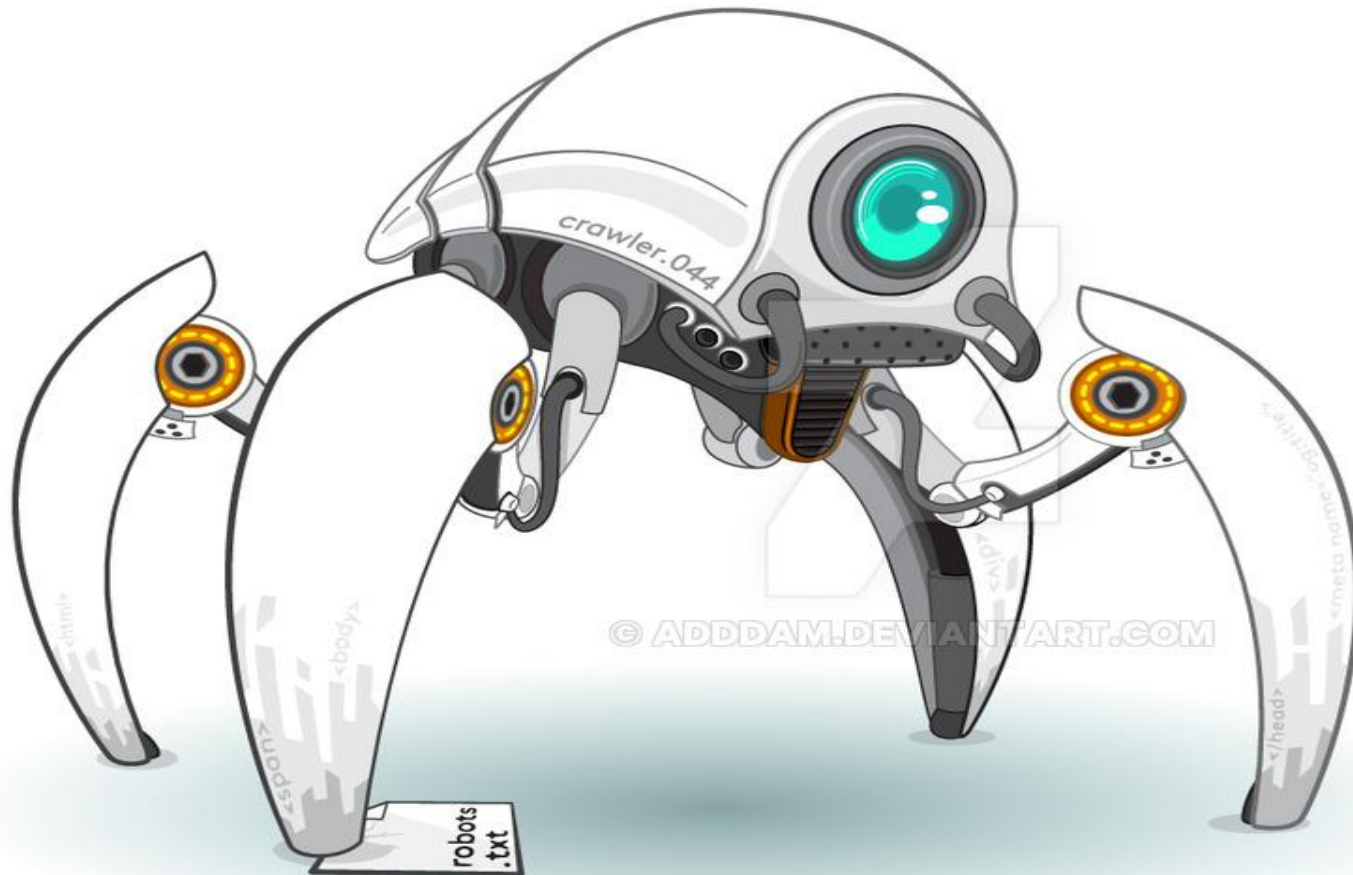


NORAISHA YUSUF (WQD180008)

DATA MINING

ASSIGNMENT: MILESTONE 6



Part 1: Objective



Recommend
potential stocks
for investment

Market Movers

Active	Gainers	% Gainers	Losers	% Losers
Symbol	Open	High	Low	
HSI-H6S	0.550	0.555	0.520	
HIBISCS	1.000	1.000	0.975	
EKOVEST	0.760	0.785	0.755	
HSI-C5J	0.150	0.175	0.145	
ARMADA	0.195	0.200	0.190	
SAPNRG	0.290	0.295	0.285	
LAMBO	0.060	0.065	0.055	
EAH-WE	0.005	0.010	0.0	
TM-C41	0.105	0.105	0.0	
IRIS	0.165	0.165	0.150	

Part 2: Dataset



- Stock prices & volumes
- financial
- News
- Tweets
- Interests

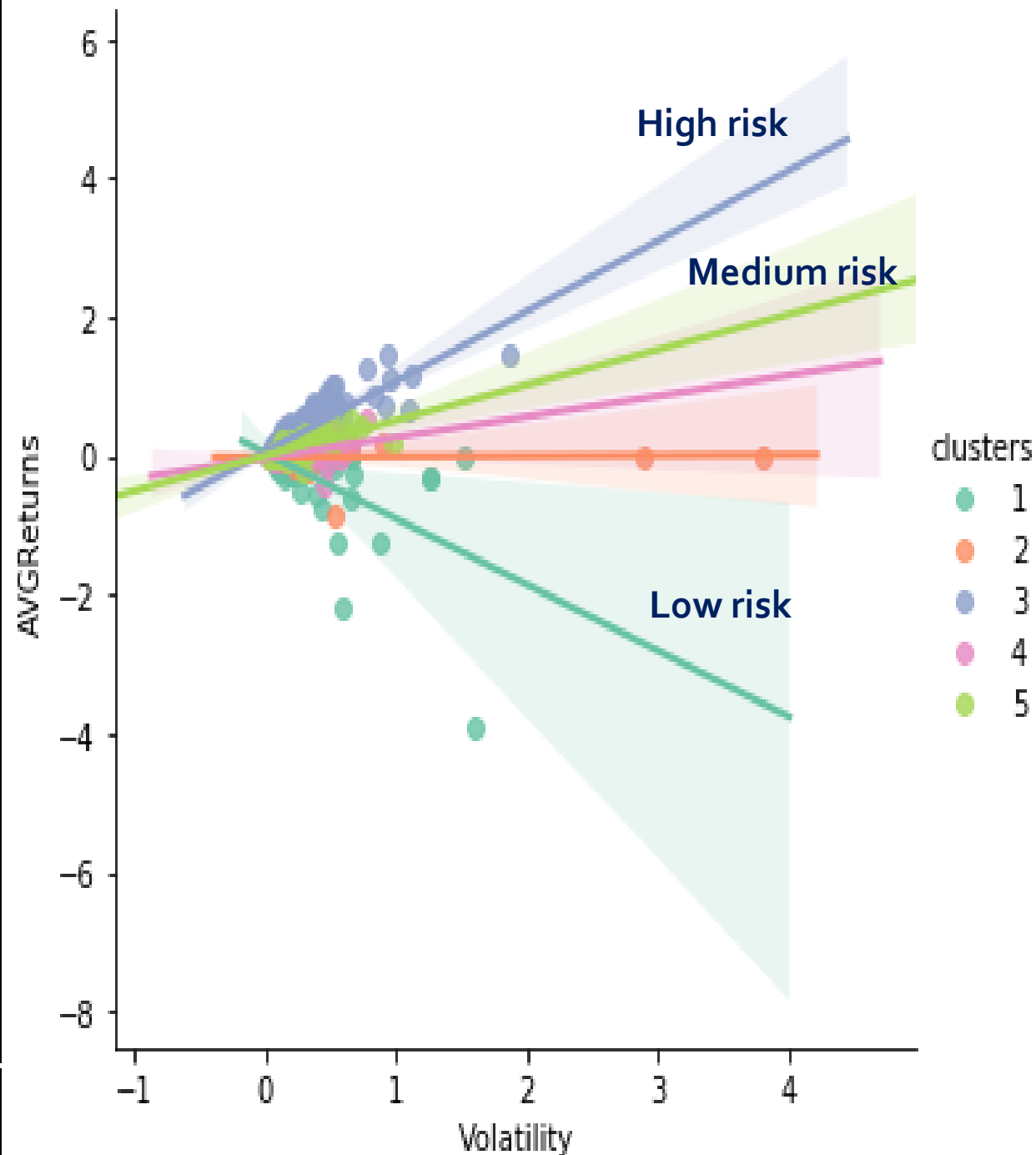
	A	B	C	D	E	F	G	H	I	J	K	L
1	name	day	close	volume	returns	risk	newspolar	newssenticlass	tweetspol	tweetsent	interest	target
2	AAX	3/1/2019	0.24	8801100	0.01	low risk	0	none	0	none	1	none
3	AAX	4/1/2019	0.26	30545800	0.02	low risk	0	none	0.209075	Positive	1	up
4	AAX	7/1/2019	0.305	84608800	0.045	low risk	0	none	0	none	0	up
5	AAX	8/1/2019	0.29	69337400	-0.015	low risk	0	none	0	none	0	down
6	AAX	9/1/2019	0.29	15586200	0	low risk	0	none	0.168395	Positive	0	none
7	AAX	10/1/2019	0.29	28305200	0	low risk	0	none	0	none	0	none
8	AAX	11/1/2019	0.29	16799600	0	low risk	0	none	0	none	0	none
9	AAX	14/1/2019	0.285	22089100	-0.005	low risk	0.2827	Positive	0	none	2	none
10	AAX	15/1/2019	0.3	32894100	0.015	low risk	0.7906	Positive	0	none	2	up
11	AAX	16/1/2019	0.3	25793300	0	low risk	0	none	0	none	2	none
12	AAX	17/1/2019	0.295	11649500	-0.005	low risk	0	none	0	none	4	none
13	AAX	18/1/2019	0.29	5171300	-0.005	low risk	-0.4404	Negative	0.4437	Positive	3	none
14	AAX	22/1/2019	0.29	48519200	0	low risk	0.4404	Positive	0.090155	Positive	3	none
15	AAX	23/1/2019	0.285	12168600	-0.005	low risk	0.39225	Positive	0	none	1	none
16	AAX	24/1/2019	0.285	8826500	0	low risk	-0.3548	Negative	0	none	3	none
17	AAX	25/1/2019	0.29	9333000	0.005	low risk	0.4939	Positive	0	none	1	none
18	AAX	28/1/2019	0.29	19275900	0	low risk	0.5574	Positive	0	none	1	none
19	AAX	29/1/2019	0.315	41545600	0.025	low risk	0.1971	Positive	0	none	1	up
20	AAX	30/1/2019	0.295	24305100	-0.02	low risk	0.37895	Positive	0	none	2	down
21	AAX	31/1/2019	0.29	13397800	-0.005	low risk	-0.3477	Negative	0.059395	Positive	1	none
22	AAX	4/2/2019	0.295	2467000	0.005	low risk	-0.83145	Negative	0	none	0	none
23	AAX	7/2/2019	0.29	7705500	-0.005	low risk	0	none	0	none	0	none
24	AAX	8/2/2019	0.285	9437900	-0.005	low risk	0	none	0	none	0	none

Part 3: Methodology & Results



What are the stock risks?

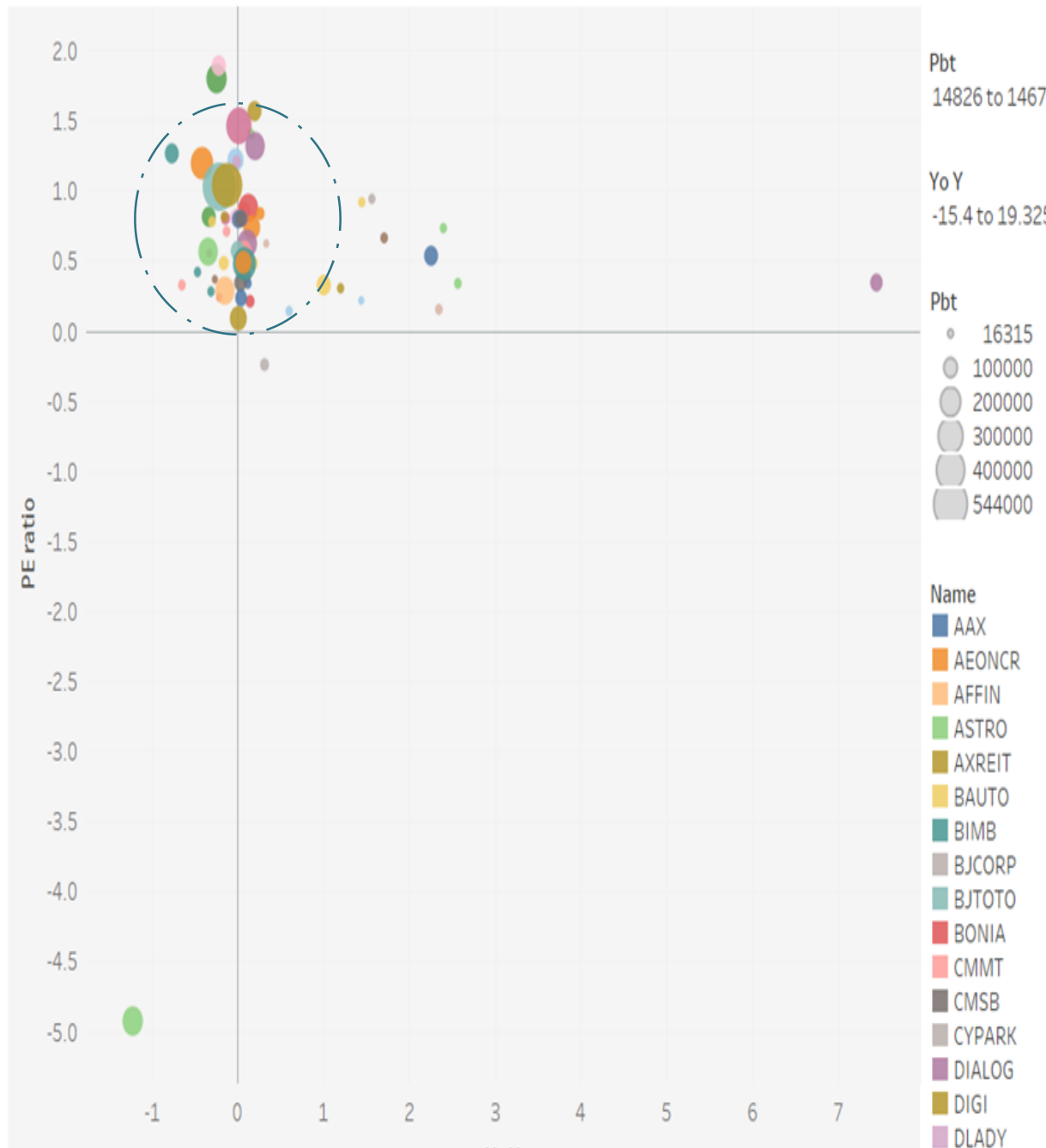
- Concept of portfolio diversification
- Clustering, correlation
- volatility vs. returns



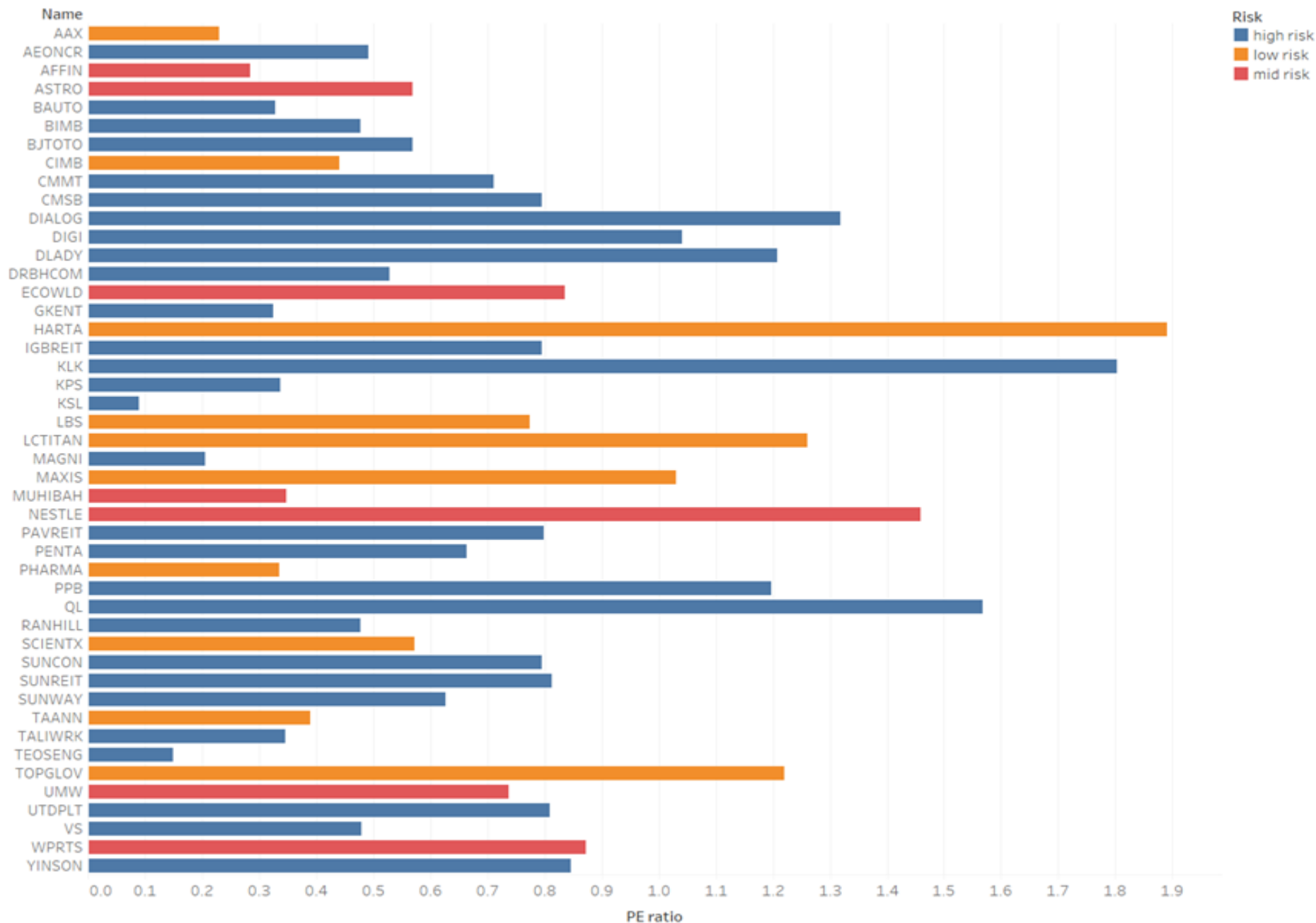
Is the stock valuable?

- Compute latest P/E ratio
- Company's growth

PE ratio vs. company growth

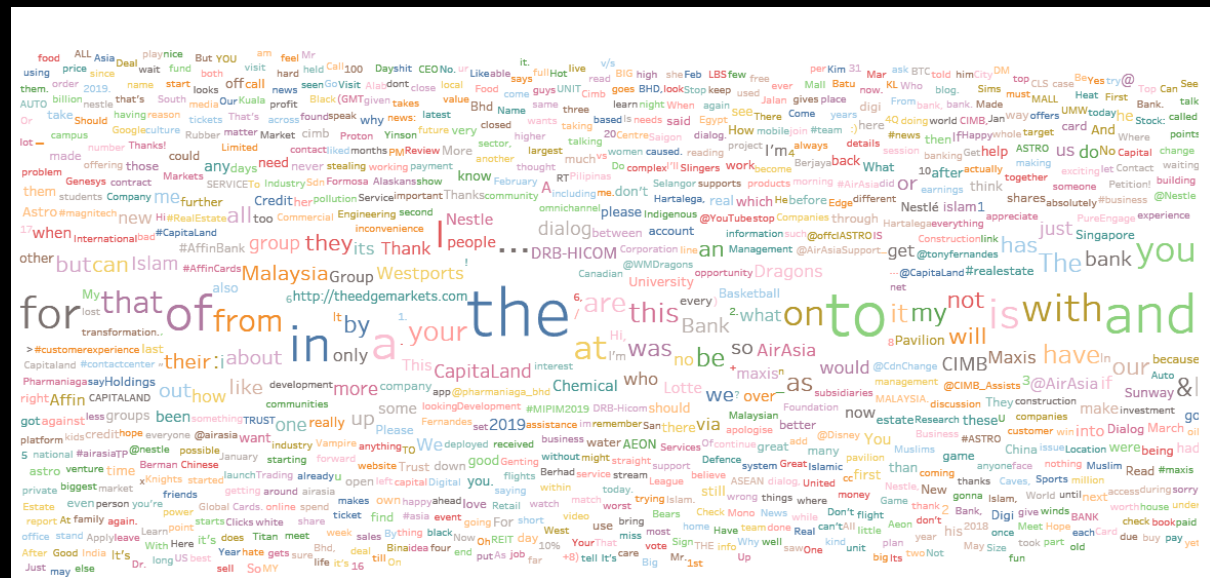


Potential Stocks



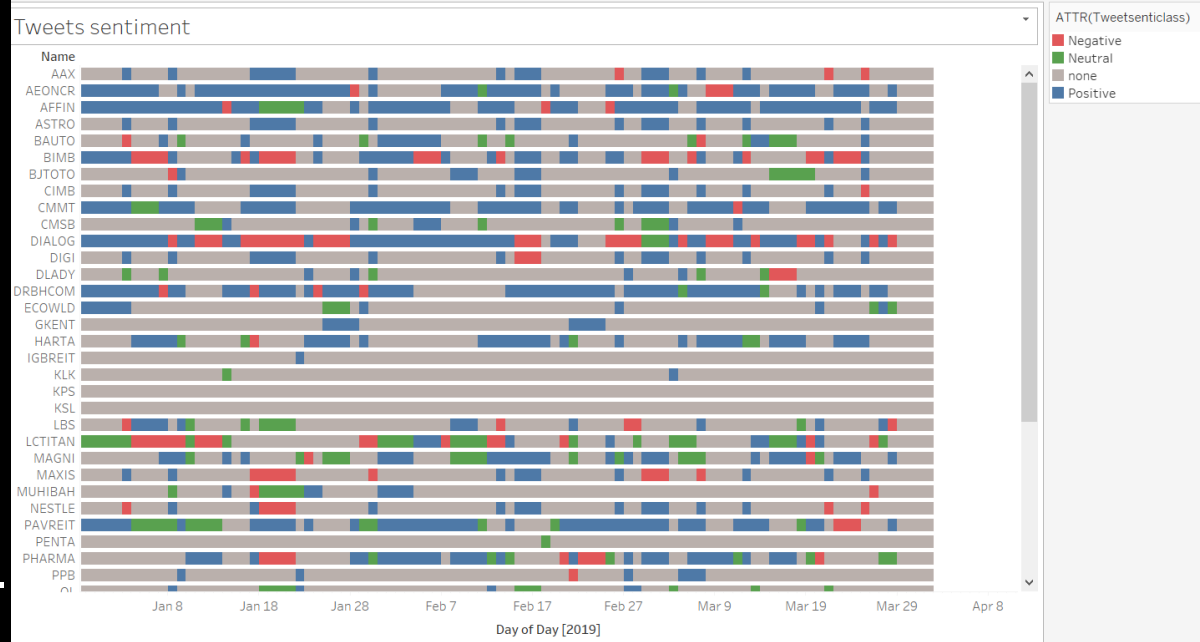
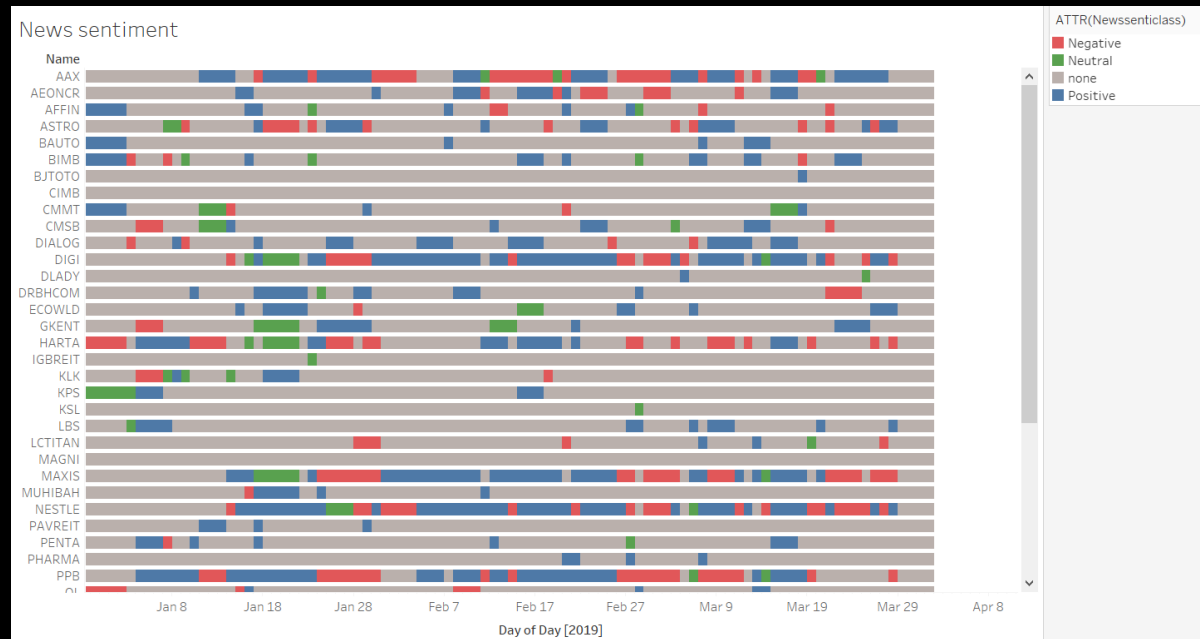
PE ratio as an attribute for each Name. Color shows details about Risk. The data is filtered on PE ratio, Pbt Norm and YoY norm. The PE ratio filter ranges from 0 to 3. The Pbt Norm filter ranges from 0 to 6.006291215. The YoY norm filter ranges from -1 to 19.3253.

- News sentiment
- Tweets sentiment



Does sentiment affect stock prices?

- Polarity scoring





Name

☐ (All)

☒ AAX

☐ AEONCR

☐ MAXIS

☐ NESTLE

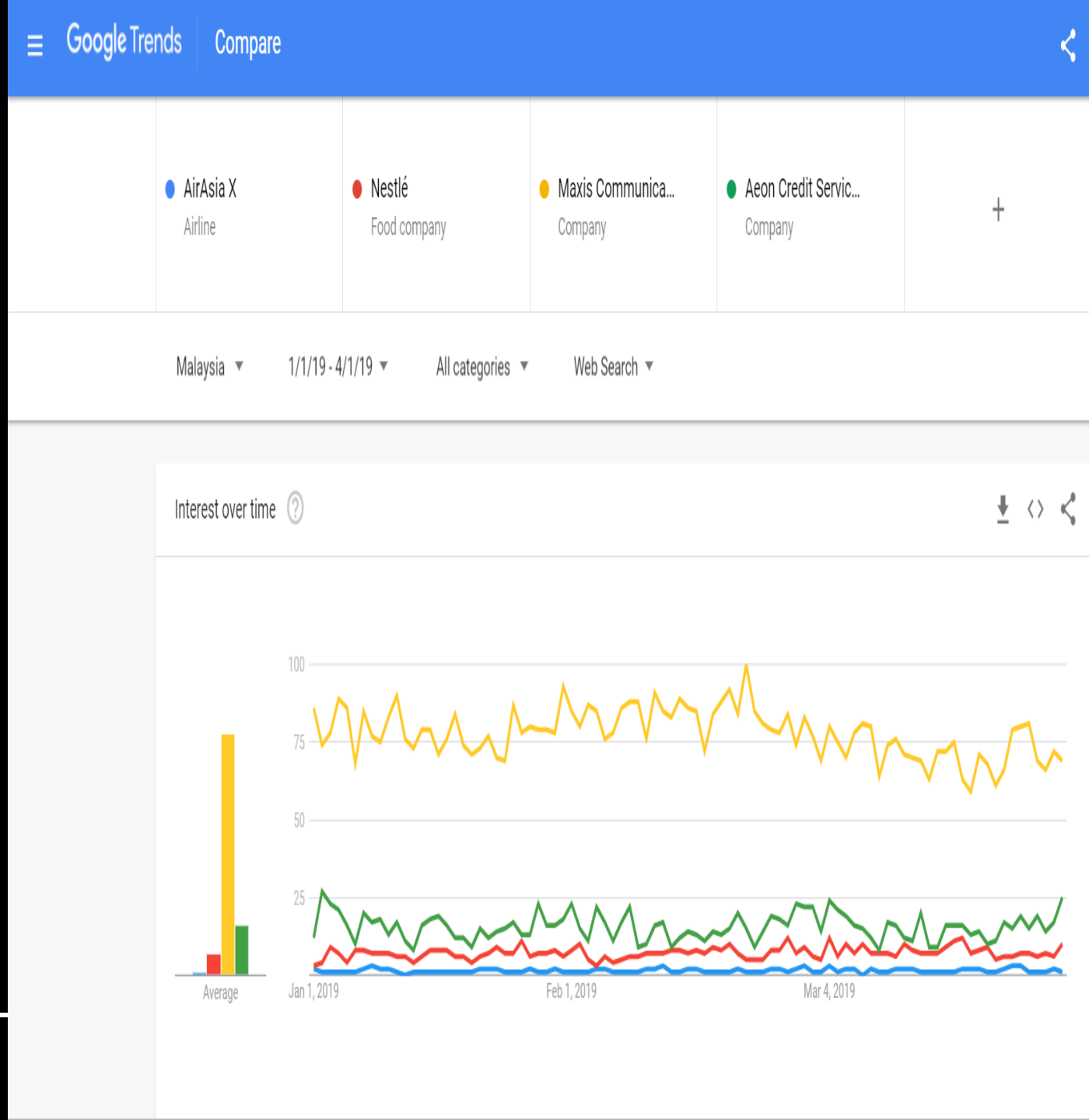
Name

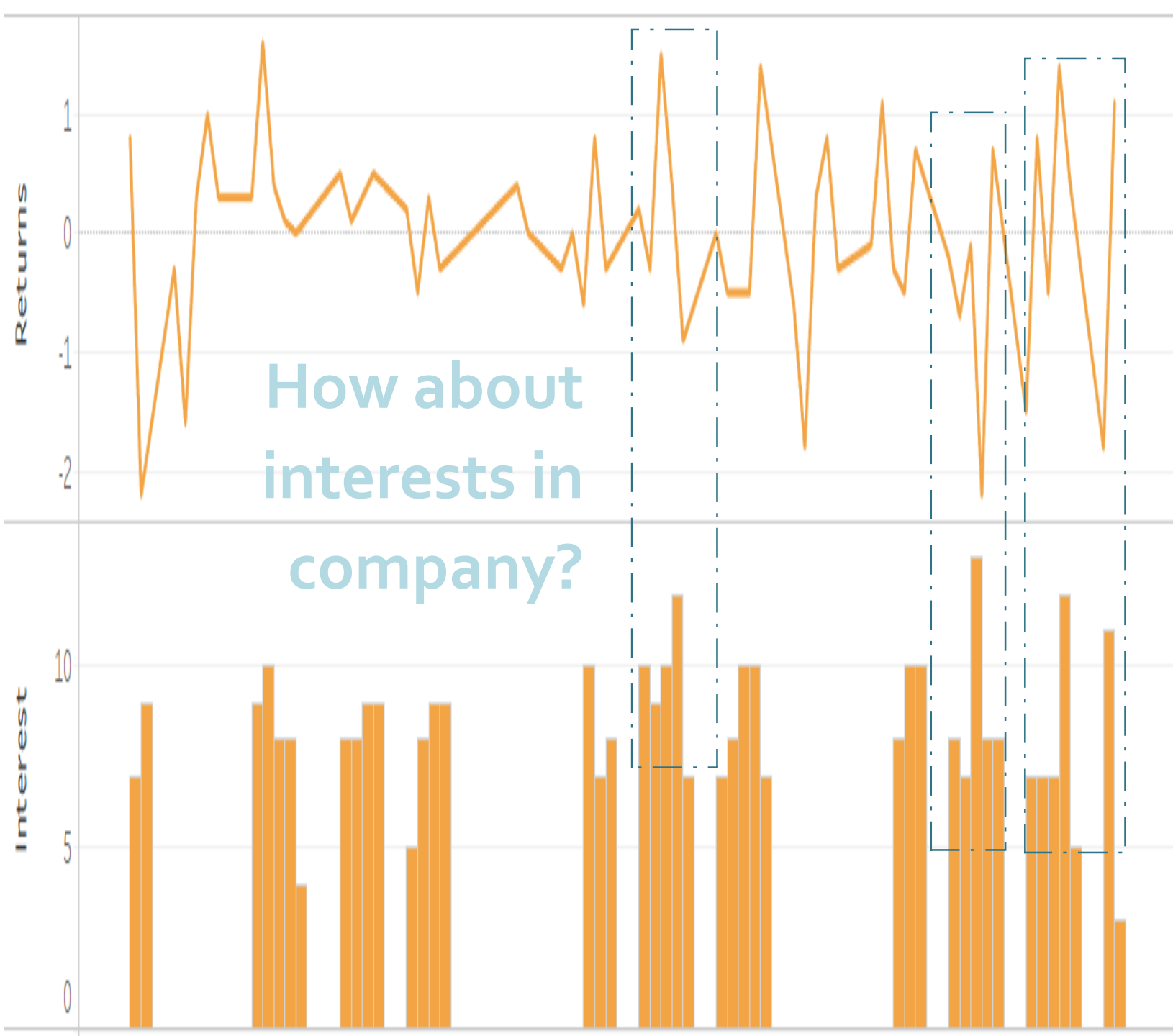
AAX

Does
sentiment
affect
stock
prices?

How about interests in company?

- Google Trends





Name

☐ (All)

☐ AAX

☐ AEONCR

☐ MAXIS

☒ NESTLE

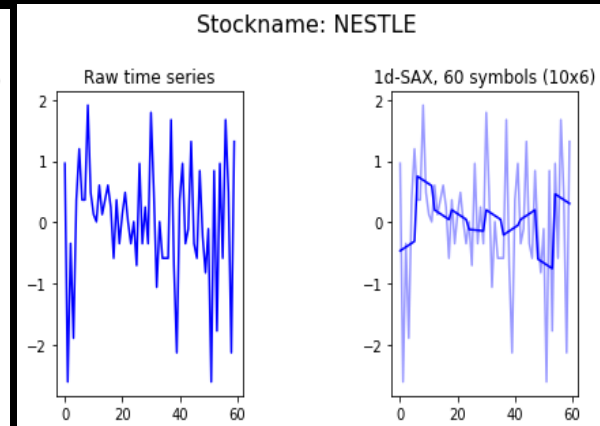
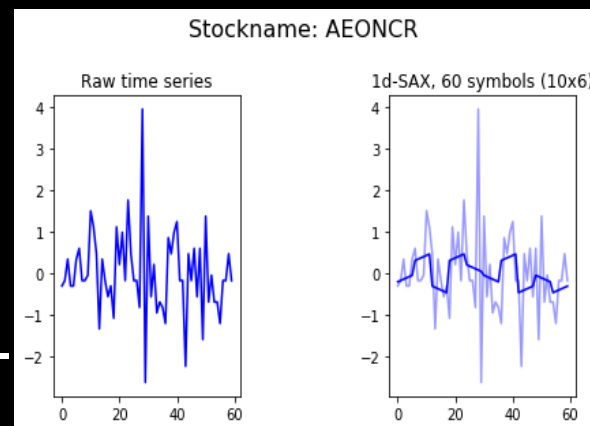
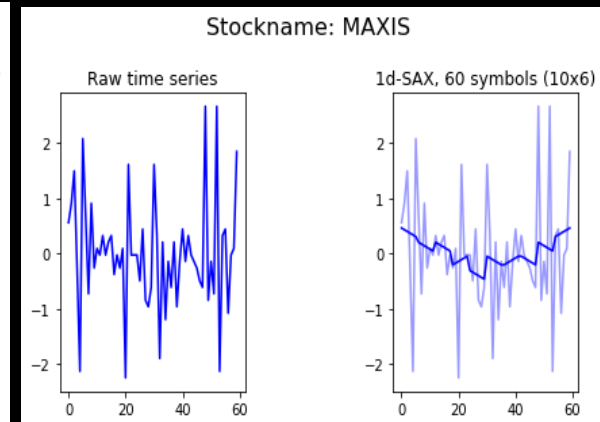
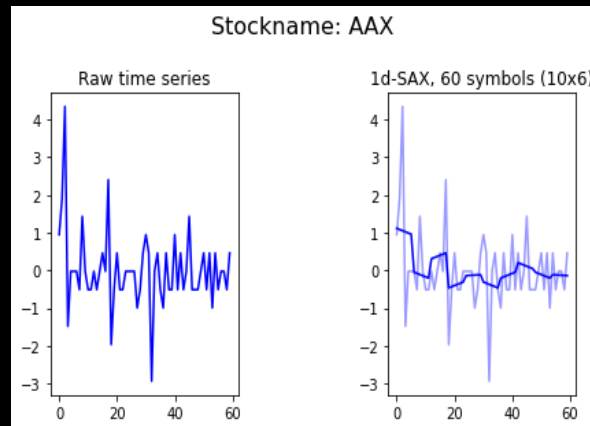
Name

☒ NESTLE

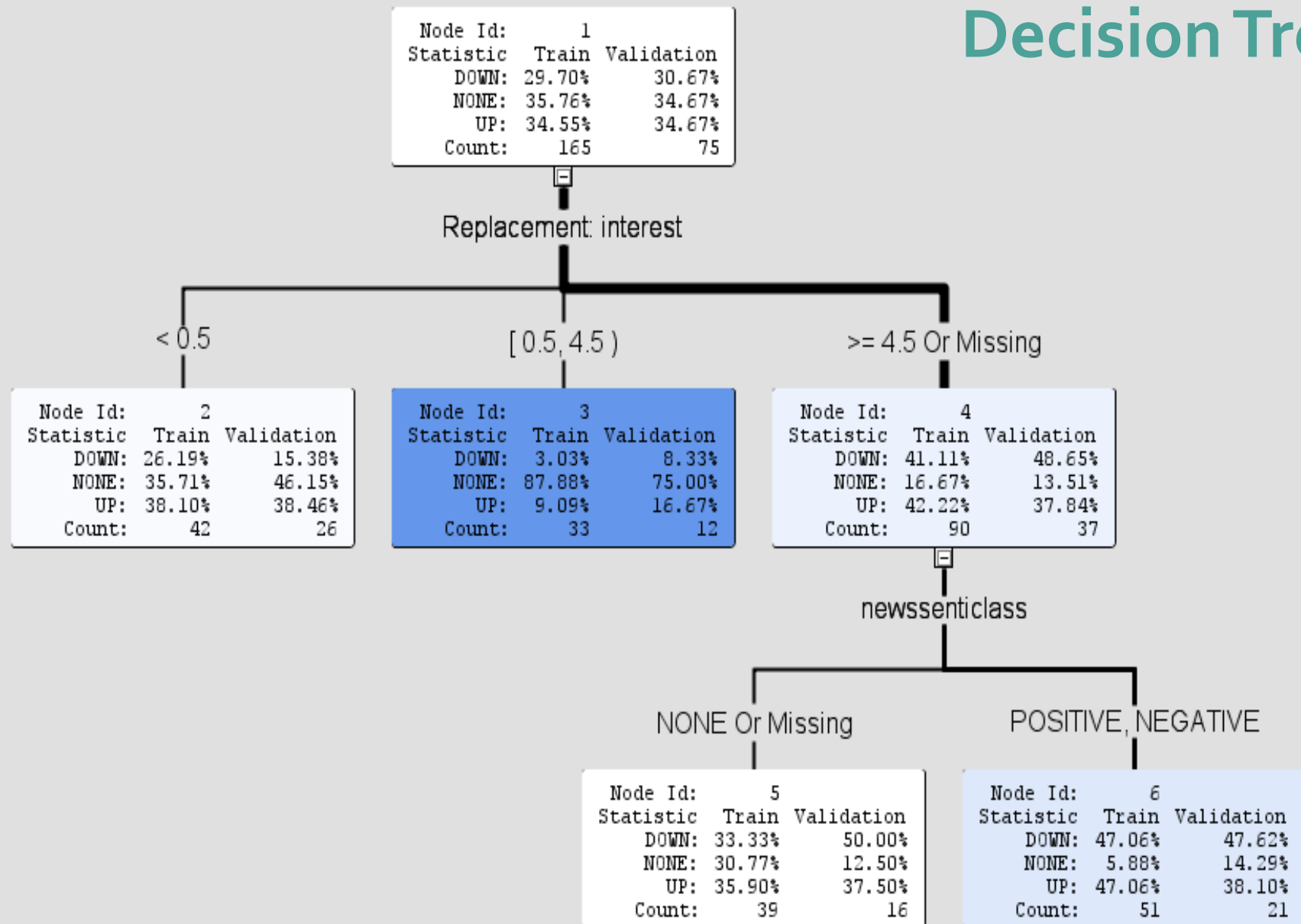
Sentiment & interests on stocks?

- ML classification –
Decision tree

Stock name	Company	Business sector	Risk category
AAX	AirAsia X Bhd	Aviation	Low risk
MAXIS	MAXIS Bhd	Telco	Low risk
NESTLE	NESTLE (Malaysia) Bhd	Consumer	Medium risk
AEONCR	Aeon Credit Service (M) Bhd	Finance	High risk



Decision Tree 1



Decision Tree 2

Node Id:	1	
Statistic	Train	Validation
DOWN:	29.70%	30.67%
NONE:	35.76%	34.67%
UP:	34.55%	34.67%
Count:	165	75

Replacement: interest

< 0.5

[0.5, 4.5)

[4.5, 61) Or Missing

[61, 69)

[69, 82.5)

>= 82.5

Node Id:	2	
Statistic	Train	Validation
DOWN:	26.19%	15.38%
NONE:	35.71%	46.15%
UP:	38.10%	38.46%
Count:	42	26

Node Id:	3	
Statistic	Train	Validation
DOWN:	3.03%	8.33%
NONE:	87.88%	75.00%
UP:	9.09%	16.67%
Count:	33	12

Node Id:	4	
Statistic	Train	Validation
DOWN:	42.86%	47.83%
NONE:	15.87%	8.70%
UP:	41.27%	43.48%
Count:	63	23

Node Id:	5	
Statistic	Train	Validation
DOWN:	80.00%	
NONE:	0.00%	
UP:	20.00%	
Count:	5	0

Node Id:	6	
Statistic	Train	Validation
DOWN:	28.57%	45.45%
NONE:	35.71%	27.27%
UP:	35.71%	27.27%
Count:	14	11

Node Id:	7	
Statistic	Train	Validation
DOWN:	25.00%	66.67%
NONE:	0.00%	0.00%
UP:	75.00%	33.33%
Count:	8	3

Replacement: newspolarity

< -0.0104

[-0.0104, 0.1956) Or Missing

[0.1956, 0.527)

[0.527, 0.7499)

>= 0.7499

Node Id:	14	
Statistic	Train	Validation
DOWN:	63.64%	33.33%
NONE:	0.00%	0.00%
UP:	36.36%	66.67%
Count:	11	3

Node Id:	15	
Statistic	Train	Validation
DOWN:	39.39%	58.33%
NONE:	30.30%	16.67%
UP:	30.30%	25.00%
Count:	33	12

Node Id:	16	
Statistic	Train	Validation
DOWN:	37.50%	40.00%
NONE:	0.00%	0.00%
UP:	62.50%	60.00%
Count:	8	5

Node Id:	17	
Statistic	Train	Validation
DOWN:	50.00%	100.00%
NONE:	0.00%	0.00%
UP:	50.00%	0.00%
Count:	6	1

Node Id:	18	
Statistic	Train	Validation
DOWN:	20.00%	0.00%
NONE:	0.00%	0.00%
UP:	80.00%	100.00%
Count:	5	2

Replacement: interest

< 9 Or Missing

>= 9

Node Id:	24	
Statistic	Train	Validation
DOWN:	83.33%	.
NONE:	0.00%	.
UP:	16.67%	.
Count:	6	0

Node Id:	25	
Statistic	Train	Validation
DOWN:	40.00%	33.33%
NONE:	0.00%	0.00%
UP:	60.00%	66.67%
Count:	5	3

Sentiment & interests on stocks?

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
REP interest	Replacement: inte...	2	1.0000	1.0000	1.0000
REP newspolarity	Replacement: ne...	1	0.4109	0.4157	1.0115
REP tweetspolar...	Replacement: twe...	0	0.0000	0.0000	-
newssenticlass		0	0.0000	0.0000	-
tweetssenticlass		0	0.0000	0.0000	-

Set	Decision Tree 1		Decision Tree 2	
	True positive	True negative	True positive	True negative
Training	37%	67.8%	50%	76.2%
Validation	38%	69.7%	46.2%	77.8%

Part 5: Future works



THE END