Bitcoin Orderbook

Predicting BTC price from order book pressure

Rany Khalil
www.github.com/ranykhalil

Business Understanding

- This study is aimed at cryptocurrency price prediction using market bid/ask pressures within orderbook data collected from Kaiku.
- When trading on a specific exchange there exists two different orders a trader can set:
- 1- **Market Orders:** These are orders that are settled immediately at the current market price.
- 2- **Limit Orders:** Include orders that are set to trigger at specific price levels to either buy/sell an asset

Business Understanding

- Using the orderbook:
- "The term order book refers to an electronic list of buy and sell orders for a specific security or financial instrument organized by price level. An order book lists the number of shares being bid on or offered at each price point, or market depth.” Investopedia
- Using this data the prediction of the price movement can be monitored and predicted through changes in the orderbook reflecting buying support or selling resistance.

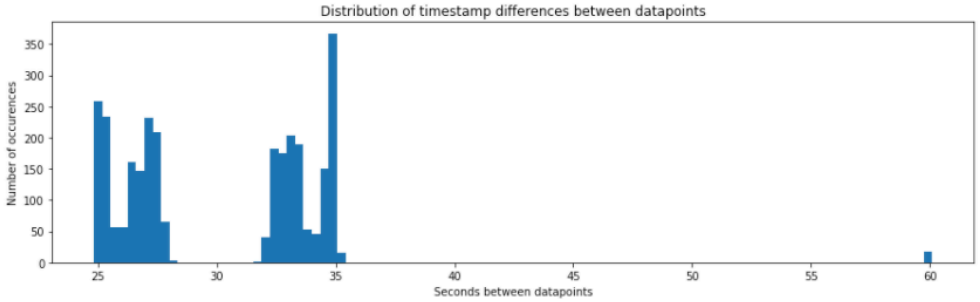
Data Understanding

- The order book data is represented to as can be seen to the right.
  - The order is identified as either a bid or an ask within the 'type' column. The price and amount are also present in the remaining two columns.
- The timestamp is in unix format with a millisecond accuracy.
- The dataset's timestamp frequency is not constant.

	date	type	price	amount
0	1567296016662	b	9600.85	0.058399
1	1567296016662	b	9600.01	0.225000
2	1567296016662	b	9600.00	0.104123
3	1567296016662	b	9600.00	0.021774
4	1567296016662	b	9600.00	1.000000
...	...	...	...	...
31007929	1567382368048	a	10743.20	0.004840
31007930	1567382368048	a	10744.40	0.001000
31007931	1567382368048	a	10745.00	0.001000
31007932	1567382368048	a	10745.46	0.001000
31007933	1567382368048	a	10747.00	1.000000

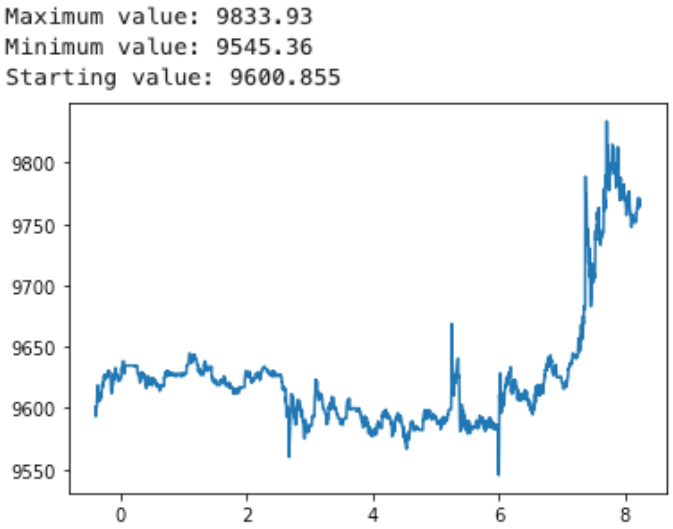
31007934 rows × 4 columns

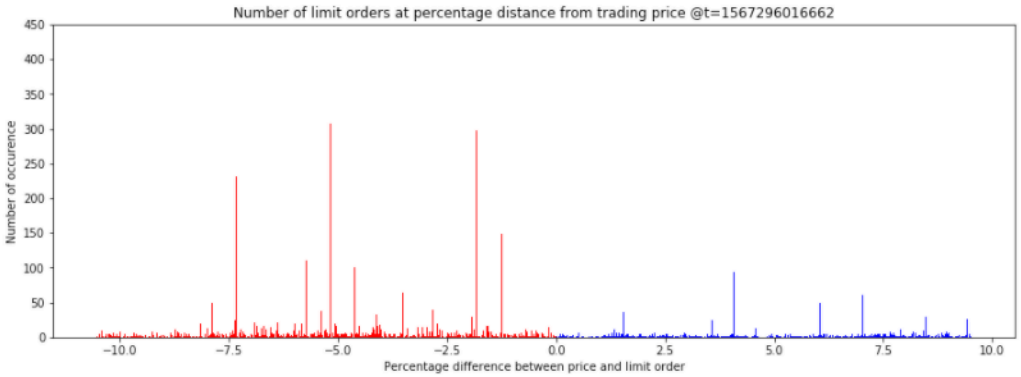
```
There are 2861 unique timestamps
The timestamp difference is not consistent and varies from: 24.839s to 60.088s
```



Data Preparation

- The price is calculated from the orderbook.
- The data amounts are then binned in bins relative to the current price with a percentage of 0.01%
- The bottom graph shows the buy pressure in red out is higher than the sell pressure in blue at the very first timestamp.
    - Validated from the top graph that saw the price move up after the first entry.





Modelling

- Since the data was not in consistent time intervals modelling included running logistic and linear regressions to ask as baseline to whether the data could be looked at as discrete values rather than continuous.

- The linear regression returned jargon, while the logistic regression returned a prediction with a 52% accuracy which was considered to be the baseline model.

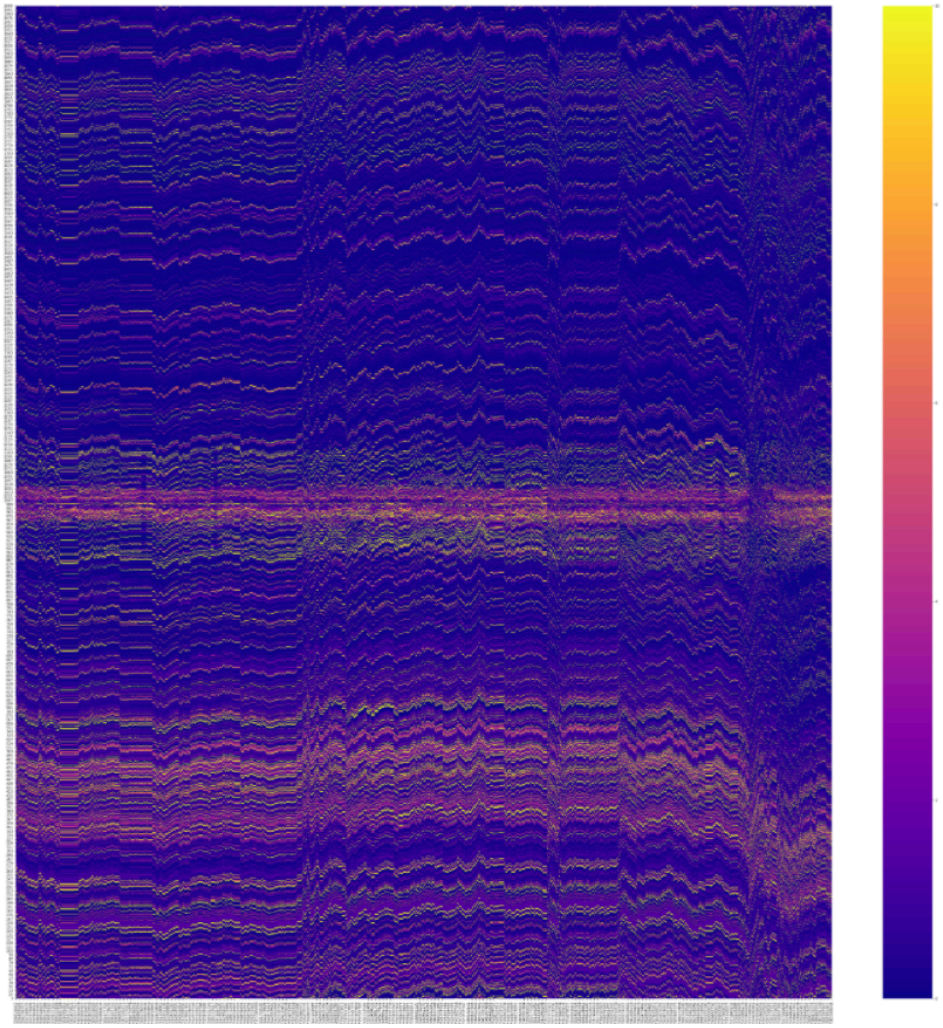
```
The model has returned its predicted values of the validation sample
The prdiction was wrong 211 out of 458 entries
An accuracy of 53.93%
```

	validation	linear_predicted	linear_abs_difference
1830	1	0	1
1831	1	0	1
1832	0	0	0
1833	1	0	1
1834	1	0	1
...	...	...	...
2283	0	0	0
2284	0	0	0
2285	1	0	1
2286	0	0	0
2287	0	0	0

458 rows × 3 columns

Evaluation

- The buying and selling pressure as time moves indicates the level of interest within the markets.
- This change of interest is directly translated to a change in price.
- The orderbook containing all this data should be capable of identifying the movement of the price as well as the price levels if prices were to move.



Future

- Combining orderbook data with actual executed orders could be considered

- Creating bins within return values as a possible range identifier with a separate model.