Comments on the project:

Deliverables

- Github repo is made available publicly and can be cloned directly on your computer.
- The SQL database was produced using the ingest script directly in python. There is though a sql file showing the structure of the SQL database and schema.
- As outputs we have the excel file automatically produced from python and this file giving additional insights on the project.
- The README is available on the GitHub directly.

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/MT, 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 for parametric VaR

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.

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

<u>Confidence Level and Horizon</u>
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. The VaR has been adjusted to be performed under parametric assumptions and by using historical method which is less constrained.

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 by Business line Parametric

<u>Zinc/Lead:</u> 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 results seem aligned with theory.

<u>Prop:</u> 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. The nominal value is high so the 1-2% changes in prices can completely justify it.

<u>Copper:</u> 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 timing difference could explain such a spike in the VaR. I would recommend using the same timing to estimate the VaR for the whole portfolio.

Results by Business line Historical

The overall analysis is similar. Results do make sense for Zinc/Lead and Prop, but are too high for Copper. The timing of the LME compared to COMEX is probably the biggest factor. Nevertheless,

we can see that the overall portfolio VaR is a lot lower with historical method. This could be explained by high volatility and not observed tail events in the past year. In this case VaR would be understated as it can only give results on what was seen during the lookback period. While the parametric method would extrapolate even unseen events. In any of the 2 methodology, VaR for copper business line is definitely too high.

Historical Methodology

I decide to use the same base code for the historical method. I added some other functions to calculate the P&L moves. I decide to look at quantiles of the P&L variation instead of quantile of returns, because for some portfolio like Zinc/Lead, Total portfolio value was so close to 0 that results would be extreme using the returns.

There were a few prices missing for copper DEC 25 future. I decided to drop these 6 P&L Value because it is a small enough amount of observations. I could implement taking the previous day price for it like I did for the parametric version. I would maybe need to understand if it was a holiday or if it is just some missing prices to make these decision.

Again, I aimed to make it modular and scalable, as we would be able to change the confidence level and lookback period.