



TokenSense

Keep on the crypto pulse

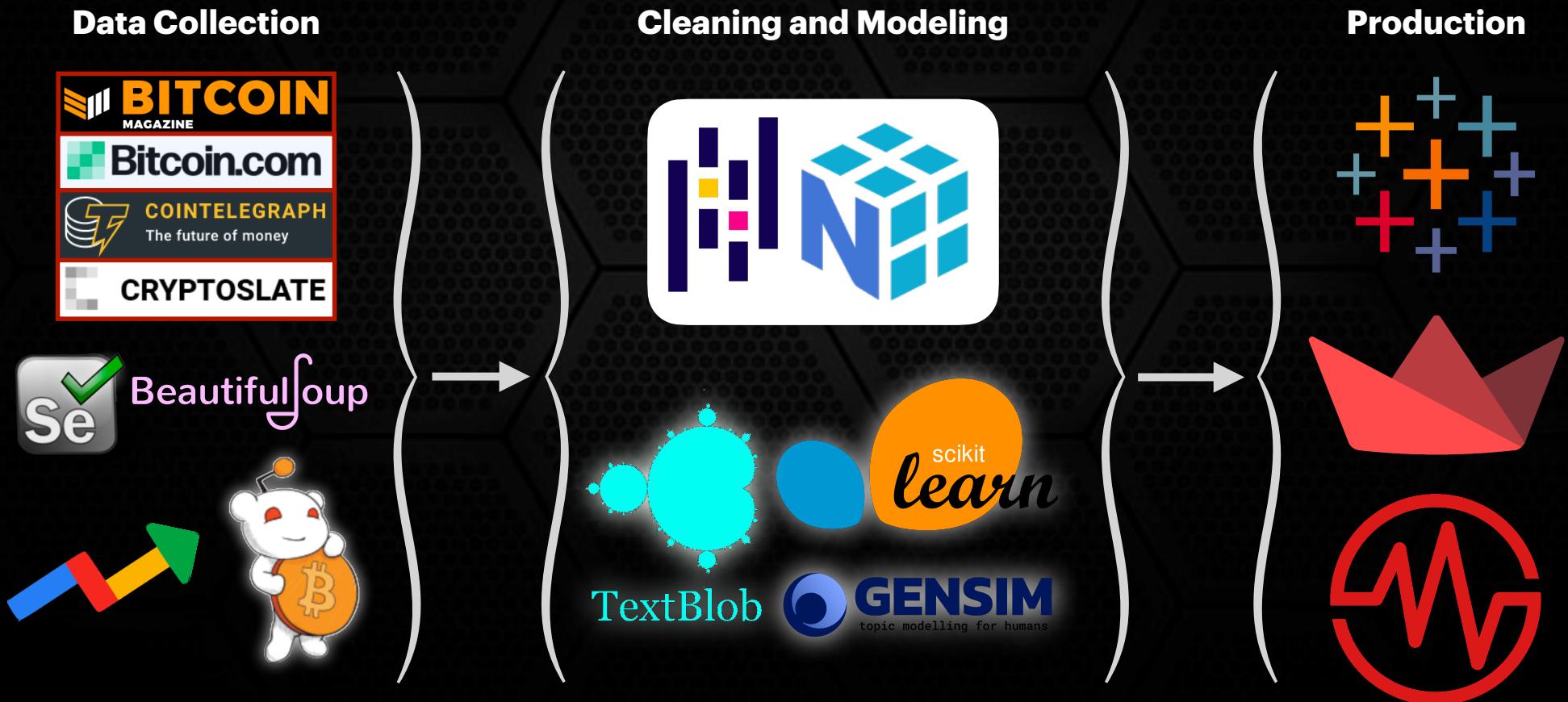
David Wismer — Metis 2021

What is TokenSense?

- ⌚ TokenSense is a web application tracking the pulse of the Bitcoin market, with two primary functions:
 - ⌚ Clustering dates by **Sentiment** and **Activity** metrics
 - ⌚ **Topic Modeling** prevalence and sentiment over time



Data Pipeline

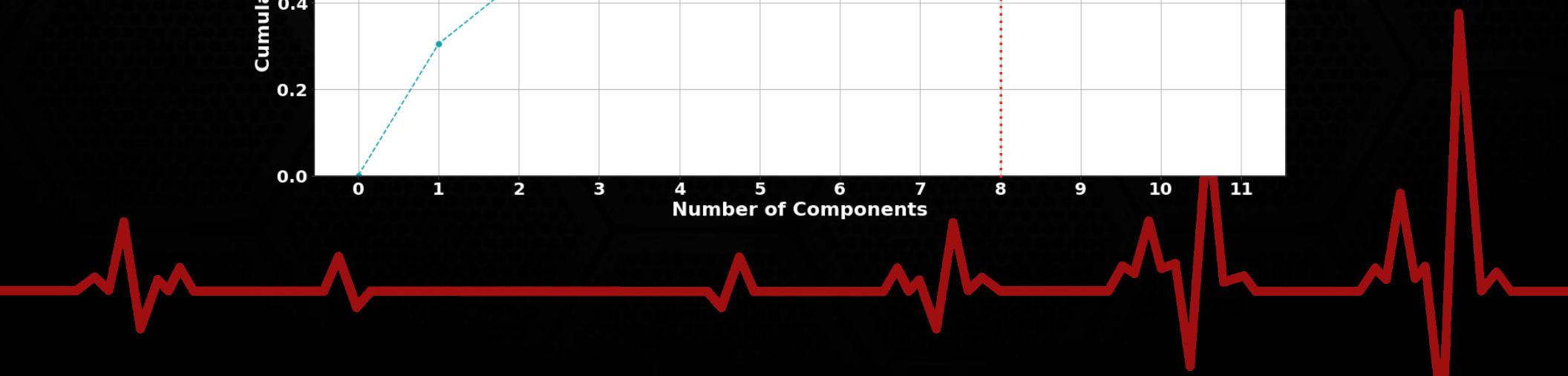
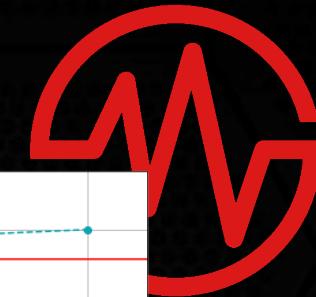
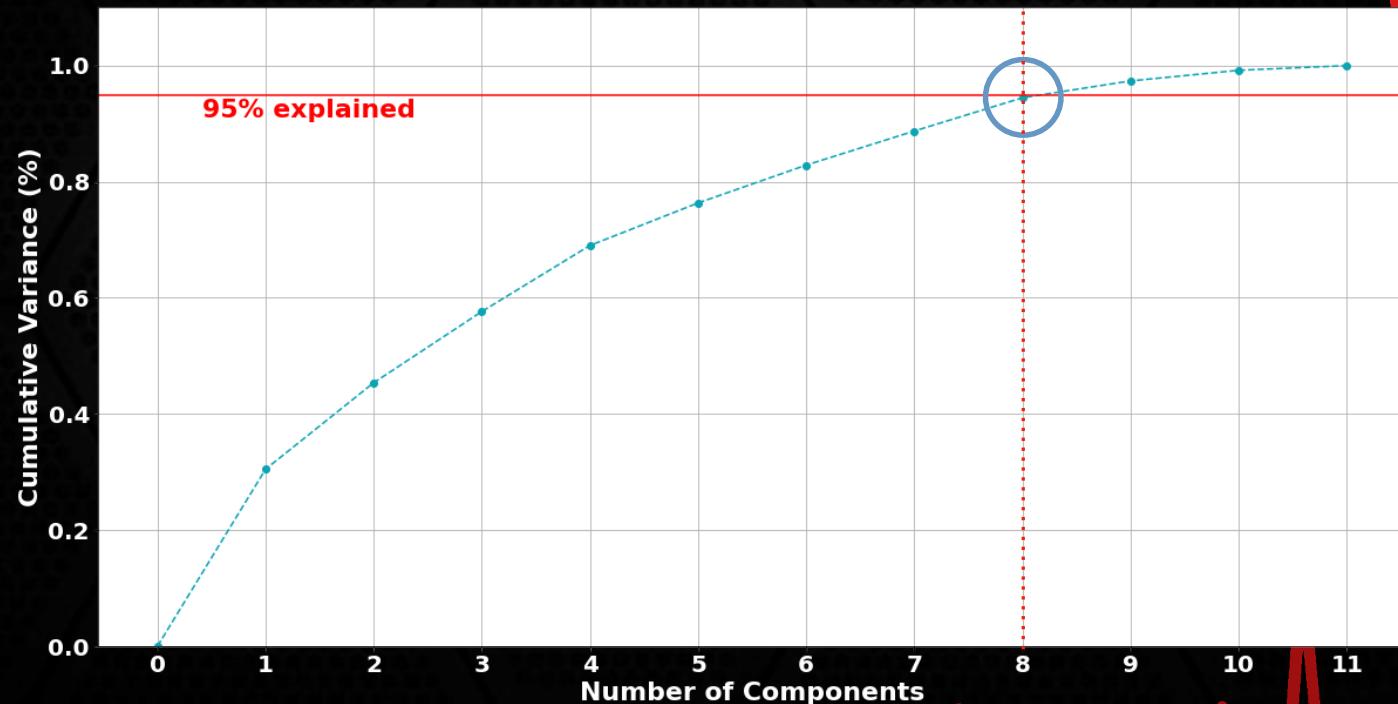


Sentiment Clusters: Feature Engineering

- ⌚ Reddit Comments >>> TextBlob >>> **Polarity and Subjectivity**
- ⌚ News Articles >>> TextBlob >>> **Polarity and Subjectivity**
- ⌚ Google Trends >>> Custom >>> **Activity vs. All-Time Median**
- ⌚ Trading Info >>> Custom >>> **Metric vs. All-Time High**
- ⌚ On-Chain Metrics >>> Unaltered >> **NUPL, NTV**

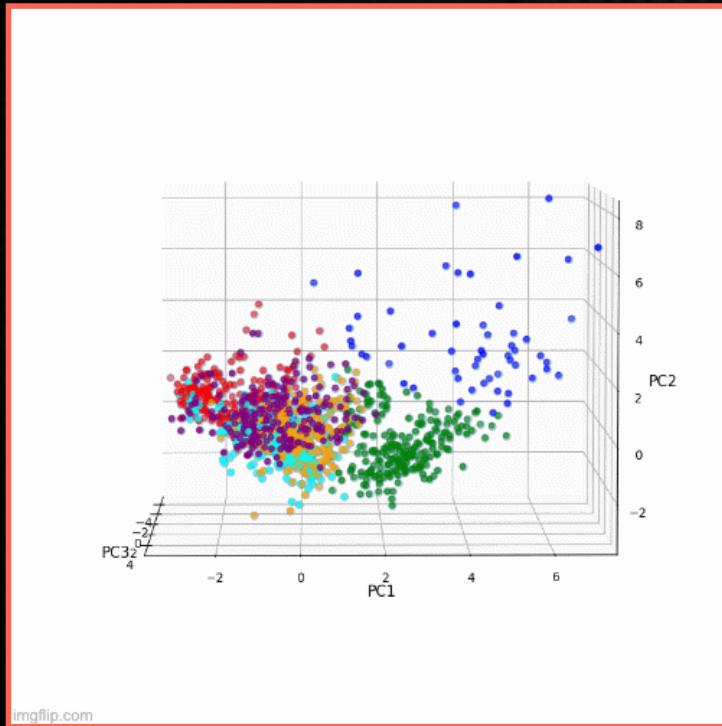


Sentiment Clusters: PCA

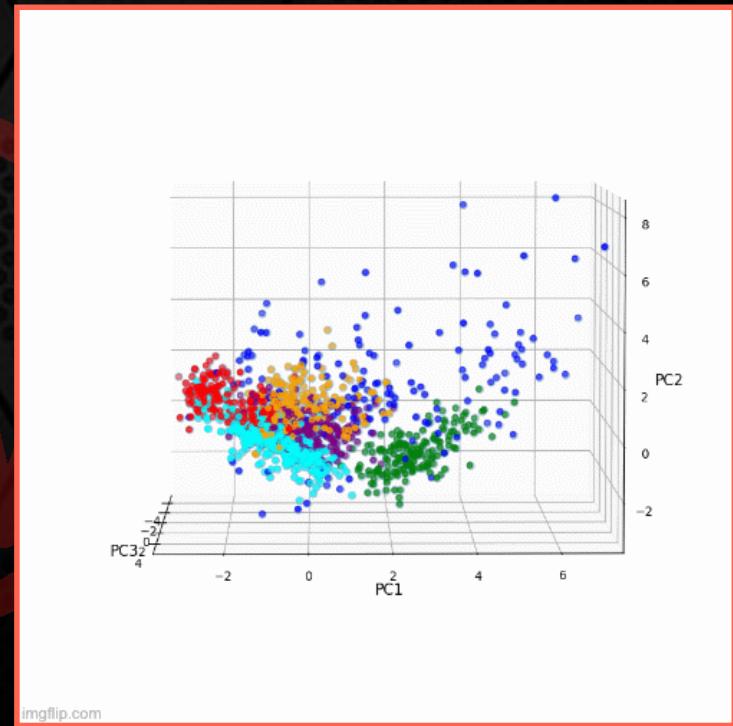


Sentiment Clusters

K-Means



Gaussian Mixture





Sentiment Clusters: Flying High vs. Laying Low



Video Demo

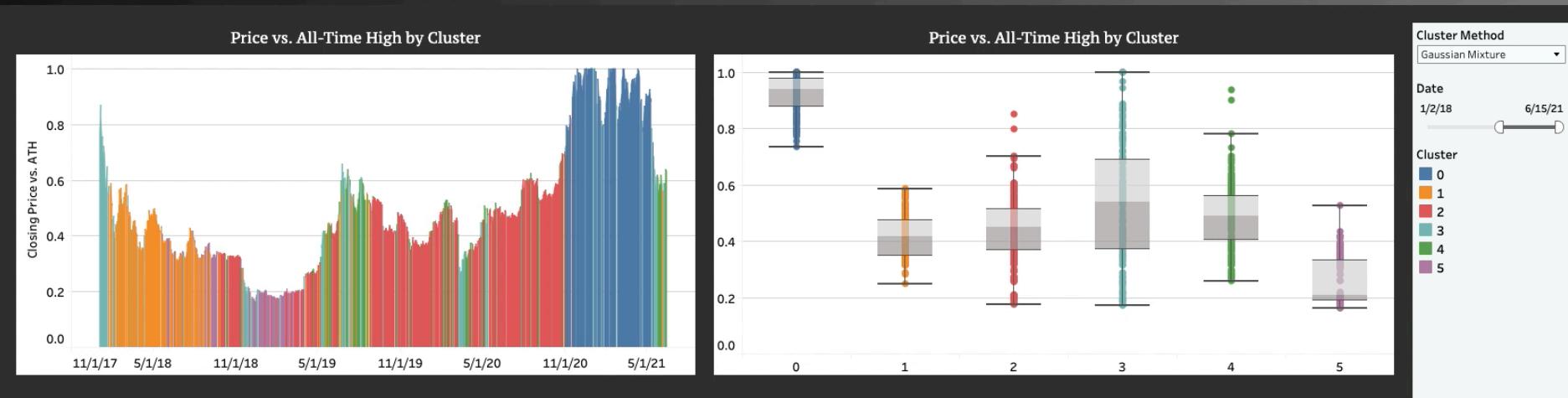
Sentiment Cluster Analysis

This dashboard is the result of machine learning clustering exercise. Colors in the area charts represent the cluster assigned to each date, while the level of each chart represents the level of the relevant metric on that date. Dates are clustered based on price levels, activity metrics, and sentiment metrics. These include Price vs. All-Time High, Net Unrealized Profit/Loss, Net Transfer Volume, Active Addresses, Google Activity, Reddit Activity, Reddit Polarity, Reddit Subjectivity, News Article Polarity, and News Article Subjectivity.

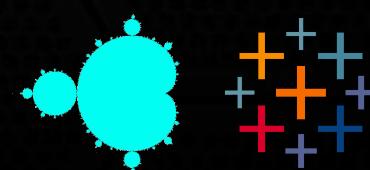
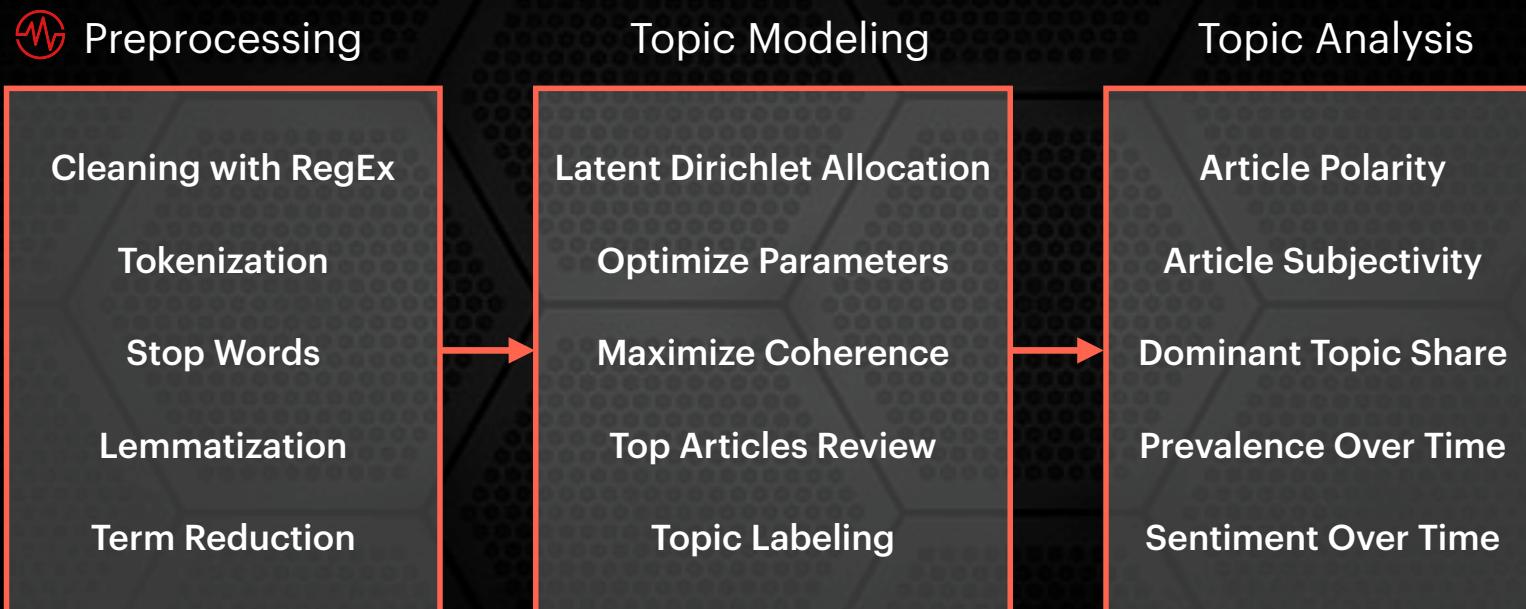
Clustering methods available include: K-Means (6 clusters), K-Means (9 clusters), Gaussian Mixture (6 clusters), Hierarchical Agglomerative Clustering (6 clusters).

Polarity is a measure of positivity/negativity. Subjectivity is a measure of how opinionated a comment or article is. These were calculated using TextBlob.

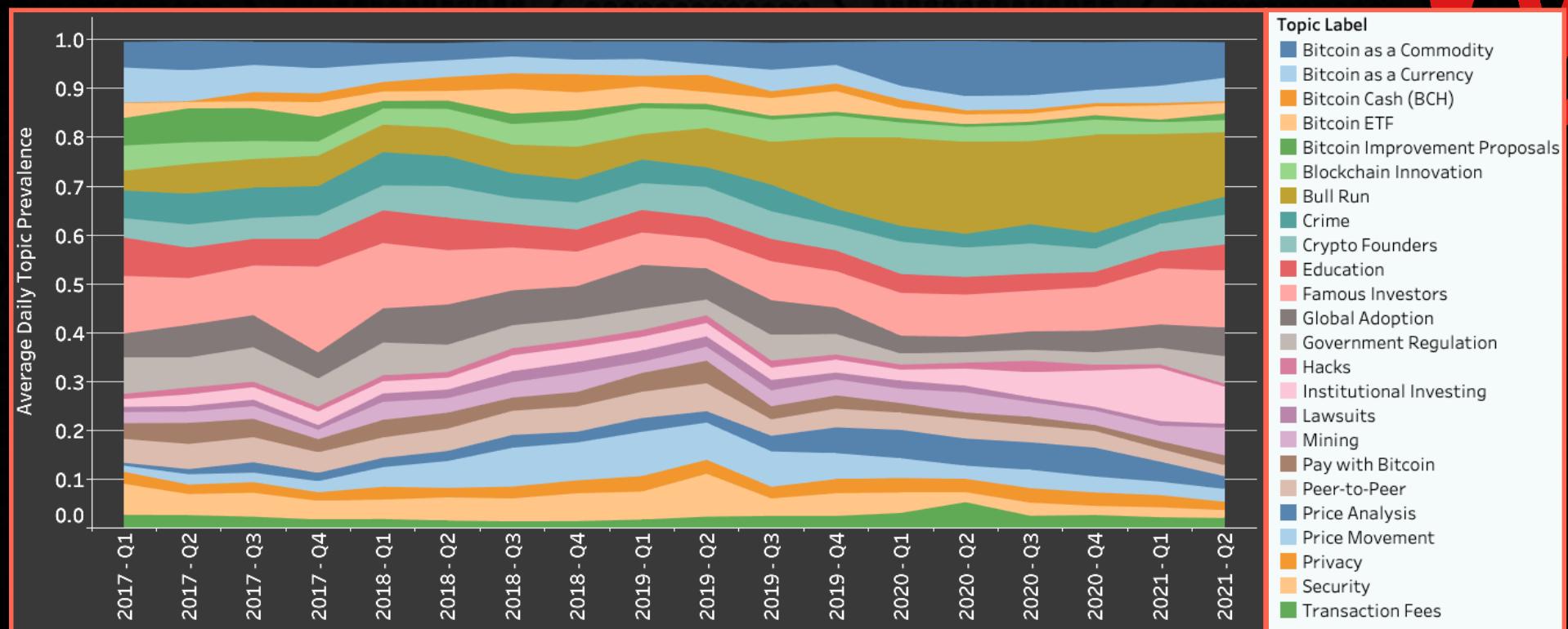
Use the date range selector to adjust the dates shown. This also adjusts the data fed to the boxplots, color coding, and the remainder of the dashboard



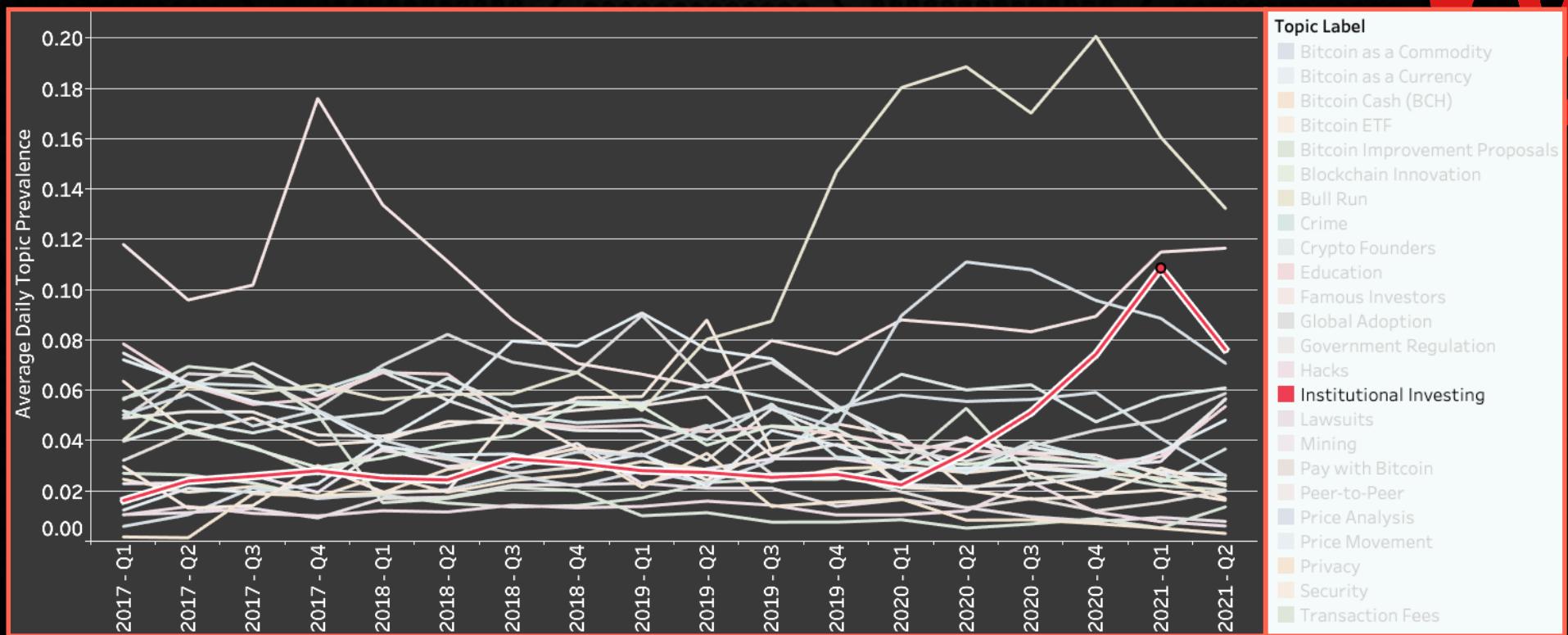
Topic Modeling: Workflow



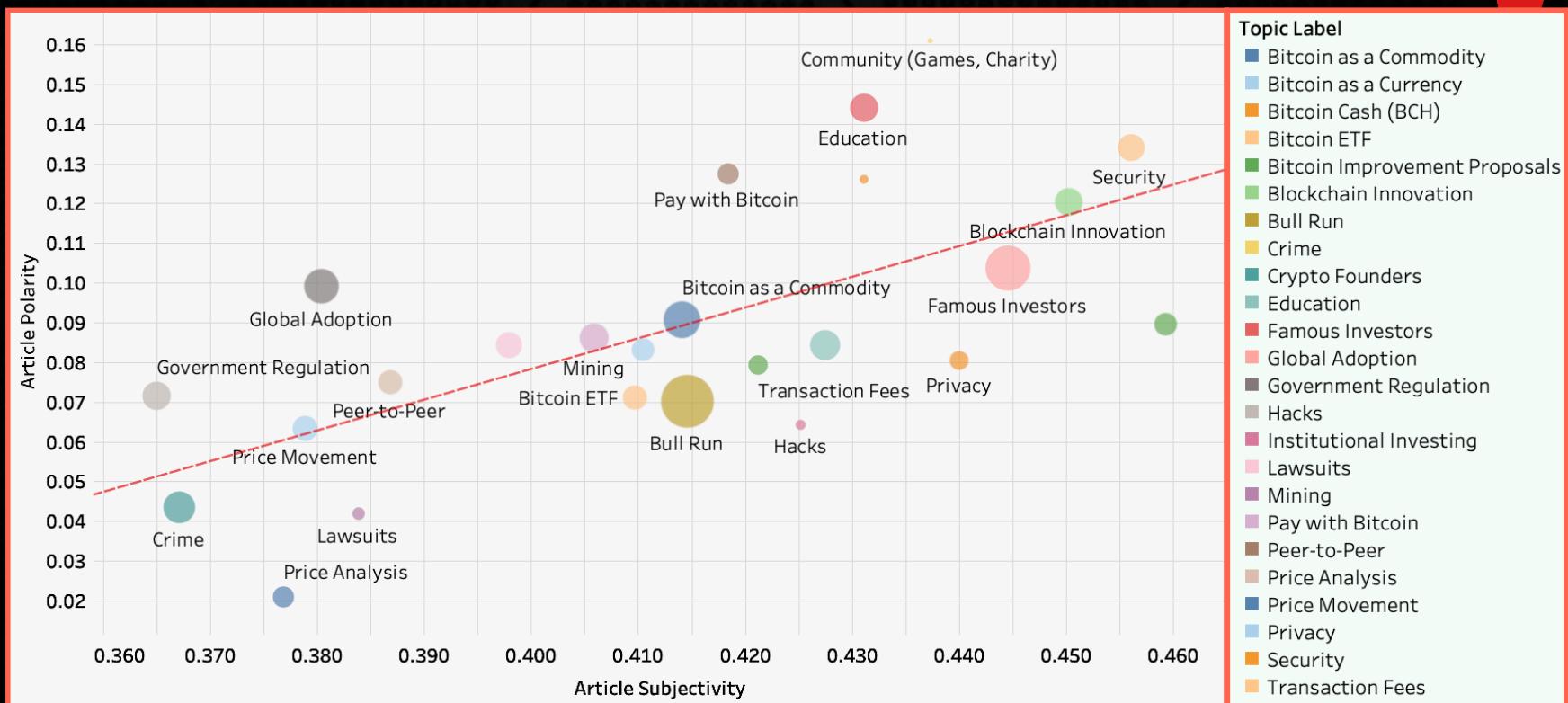
Topic Modeling: Prevalence



Topic Modeling: Prevalence



Topic Modeling: Sentiment





Topic Modeling: Institutional Investing



Video Demo

Topic Deep Dive

Take a deep dive into a single Bitcoin topic or combination of topics to view polarity, subjectivity, and prevalence over time.

Suggested Combinations: (Crime + Hacks + Lawsuits), (Bitcoin Improvement Proposals + Blockchain Innovation), (Price Analysis + Price Movement), (Privacy + Security), (Global Adoption + Government Regulation).

The table on the right, by default, shows the articles that had the highest percentage topic share for the selected topic. Dominant Share % can be adjusted in the far right container, along with the Date Range and the Date Level.



What's next for TokenSense?

- ⌚ Flesh out the data pipeline to allow for periodic updates
- ⌚ Perform topic modeling on Reddit data for comparison to news articles
- ⌚ Implement additional clustering and NLP techniques
 - ⌚ Dynamic Time Warping
 - ⌚ Named Entity Recognition





TokenSense

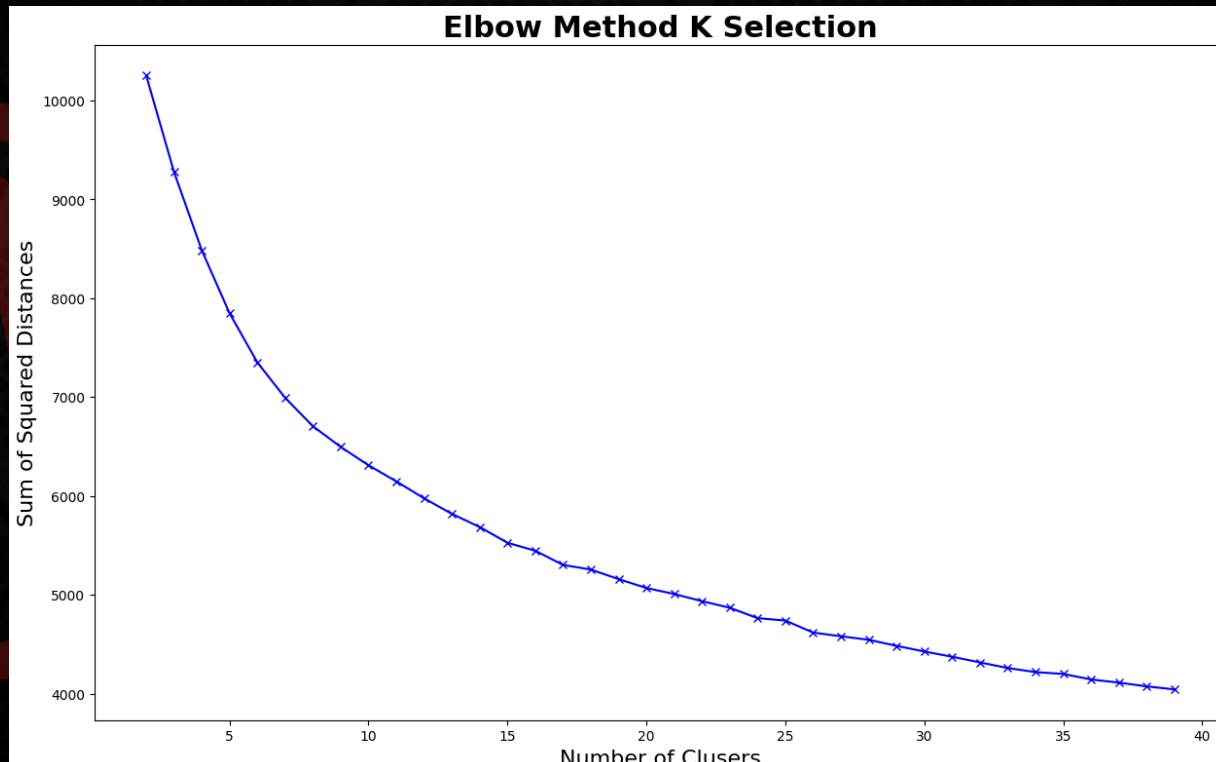
Keep on the crypto pulse





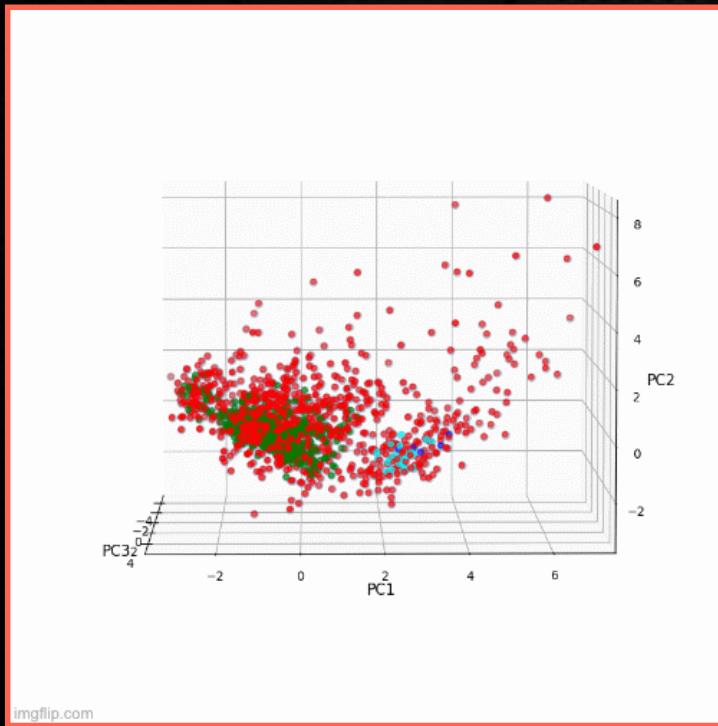
Appendices

K-Means K Selection

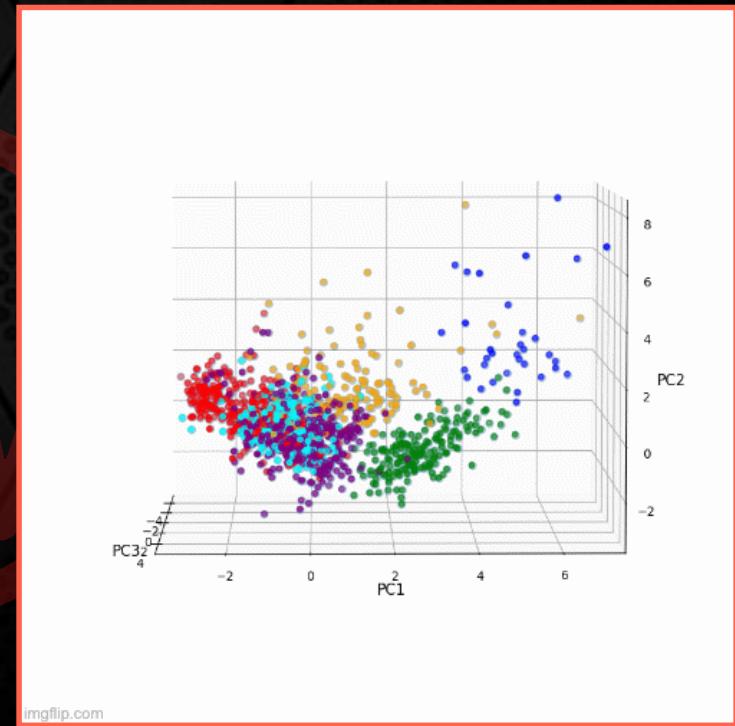


Sentiment Clusters

HDBSCSAN



Hierarchical Agglomerative



imgflip.com

imgflip.com