

How to Choose the Right Visualization for Your Data — Data Analysis



Abdallah Ashraf · Follow

6 min read · Oct 15, 2023



595



2



When visualizing data, it's crucial to use the chart or graph that best conveys your message clearly and accurately. While your data may work with multiple chart options, you must select the one ensuring your intended meaning comes across. Data only has value if you know how to display and provide context for it visually.

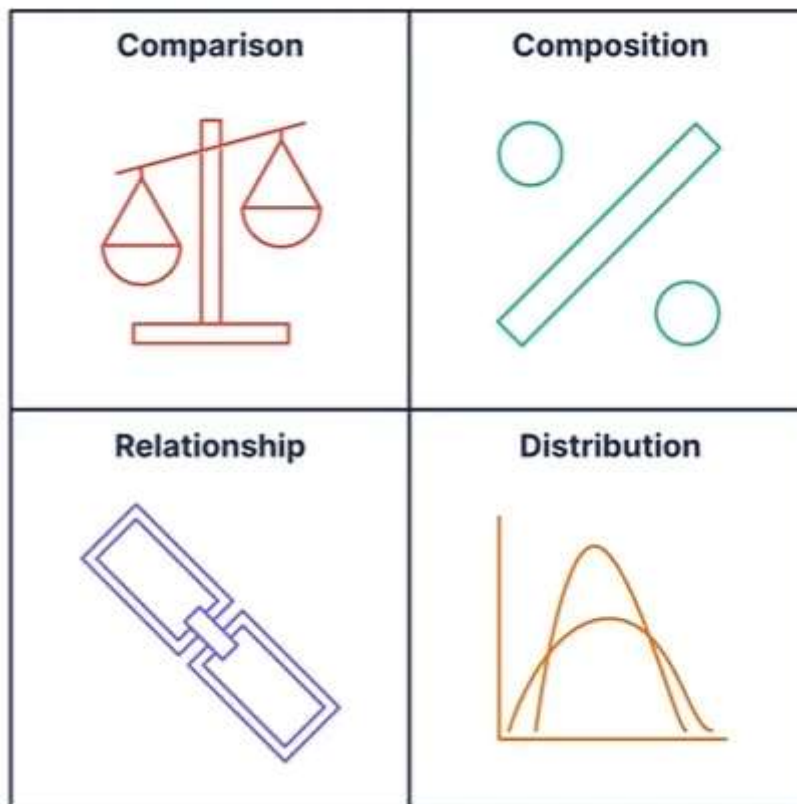
We'll overview the different chart categories and explain how to determine the right fit. First, understand the story your data tells. Charts, maps and infographics help audiences comprehend complex numbers, uncover patterns, identify trends and receive information. Consider the point you want to make with your visualization.

It's also important to follow best practices for charting. Your visual should have numbers that add up properly and be scaled suitably. What are you

aiming to showcase? Carefully examining your data and objective will help you choose the visualization type best serving your needs.

There are generally four main chart types to consider:

- **Comparison**
- **Composition**
- **Relationship**
- **Distribution**



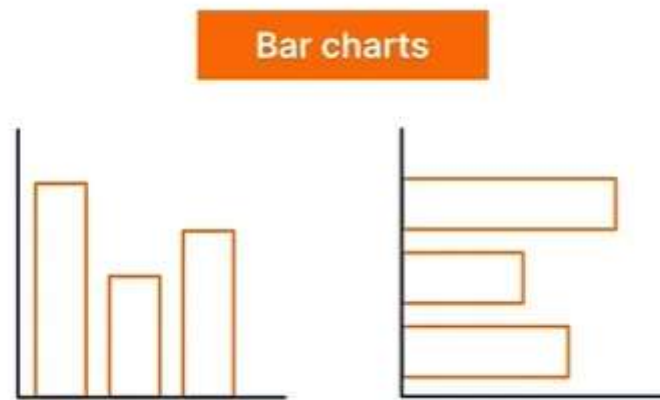
1. Comparison visualizations

They allow you to examine differences between different entities.

Bar charts and **line charts** are common comparison visualizations as they show proportions and differences between categories.

When to use Bar charts ?

- Comparing values for different categories
- Tracking changes over time
- when the changes are large



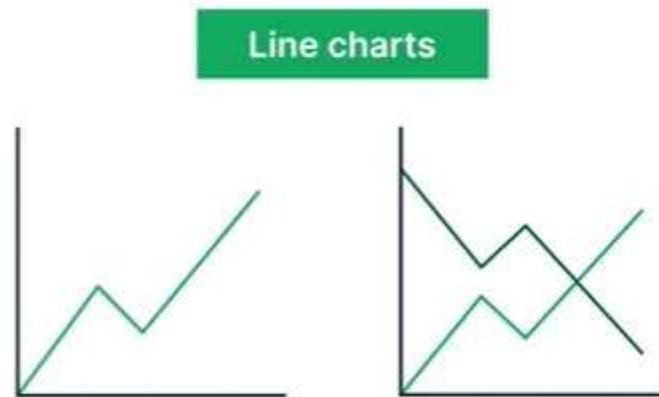
your choice between a horizontal and vertical bar chart can be guided by how much text you have to display at the foot of each bar.

If you have more text, a horizontal display is often more effective.

When to use line charts ?

- Tracking changes over time
- when the changes are small

- Additionally, you can compare changes over time for more than one group



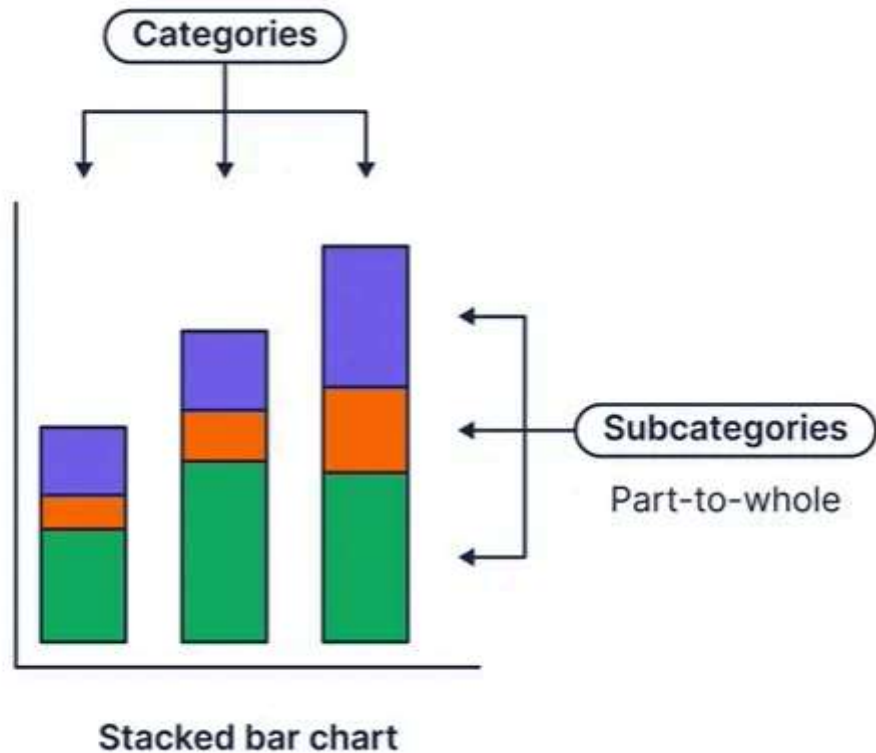
2. Composition visualizations

Composition visualizations provide insight into how individual components combine to form a whole system or dataset. These visuals are ideal for understanding the proportional makeup or hierarchical structure of categorized data.

Commonly used composition charts are **pie charts**, **tree-maps**, **stacked bar charts** and **stacked area charts**.

When to use stacked bar charts ?

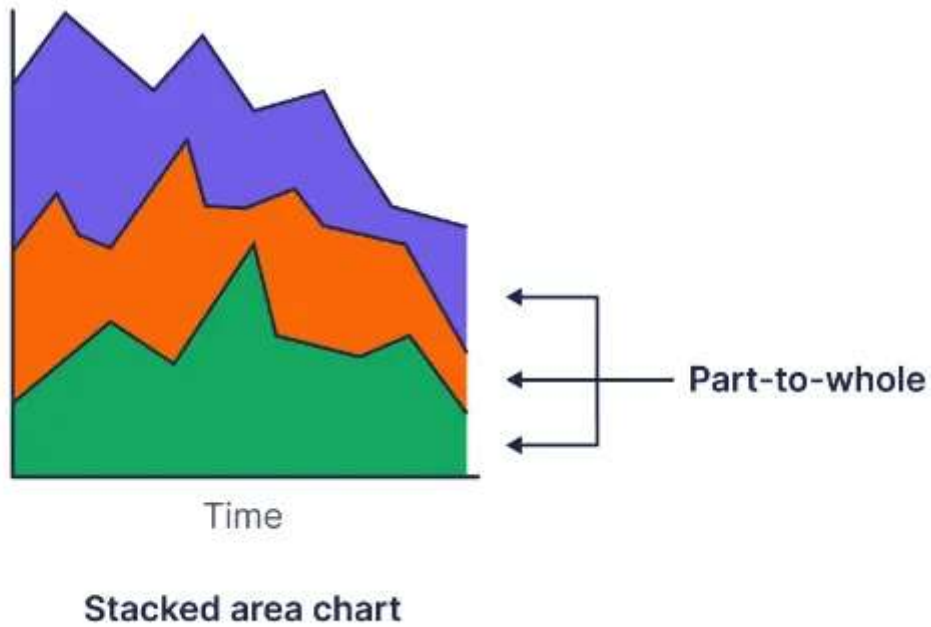
Like a bar chart, it compares different categories, each bar or category can be divided into sub-categories.



It is useful to compare totals across categories and also to show how each sub-category forms a part of the whole.

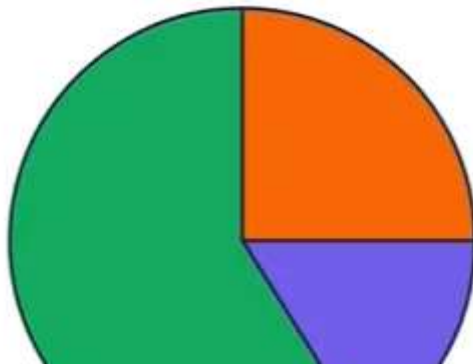
When to use stacked area chart?

Stacked area chart is an extension of the line chart, while the line chart is used to compare values of multiple categories over time, the stacked area chart can do the same while also showing the composition of each category to the whole.



The part to whole relationship will be clearly indicated by filling the sections between the lines and the x-axis with different colors. The sum of all the areas is equal to 100 %.

When to use Pie charts?



Open in app ↗

Sign up

Sign in

Medium

Search

Write



Pie chart is great for displaying the basic composition but it is limited by how many parts you can display.

When categories are close in size it is difficult to tell which is bigger and it becomes hard to read.

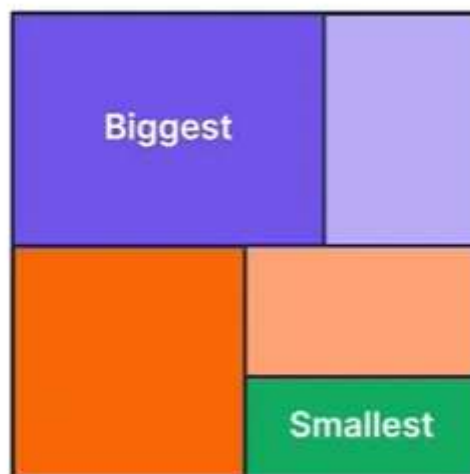
A Pie chart should not be selected if it will be divided into more than five parts.

When to use Treemaps?

When the whole consists of five or more parts, a Treemap is a good choice.

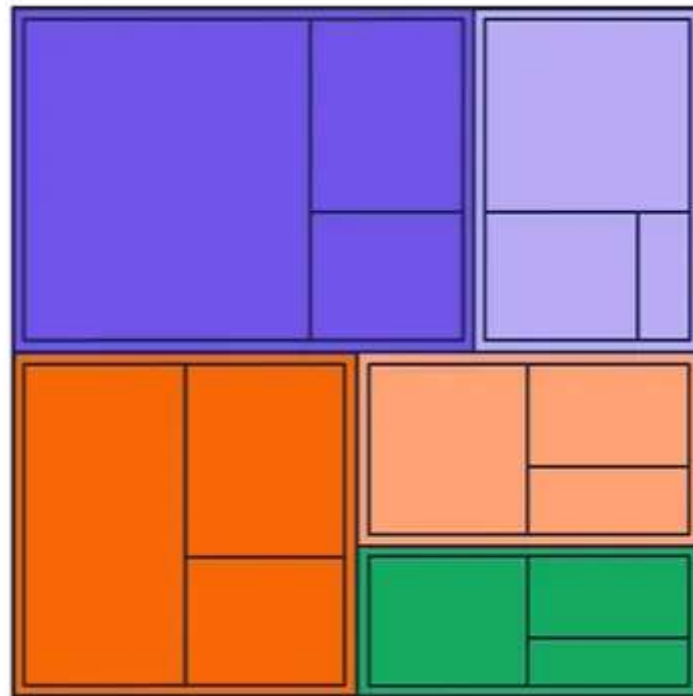
Instead of a circle a rectangle is used, this immediately makes it easier to interpret as we don't need to compare angles, but rather rectangles of different sizes.

The rectangles will be ordered in decreasing size from top left to bottom right.



Treemap

A treemap is also a multi-level visualization as each category or rectangle can be composed of sub-categories.



Multi-level treemap

3. Relationship visualizations

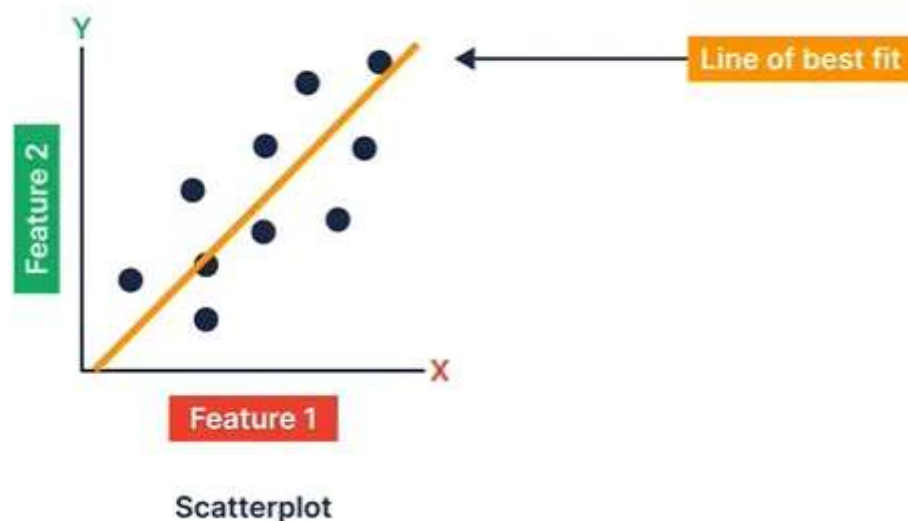
Relationship visualizations uncover relationships between variables.

Scatter plots, **Bubble charts** and **Heat maps** are good for examining correlations and trends between two or more **numerical features**. Look to these when you want to understand how changes in one variable impact another.

When to use the scatter plot?

To find out the relationship between two numeric variables.

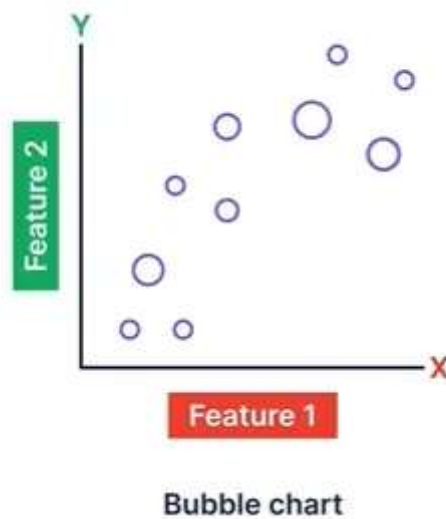
If there is a linear relationship, a line of best fit can be added to the plot.



When to use the Bubble chart?

To find out the relationship between three numeric features.

The bubble chart is plotted the same way as a scatter plot for 2 features, while the third feature will determine each plotted point size, creating bubbles of various sizes.



When to use the Heat maps?

if you have more than three numeric features in a dataset , A heat map will help us to determine which two features are related to each other in a time-efficient manner.

A heat map shows the correlation between each two features in a dataset.

Correlation is a statistic that measures how strongly or weakly two features are related and it is measured from negative 1 to positive 1, with zero representing no relationship.



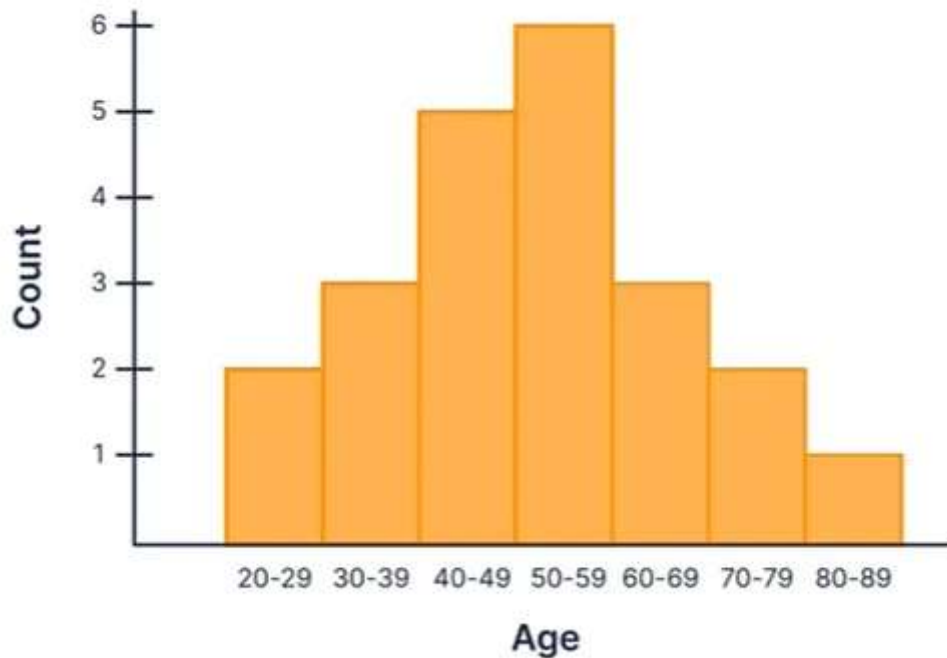
4. Distribution visualizations

Distribution visualizations portray the frequency distribution of values in a dataset. **Histograms**, **box plots**, and **density plots** effectively depict patterns in the spread and concentration of values. These help analyze **how values cluster** and **identify outliers**.

When to use histograms?

Histograms group continuous (numeric) data into bins of fixed widths and count the number of observations in each bin.

The bins are displayed as adjacent rectangles plotted on the x-axis, with height indicating frequency.

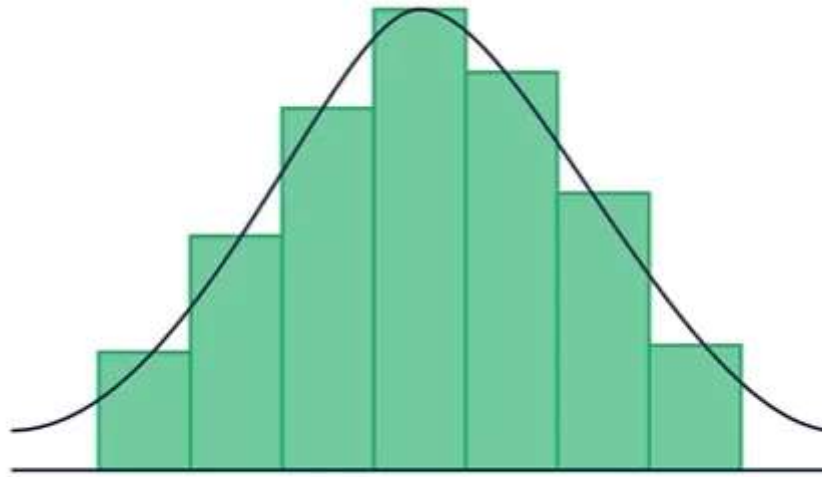


They effectively show the overall shape and clusters in a dataset's distribution.

Histograms are great if you wish to communicate the distribution of the data quickly and easily to others.

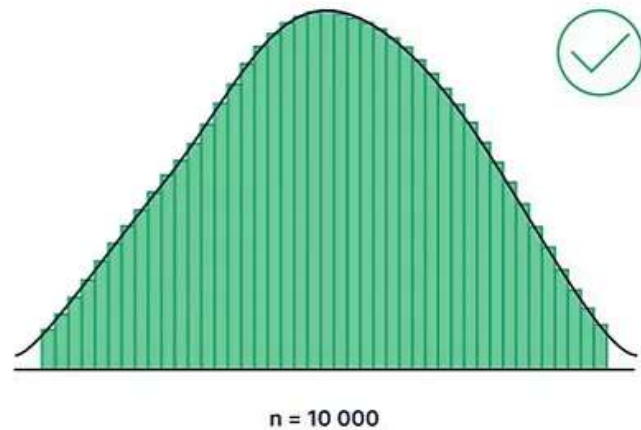
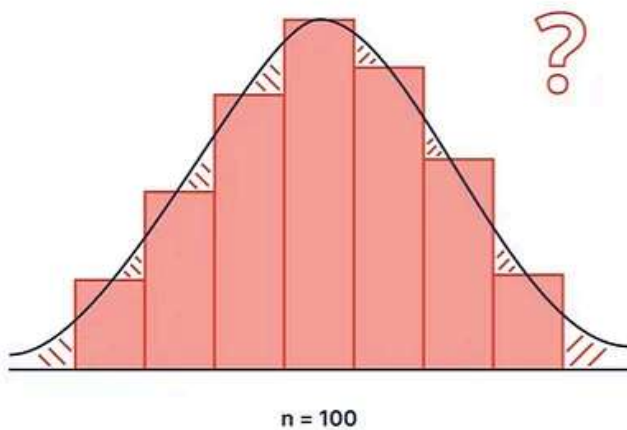
When to use Density plots?

Density plots use kernel density estimation (KDE) to smooth histogram bins into curves, revealing subtler patterns.



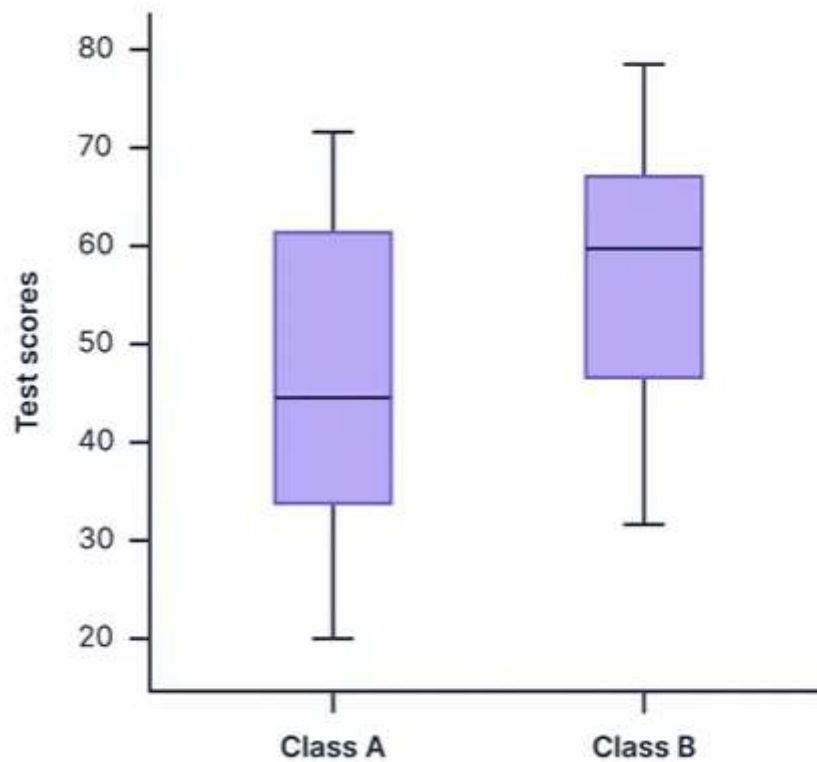
Density curves are very useful when working with large sample sizes.

At small sample sizes, the density curves can be inaccurate due to the missing gaps.



When to use Box plots?

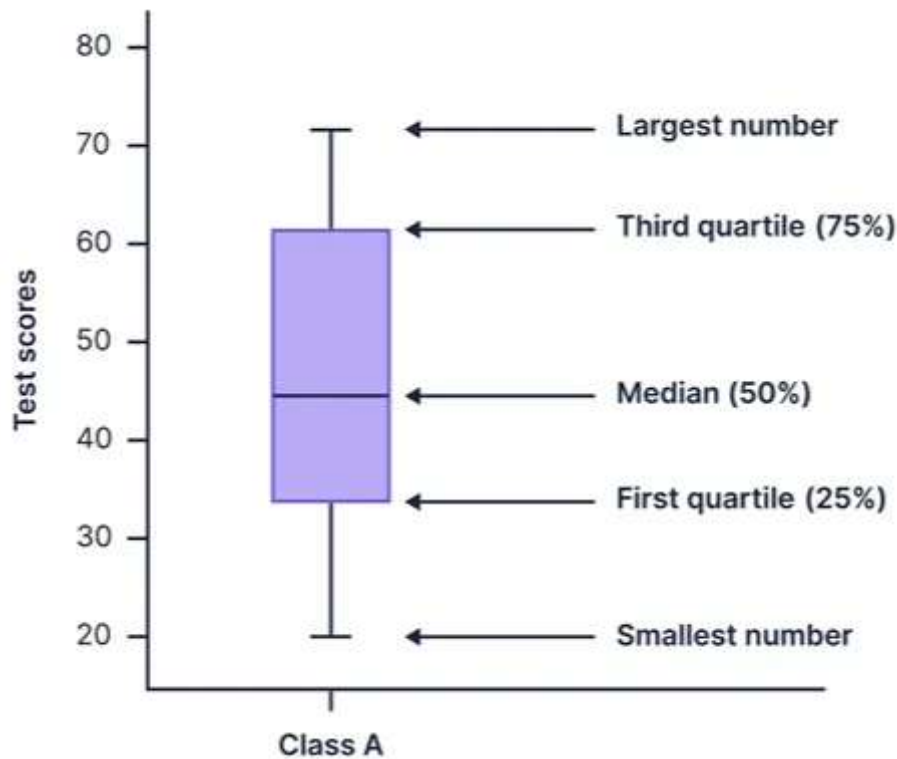
Box plots are used to show distributions of numeric data values, especially when you want to compare them between multiple groups.



At its core, the box plot depicts aspects of the “**five number summary**” — a set of values that describe a dataset’s central tendency and spread.

The Five Number Summary:

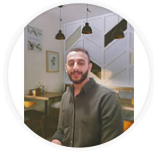
- Minimum value
- First quartile (Q1)
- Median
- Third quartile (Q3)
- Maximum value



The best practice is to explore your data and consider your goals before choosing. Ask yourself what questions you most want answered from the data. Then select a visualization type that matches the insight you need.

With practice, you'll learn to intuitively match the right visualization to your specific data and analytical needs. The key is finding visual representations that make patterns and insights easy to see at a glance.

[Data Science](#)[Data Visualization](#)[Data Analysis](#)[Machine Learning](#)[Statistics](#)



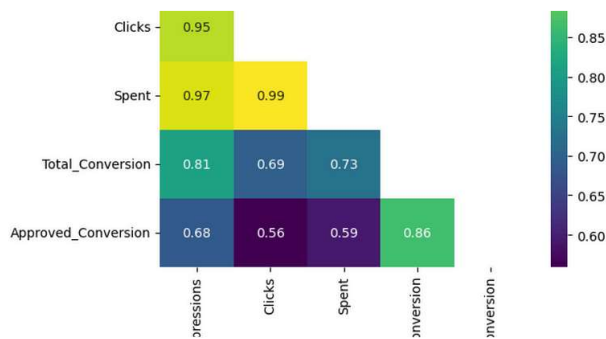
Written by Abdallah Ashraf

347 Followers



Data Analytics 📊 | Data Science | Tech enthusiast | Sharing knowledge

More from Abdallah Ashraf



Abdallah Ashraf

Correlation in machine learning — All you need to know

What is correlation?

7 min read · Sep 22, 2023

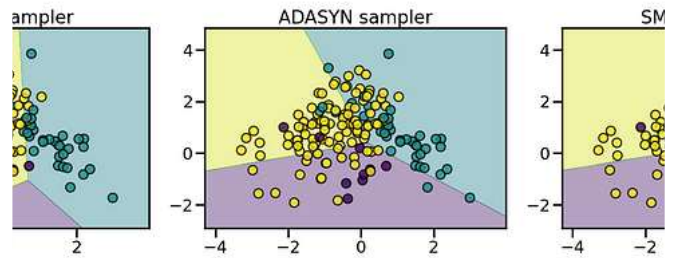
👏 258 💬 4



👏 68 💬 1



Decision function using a LogisticRegression



Abdallah Ashraf

Oversampling — Handling Imbalanced Data

Data is the lifeblood of machine learning, but traditional models struggle when fed...

8 min read · Dec 23, 2023