crypto_sentiment

May 23, 2019

```
In [1]: import os
        import pandas as pd
        %matplotlib inline
```

```
In [2]: from nltk.sentiment.vader import SentimentIntensityAnalyzer
        analyzer = SentimentIntensityAnalyzer()
```

1 News Headlines Sentiment

Use the news api to pull the latest news articles for bitcoin and ethereum and create a DataFrame of sentiment scores for each coin.

Use descriptive statistics to answer the following questions: 1. Which coin had the highest mean positive score? 2. Which coin had the highest negative score? 3. Which coin had the highest positive score?

```
In [3]: # Read your api key environment variable
```

```
In [4]: # Create a newsapi client
```

```
In [6]: # Fetch the Bitcoin news articles
```

```
In [7]: # Fetch the Ethereum news articles
```

```
In [8]: # Create the Bitcoin sentiment scores DataFrame
```

```
Out[8]:    Compound  Negative  Neutral  Positive  \
        0    0.0516     0.900    0.036     0.064
        1    0.3818     0.943    0.000     0.057
        2   -0.2263     0.888    0.065     0.047
        3    0.3612     0.937    0.000     0.063
        4   -0.6486     0.897    0.103     0.000

                                                        text
        0  Cryptocurrency exchange Binance has resumed tr...
        1  Bitcoin is now trading at around $8,130, up a ...
        2  Binance has vowed to raise the quality of its ...
        3  A new payment network called Flexa is launchin...
        4  If you thought that the theft of 7,000 bitcoin...
```

```
In [9]: # Create the ethereum sentiment scores DataFrame
```

```
Out[9]:    Compound  Negative  Neutral  Positive  \
        0   -0.3919     0.801    0.140     0.059
        1    0.1779     0.961    0.000     0.039
        2    0.0000     1.000    0.000     0.000
        3   -0.8020     0.717    0.217     0.066
        4   -0.6486     0.897    0.103     0.000

                                                        text
        0  President Trump tweeted insults at Twitter aga...
        1  After announcing that they were launching a ma...
        2  Captain Kirk and neo-Dadaists. Repugnant marke...
        3  So long as cryptocurrency exists, so too will ...
        4  If you thought that the theft of 7,000 bitcoin...
```

```
In [10]: # Describe the Bitcoin Sentiment
```

```
Out[10]:         Compound   Negative    Neutral   Positive
         count  17.000000  17.000000  17.000000  17.000000
         mean    0.024835   0.882118   0.049471   0.068412
         std     0.523138   0.082548   0.054700   0.087631
         min    -0.822500   0.691000   0.000000   0.000000
         25%    -0.318200   0.849000   0.000000   0.000000
         50%     0.000000   0.897000   0.051000   0.047000
         75%     0.381800   0.943000   0.065000   0.064000
         max     0.921700   1.000000   0.218000   0.267000
```

```
In [11]: # Describe the Ethereum Sentiment
```

```
Out[11]:         Compound   Negative    Neutral   Positive
         count  20.000000  20.000000  20.000000  20.000000
         mean    0.076370   0.873400   0.057900   0.068750
         std     0.518197   0.092956   0.080454   0.056343
         min    -0.918200   0.698000   0.000000   0.000000
         25%    -0.267700   0.815250   0.000000   0.019500
         50%     0.253500   0.877500   0.000000   0.063500
         75%     0.426225   0.940250   0.112250   0.107500
         max     0.769500   1.000000   0.240000   0.180000
```

1.0.1 Questions:

Q: Which coin had the highest mean positive score?
A:
Q: Which coin had the highest compound score?
A:
Q. Which coin had the highest positive score?
A:

2 Tokenizer

In this section, you will use NLTK and Python to tokenize the text for each coin. Be sure to: 1. Lowercase each word 2. Remove Punctuation 3. Remove Stopwords

```
In [12]: from nltk.tokenize import word_tokenize, sent_tokenize
         from nltk.corpus import stopwords
         from nltk.stem import WordNetLemmatizer, PorterStemmer
         from string import punctuation
         import re
```

```
In [13]: # Expand the default stop words list if necessary
```

```
In [14]: # Complete the tokenizer function
         def tokenizer(text):
             """Tokenizes text."""

             # Create a list of the words

             # Convert the words to lowercase

             # Remove the punctuation

             # Remove the stop words

             # Lemmatize Words into root words

             return tokens
```

```
In [15]: # Create a new tokens column for bitcoin
```

```
Out[15]:    Compound  Negative  Neutral  Positive  \
         0    0.0516     0.900    0.036     0.064
         1    0.3818     0.943    0.000     0.057
         2   -0.2263     0.888    0.065     0.047
         3    0.3612     0.937    0.000     0.063
         4   -0.6486     0.897    0.103     0.000

                                                     text  \
         0  Cryptocurrency exchange Binance has resumed tr...
         1  Bitcoin is now trading at around $8,130, up a ...
         2  Binance has vowed to raise the quality of its ...
         3  A new payment network called Flexa is launchin...
         4  If you thought that the theft of 7,000 bitcoin...

                                                   tokens
         0  [Cryptocurrency, exchange, Binance, resumed, t...
         1  [Bitcoin, trading, around, 8,130, whopping, 60...
         2  [Binance, vowed, raise, quality, security, aft...
```

```
3    [new, payment, network, called, Flexa, launchi...
4    [thought, theft, 7,000, bitcoins, one, world, ...
```

In [16]:
```python
# Create a new tokens column for ethereum
```

Out[16]:
```
    Compound  Negative  Neutral  Positive  \
0    -0.3919     0.801    0.140     0.059
1     0.1779     0.961    0.000     0.039
2     0.0000     1.000    0.000     0.000
3    -0.8020     0.717    0.217     0.066
4    -0.6486     0.897    0.103     0.000

                                                text  \
0  President Trump tweeted insults at Twitter aga...
1  After announcing that they were launching a ma...
2  Captain Kirk and neo-Dadaists. Repugnant marke...
3  So long as cryptocurrency exists, so too will ...
4  If you thought that the theft of 7,000 bitcoin...

                                              tokens
0  [President, Trump, tweeted, insult, Twitter, m...
1  [announcing, launching, managed, blockchain, s...
2  [Captain, Kirk, neo-Dadaists, Repugnant, marke...
3  [long, cryptocurrency, exists, extraordinaryle...
4  [thought, theft, 7,000, bitcoins, one, world, ...
```

---

3 NGrams and Frequency Analysis

In this section you will look at the ngrams and word frequency for each coin.

1. Use NLTK to produce the n-grams for N = 2.
2. List the top 10 words for each coin.

In [17]:
```python
from collections import Counter
from nltk import ngrams
```

In [18]:
```python
# Generate the Bitcoin N-grams where N=2
```

Out[18]:
```
[(('40', 'million'), 3),
 (('Cryptocurrency', 'exchange'), 2),
 (('exchange', 'Binance'), 2),
 (('crypto', 'asset'), 2),
 (('world', ''), 2),
 (('crypto', 'exchange'), 2),
 (('7,000', 'bitcoins'), 2),
 (('one', 'world'), 2),
```

```
 (('price', 'Bitcoin'), 2),
 (('Wall', 'Street'), 2),
 (('Street', 'Market'), 2),
 (('last', 'week'), 2),
 (('char', 'Cryptocurrency'), 2),
 (('Binance', 'resumed'), 1),
 (('resumed', 'trading'), 1),
 (('trading', 'activity'), 1),
 (('activity', 'Users'), 1),
 (('Users', 'cancel'), 1),
 (('cancel', 'open'), 1),
 (('open', 'order'), 1)]
```

In [19]:
```
# Generate the Ethereum N-grams where N=2
```

Out[19]:
```
[(('private', 'key'), 5),
 (('Ethereum', 'ETH'), 3),
 (('blockchain', 'bandit'), 2),
 (('last', 'year'), 2),
 (('weak', 'private'), 2),
 (('Elon', 'Musk'), 2),
 (('worth', 'Ethereum'), 2),
 (('JP', 'Morgan'), 2),
 (('President', 'Trump'), 1),
 (('Trump', 'tweeted'), 1),
 (('tweeted', 'insult'), 1),
 (('insult', 'Twitter'), 1),
 (('Twitter', 'morning'), 1),
 (('morning', 'time'), 1),
 (('time', 'Jack'), 1),
 (('Jack', 'took'), 1),
 (('took', 'conversation'), 1),
 (('conversation', 'platform'), 1),
 (('platform', 'White'), 1),
 (('White', 'House'), 1)]
```

In [20]:
```
# Use the token_count function to generate the top 10 words from each coin
def token_count(tokens, N=10):
    """Returns the top N tokens from the frequency count"""
    return Counter(tokens).most_common(N)
```

In [21]:
```
# Get the top 10 words for Bitcoin
```

Out[21]:
```
[('char', 16),
 ('Bitcoin', 8),
 ('', 8),
 ('price', 6),
 ('hacker', 6),
 ('exchange', 5),
```

```
 ('Binance', 5),
 ('crypto', 4),
 ('cryptocurrency', 4),
 ('million', 4)]
```

In [22]:
```python
# Get the top 10 words for Ethereum
```

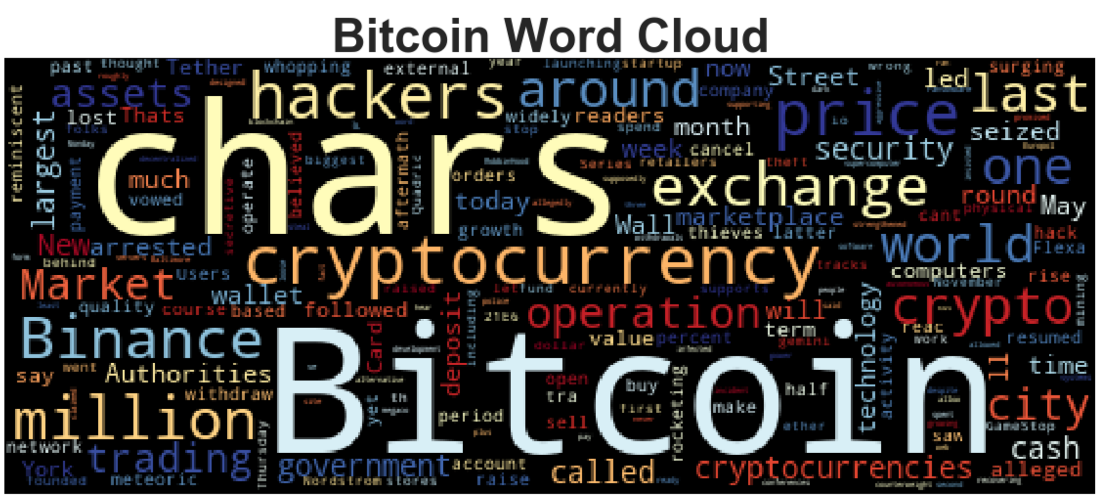
Out[22]:
```
[('char', 19),
 ('Ethereum', 9),
 ('blockchain', 8),
 ('key', 7),
 ('private', 5),
 ('world', 5),
 ('time', 4),
 ('weak', 4),
 ('million', 4),
 ('transaction', 4)]
```

4 Word Clouds

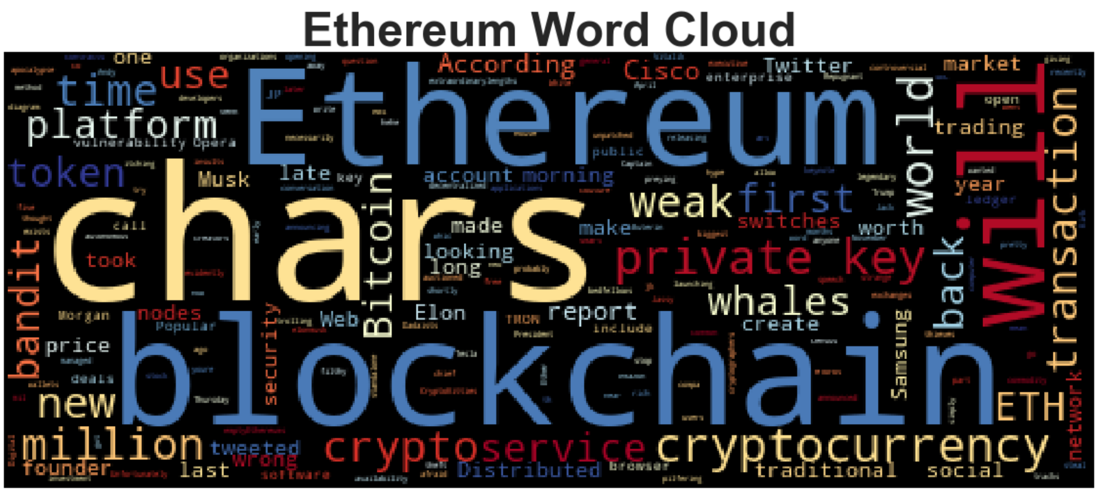
In this section, you will generate word clouds for each coin to summarize the news for each coin

In [23]:
```python
from wordcloud import WordCloud
import matplotlib.pyplot as plt
plt.style.use('seaborn-whitegrid')
import matplotlib as mpl
mpl.rcParams['figure.figsize'] = [20.0, 10.0]
```

In [25]:
```python
# Generate the Bitcoin word cloud
```



In [26]:
```python
# Generate the Ethereum word cloud
```



5 Named Entity Recognition

In this section, you will build a named entity recognition model for both coins and visualize the tags using SpaCy.

```
In [27]: import spacy
         from spacy import displacy
```

```
In [28]: # Optional - download a language model for SpaCy
         # !python -m spacy download en_core_web_sm
```

```
In [29]: # Load the spaCy model
         nlp = spacy.load('en_core_web_sm')
```

5.1 Bitcoin NER

```
In [30]: # Concatenate all of the bitcoin text together
```

```
Out[30]: 'Cryptocurrency exchange Binance has resumed trading activity. Users can now cancel ope
```

```
In [31]: # Run the NER processor on all of the text

         # Add a title to the document
```

```
In [32]: # Render the visualization
```

```
<IPython.core.display.HTML object>
```

```
In [33]: # List all Entities
```

Binance PERSON
Binance GPE
around $8,130 MONEY
60.84 percent PERCENT
the past month DATE
3,086.14 MONEY
the latter half of 2017 DATE
$40 million MONEY
Bitcoin GPE
Flexa ORG
today DATE
GameStop ORG
Nordstrom NORP
7,000 CARDINAL
Bitcoin GPE
Thursday DATE
Bitcoin GPE
6,000 MONEY
first ORDINAL
November last year DATE
$15 million MONEY
Europol PERSON
three CARDINAL
second ORDINAL
more than 550,000 MONEY
around $615,000 MONEY
last week DATE
m PERSON
May 7 DATE
about 10,000 CARDINAL
Baltimore GPE
RobbinHood ORG
13 CARDINAL
76,280 MONEY
today DATE
102,310 MONEY
Binance PERSON
last week's DATE
7,000 CARDINAL
roughly $40 million MONEY
Monday DATE
Binance GPE
one CARDINAL
about 7,000 CARDINAL
around $40 million MONEY
Bloomb PERSON
Tether PERSON
New Yorks GPE

```
month DATE
California GPE
Maine GPE
New York GPE
Albany GPE
Peter da Silva PERSON
the National Academy of Sciencies ORG
1965 DATE
Bitcoin BTC ORG
Tether ORG
Bitcoin GPE
Bitcoin GPE
the Wall Street Market FAC
European NORP
U.S. GPE
Millions CARDINAL
Bitcoin GPE
```

5.2 Ethereum NER

```
In [34]: # Concatenate all of the bitcoin text together

Out[34]: 'President Trump tweeted insults at Twitter again this morning, but this time Jack took

In [35]: # Run the NER processor on all of the text

         # Add a title to the document

In [36]: # Render the visualization
```

```
<IPython.core.display.HTML object>
```

```
In [37]: # List all Entities
```

```
Trump PERSON
Twitter GPE
this morning TIME
Jack PERSON
the White House ORG
late last year DATE
Amazon Web Services ORG
only about five months ago DATE
Andy Jassy PERSON
Kirk PERSON
neo-Dadaists NORP
```

Digital ORG
CryptoKitties ORG
2047 DATE
2026 DATE
one CARDINAL
millions CARDINAL
7,000 CARDINAL
Bitcoin GPE
Thursday DATE
Bitcoin GPE
6,000 MONEY
first ORDINAL
November last year DATE
Ether PERSON
ETH ORG
ETH ORG
Vitalik Buterin PERSON
Twitter GPE
Ethereum GPE
early this morning TIME
April 30 DATE
201 CARDINAL
chars]Have PERSON
first ORDINAL
SEC ORG
CoinDesk Korea GPE
over $6.1 million MONEY
Ethereum GPE
Independent Security Evaluators ORG
ISE ORG
ETH ORG
just 7 percent PERCENT
a third CARDINAL
Chainanalysis GPE
chars]Cisco CVE-2019-1804 PERSON
Cisco GPE
9000 PRODUCT
Cisco PERSON
hundreds of millions CARDINAL
Ion ORG
Airbnb PERSON
Facebook PERSON
TRON ORG
TRON ORG
Opera ORG
millions of dollars MONEY
Ethereum GPE
Morgan PERSON

```
Microsoft ORG
Jane Connolly PERSON
Quorum ă PERSON
Ethereum GPE
first ORDINAL
the late 1990s DATE
AllAdvantage ORG
AllAdvant PERSON
```