PREDICT $OCEAN PRICE WITH TWEET SENTIMENTS

OCEAN DATA CHALLENGE : OCEAN TOKEN SENTIMENT ANALYSIS

CONTENTS

CONTEXT

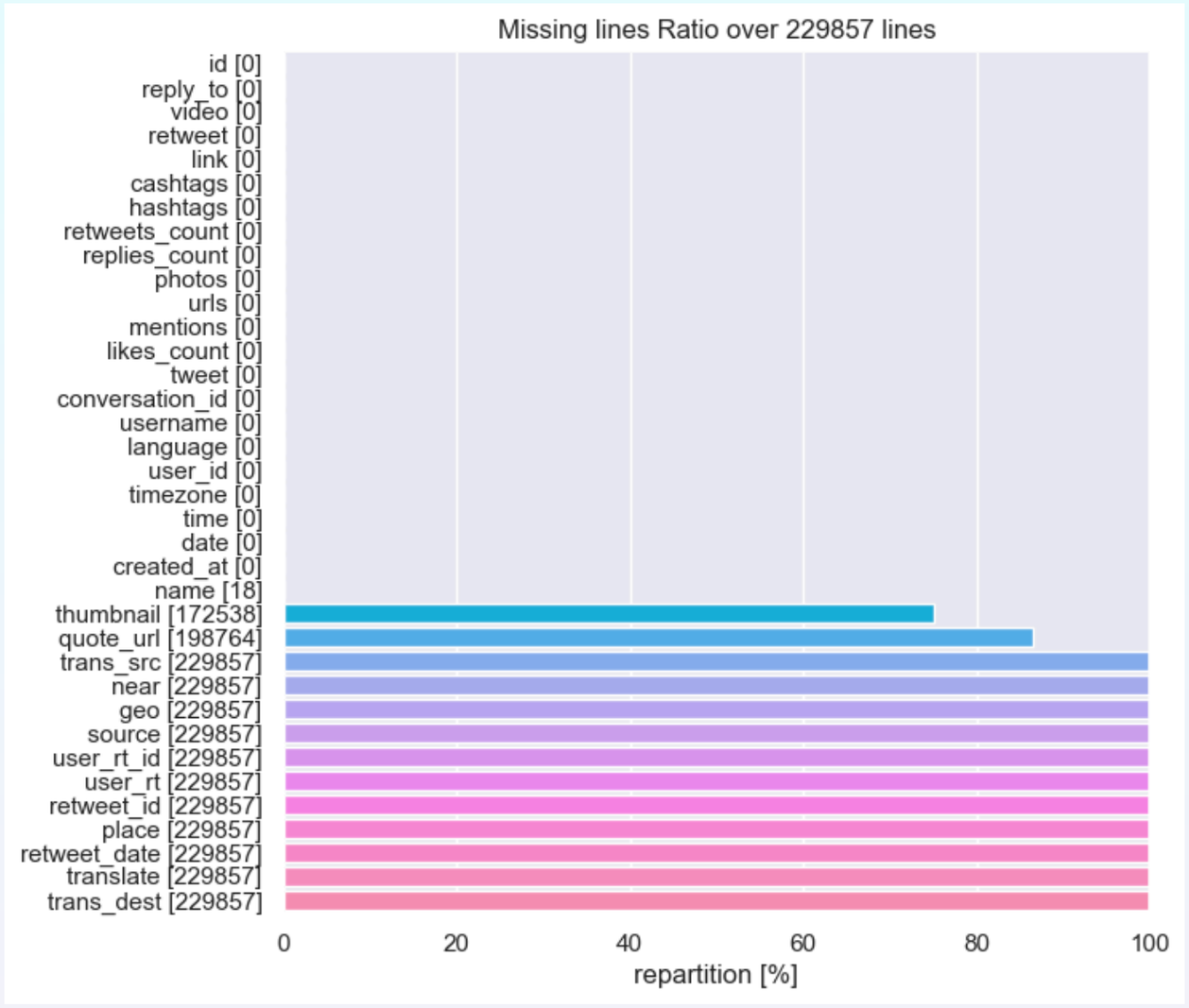
- Ocean Data Challenge
  - predict $OCEAN price with twitter sentiment
- Data used:
  - OCEAN token price dataset
  - OCEAN tweets dataset
  - not other data!

THE QUESTION AND ITS INTERPRETATION

- Find correlation with price and:
  - number of tweets
  - number of likes
  - number of retweets
  - number of individuals
  - number of influential tweets
- Model of twitter sentiment
  - Extract sentiment from tweet with a model to classify tweets
- Find correlation with price evolution

EXPLORATION : THE DATA : TWEETS

- keep columns with NaN less than 90%



229857 TWEETS

ELIMINATE

DATE UTC+3 != DATE PRICES
=> NEED TO SYNCHRONISE

EXPLORATION : THE DATA : PRICES

NB PRICES = 1024 DAYS

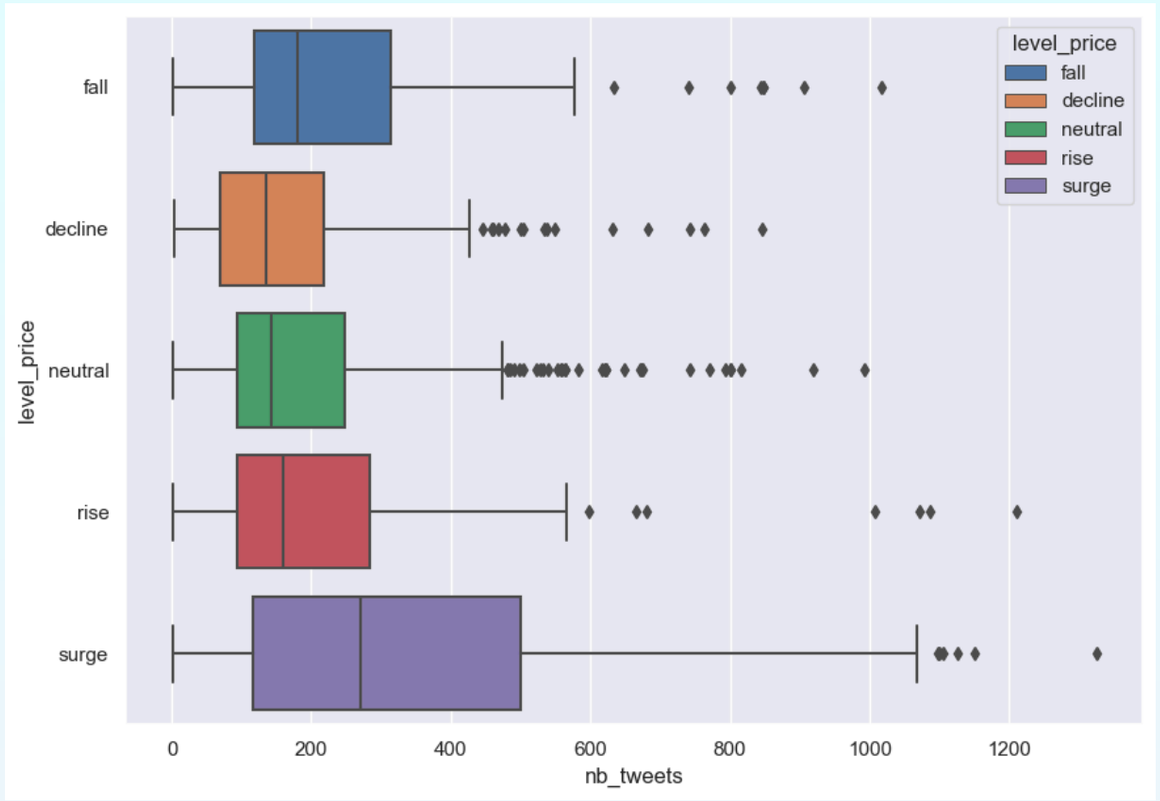
- Open / High / Low / Close / Volume : classical data
- Date reference : UTC (checked on Tradingview)
- Date range synchronise with tweets
- Add some simple insights
  - Use price variation : (Close - Open) / Open
  - Classification rules

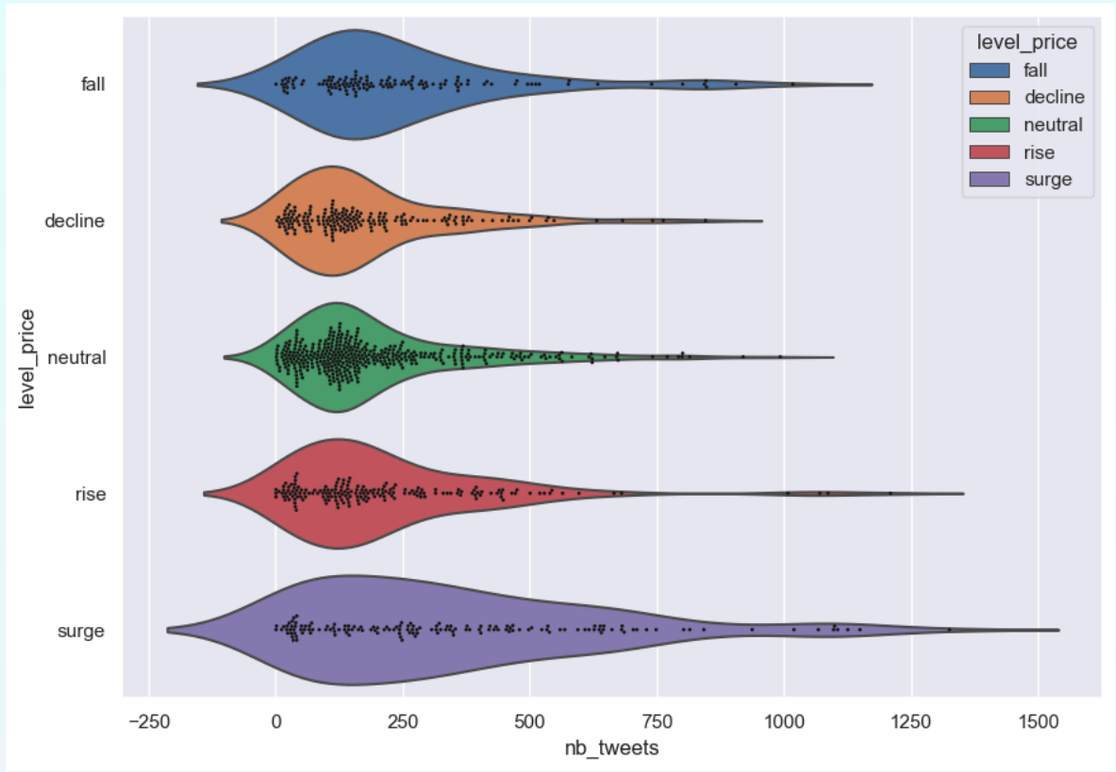
```
Levels of price variation : level_stat_price :
- neutral : var mean - 1 std < var price < var mean + 1 std
- bearish : var price < var mean - std
- bullish : var price > var mean + std
```

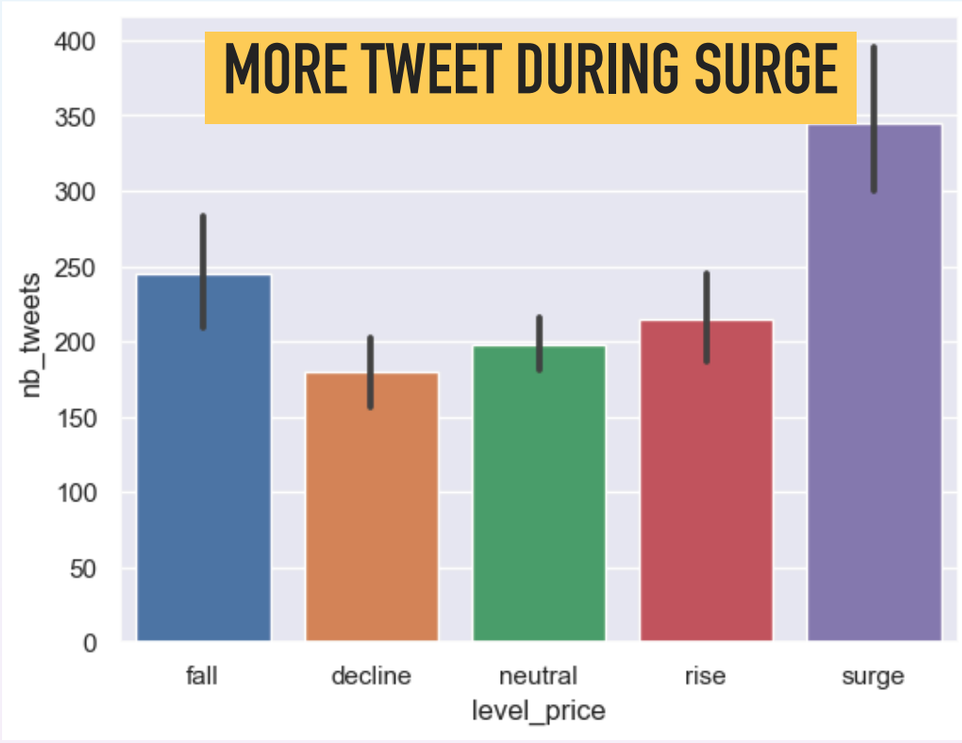
EXPLORATION : MERGE METHOD PRICES & TWEETS

- Choose to synchronize date for exploration
  - (no for last part : model prediction)
- Merge with daily prices :
  - **Aggregate each day**
    - with sum
      - number of tweets / likes / retweets / influential
    - with nunique
      - number of individuals
- For **influential**, select tweets with **more the 100 likes**

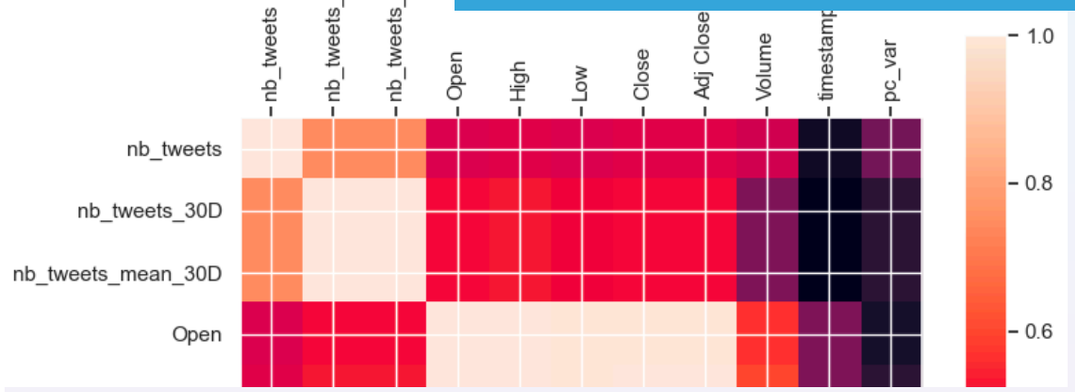
CORRELATION BETWEEN THE PRICE OF $OCEAN AND THE NUMBER OF TWEETS





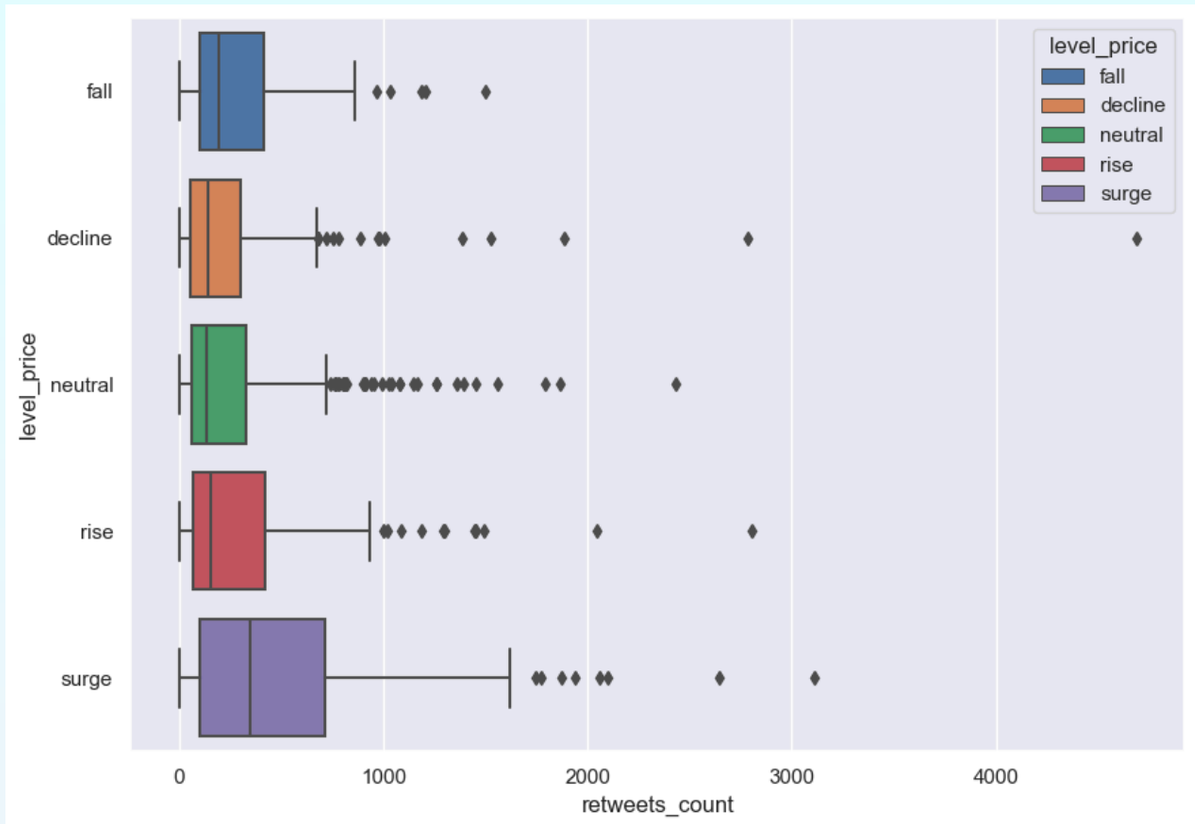


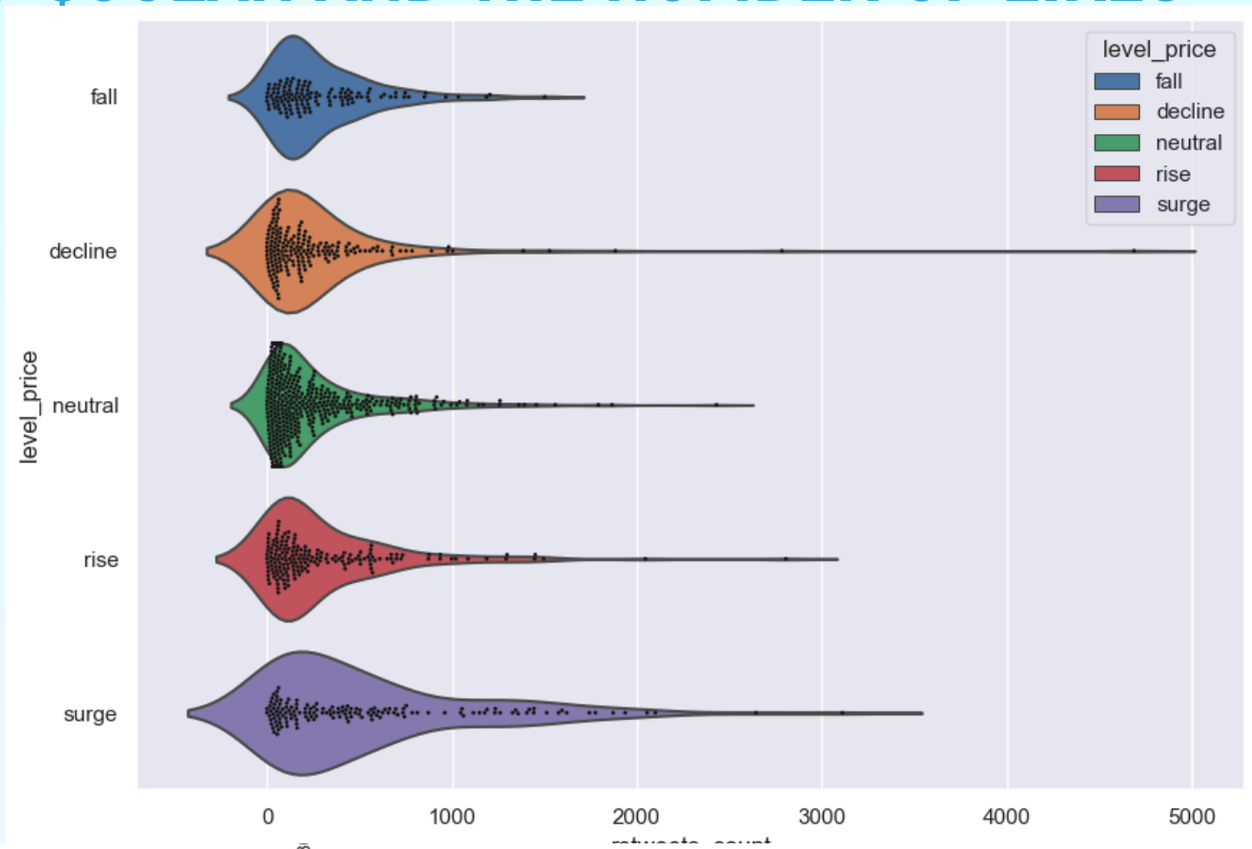
Corr coef nb_tweets vs Close: 0.44
Corr coef nb_tweets vs pc_var: 0.17

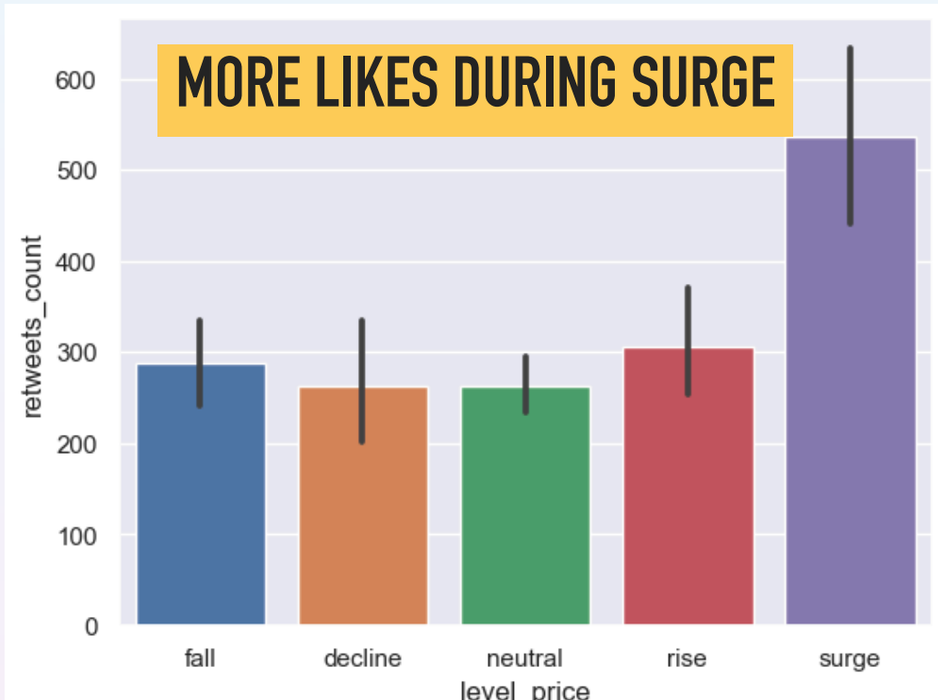


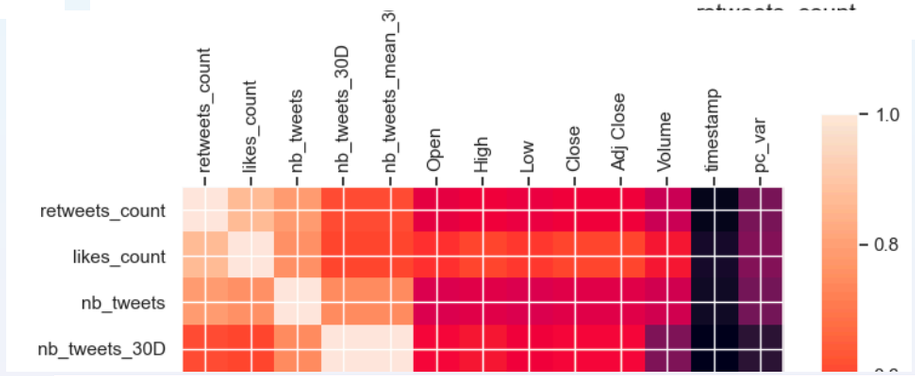
=> VERY LOW CORRELATION
MAYBE DURING SURGE

CORRELATION BETWEEN THE PRICE OF $OCEAN AND THE NUMBER OF LIKES









Corr coef nb likes vs Close: 0.59
Corr coef nb likes vs pc_var: 0.21

=> LOW CORRELATION
BUT MORE THAN SIMPLE NB TWWETSMAYBE DURING SURGE

CORRELATION BETWEEN THE PRICE OF $OCEAN

- Lower correlation with Nb influential tweets & nb individual

	Close	pc_var	Low	High
nb_influ	0.560370	0.201397	0.537332	0.559453
nb_indiv	0.499816	0.145142	0.476117	0.507835
retweets_count	0.491611	0.180578	0.473187	0.494742
likes_count	0.598143	0.211415	0.573586	0.599180
nb_tweets	0.446507	0.170824	0.421602	0.454216

=> NB LIKES = THE BEST INSIGHT BUT STILL LOW

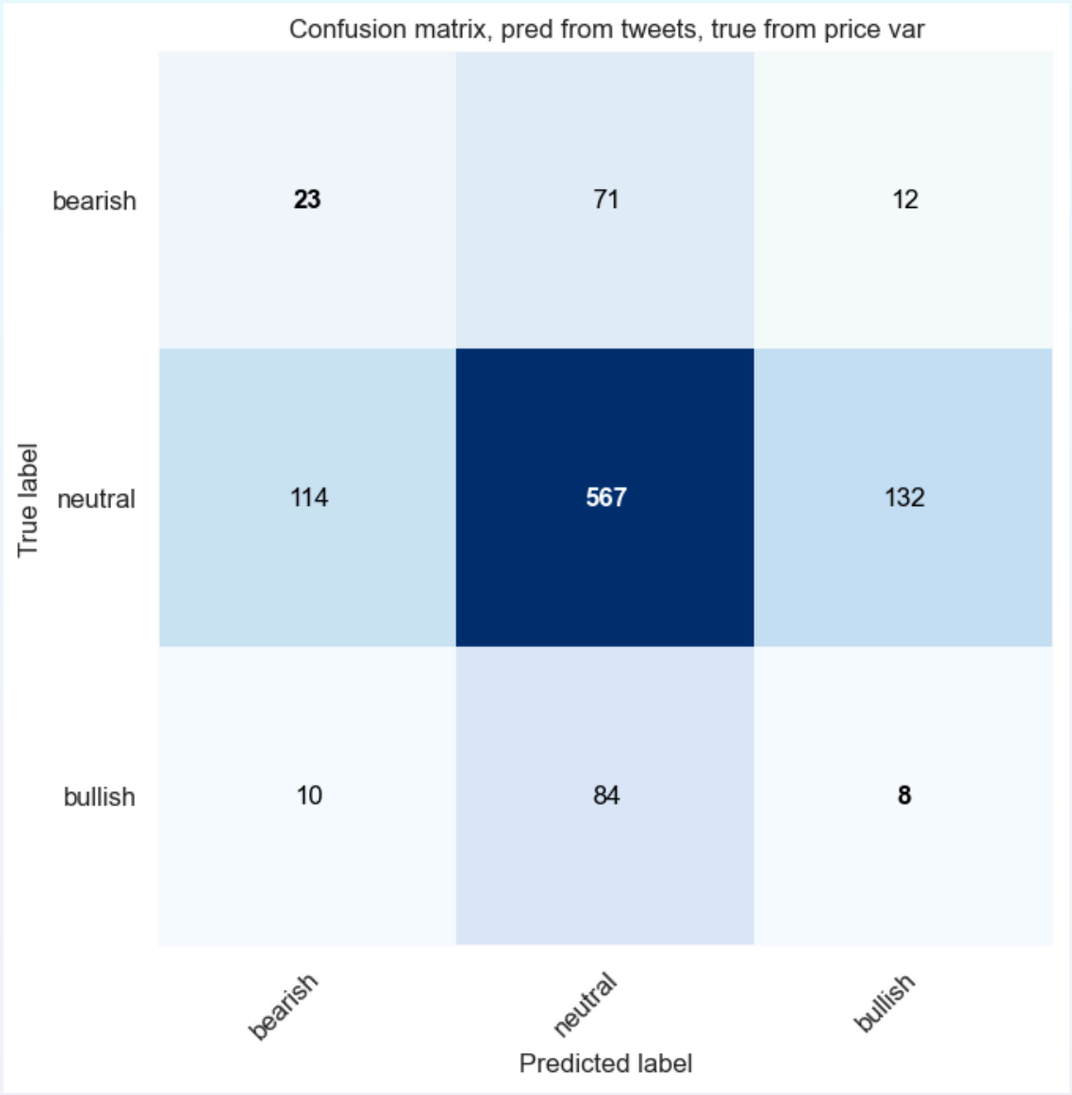
MODEL SENTIMENT: SIMPLE TEXTBLOB

- Use Textblob to extract sentiment in tweets
  - Desynchronise dates between Tweets and Prices
    - Tweets = 3 hours before

MODEL SENTIMENT: RESULTS

- Good to detect neutral
- Still bad for perish or bullish
  - predict neutral instead

CONFUSION MATRIX



MODEL SENTIMENT: RESULTS

- Sorry no time to finish !!!