Data Visualization

Assignment - 1

Theoretical Questions

1. Purpose and Usage of Bar Graphs in Data Visualization

Bar graphs are charts used for comparing categorical data. They display data in rectangular bars where the length of each bar represents the value of the category, making it easier to see differences and trends.
In Tableau, we can use bar graphs to compare categorical data by dragging a categorical field (e.g., Region) to the Columns shelf and a measure (e.g., Sales) to the Rows shelf. For example, using the 'Sample Superstore' dataset, we can create a bar graph to visualize total sales across different regions, enabling a clear comparison.

2. Limitations of Pie Charts in Data Visualization

Some limitations of pie charts are:

- Difficulty in interpreting data with multiple categories.
- Difficulty in accurately comparing slice sizes or angles.
- Difficulty in reading smaller slices.

Pie charts are use for showing part-to-whole relationships when there are fewer categories (e.g., two to five). For example, in the 'Sample Superstore' dataset, a pie chart can show the proportion of total sales by Segment effectively.

3. Significance of Scatter Plots in Data Analysis

Scatter plots are used for analyzing relationships or correlations between two variables. They allow us to detect patterns, clusters, and outliers in the data.
Using the 'Sample Superstore' dataset, we use a scatter plot to explore the relationship between Discount and Profit. This reveals whether higher discounts correlate with lower profits and identify any anomalies in the data.

4. What is a Stacked Plot?

A stacked plot is a type of of a bar graph where each bar is divided into segments representing subcategories. This allows us to visualize the total value for each category along with the contribution of individual subcategories.
For example, in the 'Sample Superstore' dataset, we use a stacked bar plot to display Sales by Region, segmented by Category. This is more beneficial than a grouped bar chart when both the total and the composition of the categories are important.

5. Level of Detail (LOD) Expressions in Tableau

LOD expressions in Tableau provide precise control over aggregations. They define the level of granularity at which a calculation is performed.

- FIXED: Aggregates data at a specific level, regardless of the view's context.
- INCLUDE: Includes additional dimensions in the calculation.
- EXCLUDE: Excludes certain dimensions from the calculation.

Example:
In the 'Sample Superstore' dataset, if we want to calculate total Sales by Region regardless of the Category dimension in the view, we use a FIXED LOD expression:
{FIXED [Region]: SUM([Sales])}

6. Types of Joins in Tableau

Types of joins in Tableau:

- Inner Join: Includes only rows with matching values in both datasets.
- Left Join: Includes all rows from the left dataset and matching rows from the right dataset.
- Right Join: Includes all rows from the right dataset and matching rows from the left dataset.
- Full Outer Join: Includes all rows where there is a match in either dataset.

Factors to Consider When Performing Joins:

- Key Fields: Ensure the datasets share a common key field for joining.
- Granularity: Check for duplicates or missing data to prevent inaccurate results.
- Join Type: Select the appropriate join type to align with the analysis goals.

7. Creating a Calculated Field in Tableau

To create a calculated field:

- We navigate to the Data Pane, click the drop-down arrow near a data source, or right-click anywhere in the pane.
- Select Create Calculated Field.
- Enter a name and write the formula using Tableau's calculation syntax.

Example:
In the 'Sample Superstore' dataset, we can calculate profit ratio with the formula:
[Profit] / [Sales]

8. Difference Between a Dimension and a Measure in Tableau

- Dimension: Represents qualitative data, such as Region or Category, which we use to group or categorize information.
- Measure: Represents quantitative data, such as Sales or Profit, which can be aggregated.

Dimensions are typically placed on the Columns/Rows shelves or used for filtering, while measures are used in charts or calculations.

9. Creating a Dashboard in Tableau

To create a dashboard in Tableau:

- We click on the New Dashboard icon in the workspace.
- Drag and drop sheets onto the dashboard canvas.
- Adjust the layout and add interactive elements like filters, parameters, or actions.

Example:
Using the 'Sample Superstore' dataset, we can build a dashboard that includes a map showing Sales by State, a bar chart of Sales by Category, and filters for Region and Segment.

10. Data Blending in Tableau

Data blending is a method of combining data from two unrelated data sources. We use it when:

- The datasets do not share the same structure or key fields.
- One dataset is the primary source, while the other provides supplemental information.

In Tableau, the primary data source fields are shown in blue, while secondary data source fields are shown in orange. For example, we blend data from the 'Sample Superstore' dataset with an external benchmark dataset to compare performance metrics.

11. Purpose of a Parameter in Tableau

A parameter in Tableau acts as a dynamic placeholder that allows us to input a single value to modify calculations, filters, or visualizations. Parameters enhance interactivity and flexibility in dashboards.
Example:
Using the 'Sample Superstore' dataset, we can create a parameter to adjust a discount threshold dynamically and visualize its impact on Profit.

12. Performing Data Aggregation in Tableau

In Tableau, data aggregation is the process of summarizing data values, such as summing, averaging, or counting. Aggregation is applied automatically when we drag a measure to the view.
Steps:

- Drag a measure (e.g., Sales) into Rows or Columns.
- Use the drop-down menu in the field to select the desired aggregation (e.g., SUM, AVG).

Example:

Aggregating Sales by Region shows total sales for each region in the 'Sample Superstore' dataset.

13. Different Types of Charts in Tableau

Tableau offers various chart types for diverse data visualization needs:

- Bar Chart: For comparing categorical data.
- Line Chart: For trends over time.
- Pie Chart: For part-to-whole relationships.
- Scatter Plot: For exploring relationships between two variables.
- Tree Map: For hierarchical data visualization.
- Heat Map: For identifying patterns in large datasets.
- Gantt Chart: For project schedules.

Example:
We can use a tree map in the 'Sample Superstore' dataset to show Sales contribution by Category and Sub-Category.

14. Sharing Tableau Workbooks with Others

We can share Tableau workbooks in the following ways:

- Tableau Public: Publish workbooks for public access.
- Tableau Server or Tableau Online: Share workbooks with secured access to team members.
- Export: Export as PDF, PowerPoint, or image files for offline sharing.
- Packaged Workbook (.twbx): Share the workbook with embedded data sources for portability.

Example:
To share analysis based on the 'Sample Superstore' dataset, we can export a dashboard as a packaged workbook for others to open in Tableau Desktop.

15. Context Filter in Tableau

A context filter creates a dependent filter hierarchy. It sets the filtering context for other filters, improving performance and relevance.
Example:
In the 'Sample Superstore' dataset, we can create a context filter for Region to limit the data to the East region, ensuring subsequent filters like Category apply only to the East's data.

16. Difference Between Joining and Blending

- Joining: Combines datasets at the database level based on shared key fields. It works best when both datasets have the same granularity and structure.
- Blending: Combines datasets at the visualization level, where one data source acts as primary and the other as secondary. This is useful when datasets lack a common key or are stored in different systems.

Example:
In the 'Sample Superstore' dataset, joining might merge sales data with a detailed customer database, whereas blending might combine sales data with an external regional benchmark.

17. Use of Dual-Axis and How to Create One

A dual-axis chart allows us to overlay two measures on the same graph with different axes, useful for comparing trends with different scales.
Steps to Create:

- Drag two measures to the Rows shelf.
- Right-click one of the axes and select Dual Axis.
- Synchronize axes if needed for better alignment.

Example:
Using the 'Sample Superstore' dataset, we can create a dual-axis chart to compare Sales and Profit trends over time.

18. Difference Between .twb and .twbx Extensions

- .twb (Tableau Workbook): Stores only the workbook structure (visualizations, dashboards) but not the data. It requires access to the original data source.
- .twbx (Tableau Packaged Workbook): Contains the workbook along with embedded data, making it portable and shareable without requiring the original data source.

Example:
We use .twb for large datasets stored in a secure database and .twbx for sharing offline visualizations, such as a 'Sample Superstore' analysis.

19. Filter vs Parameter vs Context Filter

- Filter: Directly restricts the data displayed in the visualization.
- Parameter: A dynamic input value used in calculations, filters, or visualizations.
- Context Filter: Creates a dependent hierarchy for other filters.

Example:
In the 'Sample Superstore' dataset:

- Use a filter to exclude specific Regions.
- Use a parameter to set a dynamic Discount threshold.
- Use a context filter to limit data to a Region before applying subcategory filters.

20. Handling Big Data Technologies in Tableau

Yes, Tableau can handle big data technologies. It integrates with platforms like Hadoop, Google BigQuery, Snowflake, and Amazon Redshift. By leveraging live connections or extracts, Tableau efficiently visualizes and interacts with large datasets.
Best Practices:

- Use live connections for real-time analysis.
- Optimize queries with aggregation and filtering.
- Use Tableau's Hyper engine for faster processing of extracts.