

Tata Data Visualization Virtual Experience

☰ Subject Area	CS Data Science
☰ Source/Uni Name	Forage
☰ Status	started
⌚ Website1	
⌄ Level	University
☰ Main Topic Areas Included	Data Visualization
☰ Further sources/books	
⌄ University Course (For GPA)	No
⌄ Minimum Grade Goal	
⌄ Grade Achieved	

If you would like to connect with the Tata team to find out more information please email microinternships@tata.com

- You will gain insight into how our passionate multi-disciplinary experts solve some of the most **complex business problems** using **data visualisation solutions**, such as **Tableau** and **Power BI** and create amazing stories which are hidden under the mountains of data generated.
- Sophisticated and predictive analytics, as well as ML algos, to create advanced analytics and Data Engineering Solutions, allowing the group to operate
- AI + ML
- One of the key functions is visualization practice → data visualisation stories that will empower businesses to take decisions backed by facts

- In this virtual job simulation, you will act as a **Data Visualization Analyst** at Tata iQ and address the needs of various stakeholders
- Drive real world impact by empowering businesses
- Telling stories through Data

▼ Task 1 - Preparing Questions for CEO and CMO Meeting

- We'll be focusing on the **anticipation of questions that you may be asked by your business leaders**
- Proactive thinking → the ability to frame the data and your visual mindset → important when preparing for presentations
- Being able to anticipate the types of questions and needs your leaders will have helps facilitate in depth discussions about the business objectives
- Hypothetical scenario → expected to draft questions that are important to the CEO and CMO of the business → **Allows you to be better prepared to present impactful data analysis in your meeting**
- Data Set in GitHub

Article on how to think from the perspective of a CEO

▼ Summary of Article

What is a CEO Dashboard?

- A CEO Dashboard is a **management reporting tool used to track, analyze, and visualize data.**
- It empowers CEOs and other c-level executives to make **data-driven decisions with the help of interactive, high-level metrics from sales, marketing, finance, and other areas.**
- CEOs need a **birds-eye overview of the most prominent metrics that will help them make the best possible decision in order to improve outcomes.**

- Data-driven dashboards offer business leaders a level of insight that maximizes the value of the top-level data available at their fingertips, enhancing strategic decision-making and mentorship across the board

What is a CEO Report?

- A CEO report is a critical analytical tool that serves up historical trends, and real-time insights in a way that's visual, digestible, and easy to navigate.
- These reports can be delivered in the form of an interactive dashboard or .pdf file.
- It's possible to tailor CEO dashboards to your specific needs, goals, and requirements, working with CEO KPIs that will help you uncover priceless trends and insights with ease.
- In essence, it offers benefits like improved decision-making, wider accessibility to important data, mobility through 24/7 mobile dashboard access, and improved internal communication that will make the organization stronger

The Key Benefits of CEO Reporting & Dashboards

1. **Confidence, validation & trust:** A modern CEO BI dashboard will give you the confidence to take charge of every core aspect of the business while validating your choices with actionable insights. This builds greater trust and transparency, boosting engagement and productivity in the process.
2. **Decision-making:** The accessible visual nature of CEO dashboards helps you make informed, accurate, and progressive decisions under pressure.
3. **Extensive data access:** Investing in innovative business intelligence dashboard software creates a data-powered culture throughout the organization. Any modern CEO dashboard worth its salt is customizable, allowing you to work with the key performance indicators that align with your specific goals and objectives.
4. **Flexibility & adaptability:** CEO dashboard reporting software allows you to curate, analyze, and access a raft of invaluable data from one central location. A solid modern CEO report generator is optimized for a range of devices, including mobile and desktop, allowing you to make razor-sharp decisions 24/7 wherever you are in the world.

5. **Panoramic vision & BI:** By working with historical, predictive, and real-time analytics, you can leverage strengths, identify weaknesses, spot emerging trends, and uncover insights you never thought possible. This level of BI reporting filters down throughout the entire organization, making it efficient, effective, and innovative.

Features to Look for in CEO Dashboard Software

- **Interactivity:** Look for a dashboard that allows you to **interact with the data, drilling down into the details when needed.**
- **Customizability:** The dashboard should be customizable to your specific needs, with the ability to set up **specific KPIs and metrics that matter most to your business.**
- **Real-time data:** The best dashboards offer **real-time data, allowing you to make decisions based on the most current information.**
- **Access Control:** The software should provide you with the ability to control who has access to the dashboard and what they can see or do with the information.
- **Integration Capabilities:** Look for a dashboard that can integrate with other tools and platforms you're already using in your business.
- **Visual Appeal:** A good dashboard is visually appealing and easy to understand, making it easier to digest the information and make data-driven decisions.

Article on Metrics you should be tracking

▼ Background Information on Task 1

- You are hired as a **consultant for an online retail store to review their data and provide valuable insights.**
- The business is performing well and the management wants to understand the **key factors contributing to the revenue for strategic planning.**
- Leadership is interested in **metrics from both operations and marketing perspectives.**

- As part of their expansion plans, they are **seeking guidance on areas that are performing well to maintain focus on successful strategies.**
- Management **wants to understand metrics related to demographic data available.**
- You have a scheduled meeting with the CEO and CMO next month where you will present analytics and insights to help evaluate current business performance and suggest metrics for expansion decisions.
- You are encouraged to **think from the business leaders' perspective to analyze data effectively and deliver better insights.**
- Additional resources are provided to help understand how business leaders think and approach business performance.

▼ Task 1 breakdown and Milestones

Task 1: Prepare Questions for CEO and CMO Meeting

Timeline:

- **Day 1, Hours 0-0.5:** Understand Task and Background Information
 - Review the task details and comprehend the business context
 - Understand the different perspectives of the CEO and CMO
 - Review linked resources to understand how business leaders approach performance evaluation
- **Day 1, Hours 0.5-1:** Examine Provided Dataset
 - Review the dataset and take note of the available information
 - Begin identifying potential insights relevant to the CEO and CMO
- **Day 1, Hours 1-1.5:** Draft Preliminary Questions
 - Draft potential questions that the CEO and CMO might ask or want to know answers to
 - Ensure questions are both qualitative and quantitative in nature
- **Day 1, Hours 1.5-2:** Review and Refine Questions

- Review the preliminary questions, ensure they're aligned with the roles of the CEO and CMO
- Refine questions for clarity, relevance, and depth
- **Day 1, Hours 2-2.5:** Finalize and Submit Questions
 - Finalize four questions for each business leader
 - Submit your eight questions in the text submission box

▼ What is the difference between questions that are qualitative and quantitative in nature?

▼ Quantitative Questions

- Aim to collect numerical data or data that can be transformed into usable statistics.
- Often address "**how much?", "how many?", "how often?"**.
- Measure variables and the relationship between them.
- Allow for broad overviews and statistical analysis.
- Example: "What percentage of the population has a university degree?"

▼ Qualitative Questions

- Gather non-numerical data.
- Typically answer "**why**" and "**how**" questions about human behavior.
- Explore topics in depth and detail, leading to **detailed descriptions, anecdotes, and personal narratives**.
- Provide a deeper, more detailed understanding.
- Example: "How do people **feel** about their educational experiences?"
- Love is Qualitative

▼ CEO Questions

- What are the top-selling products, and what percentage do they contribute to our overall revenue? Are there products with really low sales volumes that we should consider discontinuing?

- How are sales distributed across multiple countries and where is the most potential for expansion?
- Which customers (from CustomerID and Country) contributes most to our sales, volume and revenue?
- What are the patterns in sales volume over time? Are there any seasonal trends?

▼ CMO Questions - Chief Marketing Officer

- How are different products (from Description) performing in different markets (countries)?
- What times of day (from the InvoiceDate) do we see the highest sales volume, and can this inform our marketing and promotional efforts?
- Are there any specific products (from StockCode) that generate higher sales during certain times and how can we leverage this information for marketing campaigns?
- Do some groups of customers based on their demographic data (from Country and CustomerID) tend to buy certain products more than others, and how can we better market to these groups?

▼ Sample Answer

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/914937da-c0be-4c08-8c0b-fa8565578944/Model_Answer_-_Questions.pdf

▼ Task 2 - Choosing the Right Visuals

- The last task helped form the basis of your analysis
- Now, it's time to choose the right visual for a given scenario
- Which visual will be the most impactful to convey the story
- Choose the right visual and chart for the data that you want to present

- In this task, you will come across different business scenarios where you will need to decide which visual would be best suited for the businesses requirement
- Understand the requirements!

▼ Background information on Task

- You have been tasked by the CEO and CMO of an online retail store to provide visualizations of key business metrics.
- The senior management is interested in understanding the business's performance and identifying the company's key strengths.
- The management is also keen on spotting opportunities that could lead to business growth and increased future revenue.
- You will be given multiple visualization requests by the CEO.
- Your responsibility is to create visuals that present data and insights in the simplest and most comprehensible way.
- It is crucial that the visuals accurately represent the information you are trying to convey, as incorrect representation can lead to misinterpretation or flawed decisions by the management.
- Therefore, you need to ensure the correct visuals are used for each data set.

▼ Task 2 breakdown and milestones

1. Understand Business Requirements:

- Read the questions carefully (in the quiz) and understand what the CEO and CMO are looking for in each scenario.

2. Identify Suitable Visuals:

- Based on the business requirements, identify the perfect visuals that would best display the information being presented.
- Remember, the same data can be represented in multiple types of charts. Your task is to choose the most effective one for each scenario.

3. Use Provided Resources:

- Refer to the resources provided in the resources section to understand how to select visuals based on different scenarios.
- These resources will guide you in choosing the right chart or graph for your data.

4. Answer Multiple Choice Quiz:

- Answer the multiple-choice quiz based on your understanding and visual selection. There are 5 questions to complete in this task.

▼ Grammar of Graphics to understand the dimensions of visuals

A Comprehensive Guide to the Grammar of Graphics for Effective Visualization of Multi-dimensional Data - Towards Data Science

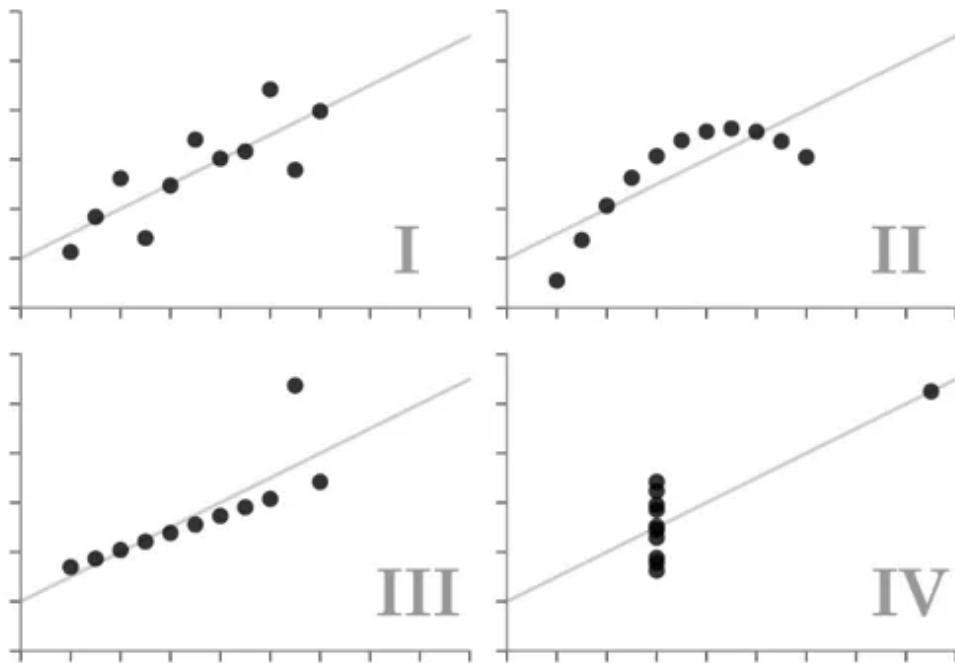
▼ Datasaurus Dozen

- A set of twelve datasets that were created to demonstrate the importance of visualizing data rather than relying solely on summary statistics.
- All twelve datasets have the same summary statistics (mean, standard deviation, and correlation), which would lead you to believe they are very similar.
- When visualized, each dataset is distinct and tells a very different story.
- Moral → always visualize your data, because relying solely on summary statistics can be misleading and miss the underlying patterns or anomalies that are present in the data.

▼ Anscombe's quartet is really similar

Anscombe's Quartet

Each dataset has the same summary statistics (mean, standard deviation, correlation), and the datasets are *clearly different, and visually distinct.*



- Always **visualize and understand your data attributes before moving on to feature engineering and building statistical, machine learning and deep learning models.**

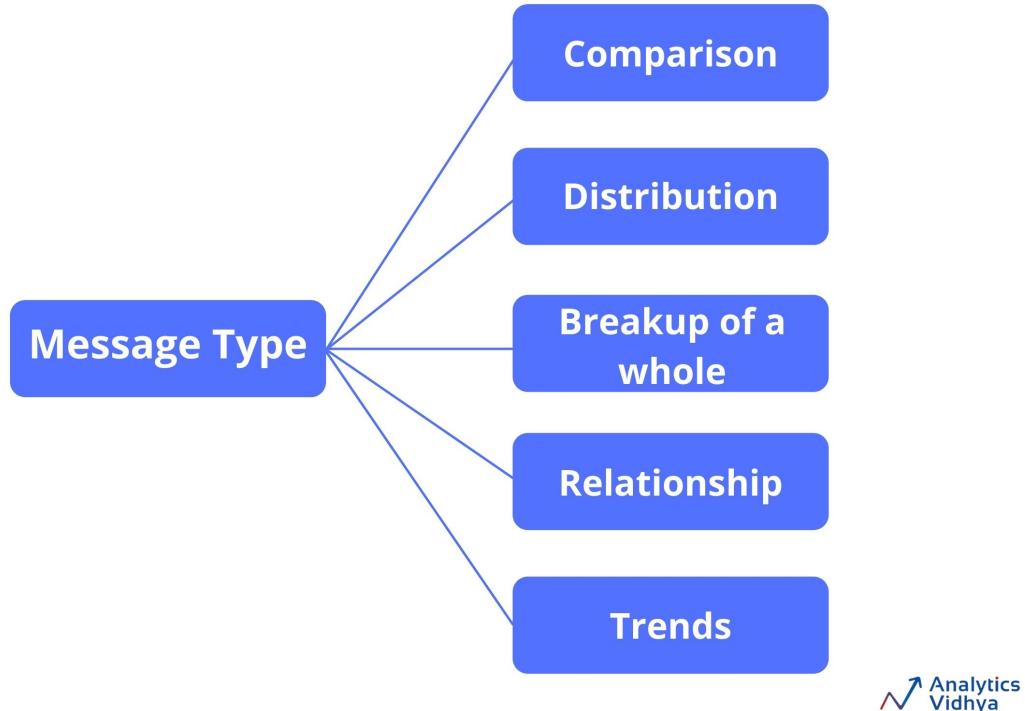
▼ How do you choose the right chart for data visualization?

- Incorrect representation can lead to a **wrong message or wrong decision** taken by the audience or whatever you've in your mind when you were creating that chart, **that message might not be conveyed to the audience.**

Some benefits of visualizing data:

1. It **helps us to convey the right message to the audience through visuals.**
2. It **helps us find outliers in our data.**
3. It **helps the business leader to take an accurate decision.**
4. It **helps us to understand how the data is distributed over time.**

Type of messages:

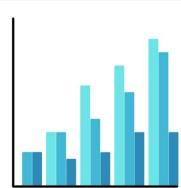


▼ Comparison Chart

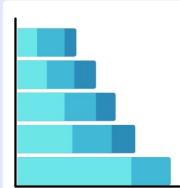
In this chart, we compare one value with the other like region-wise sales, economy rate comparison of bowler in cricket.

Comparison

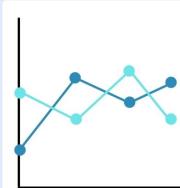
- To compare one data value with others.
- Use the following charts to represent it.



Column Chart



Bar Chart



Line Chart



Scatter Chart

- Example:
 - Sales comparison over the different regions.
 - Economy rate comparison of bowlers.

• Column Charts

- Used to compare values across multiple categories.
- Categories are on the X-axis, values on the Y-axis.
- Can show parts of a whole across categories, both in absolute and relative terms.
- Variations include stacked column charts and 100% stacked column charts.

• Bar Charts

- Similar to column charts, but values are represented on the X-axis and categories on the Y-axis.
- Typically used when category text is long or when comparing values across categories.
- Stacked bar charts can compare parts of a whole, and change over categories or time.

• Line Charts

- Popular chart type used to show trends over time or across categories.
- Categories are on the X-axis, values on the Y-axis.
- Helpful for analyzing data like sales, year-on-year profit, salary increases over time.

- **Scatter Plots**

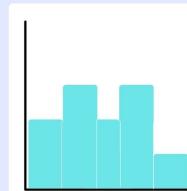
- Use numerical values along both axes.
- Useful for showing correlation between data points.
- Typically used for displaying and comparing numerical values, such as scientific or statistical data.

▼ **Distribution Chart**

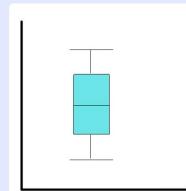
- Show the spread of the data values over categories or continuous values.
- For example Distribution of bugs found in 10 weeks of the software testing phase.

Distribution

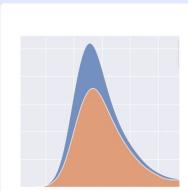
- To show the spread of data values over categorical or continuous values.
- Use the following charts to represent it.



Histogram



Boxplot

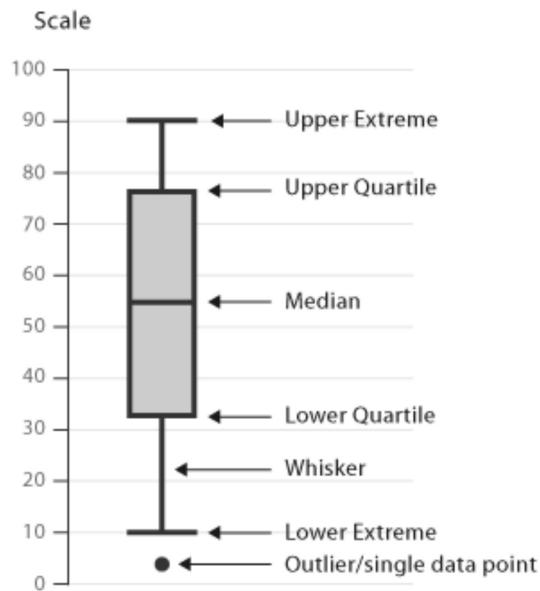


KDE plot

- Example:

- Distributions of bugs found in 10 weeks of software testing phase.

- **Histogram**
 - Used to graph the frequency over a distribution.
 - A key tool in analytics for deriving insights from data.
 - Visually, all bars are touching each other with no space in between.
- **Box Plot (also known as Box and Whiskers Plot)**
 - The line in the middle of the box represents the median value, meaning 50% of data is above it and 50% below it.
 - Median is resistant to outliers, unlike the mean.
 - The box contains 50% of the data, with 25% above the median and 25% below.
 - Useful for spotting outliers and understanding the distribution of the data.



- **KDE Plot (Kernel Density Estimation Plot)**
 - Represents a smooth version of a histogram.

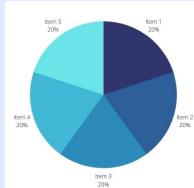
- Used for visualizing the distribution of observations in a dataset, similar to a histogram.
- Compared to histograms, KDE plots can be less cluttered and more interpretable, particularly when representing multiple distributions.

▼ The breakup of a whole chart

- Used to analyze, how various parts comprise the whole.

Break up of a whole

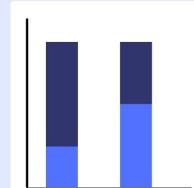
- To analyze, how various parts comprise the whole.**
- Use the following charts to represent it.**



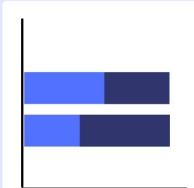
Pie chart



Donut Chart



Stacked Column chart



Stacked Bar chart

• Example:

- Revenue contribution by different regions.
- The batsman scored on which sides of the ground.

• Pie Chart

- Represents categorical data as parts of a whole.
- Each slice signifies the percentage of the category out of the total.
- Recommended when dealing with less than 5 categories.

• Donut Chart

- A variant of a pie chart with a hole in the center.
- Displays categories as arcs instead of slices.

- **Stacked Column Chart**

- Used to show relative percentages of multiple data series in stacked columns, cumulatively adding up to 100%.
- Can depict part-to-whole proportions over time, such as quarterly sales per region or monthly mortgage payments divided into interest vs. principal.

- **Stacked Bar Chart**

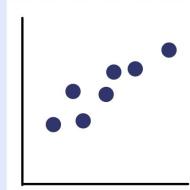
- Used to represent the relative percentage of multiple data series in a stacked bar.

▼ Relationship Charts

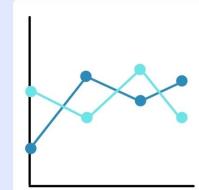
Very helpful when we want to know that what is the relation between the different variables

Relationship

- To show the relationship between two variables.
- Use the following charts to represent it.



Scatter Plot



Line chart

- Example:

- Relationship between strategic breaks and batsman dismissal.

Scatter Plot

- Utilizes numerical values on both axes to represent data points.
- Uses dots to denote the values for two different numerical variables.

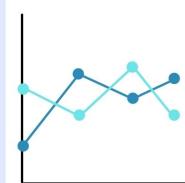
- The position of each dot on the horizontal and vertical axes indicates the value of a specific data point.
- Useful in identifying correlations between data points that might not be apparent from the data itself.
- Often used to display and compare numerical values, such as scientific or statistical data.

▼ Trend Charts

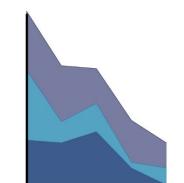
- Visualize trends of values over time and categories
- Also known as “Time Series” data in the data-driven world.

Trends

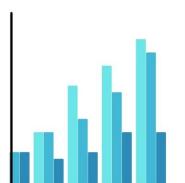
- To show trends over time, it is also termed as "Time Series Data".
- Use the following charts to represent it:



Line chart



Area chart



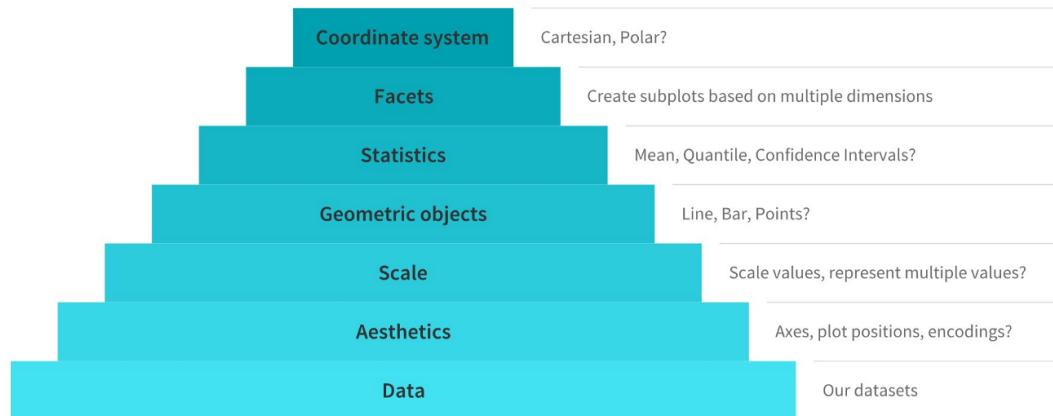
Column chart

- Example:
 - Hourly temperature variation during a day.
 - Run rate tracker over by over.

- **Line Chart**
 - Ideal for visualizing trend data.
 - Commonly used to identify trends in various domains.
- **Area Chart**
 - Used to observe the magnitude of values.

- Shows the relative significance of values over time.
 - Similar to a line chart but emphasizes the magnitude of values more due to the filled-in area between lines.
- **Column Chart**
- Used to depict trends of values over time and across categories.

Major Components of the Grammar of Graphics



- The "grammar of graphics" is a **systematic approach and layered framework to describing and creating visualizations**
 - Provides a structured method to **describe and create a wide variety of statistical graphics**
 - **Foundational principle behind many modern data visualization software and libraries**, such as R's ggplot2
1. **Data:** This is the raw material of any graphic. It is usually in a structured format like a data frame with rows and columns
 2. **Aesthetics:** Aesthetics describe how **variables in the data map to visual properties of the graphic** such as size, shape, color, position, etc → Determine axes based on data dimensions, positions of data points in the plot, and encoding (size, shape, color) for plotting multiple data dimensions.

- How you use features like color, size, shape, and position to represent data in your visualization.
3. **Scales:** Scales map the values in the data to visual properties such as the range of colors, the size range, the position along an axis, etc. → Determine if we need to **scale the potential values or use a specific scale to represent multiple values or a range.**
 4. **Geometries:** These are the visual elements used to represent the data. They could be points, lines, bars, or other shapes.
 5. **Statistics:** Decide whether to show statistical measures in the visualization → These are **statistical operations that summarise or transform the raw data.** Examples include linear regression, histogramming, etc.
 6. **Facets:** Facets are used to create **multiple similar plots based on a factor or combination of factors in the data** → Determine whether subplots based on specific data dimensions are needed
 7. **Coordinates:** The **coordinate system** determines how data coordinates are mapped to the plane of the graphic. The most common is Cartesian coordinates, but there are others like polar coordinates, map projections, etc.

▼ Additional Resources

Which Chart or Graph? A guide to data visualization

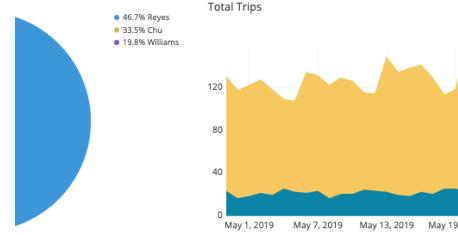
Which chart or graph should you use to communicate your data? This whitepaper explores the best ways for determining how to visualize your data to communicate information.

 <https://www.tableau.com/learn/whitepapers/which-chart-or-graph-is-right-for-you>

How to Choose the Right Data Visualization

There are many ways that charts can be used to visualize data. Read this article to learn which charts can be used for each kind of visualization task.

 <https://chartio.com/learn/charts/how-to-choose-data-visualization/>



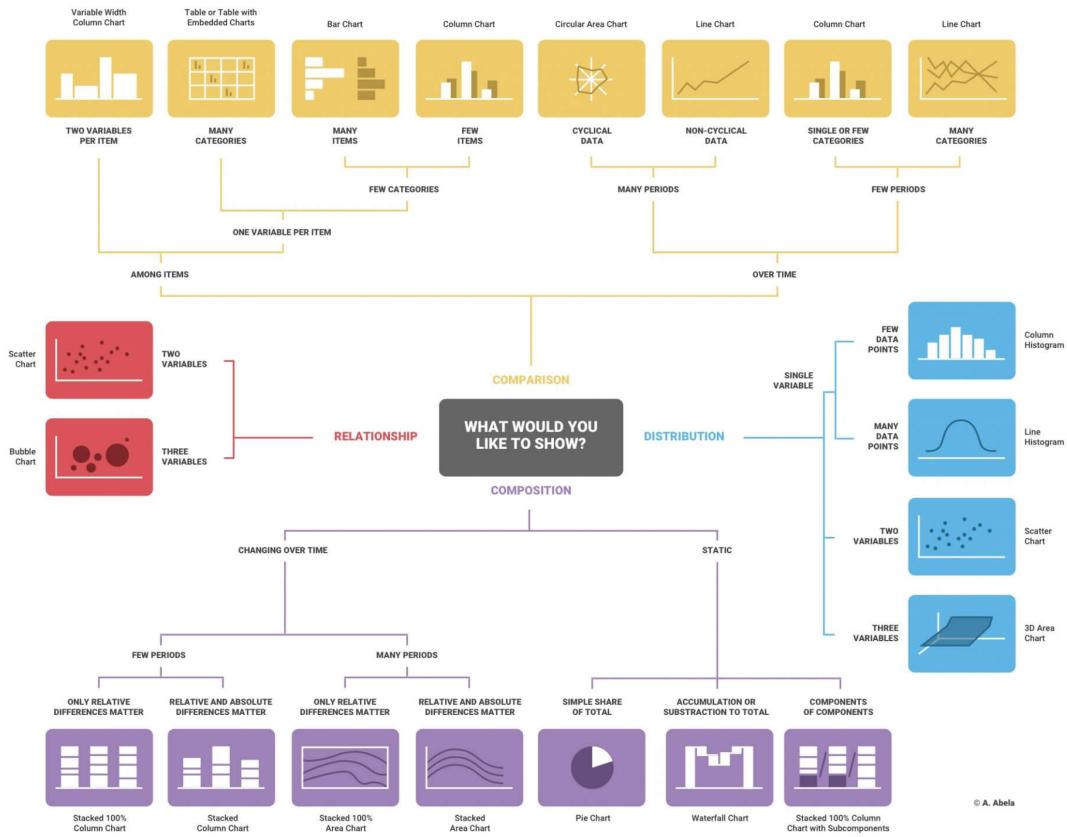
<https://towardsdatascience.com/ultimate-guide-to-choosing-the-right-visual-2a77aa8eec08>

How to Choose the Right Chart for Your Data

 <https://infogram.com/page/choose-the-right-chart-data-visualization>



CHART SUGGESTIONS - A THOUGHT-STARTER



▼ Task 3 - Creating Effective Visuals

- Now it's time to create your own visualizations for the CEO and CMO
- Tableau or Power BI to create these visuals
- **Analyse the requirements provided by CEO and CMO and create visuals that best answer the insights they are looking for**

▼ Background Information

- The **CEO and CMO want to use data to make more informed decisions.**
- They **need analysis and visuals that will help answer their questions and provide useful insights for an expansion strategy.**
- They're particularly interested in understanding trends and breakdowns of different categories to clarify revenue sources and major factors impacting the online store.
- The task involves gathering executive requirements, ensuring data quality (and cleaning if necessary), then creating visuals using either Tableau or Power BI.
- The choice of tool for creating the visuals does not impact the outcome, which is meant to assist the executives in effective decision-making and strategizing for expansion.

▼ To Do and Milestones

Todo List:

1. Setup and Preparation:

- Download and install either **Tableau** or Power BI for creating the visuals.
- Review and familiarize yourself with the data provided.

2. Data cleanup:

- Identify and exclude entries with a quantity below 1 unit. (Quantity Column)
- Identify and exclude entries with a unit price below \$0. (Unit Price Column)

- Use **conditional formulas or data transformation methods to clean the data.**

3. Visual Creation:

- **Question 1:**
 - Create a time series visual of the 2011 revenue data showing each month.
- **Question 2:**
 - Create a visual showing the top 10 countries generating the highest revenue, excluding the United Kingdom.
 - Include the quantity sold in the visual, along with the revenue generated
- **Question 3:**
 - Create a visual showing the top 10 revenue-generating customers, from highest to lowest.
- **Question 4:**
 - Create a **visual showing the demand for products across all countries**, excluding the United Kingdom.

4. Save and Upload Files:

- **Save each visual on a separate tab, naming each tab with the question number.**
- Save Tableau files in .twbx format, and Power BI files in .pbix format.
- Upload the saved files.

Milestones:

1. **Milestone 1:** Successful setup and data review.
2. **Milestone 2:** Completion of data cleanup and confirmation of its quality for analysis.
3. **Milestone 3:** Completion of visuals for all four questions.

4. **Milestone 4:** Successful saving and uploading of the visual files in the appropriate formats.

▼ Navigating Tableau Interface

- Tableau offers a 14-day free trial

Step 1:

- Install Tableau and register

Step 2: Connect to Data

- Under the **Connect pane**, you will see options to connect to **different types of data**. This could be a file (like an Excel or CSV file) or a server (like SQL Server, Oracle, etc). Click on the type of data source you are connecting to, then navigate to your file or server.

Step 3: Data Source Page

- Once your data is connected, you will see the **Data Source page**. Here, you can set up your data how you want it to appear in Tableau. This could include joining multiple tables, setting data types, renaming fields, creating calculated fields, etc.

Step 4: Go to worksheet

- At the **bottom left of the Tableau window**, you will see two tabs: "**Data Source**" and "**Sheet 1**". **Click on "Sheet 1" to start building your visualizations.**

Step 5: Drag and Drop Fields to Create a Visualization

- Tableau is a drag-and-drop tool, meaning you **create visualizations by dragging fields onto "shelves"**. The **main shelves** are **Columns, Rows, and Marks**. By dragging a field to Columns or Rows, you create an axis. By dragging a field to Marks, you add detail to the marks in the view (marks are the dots, lines, bars, etc in the view).

Step 6: Use Show Me to Change Visualization Type

- If you want to change the type of visualization, you can use the **Show Me** tool at the top right of the Tableau window. Show Me suggests different visualization types based on the fields you have selected.

▼ Data Cleanup using Tableau

- Data source filters, data roles, and calculated fields.
- **Data Source** Tab → here you will find the **Data** pane which contains all the columns → click drop-down arrow next to the column name, and select "Create Calculated Field..."
- **IF [Quantity] >= 1 THEN [Quantity] END** → will return the quantity if it is greater than or equal to 1 and null otherwise
- However, if do the above steps, it creates a new column next to the Quantity column

This is how to filter the data without creating an extra column

1. Navigate to the data source page.
2. In the Filters section in the upper-right corner of the page, click Add.
3. This will open the Edit Data Source Filter dialog box, listing any existing data source filters.
4. Click Add to open an Add Filter dialog box listing all fields in the data source.
5. Select a field to filter, then specify how the field should be filtered

You should also apply the filter in the worksheet for future visualizations

1. Locate the "**Data Pane**" in the workbook and drag the column to the **filters** shelf.
2. A dialog box will appear, allowing you flexibility on how to **aggregate** the fields → select 'at least' option an specify the options

▼ Question 1

▼ Creating line charts in Tableau

- Drag `Invoice Date` field from dimensions pane to the **Columns shelf** at the top of the screen → by default, **Tableau will aggregate this field by year.**
 - Columns shelf → defining **x-axis** of line chart as time
- You will need to create a calculated field for the `revenue`
 - right-click in the data pane on the left side and choose "Create Calculated Field" → Name field `Revenue` and then do **quantity * unit_price**
 - Drag revenue to **rows shelf**, which defines **y-axis** of line chart
- Default aggregation of `Invoice Date` is year so we have to set the **date granularity to monthly**
 - click the plus sign (+) next to the Year field on the Columns shelf → quarters → click plus sign again to break down the quarters into individual months.

▼ What is **granularity**?

- level of detail or depth of the data in a set.
- For example, if a dataset has a daily granularity, it means there is one data point for each day. If it has a monthly granularity, it means there is one data point for each month. The higher the granularity (i.e., the more frequent the data points), the more detailed the data.
- We also need to **filter the data to only 2011**
 - Drag the Invoice Date field from the Dimensions pane to the Filters shelf.
 - Choose "Years" (from dialog box) and then check the box next to 2011 and click "OK".
- We could also add a **forecast** to the visualization
 - "Analytics" pane → In the "Model" section, you'll see a "Forecast" option → Drag this option into the visualization area
- View granular data by looking into revenue for each month

- Create a line chart
 - Invoice Data - x axis
 - revenue - y axis
- $\text{revenue} = \text{quantity} * \text{unit_price}$
- Once the revenue field is calculated, you can use it to view the trend of revenue.
- Set the granularity of the data to each month and **only provide data for the year 2011.**

▼ Question 2

The CMO is interested in viewing the top 10 countries which are generating the highest revenue. Additionally, the CMO is also interested in viewing the **quantity sold along with the revenue generated.** The CMO does not want to have the United Kingdom in this visual.

▼ Building Side by Side Chart in Tableau

- There would be **two bars for each country** which would represent the **revenue and quantity for each region.**
- Add the filter to only show the top 10 countries by revenue (exclude UK)

Steps to create side by side bar chart:

- The purpose of creating a side-by-side bar chart is to **compare two measures across one dimension.**
1. Drag **revenues** to rows for vertical chart **(first measure)**
 2. Drag a Dimension, **Country**, to column → will show a separate bar for each country's revenue **(dimension)**
 3. Drag another measure, **Quantity**, close to the existing measure, and look for a double ruler icon to appear, which indicates that you're adding a second measure to the same axis → side by side chart will be created **(second measure)**
 4. Use the **showme section** to add different colors to each bar

Filtering top 10 countries by revenue

1. By clicking on the “Sort” icon (arrow down) found in the **toolbar**, you’re arranging the countries in the bar chart based on the revenue in descending order.
2. Drag the main dimension (Country) to the filter, choose top in dialog box and select 10

Excluding UK by creating a set

1. **Right-click on the Countries dimension and select Create Set.** This opens a dialog box **where you can create a new set**, which is a **custom subset of the data**.
2. **Name the set** "Countries excluding UK".
3. **Exclude the United Kingdom from the members.** In the set creation dialog, you should see a list of all the countries in your data. Find the United Kingdom in the list and deselect it, then click OK to create the set.
4. **Place the newly created set on the Filters shelf. This will apply the set as a filter to your view, excluding the United Kingdom from the chart.**

▼ Question 3

- Create a **vertical bar chart** where each bar would represent the revenue generated by the customers.
- You would need to add a filter to only display the top 10 customers.
- The customers will be identified from the field “**CustomerID**”. Make sure that the customers who do not have any customer IDs are excluded from the visual. Finally, sort the customers in descending order based on the total revenue generated.

▼ Making Bar Charts in Tableau

- Drag the "CustomerID" to Columns and the Revenue to Rows
- To display only the top 10 customers by revenue, you **need to apply a Top N filter** → drag "CustomerID" to filters → dialog box will open →

deselect null values and enter '10' in the "Top" field in the Top tab

- Click on the "CustomerID" field in the Columns shelf to select it → Then click on the "Sort Descending" button on the toolbar

▼ Question 4

- The CEO is looking to gain insights on the demand for their products.
- He wants to **look at all countries and see which regions have the greatest demand for their products.**
- Once the CEO gets an idea of the regions that have high demand, he will initiate an **expansion strategy** which will allow the company to target these areas and generate **more business from these regions.**
- He wants to view the entire data on a single view without the need to scroll or hover over the data points to identify the demand.
- **There is no need to show data for the United Kingdom as the CEO is more interested in viewing the countries that have expansion opportunities.**
- You would need to create a **map chart** here as the **map chart** would allow the CEO to view the entire map of the world and it will **highlight each country and show the total number of units sold or the name of the country.**
- You would need to filter out the United Kingdom from the data as it is not needed in this visual.
- Make sure that the name of the country or the total units sold is visible for each country.

▼ Building a Simple Map

- Add longitude (generated) and latitude (generated) to Columns and Rows respectively → This will generate a map
- Double click on "Country" Field
- On the Marks card, click the Mark Type drop-down and select Map.

- Drag **“Quantity”** to Color on the Marks card → This will color the polygons on the map to show the amount of demand using color (more darker = more quantities)
- Drag **Country** to Filters and then exclude UK
- Show labels on each country → On the Marks card, click Label, and then select Show mark labels → if countries densely packed, check **“Allow labels to overlap other marks”**

File → "Export Packaged Workbook..." → save as .tbwx

▼ Task 4 - Communicating Insights and Analysis

- Relay your findings and specify how it relates to each scenario
- Effectively communicate insights to business leaders
- Inform them of the most pertinent information that comes from the analysis
- Communicating results also offers → valuable feedback, further refine data analysis skills and continue learning how to anticipate needs from business leaders
- Use less technical jargon

▼ Background information

- You are **required to present your findings to the CEO and CMO.**
- They are **interested in understanding your thought process and your methods in handling data cleanup and visualization.**
- The presentation must be **clear, straightforward, and understandable.**
- Your task is to present the analysis of all four questions from the previous task.
- It's important that you **know the data and the conclusions you've drawn from your analysis well.**
- The primary focus of the leaders is **business expansion**, so they're interested in understanding where the most lucrative opportunities are.

▼ To Do

1. Review the four questions asked by the CEO and CMO. Understand what information they are most interested in.
2. Revisit the work done in the previous tasks, including the data analysis and visualizations you have created.
3. Identify the key findings in your data analysis that are most relevant to the questions asked by the CEO and CMO.
4. Develop a script for your video presentation:
 - Start with a brief introduction of the purpose of your presentation.
 - Explain the process you undertook to complete the analysis. This should include your initial data load and clean-up steps.
 - Present your key findings related to the four questions, explaining each visual and its significance.
 - Make sure to provide context for each scenario. **Explain how your findings relate to the questions asked by the CEO and CMO.**
 - Remember to highlight the most important points that are most likely to be of interest to the CEO and CMO.

▼ Script

Slide 1 - Introduction

[Include Company Logo]

Good morning, everyone. I'm excited to share with you some crucial insights we've deducted from our analysis of our online retail data. My aim is to present this in a way that directly relates to our business strategy and objectives

Slide 2 - Overview of Presentation

We'll be looking at four key areas today: revenue trends, country-specific revenues, customer segmentation, and product demand. I believe these insights will provide valuable direction as we consider our expansion plans.

Before we dive into the findings, I want to emphasize the diligence we have taken in preparing our data for analysis. We've cleaned the data, removing null values and applying consistent filters to ensure the reliability of our insights

Slide 3 - Answering Question 1

Let's start with our revenue trends. This line graph represents our revenue flow throughout 2011. You can see that there's an upward trend between August to November, and we have a strong market presence during November, just before the Holiday season

Slide 4 - Answering Question 2

Next, we look at the revenues from different countries. This side-by-side bar chart shows the top 10 countries generating the highest revenue, excluding the UK. As you can see, countries including Ireland, France, Germany and Netherlands outcompeted the rest of the countries in the Top 10 by a large margin. As we strategize our expansion plans, these countries are potential markets that could yield significant returns.

Slide 5 - Answering Question 3

Moving on to customer segmentation, this bar chart shows the top 10 customers generating the highest revenue. Understanding who our top customers are can help us refine our marketing strategies to better serve this segment and attract similar profiles.

Slide 6 - Answering Question 4

Lastly, this map chart highlights each country and shows the total number of units sold, along with the revenues as well. The darker the color, the more the total number of units sold in that country. This will guide us in identifying which regions have the highest demand for our products and where we may need to increase our supply.

Slide 7 - Conclusion

In conclusion, by analyzing our online retail data, we've identified key revenue trends, promising markets for expansion, our top customers, and product demand by country. These insights provide a solid foundation for our future strategy discussions. Now, I'd like to open the floor for questions or feedback. Thank you for your time."

▼ Example Script

Good Afternoon,
I'm [insert name], and I'm excited to share some insights about your business.
Thank you for
providing the guiding questions. It was helpful to see what types of insights you
are looking to
gain from the data. I hope you find the analysis compelling and helpful as you
make decisions
regarding future business opportunities.
First off, I want to assure you that I've provided the most up to date and error
free analysis. After
I loaded the data into my software, I scrubbed any records that have negative
quantities and
unit price, as these records needed to be removed in order to provide helpful
analysis.
As for your first question, the CEO has requested a trend of the revenue to see if
there is any
seasonality in the store sales. My analysis shows that there are some months of
the year where
exceptional growth is witnessed. The data shows that the revenue in the first 8
months is fairly
constant as the average revenue generated for these 8 months is around \$685k.
The increase
in revenue starts in the month of September, where the revenue increases by
40% over the
previous month. This trend continues till the month of November where it
reached 1.5 million
USD, the highest during the entire year. The data is incomplete for the month of

December, therefore, no conclusion can be drawn from it, unfortunately. This analysis shows that the retail store sales are impacted by the seasonality which usually occurs in the last 4 months of the year.

The second visual shows how the top 10 countries which have opportunities for growth are performing. This data does not include the UK as the country already has high demand and I've been told you're more focused on the countries where demand can be increased. The analysis shows that countries such as the Netherlands, Ireland, Germany and France have high volumes of units bought and revenue generated. I would suggest that these countries should be focused on to ensure that measures are taken to capture these markets even more. The third analysis has been performed on the top 10 customers who have purchased the most from the store. The data shows that there is not much of a difference between the purchases made by the top 10 customers. The highest revenue generating customer only purchased 17% more than the 2nd highest which shows that the business is not relying only on a few customers to generate the revenue. This shows that the bargaining power of customers is low and the business is in a good position.

Finally, the map chart shows the regions that have generated the most revenue compared with the regions that have not. It can be seen that apart from the UK, countries such as Netherlands, Ireland, Germany, France and Australia are generating high revenue and the company should invest more in these areas to increase demand for products. The map also shows that most of

the sales are only in the European region with very few in the American region.
Africa and Asia

do not have any demand for the products, along with Russia. A new strategy
targeting these

areas has the potential to boost sales revenues and profitability.

Thanks so much for your time. If you have any questions about the analysis or
would like to see

anything a

Video in Google Drive → recorded using OBS

- Used Tableau Stories and Dashboard feature to give a presentation that has interactivity in it too as built into Tableau