Learn Something Every Day

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

DATA VISUALIZATION?

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

What is Data Visualization

- Visual display of quantitative information
- Mapping data to visual elements
- Encoding data with size, shape, color...
- Storytelling / narrative elements

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

What is Data Visualization

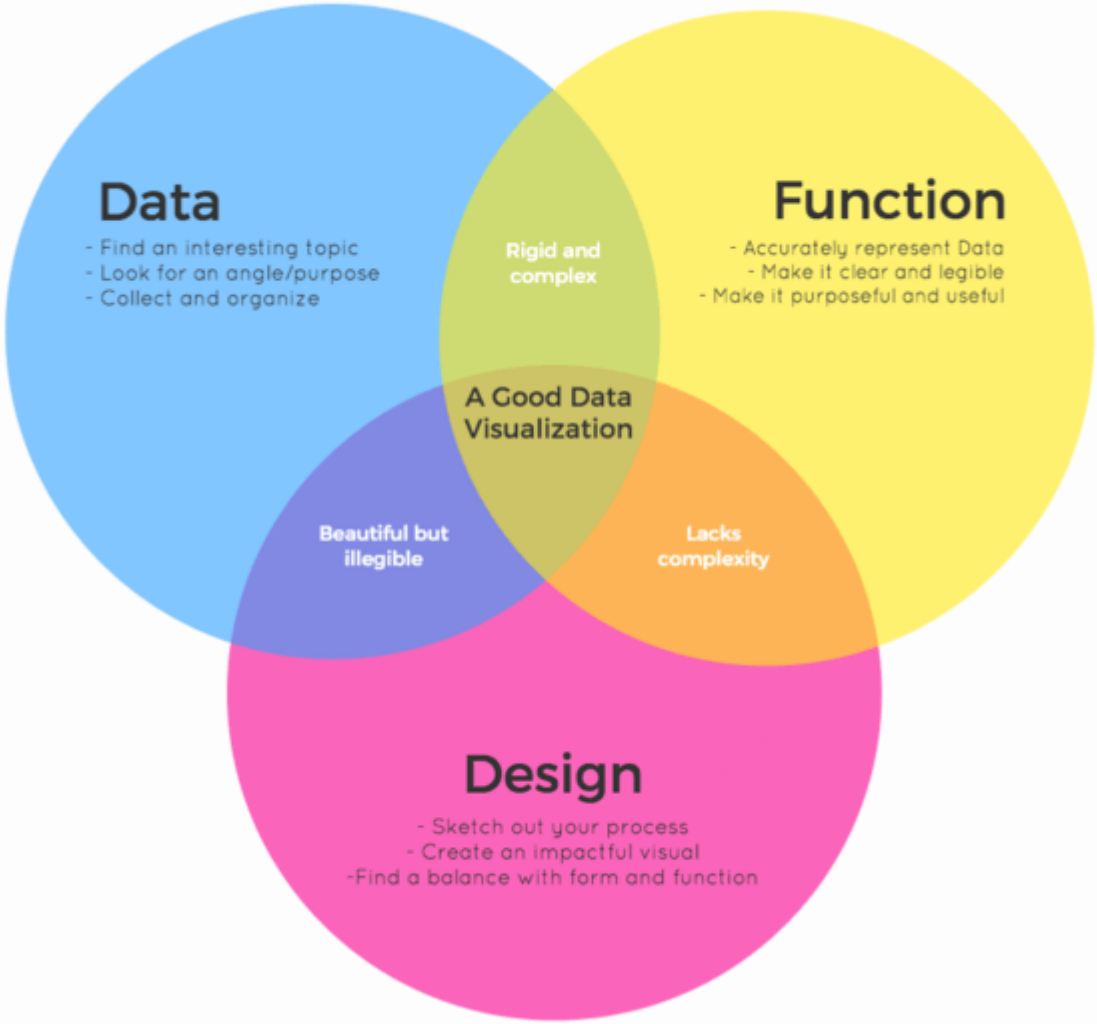
- **Data visualization** is the visual presentation of data or information.
- The goal of data visualization is to communicate data or information clearly and effectively to readers.
- Typically, data is visualized in the form of a chart, infographic, diagram, map and more.

How does it help?

•Identify trends and outliers
•Tell a story within the data
•Reinforce an argument or opinion
•Highlight an important point in a set of data

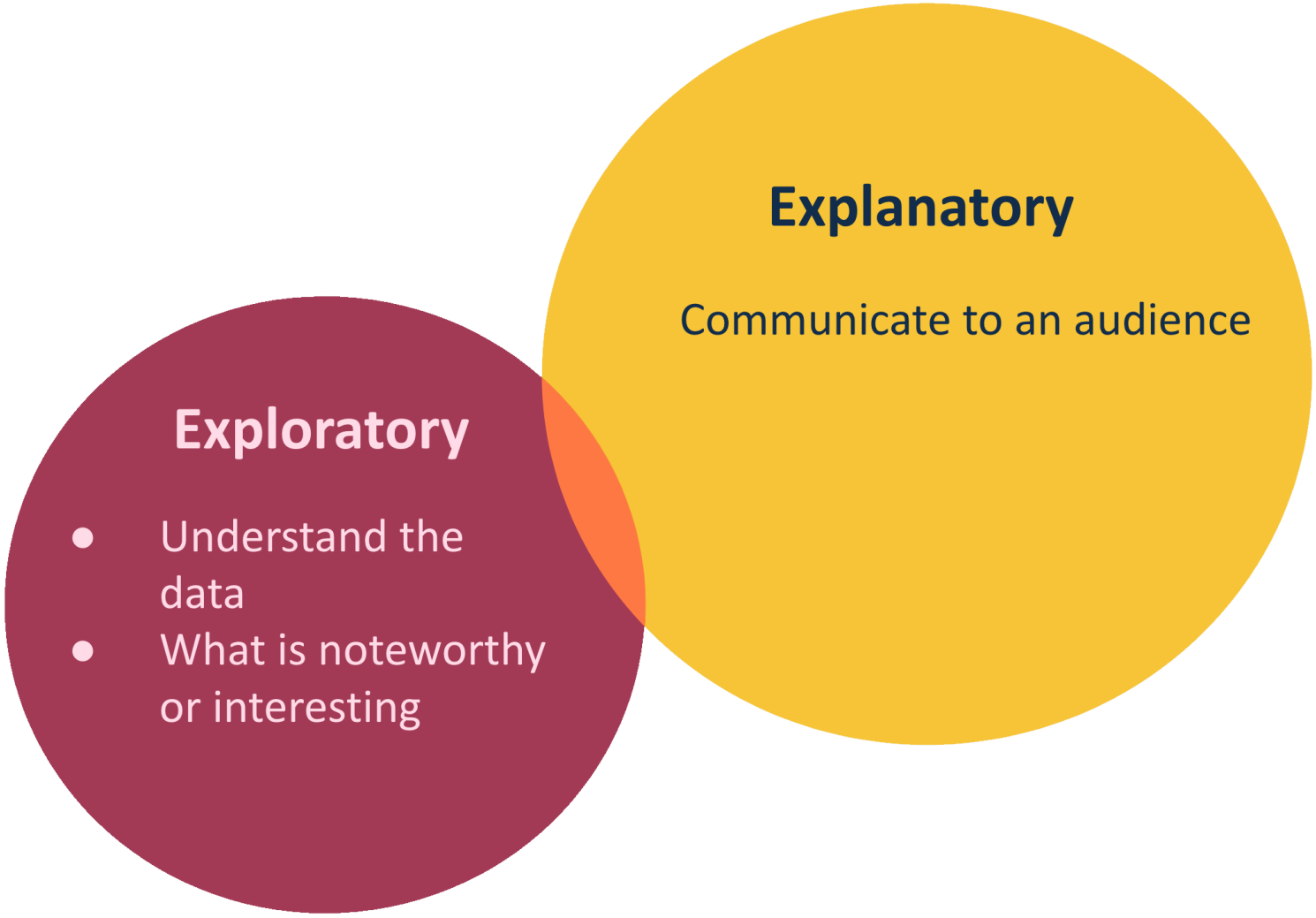
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

What makes good Data Visualization



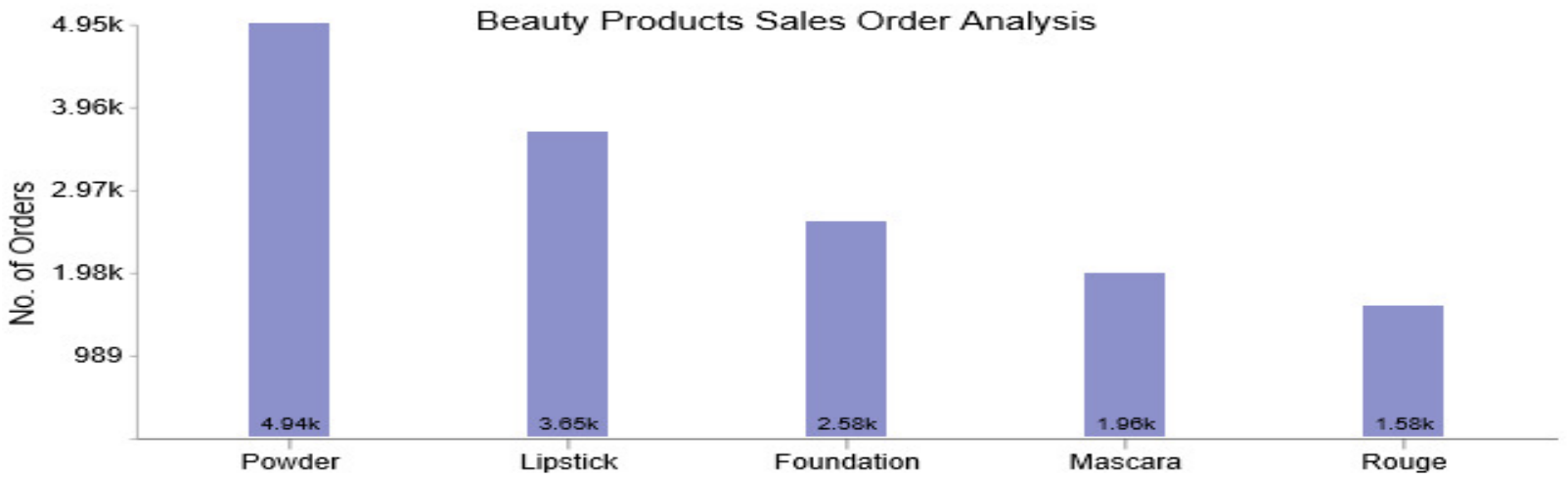
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

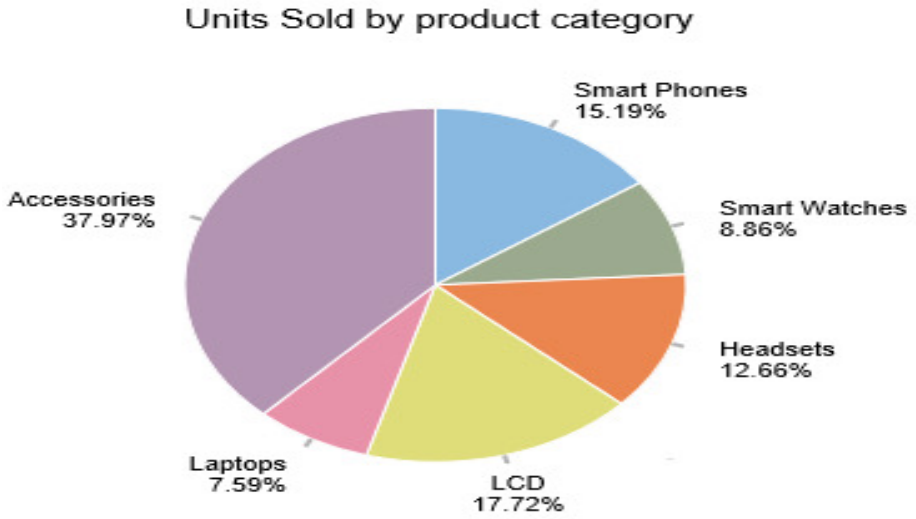
Types of Data Analysis

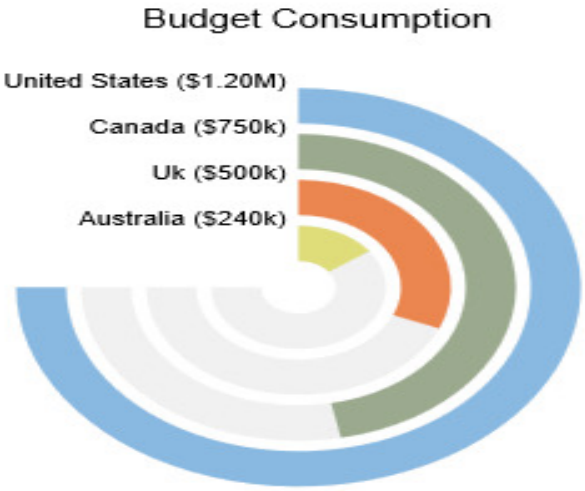


CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Exploratory Data Analysis





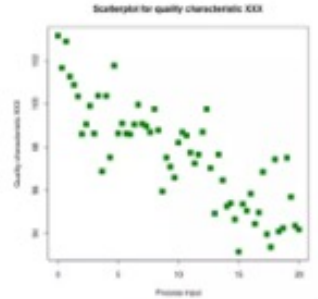


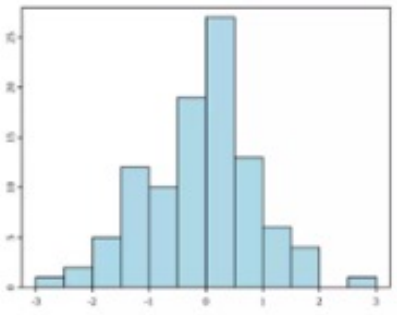
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

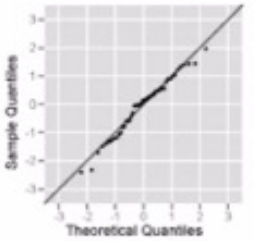
Exploratory data visualization

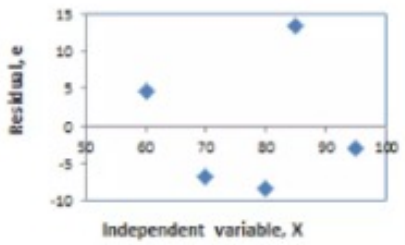
Graphical

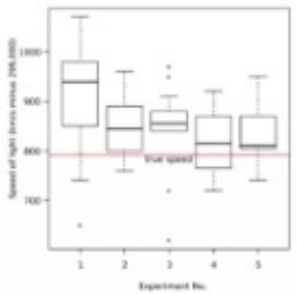
- Scatter plots
- Histograms
- Probability plots
- Residual plots
- Box plots
- Block plots

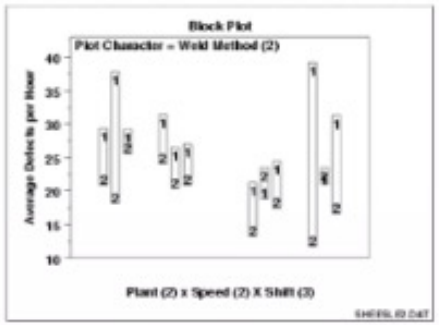












CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Exploratory data visualization

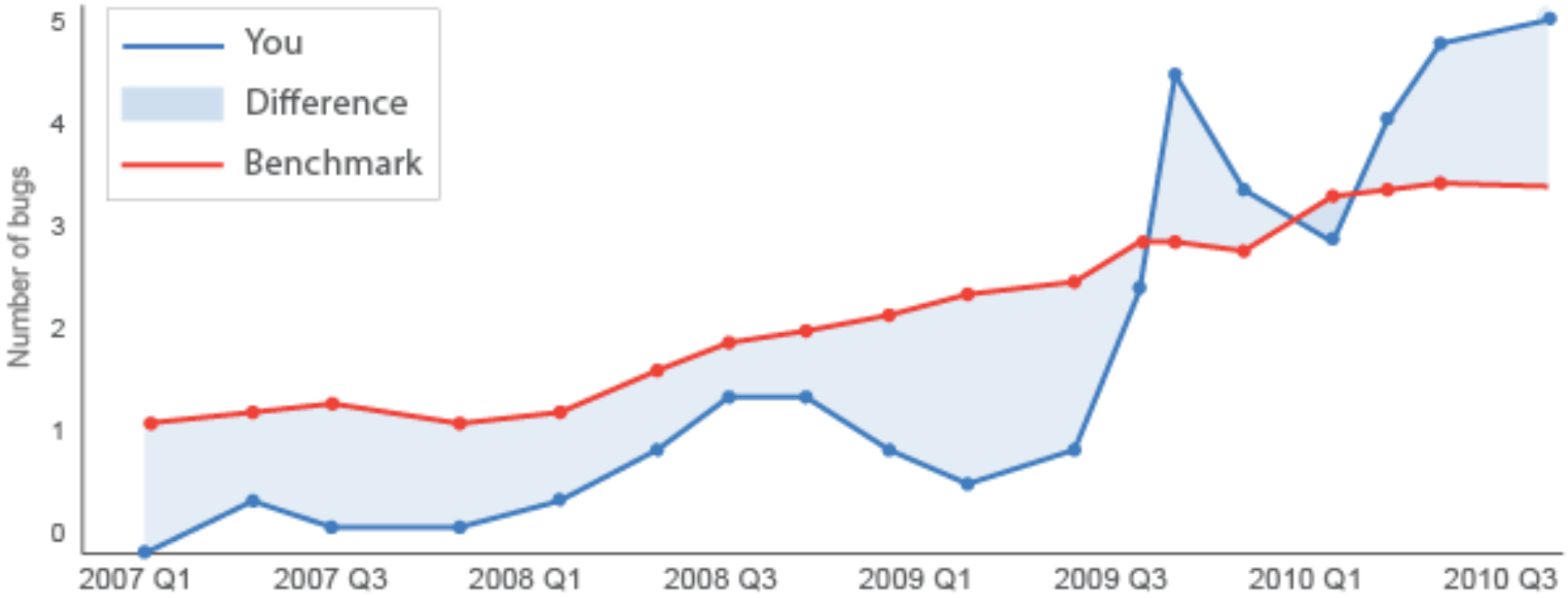
Graphical analysis procedures:

- Testing assumptions
- Model selection
- Model validation
- Estimator selection
- Relationship identification
- Factor effect determination
- Outlier detection

MUST USE for deriving insights from data

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Explanatory Data Analysis



CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Explanatory data visualization

- Design
- Engineering
- Journalism

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Data Visualization in Data Science	Graph	Bar Graph
		Stack Bar Graph
	Plot	Scatter Plot & Line Plot
		Box Plot
		Histogram Plot
	Chart	Pie Chart

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Libraries required for data visualization in Python

1. Matplotlib

Matplotlib is one of the best python data visualization libraries for generating powerful yet simple visualization. It is a 2-D plotting library that can be used in various ways, including Python, iPython sheets, and Jupyter notebooks.

Key Features

- It supports various types of graphical representation, including line graphs, bar graphs, and histograms.
- It can work with the NumPy arrays and border SciPy stack.
- It has a huge amount of plots for understanding trends and making correlations.

Pros And Cons

- Interactive platform
- Versatile library

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Libraries required for data visualization in Python

2. Plotly

The most popular data visualization library in Python is Plotly, which delivers an interactive plot and is easily readable to beginners. It is widely used for handling financial, geographical, statistical, and scientific data.

Key Features

- Its robust API functions effectively in both local and web browser modes.
- It is an interactive, open-source, and high-level visualization library.
- It can be viewed in Jupyter notebooks, standalone HTML files, or even hosted online.

Pros And Cons

- Offers contour plots, dimension chars, and dendrograms.
- Allows 40 unique chart and plot types
- Difficult to use

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Libraries required for data visualization in Python

3. Seaborn

Seaborn is the best python library for data visualization, which offers a variety of visualized patterns. It is designed to work more compatible with Pandas data form and is widely used for statistical visualization.

Key Features

- It performs the necessary mapping and aggregation to form information visuals.
- It is integrated to explore and understand data in a better and more detailed way.
- It offers a high level of a crossing point for creating beautiful and informative algebraic graphics.

Pros And Cons

- Much more visually appealing representation
- Switch to any other data format
- Limited customizable options

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Libraries required for data visualization in Python

4. Ggplot

GGplot is another popular data visualization library in Python, known as the python implementation of graphics grammar. It refers to the map of the data, with its aesthetic attributes including color, shape, and geometric objects like points and bars.

Key Features

- It allows you to build informative visualization substantially with improved representations
- It is integrated with Panda to store data in a data frame.
- It is based on ggplot2, an R programming language plotting system.

Pros And Cons

- Documentation is simple and easy to follow.
- Save method to discuss and exhibit plots
- Not suitable for creating highly customized graphics.

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Libraries required for data visualization in Python

5. Bokeh

Bokeh is another interactive python library for data visualized for modern web browsers. This is best suitable for developing interactive plots and dashboards for complex or streaming data assets.

Key Features

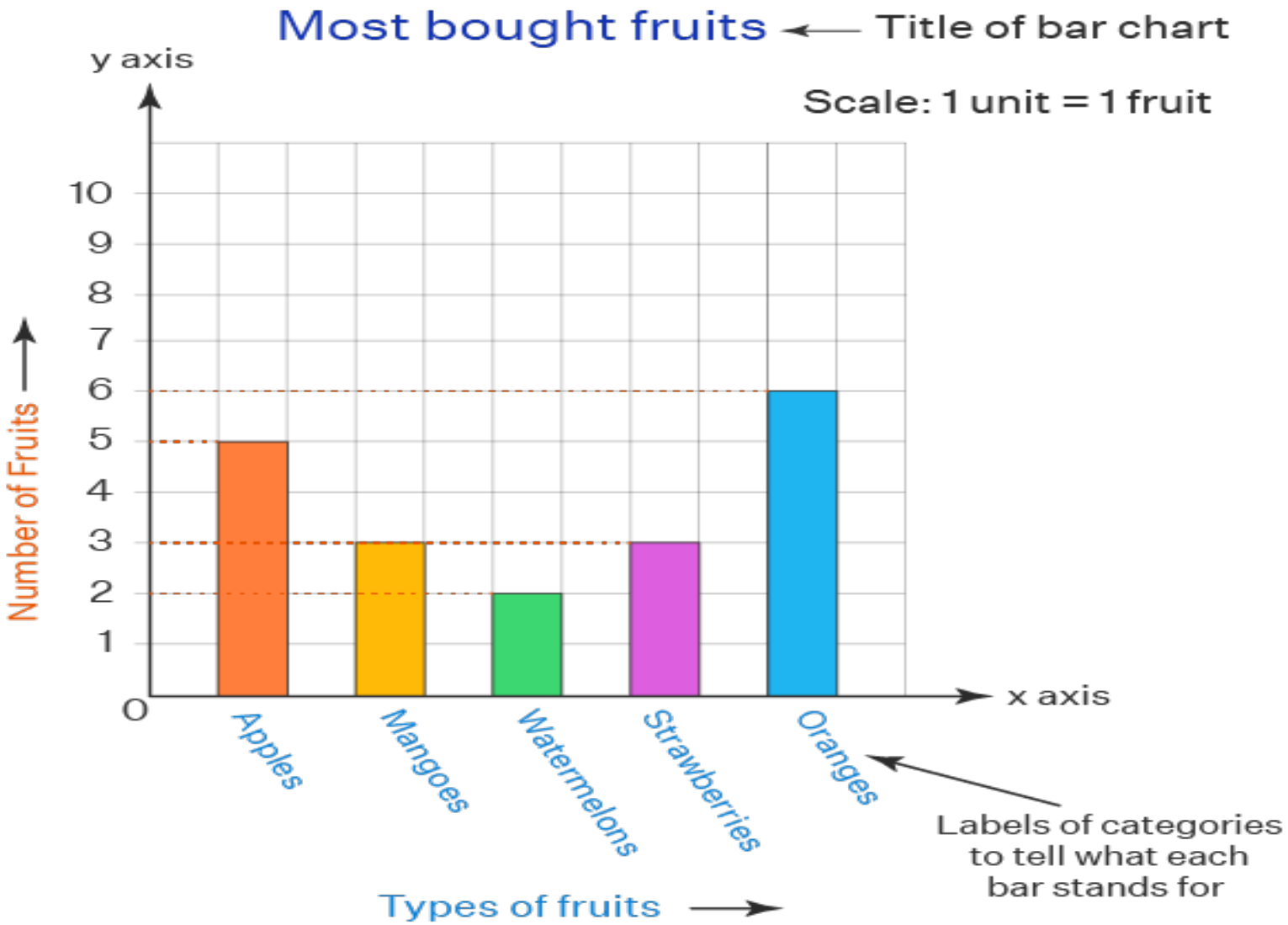
•It has a broad range of intuitive graphs which can be leveraged to form solutions.
•It is well-known for creating custom-made visualizations.
•It includes various generation and plot chart methods, including box plots, bar plots, and histograms.

Pros And Cons

•Highest level of control for the rapid creation of charts
•Many graphs with fewer codes and higher resolution
•No pre-set defaults, and users have to define them each time.

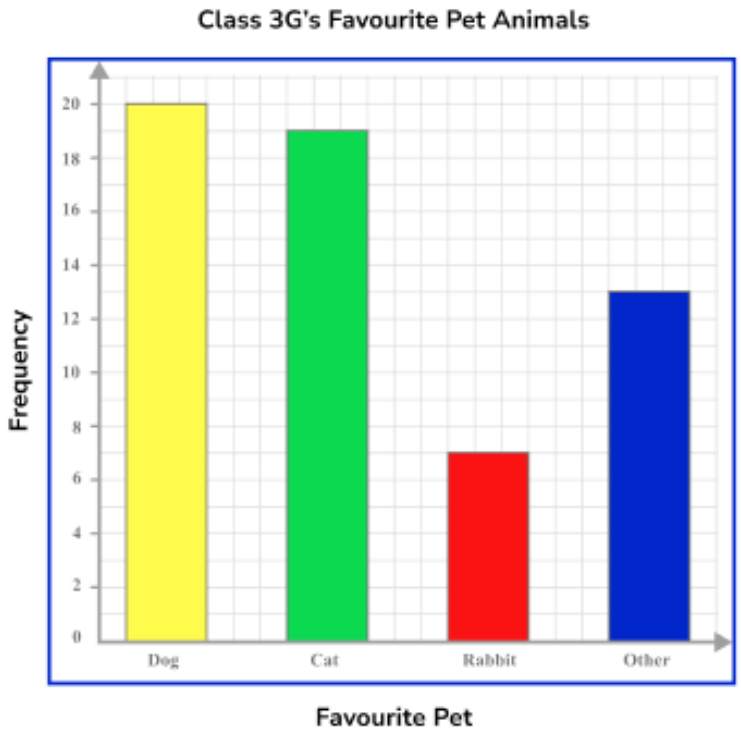
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Bar Graph



CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

A **bar chart** represents a data set by using vertical or horizontal bars. The larger the bar, the higher the value for the individual category.

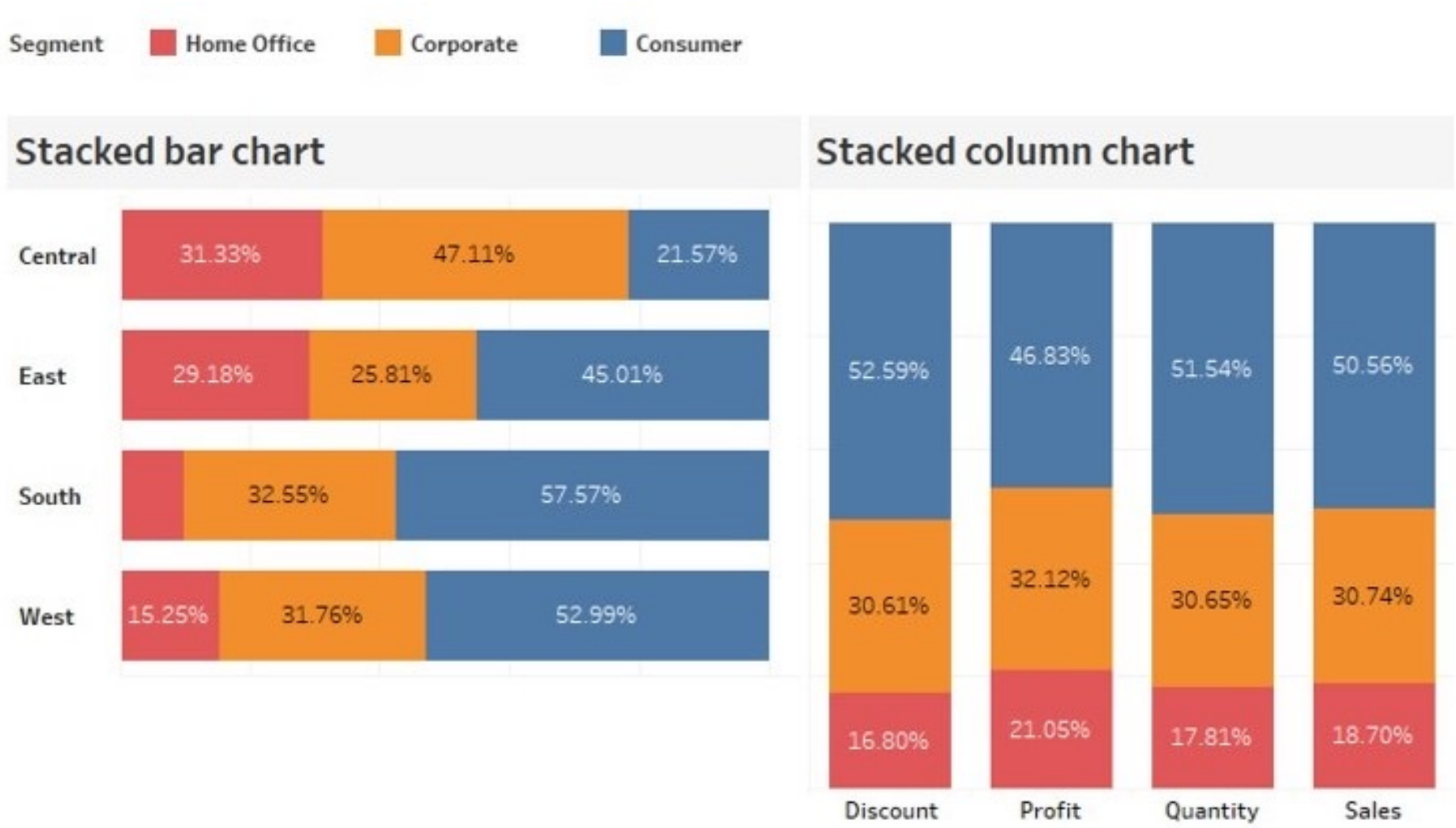


To draw a **bar chart** we need the following:

- A pair of axes. Usually the horizontal axis is labelled with the categories of the data set and the vertical axis is the frequency. Your axes must be labelled.
- The frequencies need to be labelled on the vertical axis in equal intervals.
- The bars need to have equal gaps between them as it is representing discrete data.
- The bars need to be of equal width.
- The chart needs a title.

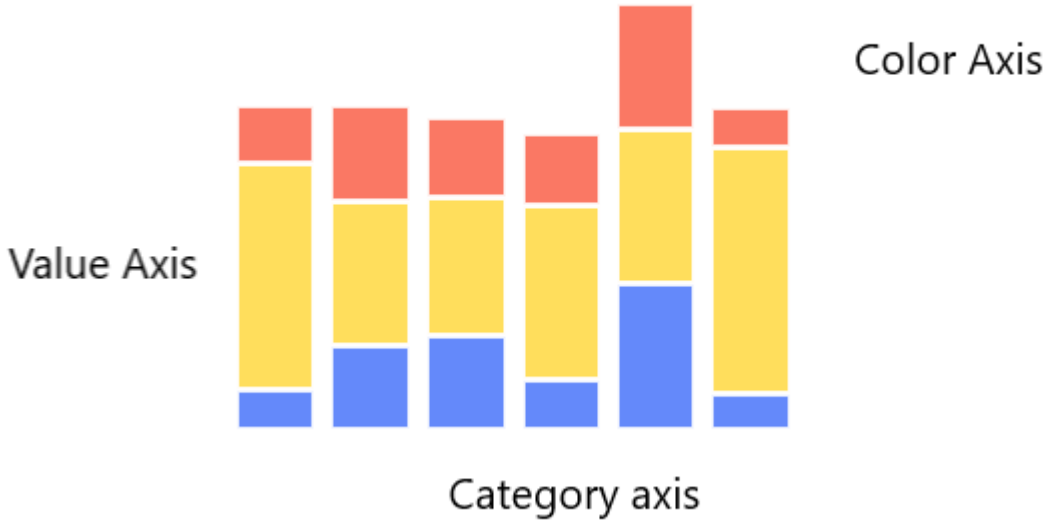
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Stacked Bar Graph



CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

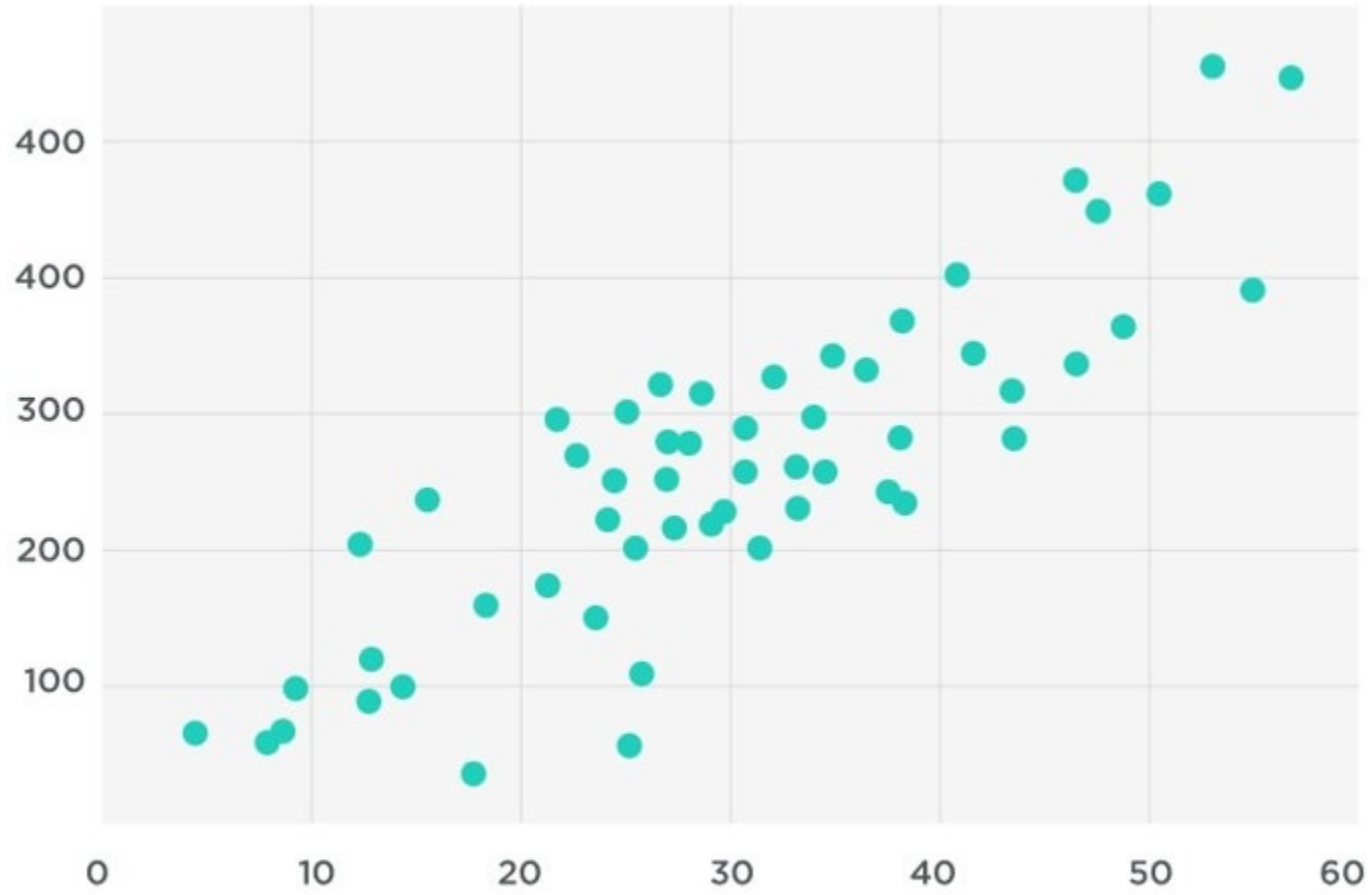
Stack Bar Graph



Stack Bar Graph is a form of bar chart that shows the composition and comparison of a few variables, either relative or absolute, over time. Also called a stacked bar or column chart, they look like a series of columns or bars that are stacked on top of each other.

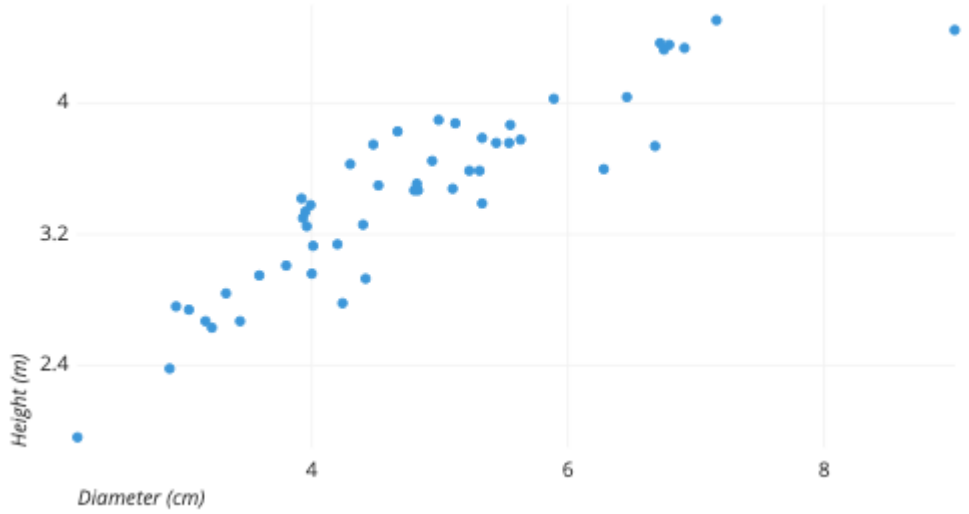
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Scatter Plot



CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

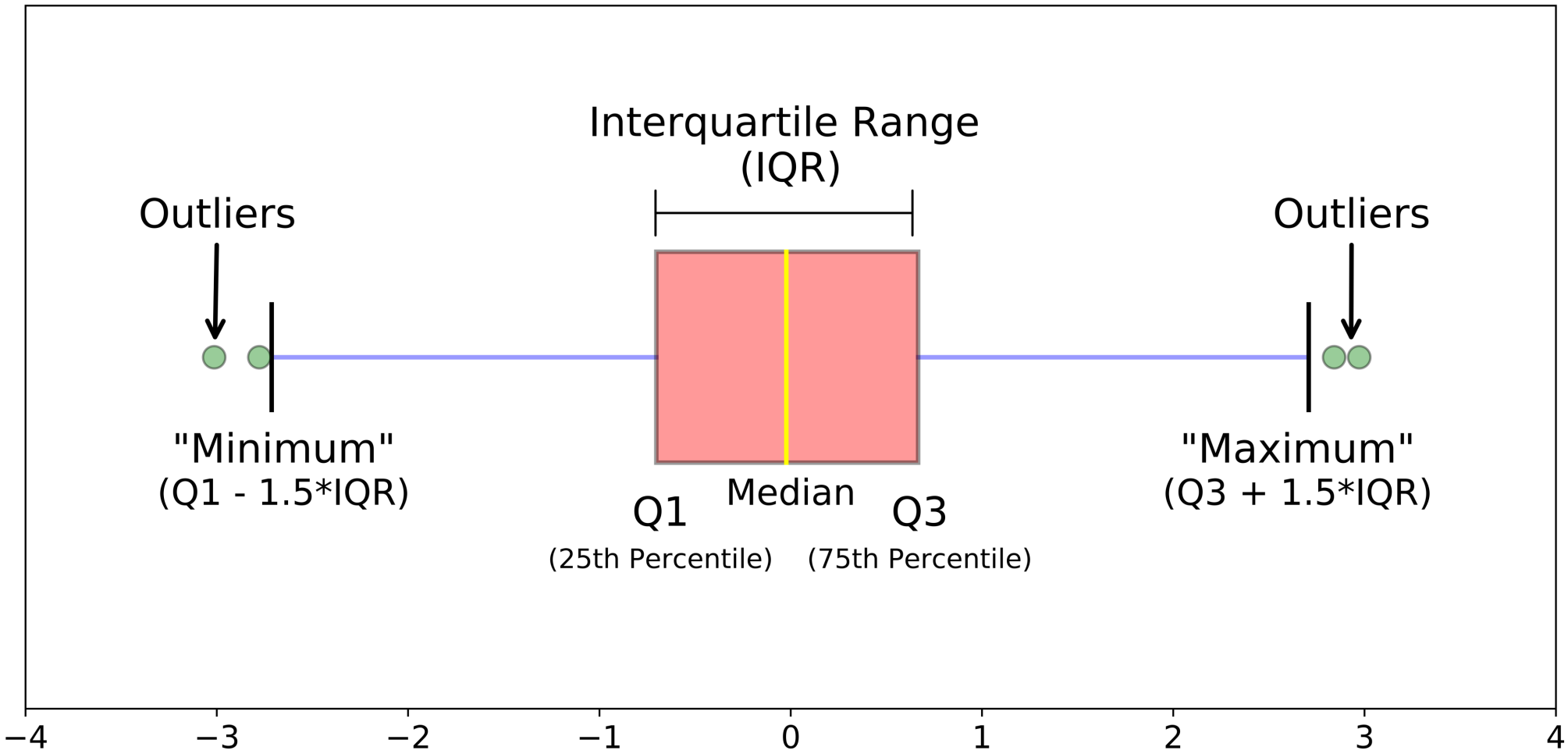
Scatter Plot



A scatter plot uses dots to represent values for two different numeric variables. The position of each dot on the horizontal and vertical axis indicates values for an individual data point. Scatter plots are used to observe relationships between variables.

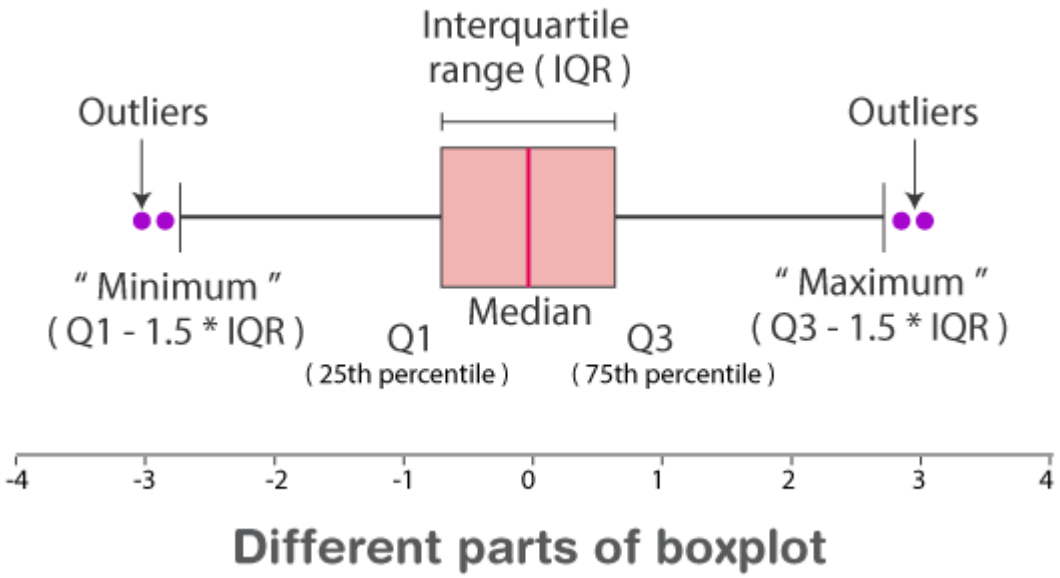
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Box Plot



CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Box Plot



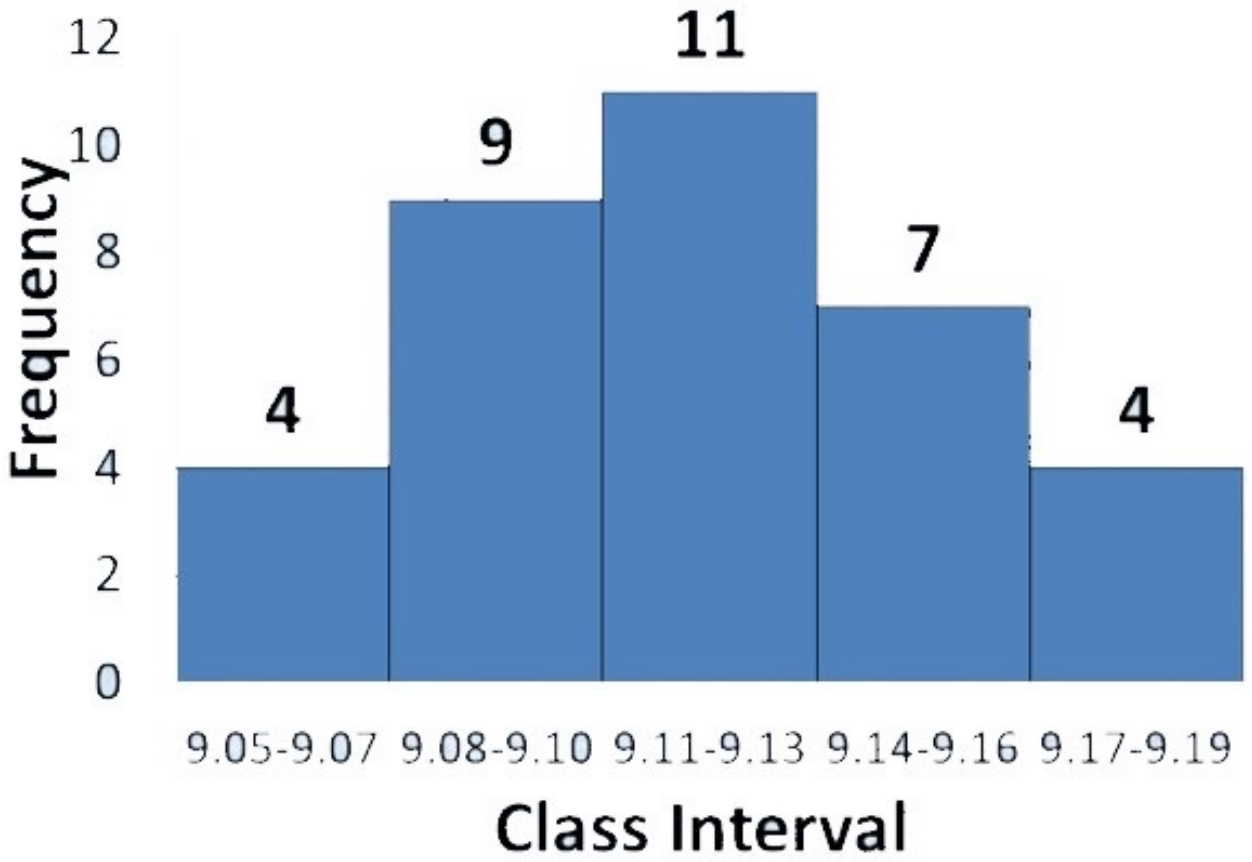
Different parts of boxplot

A box plot is a chart that shows data from a five-number summary including one of the measures of central tendency.
It helps to find out how much the data values vary or spread out with the help of graphs.

Minimum: The minimum value in the given dataset
First Quartile (Q1): The first quartile is the median of the lower half of the data set.
Median: The median is the middle value of the dataset, which divides the given dataset into two equal parts. The median is considered as the second quartile.
Third Quartile (Q3): The third quartile is the median of the upper half of the data.
Maximum: The maximum value in the given dataset.
Apart from these five terms, the other terms used in the box plot are:
Interquartile Range (IQR): The difference between the third quartile and first quartile is known as the interquartile range. (i.e.) IQR = Q3-Q1
Outlier: The data that falls on the far left or right side of the ordered data is tested to be the outliers. Generally, the outliers fall more than the specified distance from the first and third quartile.
(i.e.) Outliers are greater than Q3+(1.5 . IQR) or less than Q1-(1.5 . IQR).

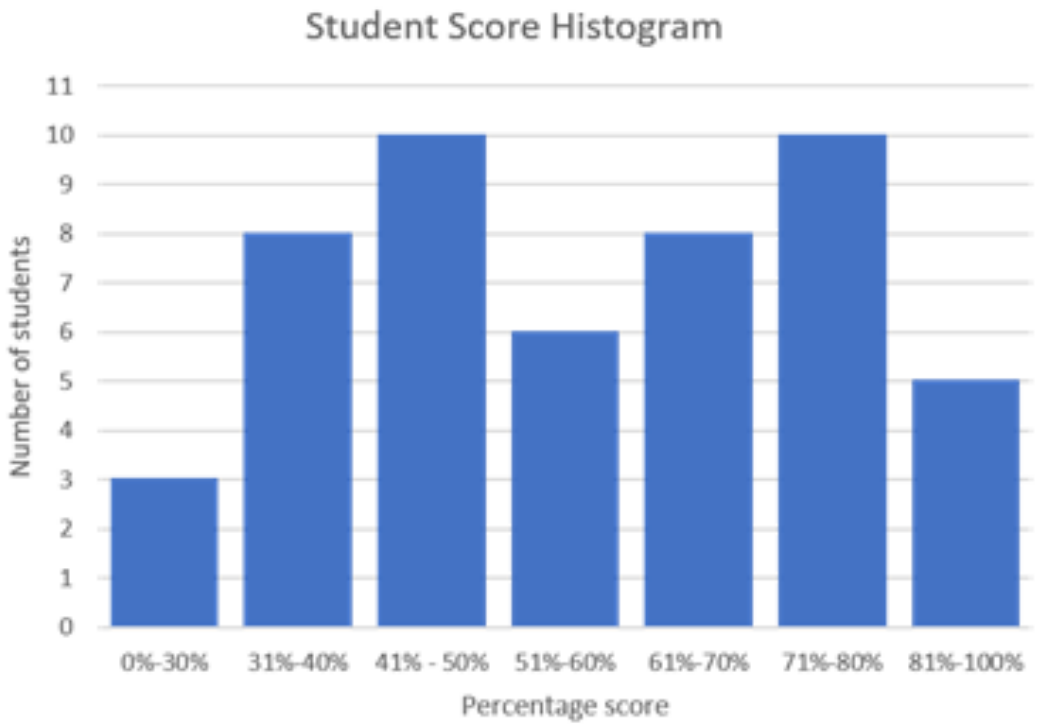
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Histogram



CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

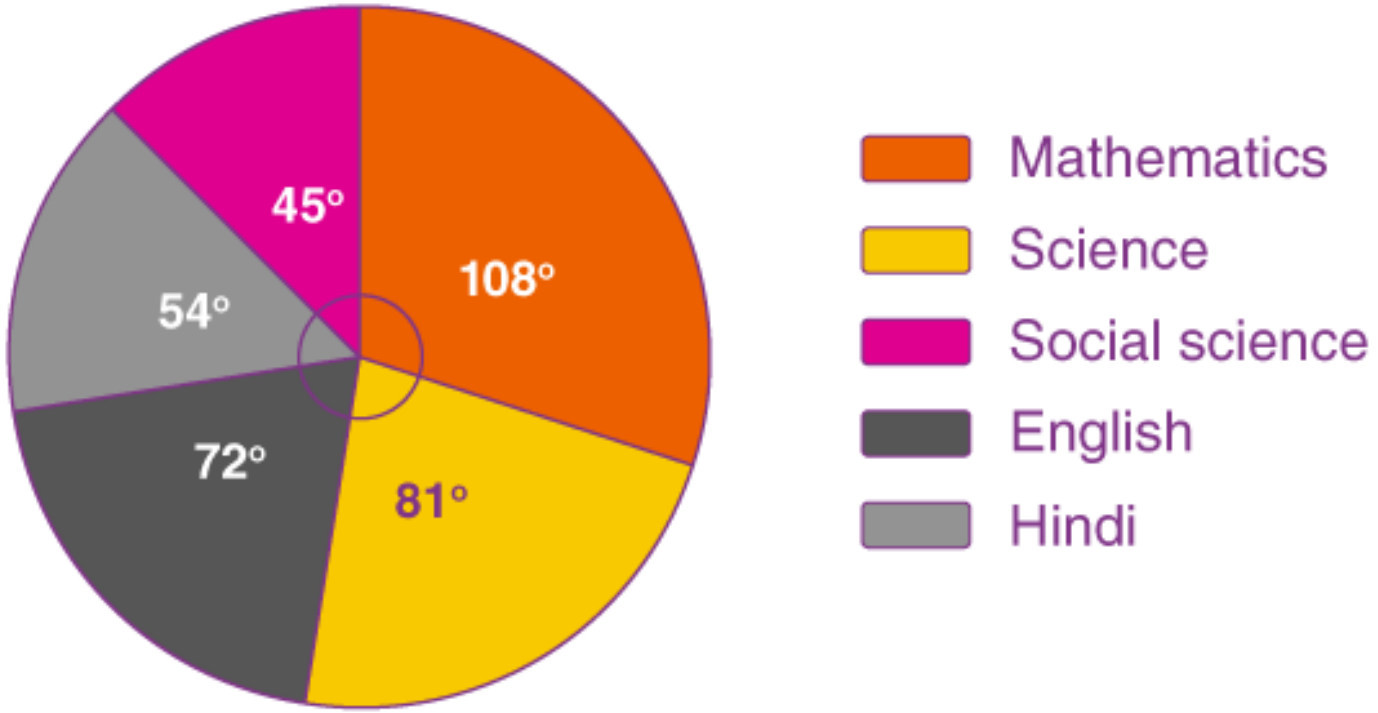
Histogram



A histogram is a graph used to represent the frequency distribution of a few data points of one variable. Histograms often classify data into various “bins” or “range groups” and count how many data points belong to each of those bins.

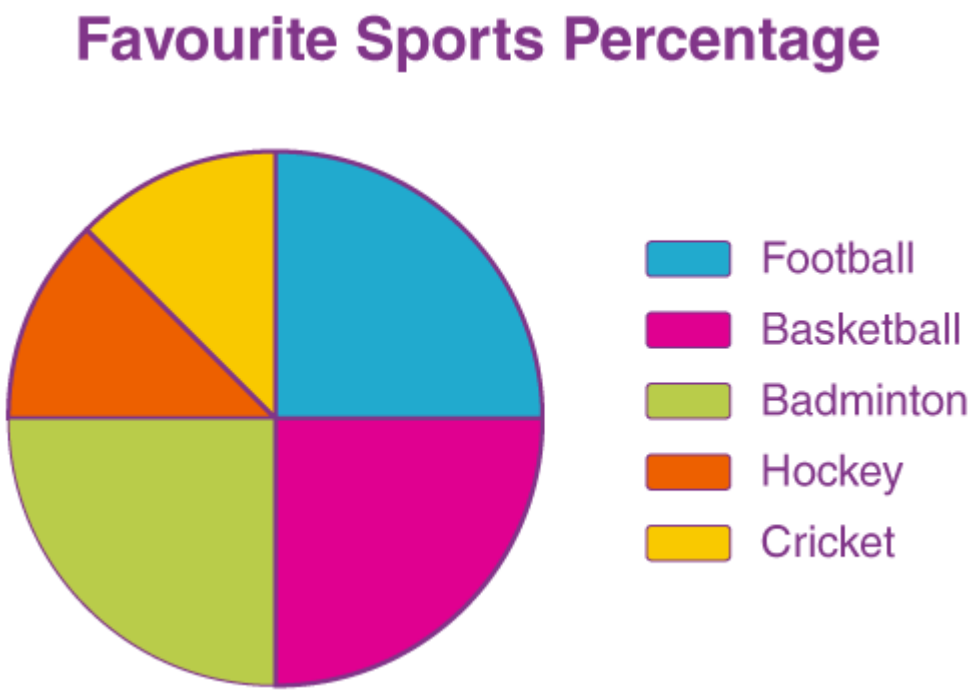
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Pie Chart



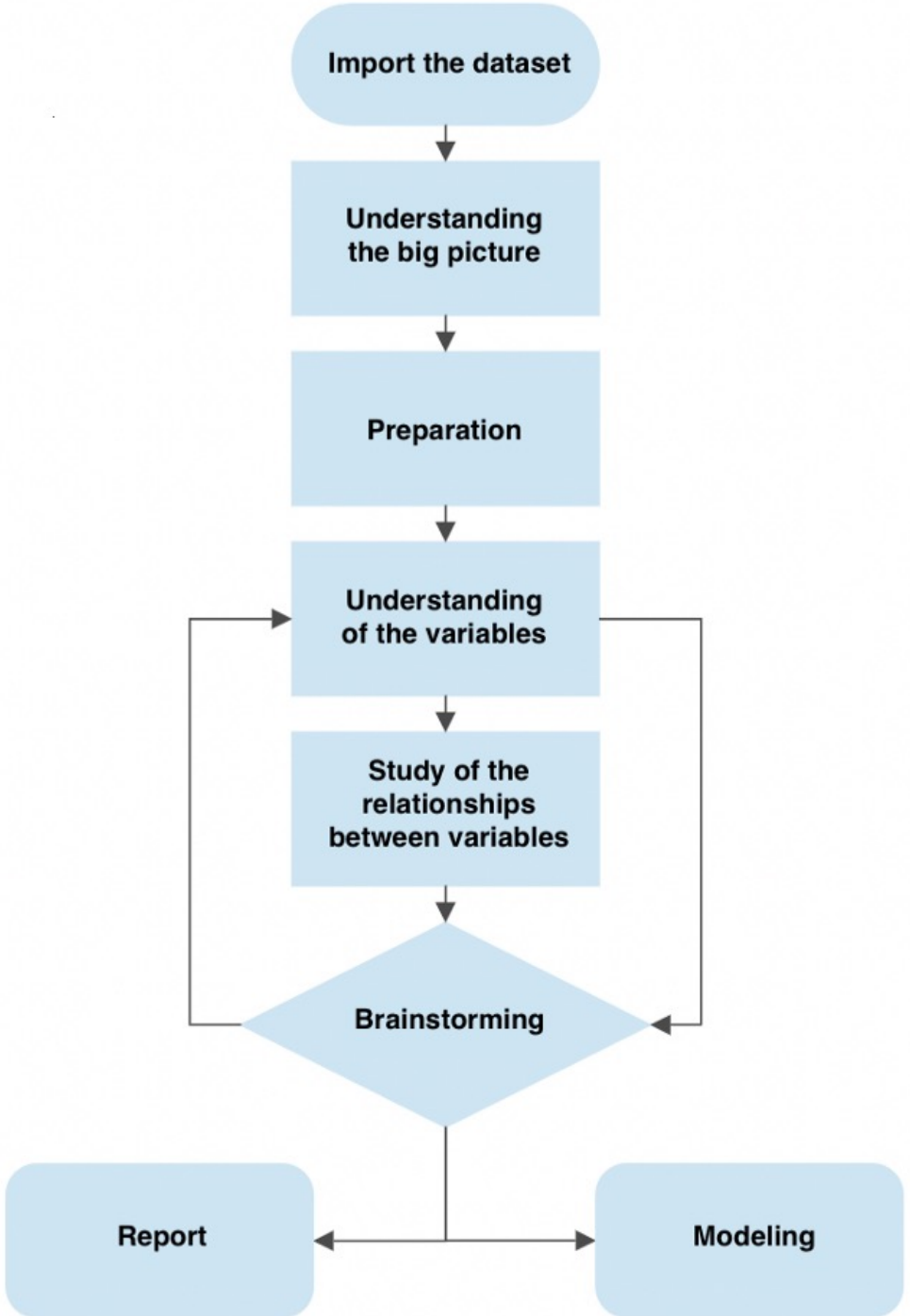
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

Pie Chart



A pie chart is a type of graph that represents the data in the circular graph. The slices of pie show the relative size of the data, and it is a type of pictorial representation of data. A pie chart requires a list of categorical variables and numerical variables. Here, the term “pie” represents the whole, and the “slices” represent the parts of the whole.

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies



_ & PROPRIETARY 2022 – Adverk Technologies

THANK YOU!

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies