Data Visualization - Assignment 1 - Oculi

Part 1: Project Description

The data visualized in this report relates to the topic of financial analysis, particularly financial markets and macroeconomics. The various perspectives provided aim to respond to several problematics faced by discrete market-makers (as compared to automatic/systematic) such as traders in banks and by investment professionals such as portfolio managers in hedge funds and alike enterprises.

We chose to focus on financial markets and macroeconomics because of the interdisciplinary nature of the field, which encapsulates elements from mathematics, computer science, sociology, history, and economics. To understand financial markets is somehow akin to fathoming a facet of human experience both from an analytical/cognitive and a psychological standpoint. The sheer dynamism of this social construct and its linkage with our societies' development - political, economical, cultural - has made it a topic not only of captivation, but of discovery as well. Our topic

Moreover, purely from a data stance, the variety of financial information and their relative ease of access makes this topic even more interesting. Several free packages implemented in Python, as well as one of group member's access to the Bloomberg Terminal, the mother of all repositories for financial data/information, smoothen the data collection process, allowing us to focus more deeply on the object at hand :

How to visualize the link between asset classes ?

The quintessential metric in any financial analysis is **return** (on investment). A second, all the more precious to market makers, is **correlation**. Keeping these two gauges of performance in mind, in what follows will be described exactly what we collected.

Dataset: **Assets Price Behavior from January 2015 to February 2024.**

Given the interdependency between different financial asset classes, we proposed to include the most important 'tickers' of all classes. All financial assets are broadly divided into: Equities, Fixed Income (Credit, Rates, Bonds), Foreign Exchange (& Cryptocurrency), and Commodities. We also included Exchange-Traded Funds (ETFs), which are instruments that track the performance of the assets it is composed of (for example, a Tech-ETF would track the performance of selected technology sector assets, such as: Apple, Nvidia, Meta, …). Regarding the geographical distribution of our chosen assets, we preferred to tailor our selection process to the market capitalization of the respective companies, rather than focusing on one specific region - globalization obliges. As such, we included companies like Samsung, Alibaba, Louis Vuitton, Airbus, Safran, amongst others.

Overall, we selected: 63 equitie (60 companies, 3 tickers), 3 Credit/Rates ETFs, 4 Commodities ETFs, 3 Cryptocurrencies:

- In Equities, we covered 4 different sectors (15 tickers each) : technology, consumer discretionary, financial services, and industrials. Technology is included as a gauge of greed, and acts as an outlier for volatility and correlation measurements. Consumer Discretionary echoes the underlying economy and the strength of the consumer. Financial Services reflect the health of the monetary system (think failure of Silicon Valley and First Republic banks last year due to exponentially increasing interest rates). Industrials act as the baseline for returns and correlations - the backbone of an economy, and where most of its fundamental health is uncovered.

- Regarding Equity Indices, we only chose American ones : S&P500 (most capitalized companies), Dow Jones (industrial), and Russel (proxy for the American economy). Other choices such as the Nikkei (Japan) and Stoxx (Europe) could have been included, but given the generally higher levels of coverage and involvement (in terms of volatility) in American indices, for analysis of correlation particularly, it is more judicious to select more mercurial instruments.
- In Credit/Rates, we included instruments tracking the performance of 1) the 20 year US bond (gauge for economic security and confidence), 2) high risk credit instruments, and 3) low risk credit instruments.
- In Commodities, we included the main ETFs sub-categories: Energy (Oil Futures Contracts), Precious Metals (Gold and Silver futures contracts), Industrial Metals (Copper futures contracts), and Agriculture (diversified selection of futures contracts : Corn, Wheat, Sugar, etc…)
- In Cryptocurrency, we selected Bitcoin, Ethereum, and XRP. We did not select any foreign exchange (euro, dollar, yen) yet - we obviously must, and will be improved in subsequent versions.

A more thorough description of all these instruments can be found in the Annex.

Despite the data perhaps appearing unbalanced, the key and most revealing macroeconomic dynamics can be revealed through looking at the Fixed Income and Commodities tickers we chose, while equities' behavior is more scattered as varying on rather idiosyncratic factors. Additionally, performance in many Fixed Income products can be 'seen' in the performance of assets in the financial services sector. Moreover, in Commodities specifically, the Agricultural ETF does a fairly good job at representing the state of inflationary, supply, and demand forces in aggregation. As for Precious Metals, they were chosen as a measure of confidence in the future (as the traditional rule stands : the higher Gold prices, the higher sense of fear. Finally, Cryptocurrencies were included as a gauge of greed, as they exhibit extreme levels of volatility based on no sound fundamental factors.

While it is true that several factors are 'interdependent' and exhibit multicollinearity in their **absolute** behavior, lags in performance make their **relative** behavior different. Financial markets are a science of flow, and unless Federal Reserve Banks start printing money as no tomorrow exists (hm, Covid), capital cannot flow from everywhere to everywhere : this lag in return, essentially, is a key feature of markets - co-integrated movements occur in sequential waves. More must be studied about this feature, but this project is a good starting point to exploring the nature of this human construct.

Data Transformation:

The data obtained from Yahoo Finance and Bloomberg was unfortunately not cleaned. Several issues existed, and in the following we list them, along with their respective solutions:

Note : Unfortunately, and due to access limitations, the smallest frequency accessible is daily. No high frequency analysis was thus performed.

- Heterogeneous datasets retrieved for similar start and end date. This has to do with structural market factors (whether the exchange is open on a date or not,...). For example, Cryptocurrencies trade 24/7 while all exchanges are closed during weekends, and bank holidays differ between countries.
  Solution : Match datasets by date index.
- Different prices shown : Open, High, Low, Close. Also, overnight (i.e., when the market is closed to the 'public') moves were retrieved as Open_t+1 - Close_t .
  Solution : Focus on Closing Day Prices.
- Dealing with NaN values for common dates.
  Solution : Rolling means between two engulfing dates. Harder when there are 3 dates missing, solve it with a recursive mean strategy : assume 5 dates, and date_1 and date_5 available. Calculate Date_3 as mean between date_1 and date_5, and date_2 as mean between date_1 and date_3. Harder when the number of missing dates is even.

Computation of which the results were used for our problematics :

- Calculate log-returns from prices : LN(Price_t / Price_t-1)
- Calculate PNDV : Price Normalized by Daily Volume (cool indicator, not that used traditionally for financial analysis, and not included (for now) in answer our questions). Yet, in our python file, very interesting graphs could be plotted by overlapping price movements and PNDV.

Note:
There are no URL/references to the dataset used.
The merged file can be sent if needed.
To retrieve any stock price :

import yfinance as yf
df = yf.download(ticker, start_date, end_date) – default frequency is Day if date range is superior than 7 Days.

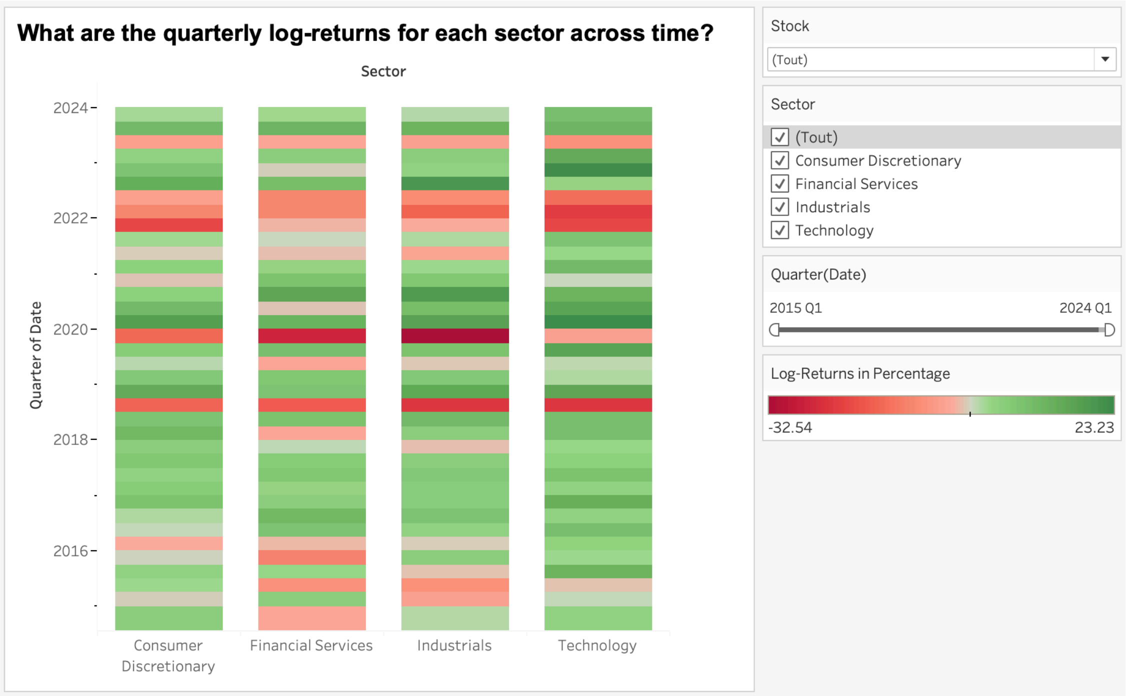
Questions tackled by each group member :

Killian : What are the quarterly log-returns for each sector across time?
(**Sectoral Return Analysis)**

Nishant : What is the proportion of returns attributable to a specific stock from investments on a basket of stocks ? **(Idiosyncratic Return Analysis)**

Ghali : How are different assets correlated with one another ?
(Cross-correlations Analysis)

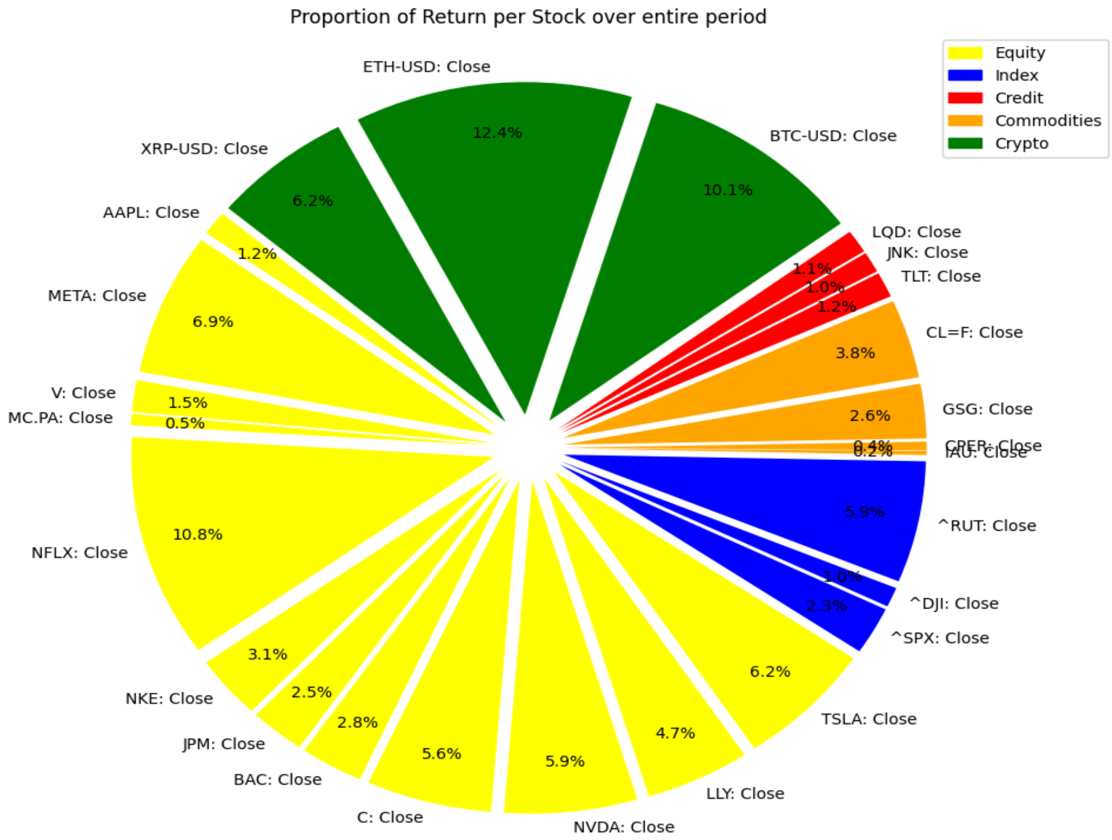
Killian LE METAYER



This heat map undertakes a comparative review of quarterly log-returns from four distinctive sectors: Consumer Discretionary, Financial Services, Industrials, and Technology, from 2015 to 2024. The magnitude of log-return in each cell is presented by the intensity of the used color for the corresponding quarter and sector. It is green when returns are positive and red when returns are negative. The color pattern across the heatmap provides a visual narration of the performance dynamics of each sector and indicates periods when the economy grows, stabilizes, or downturns. The above visualization summarizes nearly a decade's financial data in such a way that all stakeholders could understand at a glance not only which of the sectors shows more volatility and, therefore, higher risk but also general tendencies within each sector and deviations from a tendency that occurs in a selected period of time.

The use of heat maps seems appropriate to the data and research question, since it enables simple comparison across time (quarters) and at the same time across sectors. This shall be done using a diverging color scale from green (with the possibility to change to blue in case of color vision deficiencies) through white to red, which would help in immediate interpretation of positive, neutral, and negative financial return rates. This color encoding leverages intuitive associations (especially in finance) of green with growth and red with decline, thus furthering the ease of interpretation. The vertical layout of the temporal axis, therefore, gives a natural line of the eye, just as tracking changes through time, as in reading a timeline. Sectors are spread on a horizontal plane, allowing for easy comparisons between them. By standardizing the color encoding across all sectors, it facilitates the detection of outliers and patterns without the need for cross-referencing or external data interpretation tools. The design choices coalesce to transform a dense dataset into a story about sectoral resilience and economic cycles, promoting a rapid yet comprehensive understanding of complex market behaviors over an extended period.

Nishant DAVE



Pie Chart - Proportion of Return per Stock Over Entire Period

For the pie chart, we are visualizing the contribution of each stock to the portfolio's return on the last day of the period. Here's how it was created:

Final Day Returns: We take the returns of each stock on the final day and apply the equal weight to them to reflect their contribution to the portfolio's return.

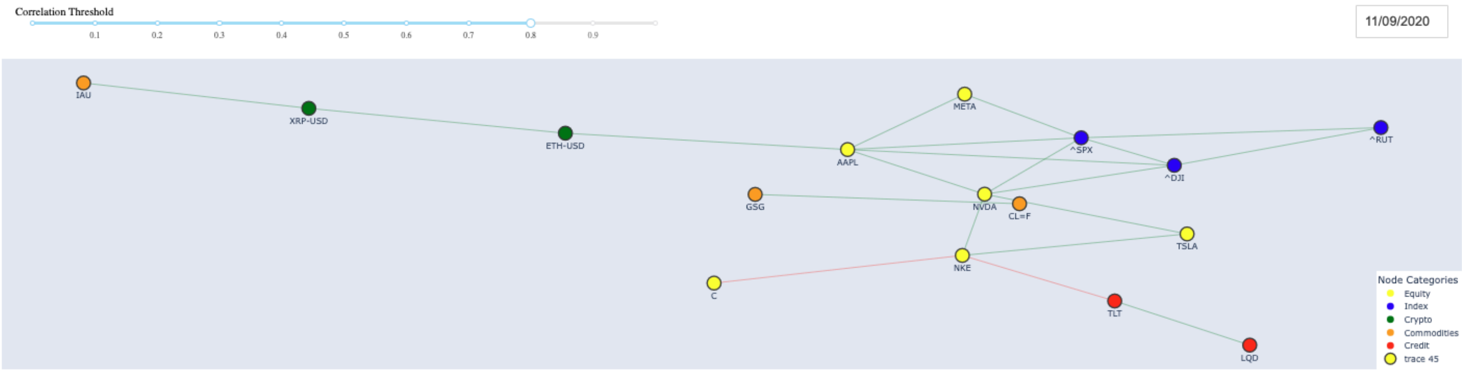
Pie Chart Creation: The pie chart is then constructed to show these contributions as proportions of the whole. Each 'slice' of the pie corresponds to a stock, with the size of the slice representing the stock's absolute contribution to the portfolio's return on the last day.

Visualization: This visualization helps to quickly identify which stocks had the largest impact on the portfolio's performance for the final day. It's particularly useful for assessing which assets and which asset classes drove the performance.

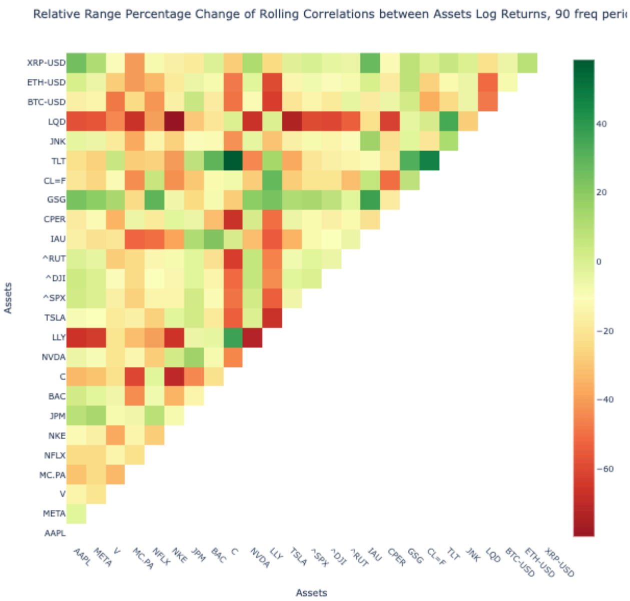
Ghali LARAQUI

Some plots

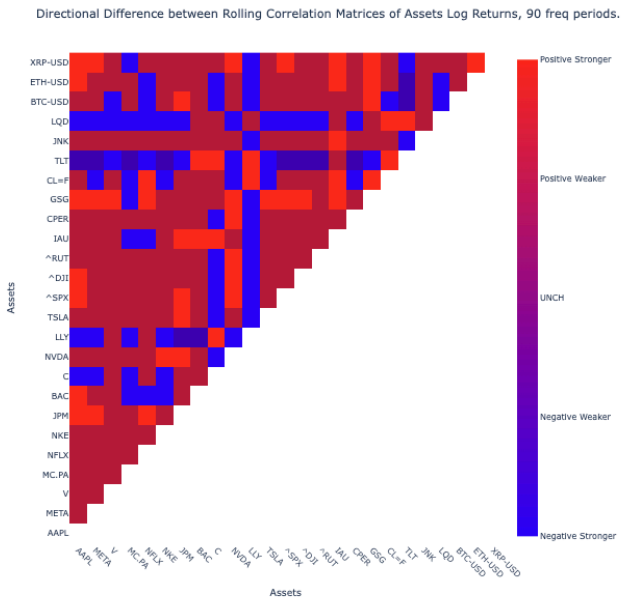
Network Graph Visualization



Heatmap, Relative



Heatmap, Directional



This rendering is a WebApp created with PyDash. Two inputs are chosen: correlation threshold, and date. To understand these figures, I must first elucidate what is calculated in the first place.

All correlations depend on a variable called ‘span’, in this case span = 90 for quarterly log-return analysis (not yet modifiable on the dashboard). What this does is, given a date chosen by the user, a correlation matrix is calculated for the previous *‘span’* days, and then visualized in the network graph. What is more interesting is what happens in the heatmap (they are both needed).

The first computes the relative range % change of rolling correlations between assets’ log returns. What is at first computed is two correlations matrices for consecutive time periods of length *span* days. Then, I wanted to compute the % change between these correlations.

What is innovative about this visualization is the formula behind the “relative range % change”. An issue I encountered was that I could not compute % changes between a positive and negative number, and in fact, no method is generally accepted. My solution was to discretize ranges of possible values, and given correlations vary between -1 and 1 (giving a range of 2), that was relatively straightforward (no need to discretize the space). Thus, we can interpret a green square as such : the correlations of (daily) log returns between XRP-USD and AAPL over the first span_period (the older) and the second span_period has gained >60% of “correlation power”. To know the new correlation, one must do the following : old correlation + relative change * abs(range).

The second heatmap makes this analysis more specific by indicating the direction of absolute change between the correlations of different span-periods. It informs us whether the correlation has gone positive strongly or weakly, unchanged, or negative strongly or weakly. The value added here occurs when we have a stock of which the relative change exhibits a negative value, but the correlation is either weakly or strongly positive. **For example, the 90-days rolling relative range correlation change between V (Visa) and META** (prev. Facebook) **is mildly negative, but the correlation between these two instruments is actually positive, which means over the past 180 days, their returns are less correlated to one another (Heatmap 1), but they still incur positive correlation, albeit weaker (Heatmap 2).** You can notice that, for the case of extreme moves (i.e. colors), values/interpretation overlap for both heatmaps (though not always) - it is rather for mild values that heatmap 2 becomes useful as it tells us if the instruments move together or in opposite directions.

This type of analysis is useful for derivatives market-making, particularly if the products designed are of the exotic type (an example would be a variance swap between two portfolios) or simply for systematic statistical arbitrage strategies and their monitoring.

<u>Part 3: Findings</u>

The visualization carried out in this assignment reveals that there are no clear cuts to categorize price behaviors between different asset classes, but there emerges interesting insights despite this fact.

When looking at the first graph, we see that the equities sub-categories are well behaved in general : periods of positive returns for one sector are overall mirrored in others, and the reverse also holds true. Yet, there are few periods of idiosyncratic performance, when fundamental factors impacting one sector are not systemic and thus make the performance a relative ‘outlier’. It all depends on the granularity with which we look at our data.

Indeed, if we look more closely, for instance Q4 2019 more closely, we see that, while the returns for all equity sectors is positive, when looking at the figure “Heatmap, Relative”, and particularly the assets belonging to the Equity category, not all stocks move together, regardless of positive or negative direction. This divergence is due to how the first figure aggregates entire sectors, and so individual behaviors are not reflected. The pie chart finally brings this into more light: aggregating behaviors across time, we lose short term fluctuations, and see that all assets, if invested in at first, bring in positive returns.