1. Briefly state the importance of Data Visualization.

⇨ Data visualization is crucial because it helps to simplify complex data and present it in a clear and understandable visual format, such as charts, graphs, and maps. This enables quicker interpretation of trends, patterns, and outliers, supporting better decision-making. It also enhances communication by making data insights accessible to both technical and non-technical audiences. Furthermore, it assists in identifying relationships and correlations within the data that might be missed in raw formats. Overall, data visualization turns raw data into actionable insights, improving analysis efficiency and impact.

2. Define data visualization. Mention different Data visualization techniques.

⇨ Data visualization is the graphical representation of information and data using visual elements like charts, graphs, maps, and dashboards. It helps users understand trends, patterns, and insights from complex data more easily.

- **Bar Charts** – Used to compare different categories or groups.
- **Line Graphs** – Ideal for showing trends over time.
- **Pie Charts** – Represent parts of a whole or percentage data.
- **Histograms** – Show frequency distributions of continuous data.
- **Scatter Plots** – Display relationships or correlations between two variables.
- **Heatmaps** – Use color coding to show data density or intensity.

3. Why data visualization is such a powerful tool ?

⇨ Data visualization is a powerful tool because it transforms raw, complex data into visual formats that are easier to understand and analyze. It enables users to:

1. Quickly identify trends and patterns that may not be obvious in spreadsheets or text.
2. Simplify complex information, making it more accessible and understandable.
3. Enhance decision-making by clearly showing comparisons, relationships, and performance.
4. Engage users visually, improving retention and comprehension of information.
5. Future Prediction and Analysis Purpose

4. Define Data Analytics. State any two factors/drivers that are involve in the growth of data analytics and hence mention how they are involve with this growth.

⇨ Data Analytics is the process of examining, cleaning, transforming, and interpreting data to discover useful information, draw conclusions, and support decision-making.

Two Factors/Drivers Involved in the Growth of Data Analytics :-

Increase in Data Generation:
With the rise of the internet, social media, IoT devices, and digital platforms, massive amounts of data are being generated daily. This abundance of data creates the need and opportunity for businesses to analyze it to gain insights, optimize operations, and improve customer experiences.

Advancements in Technology:
The development of powerful computing systems, cloud storage, and analytical tools (like Python, R, and AI/ML algorithms) has made it easier and faster to process and analyze large datasets, fueling the growth of data analytics in various sectors.

5. Define Data Analytics Lifecycle?

⇨ The **Data Analytics Lifecycle** is a structured process that outlines the steps involved in analyzing data, from understanding the problem to delivering actionable insights. It helps ensure that data analysis is done systematically and effectively to meet business objectives.

The typical stages of the Data Analytics Lifecycle are:

1. **Discovery** – Understanding the problem, objectives, and available data sources.
2. **Data Preparation** – Collecting, cleaning, and transforming raw data for analysis.
3. **Model Planning** – Selecting the appropriate analytical techniques or algorithms.
4. **Model Building** – Developing and testing models using tools and programming languages.
5. **Evaluation** – Interpreting results and checking the model's effectiveness.
6. **Deployment** – Sharing insights through reports, dashboards, or integrating into systems.
7. **Feedback & Iteration** (optional) – Refining the process based on outcomes and feedback.

6. State different phases of data analytics lifecycle.

⇨

a) **Discovery** – Understanding the problem, objectives, and available data sources.
b) **Data Preparation** – Collecting, cleaning, and transforming raw data for analysis.
c) **Model Planning** – Selecting the appropriate analytical techniques or algorithms.
d) **Model Building** – Developing and testing models using tools and programming languages.
e) **Evaluation** – Interpreting results and checking the model's effectiveness.
f) **Deployment** – Sharing insights through reports, dashboards, or integrating into systems.
g) **Feedback & Iteration** (optional) – Refining the process based on outcomes and feedback

7. Mention how these phases are used to address big data analytics project.

⇨

- **Discovery:**
  In big data projects, this phase involves understanding the business problem and identifying the types of big data (structured, unstructured, real-time) required from various sources like sensors, logs, or social media.
- **Data Preparation:**
  Since big data is large and complex, this phase focuses on collecting data using tools like Hadoop or Spark, and then cleaning and formatting it using ETL (Extract, Transform, Load) processes.
- **Model Planning:**
  Suitable frameworks and algorithms for big data (like MapReduce, machine learning, or deep learning models) are chosen, considering the volume and variety of data.
- **Model Building:**
  Advanced tools and platforms (e.g., Spark MLlib, TensorFlow) are used to build scalable models that can handle massive datasets.
- **Evaluation:**
  The model's performance is validated using test datasets, and accuracy is checked to ensure reliable insights in a big data environment.
- **Deployment:**
  The final model is deployed into a big data system or cloud platform, often integrated with dashboards or applications to provide real-time insights and actions.

8. Mention stages of visualizing data. Explain any three steps of visualizing data in data science process.

⇨ The main stages of visualizing data in the data science process are:

- **Understanding the Objective**
- **Data Collection**
- **Data Cleaning and Preparation**
- **Selecting the Right Visualization Tools**
- **Choosing the Right Visualization Type**
- **Designing the Visualization**
- **Interpreting and Communicating Results**

Data Cleaning and Preparation:
This step involves removing errors, handling missing values, and transforming data into a suitable format for visualization. Clean data ensures accurate and meaningful visual outputs.
Choosing the Right Visualization Type:
Depending on the data and the objective, you must select an appropriate chart or graph—like bar charts for comparison, line graphs for trends, or scatter plots for relationships. The right choice makes insights easier to interpret.
Interpreting and Communicating Results:
Once the visualization is created, it's used to draw conclusions, identify patterns, and communicate insights to stakeholders. This step is essential for driving decisions based on the visualized data.

9. Explain 7 Gestalt principles of visual perception with the help of proper figure.

⇨ 7 Gestalt Principles of Visual Perception:

1. **Proximity:**

   Objects that are close together are perceived as a group.

   👉 *Used to group related data points in charts or dashboards.*

2. **Similarity:**

   Items that look similar (color, shape, size) are seen as part of the same group.

   👉 *In charts, using the same color for related categories reinforces similarity*.

3. **Continuity (Good Continuation):**

   The eye follows continuous lines or patterns.

   👉 *Line graphs use this to show smooth transitions or trends*.

4. **Closure:**

   The mind fills in missing parts to see a complete figure.

   👉 *Even if parts of a shape are missing, viewers still perceive the whole.*

5. **Figure-Ground:**

   Elements are seen as either the figure (focus) or the background.

   👉 *Used in dashboards to highlight key information over less important data*.

6. **Symmetry and Order (Prägnanz):**

   Viewers perceive objects as symmetrical and orderly.

   👉 *Clean, balanced visual design enhances readability and trust.*

7. **Common Fate:**

   Elements moving in the same direction are seen as a group.

   👉 *In animations or dynamic visuals, items moving together are interpreted as related.*

10. **Explain Tufte's Design Principles briefly. Explain the role of data ink ratio. Explain the situations where this principle contradicts the real situation.**

⇨ **Show the Data**: Emphasize the data itself over non-data elements like decorative graphics or excessive chart elements.
Maximize Data-Ink Ratio: Use as much "ink" (or digital space) as possible to represent data, minimizing non-essential ink (explained further below).
Erase Non-Data Ink: Remove elements that don't contribute to understanding the data, such as unnecessary gridlines, borders, or labels.
Erase Redundant Data-Ink: Avoid repeating data or using redundant visuals (e.g., showing the same data in multiple formats unnecessarily).
Revise and Edit: Continuously refine visualizations to improve clarity, similar to editing a written document.
Avoid Chartjunk: Eliminate decorative elements (e.g., 3D effects, excessive colors, or patterns) that distract from the data.
Integrate Words, Numbers, and Graphics: Combine textual explanations, numerical data, and visuals seamlessly to tell a cohesive story.

Role of Data-Ink Ratio :

- **Data-Ink Ratio** = **(Ink used to present actual data) / (Total ink used in the graphic)**
- A high data-ink ratio means **less clutter**, **more focus on data**, and **better clarity**.
- It encourages **removing unnecessary grids, frames, and decorative elements** to keep the focus on the actual information being presented.

Situations Where This Principle Contradicts Real Situations :

1. **Marketing Presentations:**
   Visually rich charts with colors, icons, and branding are used to **engage audiences**, even if they reduce the data-ink ratio.
2. **Beginner or Non-Technical Audiences:**
   Sometimes, **extra labels, legends, or pictorial elements** are added to make the chart more intuitive, even if it's less efficient in ink usage.
3. **Interactive Dashboards:**
   Modern tools add **interactive features, animations, and tooltips** which may increase non-data elements, but improve usability.
4. **Design for Impact:**
   In journalism or storytelling visuals, **emotional or visual impact** may be prioritized over minimalism.

13. Briefly Explain about real time issues in data gathering and preparation

⇨ **Data Inconsistency:**
Real-time data often comes from multiple sources (sensors, APIs, user inputs), which may have different formats, leading to inconsistency and difficulty in integration.

Data Quality and Noise:
Real-time streams may contain **missing, duplicate, or incorrect values**, which require immediate detection and cleaning to avoid misleading results.

Latency and Speed:
Processing large volumes of incoming data with **minimal delay** is challenging. Delays can affect the accuracy of time-sensitive decisions.

Data Storage and Scalability:
Continuous data generation puts pressure on storage systems and demands scalable architectures to handle and process incoming data efficiently.

Security and Privacy Issues:
Gathering real-time data, especially from users or devices, raises **concerns about privacy, authorization, and secure transmission**.

Q14. Explain "Data Cleaning" steps used during data pre-processing.

[Module2/PE-CS602C.2, Remember, LOCQ] [5 Marks]

Data Cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a dataset. Key steps include:

1. **Handling Missing Data** – Techniques like deletion, mean/mode imputation, or using predictive models.
2. **Removing Duplicates** – Identifying and dropping duplicate records.
3. **Correcting Inconsistencies** – Fixing data entry errors, format mismatches (e.g., date formats).
4. **Filtering Outliers** – Identifying values that deviate significantly from the dataset (using IQR, Z-score).
5. **Standardizing Data** – Converting data into a common format (e.g., all lowercase text, unit conversions).

Q15. Explain the requirement of dimensionality reduction of a dataset. Mention different data reduction strategies used during data pre-processing.

[Module2/PE-CS602C.2, Understand, LOCQ] [5 Marks]

Dimensionality Reduction is required to:

- Reduce **model complexity** and computation time.
- Improve **model performance** by eliminating noise.
- Enhance **data visualization** in 2D/3D.

- Avoid **overfitting** in machine learning models.

Data Reduction Strategies:

1. **Feature Selection** – Selecting important features using methods like filter, wrapper, and embedded methods.
2. **Principal Component Analysis (PCA)** – Projects data to a lower dimension using eigenvectors.
3. **Sampling** – Selecting a subset of data to represent the whole dataset.
4. **Aggregation** – Combining data (e.g., monthly to quarterly).
5. **Discretization and Binning** – Reducing continuous features into categorical bins.

Q16. State PCA algorithm for dimensionality reduction.

[Module2/CO2/Understand-IOCQ] [6 Marks]

Principal Component Analysis (PCA) Algorithm:

1. **Standardize the Data** – Normalize features to mean = 0 and variance = 1.
2. **Compute Covariance Matrix** – Captures the relationship between variables.
3. **Calculate Eigenvalues and Eigenvectors** – Eigenvectors determine principal components; eigenvalues show variance.
4. **Sort Eigenvectors** – Based on decreasing eigenvalues to rank importance.
5. **Select Top k Components** – Choose the top k eigenvectors.
6. **Transform Data** – Multiply original data by top k eigenvectors to get reduced dimensions.

Q17. Functions of matplotlib are similar to which programming language/package? Write the sequence of functions to draw a line chart with axis headings, graph heading, and red square markers.

[Module2/CO2/Apply-IOCQ] [5 Marks]

- **Matplotlib** functions are similar to **MATLAB**.

Python Code (matplotlib.pyplot):

```
import matplotlib.pyplot as plt


x = [1, 2, 3, 4]

y = [2, 4, 6, 8]


plt.plot(x, y, 'rs-')  # 'r' = red, 's' = square, '-' = line

plt.xlabel('X Axis')
```

```
plt.ylabel('Y Axis')

plt.title('Line Chart Example')

plt.show()
```

Q18. Define COV(X,Y) and Calculate COV(X,Y) for dataset. Explain the meaning of covariance values.

[Module2/CO2/Apply-IOCQ] [2+4+3 = 9 Marks]

Covariance (COV): Measures how two variables change together.

$$COV(X,Y)=\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

Dataset:

X	**16**	**10**	**30**	**20**	**24**	**6**	**34**
Y	10	7	17	12	14	5	19

Calculation (Summary):

- Mean of X = 20, Mean of Y = 12
- COV(X,Y) = 37.14

Interpretation:

- **COV(X,Y) = 0** → No linear relationship.
- **COV(X,Y) > 0** → Positive correlation; as X increases, Y increases.
- **COV(X,Y) < 0** → Negative correlation; as X increases, Y decreases.

Q19. Explain 5 principles of improving a vision of a plot/chart.

[Module2/CO2/Understand-IOCQ] [5 Marks]

1. **Simplify the Chart** – Avoid clutter, remove unnecessary gridlines.
2. **Use Consistent Colors** – Use meaningful and distinct colors.
3. **Add Labels and Titles** – Properly label axes, title the graph clearly.
4. **Choose the Right Chart Type** – Pick appropriate chart for data type (bar, line, pie, etc.).
5. **Highlight Key Data Points** – Use color/annotations to emphasize insights.

Q20. Explain the requirement of data cleaning during data preprocessing.

[Module2/CO2/Understand-IOCQ] [3 Marks]

Data Cleaning is essential to:

- Improve data **quality** and **reliability**.
- Ensure **accurate analysis** and model results.
- Remove **inconsistencies, noise, missing values** or duplicates.
- Enhance **model performance** and reduce errors.

Q21. What is meant by categorical variable? Mention string manipulation functions used for noise reduction with examples.

[Module2/CO2/Understand-IOCQ]

Categorical Variable: A variable with a fixed set of possible values (e.g., "Male", "Female").

String Manipulation Functions:

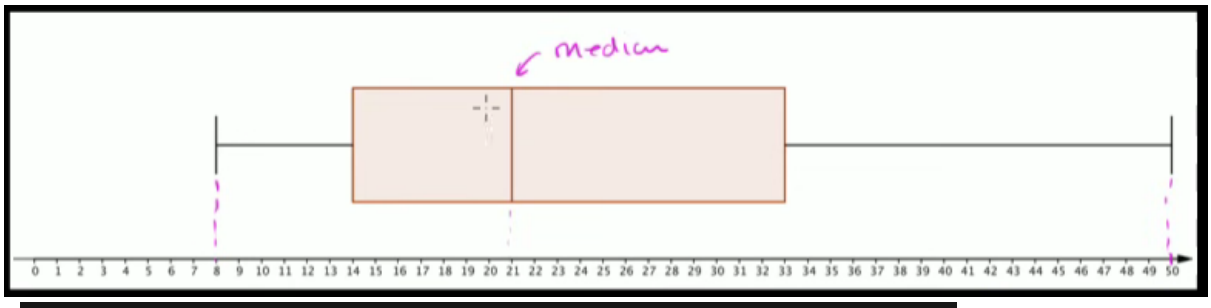
- str.lower() – Converts to lowercase → "MALE" → "male"
- str.strip() – Removes whitespace → " Female " → "Female"
- str.replace(" ", "_") – Replaces characters → "New York" → "New_York"
- str.title() – Capitalizes words → "data science" → "Data Science"

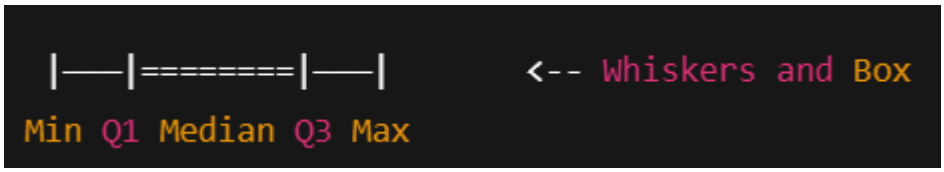
Purpose: To reduce noise and bring uniformity to categorical text data.

Q22. Diagrammatically represent Box & Whisker plot and mention measurable parameters.

[Module3/CO3/Understand-LOCQ] [5 Marks]

Box Plot:





Parameters:

1. Minimum
2. Q1 (25th percentile)
3. Median (Q2 – 50th percentile)
4. Q3 (75th percentile)

5. Maximum
6. IQR = Q3 - Q1
7. Outliers (if any)

Q23. Define Analytics Architecture. Mention and briefly discuss any four capabilities.

[Module3/CO3/Understand-LOCQ] [5 Marks]

Analytics Architecture – A framework that supports data storage, processing, and analysis.

Four Capabilities:

1. **Data Ingestion** – Importing from multiple sources.
2. **Data Storage** – Storing structured and unstructured data (e.g., Data Lakes).
3. **Data Processing** – Using ETL, Spark, MapReduce.
4. **Visualization/Reporting** – Display insights via dashboards (Tableau, Power BI).

Q24. Differentiate between Data Exploration and Explanation Analysis.

[Module3/CO2/Understand-IOCQ] [5 Marks]

Feature	Exploratory Analysis	Explanatory Analysis
Purpose	Discover patterns & trends	Communicate specific insight
Audience	Analysts, data scientists	General/business users
Focus	Open-ended exploration	Focused message or insight
Examples	Heatmaps, clustering, scatter plots	Annotated bar chart, story dashboard

---

Q25. Contrast: i) Stacked vs Grouped Bar Chart, ii) Line vs Area Chart

[Module3/CO3/Understand-LOCQ] [2+2 = 4 Marks]

i) Stacked vs Grouped Bar Chart:

Feature	Stacked Bar Chart	Grouped Bar Chart
Layout	Bars stacked on each other	Bars side by side
Focus	Part-to-whole comparison	Comparing across groups

ii) Line Chart vs Area Chart:

Feature	Line Chart	Area Chart
Use	Shows trends	Shows magnitude + trends
Design	Clean with lines	Filled area under line

Q26. Mention two ways to visualize hierarchical data and compare them.

[Module3/CO3/Understand-LOCQ] [6 Marks]

1. **TreeMap** – Nested rectangles; size shows value.
2. **Sunburst Chart** – Concentric rings; inner circle is root, outer rings are children.

Feature	TreeMap	Sunburst Chart
Layout	Rectangle-based	Circular/radial layout
Space Efficiency	High	Moderate
Readability	Harder in deep levels	Easier to understand hierarchy

---

Q27. Three visualization mistakes and define combination chart with example.

[Module3/CO3/Remember-LOCQ] [5 Marks]

Mistakes to Avoid:

1. **Wrong chart type** – E.g., Pie chart for too many categories.
2. **Overloaded visuals** – Too much data/clutter.
3. **Missing labels/legends** – Confuses interpretation.

Combination Chart:
Combines two chart types (e.g., line + bar) for dual insights.

Example:

- Bar for sales per month.
- Line for profit margin across same months.

28. Is it possible to plot categorical variables using the functions of matplotlib? Write the functions for plotting 3 categorical variables in 3 ways (line, bar, scatter) under one plot. [5]

Answer:
Yes, categorical variables can be plotted using matplotlib.

```python
import matplotlib.pyplot as plt

x = ['A', 'B', 'C']
y1 = [10, 15, 7]
y2 = [12, 18, 9]
y3 = [5, 10, 6]

plt.figure(figsize=(10,6))

plt.plot(x, y1, label='Line Plot')
plt.bar(x, y2, alpha=0.5, label='Bar Plot')
plt.scatter(x, y3, color='red', label='Scatter Plot')

plt.title('Categorical Variable Plots')
plt.xlabel('Categories')
plt.ylabel('Values')
plt.legend()
plt.show()
```

29. "Visual representation of data should be ethical" - Justify [5]

Answer:
Ethical visualization ensures that data is not misrepresented or misleading.

- Use proper scales and avoid distortion.
- Maintain data accuracy and honesty.
- Avoid manipulating visuals to push biased narratives.
- Ensure accessibility (color blindness, readability).
  Ethical visuals build **trust** and lead to **informed decisions**.

---

30 & 32. Depending on your dataset how will you select the chart type and colour scheme to be used?
=> To select the appropriate chart type and colour scheme, consider the following factors:

- Nature of Data: For categorical data, use bar or pie charts; for continuous data, use line or scatter plots.
- Comparison: Use bar/column charts for comparing quantities.
- Trend Over Time: Use line charts to show how data changes over time.
- Distribution: Use histograms or box plots to visualize data distribution.
- Part-to-Whole Relationship: Use pie or stacked bar charts.
- Hierarchy: Use tree maps or sunburst charts.
- Colour Scheme: Use a sequential color palette for continuous data (e.g., light to dark), a diverging palette to show deviation from a midpoint, and categorical colors for distinct groups. Make sure to maintain contrast and accessibility (color-blind-friendly palettes).

---

31. List five reasons why data visualization is important.

=>

1. Simplifies Complex Data: Turns large datasets into understandable visuals.
2. Reveals Patterns and Trends: Helps detect correlations, trends, and outliers quickly.
3. Supports Decision Making: Enables faster and more accurate business decisions.
4. Enhances Data Communication: Makes presentations and reports more engaging and clear.
5. Increases Retention: Visuals are more memorable than raw numbers or text.

---

33. Mention few points in favour of using Tableau software.

=>

- User-friendly interface with drag-and-drop features.
- Real-time data analysis and dashboard updates.
- Supports large datasets and multiple data sources.
- Strong community and learning resources.

---

34. The Tableau Product Suite consists of several parts.

=>

1. Tableau Desktop – For report creation and data analysis.
2. Tableau Server – To share visualizations securely within an organization.
3. Tableau Online – Cloud-hosted version of Tableau Server.
4. Tableau Public – Free platform for sharing data visualizations publicly.
5. Tableau Prep – For data cleaning and preparation.

---

35. What are the file extensions used in Tableau?

=>

- .twb – Tableau Workbook (XML format)
- .twbx – Packaged Workbook (includes data and images)

---

36. Is it possible to use different colors depending on conditions in a line graph in Tableau? Mention steps to send graphs to PowerPoint.

=> Yes, it's possible using calculated fields and color marks.
Steps:

1. Create a calculated field based on your condition (e.g., IF [Sales] > 1000 THEN "High" ELSE "Low").
2. Drag the calculated field to the "Color" mark in the line chart.
3. The line will now display segments in different colors.
   Export to PowerPoint:

- Go to File > Export as Image or Download > PowerPoint.
- Tableau exports the current sheet or dashboard as a slide.

37. What data types does Tableau support?

=>

- String
- Number (Integer, Decimal)
- Boolean
- Date and DateTime
- Geographic values (like Country, City)

38. What are Measures and Dimensions?

=>

- Dimensions are qualitative fields used to segment data (e.g., Category, Region).
- Measures are quantitative fields that can be aggregated (e.g., Sales, Profit).

39. What are the different types of joins in Tableau?

=>

- Inner Join – Only matching records from both tables.
- Left Join – All records from the left table, and matches from the right.

- Right Join – All records from the right table, and matches from the left.
- Outer Join – All records from both tables, matched where possible.

---

40. What is disaggregation and aggregation of data?

=>

- Aggregation means summarizing data (e.g., sum of sales).
- Disaggregation shows individual data points without summarizing, useful for spotting granular patterns.

41. What happens when you join two tables in Tableau and values do not match?

⇨ When you join two tables in Tableau and values do not match, the result depends on the join type. Unmatched values may lead to missing rows (in inner joins) or NULL values (in left, right, or outer joins).

42. Define Tableau. What are the different data types in Tableau?

=> **Tableau** is a powerful data visualization tool used for converting raw data into interactive and easy-to-understand dashboards and visual reports. It helps users analyze data quickly and share insights through visualizations without the need for complex coding.

The different data types in Tableau are:

1. **String** – Text values (e.g., "India", "Product A").
2. **Number** – Numeric values (e.g., 100, 45.6).
3. **Boolean** – Logical values (TRUE or FALSE).
4. **Date/DateTime** – Calendar dates and timestamps (e.g., 2024-08-15, 2024-08-15 10:30 AM).
5. **Geographical** – Data representing locations (e.g., Country, State, City).
6. **Cluster Group** – Automatically generated fields from clustering data.

43. Compare among Worksheet, Dashboard and Story with respect to tableau.

=> 1. Worksheet

- A single view that contains **visualizations** like charts or graphs.
- Users can **drag and drop fields** to create visuals from data.
- It is the **basic building block** in Tableau.

2. Dashboard

- A collection of **multiple worksheets** and other elements (e.g., filters, images).
- Used to create an **interactive view** of related data.
- Allows users to **combine multiple visualizations** in one interface.

3. Story

- A sequence of **worksheets and dashboards** presented in a **narrative format**.
- Used to **convey a data-driven story** or analysis step-by-step.
- Helps guide the audience through insights and decision-making.

44. Mention few points in favour of using Tableau software. The Tableau Product Suite consists of several parts. Mention all of them.

=> Points in Favour of Using Tableau Software:

1. **User-Friendly Interface** – Drag-and-drop functionality makes it easy to use without programming knowledge.
2. **Powerful Data Visualization** – Creates interactive and attractive charts, graphs, and dashboards.
3. **Connects to Multiple Data Sources** – Supports connections to Excel, SQL, Google Sheets, and cloud platforms.
4. **Real-Time Data Analysis** – Enables real-time data updates and dynamic filtering.
5. **Fast Performance** – Handles large datasets efficiently with quick rendering.

Tableau Product Suite Includes:

1. **Tableau Desktop** – Used for creating reports, visualizations, and dashboards.
2. **Tableau Public** – Free version of Tableau Desktop with public data sharing.
3. **Tableau Server** – Allows sharing and collaboration on dashboards within an organization.
4. **Tableau Online** – Cloud-based version of Tableau Server.
5. **Tableau Reader** – Free desktop app to view Tableau visualizations shared by others.
6. **Tableau Prep** – Tool for data cleaning and preparation before visualization.

45. Differentiate between Dual-axis and blended axis in Tableau?

=>

Feature	Dual-Axis	Blended Axis
Definition	Displays **two different axes** on the same graph.	Uses **one shared axis** for multiple measures.
Usage	Used when measures have **different scales**.	Used when measures have the **same scale/unit**.
Visualization	Creates a **layered chart** (e.g., bar + line).	Creates **grouped bars/lines** using a single axis.
Axes	Each measure has its **own separate axis**.	Measures share the **same axis**.
Example	Sales (in ₹) and Profit % on different axes.	Sales and Quantity both shown using the same scale.

46. Who to create a calculated field in tableau ?

=>

1. **Open your Tableau worksheet.**
2. In the **Data pane**, select on any attribute you want to calculate
3. Then right click on it then click on create and select **"Create Calculated Field..."**.
4. A dialog box will appear — give your calculated field a **name**.
5. Enter the **calculation formula** using Tableau's built-in functions and fields
6. Click **OK** to save the field.

48. What are the different types of joins in Tableau?

=> 1. **Inner Join**

- Returns only the **matching rows** from both tables.
- Rows without matches are **excluded**.

2. **Left Join**
   - Returns **all rows from the left table**, and matching rows from the right table.
   - If no match, values from the right table are **NULL**.
3. **Right Join**
   - Returns **all rows from the right table**, and matching rows from the left table.
   - If no match, values from the left table are **NULL**.
4. **Full Outer Join**
   - Returns **all rows from both tables**.
   - If there's no match on either side, unmatched fields are filled with **NULLs**.

49. Design a view in a map such that if a user selects any state the cities under that state have to show profit and sales using Tableau.

=> Creating an Interactive Map with State and City-Level Data in Tableau

To design a view in Tableau where **users can select a state** and view **Profit and Sales** for cities in that state, follow these detailed steps:

1. Prepare Your Data

Ensure that your data includes:

- **States**: A field for geographical locations like states (e.g., "Maharashtra", "California").
- **Cities**: A field for cities that belong to each state (e.g., "Mumbai", "Pune").
- **Sales and Profit**: Numeric fields for Sales and Profit related to each city.

Example data format:

State	City	Sales	Profit
Maharashtra	Mumbai	100000	30000
Maharashtra	Pune	80000	20000
California	Los Angeles	120000	40000
California	San Francisco	90000	25000

2. Create the Map Visualization for States

- **Step 1**: Drag the **State** field to the **Rows** shelf.
- **Step 2**: In Tableau, if your "State" field is recognized as geographical, it will automatically create a **Map**.
- **Step 3**: Drag **Sales** and **Profit** to the **Size** and **Color** marks to represent sales and profit visually on the map.

- **Step 4**: Customize the map by adjusting colors, map style, and adding tooltips to display **Sales** and **Profit** on hover.

3. Create the City-Level View

- **Step 1**: Create a new worksheet for **City-Level** data.
- **Step 2**: Drag the **City** field to the **Rows** shelf.
- **Step 3**: Add **Sales** and **Profit** as **text** or use **bar charts** to show the values for each city.
- **Step 4**: Customize the city-level visualization by sorting the cities based on Sales or Profit, and applying color gradients for better visual appeal.

4. Build a Dashboard

- **Step 1**: Now, create a **Dashboard** by selecting **Dashboard** > **New Dashboard**.
- **Step 2**: Drag both your **State Map worksheet** and the **City-level worksheet** onto the dashboard.
- **Step 3**: Ensure both are properly sized and positioned so the user can clearly see both.

5. Add Filter Action for Interactivity

This is where Tableau's interactivity shines. When you click a state on the map, Tableau should automatically update the city-level view to show only cities in that state.

- **Step 1**: Go to **Dashboard** > **Actions** > **Add Action** > **Filter**.
- **Step 2**: Choose the **Source Sheet** as your **State Map** worksheet and the **Target Sheet** as your **City-Level worksheet**.
- **Step 3**: Set the action to **filter** the **City-level worksheet** based on the **State** selected on the map. Choose **Select** for the user action.
- **Step 4**: Under the **Target Filters**, ensure that the **State** field is used to filter the **City** field.

6. Customize Tooltips for Better User Experience

- **Step 1**: In both your **State map** and **City view**, edit the **Tooltips** to show additional information when hovering over data points.
- **Step 2**: For the map, you can display details like **State Name**, **Sales**, and **Profit**. For the city worksheet, show **City Name**, **Sales**, **Profit**, and even percentage change.

7. Test the Interactivity

- After applying the action filter, go to the **Dashboard** view.
- **Select a state** on the map (e.g., Maharashtra).
- The **City-level worksheet** should update and show **cities** like Mumbai and Pune along with their **Sales and Profit**.

50. Mention the significance of using calculated fields in Tableau. Also explain the types of calculations that can be done through these fields.

=> Significance of Using Calculated Fields in Tableau (4 Marks)

Calculated fields in Tableau are powerful because they allow you to:

1. **Customize Metrics**: Create custom calculations (e.g., profit margin, growth percentage) not available in raw data.
2. **Transform Data**: Perform data manipulation, such as string operations, extracting date parts, or rounding numbers.
3. **Advanced Aggregation**: Aggregate data in ways beyond standard functions (e.g., dynamic summations or averages).
4. **Enhance Analysis**: Provide flexibility for advanced calculations like logical conditions and mathematical formulas.

Types of Calculations in Calculated Fields (4 Marks)

1. **Basic Calculations**: Simple mathematical operations.
   - Example: `([Sales] + [Tax])`
2. **Conditional Calculations**: Use **IF-THEN** logic for categorization.
   - Example: `IF [Profit] > 0 THEN 'Profit' ELSE 'Loss' END`
3. **String Calculations**: Manipulate text fields (concatenation, extraction).
   - Example: `CONCAT([First Name], ' ', [Last Name])`
4. **Date Calculations**: Manipulate and calculate with dates.
   - Example: `DATEDIFF('day', [Order Date], TODAY())`
5. **Table Calculations**: Perform calculations over entire datasets (e.g., running totals).
   - Example: `RUNNING_SUM([Sales])`
6. **Aggregation Calculations**: Use custom aggregations like moving averages.
   - Example: `WINDOW_AVG([Sales], -2, 2)`

51. Mention the significance of creating parameters in Tableau.

=> **User Interaction**: Allows users to change values (e.g., choose a date range or metric) to update the view.

Customization: Helps create custom calculations and filters based on user input.

Flexibility: Makes dashboards more flexible by allowing users to control what data they see or how it's displayed.

52. Explain LOD in details with respect to Tableau software?

=> Level of Detail (LOD) in Tableau helps control how calculations are made, even if not all data is shown in the view. There are three types of LOD expressions:

1. **FIXED**:
   - Calculates values for specific data, no matter what you see in the view.
   - Example: {FIXED [Region]: SUM([Sales])} calculates total sales for each region.
2. **INCLUDE**:
   - Adds extra details to your calculations, even if those details aren't shown.
   - Example: {INCLUDE [Product]: AVG([Profit])} calculates the average profit per product, even if you only see regions.
3. **EXCLUDE**:
   - Removes certain details from the calculation.
   - Example: {EXCLUDE [Category]: SUM([Sales])} calculates total sales without considering the category.

53. What is the difference between a worksheet and a dashboard in Tableau?

=> **Worksheet**:

- A worksheet is a single view or chart where data is displayed, such as a bar chart, line graph, or table.
- It is the basic building block in Tableau where you can create individual visualizations using a single data source.

 Dashboard:

- A dashboard is a collection of multiple worksheets combined into one view.
- It allows you to display different types of visualizations, filters, and interactive elements together to give a broader analysis of the data.

54 is Same as 50