Congressional Tweet Analysis

 $Steve^1$

¹SQL Anonymous

October 3, 2024

Table of Contents

- 1 30000 ft overview
- 2 Two AB Tests
- **3** The most popular user on Twitter?

Table of Contents

- 1 30000 ft overview
- 2 Two AB Tests
- 3 The most popular user on Twitter?

some high level statistics

There are 1243370 total tweets.

There are 545 different screen names among the data.

The earliest tweet is indexed at 2008-08-04 and the latest tweet is indexed at 2017-06-06.

Table of Contents

- 1 30000 ft overview
- 2 Two AB Tests
- 3 The most popular user on Twitter?

What makes a tweet more retweetable?

- We ran multiple AB tests on the impact of the sentiment score on a tweet getting retweeted.
- We considered the impact of scores higher than 0.1, 0.2, 0.3, ..., 0.9, and the answer was decidedly no. In each case, we failed to reject the null hypothesis that the sentiment score had a slightly negative impact on a tweet being retweeted.
- We also considered the presence of a #hashtag in the tweet as an indicator of a tweet getting retweeted. The answer was overwhelmingly yes.

Not for sentiment score of greater than 0.1.

A/B test	A ing statis	stics						
Label			Number of successes		Number of tria	als		
control			537923		623019			Remove
Variation 1			534631		620351			Remove
Interval con	fidence level							
0.95				Use multiple	testing correct	ion: 🔽		
Compute	Add another	r group						
	Successes	Total	Success Rate			p-value	Improveme	ent
control	537,923	623,019	86% - 86% (86%)			_	_	
Variation 1	534,631	620,351	86% – 86% - (86%)		1+	0.0099	-0.32% — (-0.18	

Not for sentiment score of greater than 0.2.

AB testi		stics						
Label			Number of suc	ccesses	Number of t	rials		
Baseline			669352		774812			Remove
Variation 1			403202		468558			Remove
Interval confi	idence level							
0.95				Use multiple	testing corre	ction: 🔽		
Compute	Add another	er group						
	Successes	Total	Success Rate			p-value	Improvem	ent
Baseline	669,352	774,812	86% – 86% - 1 (86%)		+	-	_	
Variation 1	403,202	468,558	86% – 86% - (86%)		+	< 0.0001	-0.54% - (-0.3	

Not for sentiment score of greater than 0.3.

AB testi		tics						
Label			Number of succe	sses	Number of tria	als		
Baseline			762595		883572			Remove
Variation 1			309959		359798			Remove
Interval confi	idence level:							
0.95				Use multiple	e testing correct	ion: 🔽		
Compute	Add another	r group						
	Successes	Total	Success Rate			p-value	Improveme	ent
Baseline	762,595	883,572	86% - 86% - H	_	+	-	_	
Variation 1	309,959	359,798	86% – 86% · •	•	1+	0.019	-0.34% — (-0.1	

Not for sentiment score of greater than 0.4.

AB testi		stics						
Label			Number of succes	ses	Number of to	rials		
Baseline			842937		976343			Remove
Variation 1			229617		267027			Remove
Interval conf	idence level	:		Use multiple	testing correct	ction: <a>		
Compute	Add another	er group						
[Successes	Total	Success Rate			p-value	Improvem	ent
Baseline	842,937	976,343	86% - 86% (86%)		-	_	_	
Variation 1	229,617	267,027	86% – 86% - (86%)		1+	< 0.0001	-0.57% - (-0.4	

Not for sentiment score of greater than 0.5.

AB test		stics						
Label			Number of succ	esses	Number of tr	ials		
Baseline			938224		1086855			Remove
Variation 1			134330		156515			Remove
Interval conf	idence level	l:						
0.95				Use multiple	e testing correc	ction: 🔽		
Compute	Add anoth	er group						
	Successes	Total	Success Rate			p-value	Improvem	ent
Baseline	938,224	1,086,855	86% - 86% - H		+	_	_	
Variation 1	134,330	156,515	86% – 86% - (86%)		1+	< 0.0001	-0.79% - (-0.5	

Not for sentiment score of greater than 0.6.

	B test	A ing statis	stics						
La	bel			Number of succ	esses	Number of tr	ials		
Ba	seline			965691		1118706			Remove
Va	riation 1			106863		124664			Remove
0.9		Add anoth			Use multiple	testing correc	etion: 🔽		
		Successes	Total	Success Rate			p-value	Improvem	ent
Bas	seline	965,691	1,118,706	86% – 86% - 1 (86%)			_	_	
Vai	riation 1	106,863	124,664	86% – 86% - (86%)		1+	< 0.0001	-0.93% - (-0.7	

SQL Anonymous:)

Does emotionality make a tweet more retweetable?

Not for sentiment score of greater than 0.7.

A/B test	BA ting statis	stics						
Label			Number of succ	esses	Number of tria	ıls		
Baseline			994523		1152558			Remove
Variation 1	Variation 1		78031		90812			Remove
Interval con	fidence level	:						
0.95				Use multiple	e testing correcti	on: 🔽		
Compute	Add anothe	er group						
	Successes	Total	Success Rate			p-value	Improvem	ent
Baseline	994,523	1,152,558	86% – 86% - (86%)	-	-	-	_	
Variation 1	78,031	90,812	86% – 86% - (86%)		+	0.0023	-0.69% - (-0.4	

Steve Schluchter

Not for sentiment score of greater than 0.8.

AB testin		stics						
Label			Number of succ	esses	Number of tria	ıls		
Baseline			1035489		1199942			Remove
Variation 1			37065		43428			Remove
Interval confid	dence level							
0.95				Use multiple	testing correcti	on: 🔽		
Compute	Add anothe	er group						
	Successes	Total	Success Rate			p-value	Improver	nent
Baseline	1,035,489	1,199,942	86% – 86% - (86%)	-		_	_	
Variation 1	37,065	43,428	85% – 86% - (85%)		1+	< 0.0001		-0.71% 1%)

Not for sentiment score of greater than 0.9.

A/B test		stics					
Label		1	lumber of succe	esses	Number of trial	S	
Baseline			044070		1209878		Remove
Variation 1	Variation 1		28484		33492	33492	
Interval conf	idence level:						
0.95				Use multip	ole testing correction	n: 🔽	
Compute	Add another	er group					
	Successes	Total	Success Rate			p-value	Improvement
Baseline	1,044,070	1,209,878	86% – 86% (86%)	- 1	+	_	_
Variation 1	28,484	33,492	85% – 85% (85%)	-	1+	< 0.0001	-1.9% — -1% (-1.4%)

Do #hashtags make a tweet more retweetable?

Yes, they do!

AB testi		stics						
Label			Number of succ	cesses	Number of trial	s		
Baseline			690236		1866691			Remove
Variation 1			555134		620049			Remove
Interval confi	idence level							
0.95				Use multiple	e testing correction	on: 🔽		
Compute	Add another	r group						
	Successes	Total	Success Rate			p-value	Improve	ement
Baseline	690,236	1,866,691	37% – 37% (37%)	- (1+	-	_	
Variation 1	555,134	620,049	89% – 90% (90%)	-	+	< 0.0001		- 142% 42%)

Table of Contents

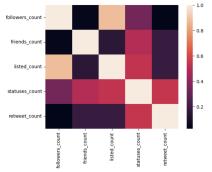
- 1 30000 ft overview
- 2 Two AB Tests

3 The most popular user on Twitter?

It depends on how you count it

First, a heatmap matrix of the Pearson correlation coefficients of some user data.

This illuminates which popularity measures are related to which.



It depends on how you count it

We focus on retweets. We could have counted friends, etc.. Here are the top 20 accounts, sorted by number of retweets, in descending order.

retweet count screen name realDonaldTrump 40852446 SenSanders POTUS 5855933 marcorubio 4964220 RepKihuen 4896177 RepDonBeyer 4809746 CoryBooker 4019408 RepMarkTakano 3827581 SenCortezMasto 3722421 RenKevinYoder 3681614 GovSandoval 3672595 RogerMarshallMD 3662501 RepAdamSchiff 3658373 RepAnnieKuster 3504108 SenWarren 3420202 RepRickLarsen 3208792 brianschatz 2942370 iahimes 2598640 RenEB1 2543544 SenSchumer 2536109

It depends on how you count it

We build a directed graph where a directed edge from one node to another means that a Twitter user tweeted at another twitter user. Here are the top 20 nodes according to the number of edges directed toward those nodes.

> 'POTUS', 505) 'WhiteHouse', 438) 'realDonaldTrump', 419) 'SpeakerRyan', 417) 'HouseGOP', 391) 'cspan', 382) 'thehill', 378) 'washingtonpost', 369) 'BarackObama', 356) 'VP', 342) 'CNN', 337) 'SpeakerBoehner', 337) 'USArmv', 332) 'USMC', 325) 'EPA', 324) 'nytimes', 322) 'USNavv', 322) 'WSJ', 313)

Steve Schluchter SQL Anonymous :)

'FoxNews', 303)
'DeptVetAffairs', 299)

Thank you!

• In any timezone! •