Bitcoin Price Classification

Springboard Data Science Track - Capstone Project 3

Julio Moreno
May 6, 2021

Project objective:

Build a model that can classify/cluster daily bitcoin price movements by finding patterns in the data.

The hypothesis is that certain clusters could be associated with or used to predict important movements in the bitcoin price.

Why should we care about the Bitcoin Price?

Rank	Name	Market Cap
1	Gold GOLD	$11.335 T
2	Apple AAPL	$2.144 T
3	Saudi Aramco 2222.SR	$1.901 T
4	Microsoft MSFT	$1.857 T
5	Amazon AMZN	$1.651 T
6	Alphabet (Google) GOOG	$1.564 T
7	Silver SILVER	$1.451 T
8	Bitcoin BTC	$1.073 T
9	Facebook FB	$893.48 B
10	Tencent TCEHY	$769.85 B

https://companiesmarketcap.com/assets-by-market-cap

Bitcoin's current market cap is similar to the most valuable companies and assets on earth.

In its 12 years of existence, it has climbed in from $0 to $57,000 per coin at the time of writing.

Bitcoin's value has increased along the growth in its usage (users, transactions, etc) and its intrinsic monetary properties (decreasing supply of new bitcoins, decentralization, censorship-resistant, etc).

Who cares?

Companies, Investors

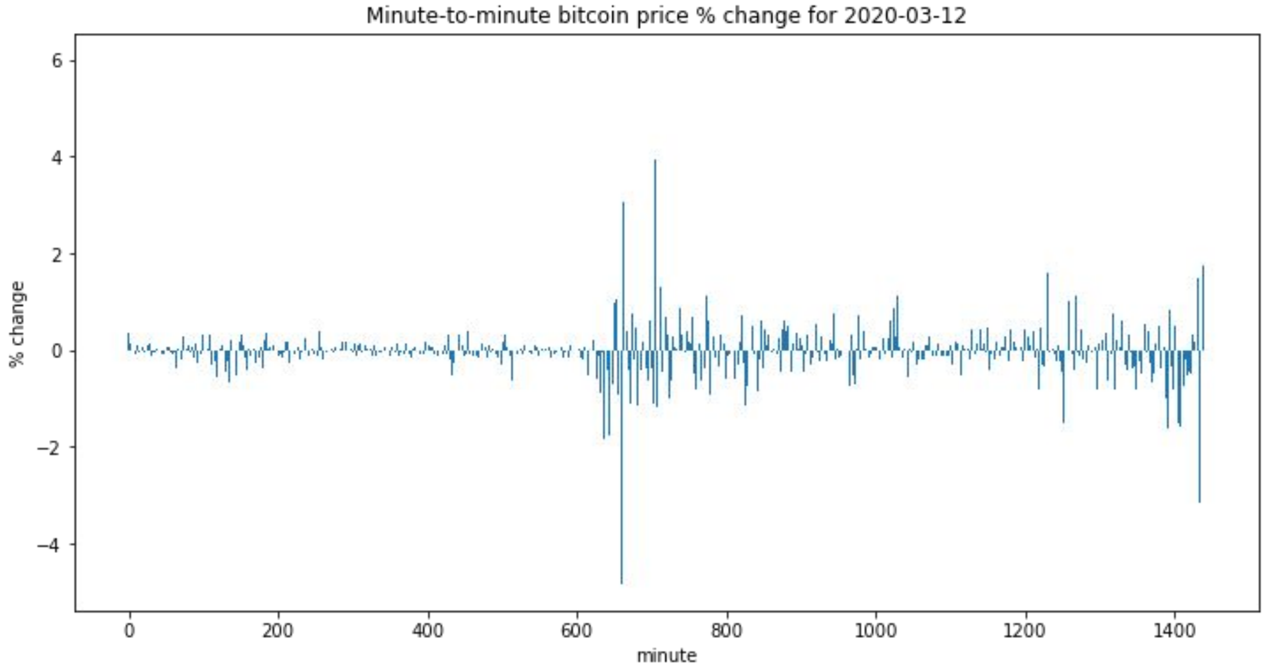
Companies, HNWI, investment funds, etc, have added bitcoin to their balance sheets or have a financial interest in bitcoin.

General Public

Around 100-130 million people around the world have put part of their savings into bitcoin, or use bitcoin to get paid on the internet, make international transfers, etc.

Data Analysis and source

For classification purposes, every sample consist of the change in the bitcoin price every minute of a given day.

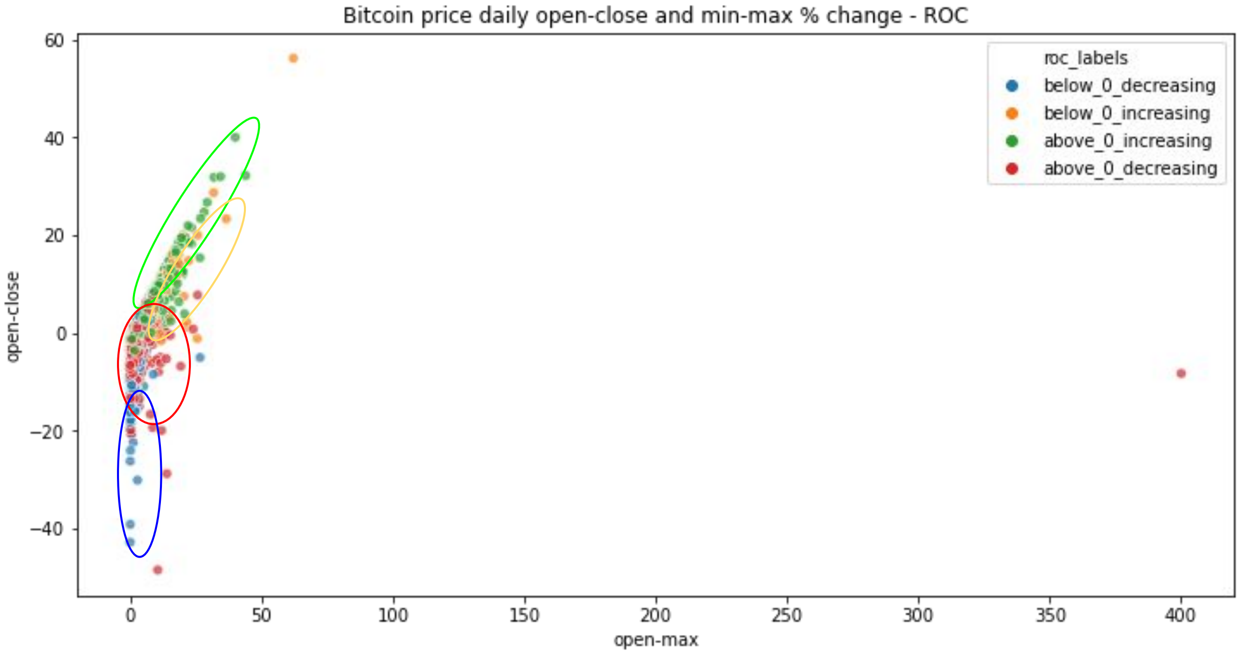


Our Data source is the minute-to-minute bitcoin price and exchange volume from Bitstamp’s API. Bitstamp is a cryptocurrency exchange company.

From this kind of data of each day we extracted the features used for modeling.

Features extracted include mean,, standard deviation, minimum, maximum, z-score, min-max range, exchange volume, technical indicators, open-close % change, etc).

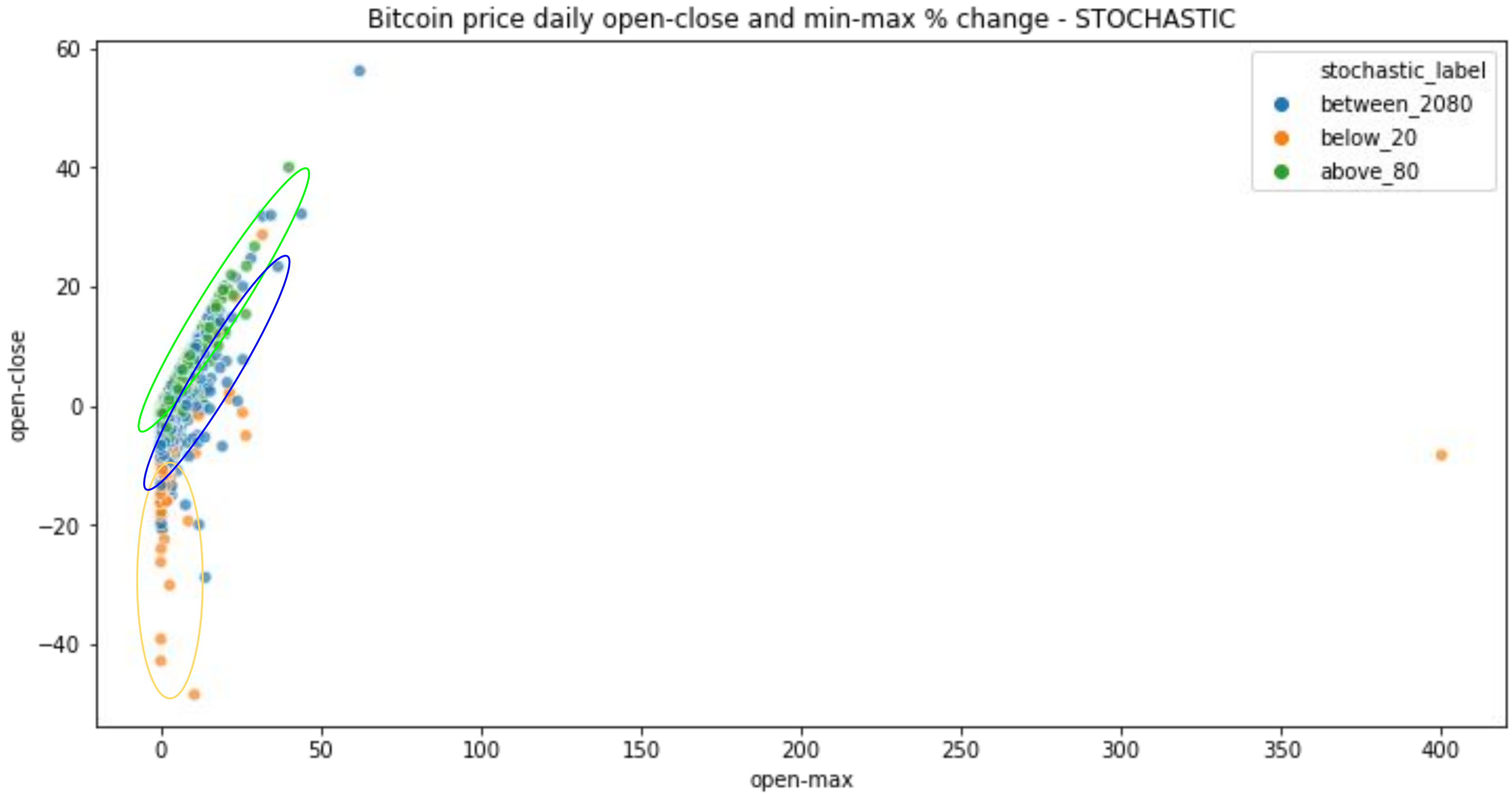
EDA was mostly focused on finding possible clusters of intraday price movement statistics and technical indicators.



The chart shows a scatter plot of the % change from the daily open to the daily close (open-close in the y-axis) price and the daily open to the daily high (open-max in the x-axis).

The color of the dots represent the level at which the Rate of Change (roc) technical indicator was in that day. We can see that the different colors (categories of the roc) seem to form clusters.

Plotting the same data using different technical indicators shows similar outcomes.



Bitcoin price daily open-close and min-max % change - STOCHASTIC
stochastic_label
between_2080
below_20
above_80
open-close
open-max
60
40
20
0
−20
−40
0
50
100
150
200
250
300
350
400

Feature engineering mostly consisted in building, categorizing and encoding technical indicators.

Building bitcoin price technical indicators.

These are indicators that traders and investors use to assess if a given asset is expensive/cheap, overbought/oversold.

The technical indicators used are: MACD, RSI, stochastic, ROC, Money flow Index. The indicators are oscillators, which can help us in the classification and clustering of daily price data.

Change technical indicators from numeric to categorical.

For example, the Money Flow Index oscillates between 0 and 100, so categories of 'above_80', 'below_20', 'between_2080' were given to the indicator.

Encoding technical indicators.

The categorical data of technical indicators were encoded using sklearn's OneHotEncoder.

Scaling.

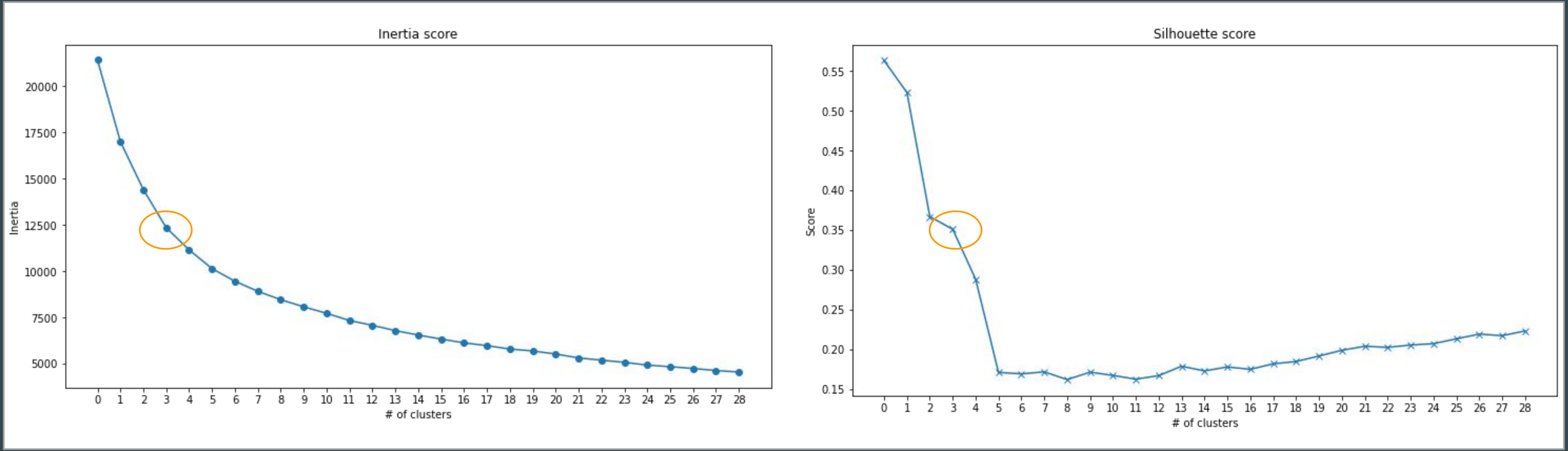
Numerical data of price movements was scaled using sklearn's StandardScaler.

Modeling and model selection

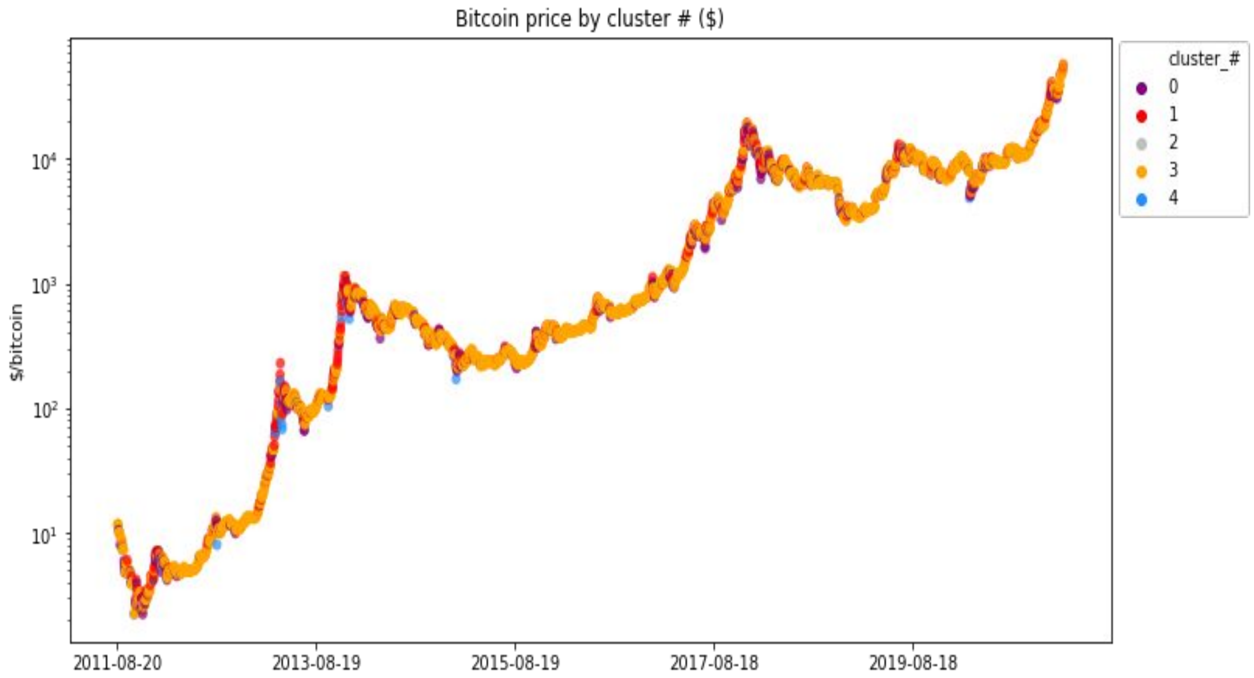
We used kmeans to find clusters in the data.

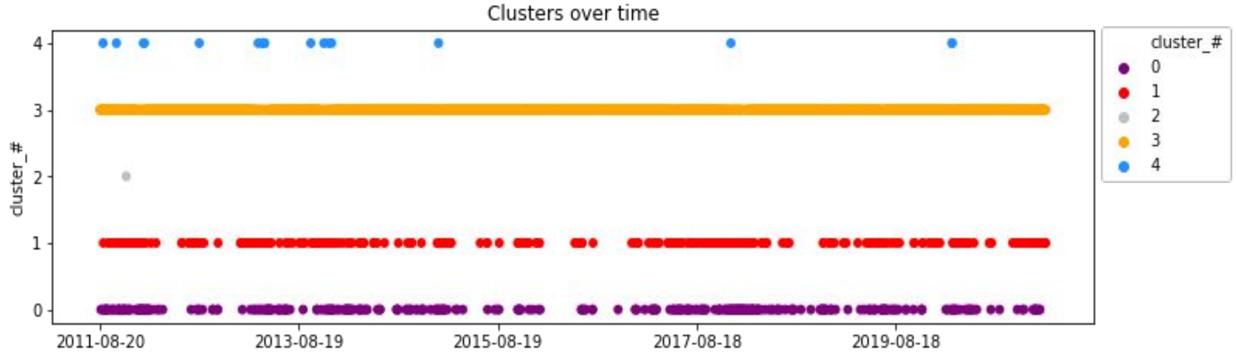
The model specification is: cluster # = f(intraday price movements, technical indicators).

According to inertia and average silhouette score, the number of clusters could be 5, where the decrease in inertia begins to slow down, but the silhouette continues to be relatively high.



Mapping cluster data to bitcoin price time series





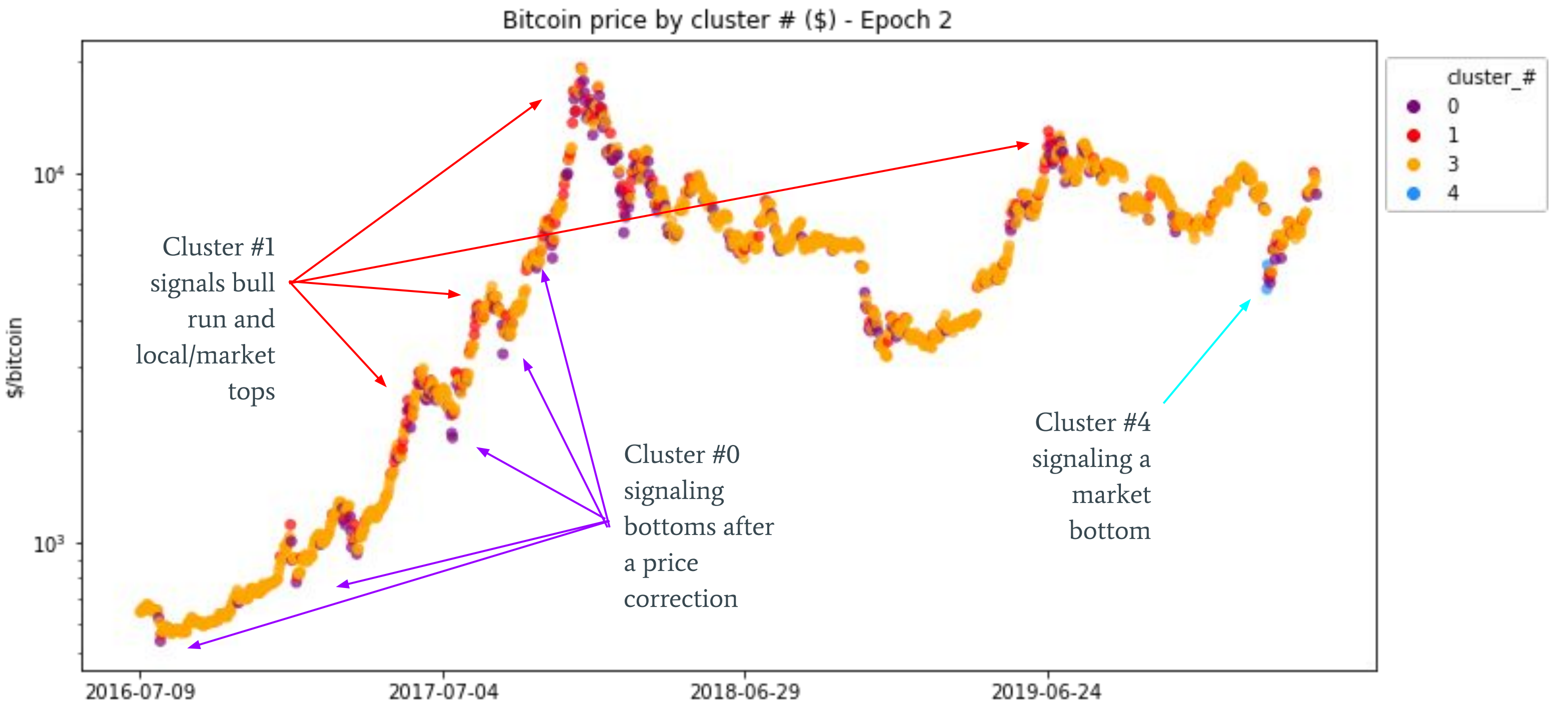
Clusters correspond to specific trends in bitcoin price.

- The majority of the days correspond to cluster #3.
- Cluster #1 shows times when the bitcoin price has increased sharply.
- Days in clusters #0 and #4 are typically seen after a correction in prices and/or at the end of bear cycles. Cluster #4 is less common.
- Cluster #2 has only one day (2011-11-25) which must be seen as an outlier.

Extracting cluster features can help us to categorize each one of the clusters

- **Cluster #3,** "normal days", low volatility, small price movements.
- **Cluster #1**, "bull days", volatility to the upside. This cluster also seems to signal market tops.
- **Clusters #0 and #4**, "reversal days", higher volatility, price reverses almost all the daily gains/losses. These type of days also seem to be also associated with correction/market bottoms.
- **Cluster #2**, "the outlier".

Zooming into a specific bitcoin price epoch we can see how clusters were related to price movements.



Conclusions

The daily bitcoin price can be categorized by its intraday movements and technical indicators.

- We found 4 types of daily price categories: “normal days”, “bull days”, “reversal days” (2 clusters) and “the outlier”.

By mapping cluster data to bitcoin price time series we can see how clusters are related to price movements.

- We found that cluster #1 (bull days) is associated with market tops, and that clusters #0 and #4 with market bottoms or end of price corrections. Cluster #3 corresponds to “normal days” (most common cluster and less volatile).

FUTURE ANALYSIS

Develop a quantitative model that predicts price movements using cluster data.	• Use supervised learning techniques to develop a model that predicts future price movements using cluster data in a specific time window.
Include technical indicators specifically related to bitcoin.	• We used technical price indicators used to analyze the price of any asset in general. However, technical indicators specifically related to bitcoin have also been developed (SOPR, Coin Days Destroyed, Puell Multiple, Realized HODL ratio, Realized profit, MVRV, etc).
Include other types of features related to the bitcoin network.	• We could add information about the bitcoin network to the dataset. This could include, daily data of active addresses, transactions, mining difficulty, issuance, etc.