hours of the day do tweets receive the most retweets?

Is there a correlation between the patterns of likes and retweets?

Key Hypothesis and Assumptions

Tweets posted during typical working hours receive more engagement (likes and retweets).

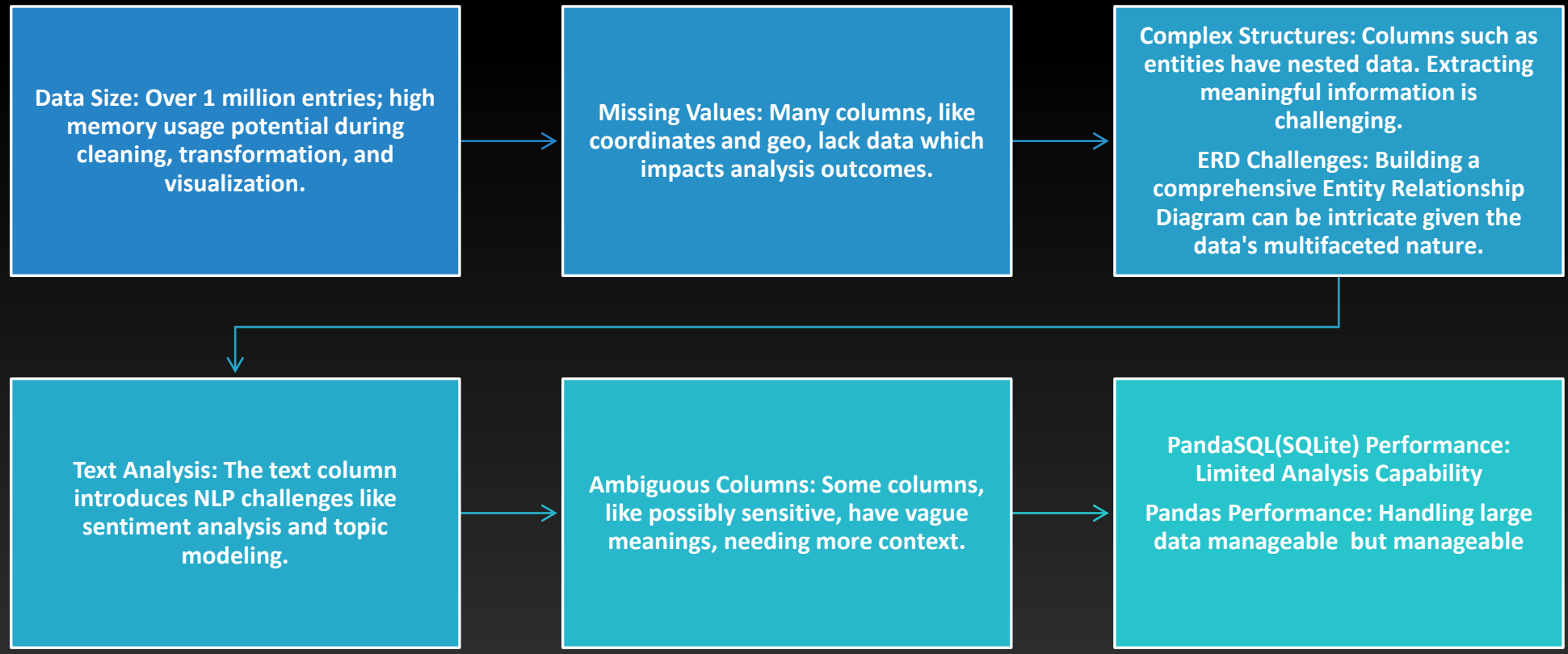
Engagement may vary with day, month, and year.

Engagement is positively correlated with posting hour.

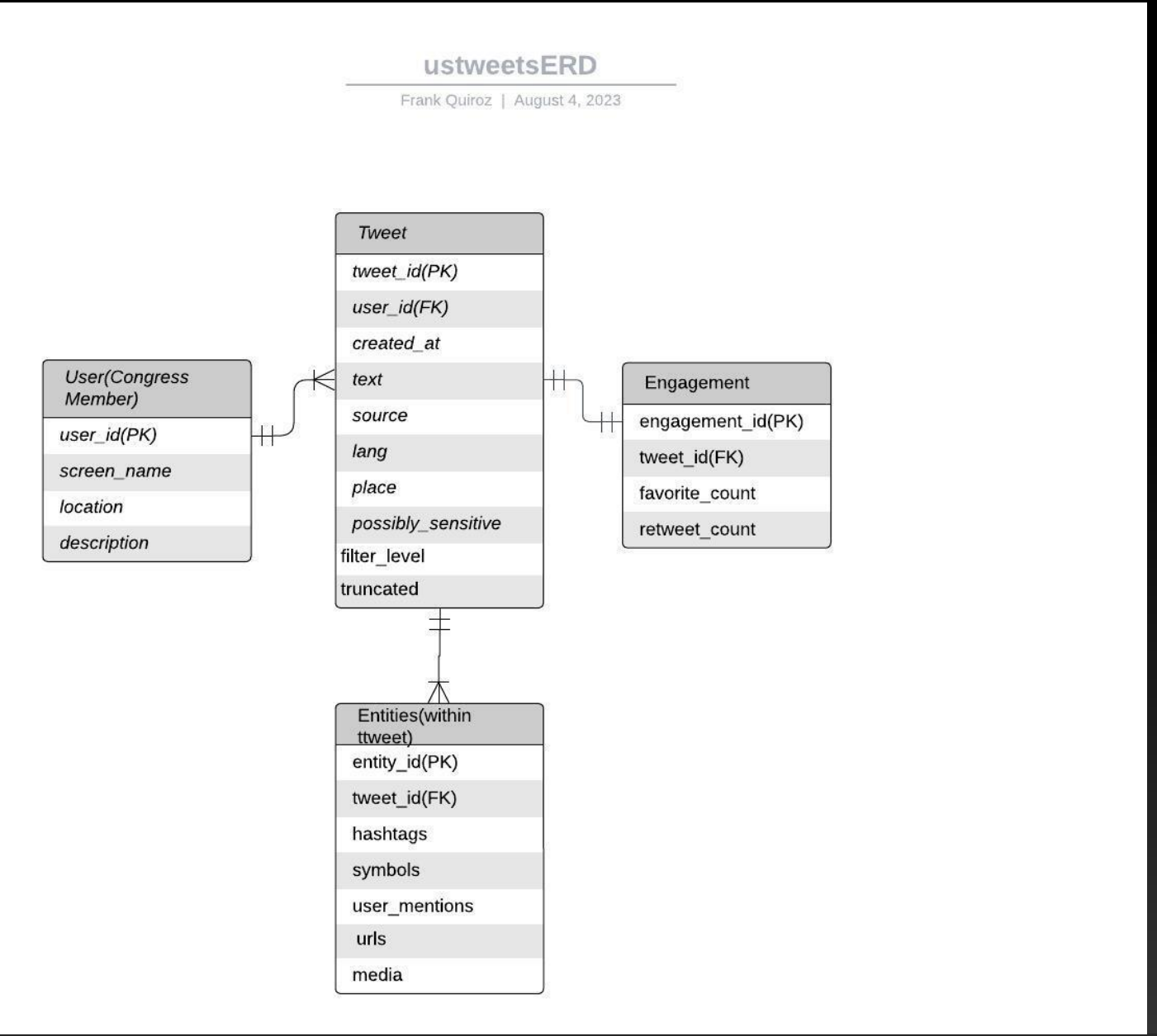
Approach:

Analyze	Analyze 'hour_created', 'favorite_count', and 'retweet_count'.
Use	Use Pearson correlation to measure the relationship between posting hour and engagement.
Use	Use t-tests to determine if there's a significant difference in engagement during working vs. non-working hours.

Technical Challenges



Entity Relationship Diagram (ERD)



Initial Findings

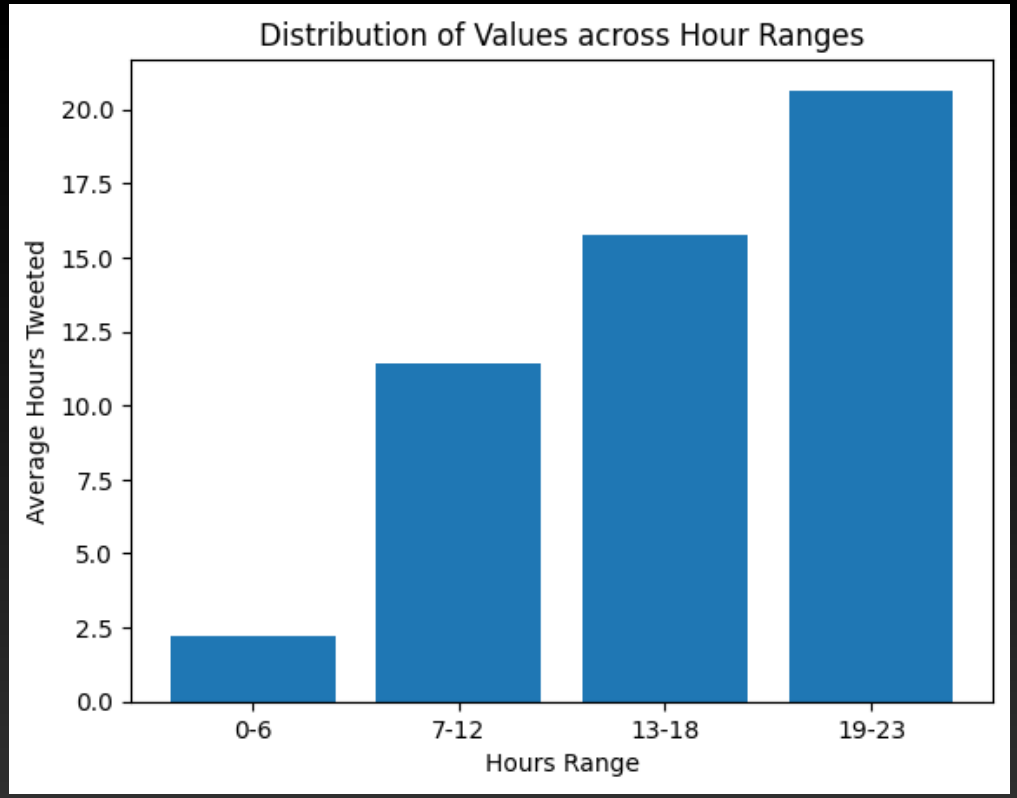
favori...	id	retweet_co...	retwee...	screen_name	trunca...	created_at_timesta...	hour_crea...
false	877418565	0	false	JohnBoozman	false	2008-08-04 17:28:51	17
false	879618172	0	false	JohnBoozman	false	2008-08-06 19:04:45	19
false	879695803						
false	880393665						
false	880474266						
false	880676101						
false	891075719						
false	893175219						
false	894551506						
false	901977122						
false	907605767						
false	909761002						
false	915252924						
false	915393190						
false	915463644						
false	917717559						

#	screen_name	favorite_count	text
1	SenSanders	984832	President Trump, you made a big mistake. By trying to divide us up by race, religion, gender and nationality you actually brought us cl...
2	realDonaldTrump	627475	Such a beautiful and important evening! The forgotten man and woman will never be forgotten again. We will all come together as n...
3	realDonaldTrump	569283	
4	SenSanders	527319	
5	realDonaldTrump	392346	
6	realDonaldTrump	347262	
7	SenatorMenendez	315624	
8	realDonaldTrump	291081	
9	realDonaldTrump	289727	
10	realDonaldTrump	269437	

screen_name	retweet_co...	text
SenSanders	461733	President Trump, you made a big mistake. By trying to divide us up by race, religion, gender and nationality you actually brought us cl...
realDonaldTrump	217557	Such a beautiful and important evening! The forgotten man and woman will never be forgotten again. We will all come together as n...
realDonaldTrump	340294	TODAY WE MAKE AMERICA GREAT AGAIN!
SenSanders	210229	.@realDonaldTrump They did. It wasn't. https://t.co/xqt29RJPEr
realDonaldTrump	82017	Peaceful protests are a hallmark of our democracy. Even if I don't always agree, I recognize the rights of people to express their views.
realDonaldTrump	139326	Happy New Year to all, including to my many enemies and those who have fought me and lost so badly they just don't know what to ...
SenatorMenendez	146679	Hey Republicans, don't worry, that burn is covered under the Affordable Care Act
realDonaldTrump	52689	THANK YOU for another wonderful evening in Washington, D.C. TOGETHER, we will MAKE AMERICA GREAT AGAINus https://t.co/V3a...
realDonaldTrump	162780	How long did it take your staff of 823 people to think that up--and where are your 33,000 emails that you deleted? https://t.co/gECL...
realDonaldTrump	70918	It all begins today! I will see you at 11:00 A.M. for the swearing-in. THE MOVEMENT CONTINUES - THE WORK BEGINS!

Deeper Analysis (Tweet Hour Engagement)

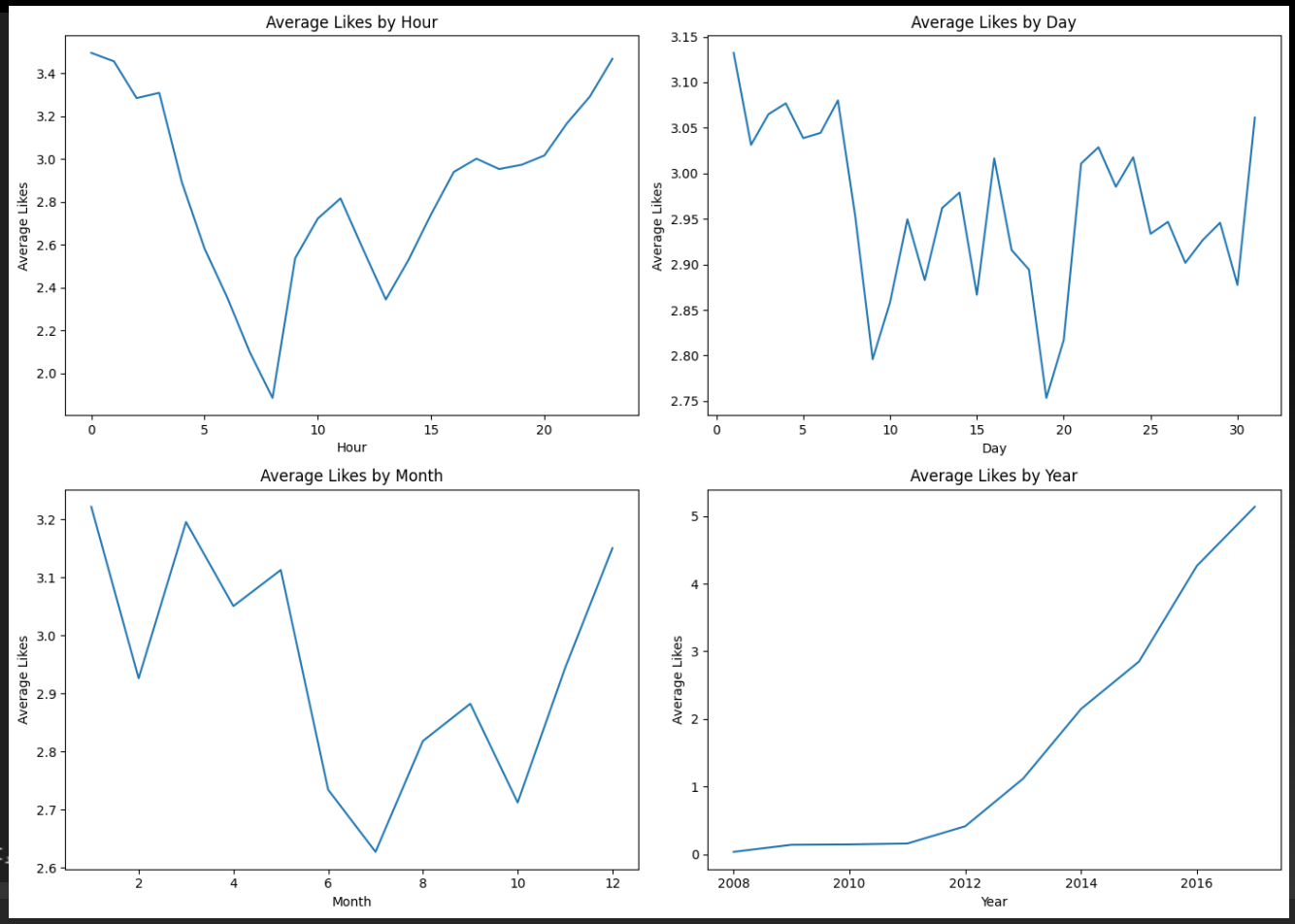
- More user engagement appeared during latter part of the day
- Does the day, month, year effect activity?



Deeper Analysis (Avg Like Engagement Time)

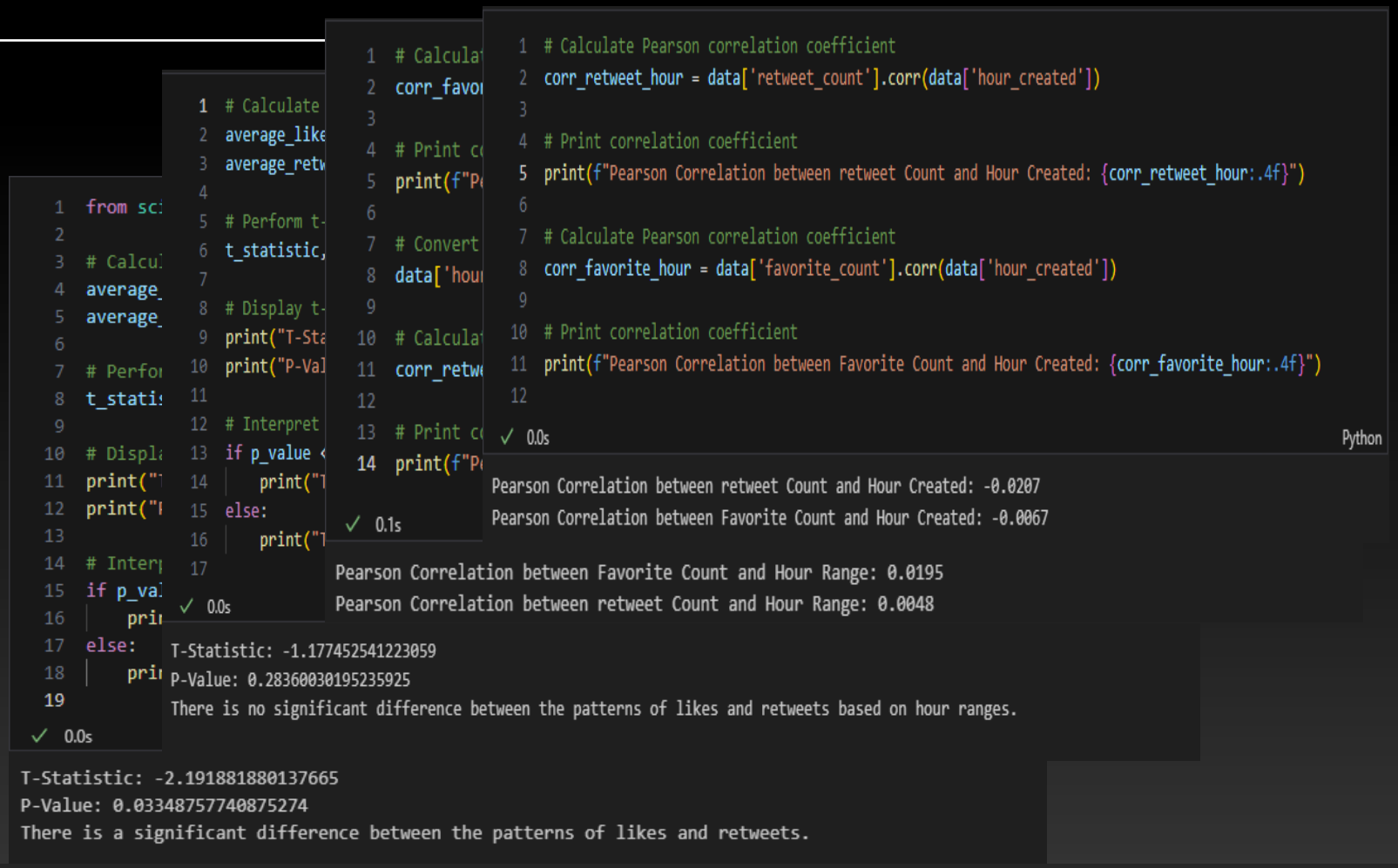
hour_created	
0	3.563484
1	3.594364
2	3.513659
3	3.566232
4	3.132356
5	2.851044
6	2.527514
7	2.195817
8	1.983553
9	2.551205
10	2.896982
11	2.888753
12	2.611377
13	2.446600
14	2.524395
15	2.690133
16	2.873003
17	2.930794
18	2.894561
19	2.908750
20	2.962888
21	3.097473
22	3.246281
23	3.482546

Name: favorite_count



Deeper Analysis (T-Test)

- **Engagement Patterns:**
  Significant difference between likes and retweets
- **Hour Ranges:**
  No significant difference in engagement
- **Favorite Count and Retweet Count:**
  Moderately positive correlation (0.374) between favorite and retweet counts.
- **Hour Range Correlations:**
  Very weak positive correlations with engagement metrics (0.0048, 0.0195).
- **Hour Created Correlations:**
  Very weak negative correlations with engagement metrics (-0.0207, -0.0067).



Deeper Analysis (T-Test)

- **Engagement Variation:** Difference in engagement patterns between likes and retweets.
- **Hourly Engagement:** While engagement doesn't significantly differ across hour ranges, observed specific time-based patterns.
- **Correlation Confirmation:** Favorite and retweet counts validation that engagement metrics often align.
- **Time-Based Patterns:** Confirmed existence of subtle time-based patterns in the data.
- **Posting Hour Influence:** Posting hour might have a minor impact on engagement.
- **Strategy Enhancement:** Supported social media strategy. Targeting specific hours and focusing on maximizing both likes and retweets, we can optimize engagement efforts.

```python
# Define working hours (9 AM to 5 PM)
working_hours = data[(data['hour_created'] >= 9) & (data['hour_created'] <= 17)]
non_working_hours = data[(data['hour_created'] < 9) | (data['hour_created'] > 17)]

# Perform A/B test
alpha = 0.05  # Significance level

retweet_working = working_hours['retweet_count']
retweet_non_working = non_working_hours['retweet_count']

t_statistic, p_value = stats.ttest_ind(likes_working, likes_non_working, equal_var=False)

print(f"T-Statistic: {t_statistic}")
print(f"P-Value: {p_value}")

if p_value < alpha:
    print("Reject the null hypothesis: There is a significant difference in average retweets.")
else:
    print("Fail to reject the null hypothesis: There is no significant difference in average retweets."
```

✓ 0.6s Python

```
T-Statistic: -43.81264790108829
P-Value: 0.0
Reject the null hypothesis: There is a significant difference in average retweets.
```

✓ 18.3s Python

```
T-Statistic: -43.81264790108829
P-Value: 0.0
Reject the null hypothesis: There is a significant difference in average likes.
```

Deeper Analysis (Avg Like & Retweet by User)

- T-Statistic: 5.773323764661414
- P-Value: 1.7987310057174 79e-05
- Reject the null hypothesis: There is a significant difference in engagement metrics.

```
1  print(avg_retweets_by_user)
```

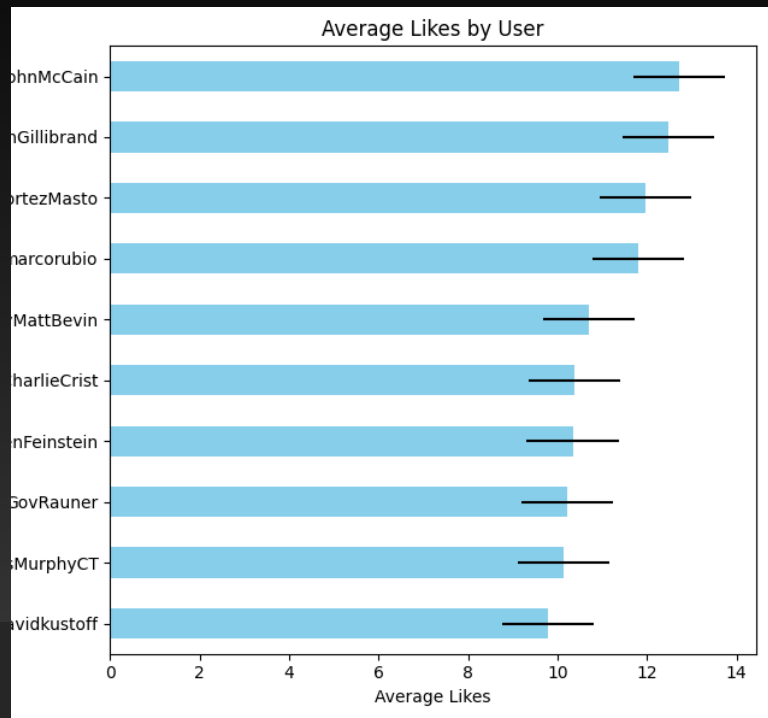
✓ 0.0s

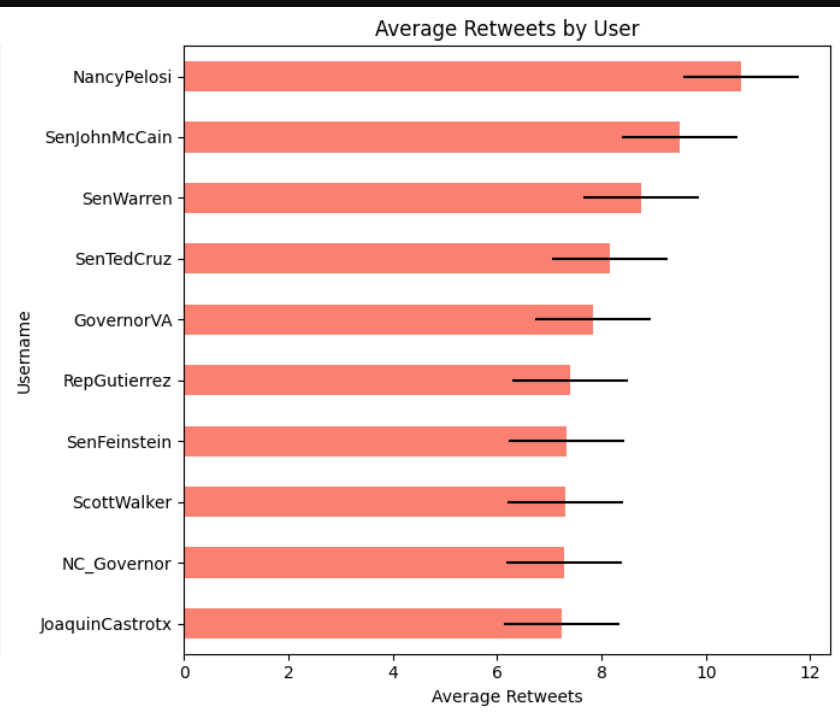
```
screen_name
NancyPelosi        10.684211
SenJohnMcCain       9.508091
SenWarren           8.750000
SenTedCruz          8.166667
GovernorVA          7.845562
RepGutierrez        7.407729
SenFeinstein        7.324017
ScottWalker         7.297980
NC_Governor         7.272727
JoaquinCastrotx     7.245283
Name: retweet_count, dtype: float64
```

```
1  print(avg_likes_by_user )
2
```

✓ 0.0s

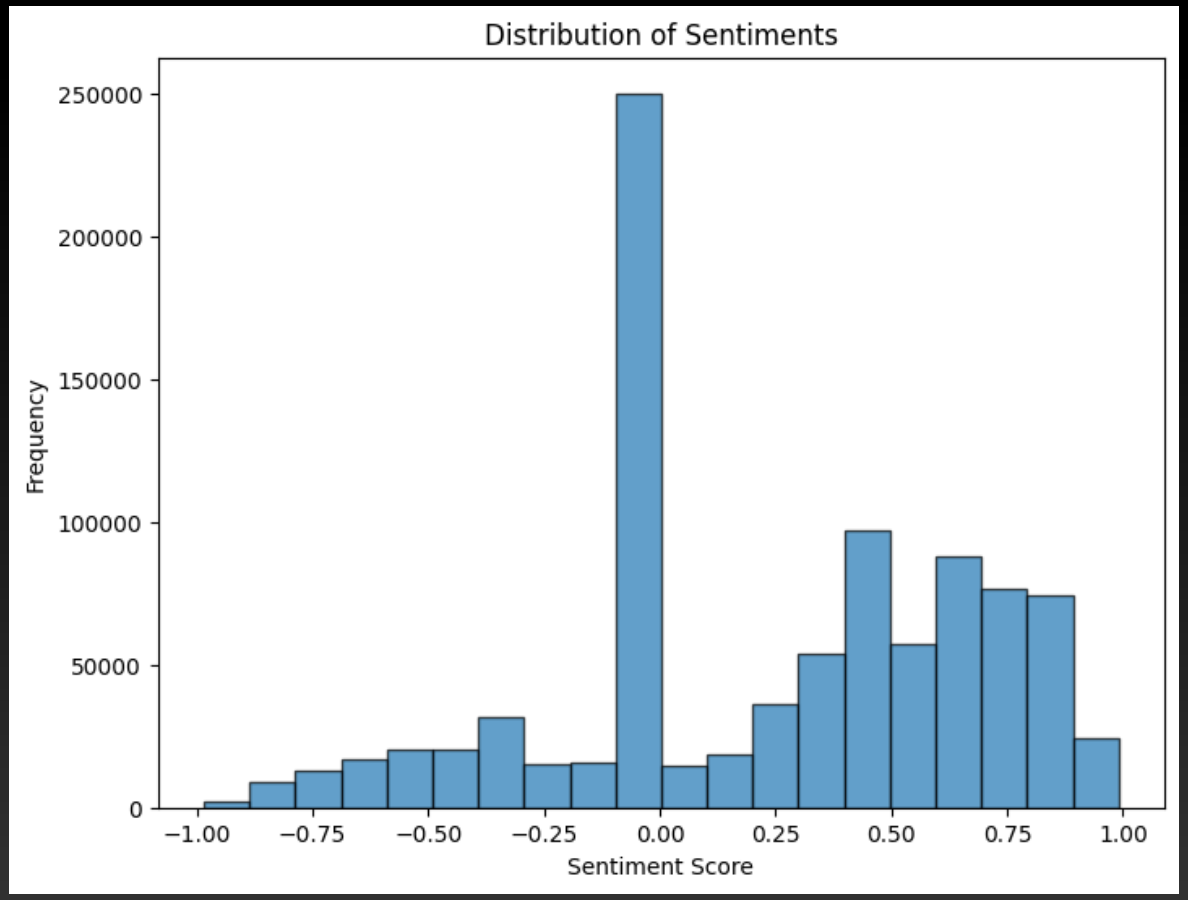
```
screen_name
SenJohnMcCain      12.728155
SenGillibrand      12.478903
SenCortezMasto     11.971014
marcorubio         11.812121
GovMattBevin       10.689655
RepCharlieCrist    10.381443
SenFeinstein       10.337474
GovRauner          10.214178
ChrisMurphyCT      10.133588
repdavidkustoff     9.790541
Name: favorite_count, dtype: float64
```



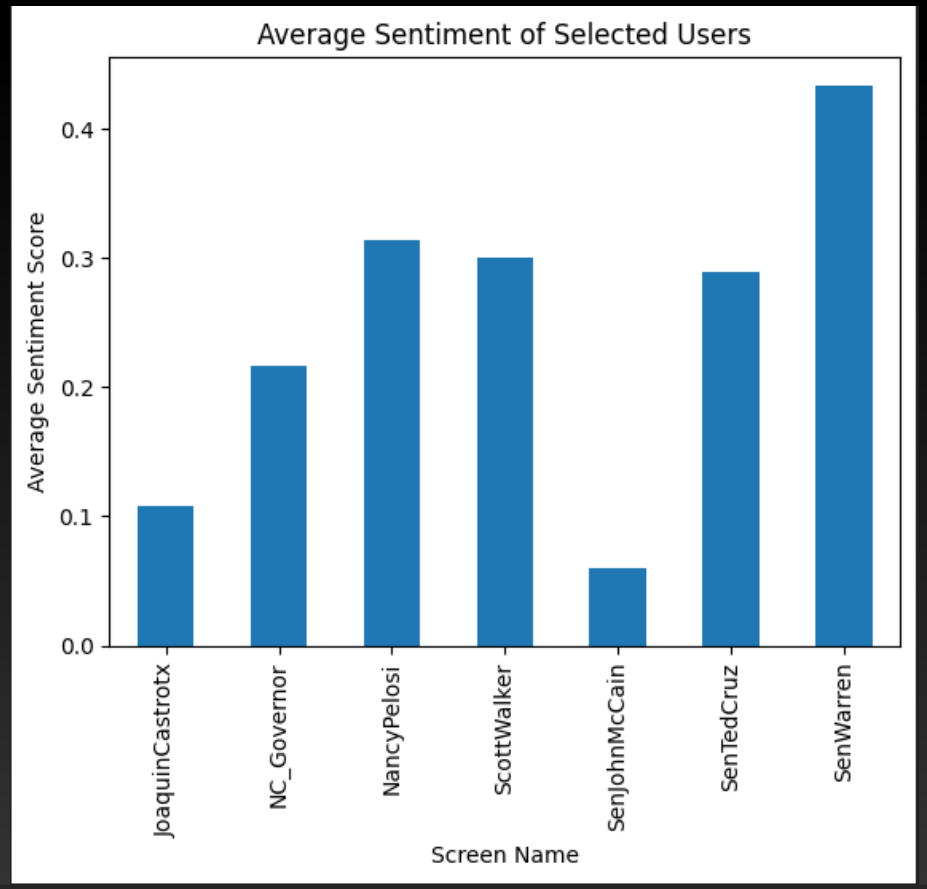


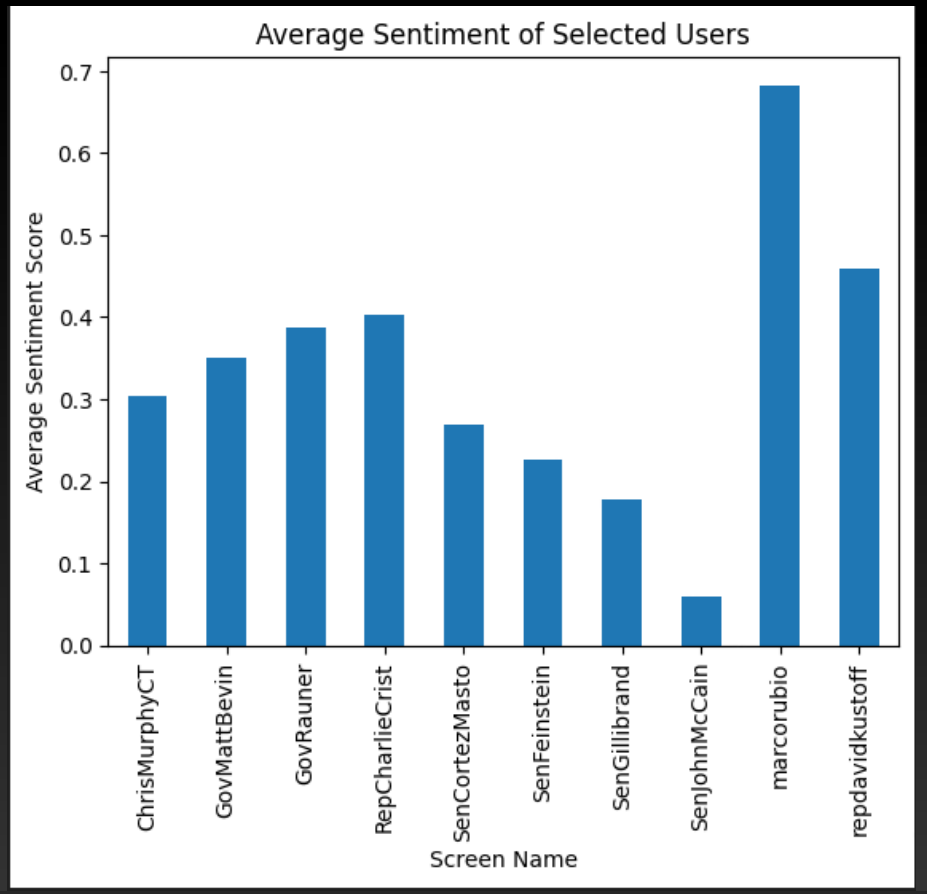
Deeper Analysis
(Avg Sentiment of Users)

- Exploring sentiment change might be a significant difference in engagement between likes and retweets

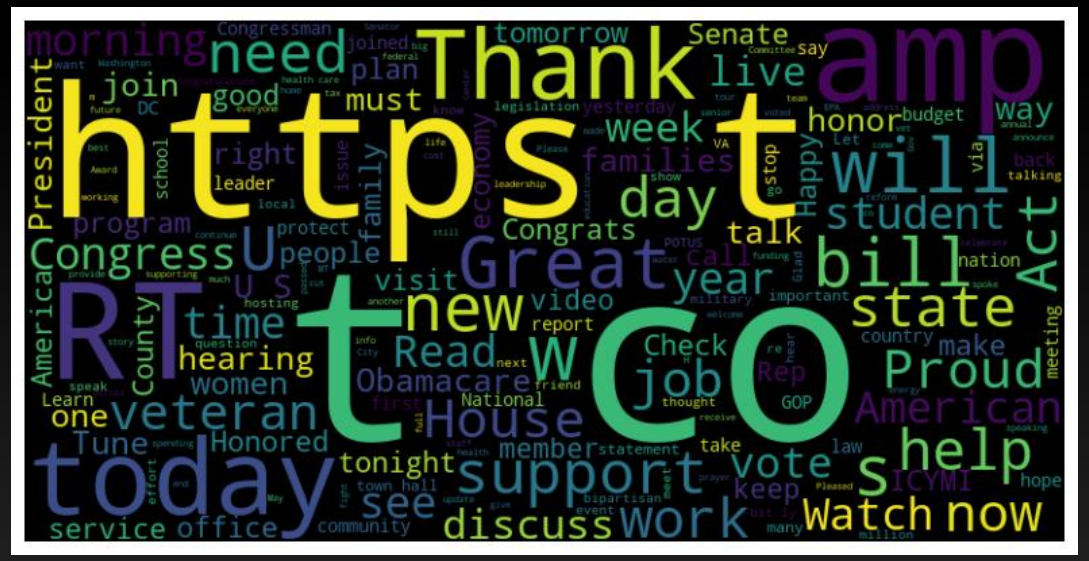


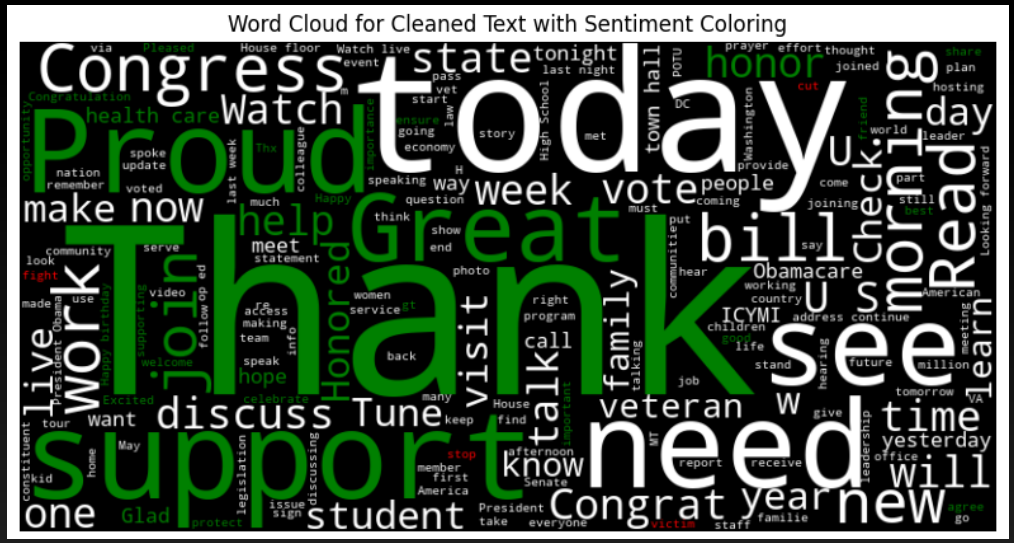
Deeper Analysis (Average Sentiment Most Liked/Retweeted Users)





Deeper Analysis (Sentiment Analysis)





Caveat to be explored later: Some terms may carry highly positive or negative sentiments; however, these terms may have either been an entity or a different connotation. See example above with URLs.

Final Findings (Hypothesis Results)

Tweets posted during typical working hours receive more engagement (likes and retweets): Partially true.

I. **Our A/B testing demonstrated that engagement patterns between working and non-working hours are indeed distinct.**
II. **When considering sentiment analysis, we observed that tweets posted during working hours tend to have a more positive sentiment, which could contribute to higher engagement during those times.**

Engagement may vary with day, month, and year: Confirmed.

I. **Sentiment analysis revealed that engagement levels can also be influenced by the emotional tone of the tweets.**
II. **Certain days or months with higher positive sentiment saw increased engagement.**

Engagement is positively correlated with posting hour: Mostly true.

I. **Tweets posted during hours of generally positive sentiment tend to attract more engagement.**
II. **Correlation between sentiment and posting hour might not be perfect, it suggests that crafting tweets with a positive emotional tone during specific posting hours can enhance engagement.**