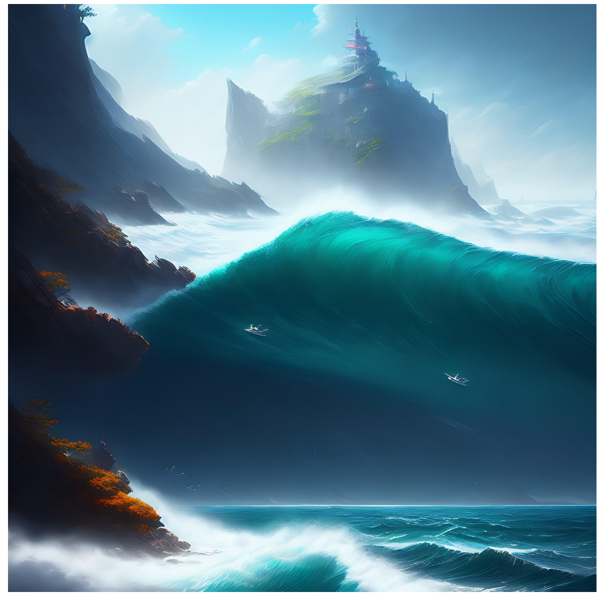
‘You Never Swim In The Same River Twice’
– Heraclit’



MDT (Market Day Type)

Problem:
A common mistake of investors/traders is to approach and treat each Market Day condition with the same strategy.
We have identified and grouped 5 basic types of days, and will try to infer/predict any common similarities and differences and if the first hour of trading (9:30-10:30 EST) can predict the type of day.

‘One of the funny things about the stock market is that every time one person buys, another sells, and both think they are astute.’ – William Feather
‘What seems that the Market wants to do- and how good a job is it doing?’ – Tom Williams

The Data Obtaining/Wrangling/Cleaning/Preparation

The raw data was exported by NinjaTrader C# script of T&S into 1 csv file-per-day and imported into pandas, comprising 280+ days of tick based values. The data consists of 2 CME Futures indexes with tickers: ES (S&P 500) and Nasdaq (NQ). Holidays, Fed news days(mostly), partial hours days and incomplete sets were discarded. Out of the approx. 250 remaining, 5 MDT groups were made:

Down, RangeDown, Range, RangeUp, Up.

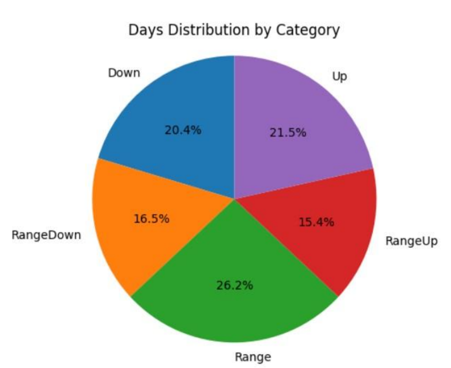
The sets had to be resampled from tick to minute data, for easier access and handling. Thus, sets from 1.5M to 2M rows per day were resampled to OHLC bars of 360 rows data frames per day, based on 12 hr days (4:20 AM - 4:20 PM) of 2 minute intervals. The sets, were approached in few different batches, as follows:

1. Dividing the NQ from ES and grouping them into 5 MDT sets made of approx. 75,000 rows for NQ and 85,000 rows for ES. Missing values were imputed, where appropriate with min, mean or 0 or were linearly interpolated. Infinite values were interpolated or ‘ffilled’/’bfilled’, as appropriate. NaN values were either dropped or ‘ffilled’/’bfilled’ imputed, for minimal effect on the frames.
2. 5 MDT groups were combined resulting in: 53 Down, 43 RangeDown, 68 Range, 40 RangeUp & 56 Up days. 40+ features were derived and later reduced by PCA to 18.

3. 5 MDT groups were normalized around 0, 35+ features (some overlapping with the previous batch, as filtered by the PCA, some newly derived), charts/plots for each were drawn and Univariate, Bivariate, and Multivariate plots were made, between features and groups to inspect and detect differences/similarities, convergences/divergences. Interesting findings, to be articulated later, were observed.
4. 5 mixed ES/NQ so called: redacted sets, partially scaled and partially original values, were produced, as large sets of approx. 170,000 rows, with potential for future analysis. Because of the vast amount of data, days, folders, rows and columns, all data sets were batch processed, meaning that folders were accessed by functions, individually tailored to specs- functions consisting of other functions, that consisted of yet another layer of functions.
   The DRY principle was unavoidable and the KISS principle often prevailing in the choices that were made.
   E.g. Special attention was made for Prices to be z-scored based on the daily range, instead of the entire range of min/max values from the entire set. This was a source of (seemingly) endless (sprouting) issues, especially when dealing with 250+ days at once, since many +inf and -inf values were sprouting, both in derived or z-scored columns.

The 5 MDT groupings do reflect general observation notions on Market habits. E.g. in the period observed (2022-2023) Up days were 0.06% more prevalent then Down days, which reflects that traditionally, Markets do advance more Up then Down. Most common were Range MDT days with 26% of days, which again makes sense, since Markets typically retrace and range, even while trending. RangeDown and RangeUp days (as defined by Range day with 100% retrace of starting Price that finishes Up/Down) were relatively symmetrically distributed with 16.5% and 15.4%, respectively. Considering that Up vs Down days, were also symmetrically distributed with 21.5 and 21.4, and Range (or mixed) days (defined as days when the Markets had moved up or down, only to retrace 100% and finish at the very starting level), we can draw conclusions, like the tenets:

- ***'Markets are a 0 sum game!'***
- ***'For every buyer, there is a seller- and vice versa!'***
- ***Central Limit Theorem (CLT) observation/justification at large.***
- ***`Efficient Market Theory` embodied!***



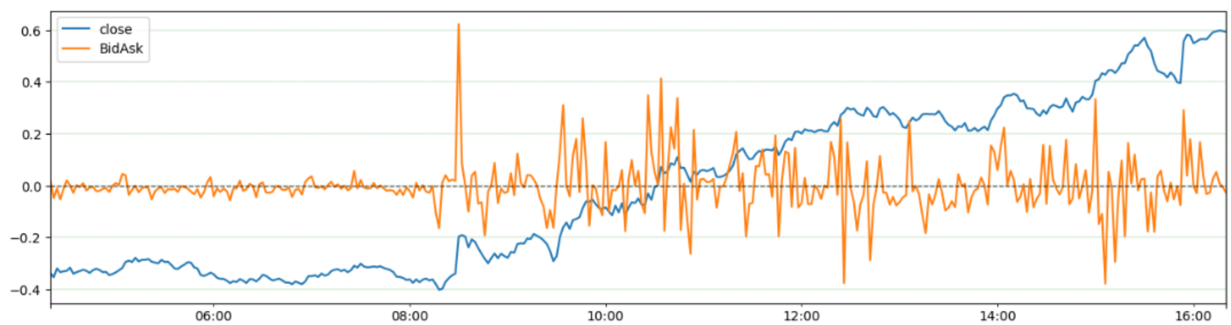
Feature Engineering

While developing batches 2, 3 and 4, out of 7 general fields (time series DateTime index, BidAsk, OHLC Price & Volume) features were developed, such as,

- **the traditional:** MACD, Simple/Exponential Averages, Stochastic Indicators, OnBalanceVolume, ADX, etc
- and **uncommon derivatives** such as: BidAsk Oscillators, BidAsk means, BidAsk-Volume combinations, BidAsk/ PriceDifference, BidAsk/PriceChange/Volume, etc.

E.g. a viable plot observable conclusion example that offers a possible trading/investing edge would be:

- On normalized data, Up days are featured with greater activity of BidAsk above the 0th axis, as shown in the pic. When and if, Price can not overcome the divergence between larger negative BidAsk pull and insufficient Price drop from that pull- then Price will continue upward. The same conclusions will apply in a Down MDT days. This sole conclusion can be useful in establishing sufficient market edge, provided it is followed with proper trade management (limited drawdown StopLoss) and patient observation without (FOMO).



More of other observations and conclusions are delineated in the Jupyter nb's, along with examples, as EDA unfolded.

ML Models in predictive analysis & the metrics

Unfortunately, in spite of the attempt to segment the Markets onto MDT days and find possible congruent predictors of Price direction, no positive predictive models worked to distinguish with any merits greater than 51% if the hours of 0:30 Am to 10:30 Am may predict the type of days that will unfold. Actual results between various models were:

- 34 f1-score with the SVM (lowest from all models)
- 34 to 39 f1-score in ANN (they underperformed in this problem like the SVM)

- 47 f1-score with the XGBoost Classifier
- 43 to 47 f1-score in Decision Trees classifiers
- 48 f1-score with the Random Forest Classifier

On this MDT classification problem, simpler models performed better.
More precise metrics are available in the git repository nb's.

For exercise, out of curiosity, ANN models were applied to a Strategy, based on inputs of simpler features: OHLC, returns projected in the future (return.shift(-nth)) with slightly better results, then the MDT classification of f1-score of 50. This points to the fact that some predictive room may be open to Models, dealing with Strategizing and Selective, rule-based approach to trading/investing. After all, that's what common sense teaches as well: a good, well-timed strategy will overperform in the long run.

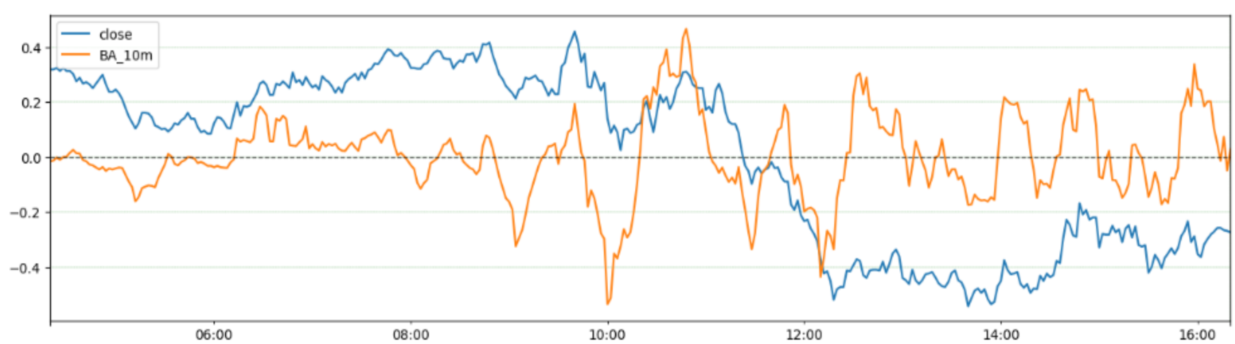
General Conclusions

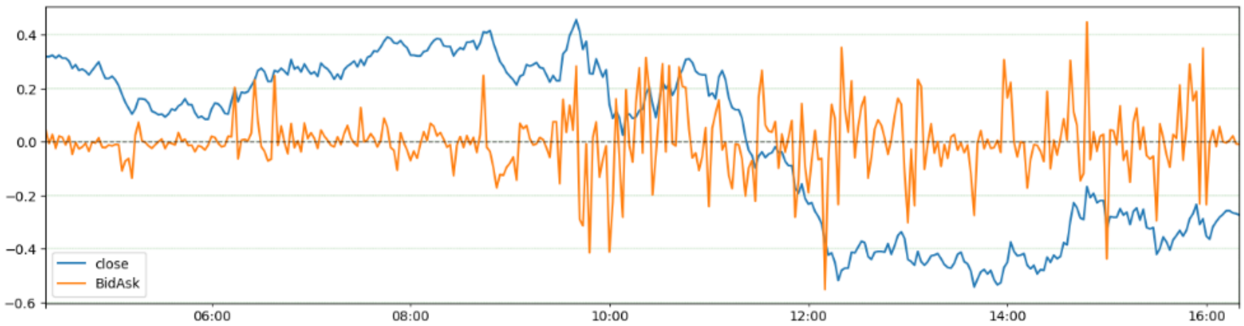
Many of the intrinsic desires and plans for this project did not pan out to be performed or when performed did not produce results that warm the heart. Like the SVD analysis was performed on too many features and did not produce results that would point to meaningful signature type of comparison, like we had in mind at the outset of the project. Yet, some possible footprints of these MDT Signatures showed some interesting traces yet remaining to be examined, but this study will be exempt from it, since that in itself is a project and more of great lengths.

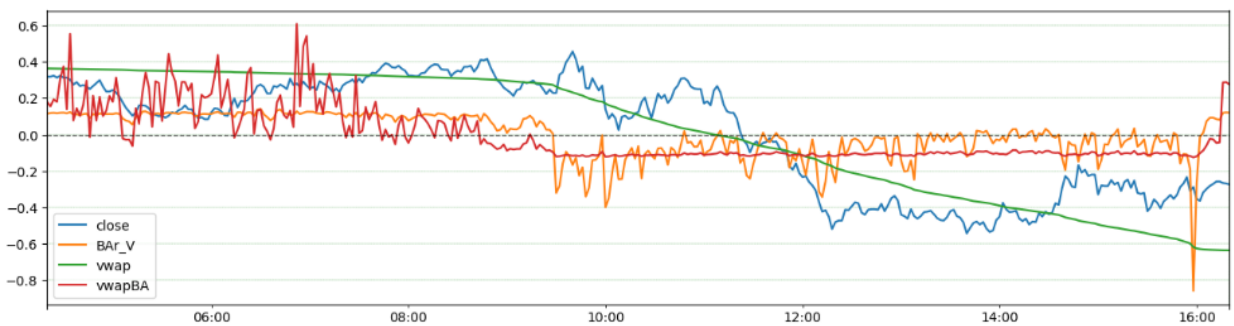
On the positive side, most interesting were visual chart examinations for scaled (normalized) features, vs 1, 2 or more derivative features, showing a promise of creating a new indicator/oscillator, that when read a certain way, filtered by certain condition, may offer edge in trading/investing. Such are the following observations, many of which offer competitive edge, and which for the same reason are stated in their cryptic original form:

```
-BidAsk in Range (in relation to VWAP) goes to 1300 while in Up goes to
2200. Extreme continuations of trend warrant greater intensity of trades/
volume. Of course, we could have learned that from basic info() and
describe().
- On Down days, intense activity in 'BA_5m', 'BA_cs' under 0, pulling
Price down. Possible Indicator/Oscillator. TBD
- Look for Intersection of 'close','OBV', 'obBA' for imminent breakout D.
```

```
- vwapBA most of Down day under 0. Compare to others!!!
- Look for Divergences between 'close' and'BA_10m'. Inability of BA_10m to
elicit significant retrace has propensity to result in a Downward move.
- 'BAr_V' and 'vwapBA' mostly constant under 0 while Down
- 'BA_5m' more under 0; if strong
- Important: Repeat of same level of 'BA_5m' @ -0.4 or 0.4, tends as
Res/Sup and Price retraces
- Huge swing in BA_10m (from -.6 to .4) without much effect on Price
retrace entails continuation of Trend, in this case Down (TRADE:
DIVERGENCE BETWEEN PRICE TOPS AND  BA_10m TOPS ENTAILS BIGGER PRICE
CONTINUATION DOWN)
- 'BAr_V' & 'vwapBA' before Down move 'Collect Contracts' above 0, and
sink under 0 before Price plummets
- Important!!! (TRADE: 'BA_cs' ABOVE AND BELOW 'close' MAY INDICATE
REVERSALS AND TEND TO CONTINUE- LOOK FOR REVERSALS AT  'BA_cs' EXTREMES,
LIKE -.6 IN DOWN M. OR .4 WHEN SWANG IN D.M.)
- Important!!! (2-ND/3 SWING UP IN BA_5M, FROM DOWN INTO ABOVE 0 INDICATE
REVERSAL AND RETRACE INTO POSITIVE TERRITORY) BA_5M MAY CONSTITUTE AN
INDICATOR PREDICTING FUTURES PRICE DIRECTIONAL MOVES.
- BA_cs crossed down by Price indicates imminent directional Price move.
Also important: How far is Price below/above BA_cs.
```







These pics are samples of, in the author's opinion, competitive conclusions drawn for Down Market days, as featured in the cryptic notes above. We can describe them dynamically, with a question: How does Price behave in relation to the derived features in the pictures? Are there signs of early activity above 0, visually preparing the fall at 9:30, the start of day?

However, the negatively predictive results of the MDT analysis do align with, and justify the tenets of the 'Efficient Market Theory', stating that markets are unpredictable and that Price optimally reflects information at any given point of time.
Based on the first 30-60 minutes of the day, Markets **can not be predicted**, for the rest of the day, with greater than 48% accuracy. That falls short of even random choice, or is at least very near a random walk observation result.

Why is that the case?

One can speculate on the reasons behind it:

- Because of the presence of the retail sector and institutional sectors, made of small, mid and long term type of investors/traders, including the almost instantaneous reactionary moves of ML Algos, Markets can change "mood" and direction at any time- instantaneously. New funds can be injected at any time or siphoned out, negating any previous slow seemingly meticulous progress in any direction.

- Because Markets intrinsically do not know ahead of time what they will do or what kind of day they may turn out to be. They unfold, like a river affected by gravity in its downstream, by a least-resistance fluctuations, only to surprise by its opposite swing, like, e.g. too much accumulation of negative bias, may get converted into positive over-reaction, in the so called 'short squeeze', when no more sellers are left and the Market changes direction, using as upward fuel the losses of exiting trapped sellers from minutes ago. The same goes for the opposite: too much positive exuberance leaves no more room for further growth, and at a certain pick tumbles down, in search of realizing made profits/gains, fuelled by the losses of exiting buyers.

  And such are the cycles of the Market undulations, minute in and out, day in and out, period in and out, fractally cycling in small and large, fueled by fear, greed, passion, strife and dreams.

For future projects/study/analysis:

- **Unsupervised Clustering techniques that initially offered not much success, later on, with better prepared datasets, may offer better results.**
- **Feature means differences in descriptions between certain periods of the days vs other periods of day may offer some predictive analysis, yet to be examined.**
- **Finding Market Edge is possible, according to this author, and after all, despite the dangers and unpredictability of oceans, fishermen still do throw their nets, equipped with proper tolls, knowing when to fish and what to avoid, like in rule based strategy approaches, for a given conditional context.**
- **The nb's are full with more then 2 X 250+ images/charts/plots of days, belonging to a certain segment of the Market. In their normalized state, these samples can be a source for visual training of a CNN model, for a future project.**
- **The variables and permutative combinations of multivariate models and searches on these topics are practically endless, without necessarily promising good results- and we had to conclude the efforts on this particular Market Holy Grail search.**