**Part 1: Data Cleaning and Summary**

1. Identify and list all columns with missing values. How many missing values are there in each column?

2. Replace missing values in the following manner:

• For numerical columns, use the median of the column.

• For categorical columns, use the mode of the column.

3. Some columns appear to have numerical values stored as strings (e.g., Total Assets, Net Income after Tax). Convert these columns to appropriate numerical formats and handle any errors that arise during conversion.

**Part 2: Data Integrity Checks**

4. Check if there are any duplicate rows in the dataset. If duplicates exist, remove them.

5. Validate that all dates in the dataset are within the range of 2008 to 2023. If any dates fall outside this range, list them and remove the corresponding rows.

**Part 3: Derived Metrics (Expanded)**

6. Calculate a new column Debt Ratio, defined as the ratio of Total Liabilities to Total Assets. For rows where Total Assets is zero or missing, replace the value of Debt Ratio with NaN and explain why this adjustment is necessary in financial terms.

7. Derive a new column Earnings Quality:

• Compute it as the ratio of Net Income after Tax to Operating Expenses - Total.

• Classify firms into three categories based on this value:

• High Quality (>= 1)

• Medium Quality (between 0.5 and 1)

• Low Quality (< 0.5)

8. Create a column Liquidity Index, calculated as:

• (“Cash & Cash Equivalents – Total” + “Working Capital”) / “Total Current Assets”

• Remove rows with invalid or missing inputs for this calculation. For rows with Liquidity Index below 0.1, flag them as Low Liquidity.

9. Generate a new column Profit Growth that shows the year-over-year percentage change in Net Income after Tax for each firm. For the first year in the dataset for any firm, set the value to NaN.

**Part 4: Data Aggregation (Expanded)**

10. Group the dataset by industry sector and:

• Calculate the top three industries by average Profit Margin.

• For each industry, identify the firm with the highest Market Capitalization and the corresponding year.

11. For firms with Net Income after Tax greater than the median value:

• Calculate their average Total Assets and Debt Ratio.

• Determine the percentage of such firms with High Debt Risk (calculated in Part 3).

12. Create a pivot table summarizing the total Revenue from Goods & Services, average Net Income after Tax, and standard deviation of Operating Expenses - Total, grouped by age and year.

13. For each firm:

• Calculate the 3-year moving average of Revenue from Business Activities - Total.

• Identify any firms with a declining trend in this metric over the last three available years.

**Part 5: Business Insights (Expanded)**

14. Identify the top 10 firms with the highest Profit Margin. For these firms:

• Determine their average Cost of Goods Sold / Sales, %.

• Find the year with the lowest Debt Ratio and the corresponding value.

15. Compare firms across age (industries) in terms of Market Capitalization. For each industry:

• Identify the firm with the most volatile Market Capitalization (measured as the standard deviation over years).

• Calculate the percentage change in average Market Capitalization from the earliest to the latest year.

16. For firms with a Liquidity Index above 0.5:

• Calculate their average Gross Revenue from Business Activities - Total.

• Identify the firm with the largest gap between Cash & Cash Equivalents - Total and Short-Term Debt & Current Portion of Long-Term Debt.

**Part 6: Advanced Analysis and Challenges (Expanded)**

17. Standardize the following columns using the z-score method: Net Income after Tax, Total Assets, and Market Capitalization. Store the results in new columns. Then:

• Identify firms where at least two of these standardized metrics exceed 2.5 (outliers).

• For these firms, calculate the average Depreciation - Total.

18. Filter the dataset to include only firms that meet the following criteria:

• Positive Working Capital.

• Debt Ratio below the median value for the dataset.

• Profit Margin above 0.1.

Save this filtered dataset as a new DataFrame and describe its characteristics (e.g., number of firms, industries, and average financial metrics).

19. Yearly Revenue and Profit Analysis

For each firm:

• Calculate the year-over-year percentage change in both Revenue from Business Activities - Total and Net Income after Tax.

• Classify firms based on their trends over the years:

• Consistently Growing Firms: Positive percentage change in both revenue and profit for at least 80% of the years.

• Fluctuating Firms: Positive percentage change in either revenue or profit for between 40% and 80% of the years.

• Declining Firms: Positive percentage change in either revenue or profit for less than 40% of the years.

• For each category, summarize:

• Average Operating Expenses - Total.

• Total Market Capitalization.

• Percentage of firms with a Debt Ratio greater than 0.5.

20. Perform a year-over-year comparison of Operating Expenses - Total and Revenue from Business Activities - Total:

• Calculate the expense-to-revenue ratio for each year.

• Identify years where the ratio increased compared to the previous year for more than 50% of firms.

**Instructions:**

• Submit your Python code as a .ipynb file, with inline comments explaining your approach for each question.

• Provide the final cleaned dataset and summary tables in .xlsx format.

• Ensure your answers include explanations for financial interpretations of derived metrics and preprocessing steps.

• Put all of your files in a folder, named: studentID\_name, and compress that folder, submit into LMS.