Analysis on Ethereum Market Cap

2013171033 Oh Han Seung, 2016121114 Cho hyeon Jae, 2019840738 XU KEVIN CHEN

1. Problem

1-1. Predicting and analyzing the main factors which mostly impact the market cap of Ethereum

Our purpose is to make it easy and reliable for those who are interested in Ethereum, the second largest cryptocurrency in the world, to understand and help their decision-making process. There are so many factors that compose the market cap of the cryptocurrency.

2. Method

2-1. Find the reliable data sets

For our project, two types of data sets are needed. One is the transaction data of Ethereum blockchain. This data set should be real time basis. Otherwise, our prediction would be unreliable and less-meaningful. The other dataset is crawled data of Ethereum market capitalization. This data is mainly about the OHLCV data of Ethereum on daily basis. We selected the closed price data as price.

2-2. Preprocessing based on data visualization.

The next step is the preprocessing and cleaning of the data. To understand the overall data structure and relation, we used some data visualizations in order to make data more visible for us. We have included some plots on the next page.

Knowing that too many features can lead to incorrect model, if there is a chance to reduce data dimensions, the data should be processed prior to construction of the model. For dimension reduction, there would be some algorithms – visualization, data summary, pivot table, PCA and etc.. We can choose some methods among them considering which is more appropriate for our dataset. Then, we need to split the data into training, validation and test datasets.

2-3. Further steps: supervised learning, model evaluation, model deployment

The current state of our project is described as it is in this report. From now on, we need to build proper model based on supervised learning. Then, compare several models and pick the best performing model. Lastly, we should deploy the model and evaluate it.

3. Current Result

3-1. Data Set

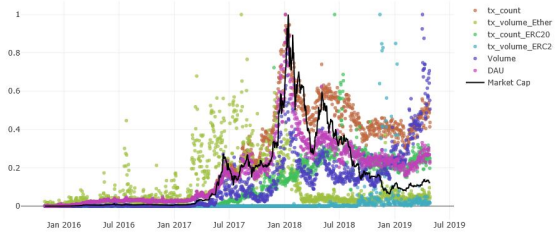
The transaction data of Ethereum blockchain platform has been acquired from BigQuery, and the other from representative real-time cryptocurrency price data platform, ‘CoinMarketCap’.

	tx_date	tx_count	Tx_volume_Ether	tx_count_ERC20	Tx_volume_ERC20	Volume($)	Market Cap($)	DAU	CUM
0	2015-10-30	7941	1.209372e+06	3.0	2.000000e-15	2429200	77401817	7501	7501
1	2015-10-31	7557	2.764179e+05	0.0	0.000000e+00	673892	68163368	7076	14577
2	2015-11-01	6915	1.455244e+05	0.0	0.000000e+00	588913	78530263	6516	21093
3	2015-11-02	6558	3.518903e+05	1.0	5.000000e-16	1145200	73654327	6223	27316
4	2015-11-03	7399	5.951720e+05	30.0	8.001248e-07	1907690	75434114	6973	34289

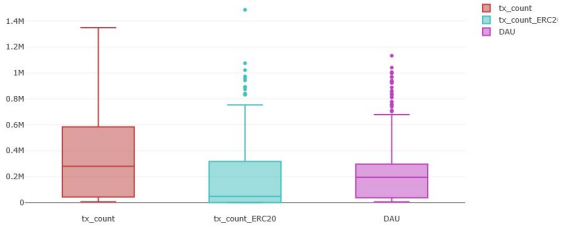
3-2. Preprocessing of the dataset

We explored the dataset by using some basic R functions like *summary()* to understand the data structure. Then, we checked if there are any missing values in the data. After that, we visualized some variables.

```
> summary(data)
                tx_date        tx_count       Tx_volume_Ether    tx_count_ERC20    Tx_volume_ERC20
 2015-10-30:    1   Min.   :   6348   Min.   :   68609   Min.   :      0   Min.   :0.000e+00
 2015-10-31:    1   1st Qu.:  43392   1st Qu.: 1218380   1st Qu.:    989   1st Qu.:6.160e+05
 2015-11-01:    1   Median : 280308   Median : 2142391   Median :  48166   Median :9.622e+09
 2015-11-02:    1   Mean   : 340542   Mean   : 5054436   Mean   : 158450   Mean   :6.509e+60
 2015-11-03:    1   3rd Qu.: 584055   3rd Qu.: 7571817   3rd Qu.: 317205   3rd Qu.:5.392e+59
 2015-11-04:    1   Max.   :1349890   Max.   :46495729   Max.   :1487305   Max.   :4.855e+62
 (Other)   :1271
   Volume...          Market.Cap...          DAU               CUM
 Min.   :1.646e+05   Min.   :5.910e+07   Min.   :   6054   Min.   :     7501
 1st Qu.:1.541e+07   1st Qu.:9.427e+08   1st Qu.:  37692   1st Qu.:  8679379
 Median :6.326e+08   Median :1.343e+10   Median : 195094   Median : 33640591
 Mean   :1.274e+09   Mean   :2.135e+10   Mean   : 196994   Mean   : 79697097
 3rd Qu.:1.943e+09   3rd Qu.:2.956e+10   3rd Qu.: 296771   3rd Qu.:160603513
 Max.   :1.062e+10   Max.   :1.354e+11   Max.   :1133228   Max.   :251561673
```



[figure 1] Scatter plot of features[1]



[figure 2] tx_count, erc20 tx_count, DAU box plot[2]

3-3. Feature selection / Dimension reduction

In this step, we found that there are two variables which may have high correlation between variables. Those variables were 'DAU' and 'tx_count'. The reason behind this was that DAU represents the number of active accounts which have occurred in at least one transaction. However, tx_count contains DAU and also the transaction that has created new contract account. Ethereum platform recognizes the creation of contract account as a 'transaction', which sets 'to_address' as 0 value. We chose 'DAU' instead of 'tx_count' because it is more related to the market cap of Ethereum platform. We also decided to delete 'CUM' feature since it just represents the cumulated values of DAU.

```
> round(cor(data.cor),2)
               tx_count Tx_volume_Ether tx_count_ERC20 Tx_volume_ERC20 Volume...  DAU   CUM
tx_count           1.00            0.12           0.79            0.14     0.79 0.96  0.75
Tx_volume_Ether    0.12            1.00          -0.16           -0.06     0.07 0.24 -0.19
tx_count_ERC20     0.79           -0.16           1.00            0.16     0.61 0.63  0.85
Tx_volume_ERC20    0.14           -0.06           0.16            1.00     0.14 0.09  0.29
Volume...          0.79            0.07           0.61            0.14     1.00 0.75  0.74
DAU                0.96            0.24           0.63            0.09     0.75 1.00  0.59
CUM                0.75           -0.19           0.85            0.29     0.74 0.59  1.00
```

[1] https://plot.ly/~myevertime/26
[2] https://plot.ly/~myevertime/24