Bitcoin price predication: linear regression

PRESENTED BY:

FATIMAH ALQAHTANI

RAWABI ALMUTAIRI

RAGHAD ALSAQER

Content

- Objective
- Methodology
- Dataset (feature exploration)
- Result
- Future work

Objective

Explore significant features → Introduce a predictive model → Find out the accuracy

Methodology

- Explored various features which may have correlation with the price of Bitcoin
- Train-Test split executed using standard Linear Regression methodology rather than as a Time Series, per project requirements
- As such, the price of bitcoin is predicted at a moment in time, rather than predicted on a time series basis

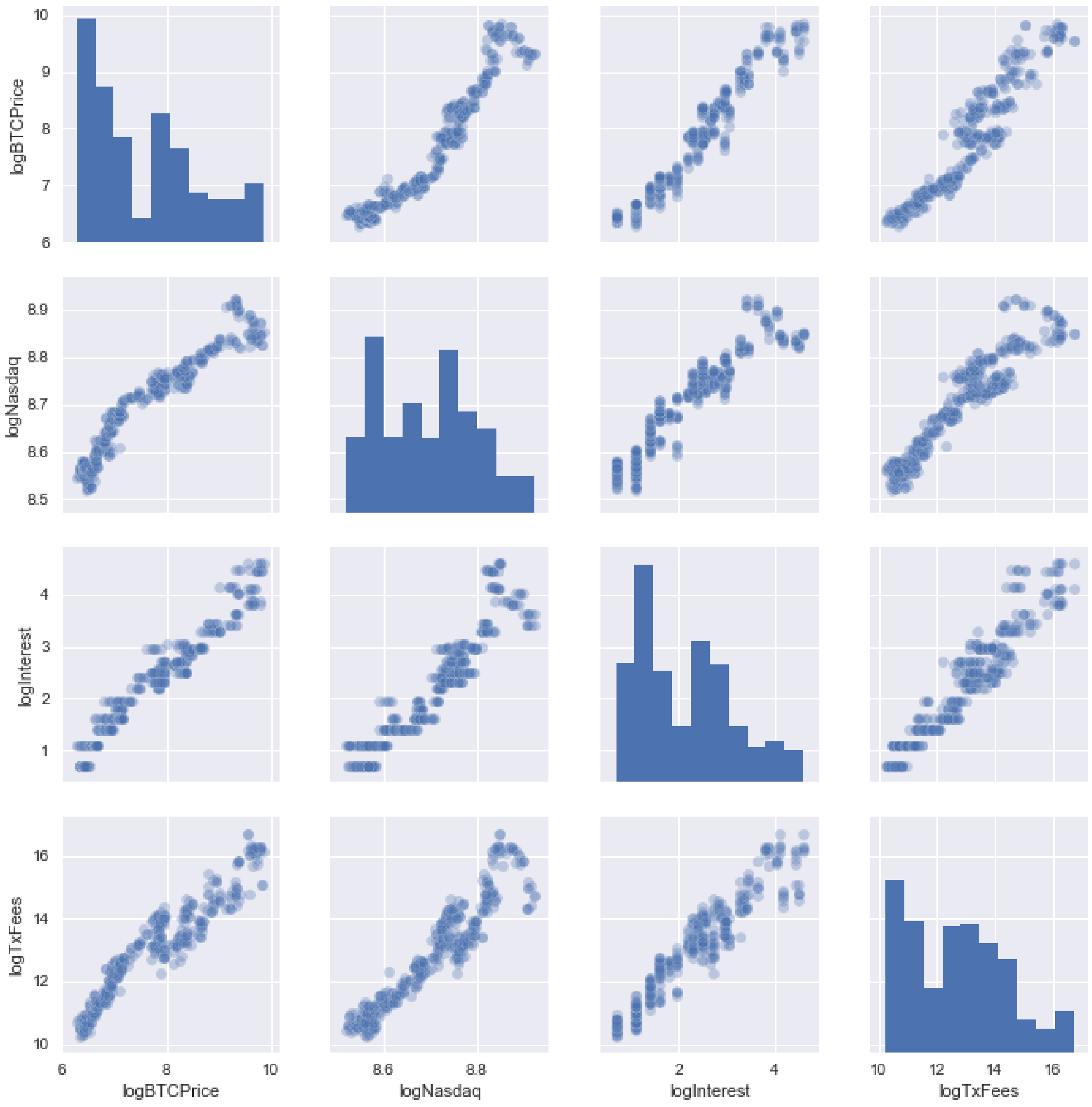
Scraping

- Bitcoin features
  - https://coinmarketcap.com/
- Google search
  - https://trends.google.com/trends/?geo=SA

Dataset (Feature Exploration)

Index	BTCPrice	logBTCPrice	logETHPrice	logBTCVol	logTxFees	logCostperTxn	logNoTxns	logAvgBlkSz	logUniqueAddresses	logHashRate	logNasdaq	logGold	logInterest	Interest	TxFees	Nasdaq
2021-06-01 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.829	46	1.651e+06	7816
2021-06-08 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.829	46	1.651e+06	7816
2021-06-15 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.829	46	1.651e+06	7816
2021-06-22 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.829	46	1.651e+06	7816
2021-06-29 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.829	46	1.651e+06	7816
2021-07-01 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.466	32	1.651e+06	7816
2021-07-06 00:00:00	7383	8.907	5.561	14.1	14.32	3.751	12.8	0.2322	13.42	17.73	8.964	5.066	3.466	32	1.651e+06	7816

Correlation



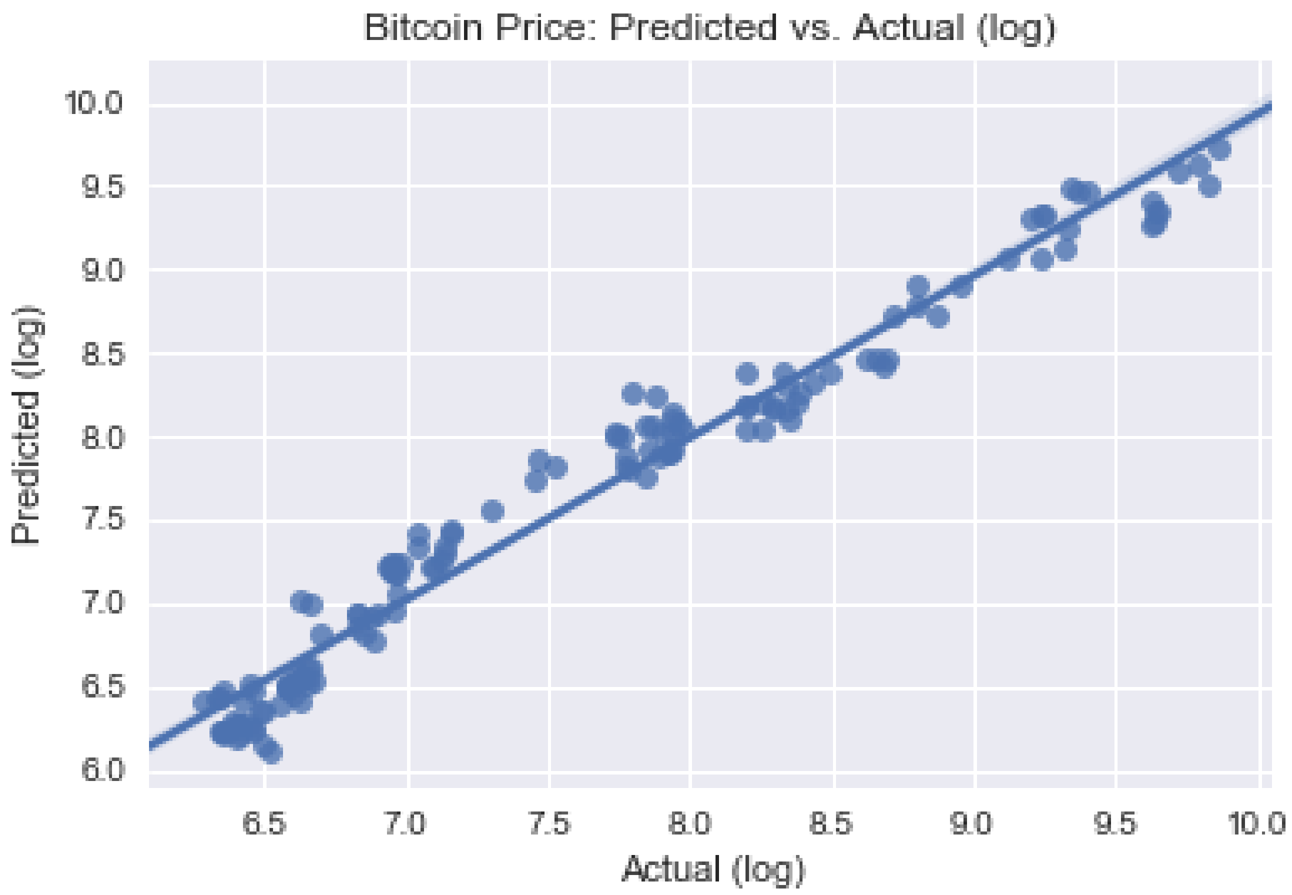
Features contribute

We found the 3 core features for the model

- Nasdaq
- Interest
- TxFees

Index	ogBTCPric	ogNasdac	ogInteres	logTxFees
2021-06-01 00:00:00	8.907	8.964	3.829	14.32
2021-06-08 00:00:00	8.907	8.964	3.829	14.32
2021-06-15 00:00:00	8.907	8.964	3.829	14.32
2021-06-22 00:00:00	8.907	8.964	3.829	14.32
2021-06-29 00:00:00	8.907	8.964	3.829	14.32
2021-07-01 00:00:00	8.907	8.964	3.466	14.32
2021-07-06 00:00:00	8.907	8.964	3.466	14.32
2021-07-13 00:00:00	8.907	8.964	3.466	14.32
2021-07-20 00:00:00	8.907	8.964	3.466	14.32
2021-07-27 00:00:00	8.907	8.964	3.466	14.32
2021-08-01 00:00:00	8.907	8.964	3.466	14.32
2021-08-03 00:00:00	8.907	8.964	3.466	14.32
2021-08-10 00:00:00	8.907	8.964	3.466	14.32
2021-08-17 00:00:00	8.907	8.964	3.466	14.32
2021-08-24 00:00:00	8.907	8.964	3.466	14.32
2021-08-31 00:00:00	8.907	8.964	3.466	14.32
2021-09-01 00:00:00	8.907	8.964	3.401	14.32

Linear Regression



3-LR Analysis

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             logBTCPrice   R-squared:                       0.846
Model:                             OLS   Adj. R-squared:                  0.845
Method:                  Least Squares   F-statistic:                     1093.
Date:                 Fri, 22 Oct 2021   Prob (F-statistic):          9.14e-242
Time:                         09:56:06   Log-Likelihood:                -86.459
No. Observations:                  600   AIC:                             180.9
Df Residuals:                      596   BIC:                             198.5
Df Model:                            3
Covariance Type:             nonrobust
===============================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept     -28.6365      1.013    -28.280      0.000     -30.625     -26.648
logInterest     0.5346      0.025     21.446      0.000       0.486       0.584
logNasdaq       3.9967      0.115     34.680      0.000       3.770       4.223
logTxFees       0.0146      0.014      1.075      0.283      -0.012       0.041
==============================================================================
Omnibus:                         8.798   Durbin-Watson:                   0.080
Prob(Omnibus):                   0.012   Jarque-Bera (JB):                8.779
Skew:                           -0.269   Prob(JB):                       0.0124
Kurtosis:                        3.247   Cond. No.                     1.46e+03
==============================================================================
```

Result

Train accuracy

- 0.8399367307539943

Test accuracy = accuracy for the model

- 0. 8568224150791021

Regularization

Train accuracy

- 0.8397597320509806

Test accuracy = accuracy for the model

- 0.8568917069026165

Future work

- Further explore seasonality in residual data
- Explore social media sentiment (e.g. Twitter) as leading indicator

Other analyses to explore

- ✓ Usage by country
- ✓ Bitcoin trading by exchange
- ✓ Bitcoin trading by currency
- ✓ SEASONALITY

Thank you, any question ?

Tools

- pandas
- numpy
- requests
- bs4 (BeautifulSoup)
- dateutil(parser)
- pprint
- json
- Datetime
- Seaborn
- matplotlib
- Statsmodels
- patsy
- sklearn
- warnings