Bitcoin Market Trends & Sentiment Analysis

Submitted by:

Aashrey Jain, Matthew Maslow, Ishanay Sharma
Course: DS 598 Engineering Big Data

1. Introduction

The Challenge

The cryptocurrency market, particularly Bitcoin, is notoriously volatile. Price movements are often abrupt and influenced by investor sentiment, real-time news, and broader macroeconomic trends. Traditional analysis methods—such as historical trend observation and technical indicators—often fail to provide early warnings of sudden shifts.

Research Gap

- Traditional models overlook real-time sentiment.
- Market tools lack integration of sentiment and live data streams.

Project Objective

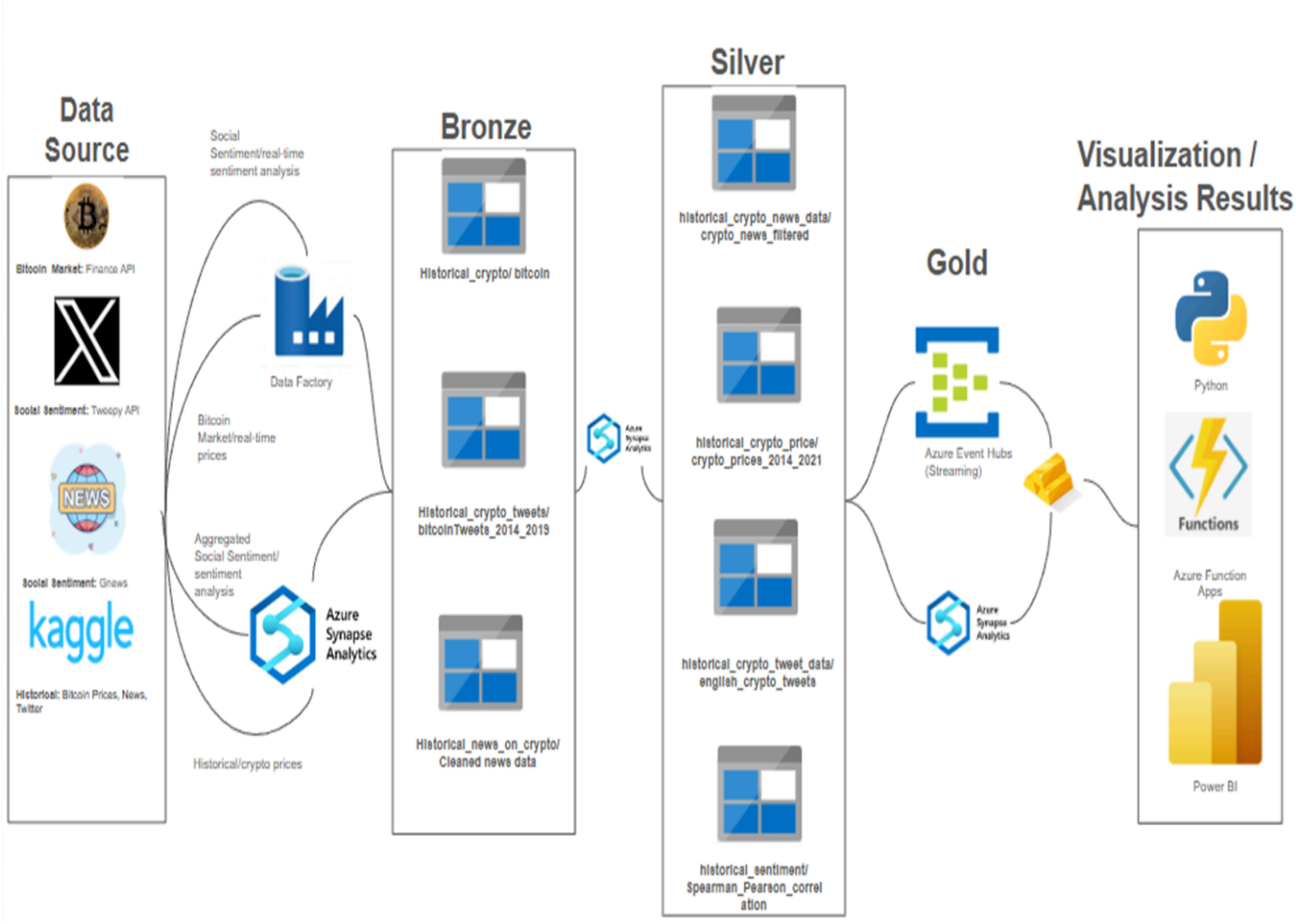
To build a **real-time analytics pipeline** on Azure that integrates:

- **Social Sentiment**: Extracted from Twitter and financial news
- **Market Data**: Live crypto prices and trading volumes
  Goal: Create an **early warning system** for traders to react swiftly to market movements influenced by public sentiment.

2. System Architecture

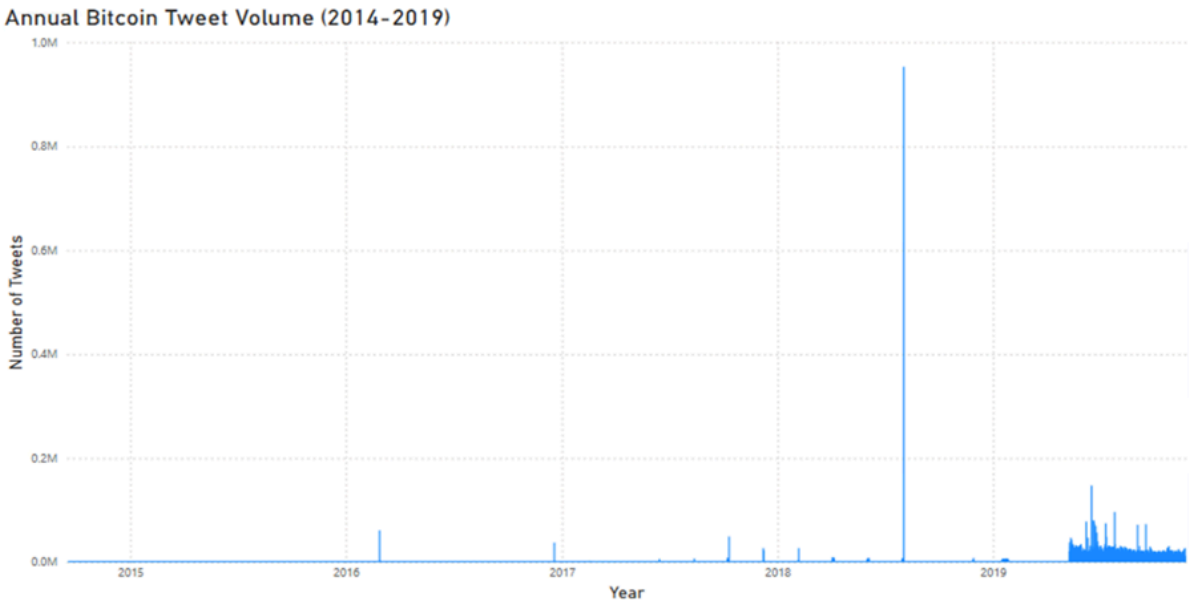
The project leverages Azure-based services for:

- Data ingestion (social media, news, market feeds)
- Stream processing and storage
- Machine learning for prediction
- Visualization dashboards (Power BI)



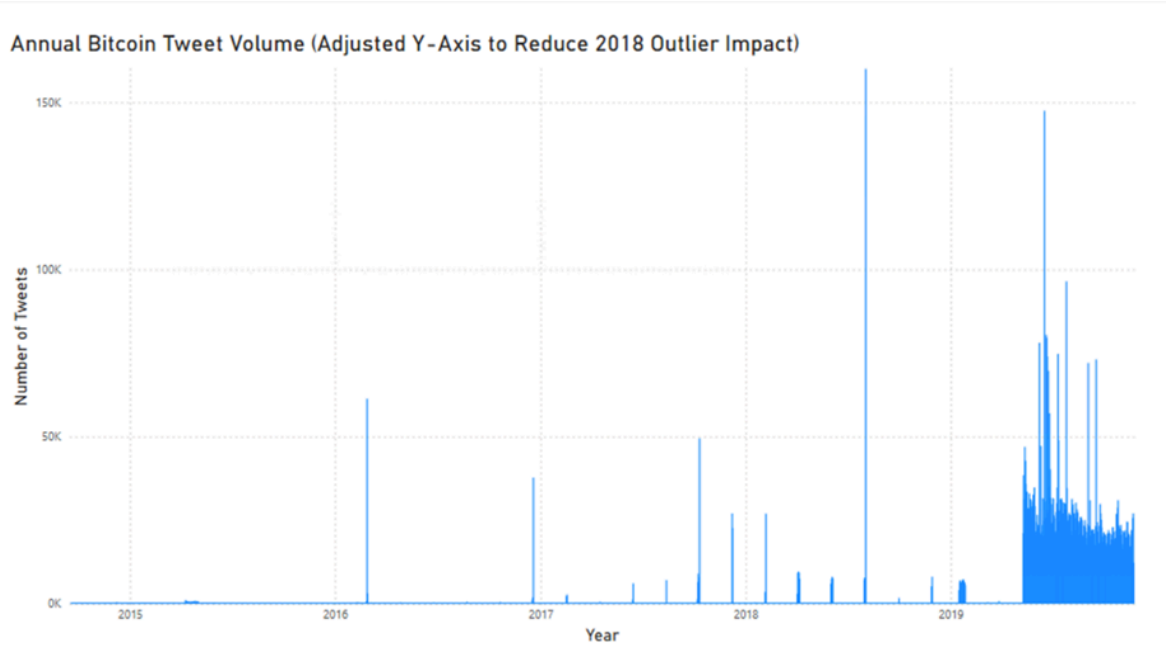
3. Exploratory Data Analysis (EDA)

Graph 1:



Explanation: The graph shows Bitcoin tweet volume from 2014 to 2019, with low activity early on, a major spike in late 2017 during Bitcoin's price surge, and steady increased interest through 2019.

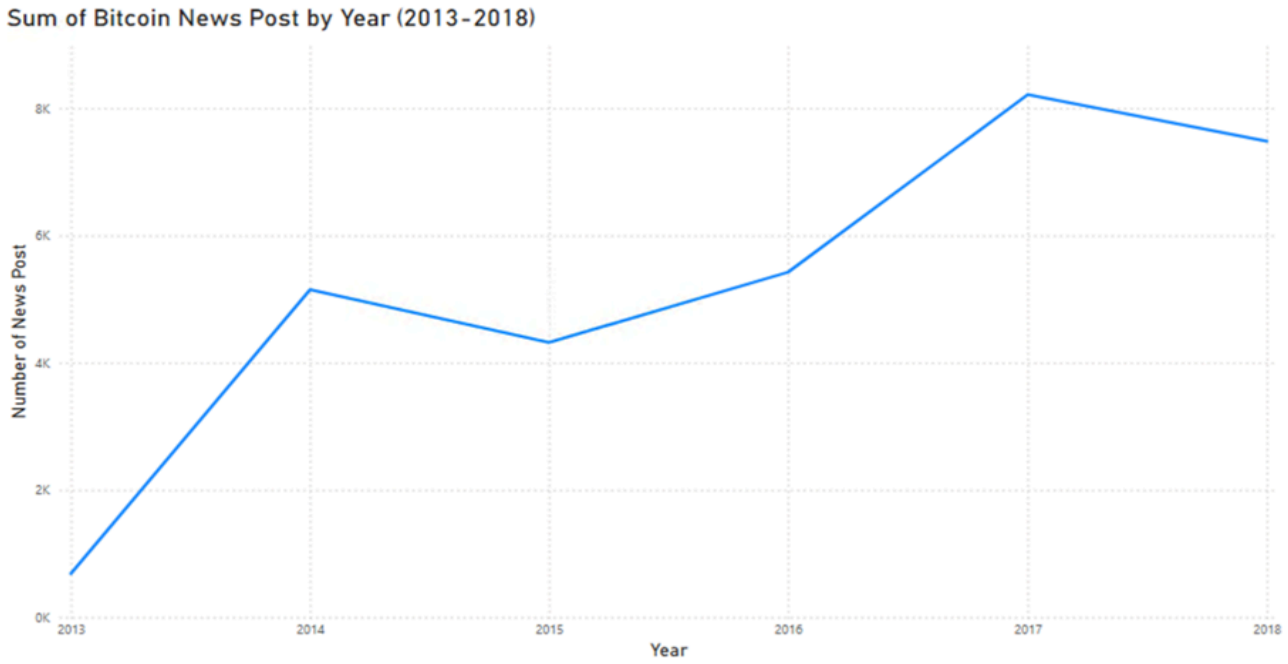
Graph 2:



Explanation: The graph shows that Bitcoin tweet volume was low from 2014 to 2017, with

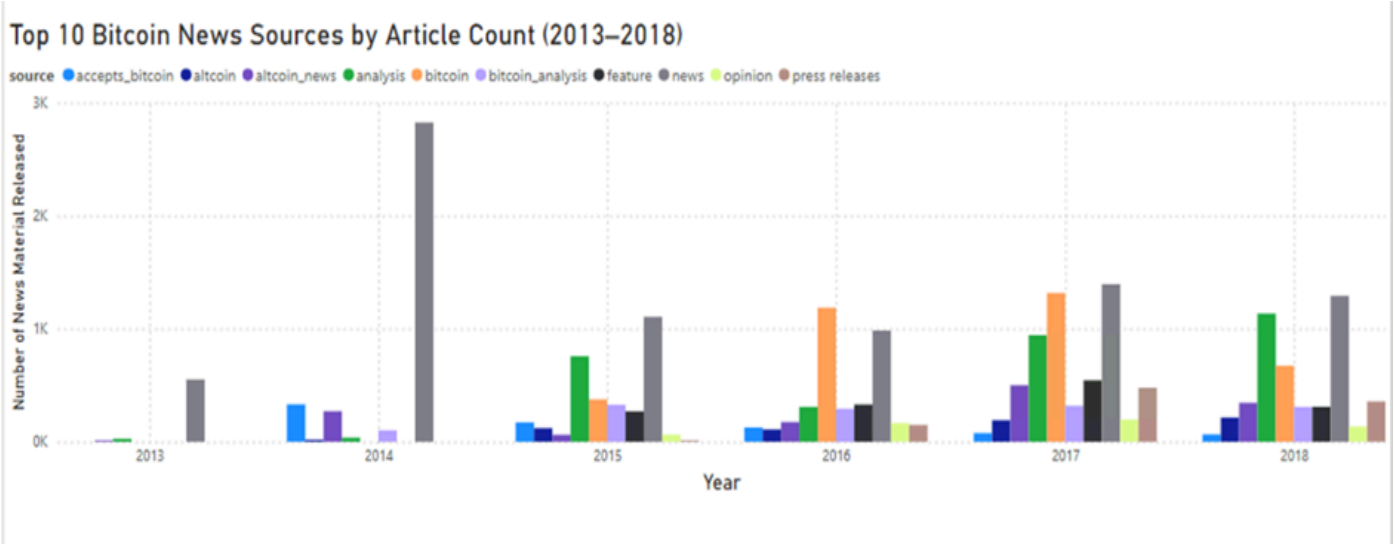
occasional spikes. A major spike occurred in 2018, and from late 2018 to 2019, tweet activity became consistently high, indicating growing and sustained public interest in Bitcoin.

Graph 3:



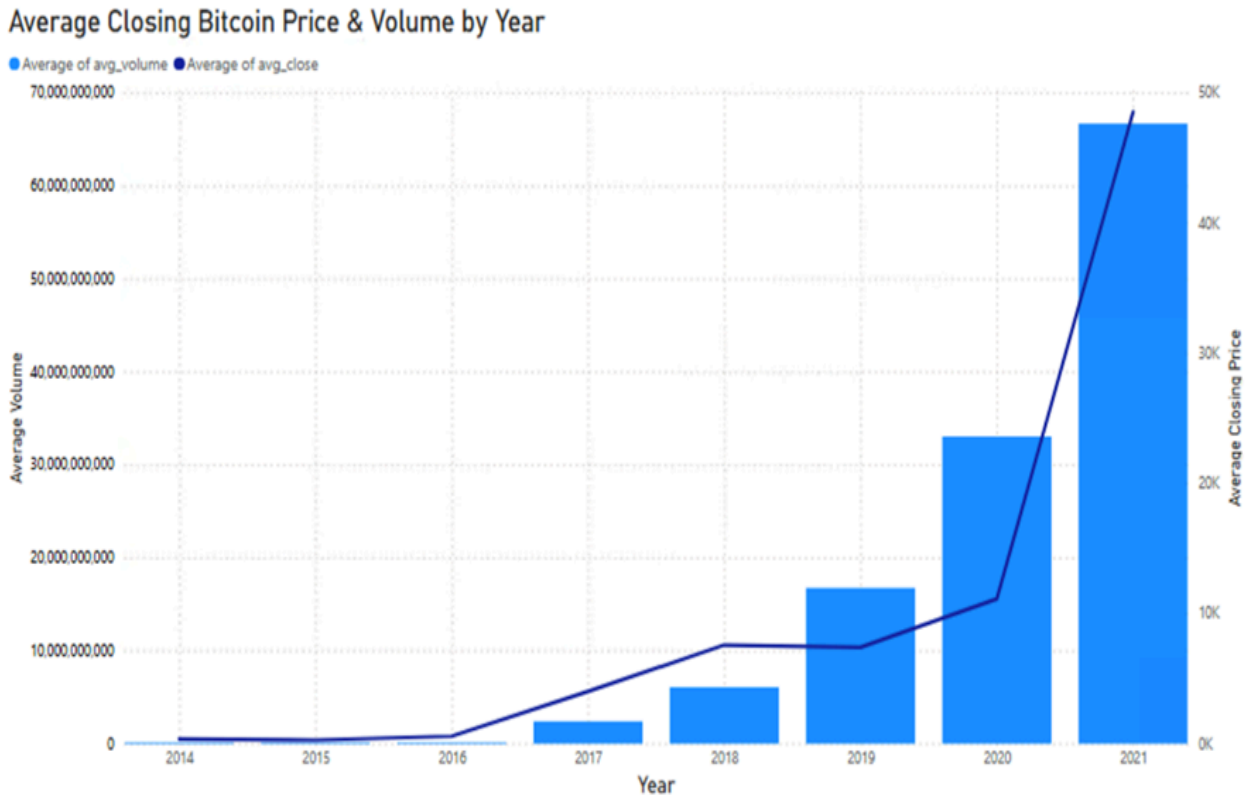
Explanation: This graph shows the number of Bitcoin news posts per year from 2013 to 2018. Bitcoin news coverage grew significantly over these years, peaking in 2017 with continued strong interest in 2018.

Graph 4:



Explanation: From 2013 to 2018, Bitcoin news articles grew steadily, with a major spike in 2014 and diverse content types (news, analysis, opinion) dominating from 2016 onward—showing both rising media attention and broader topic coverage over time.

Graph 5:



Explanation: From 2017 to 2021, both the average closing price and trading volume of Bitcoin saw exponential growth, with a sharp spike in 2021 where the price reached nearly $50,000 and trading volume peaked—indicating massive market expansion and investor interest.

4. The Sentiment Model

Model Selection: We selected the Bayesian-tuned HistGradientBoostingRegressor (HGBR) out of a variety of models we tested like Random Forest Classifier,XGBoost Classifier, Logistic Regression,etc. because it offered the best trade-off between prediction accuracy, interpretability, and robustness to missing and imbalanced data.

Among all models tested, it achieved the lowest test RMSE (0.0336), significantly outperforming baseline classifiers and regressors.

5. Feature Engineering

*** Input Features**

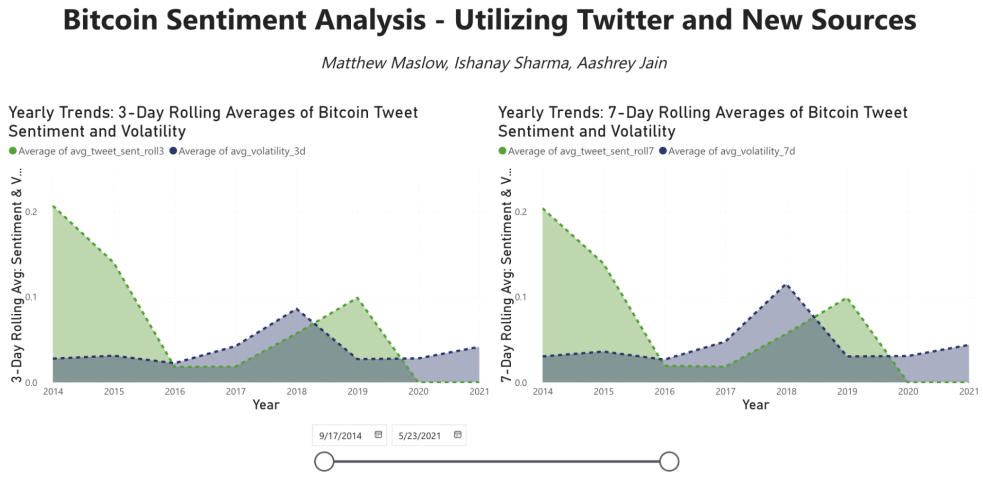
- **Sentiment:** Daily average and rolling sentiment scores from news and tweets.
- **Market:** Daily return, volatility, Simple Moving Average (SMA), and Relative Strength Index (RSI).

*** Prediction Target:** Next-day return of Bitcoin price
*** Inference Output:** A continuous return forecast
*** Decision Mapping**

- **Forecast ≥ +0.1% → Buy**
- **Forecast ≤ −0.1% → Sell**
- **Otherwise → Hold**

6. Final Dashboard Using Power BI

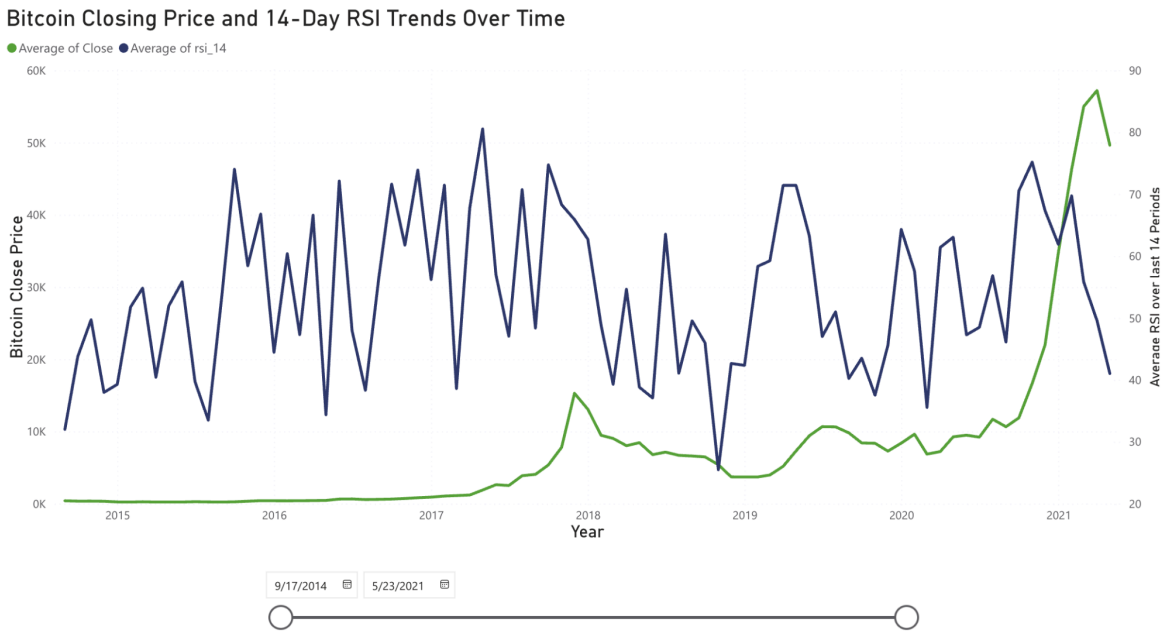
Tab 1: 3-Day and 7-Day Rolling Sentiment & Volatility



This first tab includes two side-by-side area plots:

- On the **left**, we show 3-day rolling averages of Twitter sentiment (green) and Bitcoin volatility (blue).
- On the **right**, we extend that to a 7-day rolling window.
- These plots help us examine how short-term sentiment fluctuations may correspond with market volatility. For example, during 2018 or early 2019, you can see sentiment dipping while volatility rises.
- At the bottom, a **date slider** lets us filter the visible time range, which is helpful when analyzing sentiment shifts around specific events.

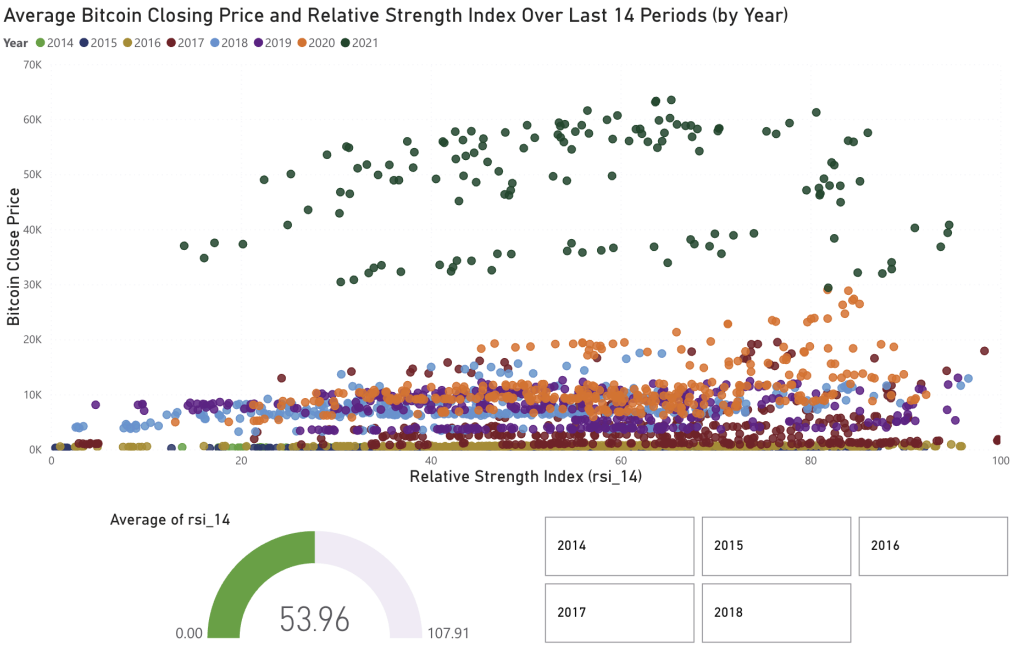
Tab 2: Bitcoin Closing Price vs 14-Day RSI



On this tab, we're comparing Bitcoin's **closing price** (green line) to the **14-day Relative Strength Index**, or RSI (blue line).

- RSI is a momentum indicator that signals overbought or oversold conditions. Spikes in RSI—especially those crossing 70—can precede downward corrections, while drops below 30 can indicate potential buying opportunities.
- Here, we include the **date range slider** to zoom into specific price movements and RSI surges. Notably, sharp RSI movements often align closely with dramatic price shifts.

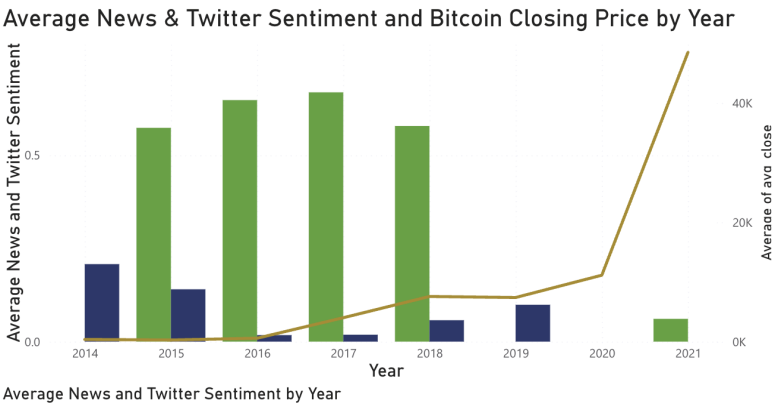
Tab 3: RSI vs Price by Year (Scatter + Gauge)

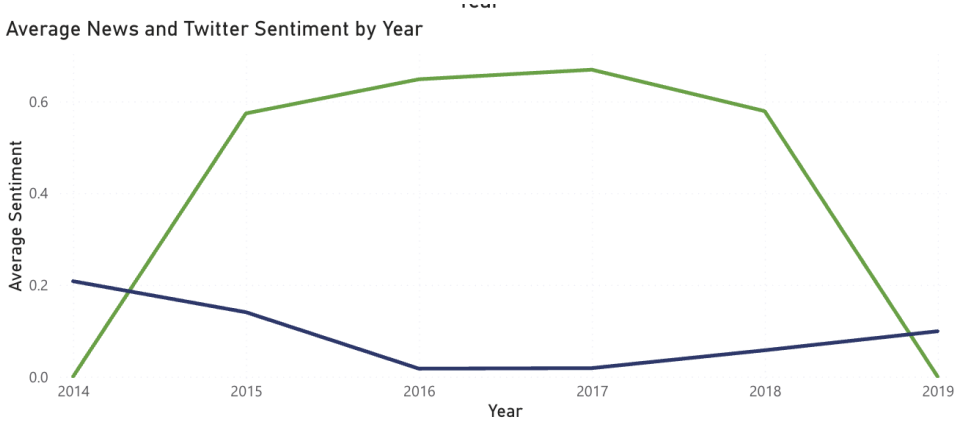


Here, we show a **scatter plot** of RSI versus Bitcoin's closing price, with each point color-coded by year. This lets us see how price and RSI cluster over time.

- For instance, in 2021 (the orange points), you'll notice a wide spread of high prices paired with high and low RSI values, highlighting volatility.
- Below the plot, we've added a **gauge chart** to show the overall average RSI across the dataset. We also included **year filter buttons** to isolate and analyze specific years, which is especially useful for tracking how sentiment and momentum varied historically.

Tab 4: Sentiment and Price Averages by Year





- This tab focuses on yearly comparisons between sentiment and market performance:
- The **top chart** shows average sentiment from Twitter and news (in bars), plotted against average Bitcoin closing price (orange line) for each year.
- The **bottom line chart** breaks down average sentiment from both sources to help highlight alignment or divergence between public opinion and price trends.

This layout helps see how sentiment may have led or lagged behind market cycles. For example, sentiment remained relatively high in 2017 and 2021, when Bitcoin reached major peaks.

7. Key Findings + Limitations

The dashboard reveals that short-term sentiment shifts (3–7 days) often coincide with spikes in volatility, highlighting the potential of social media as an early warning signal for market instability. The 14-day RSI strongly aligned with Bitcoin price trends, affirming its usefulness as a technical indicator. At the same time, scatterplots and yearly sentiment patterns, particularly in 2017 and 2021, suggest that public optimism frequently aligns with price surges. However, the analysis is limited by its reliance on historical data without real-time updates, basic polarity-based sentiment scoring that may misinterpret language, and potential bias from high-activity periods or unequal influence across data sources.

8. Sources

- https://www.kaggle.com/datasets/kashnitsky/news-about-major-cryptocurrencies-20132018-40k
- https://www.kaggle.com/datasets/gauravduttakiit/bitcoin-tweets-16m-tweets-with-sentiment-tagged