Procedure Followed to Solve this Problem.

DATA ACCUMULATION AND STUDY:

1. The data was diversified across multiple tables. Almost each table had some key column that was related for our business problem. The data was in CSV.
2. This database was created using duckDB in which these tables were injected along with logging functionality.
3. All the columns from different tables which were of some benefit to the problem were noted.
4. The tables which were for related to our interests were joined only on the related columns leaving all other non-relevant columns and filtered to create one single aggregated table using sql query embedded in python so that our analysis can be done in better way.
5. Now this whole procedure was taken care of using python scripts. These scripts were saved to run them again automation if there arises need to ingested data on intervals.
6. The insights obtained after this aggregation to optimize the performance were:

Performance Optimization

- The query involves heavy joins and aggregations on large datasets like sales and purchases.
- Storing the pre-aggregated results avoids repeated expensive computations.
- Helps in analyzing sales, purchases and prices for different vendors and brands.
- Future Benefits of storing this data for faster Dashboarding And reporting.
- Instead of running expensive queries each time, dasbboards can fetch data quickly from vendor_sales_summary

7. After the aggregation, the final table was checked whether it had any inconsistensies like missing values or improper data types.
8. Some extra columns were created that will help us solve our problem in better manner like GrossProfit and ProfitMargin.
9. A new table was created in the database called vendor_sales_summary. Which had same columns as the aggregated table. But no data yet.
10. This final vendor_sales_summary table was ingested into the database under the newly created table.
11. In this manner, the final workable data was created.

EXPLORATORY DATA ANALYSIS - PYTHON:

1. Now, during this reasearch we found out there were some entries that were of no benefit to us like the entries in which totalSalesQuantity = 0. So this time the aggregated table was fetch using filteration with WHERE clause in the embed SQL in python.
2. A HISTOGRAM was plotted to check the data dispersion inside the table. We found out that most of the tables were right skewed which meant there were more smaller values and less bigger values. Maybe because of premium brands.
3. A BOXPLOT was plotted using seaborn to check for the outliers in each column. We realised that there were negative entries in TotalSalesPrice and GrossProfit columns. We eliminated the bad performing entries from the table as they will not benefit us in any manner.
4. Then we plotted a CORRELATION HEATMAP in which we realized facts like which columns are closely correlated and which are not correlated at all.
5. Then we plotted the top 10 vendors and brands that were performing better than others.

EDA:

1. We identified the brands that needed promotional or pricing adjustments which exibit lower sales performance but higher profit margins.
- We did this by determining the brands which had low totalSalesDollars but very high Profit margin. We plotted them on SCATTER PLOT.
2. We plotted the top performing brands and sales using BAR PLOT and determined how much of the total data was help by these top performing vendors which was found to be 66%.
3. A PERETO CHART was created to demonstrate the purchase contribution and the cumulative contribution of these top brands.
4. Using a BOX PLOT we understood the benefits of getting the products in bulk.
5. We solved the rest of the problem questions accordingly.
6. Then we identified what are the confidence intervals between the top-performing vendors and low-performing vendors.
7. We also performed the hypothesis testing for the difference in profit-margin with respect to the top-performing and low-performing vendors.

DASHBOARD -TABLEAU:

1. We created a two panel tableau dashboard which displays the KPI's as main message of the dashboard.
2. On first panel we showed target brands and top performing vendors and brands along with filters.
3. On the second panel we showed the data covered by these top brands and the PERETO CHART to demonstrate the Purchase contribution and cummulative contribution to the totalPurchases.

REPORT:

1. We created a report That contained all our insights along with some screen captures that showed our process of working and the need of doing what we did.
2. We showed different charts as well.
3. In the end, We have advised some Final Recommendations to the stake holder on this business problem

----------xxxx----------xxxxxxxx----------xxxxx---------