**Project Group 17**

**Members:**

Hadrien Nguyen - 1001636184

Jay Patel - 1001870971

Sai Charan Damarapati Subramanyam - 1002026508

**Professor :**  Rama Krishna Koganti

CSE 5301-003

04-30-2022

**Twitter, NFTs and the Power of Social Media: An Exploration of the Impact of Virtual Engagement on Virtual Commodities**

The discourse around cryptocurrencies in the past decade has been contentious and misinformed for the most part. The fact that it is a technology that few understand makes it more difficult to talk about it. What started with bitcoin in 2009 is now a multi billion dollar economy with an increasing number of investors. What we are looking for in this research is how this virtual currency impacts social media like Twitter and vice-versa. As a lot of the conversations nowadays happen online, it is important to know how to react to misinformation and viral topics.

# **Twitter**

Twitter was launched in July of 2006 as an application in both the social media space (which includes other applications/websites such as Instagram, Facebook, LinkedIn and others) and microblogging. Microblogging is a medium that allows for smaller and more frequent updates than blogging. Twitter allows users to post messages publicly (which are referred to as “tweets”) with a maximum length of 140 characters.

In November of 2017 that limit was doubled to 280 characters. In addition, users can add “hashtags” to a tweet, which is the “#” symbol followed by a consecutive string of characters. This is used to identify the topic or theme of a tweet and to make them searchable. This is used later when we collect the tweets in the data section.

Since its launch in 2006 Twitter has grown rapidly in popularity and into a very rich source of data on how people feel about nearly any given topic. With the ability to see when a tweet was posted it is also possible to tell how those feelings change over time. This makes Twitter an excellent resource to collect text data on a topic like cryptocurrencies to explore the possible relationships between that and prices.

# Cryptocurrency

​​A cryptocurrency is an encrypted data string that denotes a unit of currency. It is monitored and organized by a peer-to-peer network called a blockchain, which also serves as a secure ledger of transactions, e.g., buying, selling, and transferring. Unlike physical money, cryptocurrencies are decentralized, which means they are not issued by governments or other financial institutions.

Cryptocurrencies are created (and secured) through cryptographic algorithms that are maintained and confirmed in a process called mining, where a network of computers or specialized hardware such as application-specific integrated circuits (ASICs) process and validate the transactions. The process incentivizes the miners who run the network with the cryptocurrency. Bitcoin, Ether, Litecoin are popular cryptocurrencies.

A non-fungible token (NFT) is a non-interchangeable unit of data stored on the blockchain that can be sold and traded.Types of NFT data units may be associated with digital files such as photos, videos, and audio. Because each token is uniquely identifiable, NFTs differ from most cryptocurrencies, such as Bitcoin which are fungible.

At a very high level, most NFTs are part of the Ethereum blockchain. Ethereum is a cryptocurrency, like bitcoin or dogecoin, but its blockchain also supports these NFTs, which store extra information that makes them work differently from, say, an ETH coin. It is worth noting that other blockchains can implement their own versions of NFTs. NFTs can really be anything digital (such as drawings, music, your brain downloaded and turned into an AI), but a lot of the current excitement is around using the tech to sell digital art.

# **Asking Twitter**

The first step to the project was getting the data from Twitter through the Twitter APIv2. To do so, we needed to determine exactly what we were looking for through a query. A query has multiple parameters: the text of the query (that determines what filters the search is going to use), the ‘tweet.fields’ (the fields that are returned by the API), the ‘max\_results’, and ‘next\_token.’ The query parameters that we used are:

query\_params = {

**'query'**: '(NFT OR "non-fungible token") -gn -gm -tag -DROP -WINNER -giveaway -claim -is:retweet -is:reply -has:links',

**'tweet.fields'**: 'id,text,author\_id,created\_at,public\_metrics,source',

**'max\_results'**: 100,

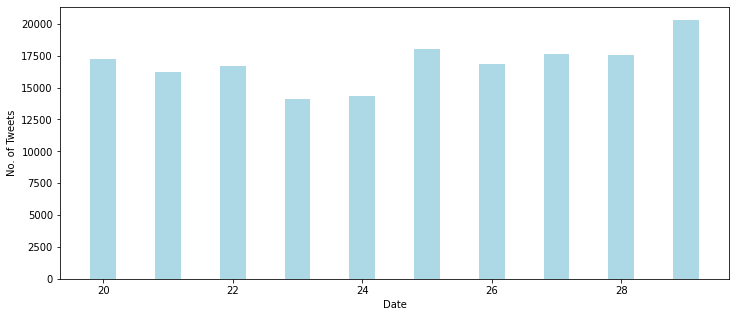
**'next\_token'**:'[ ]'}

The query text is separated into several parts. The first part: (NFT OR "non-fungible token") means that we will only get tweets that contain either those two words. The second part: -gn -gm -tag -DROP -WINNER -giveaway -claim means that we will ignore every tweet that contains each of these words. This was decided by inspection in order to limit the data, for example, there are a lot of bots that send daily, sometimes hourly tweets to “giveaway” NFTs. We did not determine that these tweets were relevant for our purposes. The last part is similar, though it takes out tweets that are replies and that have links instead of looking for keywords. Once again, this is to reduce the amount of “noise” we end up getting, hopefully filtering out some bots.

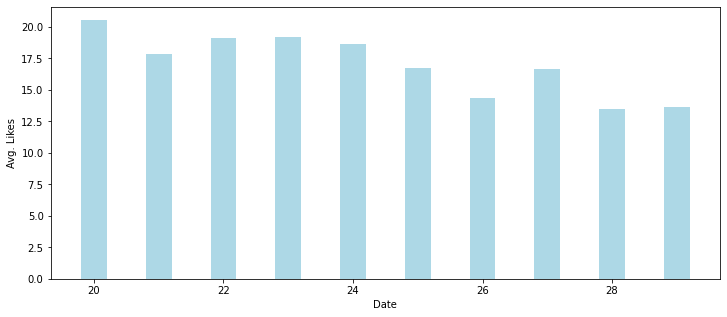
We are using official code from twitter in python in order to query, though modified for our purposes (<https://github.com/twitterdev/Twitter-API-v2-sample-code>). The limitations of using the twitter API is that it only allows for the last 7 days of data. After getting the data we are storing it in a csv file and then plotting some graphs that show some relation with the sales of NFTs.

Upon analyzing the data we get the following results:

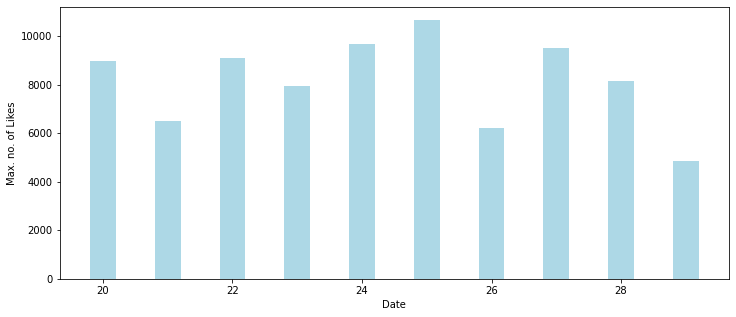
1. No. of Tweets per day :- The graph shows the number of tweets posted on Twitter related to the word NFTs per day.



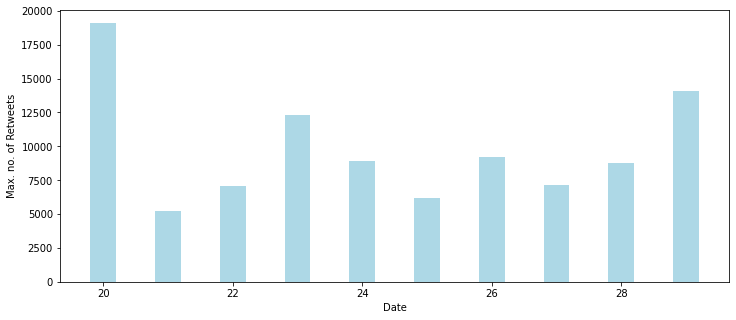
1. Average likes : The Image represents the average number of Likes on tweets related to NFTs per day.



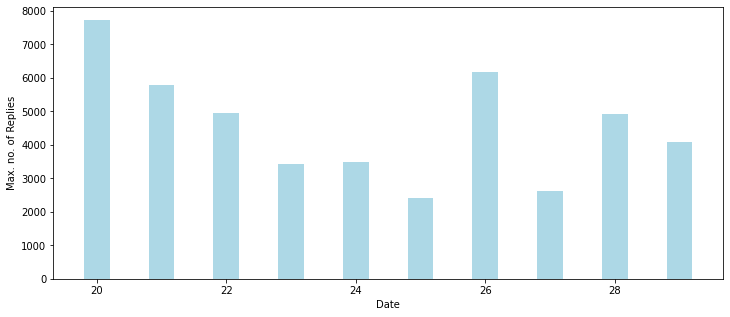
1. Maximum no. of Likes : The graph represents the maximum number of Likes received on tweets related to NFTs per day



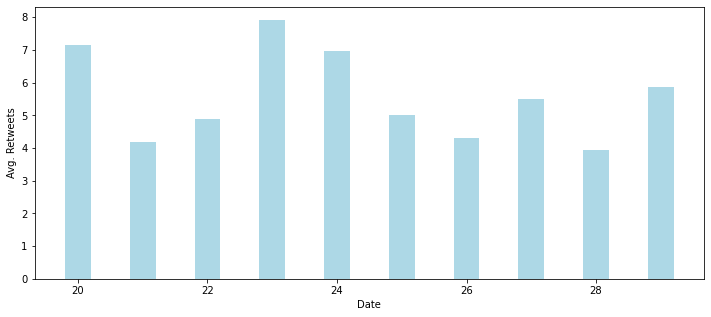
1. Maximum no. of Retweets : The graph shows the maximum number of Retweets related to NFTs per day.



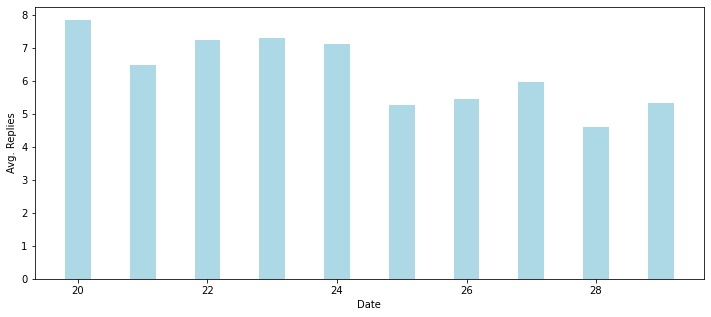
1. Maximum no. of Replies : The graph shows the maximum number of Replies to tweets related to the word NFTs per day.



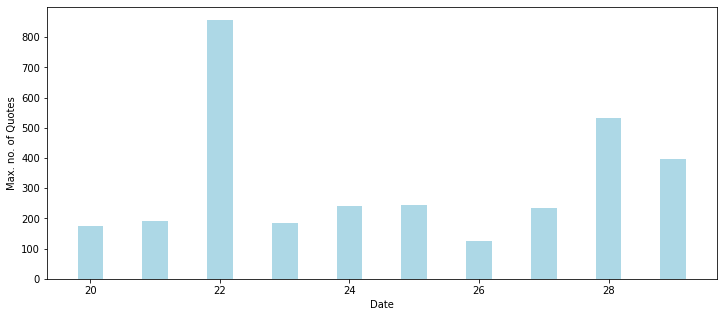
1. Average Retweets :- The graph shows the average number of Retweets related to the word NFTs per day.



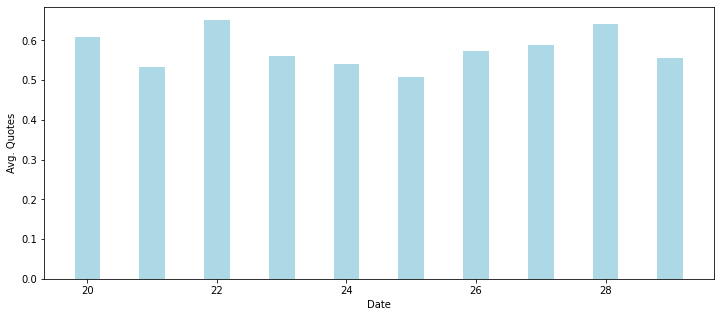
1. Average Replies :- The graph shows the average number of Replies to tweets related to the word NFTs per day.



1. Maximum Quotes :- The graph represents the maximum number of Quotes on Twitter related to the word NFTs per day

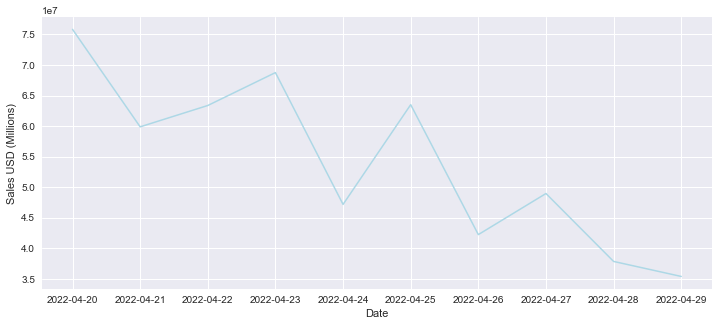


1. Average Quotes :- The graph represents the average number of Quotes on Twitter related to the word NFTs per day.



Real-Time sales of NFTs and its Relation with Twitter :

The graph shows the Sales of NFTs in USD per day.

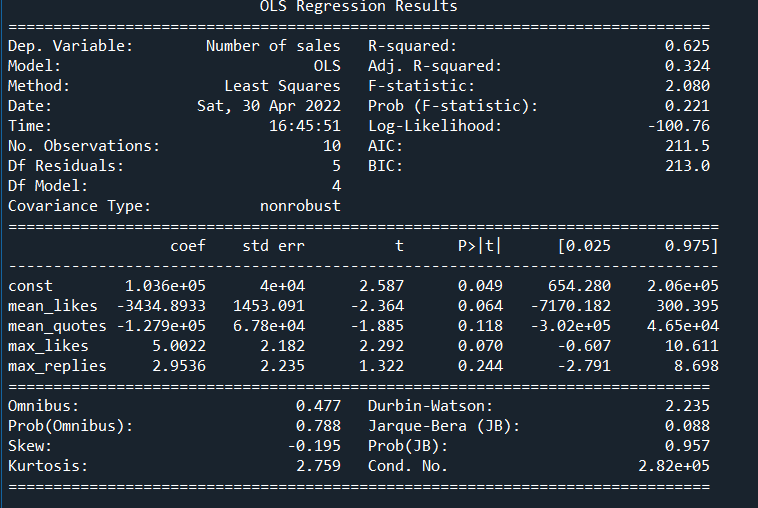


We have a total result of 221,162 tweets across 10 days. We have filtered all the tweets with no engagement at all to make our graphs more meaningful and ended up with 166,102 tweets.

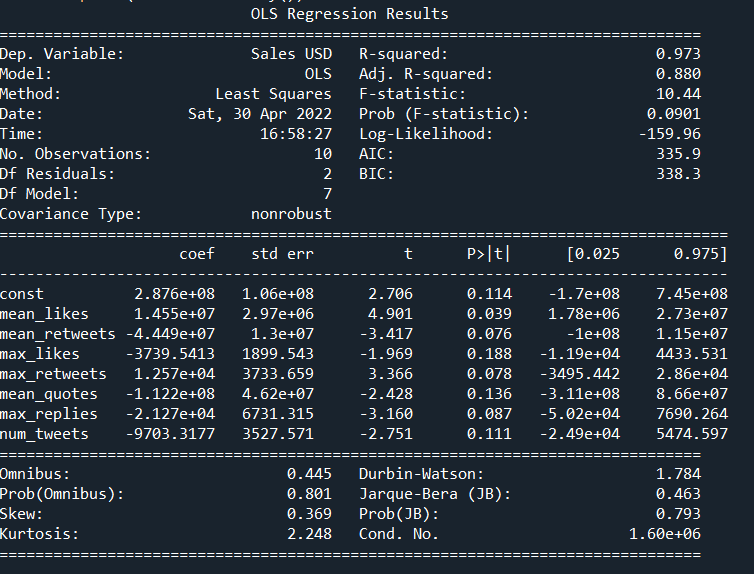
# **Linear Regression**

Our analysis is based on a multivariate linear regression. We are going to compute three models: the Number of Sales in a day, the Sales in US Dollars in a day,and the number of tweets with engagements in a day. In order to choose the variable, we are choosing Backward elimination (Baron 393), meaning we remove the least significant predictor at each iteration until all predictors are statistically relevant.

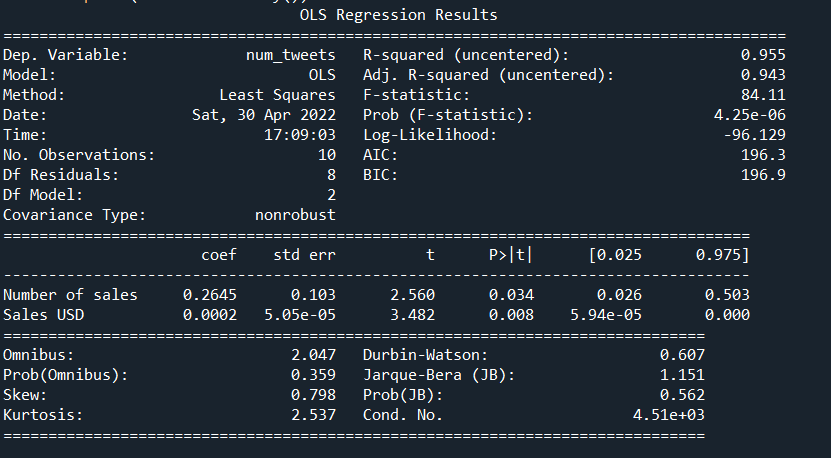
For the first model, we are trying to predict the number of sales of NFTs on a certain day. The original variables were 'mean\_likes', 'mean\_retweets', 'mean\_quotes', 'mean\_replies', 'max\_likes', 'max\_retweets', 'max\_quotes','max\_replies', and 'num\_tweets.' After Backwards elimination, we have the best fit being 'mean\_likes', 'mean\_quotes', 'max\_likes',and 'max\_replies.' The results are inconclusive (p-value of 0.221), so we determine that there is no meaningful relationship between tweet engagement and Number of sales.



For the second model, we are trying to predict the total sales in US Dollars on a certain day. The original variables were 'mean\_likes', 'mean\_retweets', 'mean\_quotes', 'mean\_replies', 'max\_likes', 'max\_retweets', 'max\_quotes','max\_replies', and 'num\_tweets.' After Backwards elimination, we have the best fit being 'mean\_likes', 'mean\_retweets', 'mean\_quotes', 'max\_likes', 'max\_retweets', 'max\_replies', and 'num\_tweets.' The results are conclusive at the 90% level (p-value is 0.0901) and the Adjusted R-squared is high, meaning we can determine that there is a relationship between the variables and the Sales USD.



For the last test, we are trying to predict the number of relevant tweets (with at least one like, retweet, reply, or quote) on a certain day. The initial variables are 'Sales USD', and 'Number of Sales.' We do not remove any of the variables as they are both relevant. The final model is conclusive with a p-value of 4.25\*10^-6. There is a strong correlation between the sale of NFTs and the number of tweets.



# **Conclusion**

Based on the test that we have conducted, there is a relationship between real life sales of NFTs and Twitter engagement, though it is only with certain parameters. It is possible to predict the total sales in US dollars using tweets and their engagement because the community on Twitter is very active and has a lot of stake in sharing NFTs to the world. What we learned from this research is that a difficult issue with this type of research is determining what data is relevant. We made a lot of assumptions on the way because we were limited in time and money. A better exploration of this topic would include sentiment analysis as the most popular tweets are in fact negative when it comes to NFTs. It would be worth looking into how the backlash against NFTs affected the actual sales. If NFTs are harmful, is the best decision to talk a lot about the dangers, make fun of them, or ignore them when it comes to Twitter? As the field grows, these questions will be important when it comes to strategy from both sides.

Work Cited

Baron, Michael. *Probability and Statistics for Computer Scientists*. Chapman and Hall/CRC, 2014.