ALY6030 : Analysis of Cryptocurrencies

Contents

Introduction

This file contains the R code used to Clean and create visualizations of the dataset.

Loading Libraries

Data Cleaning

Analysis

Overview of Dataset

Initial Analysis is performed to better understand the dataset and variables under observation,the HEAD and TAIL of the dataset, and the structure of dataset is looked-up.

```
df<-raw_data
# Head,Tail and Structure of Dataset
head(df)
```

```
##     id      name symbol num_market_pairs               date_added   max_supply
## 1    1   Bitcoin    BTC             9431 2013-04-28T00:00:00.000Z     21000000
## 2 1027 Ethereum     ETH             5715 2015-08-07T00:00:00.000Z      -100000
## 3  825    Tether   USDT            33404 2015-02-25T00:00:00.000Z      -100000
## 4 3408 USD Coin   USDC             3978 2018-10-08T00:00:00.000Z      -100000
## 5 1839       BNB    BNB              848 2017-07-25T00:00:00.000Z    165116760
## 6   52       XRP    XRP              721 2013-08-04T00:00:00.000Z 100000000000
##   circulating_supply total_supply cmc_rank        price  volume_24h
## 1           19041375     19041375        1 2.975836e+04 33050126432
## 2          120800743    120800743        2 2.015403e+03 20005339261
## 3        75752120651  79713622671        3 9.990728e-01 62815377734
## 4        51002906718  51002906718        4 9.998124e-01  6456376506
## 5          163276975    163276975        5 2.970114e+02  1734215569
## 6        48343101197  99989535142        6 4.164708e-01  1742870711
##   volume_change_24h percent_change_1h percent_change_24h percent_change_7d
## 1           16.9329       -1.11925016        -1.33688771       -9.71992518
## 2           41.6235       -1.46387163        -2.58214902      -16.01911485
## 3           10.8443        0.00717129         0.00531811       -0.09199020
```

```
## 4           27.4036       -0.06923264        -0.04835565        0.03524228
## 5            2.0641       -1.57199529        -1.10532376      -10.40013281
## 6            7.7975       -1.42418103        -2.95732608      -22.27725638
##   percent_change_30d percent_change_60d percent_change_90d   market_cap
## 1       -26.39118788       -27.35857161       -32.66720536 566640002997
## 2       -33.67710395       -28.11097724       -35.11523602 243462140262
## 3        -0.11774563        -0.13237996        -0.14339447  75681882724
## 4         0.04125619        -0.00305729        -0.02917071  50993337983
## 5       -28.45273254       -23.93411936       -30.74394665  48495114806
## 6       -46.43812985       -47.51725263       -49.82558922  20133491375
##   market_cap_dominance fully_diluted_market_cap mineable exchange payments
## 1              44.3870             624925461682      Yes       No       No
## 2              19.0580             243462140262      Yes       No       No
## 3               5.9243              79639711610       No       No      Yes
## 4               3.9917              50993337983       No      Yes       No
## 5               3.7991              49041551944       No      Yes      Yes
## 6               1.5789              41647082782       No      Yes       No
```

```
tail(df)
```

```
##         id           name   symbol num_market_pairs               date_added
## 4995  7669           UNCL     UNCL                3 2020-11-13T00:00:00.000Z
## 4996 14809           Zada     ZADA                1 2021-11-19T03:25:19.000Z
## 4997 17800       Shintama SHINTAMA                4 2022-02-02T02:54:25.000Z
## 4998 10556     B.Protocol     BPRO                9 2021-06-21T00:00:00.000Z
## 4999 19906     ChoccySwap      CCY                1 2022-05-03T03:53:12.000Z
## 5000 12698 Ninja Protocol    NINJA                4 2021-10-14T02:52:18.000Z
##       max_supply circulating_supply total_supply cmc_rank        price
## 4995  1.7034e+05                  0 0.000000e+00     4995 2.166746e+01
## 4996  1.0000e+12                  0 1.000000e+12     4996 1.146176e-06
## 4997 -1.0000e+05                  0 0.000000e+00     4997 1.959466e-17
## 4998  1.0000e+07                  0 2.194432e+06     4998 1.042482e+00
## 4999 -1.0000e+05                  0 0.000000e+00     4999 1.505606e-02
## 5000  5.0000e+07                  0 5.000000e+07     5000 2.211259e-02
##      volume_24h volume_change_24h percent_change_1h percent_change_24h
## 4995   19801.88           -5.0042       -0.51192217          -3.930204
## 4996   19772.76          -27.5065       -0.02239866          -2.150332
```

```
## 4997   19756.34          -14.0515       -1.33605640          -2.562610
## 4998   19747.28          -23.1379       -2.37697824          -3.786820
## 4999   19710.65           29.5415       -3.76718119           5.643916
## 5000   19688.05          130.7254       -0.10566668           8.525406
##      percent_change_7d percent_change_30d percent_change_60d percent_change_90d
## 4995         -26.36190          -36.48762          -18.52411          -41.98006
## 4996         -27.70020          -55.01843          -70.21234          -78.57549
## 4997         -26.30816          -99.99998          -92.93683          -91.70596
## 4998         -17.43283          -73.75611          -74.00641          -81.44987
## 4999         -26.85643          -25.89154          -25.89154          -25.89154
## 5000         -69.30090          -83.46014          -88.17120          -91.29151
##      market_cap market_cap_dominance fully_diluted_market_cap mineable exchange
## 4995          0                    0                  3690836       No       No
## 4996          0                    0                  1146176       No       No
## 4997          0                    0                        0       No       No
## 4998          0                    0                 10424823       No       No
## 4999          0                    0                        0       No       No
## 5000          0                    0                  1105629       No       No
##      payments
## 4995       No
## 4996       No
## 4997       No
## 4998       No
## 4999       No
## 5000       No
```

```
str(df)
```

```
## 'data.frame':    5000 obs. of  24 variables:
##  $ id                       : int  1 1027 825 3408 1839 52 2010 5426 4687 74 ...
##  $ name                     : chr  "Bitcoin" "Ethereum" "Tether" "USD Coin" ...
##  $ symbol                   : chr  "BTC" "ETH" "USDT" "USDC" ...
##  $ num_market_pairs         : int  9431 5715 33404 3978 848 721 440 310 3699 473 ...
##  $ date_added               : chr  "2013-04-28T00:00:00.000Z" "2015-08-07T00:00:00.000Z" "2015-02-25T0
##  $ max_supply               : num  21000000 -100000 -100000 -100000 165116760 ...
##  $ circulating_supply       : num  1.90e+07 1.21e+08 7.58e+10 5.10e+10 1.63e+08 ...
##  $ total_supply             : num  1.90e+07 1.21e+08 7.97e+10 5.10e+10 1.63e+08 ...
```

```
##  $ cmc_rank                : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ price                   : num  2.98e+04 2.02e+03 9.99e-01 1.00 2.97e+02 ...
##  $ volume_24h              : num  3.31e+10 2.00e+10 6.28e+10 6.46e+09 1.73e+09 ...
##  $ volume_change_24h       : num  16.93 41.62 10.84 27.4 2.06 ...
##  $ percent_change_1h       : num  -1.11925 -1.46387 0.00717 -0.06923 -1.572 ...
##  $ percent_change_24h      : num  -1.33689 -2.58215 0.00532 -0.04836 -1.10532 ...
##  $ percent_change_7d       : num  -9.7199 -16.0191 -0.092 0.0352 -10.4001 ...
##  $ percent_change_30d      : num  -26.3912 -33.6771 -0.1177 0.0413 -28.4527 ...
##  $ percent_change_60d      : num  -27.35857 -28.11098 -0.13238 -0.00306 -23.93412 ...
##  $ percent_change_90d      : num  -32.6672 -35.1152 -0.1434 -0.0292 -30.7439 ...
##  $ market_cap              : num  5.67e+11 2.43e+11 7.57e+10 5.10e+10 4.85e+10 ...
##  $ market_cap_dominance    : num  44.39 19.06 5.92 3.99 3.8 ...
##  $ fully_diluted_market_cap: num  6.25e+11 2.43e+11 7.96e+10 5.10e+10 4.90e+10 ...
##  $ mineable                : chr  "Yes" "Yes" "No" "No" ...
##  $ exchange                : chr  "No" "No" "No" "Yes" ...
##  $ payments                : chr  "No" "No" "Yes" "No" ...
```

```
describe(df)
```

```
##                          vars    n          mean           sd       median
## id                          1 5000  9.398840e+03 5.705670e+03      8770.00
## name*                       2 5000  2.491370e+03 1.439780e+03      2490.50
## symbol*                     3 5000  2.306580e+03 1.332990e+03      2306.50
## num_market_pairs            4 5000  2.804000e+01 5.179600e+02         5.00
## date_added*                 5 5000  1.236120e+03 7.803500e+02      1069.00
## max_supply                  6 5000  1.267392e+15 2.680090e+16 100000000.00
## circulating_supply          7 5000  3.121563e+14 1.500169e+16   2436206.35
## total_supply                8 5000  1.457883e+16 9.759416e+17 135572005.60
## cmc_rank                    9 5000  2.500500e+03 1.443520e+03      2500.50
## price                      10 5000  2.022300e+02 4.932980e+03         0.02
## volume_24h                 11 5000  3.845968e+11 1.777213e+13     84622.07
## volume_change_24h          12 5000  3.628867e+08 1.327719e+10         0.00
## percent_change_1h          13 5000 -3.800000e-01 9.930000e+00        -0.57
## percent_change_24h         14 5000  3.180000e+00 1.489100e+02        -1.37
## percent_change_7d          15 5000 -1.513000e+01 3.011200e+02       -22.54
## percent_change_30d         16 5000  6.276551e+04 3.707133e+06       -45.10
## percent_change_60d         17 5000  2.578390e+03 1.837644e+05       -45.75
```

```
## percent_change_90d         18 5000  1.211928e+04 6.835482e+05       -57.29
## market_cap                 19 5000  2.707402e+08 8.873538e+09     38319.88
## market_cap_dominance       20 5000  2.000000e-02 6.900000e-01         0.00
## fully_diluted_market_cap   21 5000  3.582016e+12 1.697423e+14   5691501.89
## mineable*                  22 5000  1.080000e+00 2.700000e-01         1.00
## exchange*                  23 5000  1.040000e+00 2.000000e-01         1.00
## payments*                  24 5000  1.020000e+00 1.500000e-01         1.00
##                               trimmed          mad        min          max
## id                            9167.76      7085.35       1.00 2.014100e+04
## name*                         2490.98      1848.80       1.00 4.985000e+03
## symbol*                       2305.29      1711.66       1.00 4.628000e+03
## num_market_pairs                 6.50         4.45       1.00 3.340400e+04
## date_added*                   1176.34       765.02       1.00 3.016000e+03
## max_supply               838241502.15 148408260.00 -100000.00 1.000000e+18
## circulating_supply        99452414.77   3611919.54       0.00 9.818468e+17
## total_supply            1161578481.88 200999055.50       0.00 6.900000e+19
## cmc_rank                      2500.50      1853.25       1.00 5.000000e+03
## price                            0.20         0.03       0.00 3.068975e+05
## volume_24h                  369164.62    125460.68       0.00 1.068732e+15
## volume_change_24h                5.60        28.60    -100.00 6.136198e+11
## percent_change_1h               -0.63         0.91     -85.50 5.639200e+02
## percent_change_24h              -0.94         3.83     -99.38 1.010783e+04
## percent_change_7d              -22.73        19.01    -100.00 2.058222e+04
## percent_change_30d             -43.11        23.70    -100.00 2.555570e+08
## percent_change_60d             -43.13        27.24    -100.00 1.299406e+07
## percent_change_90d             -53.48        27.99    -100.00 4.626783e+07
## market_cap                 2151091.82     56813.05       0.00 5.666400e+11
## market_cap_dominance             0.00         0.00       0.00 4.439000e+01
## fully_diluted_market_cap   30232309.74   8409502.61       0.00 8.824270e+15
## mineable*                        1.00         0.00       1.00 2.000000e+00
## exchange*                        1.00         0.00       1.00 2.000000e+00
## payments*                        1.00         0.00       1.00 2.000000e+00
##                                range  skew kurtosis           se
## id                      2.014000e+04  0.28    -1.08 8.069000e+01
## name*                   4.984000e+03  0.00    -1.20 2.036000e+01
## symbol*                 4.627000e+03  0.00    -1.20 1.885000e+01
```

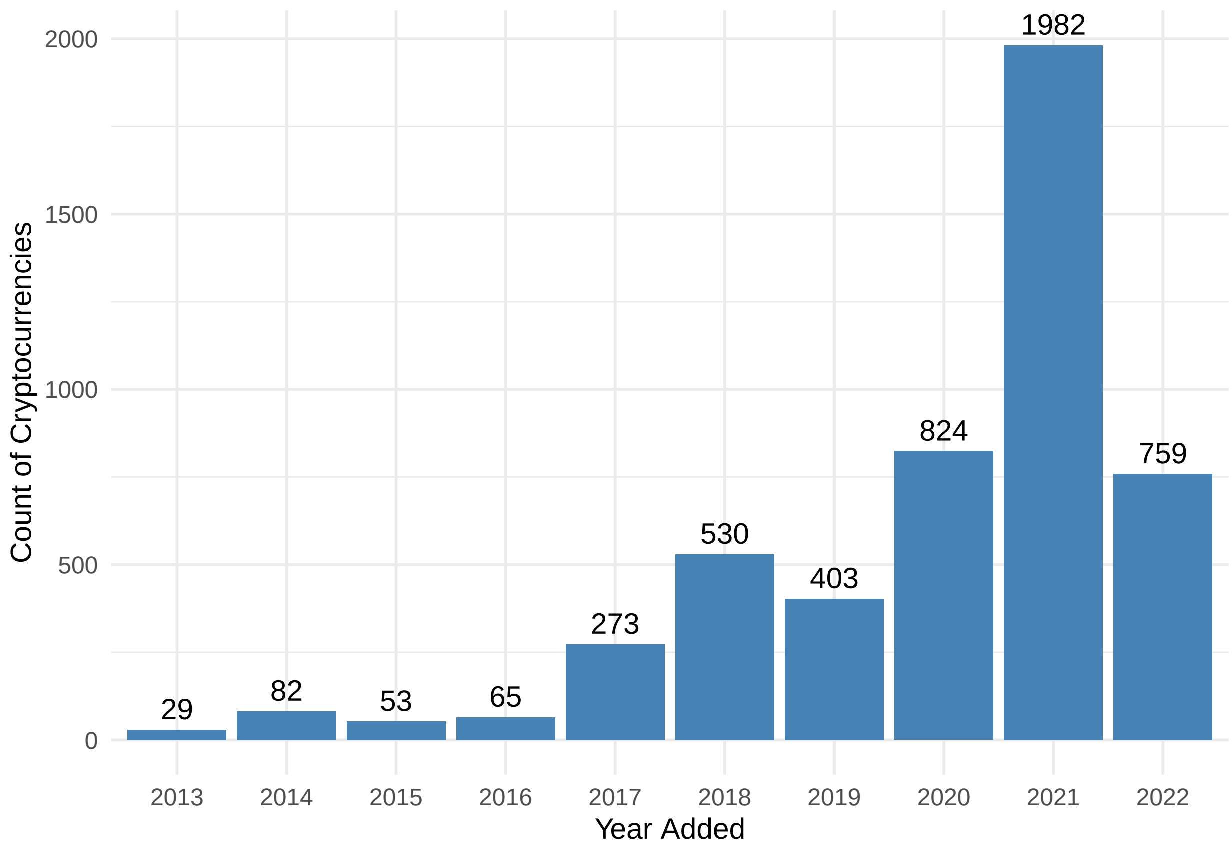
```
## num_market_pairs         3.340300e+04 55.70  3477.44 7.330000e+00
## date_added*              3.015000e+03  0.64    -0.56 1.104000e+01
## max_supply               1.000000e+18 32.86  1179.22 3.790220e+14
## circulating_supply       9.818468e+17 59.63  3756.52 2.121560e+14
## total_supply             6.900000e+19 70.64  4990.03 1.380190e+16
## cmc_rank                 4.999000e+03  0.00    -1.20 2.041000e+01
## price                    3.068975e+05 50.94  3033.24 6.976000e+01
## volume_24h               1.068732e+15 53.28  2966.85 2.513358e+11
## volume_change_24h        6.136198e+11 40.63  1698.72 1.877678e+08
## percent_change_1h        6.494200e+02 39.31  2128.83 1.400000e-01
## percent_change_24h       1.020720e+04 63.28  4248.21 2.110000e+00
## percent_change_7d        2.068222e+04 64.50  4380.26 4.260000e+00
## percent_change_30d       2.555571e+08 66.25  4521.61 5.242678e+04
## percent_change_60d       1.299416e+07 70.67  4992.97 2.598820e+03
## percent_change_90d       4.626793e+07 63.68  4225.77 9.666830e+03
## market_cap               5.666400e+11 56.33  3430.55 1.254908e+08
## market_cap_dominance     4.439000e+01 56.48  3443.56 1.000000e-02
## fully_diluted_market_cap 8.824270e+15 49.63  2479.47 2.400518e+12
## mineable*                1.000000e+00  3.16     7.97 0.000000e+00
## exchange*                1.000000e+00  4.54    18.62 0.000000e+00
## payments*                1.000000e+00  6.22    36.68 0.000000e+00
```

```r
top_ranked <- head(df)
bottom_ranked <- tail(df)
```

Barplot of Year in which Cryptocurrency was added

```r
df$year <- substr(df$date_added, 0, 4)

ggplot(data=df, aes(x=year)) +
  geom_bar(stat="count", fill="steelblue")+
  geom_text(stat='count', aes(label=..count..), vjust=-0.5)+
  xlab("Year Added") + ylab("Count of Cryptocurrencies")+
  theme_minimal()
```

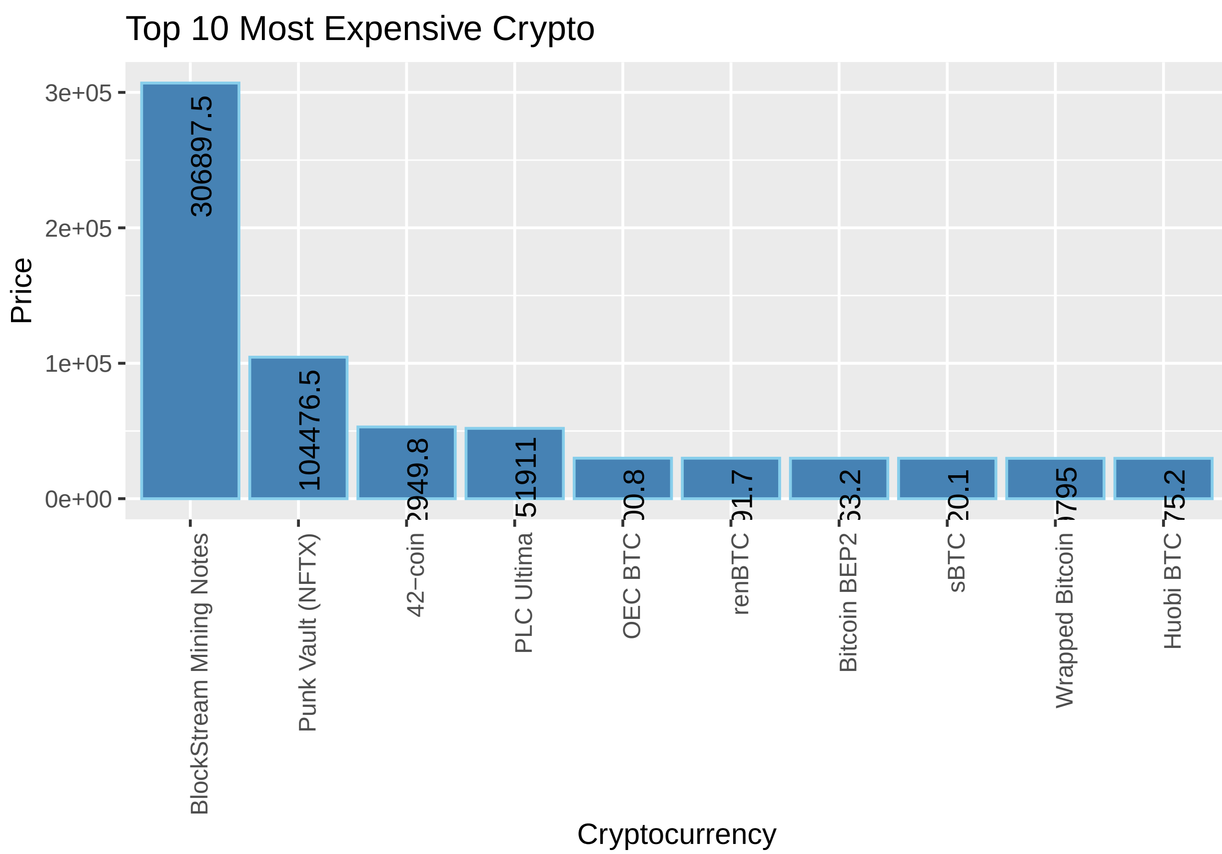


Top 10 Most Expensive Crypto

```
## OrderBy Price (DESC)
df2 <- df[order(-df$price),]

## Filter Top 10
priceDesc <- top_n(df2,10,df2$price)

p <- ggplot(priceDesc, aes(x = reorder(name, -price), y = price)) + geom_bar(stat="identity", color='sky
  labs(y="Price", x="Cryptocurrency", title="Top 10 Most Expensive Crypto")+
theme(axis.text.x=element_text(angle=90, hjust=1))+
  geom_text(label=round(priceDesc$price,digit=1),position=position_dodge(width=0.3),
                                                  hjust=1.1,angle=90,vjust=1)
p
```

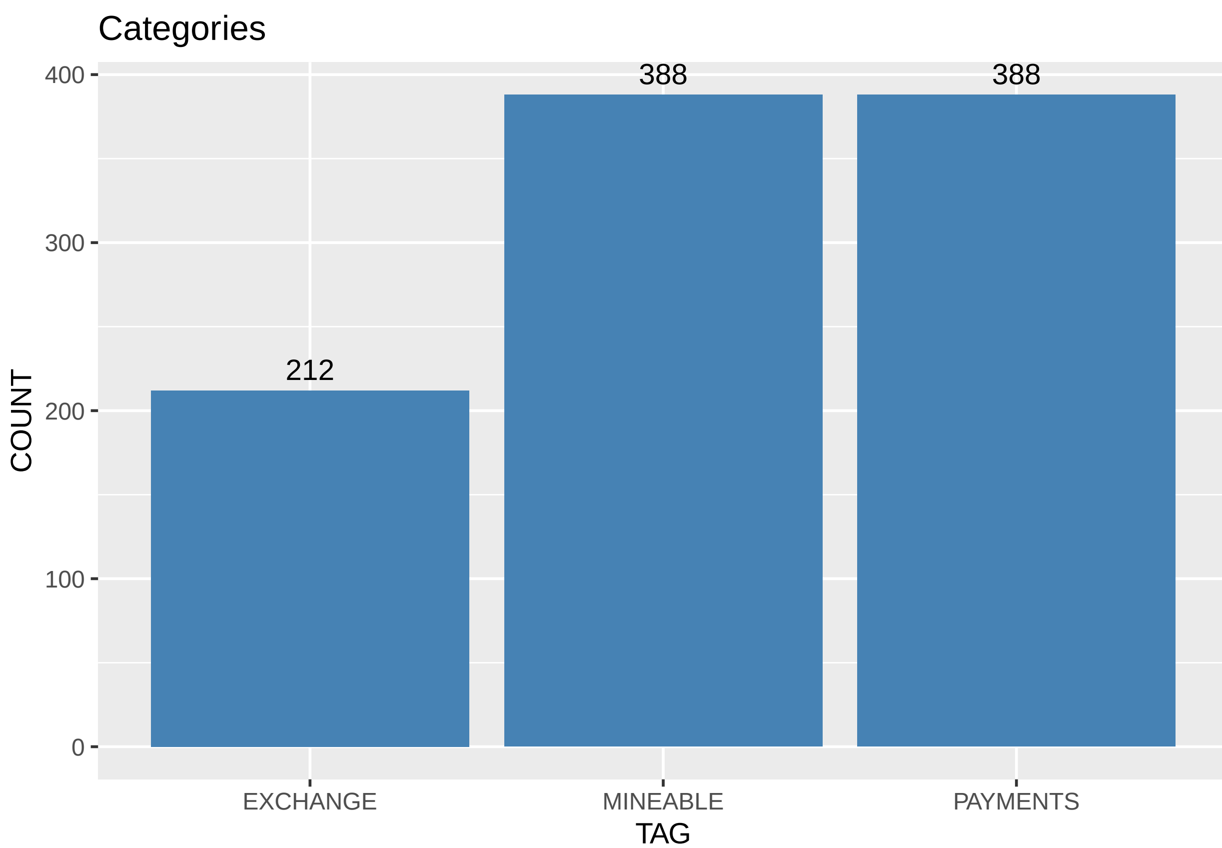


Count of Tags

```
catCount <-c()
for (i in 1:nrow(df))
  {
  if(identical(df[i,22],"Yes")){
    catCount <- append(catCount,"MINEABLE")
  }
 if(identical(df[i,23],"Yes")){
    catCount <- append(catCount,"EXCHANGE")
  }
  if(identical(df[i,22],"Yes")){
    catCount <- append(catCount,"PAYMENTS")
  }


}
```

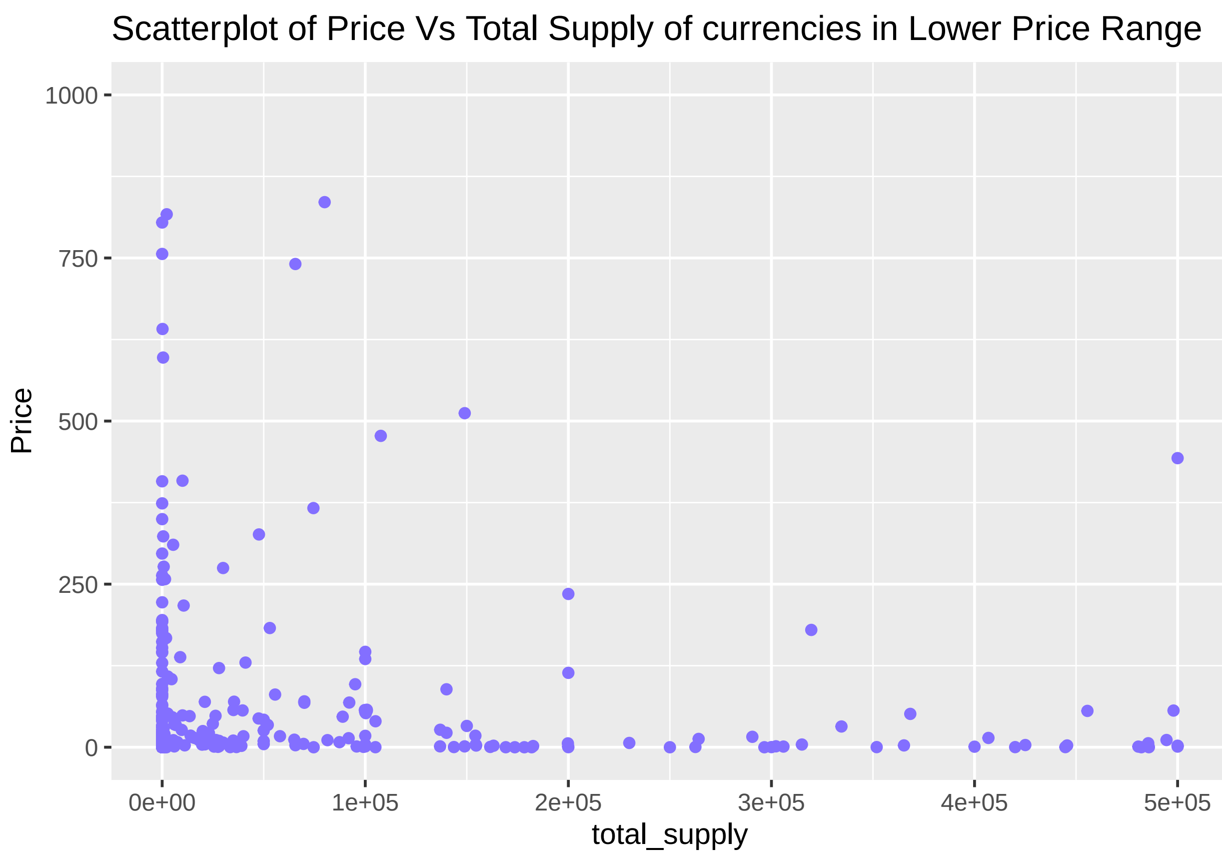
```
q <- ggplot(data.frame(catCount), aes(x=catCount)) +
  geom_bar(stat="count", fill="steelblue")+
  geom_text(stat='count', aes(label=..count..), vjust=-0.5)+
  labs(y="COUNT", x="TAG", title="Categories")
q
```



Scatterplot of Price Vs Total Supply of currencies in Lower Price Range

```
ggplot(data=df,aes(total_supply,price)) +
   geom_point(color="slateblue1") +
   ggtitle("Scatterplot of Price Vs Total Supply of currencies in Lower Price Range") + labs(y="Price",:
```

```
## Warning: Removed 4164 rows containing missing values (geom_point).
```

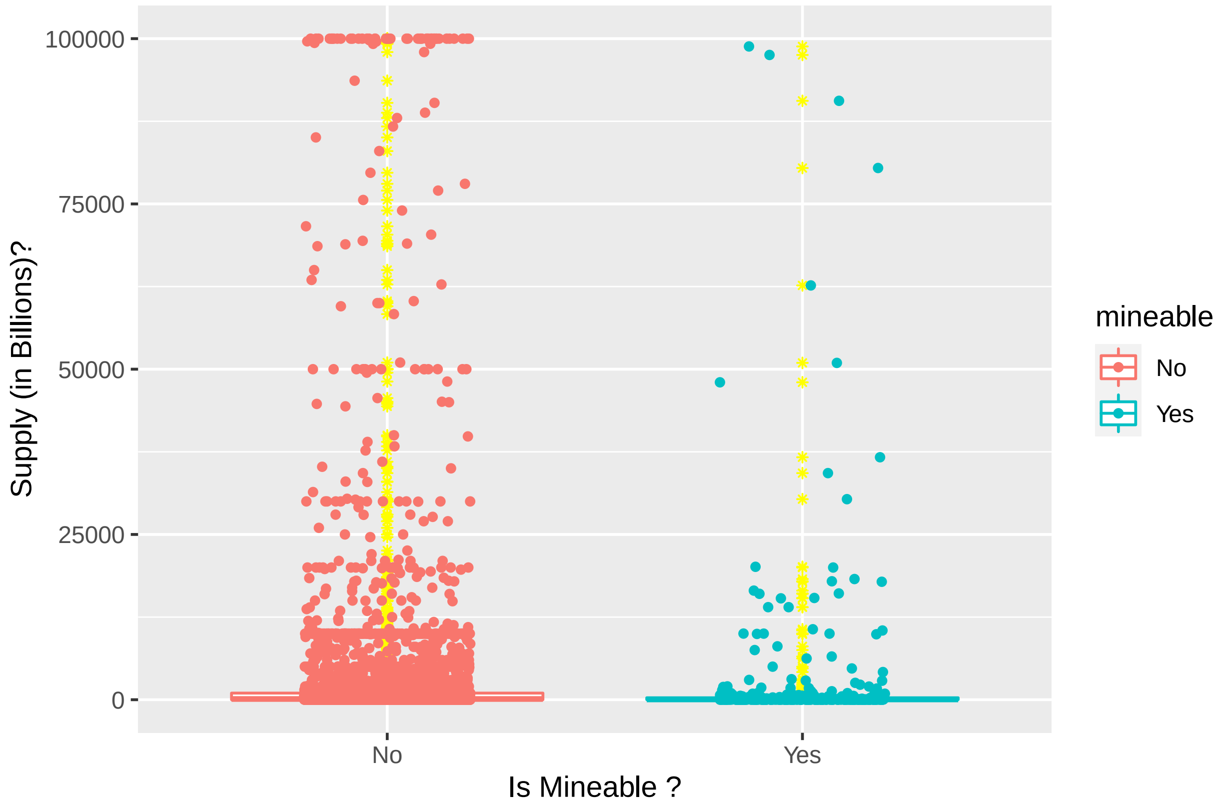


Jitterplot of Total Supply Vs Mineable

```
ggplot(df, aes(x=mineable, y=total_supply/1000000, color=mineable)) +
  geom_boxplot(notch=FALSE,outlier.colour="yellow", outlier.shape=8,
               outlier.size=1) +
  geom_jitter(shape=16, position=position_jitter(0.2)) +
  labs(caption="Jitterplot of Total Supply Vs Mineable",
  x="Is Mineable ?", y = "Supply (in Billions)?") +
  theme(plot.caption = element_text(hjust = 0.5))+ylim(c(0,100000))
```

```
## Warning: Removed 256 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 568 rows containing missing values (geom_point).
```



Jitterplot of Total Supply Vs Mineable

ANOVA Test

```r
# One-way ANOVA Test

# Set significance level
alpha <- 0.05
# Dataframe for 30 Days
thirtyDays <- data.frame('variation' = df$percent_change_30d,
                         'timeperiod' = rep('thirtyDays',5000), stringsAsFactors = FALSE)
# Dataframe for 60 Days
sixtyDays <- data.frame('variation' = df$percent_change_60d,
                        'timeperiod' = rep('sixtyDays',5000), stringsAsFactors = FALSE)
# Dataframe for 90 Days
ninetyDays <- data.frame('variation' = df$percent_change_90d,
                         'timeperiod' = rep('ninetyDays',5000), stringsAsFactors = FALSE)
# Combine the Dataframe
```

```r
variation <- rbind(thirtyDays,sixtyDays,ninetyDays)
variation$timeperiod <- as.factor(variation$timeperiod)

# Hypotheses
# H0: Mean Variation(30 Days) =  Mean Variation(60 Days) =  Mean Variation(90 Days)
# H1: Atleast one mean is different from others

anova <- aov(variation ~ timeperiod, data = variation)
a_summ <-summary(anova)
# Critical Value
qf(1-alpha,a_summ[[1]][1,1],a_summ[[1]][2,1])
```

```
## [1] 2.996331
```

```r
# Test Value
F.value <- a_summ[[1]][[1,"F value"]]
F.value
```

```
## [1] 1.101977
```

```r
# Compare p-value and alpha to make decision
p.value <- a_summ[[1]][[1,"Pr(>F)"]]
p.value
```

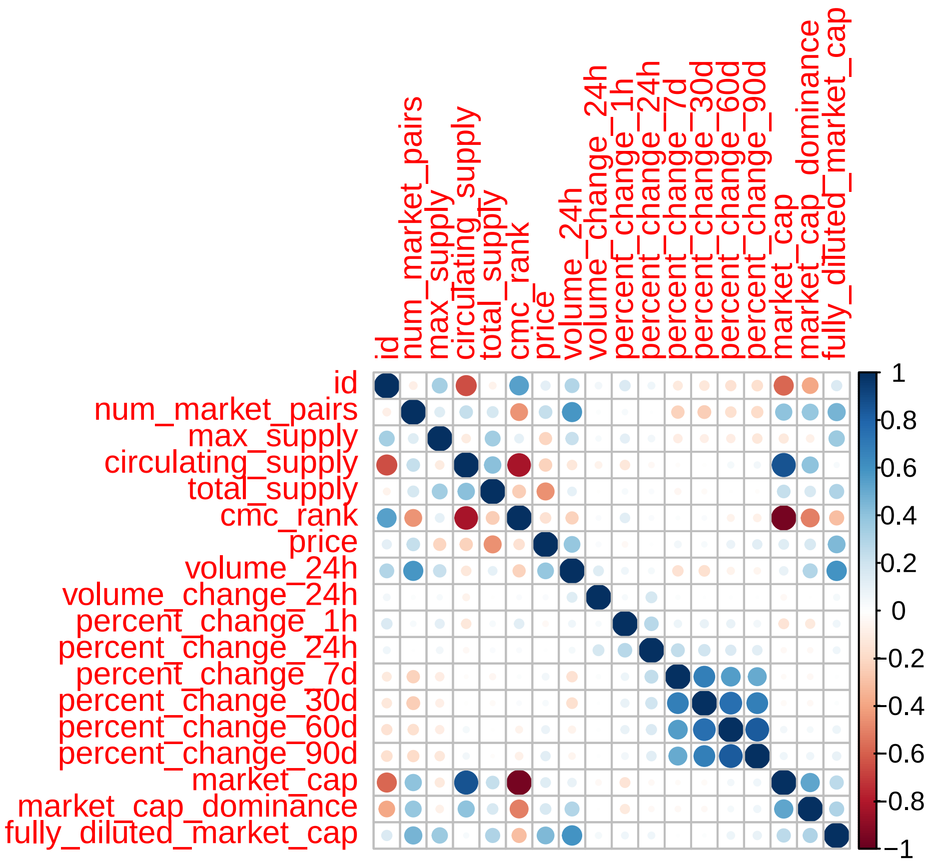
```
## [1] 0.3322404
```

```r
ifelse(p.value > alpha,"Failed to reject Null Hypothesis","Reject Null Hypothesis")
```

```
## [1] "Failed to reject Null Hypothesis"
```

Correlation Analysis of Numeric values

```r
numeric_data <- select_if(df, is.numeric)
data.cor = cor(numeric_data, method = c("spearman"))
corrplot(data.cor)
```


id
num_market_pairs
max_supply
circulating_supply
total_supply
cmc_rank
price
volume_24h
volume_change_24h
percent_change_1h
percent_change_24h
percent_change_7d
percent_change_30d
percent_change_60d
percent_change_90d
market_cap
market_cap_dominance
fully_diluted_market_cap
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1