**Rationale on normalized image generation from a universal LOB snapshot**

Following were the main problems that we encountered at designing conversion of LOB to an image where convolutions can be applied.

1. Number of price levels are not constant?

* Number of price levels vary throughout the day.
* Data deep down in the book are generally not useful as the liquidity in here is far from executable at this point.
* Possible options to solve,

1. Look at the stock you are dealing with and cap it at a number that seems reasonable for this stock. Drawback is you will have different values for different stocks.
2. [Slightly preferred approach] Cap all stocks at a certain number. e.g.20. If you don’t have 20 price levels and only have 5, set the remainder to 0. Do note that majority of the trading firms consume top 10 OB levels only and from previous research I think they’ve found that going beyond 10 levels is not giving them an advantage. So, you can look at top 10 levels too depending on the use case
3. Volume could be either order count on a price level or sum of actual volumes

Possible options,

* Sum of all volumes at a price point.
* Volume per order -> Color contrast to represent per order volume
* Count can be a separate variable but shouldn’t be volume.

1. Gap between price levels are not consistent

* Should be fine for now.
* Address if any issue arises at modeling time

1. What normalization methods we use for price and volume axis?

* Volume -> min max scaler (Consider outliers in volume)
* Price -> Robust scaler (Remove outliers)
* Scales must be same e.g. (0 - 1 or 0 - 100) or so.
* I’ve also experimented with min max scaler for both volume and price and it also worked reasonably well.
* My guess is that volumes had outliers. Prices didn’t so much in an order book.
* Also note that in this case you can do two things - normalize them for individual order books. Normalize across entire order book and then have individual book states. Both methods are good but depending on what results you need or what you are trying to find one maybe better than the other

1. Do we need a differentiation between buy and sell sides on the image?

* Differentiate by color
* Select 2 contrasting colors
* Can try without colors initially to verify that this improves model performance

1. Can we use a chart (Ex: - Bar chart) itself to generate images?

* Bar charts (widely used, friendly to human eye, good novel experiment)
* Histogram
* Line chart with multiple price levels.

1. Non-even sampling on time axis?

* This is okay for now. Temporal dynamic might have an impact on low liquid stocks.