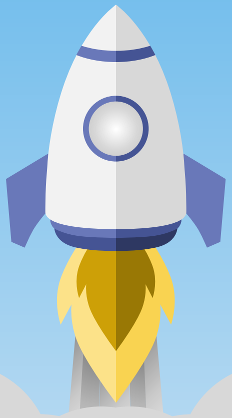


# PARADIGM + DATA-X @ BERKELEY

A project geared towards determining the effect of news articles on crypto-currency prices

# What is Paradigm?

Paradigm provides a conversational marketplace for the crypto trading market. It features AI driven tools like automated over-the-counter trading and  provides a sophisticated chat tool for institutional traders.

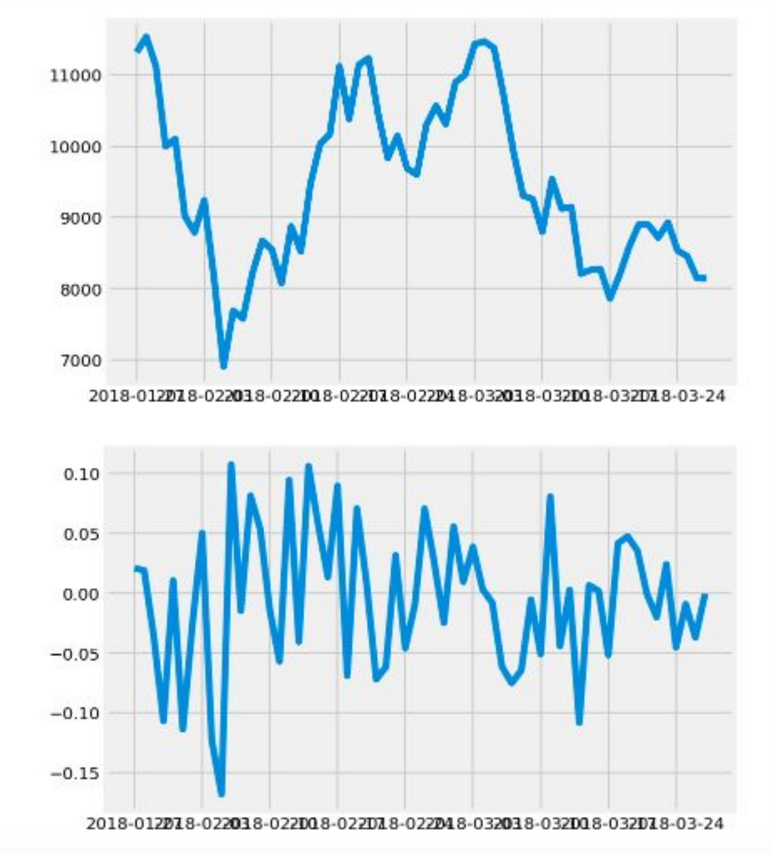
The chat is a combination of several trader tools including counter party discovery,  text-to-trade recognition, etc.

# The Project Task

Predict whether a news article will have a significant impact on the underlying crypto-asset mentioned in the article.

Details: 1) Event Detection. (2) Quantify significance of event on crypto prices. (3) Rank importance/influence of article on event. (4) Visually display this importance.

# Data Sample


11000
10000
9000
8000
7000
2018-03-24
0.10
0.05
0.00
−0.05
−0.10
−0.15
2018-03-24

# Hypothesis-driven EDA

**H1: Strong events (in terms of price change) which are caused by news exist and are common**

A strong event caused by news should be easily identifiable since the lag between news and events is expected to be short.

**H1 must be rejected**

For 1 min , 15 min and 60 min data, one cannot establish a relationship between single news and strong events. The causal impact of a SINGLE article on news does not exist for strong events. Given that it exists, it is very hard to identify.

**Decisions:**
**move from intra-day data to daily data**
**investigate impact of the mass of articles ("sentiment") on price**

# Hypothesis-driven EDA

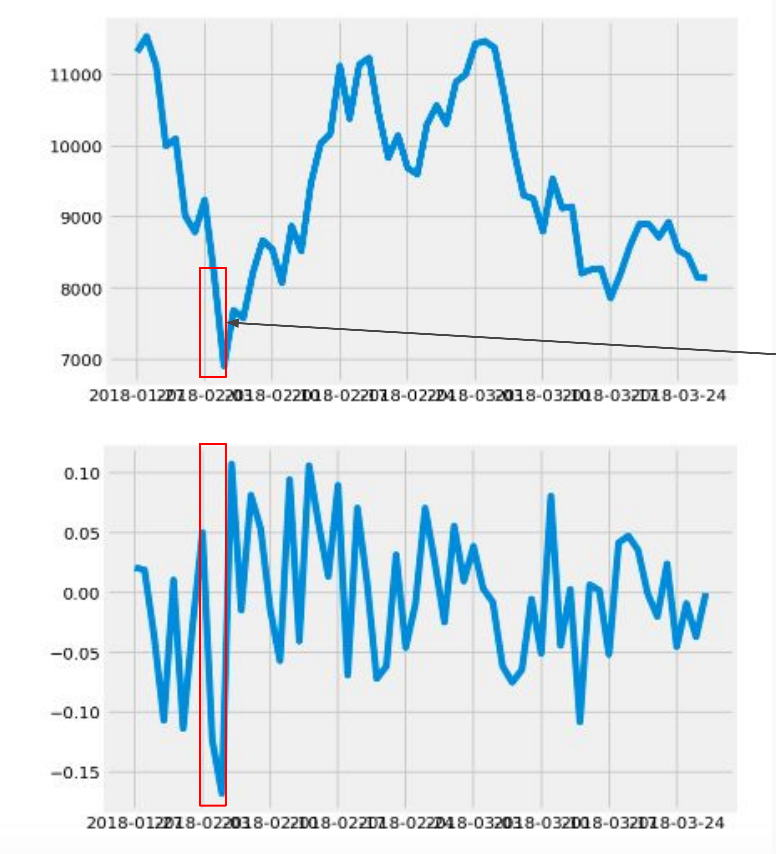
**H2: News reflect overall sentiment of the market**

Majority of news rather reflect past price movements or repeating other news, possibly building up expectations on future price movements ("sentiment")

**Our data sample has both a big upward and downward trend for which clear sentiment can be identified.**

E.g., when the market is falling for 20 days straight, news consistently report "Bitcoin bubble has burst, etc."

# Negative Sentiment




```
3513    "It's hard to put in words how life-changing b...
3514    As Regulators Tighten The Noose, Cryptocurrenc...
3515    Build a Cryptocurrency Price Tracker in 5 Minutes
3516     The Objective-Realist vs. the Crypto-Evangelist.
3517    Bitcoin set for worst week since 2013 as crypt...
3518         Could Price Manipulation Be Killing Bitcoin?
3519    Bitcoin Is the ‘Biggest Bubble in Human Histor...
3520    Unfazed by cryptocurrency crash, miners flocks...
3521             Initial coin offerings, explained - CNET
3522    Bitcoin Price Hits Floor, Bounces Back Above $...
3523    Bitcoin Investors Have Lost Nearly $87 Billion...
3524    Bitcoin falls below $8,000 as value plummets f...
3525                    Smell of doubt in Bitcoins online
3526    Bill Gates once stayed up until 4am to write a...
3527    Bitcoin set for worst week since 2013 as crypt...
3528    8 Reasons Bitcoin Has Lost $175 Billion in Mar...
3529           Here's Why Bitcoin Is Plunging Again Today
3530    Bitcoin value plunges after misleading report ...
3531    Bitcoin's huge arbitrage play just vanished as...
3532    Espionage malware snoops for passwords, mines ...
3533    Fileless WannaMine Cryptojacking Malware Using...
3534    Ex-ice tea maker Long Blockchain backs off bit...
3535              Weekly Threat Report 2nd February 2018
3536    19-year-old bitcoin millionaire: You should in...
3537    Bitcoin Has Worst Week Since 2013, Briefly Tum...
Name: title, Length: 84, dtype: object
```

# Hypothesis-driven EDA

**H3: Since H2 = true, there might be a causal impact of sentiment on price (e.g. "self-fulfilling prophecy")**

Sentiment reflects both past price movements but might also be an indicator of future price movement -> e.g. Self-fulfilling prophecy.

Given that overall sentiment is indicative of future price movements, we could label those articles which share the sentiment to be important.
E.g. negative articles will be important when overall sentiment is negative.
We can make a scoring if we can determine how "negative" an article is.

# Decision on next steps

"Event-based approach": try to label news articles (significant vs. not) and then predict the label via NLP

Major advantage: Given that we have labels, we can accurately predict the label (see last year's solution).

Major concern: it is extremely hard to label data without introducing enormous bias.

"Sentiment-based approach": consider an article to be important if it shares the overall sentiment.

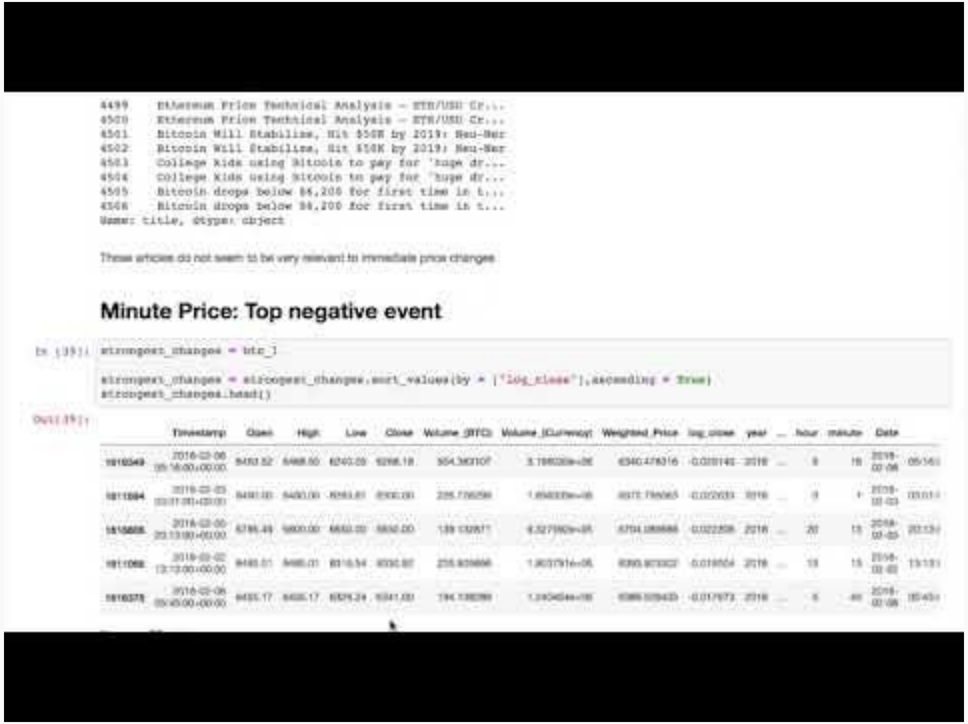
Major advantage: we avoid labeling data and possibly establish a causal relation between sentiment and price.

Major concern: no variance in importance of articles. All articles might be important if sentiment persists for a long time.

both approaches involve extensive NLP, so we can start off with NLP and postpone final decision.

# How do we ensure reliability and accuracy of our process and pipeline?

# Baseline Modeling: Codebase Demo




Link to Youtube Video Presentation

# Workflow of ML Approach

## RNN Model: Predict sentiment of articles

•Built an RNN model from scratch
•Used NLP to convert articles into input

OUTPUT:
•Articles defined as having a predicted sentiment of positive or negative
•A metric that represents the relation of the article to crypto-currency price from [0,1]
•Model has **70% accuracy**

## RNN Model: Predict sentiment of articles

•Leverage the RNN model to label articles
•Added features to define relationship between the sentiment and impact

OUTPUT:
•Added features relating to the four quadrants below:

| | | Impact on Price | |
|---|---|---|---|
| | | Increase | Decrease |
| Sentiment | Positive | | |
| | Negative | | |

## RNN Model: Predict sentiment of articles

NEXT STEPS:
•Use all article features and engineered features in ML model
•Use traditional machine learning to predict impact on price
•Try different models (Logistic Regression, Random Forest etc.)

OUTPUT:
•Predicted degree of impact of an article on crypto-currency price