Group_1
Kelly Zhang, Yinghao Liu, Jingyi Huang

*** Note:
Please open our files in sequence
1. Data cleaning(This notebook will give you the cleaned data to proceed to the next step)
2. Prediction
3. Classification
4. Sentiment Analysis

Our Presentation Link: https://youtu.be/pU8QJ8RZv8c

3 research questions:
1. Prediction: Use models to check if certain cryptos are overpriced or underpriced.
2. Classification: identify trading opportunity to see if certain cryptos deviates from its category's market behavior
3. Sentiment analysis: track twitter sentiments to imply market trends for certain cryptos, market's attention and overall attitude

Datasets :

For question 1: https://data.world/dmikebishop/crypto-investment-fund-directory and https://www.kaggle.com/kaushiksuresh147/bitcoin-tweets In this question we separately analyzed different datasets. This data cleaning is in the same file with sentiment analysis.

For question 2: We are using datasets scraped from
https://coinmarketcap.com/historical/20211001/ (raw data: crypto.csv)
Column description :

Column Name	Column Description
cmc_rank	The general ranking of popularity by Coin Market Cap
volumne_24h	The sum total of actual trades taking place.
percent_change_7d	7-day percentage of change in price
market_cap	The total value of all the coins that have been mined.
'binance-smart-chain'	1: In BSC. 0: Not in BSC.

	Binance Smart Chain —BSC— is a blockchain system from the crypto-trading platform Binance.
'collectibles-nfts'	1:Digitally Collectible. 0: Not digitally collectible. (Collectibles are going digital, using a technology called non-fungible tokens (NFTs). These are digital assets — they can take forms such as photographs, music or video clips.)
'defi'	1: Decentralized. 0: Not decentralized.
'mineable'	1: Mineable. 0: Not mineable.
circulating/total_supply	Circulating supply/Total supply of that cryptocurrency
price	Price of the cryptocurrency

Deleted Columns

Slug	The name of the cryptocurrency
platform	Trading Platform