DATA MINING AND PREDICTIVE MODELING FOR CRYPTO-CURRENCIES

LAUREN DEANS
JAMES BEEDLE
THEO SHIN
INDERPAL DHILLON

Our Main Questions

How would a 1$ increase in Bitcoin affect each of the other cryptocurrencies in our data set?

How to best set up a dataset for predictive modeling methods?

Data Preparation & Work

Dataset obtained from:

https://www.kaggle.com/albala/ticks-bitcoin-ethereumlitecoin-ripple

- CSV file with 1,314,830 data points
- Contains 39 days of transaction data for:

    Bitcoin (BTC)

    Bitcoin Cash (BCH)

    Ethereum (ETH)

    Litecoin (LTC)

    Ripple (XRP)

Data Preparation & Work (Continued)

- First transformed dataset into pivot table
- Also created a separate data matrix for Multilinear Regression Analysis
- Normalized values
- To fill in missing data, utilized Front-fill and Back-fill methods to propagate next values backward or previous values forward

coin	timestamp	Bitcoin	Bitcoin Cash	Ethereum	Litecoin	Ripple
0	2018-08-15 18:45:56	NaN	NaN	297.56	NaN	NaN
1	2018-08-15 18:46:15	6548.0	NaN	NaN	NaN	NaN
2	2018-08-15 18:46:24	6548.0	NaN	NaN	NaN	NaN
3	2018-08-15 18:46:38	6547.3	NaN	NaN	NaN	NaN
4	2018-08-15 18:46:41	NaN	NaN	NaN	NaN	0.29083

There are 1314830 rows and 3 columns

⬇ No more 'NaN' values ⬇

coin	timestamp	Bitcoin	Bitcoin Cash	Ethereum	Litecoin	Ripple
0	2018-08-15 18:45:56	6548.00	529.41	297.56	58.04	0.29083
1	2018-08-15 18:46:15	6548.00	529.41	297.81	58.04	0.29083
2	2018-08-15 18:46:24	6548.00	529.41	297.81	58.04	0.29083
3	2018-08-15 18:46:38	6547.30	529.41	297.81	58.04	0.29083
4	2018-08-15 18:46:41	6547.30	529.41	297.81	58.04	0.29083
5	2018-08-15 18:46:42	6547.30	529.41	297.81	58.04	0.29083

TOOLS

- Jupyter Notebooks
- Python
  - matplotlib
 - pandas
  - numpy
  - seaborn
  - statsmodels
  - sklearn
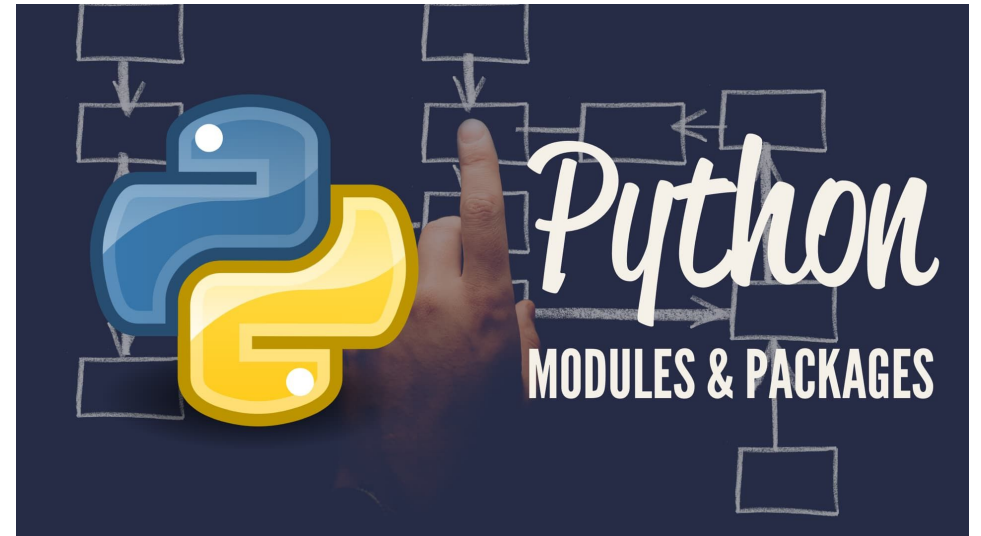
Data & Pairplot

- Seaborn library sns.pairplot

- Descriptive statistics

	bch	btc	eth	ltc	xrp
count	657396.000000	657396.000000	657396.000000	657396.000000	657396.000000
mean	510.645810	6584.879991	249.461496	57.258313	0.328438
std	56.228144	298.497715	39.861116	3.859021	0.069438
min	407.700000	6094.470000	167.000000	47.090000	0.253000
25%	464.510000	6380.510000	211.880000	54.790000	0.283210
50%	518.840000	6477.610000	262.510000	56.580000	0.322700
75%	545.500000	6727.280000	286.000000	59.030000	0.338200
max	660.070000	7411.000000	321.180000	69.360000	0.764400
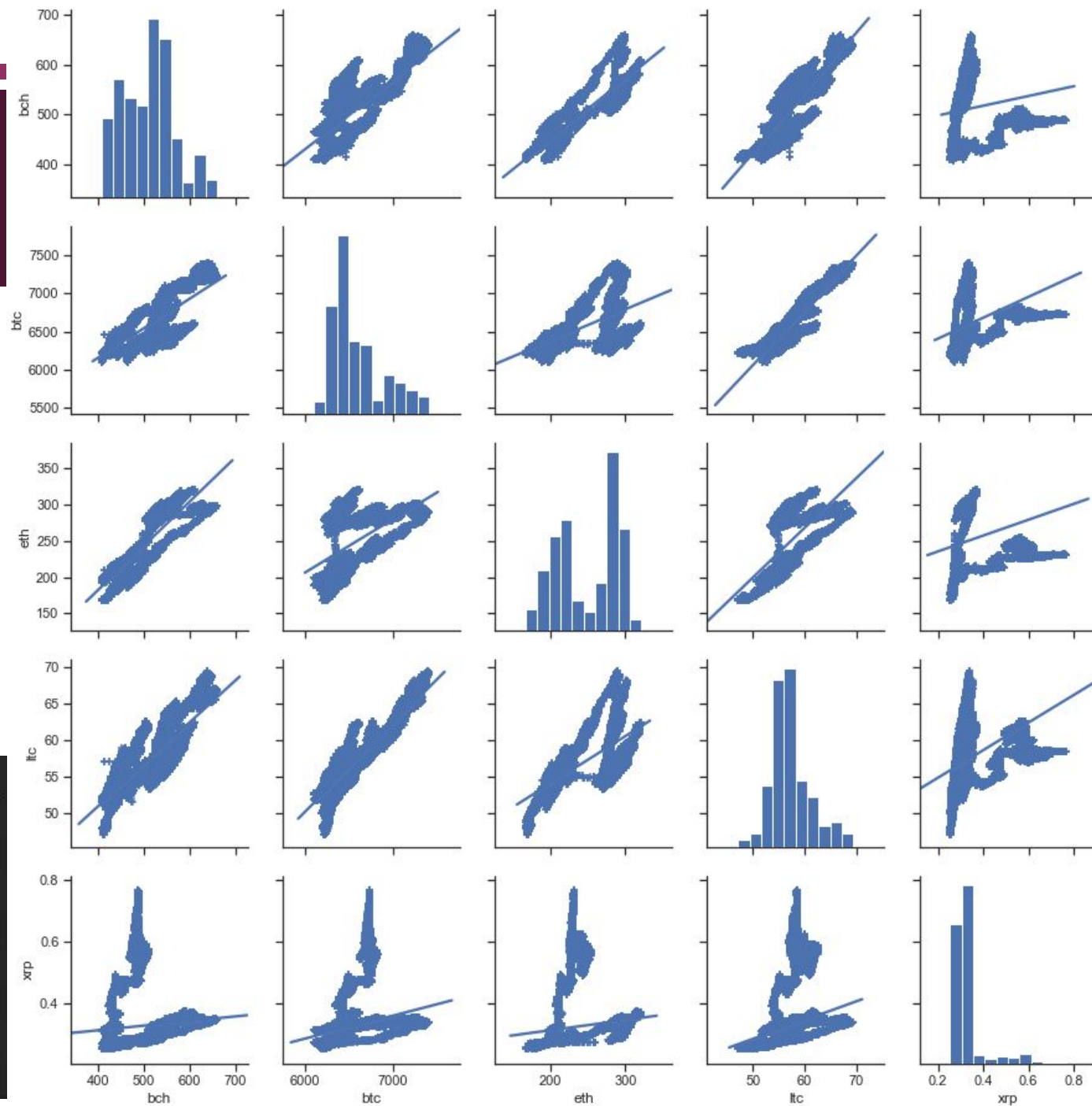
Multilinear Regression

■ Bitcoin

	coef	std err	t	P>|t|	[0.025	0.975]
Dep. Variable:	btc		R-squared:			0.877
Model:	OLS		Adj. R-squared:			0.877
Method:	Least Squares		F-statistic:			1.168e+06
Date:	Thu, 12 Dec 2019		Prob (F-statistic):			0.00
Time:	20:27:24		Log-Likelihood:			-3.9913e+06
No. Observations:	657396		AIC:			7.983e+06
Df Residuals:	657391		BIC:			7.983e+06
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2284.4780	2.096	1090.009	0.000	2280.370	2288.586
bch	-0.5061	0.007	-73.678	0.000	-0.520	-0.493
eth	-0.5595	0.007	-82.211	0.000	-0.573	-0.546
ltc	82.4846	0.071	1163.890	0.000	82.346	82.623
xrp	-74.6012	2.175	-34.296	0.000	-78.865	-70.338

	coef	std err	t	P>|t|	[0.025	0.975]
Dep. Variable:	bch		R-squared:			0.889
Model:	OLS		Adj. R-squared:			0.889
Method:	Least Squares		F-statistic:			1.314e+06
Date:	Thu, 12 Dec 2019		Prob (F-statistic):			0.00
Time:	20:25:34		Log-Likelihood:			-2.8596e+06
No. Observations:	657396		AIC:			5.719e+06
Df Residuals:	657391		BIC:			5.719e+06
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-40.7542	0.626	-65.116	0.000	-41.981	-39.528
btc	-0.0162	0.000	-73.678	0.000	-0.017	-0.016
eth	0.7453	0.001	923.739	0.000	0.744	0.747
ltc	8.9936	0.019	468.555	0.000	8.956	9.031
xrp	-130.6996	0.354	-368.855	0.000	-131.394	-130.005

■ Bitcoin cash

Multilinear Regression

- Ripple

Dep. Variable:	xrp	R-squared:	0.268
Model:	OLS	Adj. R-squared:	0.268
Method:	Least Squares	F-statistic:	6.029e+04
Date:	Thu, 12 Dec 2019	Prob (F-statistic):	0.00
Time:	20:29:30	Log-Likelihood:	9.2340e+05
No. Observations:	657396	AIC:	-1.847e+06
Df Residuals:	657391	BIC:	-1.847e+06
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.0775	0.002	-39.004	0.000	-0.081	-0.074
btc	-2.394e-05	6.98e-07	-34.296	0.000	-2.53e-05	-2.26e-05
bch	-0.0013	3.56e-06	-368.855	0.000	-0.001	-0.001
ltc	0.0177	6.68e-05	265.071	0.000	0.018	0.018
eth	0.0009	3.72e-06	237.450	0.000	0.001	0.001

Dep. Variable:	eth	R-squared:	0.775
Model:	OLS	Adj. R-squared:	0.775
Method:	Least Squares	F-statistic:	5.664e+05
Date:	Thu, 12 Dec 2019	Prob (F-statistic):	0.00
Time:	20:28:12	Log-Likelihood:	-2.8651e+06
No. Observations:	657396	AIC:	5.730e+06
Df Residuals:	657391	BIC:	5.730e+06
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	42.4777	0.631	67.320	0.000	41.241	43.714
btc	-0.0182	0.000	-82.211	0.000	-0.019	-0.018
bch	0.7579	0.001	923.739	0.000	0.756	0.759
ltc	-1.5653	0.022	-70.283	0.000	-1.609	-1.522
xrp	89.4539	0.377	237.450	0.000	88.716	90.192

- Ethereum

Multilinear Regression & Correlation Matrix

- Litecoin

Dep. Variable:	ltc	R-squared:		0.927
Model:	OLS	Adj. R-squared:		0.927
Method:	Least Squares	F-statistic:		2.086e+06
Date:	Thu, 12 Dec 2019	Prob (F-statistic):		0.00
Time:	20:28:48	Log-Likelihood:		-9.6045e+05
No. Observations:	657396	AIC:		1.921e+06
Df Residuals:	657391	BIC:		1.921e+06
Df Model:	4			
Covariance Type:	nonrobust			

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-11.3083	0.032	-353.088	0.000	-11.371	-11.246
btc	0.0082	7.01e-06	1163.890	0.000	0.008	0.008
bch	0.0278	5.94e-05	468.555	0.000	0.028	0.028
eth	-0.0048	6.78e-05	-70.283	0.000	-0.005	-0.005
xrp	5.4566	0.021	265.071	0.000	5.416	5.497
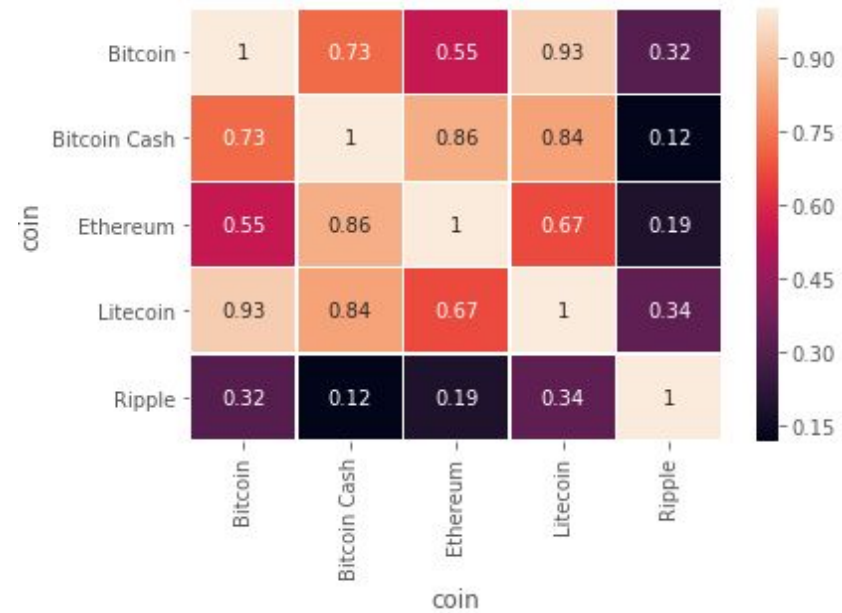
- Correlation heatmap
- Pearson method

Correlation Matrix - Spearman

- Correlation Heatmap
- Spearman method
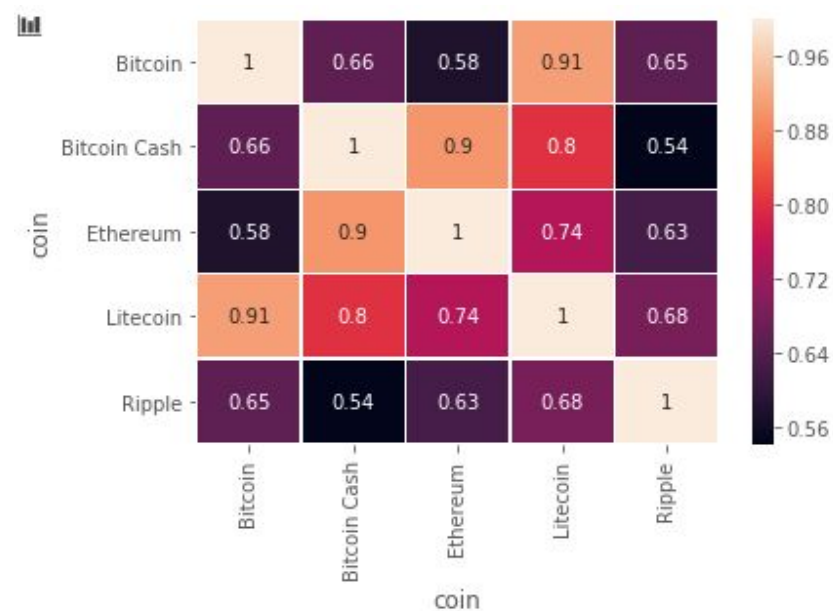
Question: How does a 1$ increase in BTC effect other coins?

Given a 1$ increase in Bitcoin

BCH = - $0.0162
ETH = - $0.0182
LTC = + $0.0082
XRP = - 2.394e-05

Given 1$ increase in Litecoin

BTC = + 0.082
BCH = + 0.0278
ETH = -.0048
XRP = + 5.4566

Some Additional Thoughts

To Conclude