

# DATA MINING

## ASSIGNMENT: MILESTONE 4

---

NORAISHA YUSUF (WQD180008)

### *UPDATE*

*Crawl tweets  
Check sentiment and  
interpret tweets  
Merge dataset*



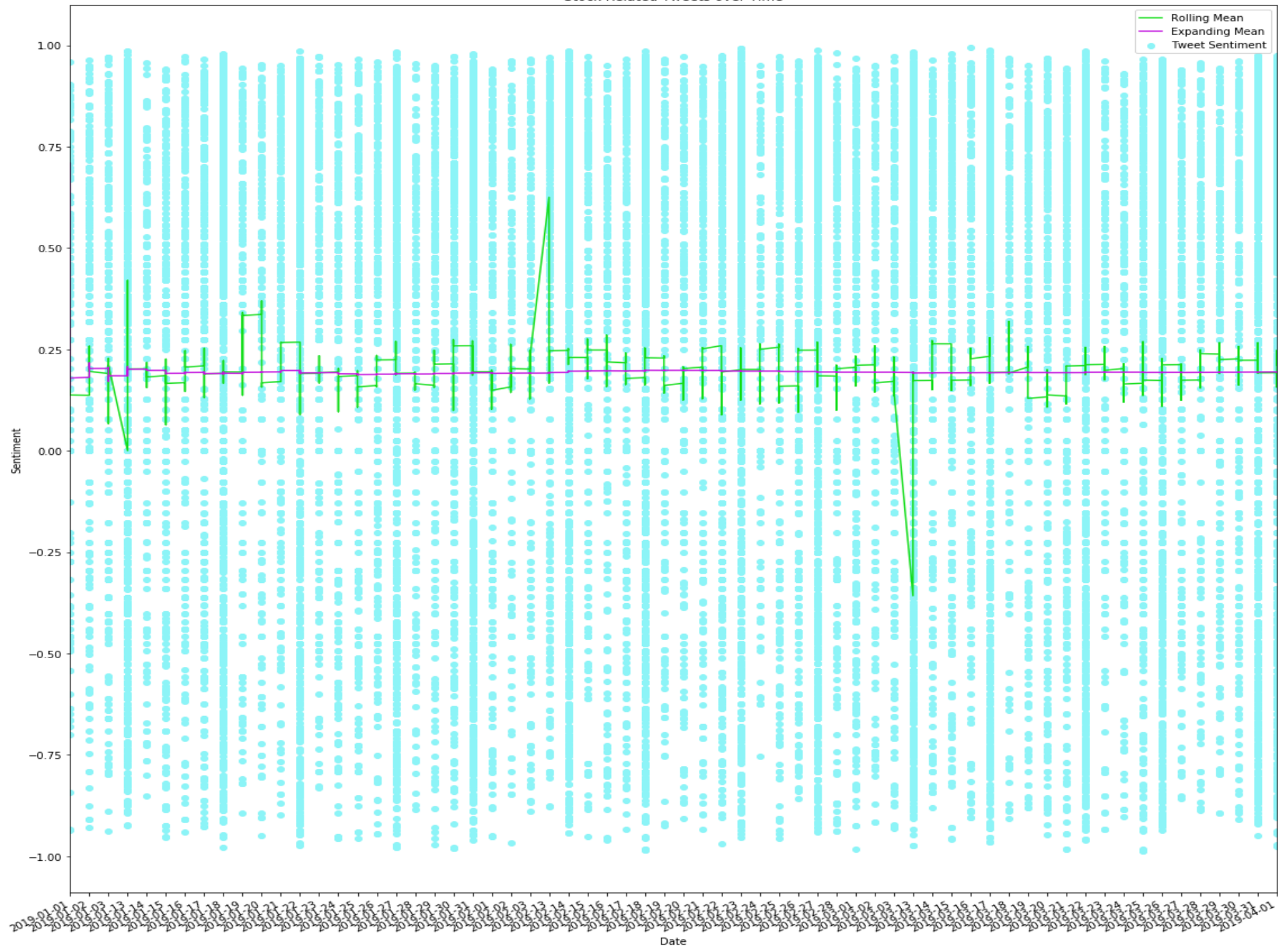
# *Part 1: Interpreting tweets*



- Sample tweets

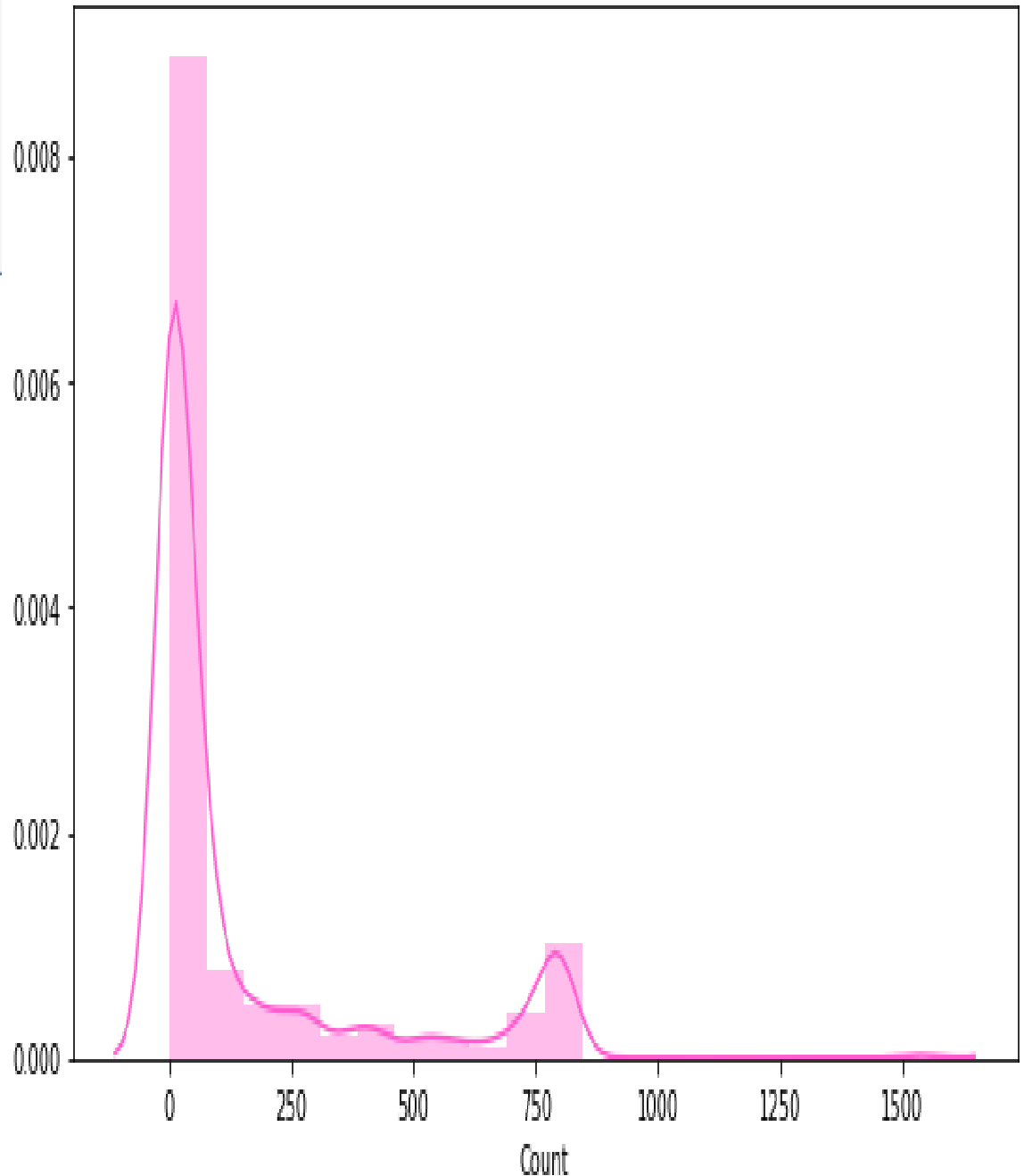
2/1/2019 0:01	PESTECH OR PESTE	StarBiz	<p	1.09E+18	1	0	2	Pestech secures TNB contracts worth RM280.89mil
2/1/2019 0:34	SAM ENGINEERING	Samuel.J.E	<p	1.09E+18	0	0	0	Oje wa ka eng ( engineering edition)
2/1/2019 0:38	AMBANK OR AMME	E. L. I. A.	<p class="	1.09E+18	1	0	0	I love you @ammb_nicole pic.twitter.com/bLKtUsrXBd
2/1/2019 0:39	BIOHLDG OR BIOAL	Sable Mug	<p class="	1.09E+18	0	0	0	Say it louder for the people in the back and I've never played it on the PS2 just it's emulator and DS
2/1/2019 0:45	HLFG OR HONG LEC	JediMASTE	<p class="	1.09E+18	0	0	2	And this Datin Sri isnâ€™t married. And she only sells insurance for Hong Leong! Now whereâ€™s Tax Revenue? How did she manage to own shares
2/1/2019 0:48	DRBHCOR OR DRB	Râ€™tBâ€™tK	<p class="	1.09E+18	0	0	0	friyay! (@ DRB-Hicom Defence Services) <a href="https://www.swarmapp.com/c/h4lksTjaWLRâ">https://www.swarmapp.com/c/h4lksTjaWLRâ</a>
2/1/2019 1:00	POS MALAYSIA	Pos Malay	<p	1.09E+18	5	8	19	[ANNOUNCEMENT]
2/1/2019 1:00	MALAYSIA AIRPORT	Firdaus Ha	<p class="	1.09E+18	2	0	1	AirAsia ups claim against Malaysia Airports by MYR80 million <a href="https://www.flightglobal.com/news/articles/airasia-ups-claim-against-malaysia-airpor">https://www.flightglobal.com/news/articles/airasia-ups-claim-against-malaysia-airpor</a>
2/1/2019 1:01	AEMULUS OR AEMI	emi	<p class="	1.09E+18	0	0	0	im starting a petition for @JTHS_204 to cancel school tomorrow
2/1/2019 1:01	CRESBLD OR CREST	StarBiz	<p	1.09E+18	0	0	0	Crest Builder unit wins RM99.6mil construction job
2/1/2019 1:11	GKENT OR GEORGE	Saeger Ryr	<p	1.09E+18	14	0	0	Considering how one-dimensional and boring Superman *could* be, he's been blessed with a lot of outstanding actors.
2/1/2019 1:24	TENAGA NASIONAL	EV Club M	<p	1.09E+18	1	1	0	Seems like a no-brainer for @Tenaga_Nasional to offer something similar, or at least a one-time rebate for home EVSE installation. Like the Gillette r
2/1/2019 4:47	FBMKLCI-H6P OR B	Palm Oil A	<p class="	1.09E+18	3	0	0	Bursa Malaysia is closed today Friday 1st of February for Federal Territory day. There will be no CPO Futures trading. Trading resumes on Monday 4th
2/1/2019 4:47	KELINGTON GROU	Darren Ha	<p class="	1.09E+18	2	1	0	One factor you canâ€™t overlook is the calibre of that blueline with Valimaki, Kylington and Andersson all in the NHL. No future NHLers in the group
2/1/2019 4:47	BURSA MALAYSIA	Palm Oil A	<p class="	1.09E+18	3	0	0	Bursa Malaysia is closed today Friday 1st of February for Federal Territory day. There will be no CPO Futures trading. Trading resumes on Monday 4th
2/1/2019 4:48	MHB OR MALAYSIA	KB Associa	<p	1.09E+18	1	0	0	Registration Closing Soon for IMCA Trainee AIR Diving Supervisor course that is scheduled on the 11-15 Feb 2019 in Malaysia, KL.
2/1/2019 4:50	HLFG OR HONG LEC	ITâ€™S SL	<p class="	1.09E+18	17	0	7	No matter what learning is necessary and can be fun too! If youâ€™re a learner and wanted to explore more then you should explore Hong Leong So
2/1/2019 4:59	SP500-HG OR SP50	IANS Twee	<p	1.09E+18	2	0	0	The #Dow Jones Industrial Average fell 15.19 points, or 0.06 per cent, to 24,999.67. The #SP500 increased 23.05 points, or 0.86 per cent, to 2,704.10
2/1/2019 5:08	AIRASIA OR AIRASI	Mineski Pr	<p	1.09E+18	8	0	0	We advance to winners match of group A!
2/1/2019 5:15	MALAYSIA AIRPORT	TheMalays	<p class="	1.09E+18	0	0	1	Malaysia Airports expects 16% more passenger traffic during CNY <a href="https://www.themalaysianinsight.com/s/130011â">https://www.themalaysianinsight.com/s/130011â</a>
2/1/2019 5:15	DATAPRP OR DATA	Lynn Langi	<p	1.09E+18	3	1	0	so impressed with @GCPcloud #Dataprep
2/1/2019 5:24	DATAPRP OR DATA	Vikram Â	<p class="	1.09E+18	1	0	0	Dataprep is awesome! Esp since you can schedule it and forget about it and it keeps working for you :)

Stock Related Tweets over Time



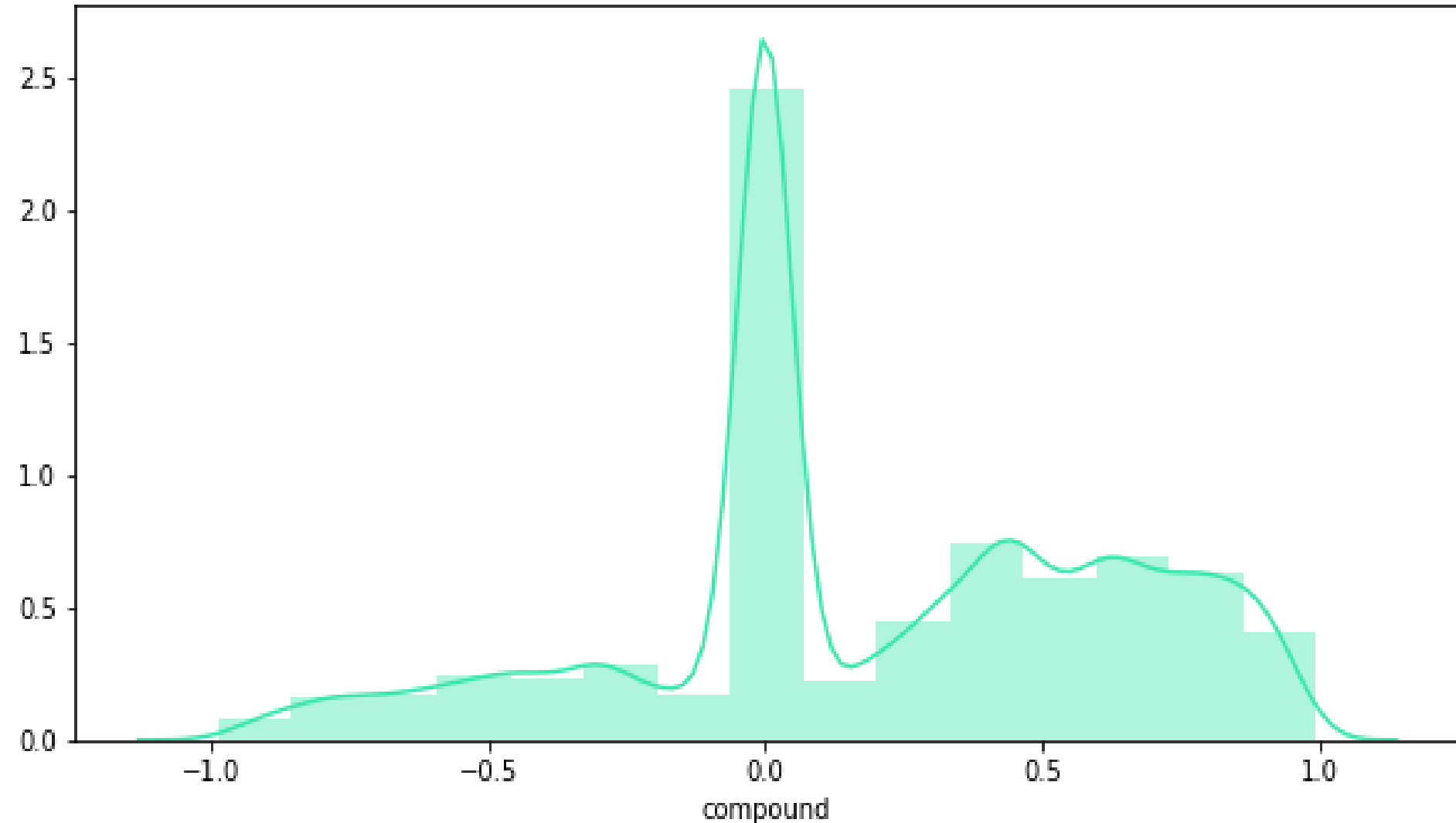
- Distribution of number of tweets by stocks.
- Majority of stocks have lesser than 10 tweets.

	filters	Count
0	AAX OR AIRASIA X	406
1	AAX-WA OR AIRASIA X	393
2	ABMB OR ALLIANCE BANK AND MALAYSIA	28
3	ACOSTEC OR ACOUSTECH	9
4	ADVCON OR ADVANCECON	17
5	AEMULUS OR AEMULUS	106
6	AEON OR AEON CO	88
7	AEONCR OR AEON CO	88
8	AFFIN OR AFFIN BANK	184
9	AIRASIA OR AIRASIA GROUP	282
10	AIRASIAC67 OR AIRASIA	797
11	ALAM MARITIM	37
12	ALLIANZ OR ALLIANZ MALAYSIA	155
13	AMBank OR AMMB	568
14	AMFIRST	62
15	ANCOM	788
16	ANNJOO OR ANN JOO RESOURCES	4
17	ARREIT OR AMANAH RAYA	46
18	ASDION OR ASDION	101
19	ASIAPAC OR ASIAN PAC	163
20	ASTRO MALAYSIA	528
21	ATAIMS OR ATA IMS	14
22	ATURMJU OR ATURMAJU RESOURCES	4
23	AWC-WA OR AWC	735
24	AXIATA OR AXIATA GROUP BERHAD	5
25	AXREIT OR AXIS REITS	17
26	AZRB OR AHMAD ZAKI RESOURCES	6
27	BAHVEST OR BAHVEST RESOURCES	2

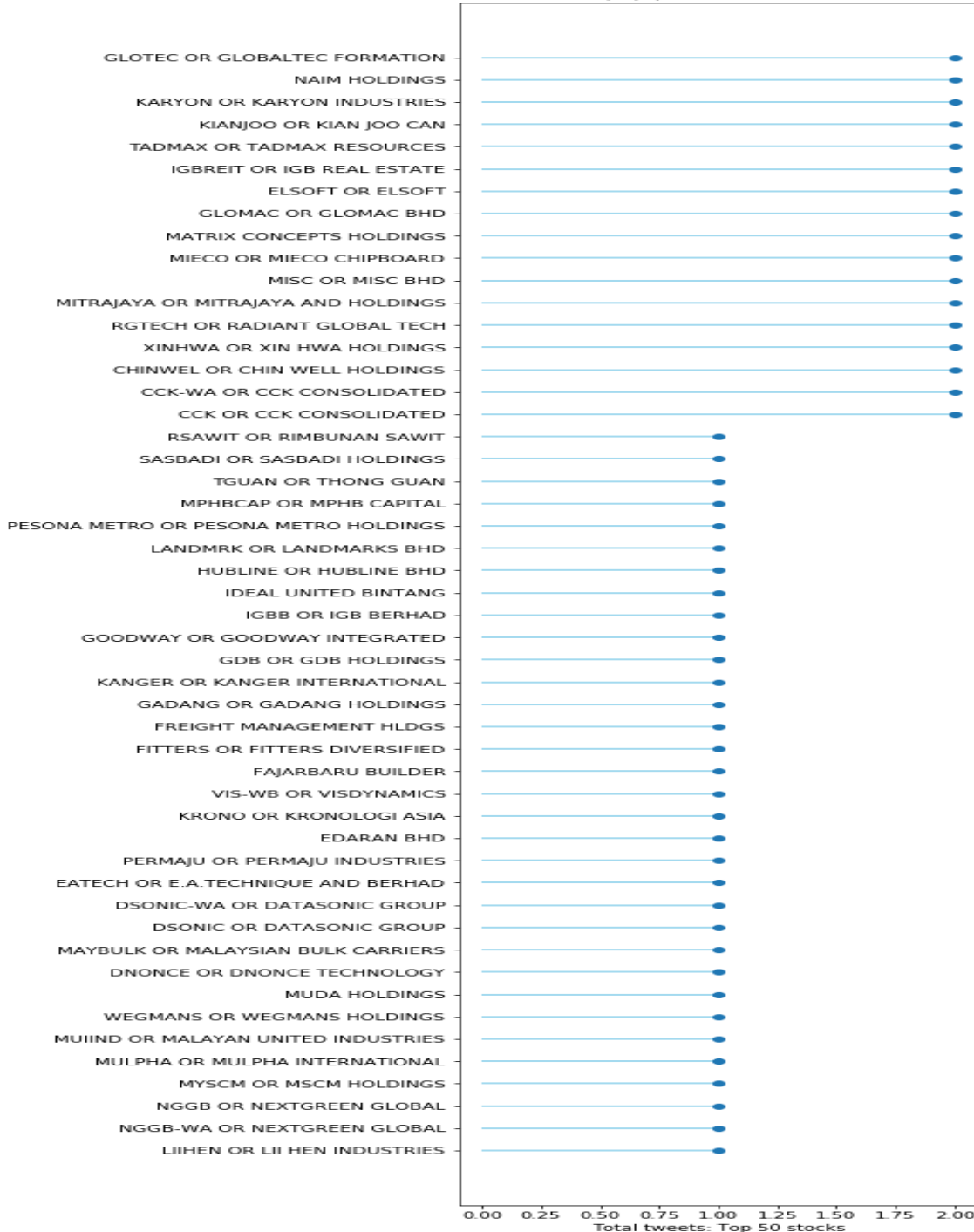


- Majority obtain sentiment scores of near zero, which indicates near-neutral sentiment

Overall Distribution of Sentiment Scores



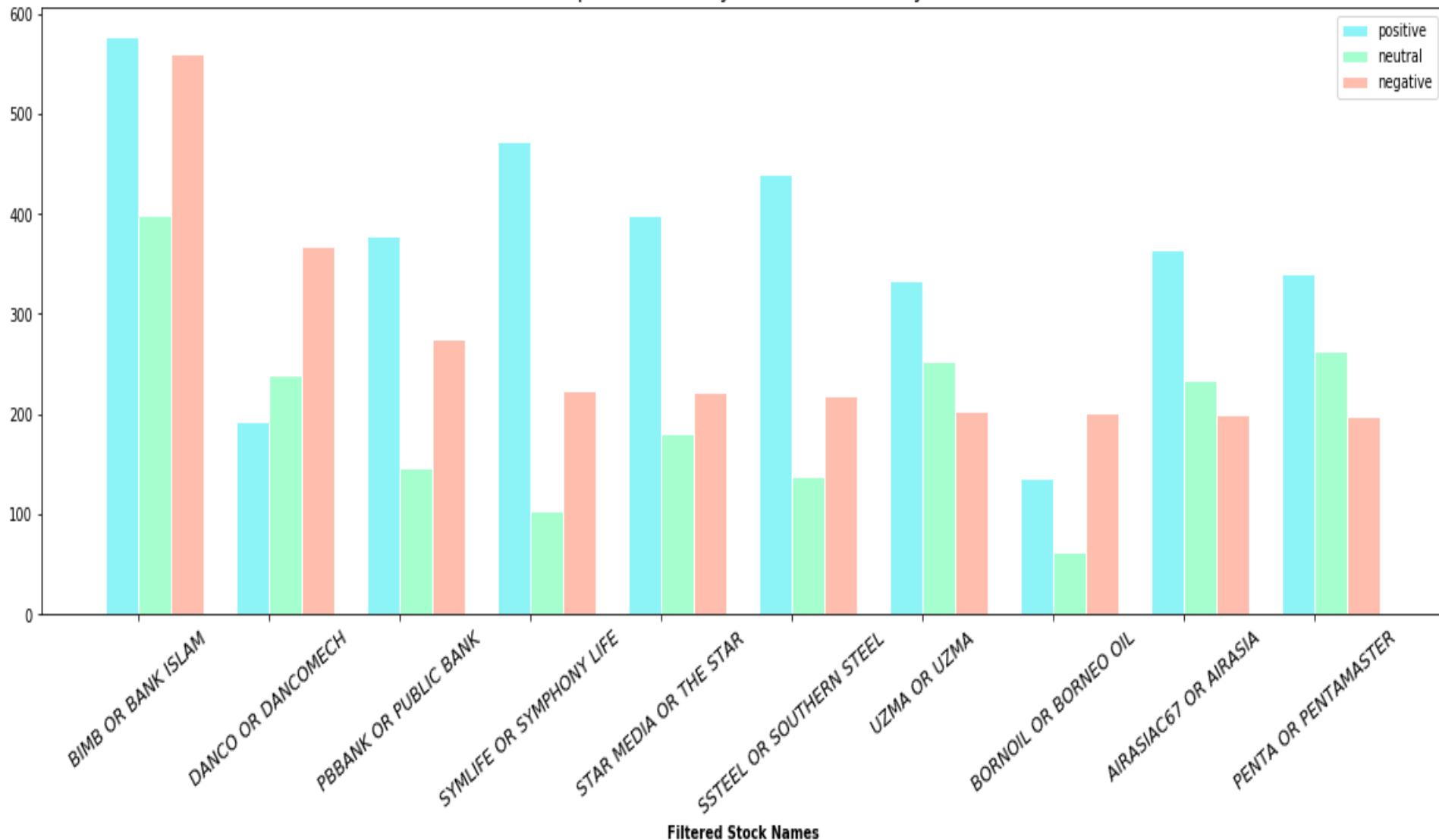
A vertical lolipop plot



Top 50 tweeted stocks  
(comparison of relative  
value)

- Top 10 tweeted stocks with sentiment polarity labels

Stock: Top 10 Tweets by Sentiment Polarity Counts





## *Part 2: Interpreting merged dataset*

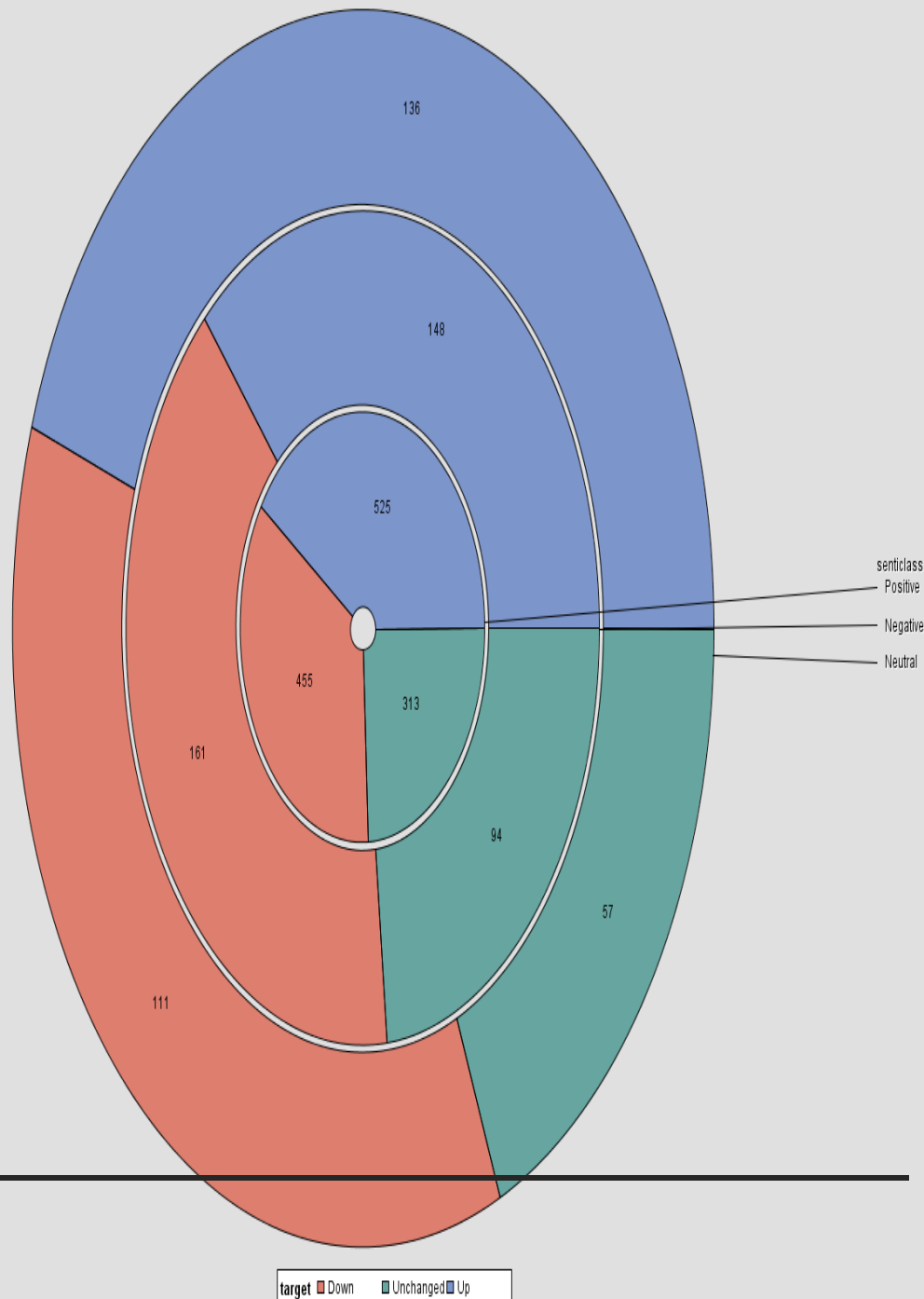


- Target variable – for daily changes of closing stock prices

- ☐ Up – price is increasing
- ☐ Down – price is decreasing
- ☐ Unchanged – no price changes

Sample Table

filters	day	name	board	close	open	high	low	volume	dif	difpct	avg_senti	senticlass	target
AAX OR AIRASIA ...	2019-01-03	AAX	Main	0.24	0.23	0.24	0.23	8801100	0.009999999999...	0.043478260869...	0.44853043478...	Positive	Up
AAX OR AIRASIA ...	2019-01-04	AAX	Main	0.26	0.24	0.26	0.24	30545800	0.020000000000...	0.083333333333...	0.10046666666...	Positive	Up
AAX OR AIRASIA ...	2019-01-07	AAX	Main	0.305	0.26	0.31	0.26	84608800	0.044999999999...	0.17307692307...	0.08164285714...	Positive	Up
AAX OR AIRASIA ...	2019-01-08	AAX	Main	0.29	0.3	0.305	0.28	69337400	-0.015000000000...	-0.049180327868...	0.17878666666...	Positive	Down
AAX OR AIRASIA ...	2019-01-09	AAX	Main	0.29	0.29	0.3	0.285	15586200	0	0	0.1779	Positive	Unchanged
AAX OR AIRASIA ...	2019-01-10	AAX	Main	0.29	0.29	0.3	0.285	28305200	0	0	-0.06743333333...	Negative	Unchanged
AAX OR AIRASIA ...	2019-01-11	AAX	Main	0.29	0.295	0.295	0.28	16799600	0	0	0.5325	Positive	Unchanged
AAX OR AIRASIA ...	2019-01-14	AAX	Main	0.285	0.29	0.295	0.28	22089100	-0.005000000000...	-0.017241379310...	0.39279999999...	Positive	Down
AAX OR AIRASIA ...	2019-01-15	AAX	Main	0.3	0.285	0.3	0.285	32894100	0.015000000000...	0.052631578947...	0.2056	Positive	Up
AAX OR AIRASIA ...	2019-01-16	AAX	Main	0.3	0.3	0.305	0.29	25793300	0	0	0	Neutral	Unchanged
AAX OR AIRASIA ...	2019-01-17	AAX	Main	0.295	0.3	0.3	0.29	11649500	-0.005000000000...	-0.01666666666...	0	Neutral	Down
AAX OR AIRASIA ...	2019-01-18	AAX	Main	0.29	0.295	0.295	0.29	5171300	-0.005000000000...	-0.016949152542...	0.12040000000...	Positive	Down
AAX OR AIRASIA ...	2019-01-22	AAX	Main	0.29	0.29	0.31	0.28	48519200	0	0	0.100475	Positive	Unchanged
AAX OR AIRASIA ...	2019-01-23	AAX	Main	0.285	0.285	0.29	0.28	12168600	-0.005000000000...	-0.017241379310...	-0.12254000000...	Negative	Down
AAX OR AIRASIA ...	2019-01-24	AAX	Main	0.285	0.28	0.285	0.285	8801100	0	0	0.00499999999...	Negative	Unchanged



- As seen in the multi-layer pie chart,
  - The biggest proportion are neutral tweets (outer pie chart layer), which are dominated by increasing class
  - Followed by negative tweets (2<sup>nd</sup> layer), dominated by decreasing class
  - Lastly, the positive tweets pie chart (inner most layer) dominated by increasing stock prices

*THE END*