In [486]:

```python
import os
import pandas as pd
%matplotlib inline
from newsapi import NewsApiClient
```

In [487]:

```python
from nltk.sentiment.vader import SentimentIntensityAnalyzer
analyzer = SentimentIntensityAnalyzer()
```

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 [488]:

```python
# Read your api key environment variable
api_key = os.getenv("NEWS_API_KEY")
```

In [489]:

```python
type(api_key)
```

Out[489]:

```
str
```

In [490]:

```python
# Create a newsapi client
newsapi = NewsApiClient(api_key=api_key)
```

In [491]:

```
newsapi
```

Out[491]:

```
<newsapi.newsapi_client.NewsApiClient at 0x1a32f87690>
```

In [492]:

```python
# Fetch the Bitcoin news articles
btc_headlines = newsapi.get_everything(
    q= 'Bitcoin' ,
    language="en"
)
# Print total articles
print(f"Total articles about Bitcoin \"BTC\": {btc_headlines['totalResults']}")

# Show sample article
btc_headlines["articles"][0]
```

```
Total articles about Bitcoin "BTC": 3628
```

Out[492]:

```
{'source': {'id': None, 'name': 'Gizmodo.com'},
 'author': 'Tom McKay',
 'title': 'God Help Us, the Winklevoss Twins Are Co-Producing a Movi
e About Their Bitcoin Journey',
 'description': 'The Winklevoss twins—the brothers Mark Zuckerberg a
llegedly stole the idea for Facebook from—are making a movie about t
hemselves, Deadline reported this week. If that doesn't seem masturb
atory enough, rejoice: said film is about how they amassed a net wea
lth …',
 'url': 'https://gizmodo.com/god-help-us-the-winklevoss-twins-are-co
-producing-a-mo-1843973511',
 'urlToImage': 'https://i.kinja-img.com/gawker-media/image/upload/c_
fill,f_auto,fl_progressive,g_center,h_675,pg_1,q_80,w_1200/iujekjzgr
ttsevxxzbop.jpg',
 'publishedAt': '2020-06-09T23:45:00Z',
 'content': 'The Winklevoss twinsthe brothers Mark Zuckerberg allege
dly stole the idea for Facebook fromare making a movie about themsel
ves, Deadline reported this week. If that doesnt seem masturbatory e
nough, r… [+2657 chars]'}
```

In [493]:

```python
# Fetch the Ethereum news articles
eth_headlines = newsapi.get_everything(
    q="ethereum AND +ethereum AND ETH",
    language="en"
)
# Print total articles
print(f"Total articles about Ethereum \"ETH\": {eth_headlines['totalResults']}")

# Show sample article
eth_headlines["articles"][0]
```

Total articles about Ethereum "ETH": 391

Out[493]:

```
{'source': {'id': 'vice-news', 'name': 'Vice News'},
 'author': 'Kevin Truong, Jordan Pearson',
 'title': 'Someone Mysteriously Paid $2.5M to Send $133 in Cryptocur
rency',
 'description': 'Was it simply a catastrophically expensive mistake,
or something more nefarious? The mining pool is holding on to the fe
e in case the sender speaks up.',
 'url': 'https://www.vice.com/en_us/article/z3e4v8/someone-mysteriou
sly-paid-25m-to-send-133-in-ethereum-cryptocurrency',
 'urlToImage': 'https://video-images.vice.com/test-uploads/articles/
5ee10c76f934b8009de6a4b6/lede/1591807361583-man-in-white-shirt-using
-macbook-pro-52608.jpeg?crop=1xw:0.8408xh;0xw,0.1592xh&resize=1200:*
',
 'publishedAt': '2020-06-10T16:44:10Z',
 'content': 'Imagine sending $133 to a friend and being charged a $2
.5 million fee. That seems to be the case for a cryptocurrency user
who paid a $2.5 million transaction fee during a trade on Wednesday
morning … [+2684 chars]'}
```

In [494]:

```python
# Create the Bitcoin sentiment scores DataFrame
btc_sentiments = []

for article in btc_headlines["articles"]:
    try:
        text = article["title"]
        date = article["publishedAt"][:10]
        sentiment = analyzer.polarity_scores(text)
        compound = sentiment["compound"]
        pos = sentiment["pos"]
        neu = sentiment["neu"]
        neg = sentiment["neg"]

        btc_sentiments.append({
            "text": text,
            "date": date,
            "compound": compound,
            "positive": pos,
            "negative": neg,
            "neutral": neu

        })

    except AttributeError:
        pass

# Create DataFrame
btc_df = pd.DataFrame(btc_sentiments)

# Reorder DataFrame columns
cols = ["date", "text", "compound", "positive", "negative", "neutral"]
btc_df = btc_df[cols]

btc_df
```

Out[494]:

	date	text	compound	positive	negative	neutral
0	2020-06-09	God Help Us, the Winklevoss Twins Are Co-Produ...	0.5859	0.304	0.000	0.696
1	2020-06-04	How to Buy a Burner Phone	0.0000	0.000	0.000	1.000
2	2020-05-28	Goldman Sachs lists 5 reasons why bitcoin is '...	-0.2755	0.000	0.123	0.877
3	2020-06-02	'Get on the train': Famed investor Michael Nov...	-0.3612	0.000	0.111	0.889

4	2020-05-25	Satoshi Nakaboto: 'Bitcoin continues downtrend...	-0.3182	0.000	0.187	0.813
5	2020-05-26	Satoshi Nakaboto: 'Visa approves launch of cry...	0.4019	0.213	0.000	0.787
6	2020-05-22	8 Reasons Why This Could Be the Time to Take B...	-0.1779	0.000	0.145	0.855
7	2020-06-08	Bitcoin Emerges as New Standard in Post-COVID ...	0.0000	0.000	0.000	1.000
8	2020-05-28	Satoshi Nakaboto: 'Bitcoin is a bad investment...	-0.5423	0.000	0.280	0.720
9	2020-05-27	Satoshi Nakaboto: 'S&P 500 outperformed Bitcoi...	0.0000	0.000	0.000	1.000
10	2020-05-22	Satoshi Nakaboto: 'Bitcoin whale moves $400M, ...	0.0000	0.000	0.000	1.000
11	2020-06-09	Satoshi Nakaboto: 'Avocado prices indicated se...	0.0000	0.000	0.000	1.000
12	2020-06-03	Satoshi Nakaboto: 'Bitcoin plummets almost $70...	0.0000	0.000	0.000	1.000
13	2020-06-18	Satoshi Nakaboto: 'Bitcoin broker sees 77% of ...	0.3818	0.191	0.000	0.809
14	2020-06-05	Satoshi Nakaboto: 'Bitcoin on its way back to ...	0.0000	0.000	0.000	1.000
15	2020-06-02	Satoshi Nakaboto: 'Bitcoin hits $10,000 again ...	0.0000	0.000	0.000	1.000
16	2020-06-08	Satoshi Nakaboto: 'Daily Ethereum network fees...	0.0000	0.000	0.000	1.000
17	2020-06-11	Satoshi Nakaboto: 'Microsoft releases Bitcoin-...	-0.3400	0.000	0.211	0.789
18	2020-05-18	Satoshi Nakaboto: 'Bitcoin transaction fees su...	0.0000	0.000	0.000	1.000
19	2020-05-20	Satoshi Nakaboto: 'VC firm Andreessen Horowitz...	0.0000	0.000	0.000	1.000

In [497]:

```
eth_sentiments[0]['text']
```

Out[497]:

'Imagine sending $133 to a friend and being charged a $2.5 million fee. That seems to be the case for a cryptocurrency user who paid a $2.5 million transaction fee during a trade on Wednesday morning … [+2684 chars]'

In [498]:

```python
# Create the ethereum sentiment scores DataFrame
eth_sentiments = []

for article in eth_headlines["articles"]:
    try:
        text = article["content"]
        date = article["publishedAt"][:10]
        sentiment = analyzer.polarity_scores(text)
        compound = sentiment["compound"]
        pos = sentiment["pos"]
        neu = sentiment["neu"]
        neg = sentiment["neg"]

        eth_sentiments.append({
            "text": text,
            "date": date,
            "compound": compound,
            "positive": pos,
            "negative": neg,
            "neutral": neu

        })

    except AttributeError:
        pass

# Create DataFrame
eth_df = pd.DataFrame(eth_sentiments)

# Reorder DataFrame columns
cols = ["date", "text", "compound", "positive", "negative", "neutral"]
eth_df = eth_df[cols]

eth_df.head()
```

Out[498]:

	date	text	compound	positive	negative	neutral
0	2020-06-10	Imagine sending $133 to a friend and being cha...	0.3400	0.089	0.050	0.861
1	2020-06-12	Intel announced CrossTalk, a new side-channel ...	-0.6705	0.000	0.143	0.857
2	2020-06-10	POLAND - 2020/03/13: In this photo illustratio...	0.0000	0.000	0.000	1.000
3	2020-06-11	Photo taken in New York City, United States\r\...	0.3291	0.090	0.000	0.910
4	2020-06-03	Trail partners Ethereum and Ethereum Classic a...	0.1280	0.102	0.116	0.782

In [499]:

```
# Describe the Bitcoin Sentiment
btc_df.describe()
```

Out[499]:

	compound	positive	negative	neutral
count	20.000000	20.000000	20.000000	20.00000
mean	-0.032275	0.035400	0.052850	0.91175
std	0.269375	0.088616	0.089001	0.10879
min	-0.542300	0.000000	0.000000	0.69600
25%	-0.202300	0.000000	0.000000	0.81200
50%	0.000000	0.000000	0.000000	1.00000
75%	0.000000	0.000000	0.114000	1.00000
max	0.585900	0.304000	0.280000	1.00000

In [500]:

```
# Describe the Ethereum Sentiment
eth_df.describe()
```

Out[500]:

	compound	positive	negative	neutral
count	20.000000	20.000000	20.000000	20.000000
mean	0.292790	0.084400	0.022600	0.893000
std	0.360966	0.059504	0.041826	0.057333
min	-0.670500	0.000000	0.000000	0.782000
25%	0.096000	0.049500	0.000000	0.851750
50%	0.334550	0.089500	0.000000	0.888000
75%	0.567900	0.128500	0.040750	0.931500
max	0.700300	0.178000	0.143000	1.000000

Questions:

Q: Which coin had the highest mean positive score?

- A: Ethereum
Q: Which coin had the highest compound score?

- A: Ethereum
Q. Which coin had the highest positive score?

- A: Bitcoin

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 [501]:

```python
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

import nltk
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     /Users/franciscolopezbarrera/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[501]:

True

In [502]:

```python
lemmatizer = WordNetLemmatizer()
```

In [503]:

```python
article = btc_headlines["articles"][0]['content']
```

In [504]:

```python
sentence = sent_tokenize(article)[1]
print(sentence)
```

If that doesnt seem masturbatory enough, r… [+2657 chars]

In [505]:

```
words = word_tokenize(sentence)
print(words)
```

```
['If', 'that', 'doesnt', 'seem', 'masturbatory', 'enough', ',', 'r…'
, '[', '+2657', 'chars', ']']
```

In [506]:

```
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you',
"you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself',
'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her',
'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 't
heir', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this
', 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'w
ere', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', '
does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', '
because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', '
about', 'against', 'between', 'into', 'through', 'during', 'before',
'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out',
'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'h
ere', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both',
'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor',
'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't',
'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now',
'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'coul
dn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn
't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mi
ghtn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "
shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren't
", 'won', "won't", 'wouldn', "wouldn't"]
```

In [507]:

```
from collections import Counter
Counter(" ".join(btc_df["text"]).split()).most_common(100)
```

Out[507]:

```
[('Satoshi', 14),
 ('Nakaboto:', 14),
 (''Bitcoin', 8),
 ('Bitcoin', 6),
 ('in', 6),
 ('the', 5),
 ('to', 5),
 ('a', 3),
```

```
('Goldman', 2),
('bitcoin', 2),
('is', 2),
('an', 2),
('(BTCUSD)', 2),
('on', 2),
('$10,000', 2),
('as', 2),
('one', 2),
('of', 2),
('back', 2),
('its', 2),
('fees', 2),
('God', 1),
('Help', 1),
('Us,', 1),
('Winklevoss', 1),
('Twins', 1),
('Are', 1),
('Co-Producing', 1),
('Movie', 1),
('About', 1),
('Their', 1),
('Journey', 1),
('How', 1),
('Buy', 1),
('Burner', 1),
('Phone', 1),
('Sachs', 1),
('lists', 1),
('5', 1),
('reasons', 1),
('why', 1),
("'not", 1),
('asset', 1),
("class',", 1),
('nor', 1),
("'a", 1),
('suitable', 1),
("investment'", 1),
("'Get", 1),
("train':", 1),
('Famed', 1),
('investor', 1),
('Michael', 1),
('Novogratz', 1),
('says', 1),
('will', 1),
('soon', 1),
('break', 1),
```

```
        ('threshold', 1),
        ('US', 1),
        ('turmoil', 1),
        ('spreads', 1),
        ('continues', 1),
        ('downtrend', 1),
        ('with', 1),
        ('5%', 1),
        ('loss', 1),
        ('day'', 1),
        (''Visa', 1),
        ('approves', 1),
        ('launch', 1),
        ('crypto', 1),
        ('debit', 1),
        ('card', 1),
        ('Europe'', 1),
        ('8', 1),
        ('Reasons', 1),
        ('Why', 1),
        ('This', 1),
        ('Could', 1),
        ('Be', 1),
        ('Time', 1),
        ('Take', 1),
        ('Seriously', 1),
        ('Emerges', 1),
        ('New', 1),
        ('Standard', 1),
        ('Post-COVID', 1),
        ('Economy?', 1),
        ('bad', 1),
        ('investment', 1),
        ('according', 1),
        ('Sachs'', 1),
        (''S&P', 1),
        ('500', 1),
        ('outperformed', 1),
        ('over', 1),
        ('past', 1),
        ('year'', 1),
        ('whale', 1)]
```

In [508]:

```python
from collections import Counter
Counter(" ".join(eth_df["text"]).split()).most_common(100)
```

Out[508]:

[('the', 29),

```
('a', 22),
('chars]', 20),
('to', 14),
('of', 12),
('and', 11),
('for', 10),
('Ethereum', 10),
('on', 9),
('its', 8),
('an', 7),
('is', 7),
('The', 7),
('be', 6),
('has', 6),
('in', 5),
('cryptocurrency', 4),
('that', 4),
('crypto', 4),
('out', 4),
('user', 3),
('who', 3),
('transaction', 3),
('new', 3),
('can', 3),
('data', 3),
('by', 3),
('this', 3),
('are', 3),
('it', 3),
('time', 3),
('after', 3),
('more', 3),
('than', 3),
('with', 3),
('being', 2),
('$2.5', 2),
('million', 2),
('fee', 2),
('…', 2),
('announced', 2),
('from', 2),
('CPU', 2),
('other', 2),
('a…', 2),
('Ethereum,', 2),
('seen', 2),
('over', 2),
('United', 2),
('worth', 2),
('at', 2),
```

```
('days', 2),
('above', 2),
('key', 2),
('level', 2),
('resistance', 2),
('against', 2),
('second', 2),
('It', 2),
('now', 2),
('up', 2),
('high', 2),
('company', 2),
('both', 2),
('assets', 2),
('have', 2),
('Ethereum-based', 2),
('one', 2),
('scaling', 2),
('Nexus', 2),
('pool', 2),
('Foundation', 2),
('application', 2),
('transactions', 2),
('Imagine', 1),
('sending', 1),
('$133', 1),
('friend', 1),
('charged', 1),
('fee.', 1),
('That', 1),
('seems', 1),
('case', 1),
('paid', 1),
('during', 1),
('trade', 1),
('Wednesday', 1),
('morning', 1),
('[+2684', 1),
('Intel', 1),
('CrossTalk,', 1),
('side-channel', 1),
('attack', 1),
('leak', 1),
('buffers.', 1),
('It's', 1),
('same', 1),
('story', 1),
('we've', 1),
('heard', 1)]
```

In [512]:

```python
sw = set(stopwords.words('english'))
```

In [513]:

```python
# We can define our own list of stopwords to add to the default nltk words
sw_addon = {'chars','noon', 'whats', 'this', 'that', 'writes', 'fucking', 'Satoshi', 'Nakaboto'}
```

In [514]:

```python
# Complete the tokenizer function
def tokenizer(text):
    """Tokenizes text."""
    sw = set(stopwords.words('english'))
    sw_addons = {'chars','noon', 'whats', 'this', 'that', 'writes', 'fucking', 'Satoshi', 'Nakaboto'}
    regex = re.compile("[^a-zA-Z ]")
    re_clean = regex.sub('', article)

    # Lemmatize Words into root words
    words = word_tokenize(re_clean)
    output = [word.lower() for word in words if word.lower() not in sw.union(sw_addons)]
    output = [lemmatizer.lemmatize(word) for word in output]

    return output
```

In [515]:

```
btc_tokens = []
for article in btc_df['text']:
    btc_tokens.append(tokenizer(article))
btc_df['tokens'] = btc_tokens
btc_df.head()
```

Out[515]:

	date	text	compound	positive	negative	neutral	tokens
0	2020-06-09	God Help Us, the Winklevoss Twins Are Co-Produ...	0.5859	0.304	0.000	0.696	[god, help, u, winklevoss, twin, coproducing, ...]
1	2020-06-04	How to Buy a Burner Phone	0.0000	0.000	0.000	1.000	[buy, burner, phone]
2	2020-05-28	Goldman Sachs lists 5 reasons why bitcoin is '...	-0.2755	0.000	0.123	0.877	[goldman, sachs, list, reason, bitcoin, asset,...]
3	2020-06-02	'Get on the train': Famed investor Michael Nov...	-0.3612	0.000	0.111	0.889	[get, train, famed, investor, michael, novogra...]
4	2020-05-25	Satoshi Nakaboto: 'Bitcoin continues downtrend...	-0.3182	0.000	0.187	0.813	[satoshi, nakaboto, bitcoin, continues, downtr...]

In [516]:

```python
# Create a new tokens column for ethereum
eth_tokens = []
for article in eth_df['text']:
    eth_tokens.append(tokenizer(article))
eth_df['tokens'] = eth_tokens
eth_df.head()
```

Out[516]:

	date	text	compound	positive	negative	neutral	tokens
0	2020-06-10	Imagine sending $133 to a friend and being cha...	0.3400	0.089	0.050	0.861	[imagine, sending, friend, charged, million, f...
1	2020-06-12	Intel announced CrossTalk, a new side-channel ...	-0.6705	0.000	0.143	0.857	[intel, announced, crosstalk, new, sidechannel...
2	2020-06-10	POLAND - 2020/03/13: In this photo illustratio...	0.0000	0.000	0.000	1.000	[poland, photo, illustration, ethereum, crypto...
3	2020-06-11	Photo taken in New York City, United States\r\...	0.3291	0.090	0.000	0.910	[photo, taken, new, york, city, united, states...
4	2020-06-03	Trail partners Ethereum and Ethereum Classic a...	0.1280	0.102	0.116	0.782	[trail, partner, ethereum, ethereum, classic, ...

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 [517]:

```python
from collections import import Counter
from nltk import ngrams
```

In [518]:

```python
bigram_counts = [Counter(ngrams(tokens, n=2)) for tokens in btc_df.tokens]
dict(bigram_counts[0].most_common(10))
```

Out[518]:

```
{('god', 'help'): 1,
 ('help', 'u'): 1,
 ('u', 'winklevoss'): 1,
 ('winklevoss', 'twin'): 1,
 ('twin', 'coproducing'): 1,
 ('coproducing', 'movie'): 1,
 ('movie', 'bitcoin'): 1,
 ('bitcoin', 'journey'): 1}
```

In [519]:

```python
# Generate the Ethereum N-grams where N=2
eth_bigram_counts = [Counter(ngrams(tokens, n=2)) for tokens in eth_df.tokens]
dict(eth_bigram_counts[0].most_common(10))
```

Out[519]:

```
{('imagine', 'sending'): 1,
 ('sending', 'friend'): 1,
 ('friend', 'charged'): 1,
 ('charged', 'million'): 1,
 ('million', 'fee'): 1,
 ('fee', 'seems'): 1,
 ('seems', 'case'): 1,
 ('case', 'cryptocurrency'): 1,
 ('cryptocurrency', 'user'): 1,
 ('user', 'paid'): 1}
```

In [520]:

```python
def token_count(tokens, N=10):

    return Counter(tokens).most_common(N)
```

In [521]:

```python
# Get the top 10 words for Bitcoin
btc_big_string = []
for tokens in btc_df.tokens:
    btc_big_string += tokens
btc_token_count = token_count(big_string)
btc_token_count
```

Out[521]:

```
[('bitcoin', 117),
 ('btc', 47),
 ('satoshi', 43),
 ('nakaboto', 38),
 ('today', 21),
 ('robot', 19),
 ('colleague', 19),
 ('writes', 19),
 ('every', 19),
 ('fucking', 19)]
```

In [522]:

```python
# Get the top 10 words for Ethereum
eth_big_string = []
for tokens in eth_df.tokens:
    eth_big_string += tokens
eth_token_count = token_count(eth_big_string)
eth_token_count
```

Out[522]:

```
[('ethereum', 12),
 ('cryptocurrency', 5),
 ('transaction', 5),
 ('crypto', 5),
 ('user', 4),
 ('new', 4),
 ('million', 3),
 ('fee', 3),
 ('data', 3),
 ('level', 3)]
```

Word Clouds

In this section, you will generate word clouds for each coin to summarize the news for each coin
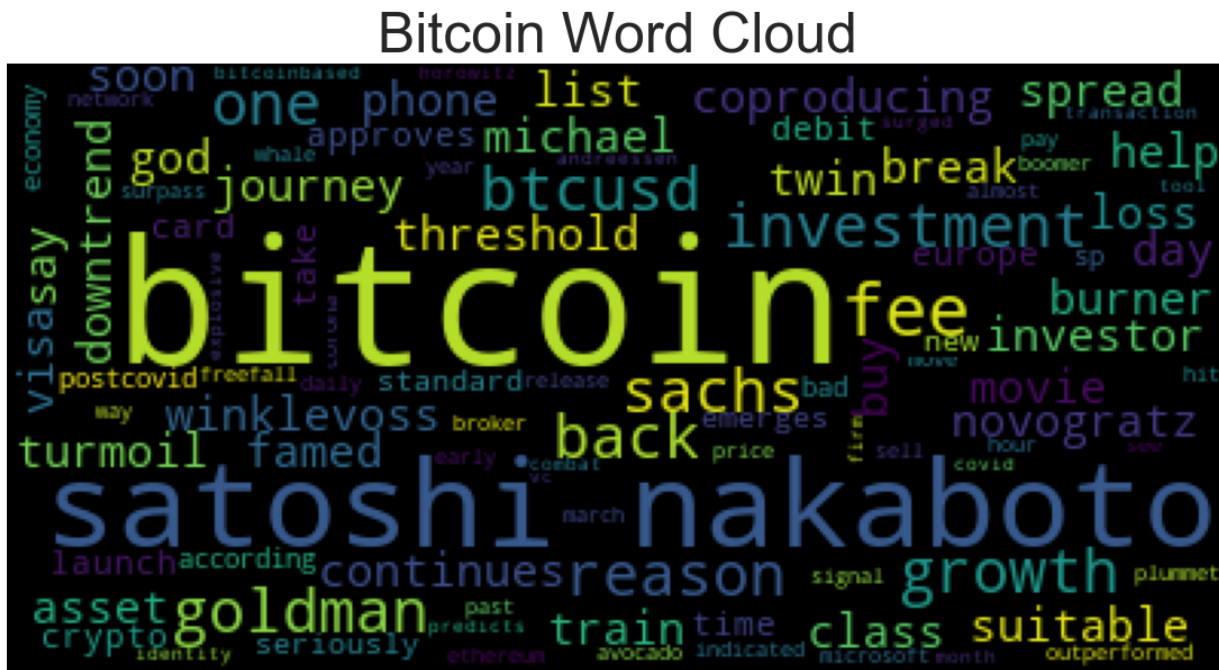
In [523]:

```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 [524]:

```python
# Generate the Bitcoin word cloud

wc = WordCloud().generate(' '.join(btc_big_string))

plt.imshow(wc)
plt.axis('off')
plt.title('Bitcoin Word Cloud', fontsize=50)
```
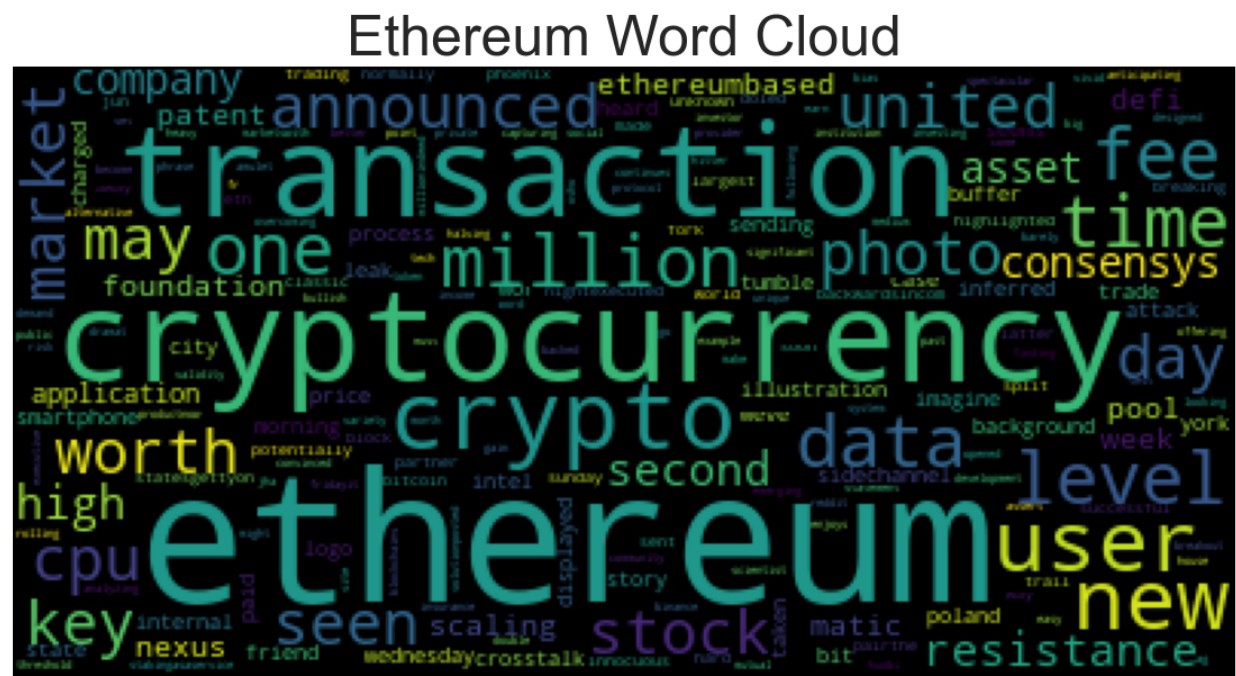
Out[524]:

```
Text(0.5, 1.0, 'Bitcoin Word Cloud')
```

Bitcoin Word Cloud

In [525]:

```python
# Generate the Ethereum word cloud
wc = WordCloud().generate(' '.join(eth_big_string))
plt.imshow(wc)
plt.axis('off')
plt.title('Ethereum Word Cloud', fontsize=50)
```

Out[525]:

```
Text(0.5, 1.0, 'Ethereum Word Cloud')
```


Ethereum Word Cloud

Named Entity Recognition

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

In [526]:

```python
import spacy
from spacy import displacy
```

In [527]:

```python
# Optional - download a language model for SpaCy
# !python -m spacy download en_core_web_sm
```

In [528]:

```
# Load the spaCy model
nlp = spacy.load('en_core_web_sm')
```

Bitcoin NER

In [556]:

```
# Concatenate all of the bitcoin text together
all_btc=[]

for text in btc_df['text']:
    all_btc.append(text)
all_btc = "".join(all_btc)
all_btc
```

Out[556]:

"God Help Us, the Winklevoss Twins Are Co-Producing a Movie About Th
eir Bitcoin JourneyHow to Buy a Burner PhoneGoldman Sachs lists 5 re
asons why bitcoin is 'not an asset class', nor 'a suitable investmen
t' (BTCUSD)'Get on the train': Famed investor Michael Novogratz says
bitcoin will soon break the $10,000 threshold as US turmoil spreads
(BTCUSD)Satoshi Nakaboto: 'Bitcoin continues downtrend with 5% loss
in one day'Satoshi Nakaboto: 'Visa approves launch of crypto debit c
ard in Europe'8 Reasons Why This Could Be the Time to Take Bitcoin S
eriouslyBitcoin Emerges as New Standard in Post-COVID Economy?Satosh
i Nakaboto: 'Bitcoin is a bad investment according to Goldman Sachs'
Satoshi Nakaboto: 'S&P 500 outperformed Bitcoin over the past year'S
atoshi Nakaboto: 'Bitcoin whale moves $400M, pays $2.50 fee'Satoshi
Nakaboto: 'Avocado prices indicated sell signal for Bitcoin in early
March'Satoshi Nakaboto: 'Bitcoin plummets almost $700 in an hour, ba
ck below $10K'Satoshi Nakaboto: 'Bitcoin broker sees 77% of its 2020
growth from boomers'Satoshi Nakaboto: 'Bitcoin on its way back to $1
0,000'Satoshi Nakaboto: 'Bitcoin hits $10,000 again after corona fre
efall'Satoshi Nakaboto: 'Daily Ethereum network fees surpass Bitcoin
's'Satoshi Nakaboto: 'Microsoft releases Bitcoin-based identity tool
to combat COVID-19'Satoshi Nakaboto: 'Bitcoin transaction fees surge
d 800% in one month'Satoshi Nakaboto: 'VC firm Andreessen Horowitz p
redicts explosive Bitcoin growth'"

In [557]:

```python
# Run the NER processor on all of the text
btc_ner = nlp(all_btc)

# Add a title to the document
btc_ner.user_data["title"] = "Bitcoin NER"
```

In [558]:

```python
# Render the doc = nlp(all_bitcoin)
displacy.render(btc_ner, style='ent')
```

Bitcoin NER

God Help Us, the Winklevoss Twins Are Co-Producing a Movie About Their Bitcoin **ORG** JourneyHow to Buy a Burner PhoneGoldman Sachs lists 5 **CARDINAL** reasons why bitcoin **GPE** is 'not an asset class', nor 'a suitable investment' (BTCUSD)'Get on the train': Famed investor Michael Novogratz **PERSON** says bitcoin **PERSON** will soon break the $ 10,000 **MONEY** threshold as US **GPE** turmoil spreads (BTCUSD)Satoshi Nakaboto: ' Bitcoin **PERSON** continues downtrend with 5% **PERCENT** loss in one **CARDINAL** day'Satoshi Nakaboto: ' Visa **ORG** approves launch of crypto debit card in Europe'8 Reasons Why This Could Be the Time to Take Bitcoin SeriouslyBitcoin Emerges as New Standard in Post-COVID Economy?Satoshi Nakaboto: ' Bitcoin **PERSON** is a bad investment according to Goldman Sachs'Satoshi Nakaboto **ORG** : 'S&P 500 outperformed Bitcoin **GPE** over the past year'Satoshi Nakaboto: ' Bitcoin **GPE** whale moves $ 400 **MONEY** M, pays $ 2.50 **MONEY** fee'Satoshi Nakaboto: ' Avocado **ORG** prices indicated sell signal for Bitcoin **GPE** in early March'Satoshi Nakaboto: ' Bitcoin **GPE** plummets almost $700 **MONEY** in an hour **TIME** , back below $10K'Satoshi Nakaboto: ' Bitcoin **GPE** broker sees 77% **PERCENT** of its 2020 **DATE** growth from boomers'Satoshi Nakaboto: ' Bitcoin **GPE** on its way back to $10,000'Satoshi Nakaboto: ' Bitcoin **PERSON** hits $ 10,000 **MONEY** again after corona freefall'Satoshi Nakaboto: ' Daily Ethereum **DATE** network fees surpass Bitcoin's'Satoshi Nakaboto: ' Microsoft **ORG** releases Bitcoin **GPE** -based identity tool to combat COVID-19'Satoshi Nakaboto **PERSON** : ' Bitcoin **GPE** transaction fees surged 800% **PERCENT** in one **CARDINAL** month'Satoshi Nakaboto: ' VC **ORG** firm Andreessen Horowitz **PERSON** predicts explosive Bitcoin **NORP** growth'

In [568]:

```python
# List all Entities
print([(ent.text, ent.label_) for ent in btc_ner.ents])
```

[('the Winklevoss Twins Are Co-Producing a Movie About Their Bitcoin
', 'ORG'), ('5', 'CARDINAL'), ('bitcoin', 'GPE'), ('Michael Novograt
z', 'PERSON'), ('bitcoin', 'PERSON'), ('10,000', 'MONEY'), ('US', 'G
PE'), ('Bitcoin', 'PERSON'), ('5%', 'PERCENT'), ('one', 'CARDINAL'),
('Visa', 'ORG'), ('Bitcoin', 'PERSON'), ('Goldman Sachs'Satoshi Naka
boto', 'ORG'), ('Bitcoin', 'GPE'), ('Bitcoin', 'GPE'), ('400', 'MONE
Y'), ('2.50', 'MONEY'), ('Avocado', 'ORG'), ('Bitcoin', 'GPE'), ('Bi
tcoin', 'GPE'), ('almost $700', 'MONEY'), ('an hour', 'TIME'), ('Bit
coin', 'GPE'), ('77%', 'PERCENT'), ('2020', 'DATE'), ('Bitcoin', 'GP
E'), ('Bitcoin', 'PERSON'), ('10,000', 'MONEY'), ('Daily Ethereum',
'DATE'), ('Microsoft', 'ORG'), ('Bitcoin', 'GPE'), ('Nakaboto', 'PER
SON'), ('Bitcoin', 'GPE'), ('800%', 'PERCENT'), ('one', 'CARDINAL'),
('VC', 'ORG'), ('Andreessen Horowitz', 'PERSON'), ('Bitcoin', 'NORP'
)]

Ethereum NER

In [569]:

```python
# Concatenate all of the bitcoin text together
all_eth=[]

for text in eth_df['text']:
    all_eth.append(text)
all_eth = "".join(all_eth)
all_eth
```

Out[569]:

'Imagine sending $133 to a friend and being charged a $2.5 million f
ee. That seems to be the case for a cryptocurrency user who paid a $
2.5 million transaction fee during a trade on Wednesday morning … [+
2684 chars]Intel announced CrossTalk, a new side-channel attack that
can leak data from CPU buffers. It's the same story we've heard befo
re. Bits of internal CPU state can be inferred by other processes. T
his a… [+5744 chars]POLAND – 2020/03/13: In this photo illustration
an Ethereum, cryptocurrency logo seen displayed on a ... [+] smartph
one. Stock market prices in the background as stock markets tumble a
ll over the wor… [+2937 chars]Photo taken in New York City, United S
tates\r\nGetty\r\nOn Jun. 10, an unknown crypto user sent $130 worth

of Ethereum (ETH). Such a transaction would normally be innocuous, but the user doled out 10,668… [+2505 chars]Trail partners Ethereum and Ethereum Classic are soon to split ways, highlighted by the latters successful Phoenix hard fork Sunday night.\r\nExecuted at block 10,500,839 on May 31, the backwards-incom… [+4650 chars]Ethereum is potentially just days away from breaking out above a key level of resistance against its Bitcoin trading pair.\r\nThe worlds second largest cryptocurrency has made a significant 22% gain ag… [+2358 chars]Ethereum continues to assert a bullish bias following its spectacular breakout above the $219 level of resistance on Friday.\r\nIt is now up against the $247 level, which was the high before the dramat… [+2364 chars]Ethereum development house ConsenSys is being backed by heavy hitters like Binance and Huobi to test its new staking-as-a-service offering, designed to make it easy for institutions to earn income fr… [+6404 chars]Osho Jha is an investor, data scientist and tech company executive who enjoys finding and analyzing unique data sets for investing in both public and private markets.\xa0\r\nWith the halving capturing the… [+6551 chars]Crypto assets have become a vivid example of the statement What goes up, must come down and time after time we are convinced of the validity of this phrase. Overcoming the threshold could have opened… [+2372 chars]Reddit is anticipating big demand for its Ethereum-based Community Points system only one month after rolling out the product.\r\nNow the social media site is looking for a scaling solution.\r\nPosted in… [+2596 chars]Nexus Mutual, an alternative insurance provider for a variety of Ethereum-based DeFi protocols, has seen its risk pool double over the past 90 days to more than $4 million.\r\nIndeed, Nexus can barely … [+5096 chars]"Amulets" might be a better word than "tokens" for some of the crypto assets emerging on blockchains today.\r\nVersion 2.0 of Uniswap is now live and it is, among other things, an amulet-minting machin… [+7195 chars]Matic, a sidechain scaling solution for Ethereum, began deploying its mainnet on Sunday.\r\nThe India-based Matic Foundation announced via a blog post that it had brought 10 nodes online, the first vol… [+2723 chars]The Brave web browser sells itself on privacy, security and ad-blocking. It also has its own cryptocurrency, the Basic Attention Token.\r\nAs such, it's a favourite with crypto people or ones who don't… [+4471 chars]A mining pool has called time on the wait for an ether whale to reach out after making a transaction with an unusually high fee worth in the millions of dollars last week. \r\nBitfly, the company behin… [+2737 chars]Introduction\r\nThe U.S. Patent and Trademark Office (USPTO) has recently published that Visa (V) has filed a patent application for a digital fiat currency. The application mentions Ethereum (ETH-USD)… [+8793 chars]Its costing more to use Ethereum and that may be because more users are flocking to the platform than ever before, according to one key on-chain metric. Analysts say the growth of both transactions a… [+5218 chars]ConsenSys' new regulatory compliance product will be at hand to analyze transactions in the growing decentralized finance (DeFi) space. \r\nThe Ethereum venture studio said the newly-launched Codefi Co… [+2531 chars]The Ethereum Foundation followed up on its 2019 gift to the

```
United Nations Childrens Fund (UNICEF) this week with a second crypt
ocurrency donation.\xa0\r\nUNICEF crypto portfolio manager Cecilia C
hapiro s… [+5371 chars]'
```

In [570]:

```python
# Run the NER processor on all of the text
eth_ner = nlp(all_eth)

# Add a title to the document
eth_ner.user_data["title"] = "Ethereum NER"
```

In [571]:

```python
# Render the visualization
displacy.render(eth_ner, style='ent')
```

Ethereum NER

Imagine sending $ 133 **MONEY** to a friend and being charged a $2.5 million **MONEY**

fee. That seems to be the case for a cryptocurrency user who paid a $2.5 million **MONEY**

transaction fee during a trade on Wednesday **DATE** morning **TIME** … [+2684

chars]Intel announced CrossTalk **ORG** , a new side-channel attack that can leak data

from CPU buffers. It's the same story we've heard before. Bits of internal CPU state can be

inferred by other processes. This a… [+5744 chars]POLAND - 2020/03/13: In this photo

illustration an Ethereum, cryptocurrency logo seen displayed on a ... [+] smartphone. Stock

market prices in the background as stock markets tumble all over the wor… [+2937

chars]Photo taken in New York City **GPE** , United States **GPE**

Getty

On Jun. 10 **DATE** , an unknown crypto user sent $ 130 **MONEY** worth of Ethereum

(ETH). Such a transaction would normally be innocuous, but the user doled out 10,668

CARDINAL … [+2505 chars]Trail partners Ethereum and Ethereum Classic are soon to split

ways, highlighted by the latters successful Phoenix **GPE** hard fork Sunday night **TIME**

.

Executed at block   10,500,839 **CARDINAL**   on   May 31 **DATE**   , the backwards-incom…

[+4650 chars]Ethereum is potentially   just days **DATE**   away from breaking out above a key

level of resistance against its   Bitcoin **GPE**   trading pair.

The worlds   second **ORDINAL**   largest cryptocurrency has made a significant   22%

PERCENT   gain   ag **ORG**   … [+2358 chars]Ethereum continues to assert a bullish bias

following its spectacular breakout above the $ 219 **MONEY**   level of resistance on   Friday

DATE   .

It is now up against the $ 247 **MONEY**   level, which was the high before the dramat…

[+2364 chars]Ethereum development house   ConsenSys **ORG**   is being backed by heavy

hitters like   Binance **ORG**   and   Huobi **GPE**   to test its new staking-as-a-service offering,

designed to make it easy for institutions to earn income fr… [+6404 chars]Osho   Jha

PERSON   is an investor, data scientist and tech company executive who enjoys finding and

analyzing unique data sets for investing in both public and private markets.

With the halving capturing the… [+6551 chars]Crypto assets have become a vivid example

of the statement What goes up, must come down and time after time we are convinced of

the validity of this phrase. Overcoming the threshold could have opened… [+2372

chars]Reddit is anticipating big demand for its Ethereum-based   Community Points **ORG**

system   only one month **DATE**   after rolling out the product.

Now the social media site is looking for a scaling solution.

Posted in… [+2596 chars]Nexus Mutual, an alternative insurance provider for a variety of

Ethereum-based   DeFi **ORG**   protocols, has seen its risk pool double over   the past 90

days **DATE**   to   more than $4 million **MONEY**   .

Indeed,   Nexus **GPE**   can barely … [+5096 chars]"Amulets" might be a better word than

"tokens" for some of the crypto assets emerging on blockchains   today **DATE**   .

Version 2.0 **LAW** of Uniswap is now live and it is, among other things, an amulet-minting machin… [+7195 chars]Matic, a sidechain scaling solution for Ethereum, began deploying its mainnet on Sunday **DATE** .

The India **GPE** -based Matic Foundation **ORG** announced via a blog post that it had brought 10 **CARDINAL** nodes online, the first **ORDINAL** vol… [+2723 chars]The Brave web browser sells itself on privacy, security and ad-blocking. It also has its own cryptocurrency, the Basic Attention Token **LAW** .

As such, it's a favourite with crypto people or ones who don't… [+4471 chars]A mining pool has called time on the wait for an ether whale to reach out after making a transaction with an unusually high fee worth in the millions of dollars **MONEY** last week **DATE** .

Bitfly, the company behin… [+2737 chars]Introduction

The U.S. Patent and Trademark Office **ORG** (USPTO **ORG**) has recently published that Visa **ORG** (V) has filed a patent application for a digital fiat currency. The application mentions Ethereum (ETH-USD)… [+8793 chars]Its costing more to use Ethereum and that may be because more users are flocking to the platform than ever before, according to one **CARDINAL** key on-chain metric. Analysts say the growth of both transactions a… [+5218 chars]ConsenSys' new regulatory compliance product will be at hand to analyze transactions in the growing decentralized finance (DeFi) space.

The Ethereum venture studio said the newly-launched Codefi Co… **FAC** [+2531 chars]The Ethereum Foundation **ORG** followed up on its 2019 **DATE** gift to the United Nations Childrens Fund **ORG** (UNICEF **ORG**) this week **DATE** with a second **ORDINAL** cryptocurrency donation.

UNICEF **ORG** crypto portfolio manager Cecilia Chapiro **PERSON** s… [+5371 chars]

In [572]:

```python
# List all Entities
print([(ent.text, ent.label_) for ent in eth_ner.ents])
```

[('133', 'MONEY'), ('$2.5 million', 'MONEY'), ('$2.5 million', 'MONEY'), ('Wednesday', 'DATE'), ('morning', 'TIME'), ('CrossTalk', 'ORG'), ('New York City', 'GPE'), ('United States', 'GPE'), ('Jun. 10', 'DATE'), ('130', 'MONEY'), ('10,668', 'CARDINAL'), ('Phoenix', 'GPE'), ('Sunday night', 'TIME'), ('10,500,839', 'CARDINAL'), ('May 31', 'DATE'), ('just days', 'DATE'), ('Bitcoin', 'GPE'), ('second', 'ORDINAL'), ('22%', 'PERCENT'), ('ag', 'ORG'), ('219', 'MONEY'), ('Friday', 'DATE'), ('247', 'MONEY'), ('ConsenSys', 'ORG'), ('Binance', 'ORG'), ('Huobi', 'GPE'), ('Jha', 'PERSON'), ('Community Points', 'ORG'), ('only one month', 'DATE'), ('DeFi', 'ORG'), ('the past 90 days', 'DATE'), ('more than $4 million', 'MONEY'), ('Nexus', 'GPE'), ('today', 'DATE'), ('Version 2.0', 'LAW'), ('Sunday', 'DATE'), ('India', 'GPE'), ('Matic Foundation', 'ORG'), ('10', 'CARDINAL'), ('first', 'ORDINAL'), ('the Basic Attention Token', 'LAW'), ('the millions of dollars', 'MONEY'), ('last week', 'DATE'), ('The U.S. Patent and Trademark Office', 'ORG'), ('USPTO', 'ORG'), ('Visa', 'ORG'), ('one', 'CARDINAL'), ('Codefi Co…', 'FAC'), ('+2531 chars]The Ethereum Foundation', 'ORG'), ('2019', 'DATE'), ('the United Nations Childrens Fund', 'ORG'), ('UNICEF', 'ORG'), ('this week', 'DATE'), ('second', 'ORDINAL'), ('UNICEF', 'ORG'), ('Cecilia Chapiro', 'PERSON')]

In [ ]: