PROJECT 1:

AN ANALYSIS OF
5 NASDAQ STOCKS
AND MEDIA SENTIMENT

TEAM CALABAR
APRIL 10, 2018

OUR TEAM

CALABAR PYTHON

KETTYL AMOAKON

ALEXA CAVENAS

AURA JOHNSON

ROBIN VASUDEVAN

BETH YODER

GUIDING QUESTIONS

1. IS IT POSSIBLE TO PREDICT STOCK MARKET BEHAVIOR?
2. WHAT FACTORS INFLUENCE SHARE COST FLUCTUATIONS?
3. DOES NEWS MEDIA COVERAGE INFLUENCE TRADING?
4. HOW CAN WE MEASURE NEWS MEDIA SENTIMENT?

HYPOTHESIS

A RELATIONSHIP EXISTS
BETWEEN STOCK PRICES
AND MEDIA COVERAGE OF THAT STOCK

OUR INITIAL RESEARCH FOCUSES ON ONLY
5 NASDAQ STOCK EXCHANGE COMPANIES
(CHOSEN ANECDOTALLY BASED ON POPULARITY)
AND 4 VARIABLES. DUE TO TIME CONSTRAINTS

PROGRESSION OF PROJECT

SATURDAY

BEGAN TO RESEARCH FIRST HYPOTHESIS USING INSTAGRAM AND GEOLOCATION

REALIZED INSTAGRAM API END POINTS WERE NOT SUFFICIENT

TUESDAY

NEW HYPOTHESIS

DECIDED TO USE TWEEPY FOR SENTIMENT ANALYSIS

THURSDAY

DISSATISFIED WITH RESULTS OF TWEEPY

RESEARCHED NEW DATA SOURCES

OBTAINED INSTITUTIONAL LICENSE FOR DATA SETS

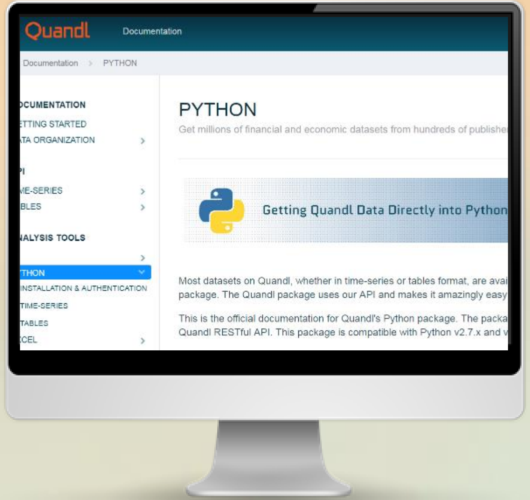
SATURDAY

DECIDED TO USE FIRST QUARTER 2018 DATA

DATA SOURCES

FINSENTS

.FINANCIAL DATA COMPANY



QUANDL API (PYTHON MODULE)

SOURCE OF FINANCIAL DATA

VARIABLES

01 CLOSING PRICE:
ADJUSTED CLOSING PRICE ACCOUNTS FOR EFFECTS OF STOCK PRICE (END OF DAY) THAT ARE CAUSED BY CORPORATE ACTIONS

QUANDL MODULE (QUANDL API FOR PYTHON) COMPILES THIS DATA DAILY

RELIABLE REAL-TIME DATA SOURCE FOR ANALYSIS OF STOCK PRICES

VARIABLES

02

SENTIMENT SCORE:
A MEASURE OF THE BULLISHNESS / BEARISHNESS OF THE LANGUAGE USED IN MEDIA COVERAGE OF A GIVEN STOCK ON A GIVEN DAY

RANGES FROM -5 (EXTREMELY NEGATIVE COVERAGE) TO +5 (EXTREMELY POSITIVE COVERAGE):
A SCORE OF 0 INDICATES AN ABSENCE OF ARTICLES FOR THAT DAY

FINSENTS ALGORITHM DETERMINES THIS NUMBER

BULL VS. BEAR MARKET

BEAR MARKET LANGUAGE IS RELATED TO FALLING PRICE TRENDS, PESSIMISM, AND OVERALL NEGATIVE INDICATORS (STOCK MARKETS, UNEMPLOYMENT, ETC) WHILE BULLISH SENTIMENT RATINGS WOULD MEAN THAT LANGUAGE USED IMPLIED UPWARD PRICE TRENDS, OPTIMISM, AND GROWTH

VARIABLES

03 NEWS VOLUME:
THE NUMBER OF NEWS ARTICLES *ABOUT THIS STOCK* PUBLISHED ON A GIVEN DAY

COMPANY NEWS EXCLUDED

FINSENTS METRIC

VARIABLES

04

NEWS BUZZ:
A MEASURE OF THE RATE OF CHANGE IN NEWS COVERAGE OF A GIVEN STOCK ON A GIVEN DAY

NORMALIZED ON A SCALE OF 1 TO 10

MEASURES THE CHANGE IN THE STANDARD DEVIATION OF PERIODIC NEWS VOLUME

CALCULATED BY FINSENTS

SERVES AS A 'RISK ALERT' INDICATOR

JUPYTER NOTEBOOK I

tuesday_newest_stock_data Last Checkpoint: a few seconds ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Trusted

Code

In [1]:

```
# Import dependencies
import pandas as pd
from scipy.misc import imread
import arrow
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from datetime import datetime
import quandl
quandl.ApiConfig.api_key = '****************************'
```

DATA ORGANIZATION

```python
# Make calls to the quandl API to get the stock data for our five chosen stocks, for only the variabl
all_stock = quandl.get_table('WIKI/PRICES', ticker=["AAPL", "AMZN", "FB", "SBUX", "TWTR"],
                             qopts={"columns":["date", "ticker","adj_open","adj_close","adj_volume"]}
                             date = {'gte': '2018-01-01', 'lte': '2018-03-31'}, paginate=True)

# Rename date column to make useful for future merges
all_stock = all_stock.rename(columns={'date':'Date'})
all_stock.head()
```

None	Date	ticker	adj_open	adj_close	adj_volume
0	2018-01-02	AAPL	170.16	172.26	25048048.0
1	2018-01-03	AAPL	172.53	172.23	28819653.0
2	2018-01-04	AAPL	172.54	173.03	22211345.0
3	2018-01-05	AAPL	173.44	175.00	23016177.0
4	2018-01-08	AAPL	174.35	174.35	20134092.0

```python
# Make calls using the FinSents API wrapper (through quandl) for our 5 chosen companies on the NASDAQ and put into sent
all_sent = quandl.get(['NS1/AAPL_US', 'NS1/AMZN_US', 'NS1/FB_US',  'NS1/SBUX_US', 'NS1/TWTR_US'], start_date='2018-01-0

# Reset index to make the date column usable for later merges
all_sent = all_sent.reset_index()
all_sent.head()
```

	Date	NS1/AAPL_US - Sentiment	NS1/AAPL_US - Sentiment High	NS1/AAPL_US - Sentiment Low	NS1/AAPL_US - News Volume	NS1/AAPL_US - News Buzz	NS1/AMZN_US - Sentiment	NS1/AMZN_US - Sentiment High	NS1/AMZN_US - Sentiment Low	NS1/AMZN_US - News Volume	...
0	2018-01-01	4.0	5.0	4.0	19.0	2.0	0.0	0.0	0.0	0.0	...
1	2018-01-02	4.0	5.0	3.0	49.0	3.0	0.0	0.0	0.0	0.0	...
2	2018-01-03	4.0	5.0	4.0	41.0	3.0	0.0	0.0	0.0	0.0	...
3	2018-01-04	4.0	5.0	4.0	21.0	2.0	0.0	0.0	0.0	0.0	...
4	2018-01-05	4.0	5.0	4.0	64.0	9.0	0.0	0.0	0.0	0.0	...

In [6]:

```python
# Rename columns for readability
all_sent_rename = all_sent.rename(index=str, columns={"NS1/AAPL_US - Sentiment": "AAPL Sentiment",
                                  "NS1/AAPL_US - News Volume": "AAPL News Volume",
                                  "NS1/AAPL_US - News Buzz": "AAPL News Buzz",
                                  "NS1/AMZN_US - Sentiment": "AMZN Sentiment",
                                  "NS1/AMZN_US - News Volume": "AMZN News Volume",
                                  "NS1/AMZN_US - News Buzz": "AMZN News Buzz",
                                  "NS1/FB_US - Sentiment": "FB Sentiment",
                                  "NS1/FB_US - News Volume": "FB News Volume",
                                  "NS1/FB_US - News Buzz": "FB News Buzz",
                                  "NS1/SBUX_US - Sentiment": "SBUX Sentiment",
                                  "NS1/SBUX_US - News Volume": "SBUX News Volume",
                                  "NS1/SBUX_US - News Buzz": "SBUX News Buzz",
                                  "NS1/TWTR_US - Sentiment": "TWTR Sentiment",
                                  "NS1/TWTR_US - News Volume": "TWTR News Volume",
                                  "NS1/TWTR_US - News Buzz": "TWTR News Buzz"})
all_sent_rename.head()
```

Out[6]:

	Date	AAPL Sentiment	AAPL News Volume	AAPL News Buzz	AMZN Sentiment	AMZN News Volume	AMZN News Buzz	FB Sentiment	FB News Volume	FB News Buzz	SBUX Sentiment	SBUX News Volume	SBUX News Buzz	TWTR Sentiment	TWTR News Volume
0	2018-01-01	4.0	19.0	2.0	0.0	0.0	0.0	4.0	3.0	4.0	3.0	6.0	5.0	4.0	156.0
1	2018-01-02	4.0	49.0	3.0	0.0	0.0	0.0	4.0	24.0	7.0	4.0	3.0	1.0	4.0	151.0

INITIAL DATAFRAMES WERE CREATED FROM CALLS TO EACH API/MODULE

DATA WAS ORGANIZED AND CLEANED

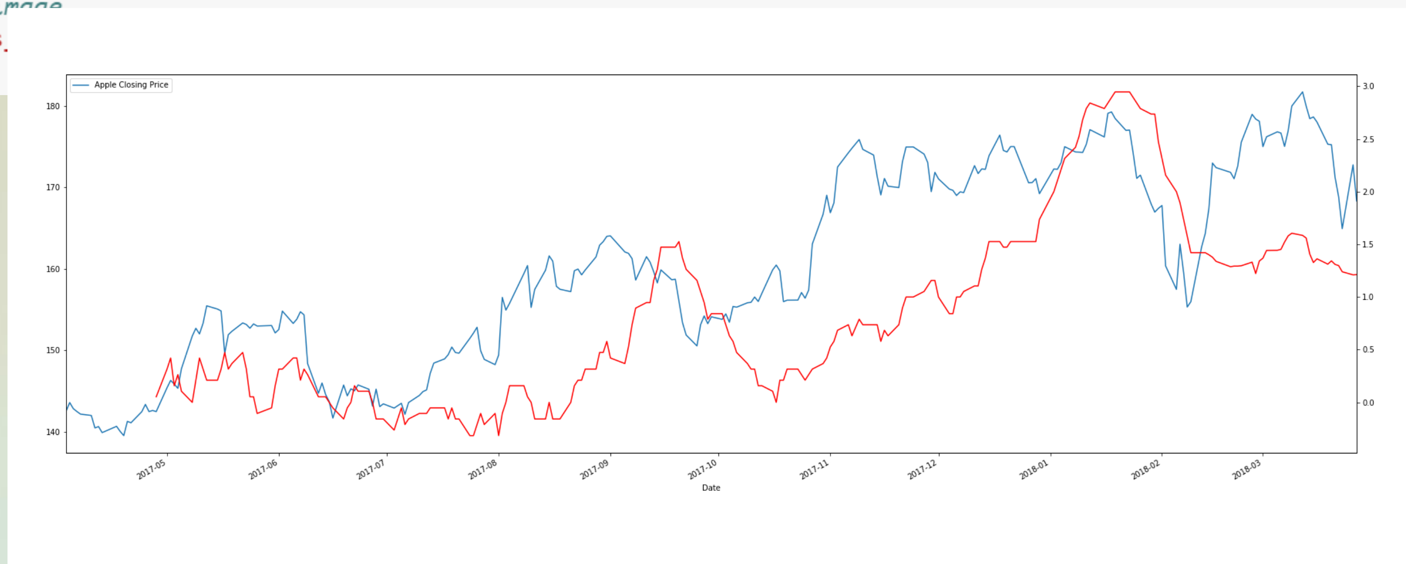
DATA ANALYSIS

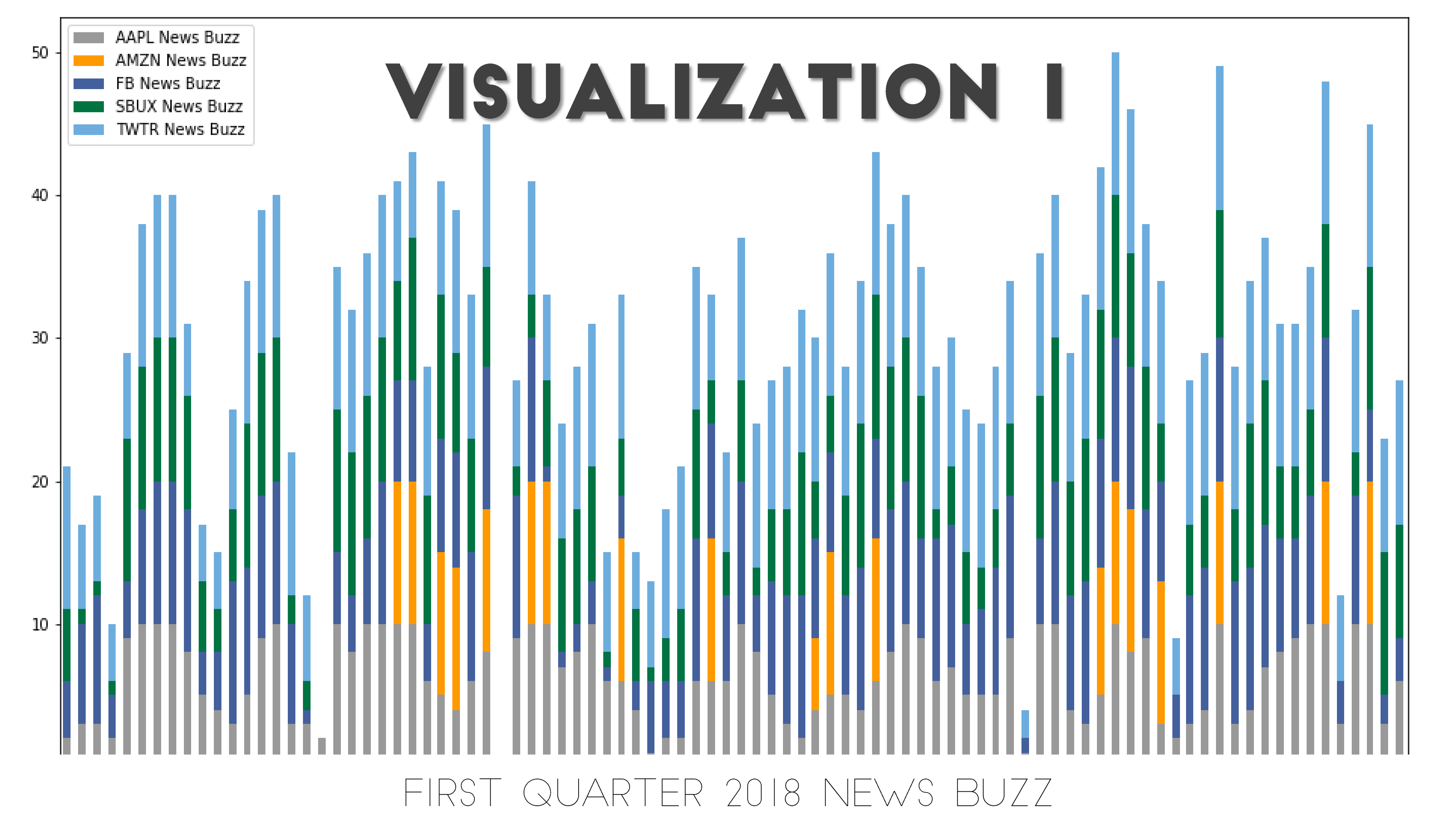
```
#Run Additional Stats on Apple
model = LinearRegression().fit(Apple_merged[['AAPL_Sentiment']], Apple_merged[['adj_close']]
m = model.coef_[0]
b = model.intercept_
#equation of the line
print("Equation of the line: Y =
result = sm.ols(formula="adj_clo
print(result.params)
print(result.summary())
```

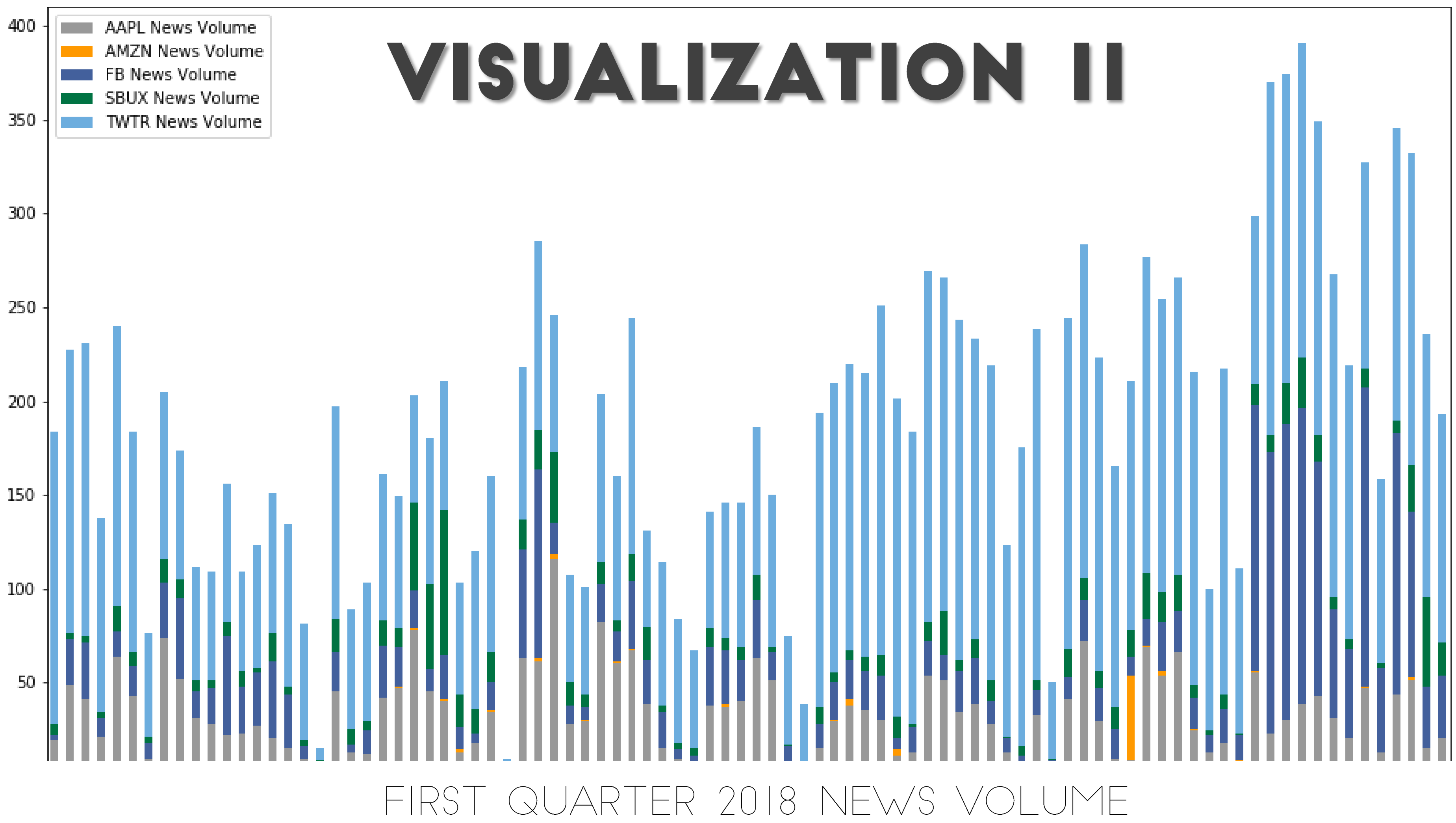
```
Intercept        157.245575
AAPL_Sentiment     3.644333
dtype: float64
                            OLS Regression Results
==============================================================================
Dep. Variable:              adj_close   R-squared:                       0.228
Model:                            OLS   Adj. R-squared:                  0.225
Method:                 Least Squares   F-statistic:                     72.22
Date:                Tue, 10 Apr 2018   Prob (F-statistic):           1.93e-15
Time:                        02:48:55   Log-Likelihood:                -923.17
No. Observations:                 246   AIC:                             1850.
Df Residuals:                     244   BIC:                             1857.
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept        157.2456      0.742    212.017      0.000     155.785     158.706
AAPL_Sentiment     3.6443      0.429      8.498      0.000       2.800       4.489
==============================================================================
Omnibus:                        4.652   Durbin-Watson:                   0.378
Prob(Omnibus):                  0.098   Jarque-Bera (JB):                3.483
Skew:                          -0.161   Prob(JB):                        0.175
Kurtosis:                       2.514   Cond. No.                         2.12
==============================================================================
```

DATA ANALYSIS II

```
# Apple Sentiment analysis
fig, ax = plt.subplots()
ax2 = ax.twinx()
plt.hold(False)
Apple_merged["adj_close"].plot(ax=ax,kind="line",figsize=(20,10), label='Apple Closing Price')
Apple_merged["Rolling_mean_19"].plot(ax=ax2, style='r-', secondary_y=True, figsize=(30,10), label='Sentiment Index')
ax.legend(loc='best')
# Save the figure as png image
plt.savefig("apple_sent_vs
plt.show()
```

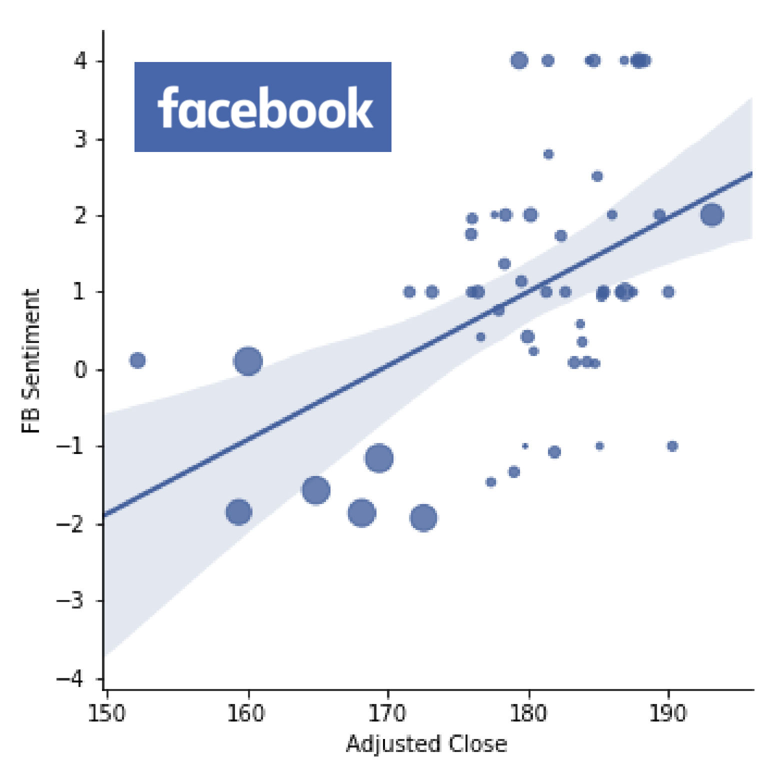



AAPL News Buzz
AMZN News Buzz
FB News Buzz
SBUX News Buzz
TWTR News Buzz
VISUALIZATION I
50
40
30
20
10
FIRST QUARTER 2018 NEWS BUZZ



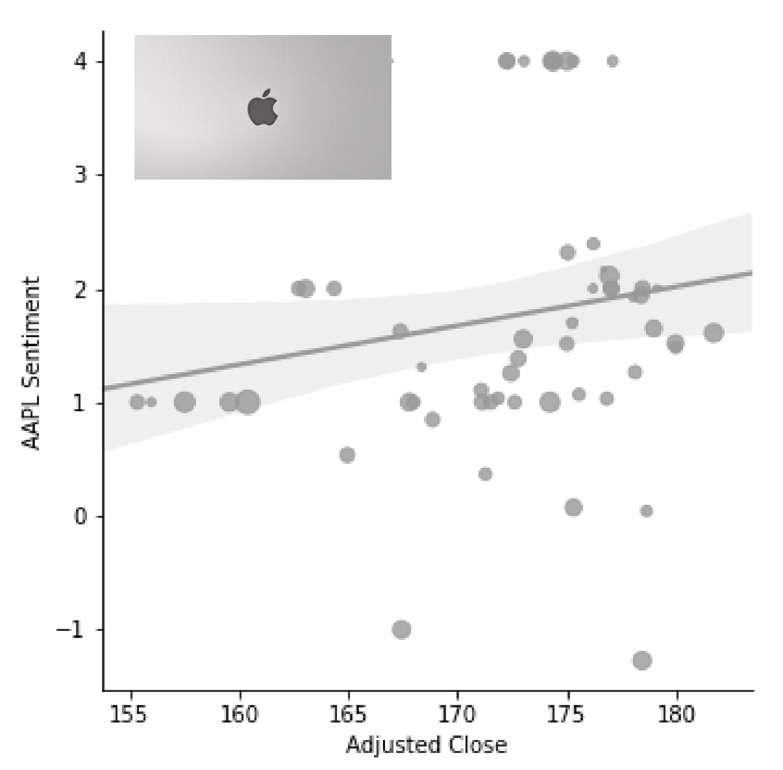
VISUALIZATION II
AAPL News Volume
AMZN News Volume
FB News Volume
SBUX News Volume
TWTR News Volume
400
350
300
250
200
150
100
50
FIRST QUARTER 2018 NEWS VOLUME

VISUALIZATIONS



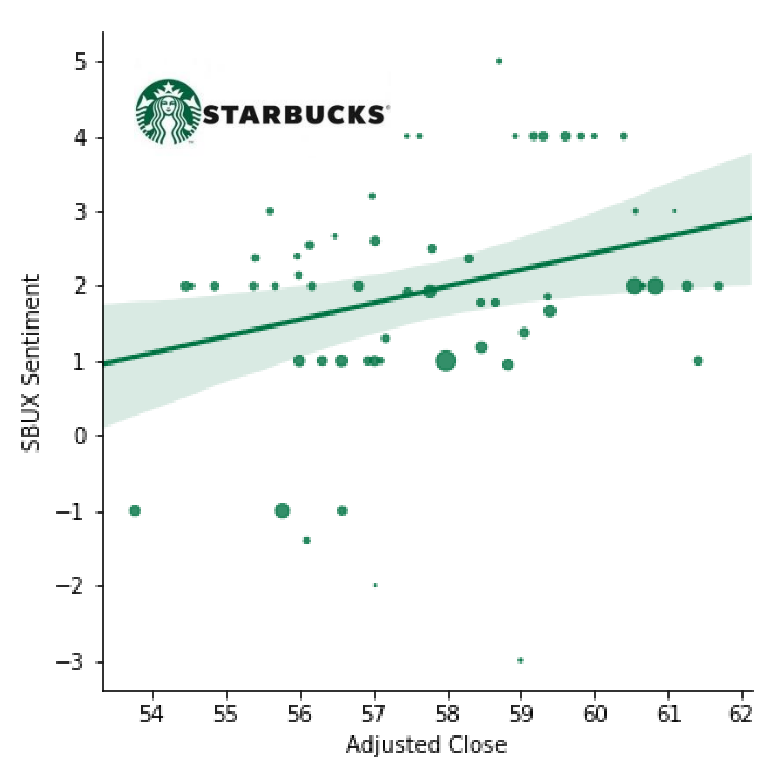
FIRST QUARTER 2018: FACEBOOK CLOSING PRICE VS STOCK SENTIMENT, AS A FUNCTION OF NEWS VOLUME

VISUALIZATIONS CONTINUED



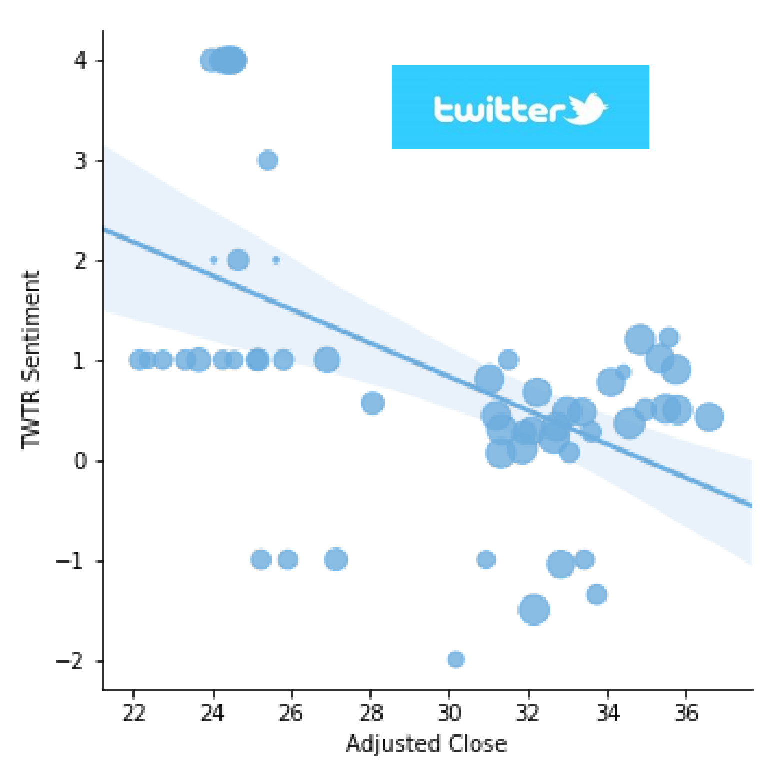
FIRST QUARTER 2018: APPLE CLOSING PRICE VS STOCK SENTIMENT, AS A FUNCTION OF NEWS VOLUME

VISUALIZATIONS CONTINUED



FIRST QUARTER 2018: STARBUCKS CLOSING PRICE VS STOCK SENTIMENT, AS A FUNCTION OF NEWS VOLUME

VISUALIZATIONS CONTINUED



FIRST QUARTER 2018: TWITTER CLOSING PRICE VS STOCK SENTIMENT, AS A FUNCTION OF NEWS VOLUME

RESULTS

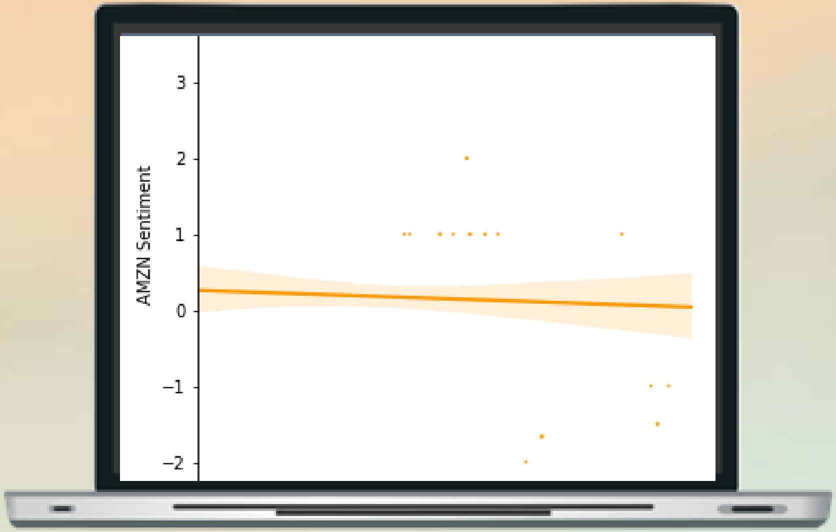
APPLE HAD THE HIGHEST R SQUARE VALUE (.23), INDICATING THE HIGHEST PREDICTIVE RELATIONSHIP BETWEEN ADJUSTED CLOSING PRICE AND MEDIA SENTIMENT

HOWEVER, OUR HYPOTHESIS IS NOT CONFIRMED BECAUSE NEWS SENTIMENT IS NOT THE BEST INDICATOR OF PRICE FLUCTUATIONS

NEWS VOLUME IN CONJUNCTION WITH SENTIMENT IS A BETTER INDICATOR, BUT MOST FLUCTUATIONS ARE DUE TO FACTORS OTHER THAN OUR CHOSEN VARIABLES

LIMITATIONS

- TIME: OVERALL PROJECT TIME TOO SHORT
- SCOPE CREEP: PROJECT GREW BEFORE FINAL FOCUS WAS AGREED UPON
- GROUP WORK: DIFFICULT TO DIVIDE WORK EFFICIENTLY: SOME WORK DUPLICATED NEEDLESSLY
- SAMPLE DATA: SET TOO SMALL. DATA MISSING FOR AMAZON
- STATISTICS: UNDERSTANDING WAS INSUFFICIENT TO CONFIDENTLY USE STATISTICAL TOOLS



SUGGESTIONS FOR FUTURE RESEARCH

RUN THE SAME ANALYSIS ON ADDITIONAL STOCKS, LIKE THE NASDAQ 100

LOOK INTO HOW TO COMPARE THE PREVIOUS DAY'S SENTIMENT (OR RECENT PAST SENTIMENT) WITH PRICE, AS THERE WOULD BE SOME INTRA-DAY TRADING BASED ON NEWS, BUT SAME DAY CLOSING PRICE MIGHT NOT BE THE BEST INDICATOR

RESEARCH SPECIFIC HIGH VOLUME NEWS DAYS TO EXTRACT MEANING

REFERENCES

HTTP://FINSENTS.COM/NEWSHOME/HELP

HTTPS://DOCS.QUANDL.COM/DOCS/PYTHON

HTTPS://WWW.INVESTOPEDIA.COM/TERMS/M/MARKETSENTIMENT.ASP

HTTP://BLOGS.LSE.AC.UK/USAPPBLOG/2017/10/14/CAN-TWITTER-SENTIMENT-PREDICT-STOCK-MARKET-BEHAVIOUR/

QUESTIONS?

HI DZMITRY