Comments on the project:

Data Quality

The Overall data quality was good. I had access to all the prices with almost no missing values. I only had missing values for some copper future prices that I filled with the previous day observation.

I had to perform some modifications to the data. I converted LB to MT and USD/LB to USD/LB, and it can be scaled easily for other cases. In more complex cases I would store the master data somewhere else to handle bigger amount of conversions. The Volume and Unit column had a naming problem with some spaces at the beginning and end of the word so I had to create handling for this.

I realised at some point that expiry for the Prop copper had a maturity that was different than other and was 30/12/2025 instead of 31st. I had no price for that so I figured I would switch to Month-Year of Maturity instead to match the prices properly.

There is a small test for outliers performed that shows no outlier. This test no where enough to analyse the dataset, and we should in a second step make a much more complex analysis of the prices, as these are fundamental in the project.

Assumptions

Normal Distribution of Returns
Asset returns are assumed to follow a multivariate normal distribution. This allows the use of standard deviation and correlation to estimate risk.

Stationarity of Returns
The historical return distribution (mean, variance, covariance) is assumed to remain stable and representative of future risk during the lookback period.

Linear Portfolio Approximation
The portfolio value is assumed to change linearly with asset price changes. Non-linear instruments (e.g., options) are not properly captured.

No Autocorrelation
Daily returns are assumed to be independent from one day to the next.

Historical Covariance
Covariance matrix is estimated from historical returns, typically over a fixed lookback window (e.g., 1 year), and used to approximate future co-movements.

Instantaneous Liquidity
Positions can be liquidated immediately without market impact or slippage.

Constant Portfolio Composition
The positions are assumed not to change during the VaR horizon (i.e., static portfolio assumption).

Confidence Level and Horizon
The VaR estimate is based on a chosen confidence level (e.g., 99%) and time horizon (e.g., 1 day). Multi-day VaR is scaled using the square root of time rule: $\sqrt{T}*VaR$

Coding part

I focused on my code being efficient and modular. For that I separated the part on loading the data from the excel in a database. I separated an other part for the processing and cleaning of the data sets. I created some functions but there is probably room for other more complex data handling. For now that dataset was quite small, as there was only 3 business lines and very small positions.

I also created separately all the utilities for the VaR calculations as this is the best practice. I decided to call all of these functions in the VaR.py main script. I did keep some functions there as the project is small, but optimally would make it even more modular to have only a few lines in the main script. I focused on making the project scalable, by allowing for selection of the confidence in the VaR and the look forward period with variable T as well as the loopback period that is set by default to 365 days. I also allowed for the set up to select other level than just the business line if needed. For example we could look at specific trader VaR.

I decided to automate the saving of the excel file output for the VaR results. In a normal set up I would have created a folder where we would be able to save for each day and each business line a specific excel file with a specific naming. This way, we would be able to plot the simulation and backtest later on.

Results

Zinc/Lead: The resulted VaR here is quite small and it makes completely sense here as the strategy in this business line is a 1 month time spread over the same product. The Result seem aligned with theory

Prop: Here doing a quick sanity check I think it makes sense as well that the VaR is a lot higher because we have a flat price position in copper, which is together with a time spread.

Copper: Here the result is really not correct in my opinion. We have an arbitrage strategy that tries to benefit from the mispricing on either LME or COMEX exchange. As the maturity and product are the same the VaR should be low. I do not know all the specs of these contracts but from what I saw the settlement time, seems really different (5 hours of difference) and that could explain the higher VaR. Since the product can vary a lot during a day, 5 hours of timid difference could explain such a spike in the VaR. I would recommend using the same timing to estimate the VaR for the whole portfolio.