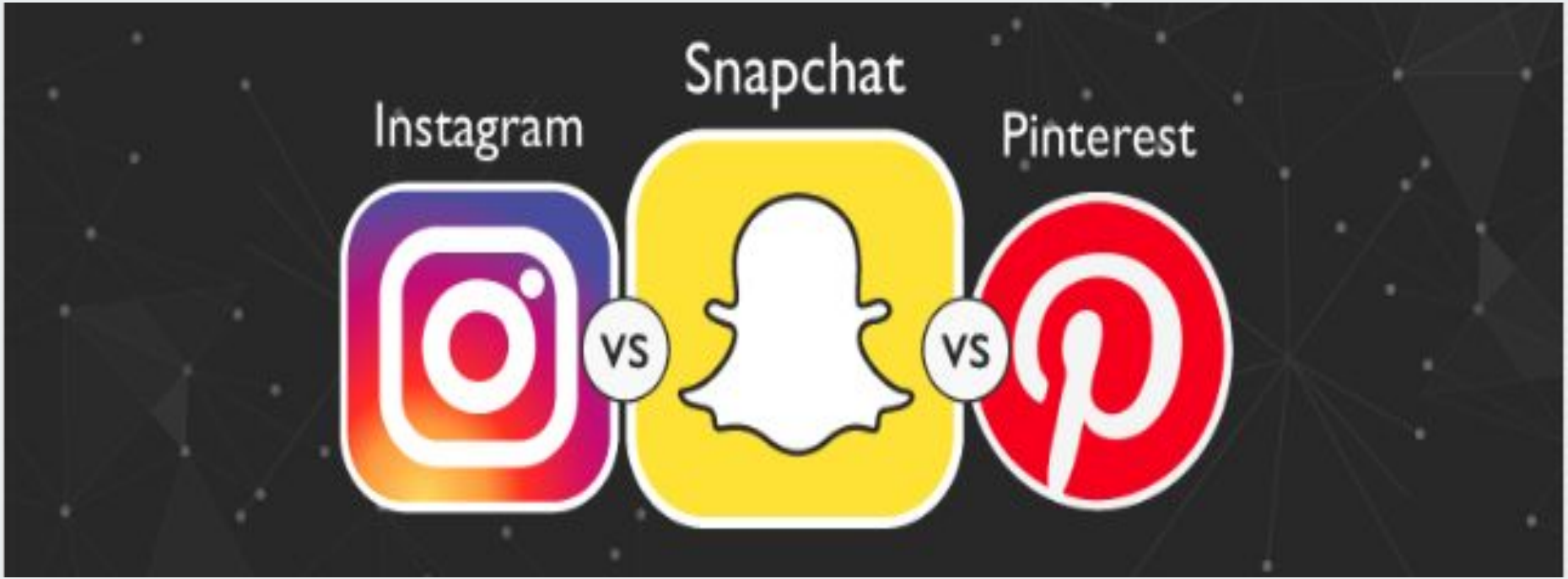


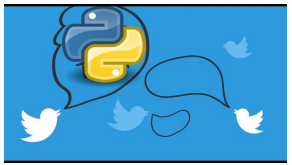
- **Popularity and Attitude Analysis Based on Tweets**

Picture source: https://www.a1future.com/blog/instagram-snapchat-pinterest-infographic/

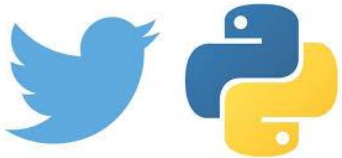
Elaine Wu

Pipeline

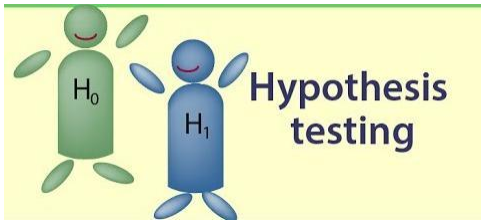
Streaming with Tweepy



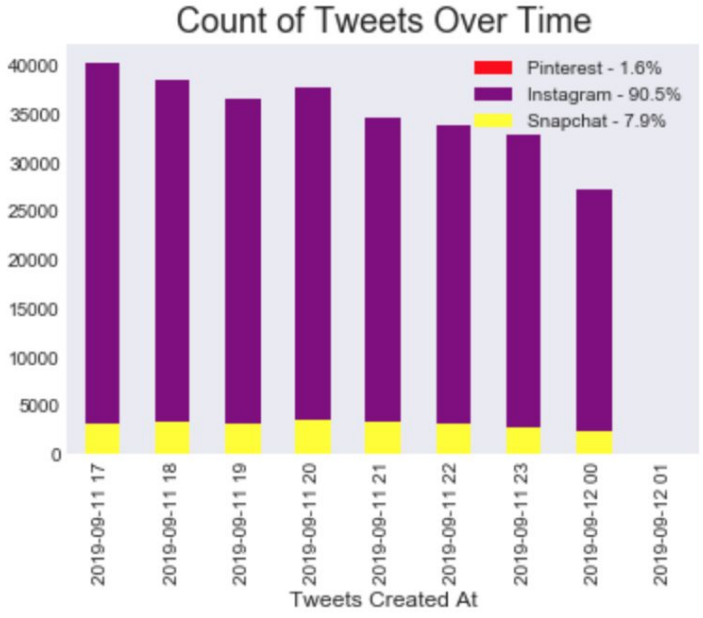
Load + Clean + Explore Data

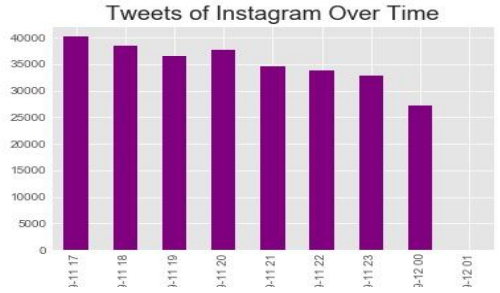


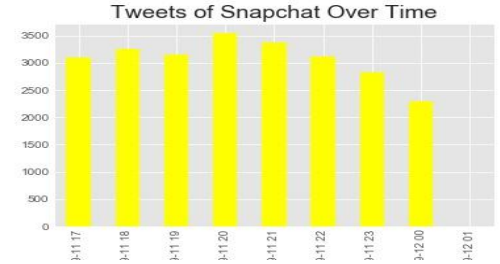
Perform Hypothesis Test

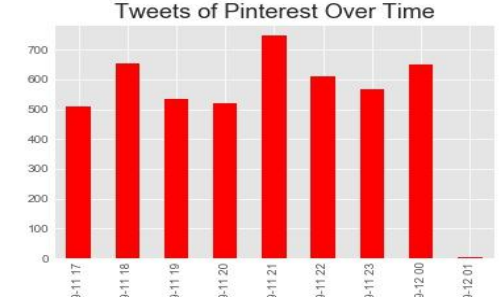


An Overview of the Data



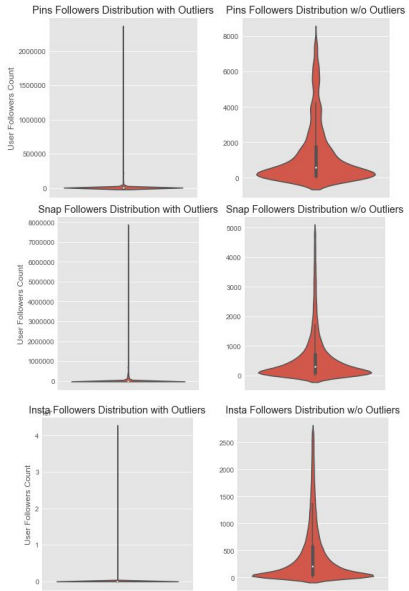






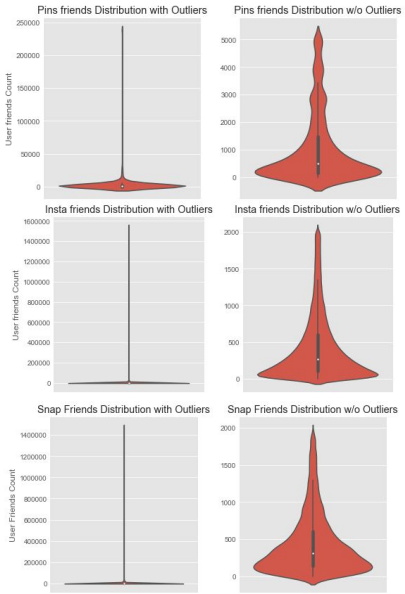
How to measure the popularity?

$$Tweet \Rightarrow User \Rightarrow Followers\ Count$$

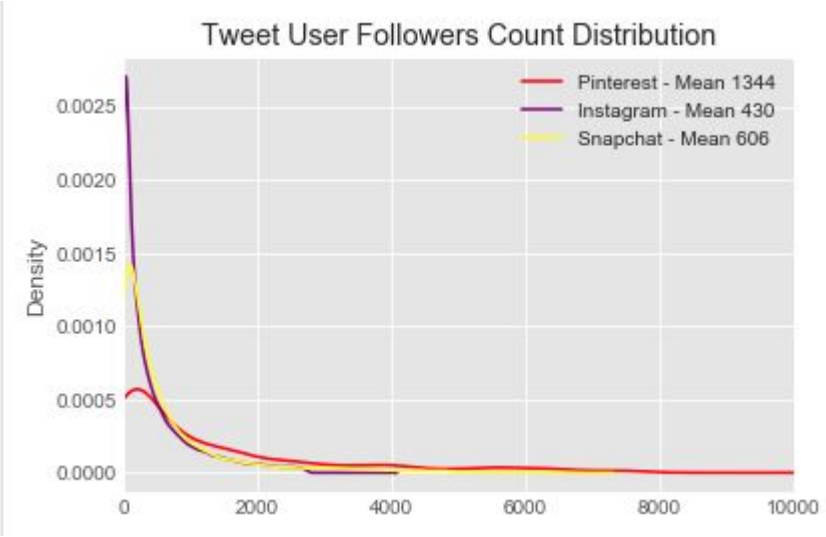


- Long tail
- Drop the outliers

$$Tweet \Rightarrow User \Rightarrow Friends\ Count$$



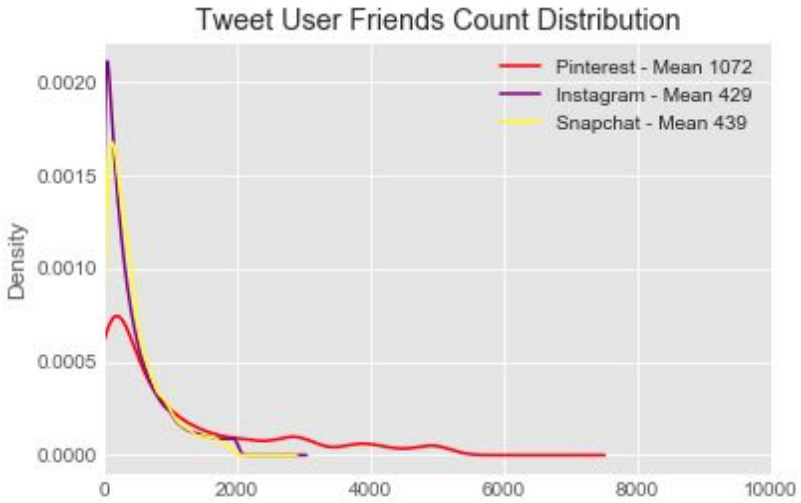
Popularity - Hypothesis Test



	Null Hypothesis*	Alternate Hypothesis	U-test P-value
Test 1	Pints_Followers = Insta_Followers	Pints_Followers > Insta_Followers	0.00000
Test 2	Pints_Followers = Snap_Followers	Pints_Followers > Snap_Followers	0.00000
Test 3	Insta_Followers = Snap_Followers	Snap_Followers > Insta_Followers	0.00000

*Pints_Followers: The mean of tweets user's followers count for pinterest ;
*Same logic for Insta_Followers and Snap_Followers
* Confidence level for the combined test = 0.05 (i.e. 0.05/3 for each test)

Popularity - Hypothesis Test



	Null Hypothesis*	Alternate Hypothesis	U-test P-value
Test 1	Pints_Friends = Insta_Friends	Pints_Friends > Insta_Friends	0.00000
Test 2	Pints_Friends = Snap_Friends	Pints_Friends > Snap_Friends	0.00000
Test 3	Insta_Friends = Snap_Friends	Snap_Friends < Insta_Friends	1.00000

*Pints_Friends: The mean of tweets user's friends count for pinterest ;
*Same logic for Insta_Friends and Snap_Friends
* Confidence level for the combined test = 0.05 (i.e. 0.05/3 for each test)

How to measure attitude?

$$Relatively\ Positive\ Ratio = \frac{\#\ of\ positive\ tweets}{\#\ of\ positive\ tweets + \#\ of\ negative\ tweets}$$

Positive Tweets: tweets with positive value > 0.5 *

text	negative	neutral	positive
I got Snapchat now yay!!!	0.0	0.412	0.588

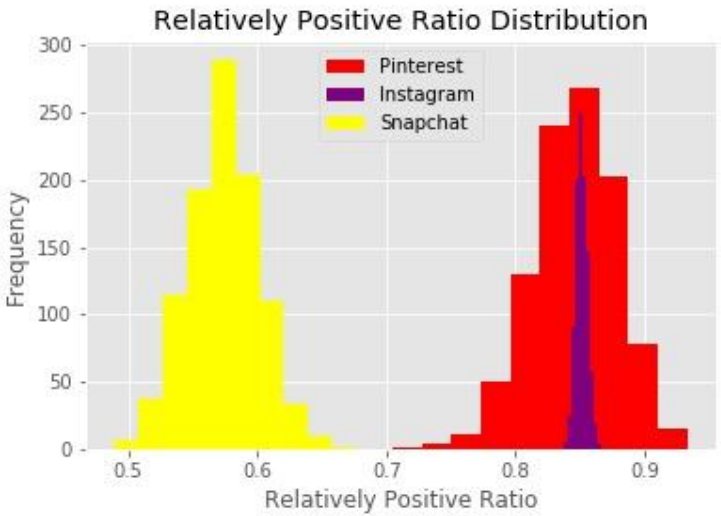
Negative Tweets: tweets with negative value > 0.5

text	negative	neutral	positive
snapchat sucks	0.714	0.286	0.0

Tweets Sentiment Summary				
	Positive Ratio	Negative Ratio	Neutral Ratio	**RPR**
Pinterest	0.0330	0.0061	0.9610	0.8443
Instagram	0.0343	0.0060	0.9597	0.8507
Snapchat	0.0239	0.0176	0.9585	0.5764

* Tool: NLTK - Source: Bird, Steven, Edward Loper and Ewan Klein (2009), Natural Language Processing with Python. O'Reilly Media Inc

Positivity - Hypothesis Test

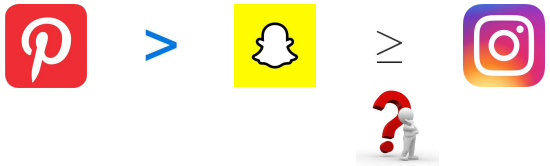


	Null Hypothesis*	Alternate Hypothesis	Welch's t-test P-value*	U-test P-value
Test 1	Pints_RPR = Insta_RPR	Insta_RPR > Pints_RPR	0.00000	0.00009
Test 2	Pints_RPR = Snap_RPR	Pints_RPR > Snap_RPR	0.00000	0.00000
Test 3	Insta_RPR = Snap_RPR	Insta_RPR > Snap_RPR	0.00000	0.00000

*Pints_RPR: The mean of tweets RPR for pinterest ;
*Insta_RPR: The mean of tweets RPR for instagram;
*Snap_RPR: The mean of tweets RPR for snapchat;
* Confidence level for the combined test = 0.05 (i.e. 0.05/3 for each test)

Conclusion

1. We used user's followers count and friends count as measurements of the platform's popularity, dropped the largest 10% outliers and performed u-test on the measurements.



2. We defined a metric (RPR) to measure the user's attitude towards each platform and performed t-test and u-test on the difference between each pair of the three platform's RPR.

Improvement and Future Work:

1. How to measure the popularity and attitude?

2. The sentiment analysis is based on score assigned by NLTK model. How accurate is it?

3. In addition to determining the positivity and negativity, can we extract more information from the text? E.g. What's the event that keeps user happy? What's the issue that makes user angry?

Thank you!

Q&A