TCS Intern

**Questions of interest to the CEO**

**Which region is generating the highest revenue, and which region is generating the lowest?**  
This question is important to the CEO as it is based on the fundamental source of income for the business, i.e., revenue. Revenue analysis is important to the CEO as top-level executives are always focused on earnings and how to increase it. Here, the CEO is interested in the viewing revenue by the regions, to assess which regions are generating the highest revenue and which regions are generating lower revenue. Using the data and  
analysis, the CEO will be able to decide on how to further generate revenue in the regions that are already generating the most revenue. For the regions that are not generating enough revenue, the CEO will then study the reasons why there is a lack of sales in those regions and try to improve the products and make them more suitable for those regions.

**What is the monthly trend of revenue, which months have faced the biggest increase/decrease?**  
A monthly trend of revenue will provide the CEO with insights on how the revenue is fluctuating each month. This will enable the CEO to analyze how the internal changes inside the company have had impact on the sales. E.g., how a new product launch has led to an increase in revenue during the month or how the introduction of a new region has led to an increase in revenue for the online store. The CEO can also analyze if there have been any delays internally that would have caused a potential decrease. Such analysis is vital for the senior management as it would enable them to plan ahead and try to make the customer experience as smooth as possible.

**Which months generated the most revenue? Is there a seasonality in sales?**  
In retail businesses, there are always months that will have a greater demand due to seasonality. There will be cases where the data will experience regular and predictable changes that recur every calendar year. Such seasonal months would be necessary to identify as the CEO would be interested in devising a strategy that would gain the maximum benefit from the months that have greater demands.

**Who are the top customers and how much do they contribute to the total revenue? Is the business dependent on these customers or is the customer base diversified?**  
This analysis is highly important as it would enable the CEO to identify what the main drivers are behind the total revenue. Looking at the top customers of the retail store would provide an idea of which customers are  
contributing the most to the revenue. The store can then derive a strategy where the top customers can be targeted with more products that they can buy. This will ensure higher revenue for the store as these customers  
are the top buyers from the store. Although having fewer customers buying in high volumes can be beneficial for a business, there can also be a drawback. Retailers would have less bargaining power with these customers because they drive the majority of the revenue for the store and can negotiate lower prices. The CEO needs to be notified of the diversification of the customers so that he can plan ahead of time. In cases wehre the business is highly dependent on a few customers, the plan would be to increase the customer base and target more customers that would bring more revenue to the store.

**Questions of interest to the CMO**

**What is the percentage of customers who are repeating their orders? Are they ordering the same products or different?**  
This question shows that the CMO is interested in viewing the trends in customer orders. He is interested to know how many customers out of the total are coming back to them and re-ordering. This analysis will help explain to the CMO what percentage of customers are buying from them more than once. Once this is identified, the CMO can come up with a strategy to target these customers with more offers and products that they would need. The analysis will also be done to see what they are buying the second time, this will provide the CMO trends into what products and sub products are in demand and then a marketing strategy can be devised to target these customers with better options.

**For the repeat customers, how long does it take for them to place the next order after being delivered the previous one?**  
This analysis will help the CMO identify the frequency of orders. This would mean determining how long the customers are taking to re-order from the store. The expectation is that those customers who have recently made a purchase would have the product on their mind and are expected to purchase or use the product again in the future. Once the information is gathered from the analysis, the CMO can create a strategy to get the recent customers to revisit the business and spend more. For the customers who have not made purchases again from the store, efforts can be made to remind them that it has been a while since they last purchased from the store. Incentivizing customers also comes into play in this scenario.

**What revenue is being generated from the customers who have ordered more than once?**  
Revenue stems from how much the customer spends to purchase the products from the store. Therefore, the analysis needs to be done to determine how much revenue is being generated from the customers who are regular buyers from the store. The CMO can devise a strategy to encourage customers who spend more money on repeat purchases to continue to do so. It is also important to note that if a customer has made a big purchase the first time, they should be encouraged to come and shop from the store again. A marketing strategy will ensure that the high paying customers will continue to bring more revenue to the store going forward, as well.

**Who are the customers that have repeated the most? How much are they contributing to revenue?**  
It is also important to assess which customers are repeating the most and how much are they contributing to the revenue. There would be customers who need the same products on a weekly or monthly basis, however, the products do not have a high monetary value. Therefore, the contribution to revenue for these customers will be low. On the other hand, there might be customers who are ordering twice a year and have very big orders in terms of revenue. These customers buy on certain months only, therefore, the management needs to ensure that enough supplies are available to accommodate their orders. The customers will high order volumes and low revenue would need to be offered more discounts so that they can buy in bulk and lead to more revenue.

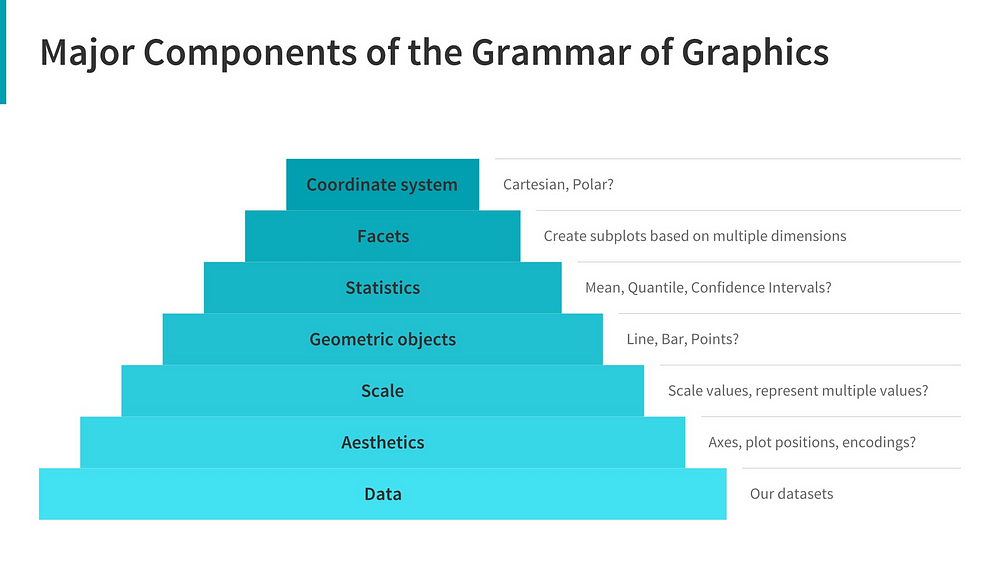
Here are my questions:

**Questions for the CEO:**

1. **Revenue Breakdown**: What are the major sources of revenue, and how have they trended over the past year?
2. **Customer Demographics**: Which customer segments have shown the most significant growth, and what demographics are driving this trend?
3. **Product Performance**: Which product categories have been the highest performers, and which ones have underperformed?
4. **Competitive Landscape**: How do we currently stand against our key competitors in terms of market share and sales growth?

**Questions for the CMO:**

1. **Campaign Effectiveness**: Which marketing campaigns have delivered the highest return on investment (ROI) and significantly contributed to customer acquisition?
2. **Digital Channel Performance**: How have different digital marketing channels (e.g., social media, email, PPC) performed in driving traffic and conversions?
3. **Customer Acquisition Cost**: What trends are we seeing in customer acquisition costs, and how are these impacting overall marketing efficiency?
4. **Customer Engagement**: What strategies have been most effective in engaging customers across various channels, and how can we enhance these efforts next year?

We illustrate the same using a pyramid architecture to show an inherent layered hierarchy of components. Typically, to build or describe any visualization with one or more dimensions, we can use the components as follows.

1. **Data**: Always start with the data, identify the dimensions you want to visualize.
2. **Aesthetics**: Confirm the axes based on the data dimensions, positions of various data points in the plot. Also check if any form of encoding is needed including size, shape, color and so on which are useful for plotting multiple data dimensions.
3. **Scale:** Do we need to scale the potential values, use a specific scale to represent multiple values or a range?
4. **Geometric objects:**These are popularly known as ‘geoms’. This would cover the way we would depict the data points on the visualization. Should it be points, bars, lines and so on?
5. **Statistics:** Do we need to show some statistical measures in the visualization like measures of central tendency, spread, confidence intervals?
6. **Facets:** Do we need to create subplots based on specific data dimensions?
7. **Coordinate system:**What kind of a coordinate system should the visualization be based on — should it be cartesian or polar?

**Data visualization and it importance:  
Table of content**

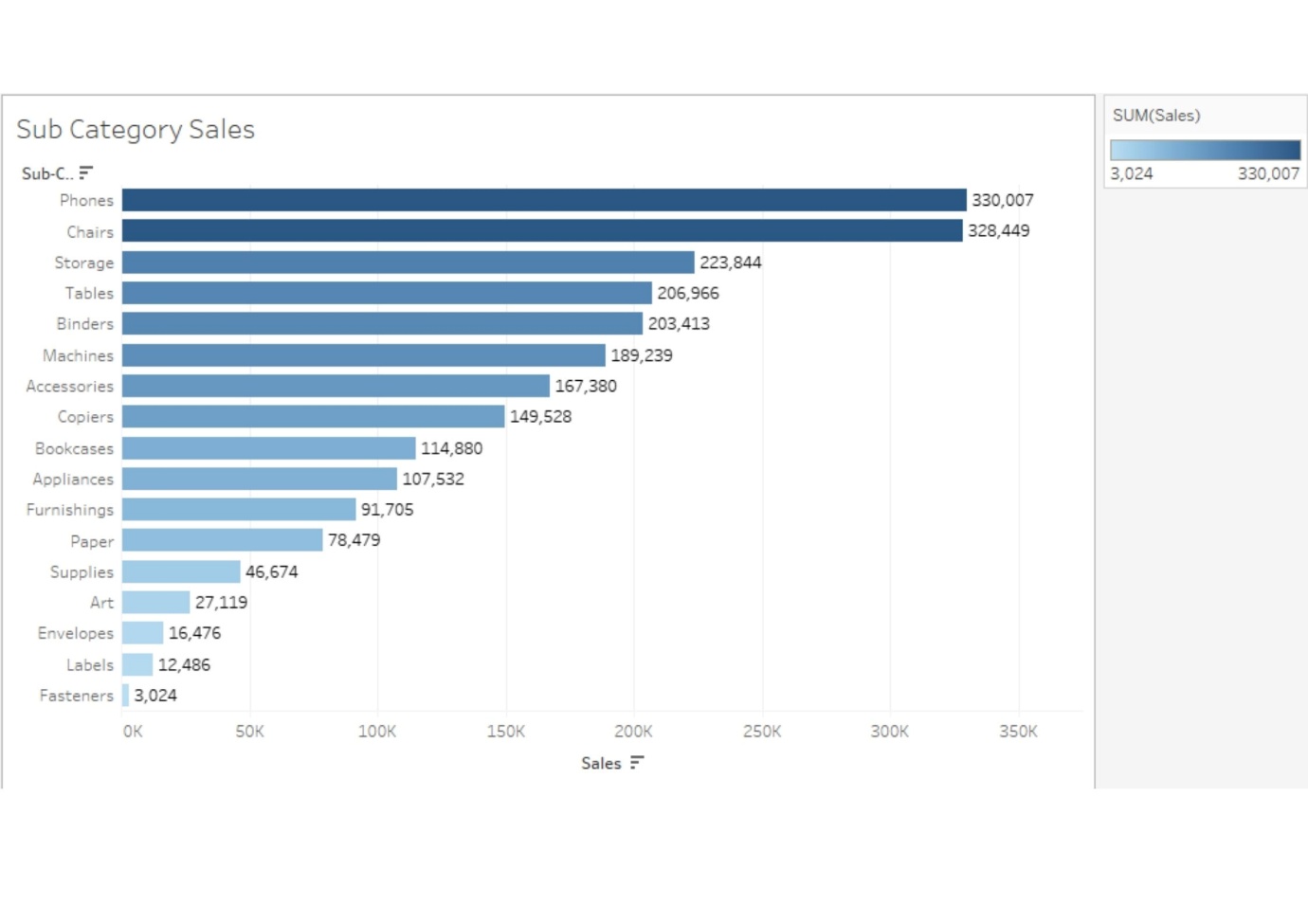
1. Importance of Data Visualization
2. The objective of your visual
3. Choosing the right visualization for your data
   1. Comparison charts
   2. Distribution charts
   3. The breakup of a whole charts
   4. Relationship charts
   5. Trend charts

**Importance of Data Visualization**

Data Visualization is a graphical representation of data and plays a vital role in understanding information in a better way. It is a way to represent data in visual content.

Look at the data that is displayed below:

**Picture 1: Doesn’t make any sense : information in table form cannot understand so we go for 2nd step.**



**Picture 2: This makes sense(because of visualization)**

What do you think, by looking at which picture, you can grasp the insights?

Of course, it is the second picture because of the graphical representation of the data.

I’ve listed down some benefits of visualization:

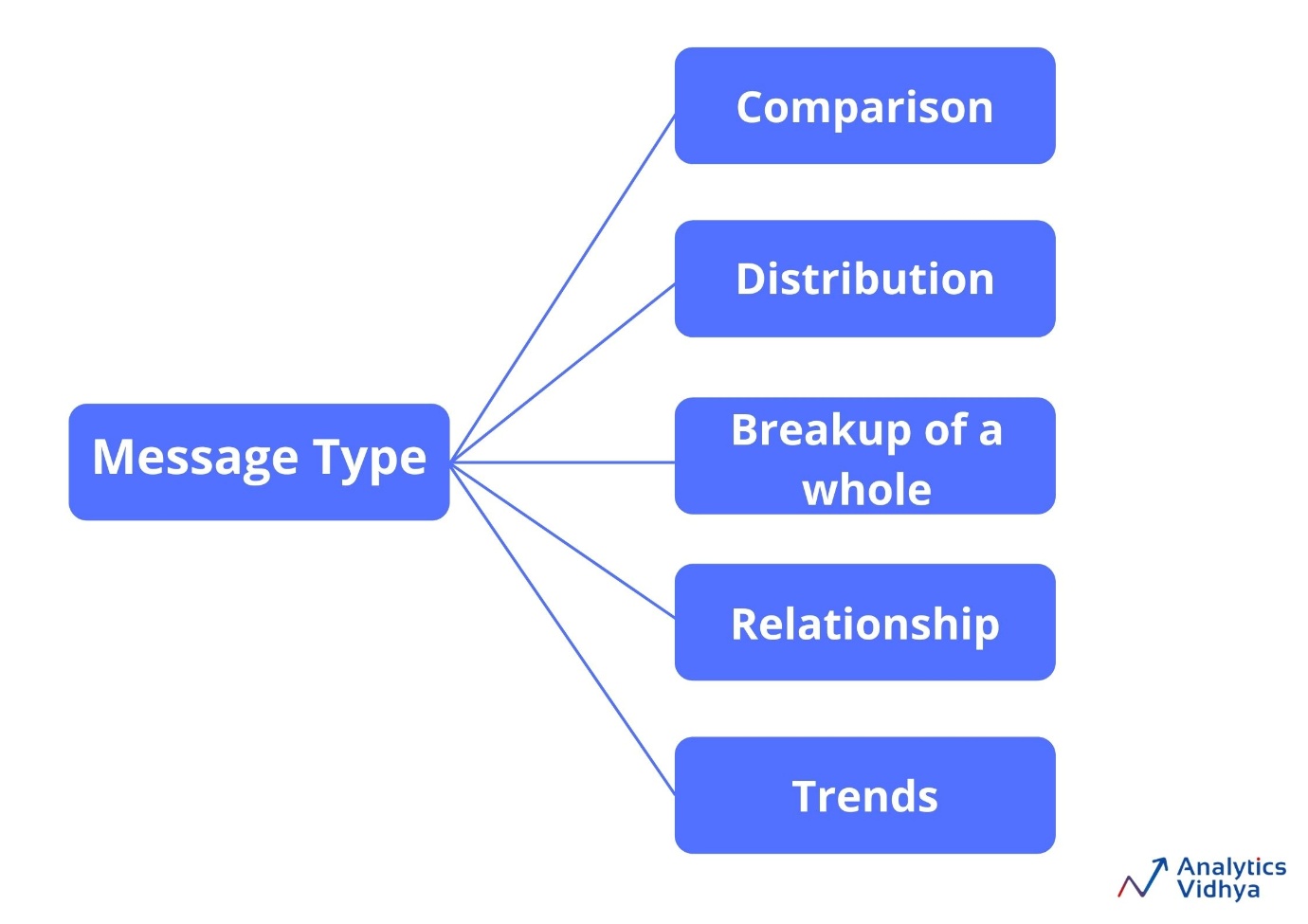
1. It helps us to convey the right message to the audience through visuals.
2. It helps us find outliers in our data.
3. It helps the business leader to take an accurate decision.
4. It helps us to understand how the data is distributed over time.

**The objective of your visual**

Before making the visualization, it is best to ask yourself what the audience will be looking for in your chart. Understand the requirements and preferences of your viewer. Know their background. Do they have enough time for a detailed visualization? How aware are they of the context of the visualization? What additional information are they looking for? Are they aware of the graphs being used? And so on. Your viewer’s information needs should be your guide in creating effective and compelling data visualizations.

**Choose the right visualization for your data**

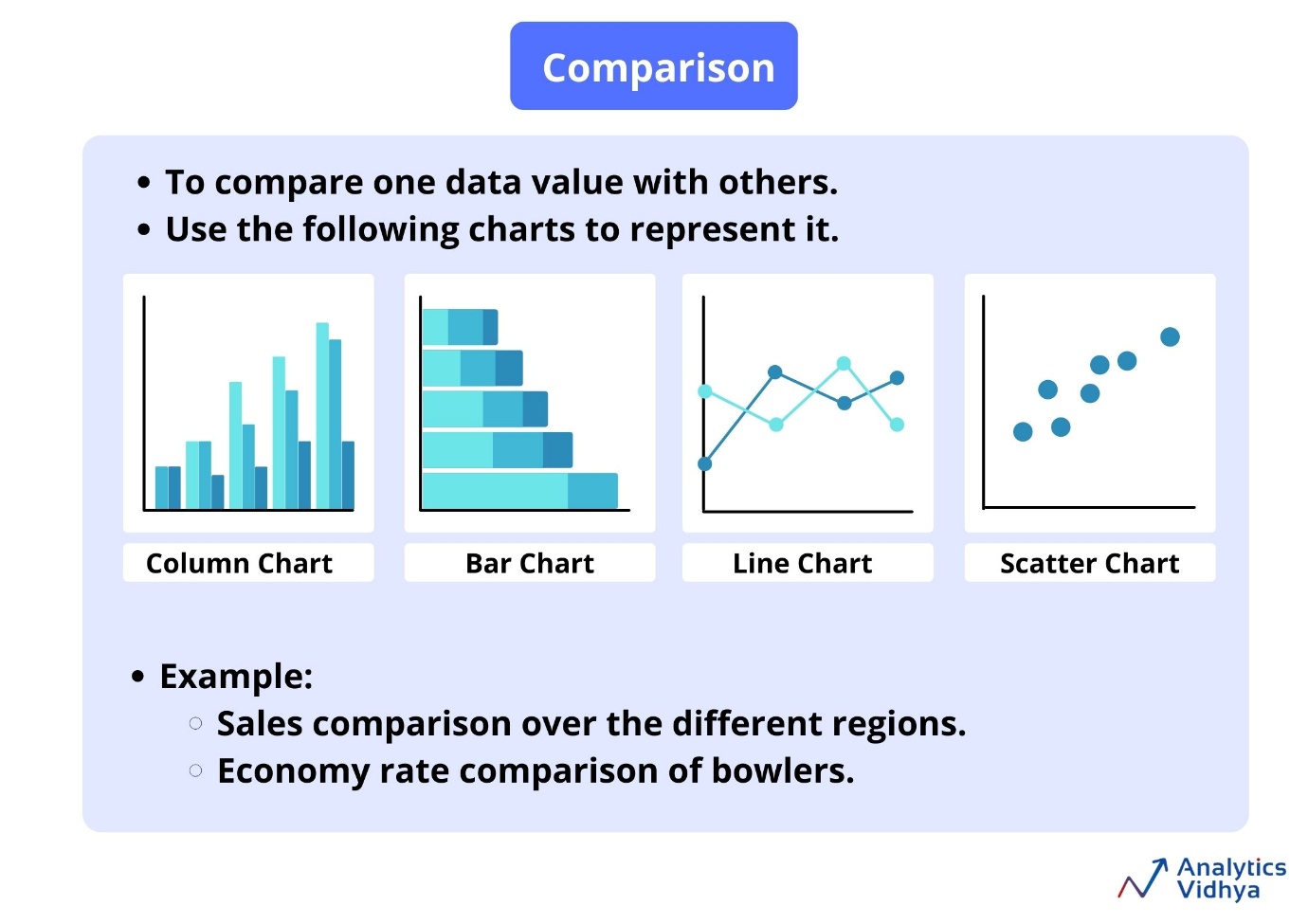
There are a tremendous number of charts available. Choosing the right visualization is paramount when you’re presenting to a senior leader. It is not easy as it sounds, because an incorrect representation can lead to a wrong message or wrong decision taken by the audience or whatever you’ve in your mind when you were creating that chart, that message might not be conveyed to the audience. Here, your focus should be on conveying the right message to your audience in an optimal way. Now let me take you through the type of messages, that we usually send out when we’re creating impactful visualizations in business.

These are the types of messages that you usually work on. Maybe you want to show a comparison of two features for example reason wise sales, the distribution of the data, maybe you want to show the breakup of the entire whole visualization, or you simply want to show trends for example sales trends.

Let’s look at all these one by one and see what kinds of charts we can use to convey the right message.

**1) Comparison Chart**

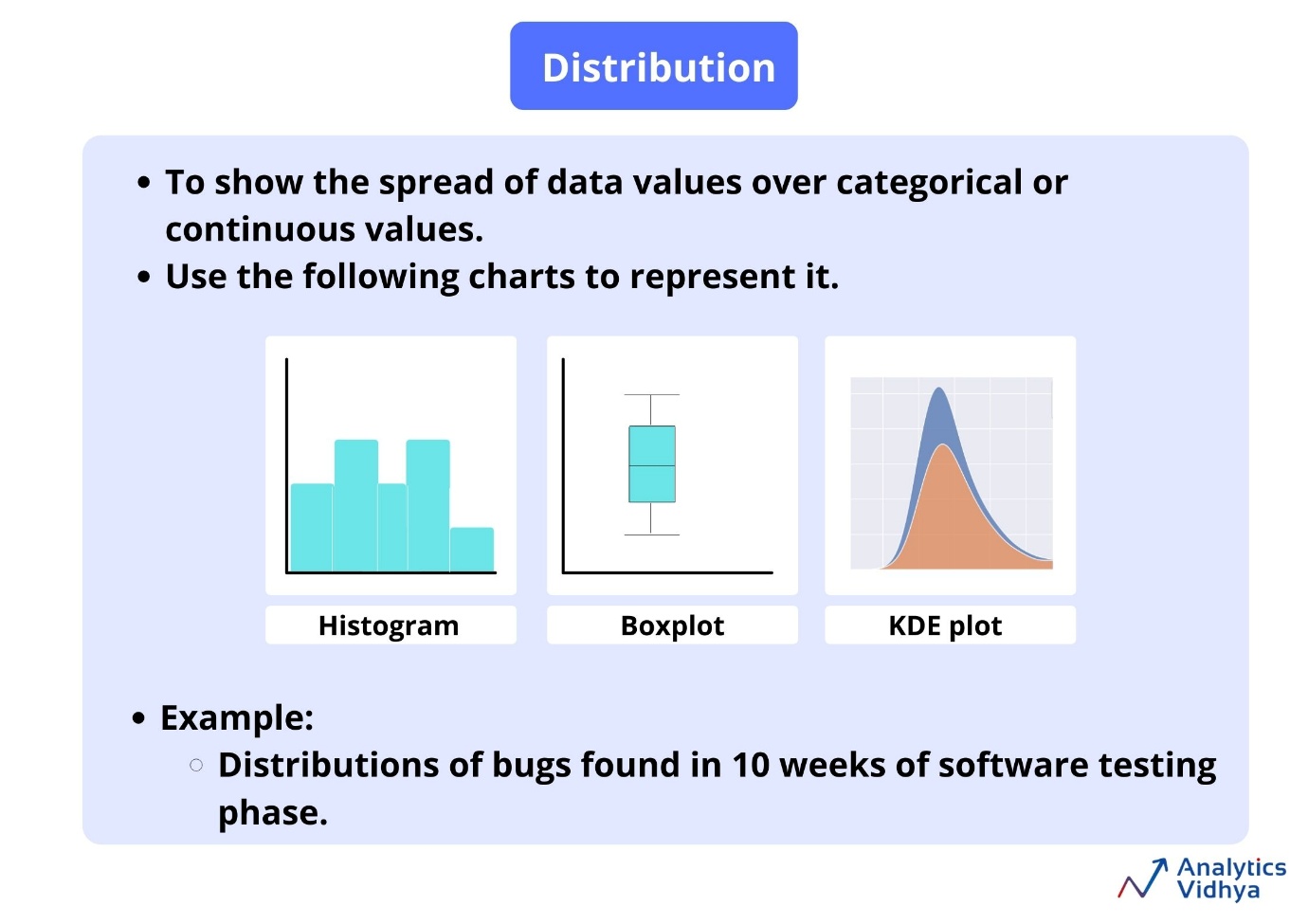
In this chart, we compare one value with the other like region-wise sales, economy rate comparison of bowler in cricket. We can use the following charts for comparison.



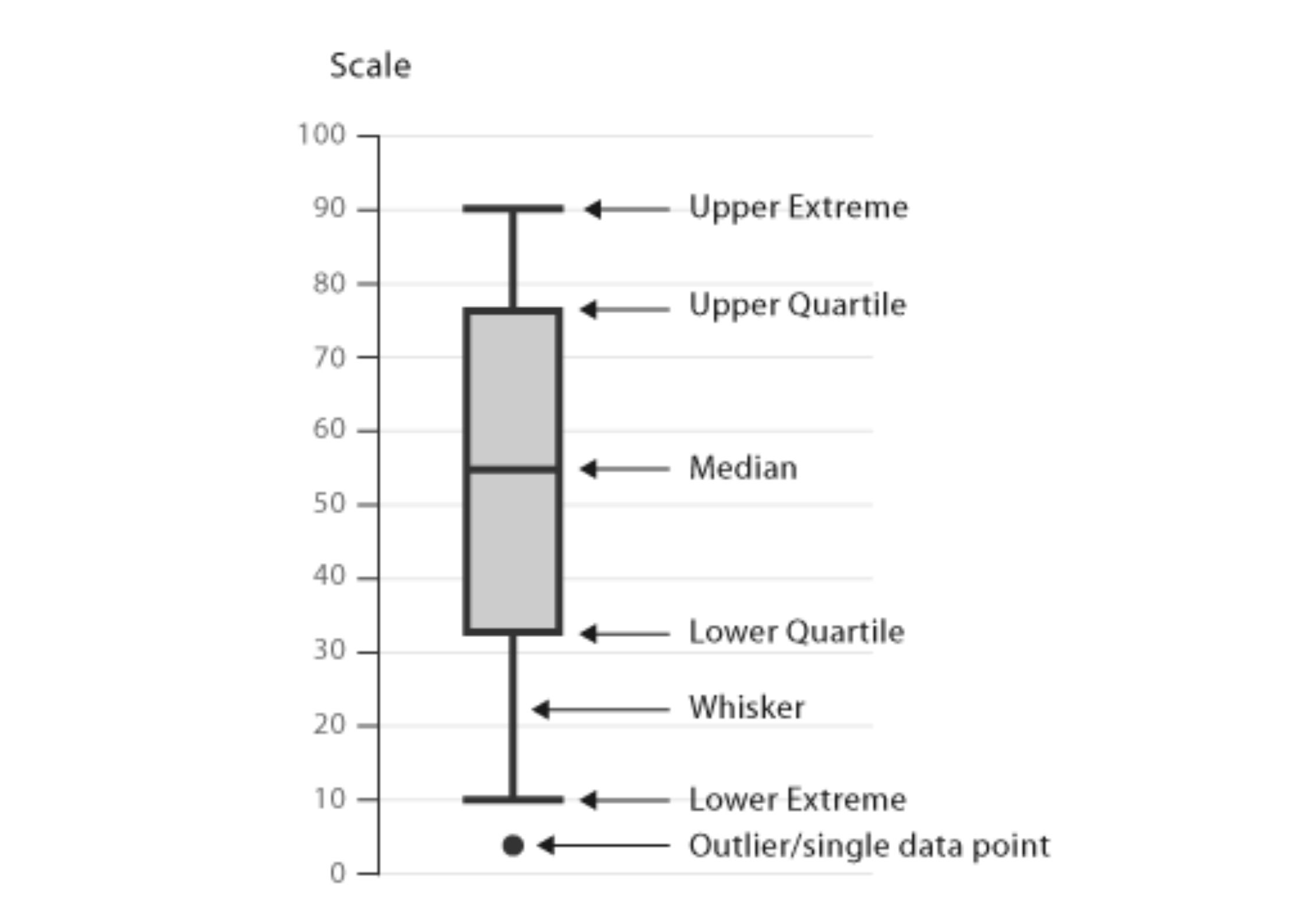
* **Column charts**
  + It is used to compare values across multiple categories.
  + Here, the category appears horizontally(X-axis) and values vertically(Y-axis).
  + In the column charts, you can also show information about parts of a whole across different categories, and you can show this in absolute value as well as relative terms. Here comes the concept of a stacked column chart and 100% stacked column charts.
* **Bar charts**
  + As you’re quite familiar with column charts, you will find that working with bar charts is very synonymous.
  + The only difference between them is that in a bar chart, values are represented on the X-axis and categories on the Y-axis.
  + We typically use a bar graph to show values across categories when the duration or category text is long.
  + Stacked bar charts are used to compare parts of a whole(relative and absolute) and compare change over categories or time.
* **Line charts**
  + It is one of the most popular charts and vitally used in most industries.
  + Whether you’re analyzing sales data, whether you’re looking at year-on-year profit, whether you’re looking at how a person’s salary increases in the last year, line charts are very helpful in these scenarios.
  + The line chart is used to show trends over time or categories.
  + Here, the category appears horizontally(X-axis) and value vertically(Y-axis).
* **Scatter plots**
  + An XY(Scatter) chart uses numerical values along both axes.
  + Scatter plots are useful for showing a correlation between the data points that may not be easy to see from the data alone.
  + It is used for displaying and comparing numerical values, such as scientific or statistical data.

**2) Distribution charts**

* These charts are used to show the spread of the data values over categories or continuous values. We can use the following charts in order to visualize the distribution of the data. For example Distribution of bugs found in 10 weeks of the software testing phase.



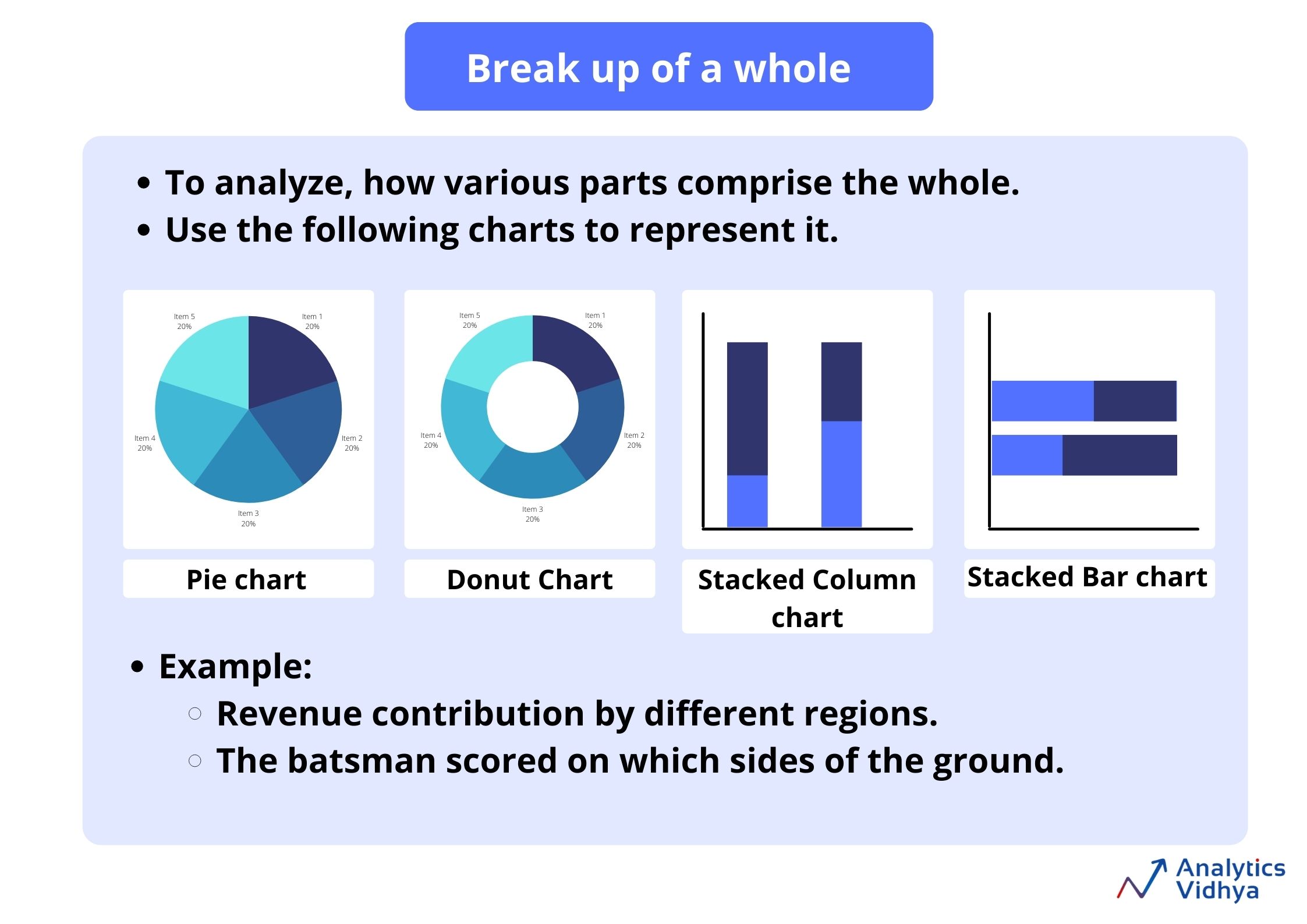
* **Histogram**
  + It is used to graphing the frequency over a distribution**.** It is a very useful graph in the analytics world and can infer many useful insights from the data.
  + Visually, all the bars are touching each other with no space between them.
* **Box plot**
  + It is also known as Box and whiskers plot.
  + The line in the middle of the box is the median value. This means that 50% of the data are above the median value and 50% of the data are below the median value.
  + Medians are useful because they’re not swayed by outliers as mean is.
  + Within the box itself, there is 25% of data above the median and 25% of data below the median, so 50% of the data is within the box.
  + By using this plot, we can easily spot outliers and the distribution of the plot.



* **KDE Plot**
  + KDE is an abbreviation for the Kernel Density Estimation plot.
  + It’s a smooth form of a histogram.
  + A kernel density estimate (KDE) plot is a method for visualizing the distribution of observations in a dataset, analogous to a histogram.
  + Relative to a histogram, KDE can produce a plot that is less cluttered and more interpretable, especially when drawing multiple distributions.

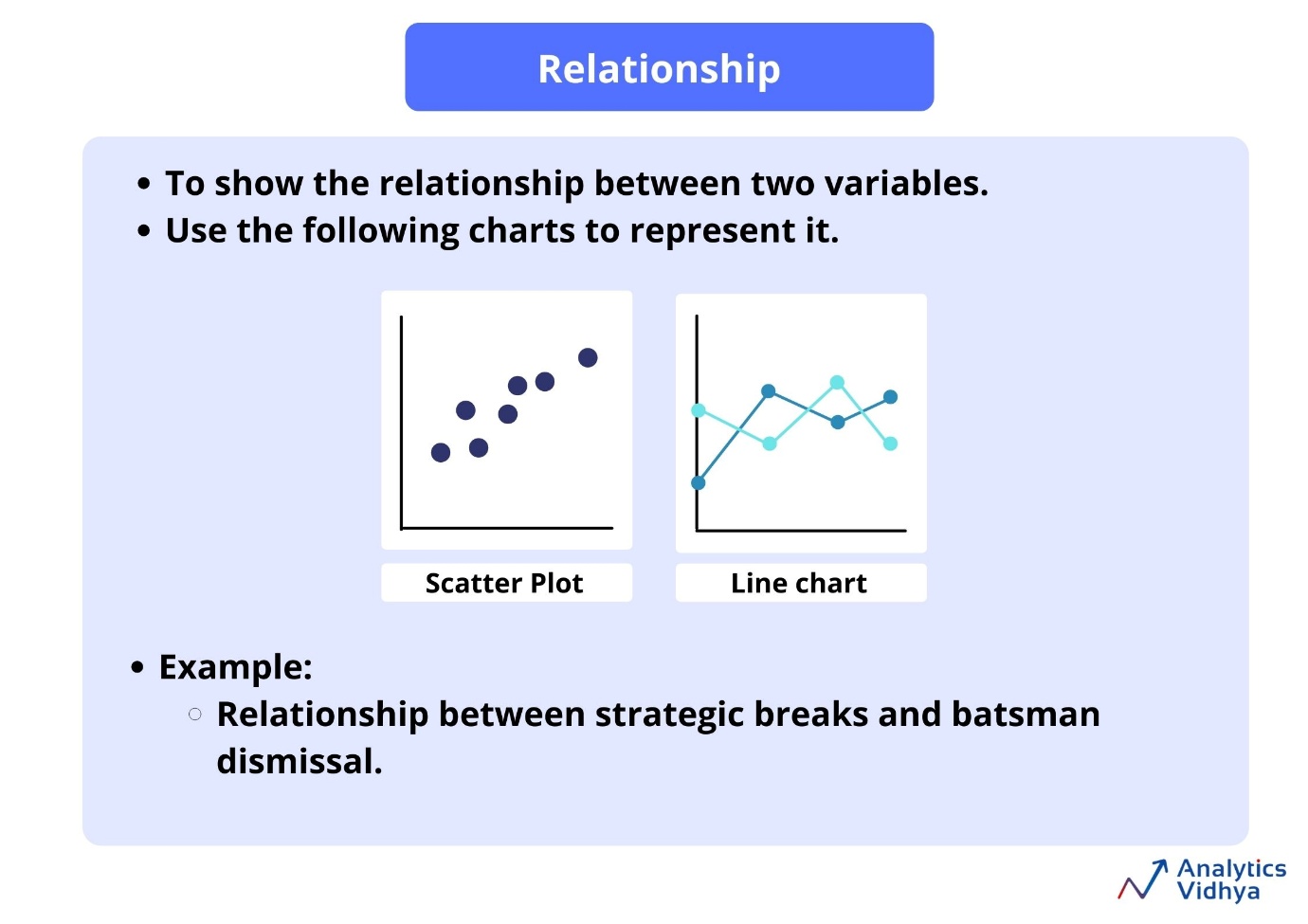
**3) The breakup of a whole chart**

These charts are used to analyze, how various parts comprise the whole. These charts are very handy in many scenarios where we have to analyze revenue contribution by different regions, batsmen scored on which sides of the ground. Charts used to represent these are listed below.



* **Pie Chart**
  + If you want to represent your categorical data as part of the whole, then you should use a pie chart.
  + Each slice represents the percentage that the given category occupies out of the whole.
  + It’s better to use a pie chart if you’re having less than 5 categories.
* **Donut Chart**
  + It is a variant of a pie chart, with the hole in the center.
  + It displays the categories as arcs rather than slices.
* **Stacked Column Chart**
  + A Stacked column chart is used when you want to show the relative percentage of multiple data series in stacked columns, the total (cumulative) of stacked columns always equals 100%.
  + The 100% stacked column chart can show the part-to-whole proportions over time, for example, the proportion of quarterly sales per region or the proportion of monthly mortgage payment that goes toward interest vs. principal.
* **Stacked Bar Chart**
  + A Stacked Bar chart is used to show the relative percentage of multiple data series in a stacked bar.

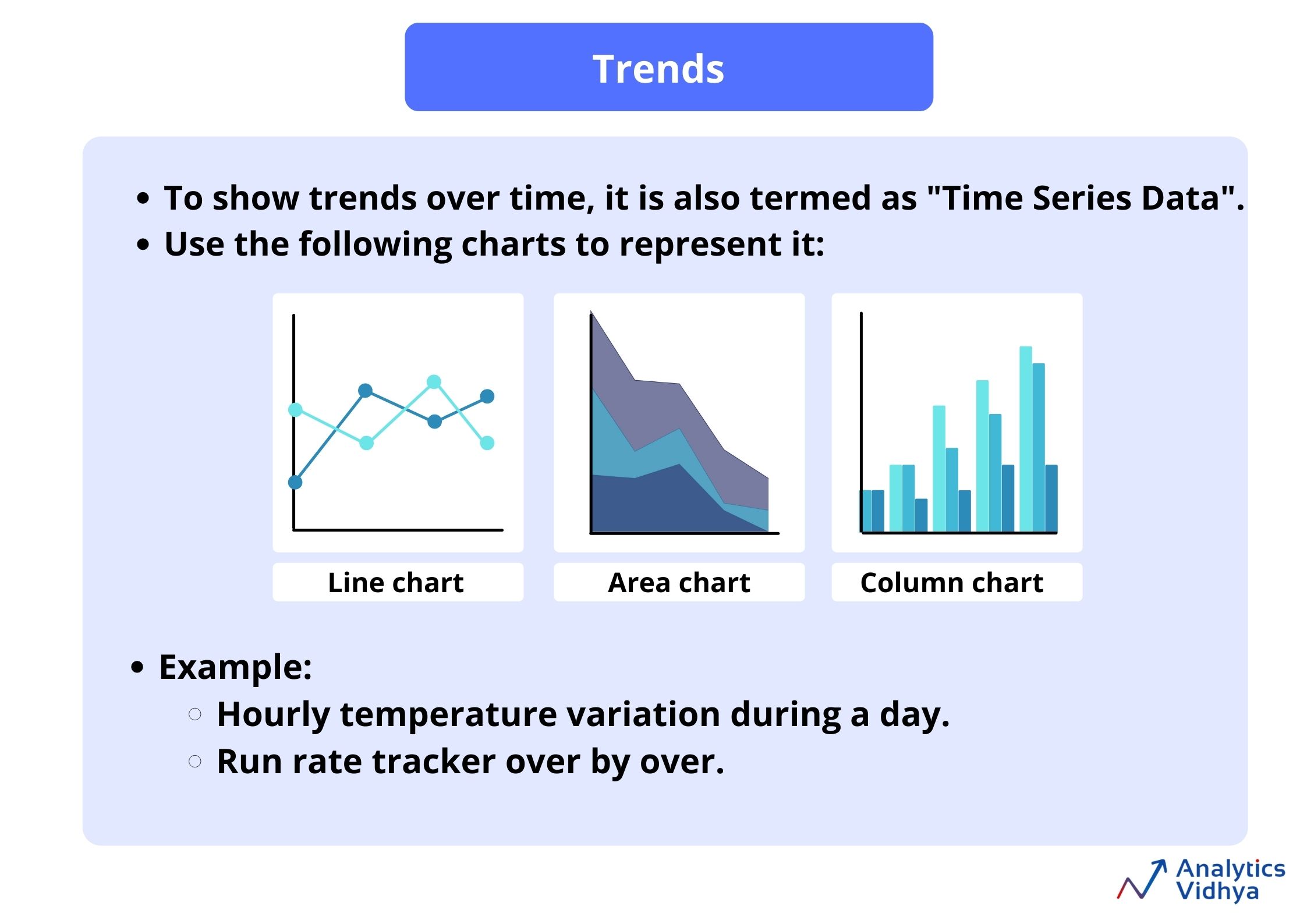
**4) Relationship charts**

These relationships charts are very helpful when we want to know that what is the relation between the different variables. Charts used to visualize the relationship between the variables are listed below.

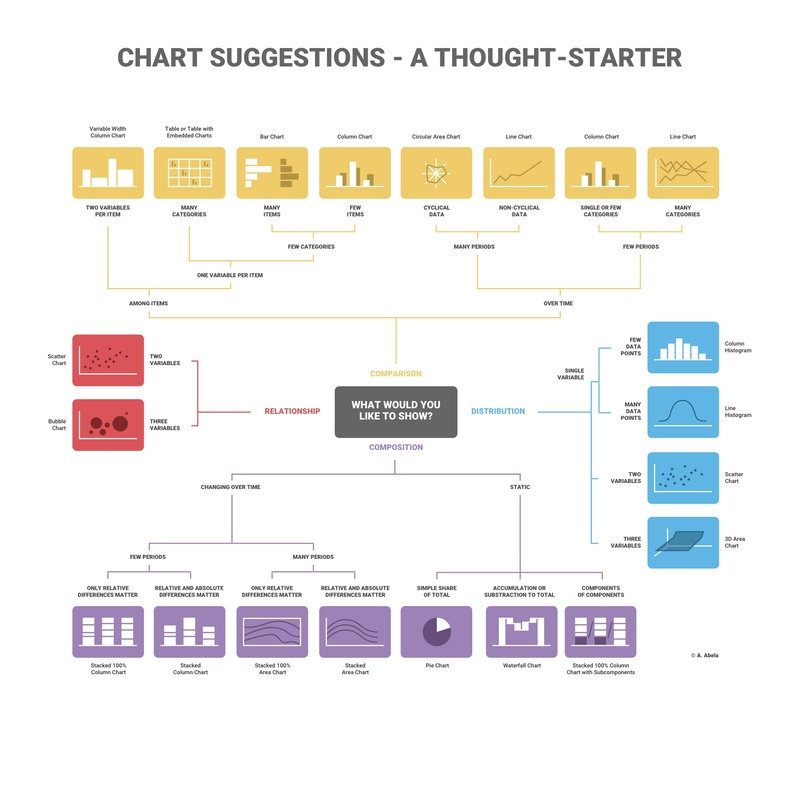
* **Scatter Plot**
  + A scatter chart uses numerical values along both axes.
  + It uses dots to represent the values for two different numerical values.
  + The position of each dot on the horizontal axis and the vertical axis signifier the value of a particular data point.
  + It is useful for showing a correlation between the data points that may not be easy to see from the data alone.
  + It is used for displaying and comparing numerical values, such as scientific or statistical data.
* **Line Chart**
  + As discussed above, a line chart is also used to find the relationship between the two variables.

**5) Trend charts**

This is used to visualize trends of values over time and categories, it is also known as “Time Series” data in the data-driven world. For example Run rate tracker over by over, Hourly temperature variation during a day. Listed below are the charts used to represent time series data.



* **Line Chart**
  + The best way to visualize trend data is by line chart.
  + Line charts are also used to see the trends in various domains.
* **Area Chart**
  + It is used to see the magnitude of the values.
  + It shows the relative importance of values over time.
  + It is similar to a line chart, but because the area between lines is filled in, the area chart emphasizes the magnitude of values more than the line chart does.
* **Column Chart**
  + A column chart as discussed above is also used to show the trends of values over time and categories.

**Chart suggestion by tcs :** 

Q and A

1 Q1/5: The CEO of the retail store is interested to view the time series of the revenue data for the entire year. The CEO is interested in viewing the seasonal trends and wants to dig deeper into why these trends occur. This analysis will be helpful for the CEO to forecast for the next year. Which visual would most likely help the CEO analyse the data?

* 1. Bar chart
  2. Line chart
  3. Pie chart
  4. Scatter plot

Ans: A line chart is used to represent repeated measurements taken over regular time intervals. Time is always displayed on the horizontal axis and values on the vertical axis. The line chart would enable the CEO to see important changes in the data, like seasonality or cyclic behaviour, which will provide a better understanding of the revenue and help forecast better.

2. The CMO is interested in viewing the top 10 countries which are generating the highest revenue. Additionally, as a subcomponent, they would also like to see which products are contributing to the total revenue being generated by each country. Which visual would enable the CMO to view the revenue for each country and the breakdown by products on a single chart?

a. Bar chart

b. Pie chart

c. Boxplot

d. Stacked bar chart

Ans : A stacked bar chart would be used here as the chart allows users to compare subcomponent pieces across different categories. The height or length of the bars will represent the total revenue generated by each country. Each bar will be divided into the products sold, where the major portion will be allocated to the products that are generating more revenue. This will allow the CMO to view the total revenue data as well as the revenue for each individual product.

3. The CEO of the online retail store wants to see how much average revenue is generated by each country. They are interested in viewing the following metrics on the visual: Minimum value First quartile value Median value Third quartile value Maximum value Which chart would you create to show the above metrics for the average revenue generated by each country?

a. Scatter plot

b. Histogram

c. Waterfall chart

d. Boxplot

Ans: the CEO’s request would best be fulfilled using a box plot. The box plot would show the distribution of data based on a five number summary (“minimum”, first quartile, median, third quartile, and “maximum”). Boxplots are used to graphically demonstrate the locality, spread and skewness groups of numerical data. Boxplots are used to graphically demonstrate the locality, spread and skewness groups of numerical data. By using the boxplot, we can easily spot outliers and the distribution of the plot.

4 . The CMO of the online retail store wants to view the information on the top 10 customers by revenue. They are interested in a visual that shows the greatest revenue-generating customer at the start and gradually declines to the lower revenue-generating customers. The CMO wants to target the higher revenue-generating customers and ensure that they remain satisfied with their products. Which visual would help the CMO understand the data on revenue generated by the top 10 customers?

A . Pie chart

b. Stacked bar chart

c . Column chart

d . Area chart

Ans: The CMO’s request would best be fulfilled by using a column chart. Column charts are used to display comparisons between different items. Alternatively, you can view a comparison of items over time. In this case, the column chart would display the top 10 customers who are contributing the most to revenue. The first bar on the left would be the longest one and would display information for the customer who bought the most goods from the online store. The revenue contributed by each customer would gradually decline as the bars move to the right.

5 . The CEO is looking to gain insights on the demand for their products. They want 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, they 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. Which chart would be most useful to provide the CEO information on the demand in each region?

A . Area chart

B . Scatter plot

c. Map chart

d . Pie chart

Ans : A map chart would be the best option for visuals here as it will allow the CEO to view the demand information on a single view. The map chart will have all the countries on a single map and the demand will be displayed by highlighting the area of each country. Dark highlights would mean that the countries have high demand for products whereas the countries that are highlighted light colour are the ones that have low demand. Geographical data is best visualised using map charts as they are very easy to view and the underlying values are also easy to analyse.