# **Exploring the Relationship between Social Media and Bitcoin Value**

Yixuan Wang, Wei Yang, Xinyi Li, Zexu Yang Instructor: Feng Mai



Business Intelligence & Analytics

jupyter

+ab|eau

#### Motivation

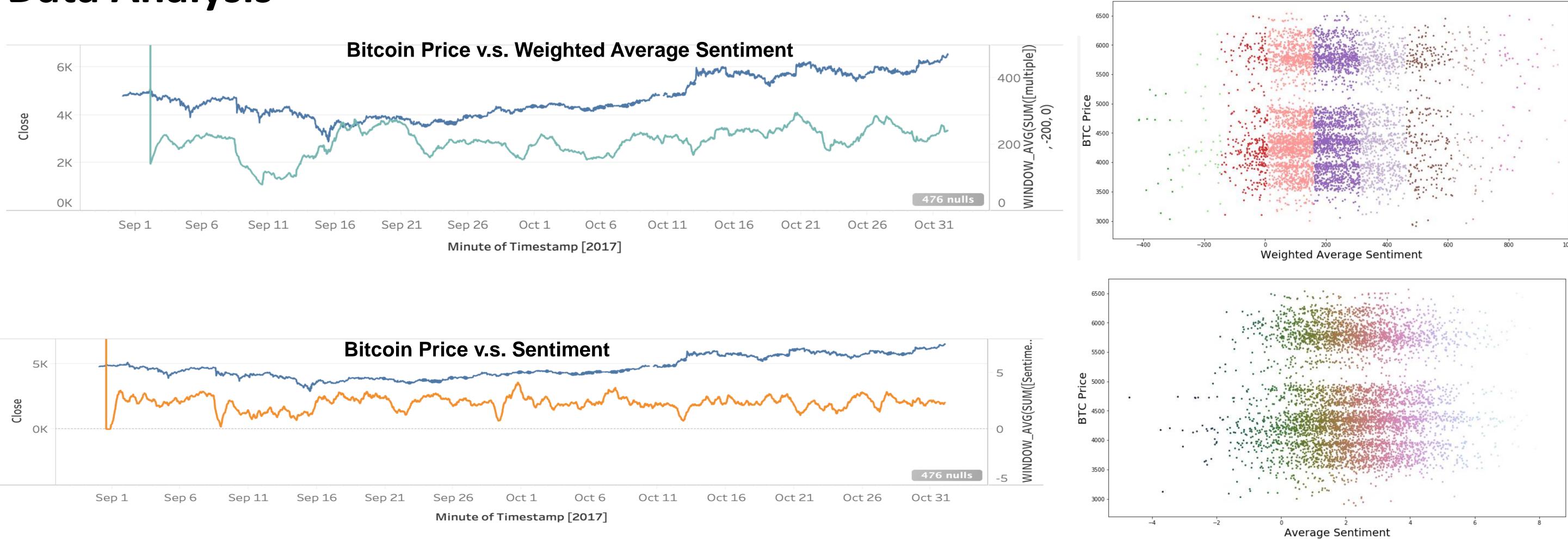
- With the popularity of Blockchain technologies and the increasing computing power invested in BTC mining, the decentralization and reliability of Bitcoin are promoting technology revolution and have been fully recognized by the market
- Based on previous researches, we hope to use textual data from Twitter, one of the largest social media platforms, to gain insights into the factors affecting the fluctuation of Bitcoin value

**Keywords:** Sentiment Analysis; Time Series Regression; Natural Language Processing (NLP)

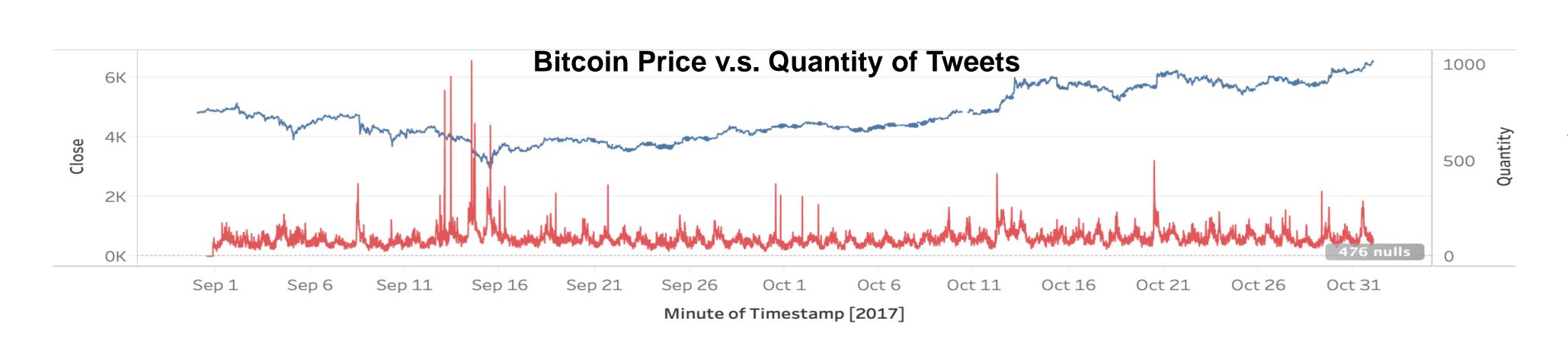
## Data Manipulation

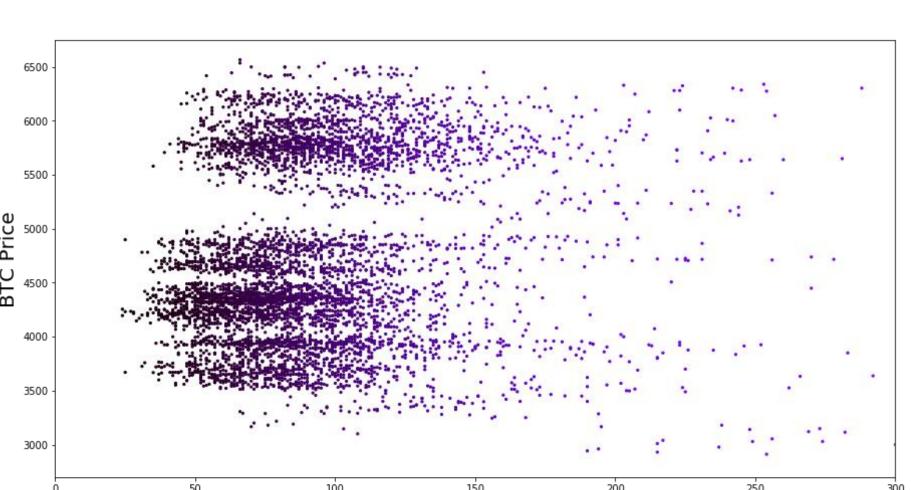
- Filtered stop-words, punctuation and whitespace
- Delivered Sentiment Analysis to distinguish positive and negative words
- Extracted and visualized the high-frequency words
- Quantified tweets features and integrated with Bitcoin value
- Analyzed the statistical relationships
- Visualized the results

### Data Analysis



• After applying the Moving Average method to these time series, we discovered that when people lose money by trading Bitcoin, they tend to publish negative words in Tweets. However, when they gain money from Bitcoin, they are less likely to showcase their success





• When Bitcoin value changes, the number of related comments from Tweet increases. Especially when Bitcoin value decreases, the number of negative words in Tweet increases dramatically

### **Future Works**

- With the collection of more labeled datasets, we will train the Convolution Neural Network (CNN) models to better capture the underlying characteristics of tweets
- Related to Bitcoin textual data, we will apply the pre-trained models of the NLU module in IBM Watson, the fastText in Facebook, and make comparison and evaluation
- For the analysis part, we will explore regression models with trading volume, other cryptocurrency price indices, and then implement them over a longer period



Word Cloud generated by Tweets contents related to Bitcoin