Data Visualisation

Table of contents
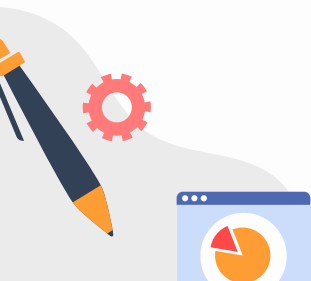
LO.1

Preparing the working environment...

Data type ,file type ,program , language

- Integer -------> Data type
- String --------> Data type
- Xlsx --------> file type
- Json-------->file type
- Csv --------->file type
- PostgreSQL----->program
- SSMS ---------->program
- SSIS ------------>program
- Power bi --------->program
- DAX-------------> language
- Power Query ----->language
- SQL ----------->language

Data Types

A) **Qualitative(Categorical):**

B) **Quantitative(Numerical):**

Data Types

A) Qualitative(Categorical):

1- a- Binary data: is the type of qualitative data that has one of two mutually exclusive values. ex: person might be male or female

2- Nominal: describes an object without assigning a numerical value to it. ex: T-shirt can be described as red or blue

3- Ordinal data: describes an object assigning a numerical value to it. ex: clothes sizes (Small, large, X large).

Data Types

B) Quantitative(Numerical):

1- Continuous (decimal values)
Ex: height, speed, time

2- Discrete (integer-valued)
Ex: number of trees

What is data visualization?

is the <u>representation of data</u> through use of common graphics, such as charts, plots and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is <u>easy to understand</u>.

Data visualization benefits:

1- Improved communication:

Can assist in **displaying complex data in an easily accessible way**, allowing stakeholders, decision-makers and analysts to communicate more effectively.

2- Improved comprehension:

Visualization can help users **spot patterns and trends** in data that might otherwise be **difficult to identify with raw data alone.**

Data visualization benefits:

3- Improved decision-making:

Data **presented in a clear and succinct manner** can aid in better decision-making.

4- Improved efficiency:

Visualizing data, analysts and decision-making can **immediately discover crucial information without having to trawl through vast amounts of data**. This can help you save time and be more efficient.

5- Increased engagement:

Can be more engaging than typical tables and charts, increasing curiosity and attention to data.
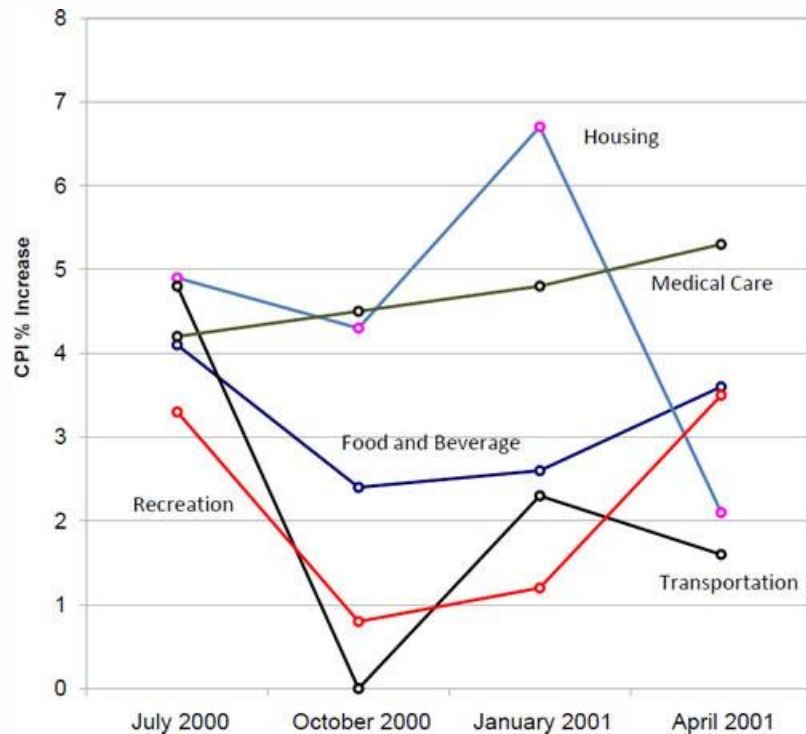
Types
of
Data
Visualization

-There are many types of data visualizations, and the best one to use will **depend on the specific data you are working with** and the message you want to convey.

Data Visualization Types

1. **Line Graph :**

- how **data changes over time**.

- Like a bar chart, it contains an 'x' and a 'y' axis; however, in a line chart, both axes will represent numerical values.

- Line charts can feature only one **line**, or can display multiple categories by plotting several lines.
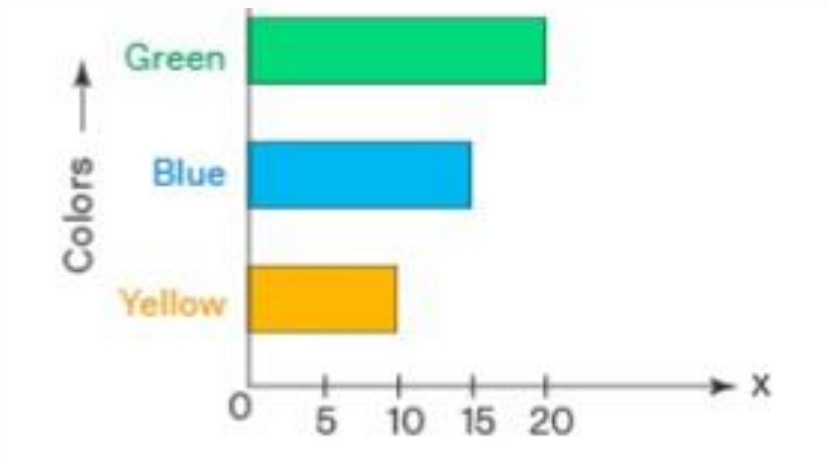
- **Ex. Profit over Time**

Data Visualization Types

2- Bar chart :

 A bar chart is a representation
of numerical data in pictorial
form of Horizontally rectangles.
 the categorical data (y-axis)
the quantity data(x-axis)

The bar and column
charts are used to show the comparisons
between different variables. It is very easy and convenient to
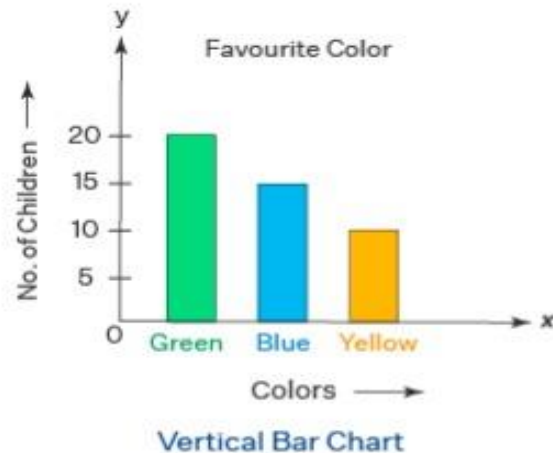visualize the parameters in a pictorial form.

Data Visualization Types

3- Column chart

column : is a representation of numerical data in pictorial form of Vertically rectangles.
 the categorical data (x-axis)
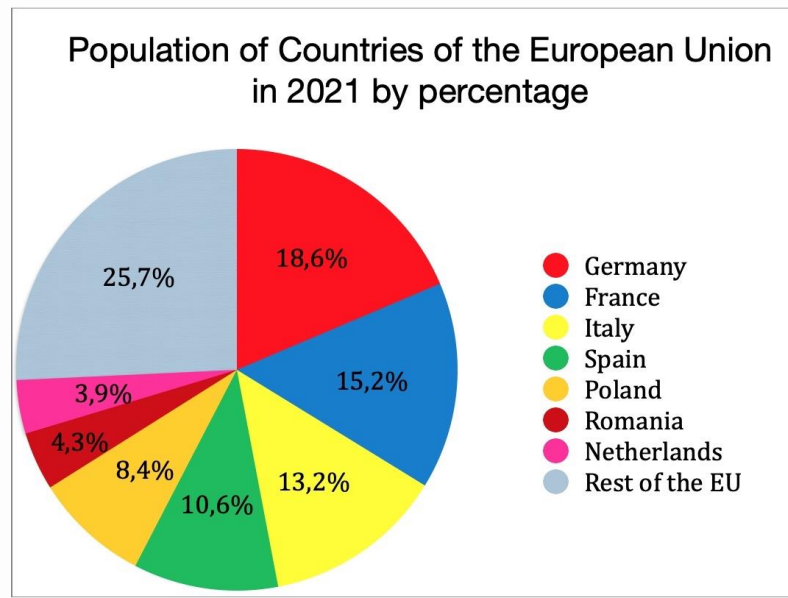the quantity data(y-axis)

Data Visualization Types

4- pie chart

- Representing data in a circular form, with each slice of the circle representing a fraction part of the whole

- The whole equaling 100 percent and 360 degrees

Population of Countries of the European Union in 2021 by percentage

- 🔴 Germany
- 🔵 France
- 🟡 Italy
- 🟢 Spain
- 🟠 Poland
- 🔴 Romania
- 🟣 Netherlands
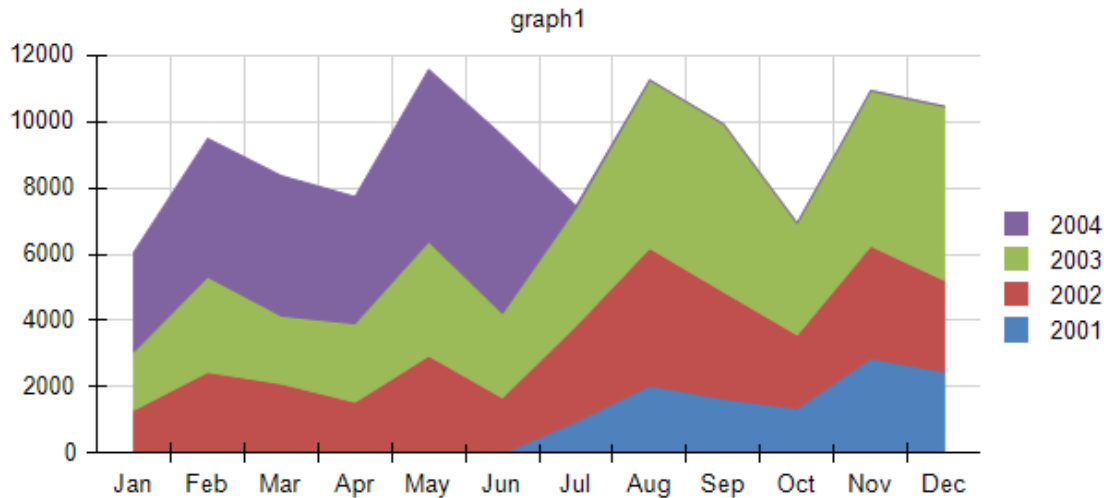- ⚪ Rest of the EU

Data Visualization Types

5- **stacked Area chart**

- Visual representation of data that utilizes both lines and filled areas to convey information

graph1

Data Visualization Types

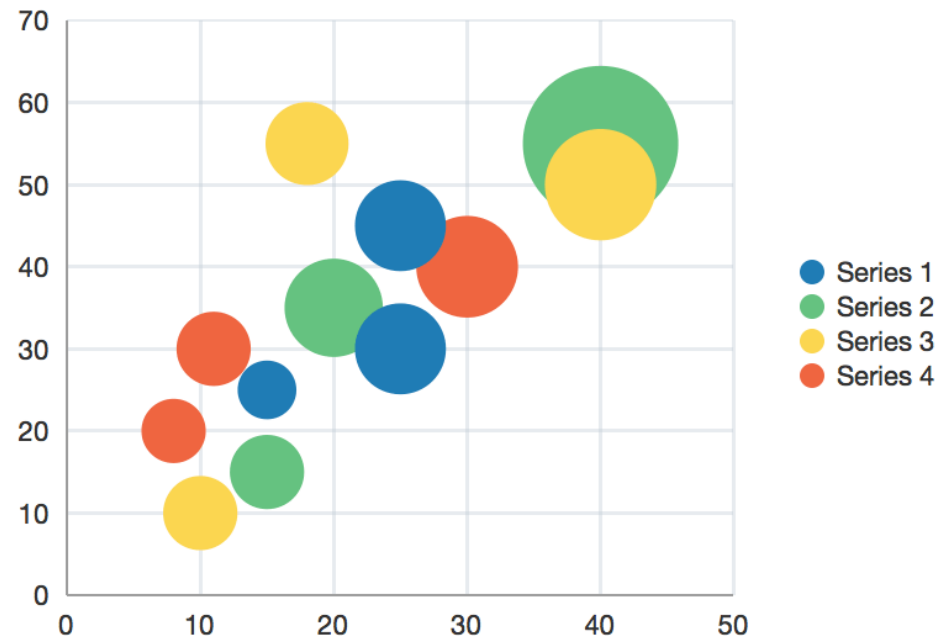6- **Bubble chart**

- An extension of the scatter plot used to look at relationships between three numeric variables.

- Each dot in a bubble chart corresponds with a single data, and values for each point are indicated by horizontal position, vertical position, dot size.

Bubble Chart

Data Visualization Types

7- **spider chart (radar)**

- To display data across several unique dimensions.
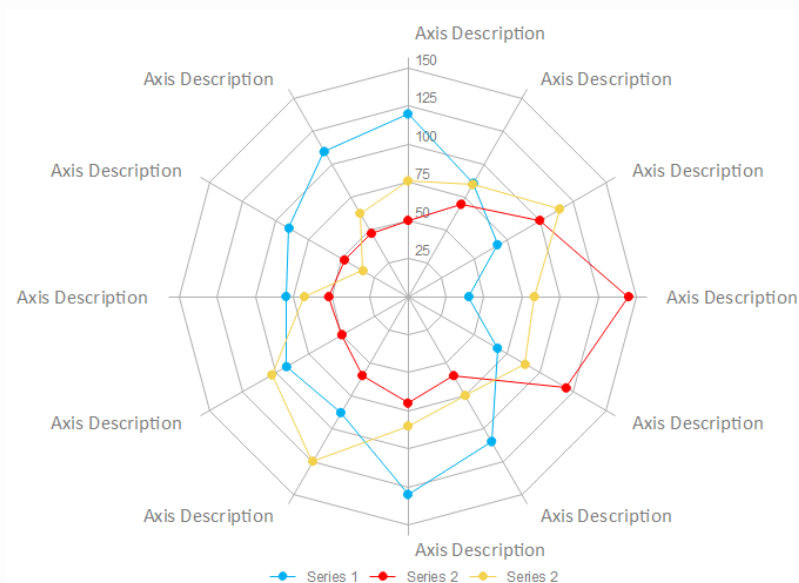
Data Visualization Types

8- **gauge chart**

- Give you a quickly see how well a given metric is performing against a target goal

Actual Sales vs Target

$0 $500k

Actual Sales vs Target
$355k

Data Visualization Types

9- histogram

•Histograms are useful when dealing with continuous data or when data points are limited to a numerical range.

•For example, temperature readings, time measurements, and weight measurements are all types of continuous data.

•Ex. Employees Age
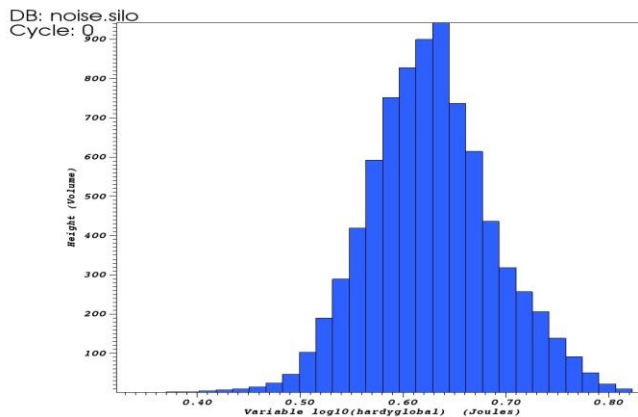
Data Visualization Types

10- **A scatter plot matrix is a grid (or matrix) of scatter plots used to visualize bivariate relationships between combinations of variables. Each scatter plot in the matrix visualizes the relationship between a pair of variables, allowing many relationships to be explored in one chart.**
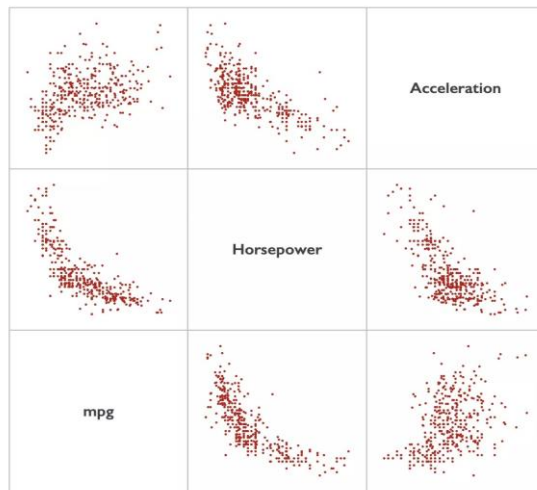
Data Visualization Types

11- step chart

- A stepped line graph (also called step chart) is a chart similar to a line graph, but with the line forming a series of steps between data points. A stepped line chart can be useful when you want to show the changes that occur at irregular intervals. For example, price rise in milk products.

Line Chart   Step Chart
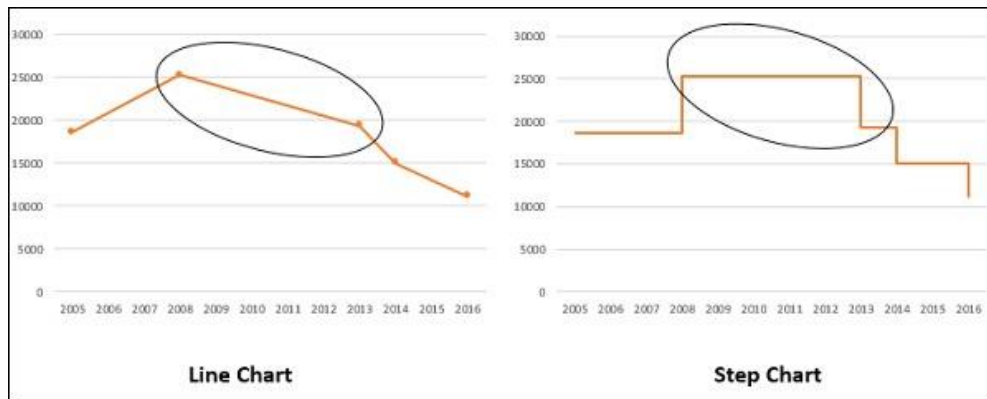
Data Visualization Types

12- **stacked bar chart**

- Type of bar chart that portrays the comparisons of several variables through time

- **Ex. Production of multiple products in each quarter**

- **Ex. Productivity of each product relation with working hours**
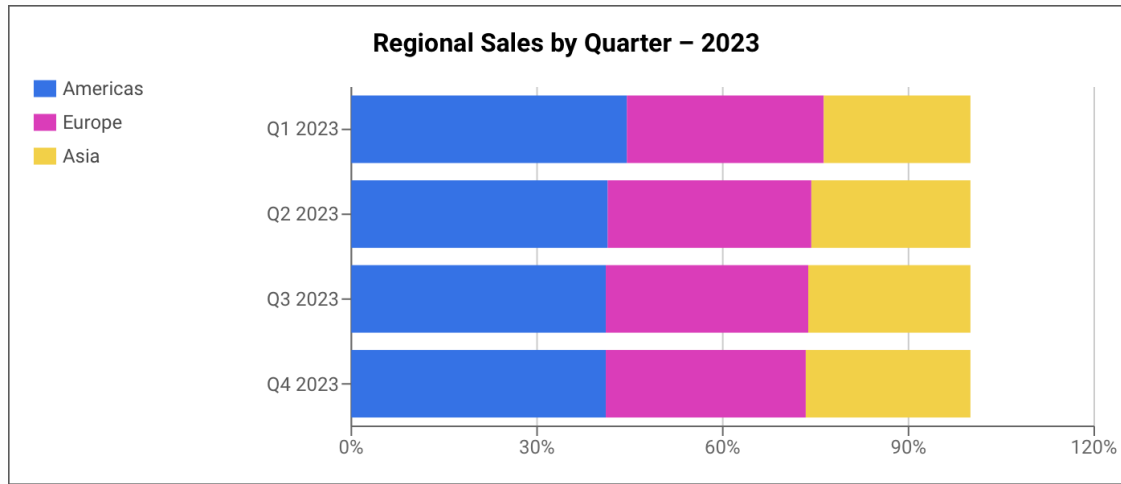
Data Visualization Types

13- **comparison chart**

- Provides a visual representation of different choices or options

	MY PRODUCT	PRODUCT 1	PRODUCT 2	PRODUCT 3	PRODUCT 4
Feature 1	✓	✓	✓	✓	✓
Feature 2	✓	✓		✓	✓
Feature 3	✓	✓	✓		
Feature 4		✓	✓	✓	✓
Price	$80 monthly	$80 monthly	$500 one-time	$70 monthly	$50 per user
Cust. Satisfaction	⭐⭐⭐⭐⭐	⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐	⭐⭐⭐⭐⭐

Data Visualization Types

 14- **Donut chart**

similar to a pie chart but with a hole in the center.

- **Representation:** It displays **data in slices**, each representing a category or proportion of the whole.

- **Color Coding:** Different colors are often used for each slice to visually differentiate and enhance readability.

- **Usage:** Donuts are useful for illustrating the contributions of various categories to a whole, but some argue they might **be less effective than other chart types for accurate**

Data Visualization Types

15- **Donut chart**

EX:

Art Print - $13,126, making up about 29% of the total revenue.

Wall Tapestry - $15,023, which is approximately 33% of the total revenue.

iPhone Case - $9,918, contributing around 22% to the total revenue.

Throw Pillow - $5,256, accounting for about 12% of the total revenue.

Mug - $2,103 which is roughly 4% of the total revenue.
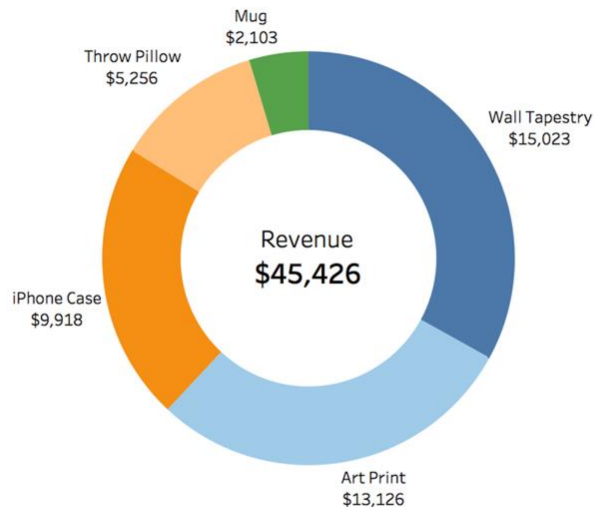
Data Visualization Types

16- Scatter plot chart

represent even groups or binaries of data, according to two coordinate axes, **vertical and horizonta**l

- The stronger the connection, the weaker the dispersion and the points are clustered closer to the line or curve.

- The **straight line** expresses the **strongest** relationship between the data.

- The relationships can either be positive or negative, non-linear or linear, and/or, strong or weak

- **Ex. Profit relation with units produced**
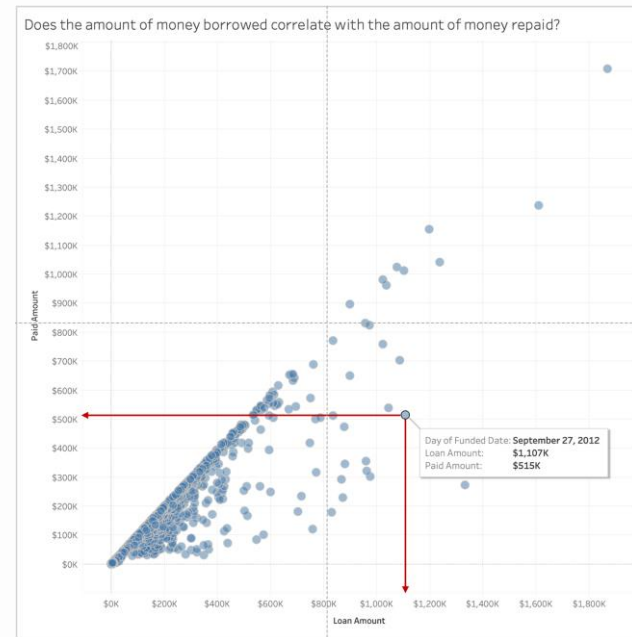
Data Visualization Types

16- dot plot:

- **Dot plots represent data points along a single axis, usually with one dot per data point.**

Number of Elective Courses Taken by Mr. Smith's Statistics Students
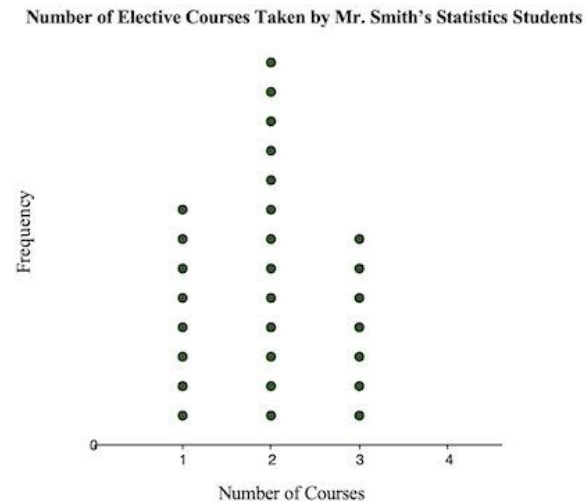
Classification of chart types

1-Comparison Charts:

1. Bar Chart
2. Column Chart
3. Line Chart
4. Area Chart

2-Distribution Charts:

1. Histogram
2. Box Plot (Box-and-Whisker Plot)
3. Dot Plot

Classification of chart types

3-Relationship Charts:

1. Scatter Plot
2. Bubble Chart
3. Radar Chart
4. Scatter Plot Matrix

4-Composition Charts:

1. Pie Chart
2. Donut Chart
3. Treemap

Classification of chart types

5-Statistical Charts:

1. Box Plot
2. Violin Plot
3. Error Bars

6-Time-Series Charts

1. Line Chart
2. Area Chart
3. Candlestick Chart

Data Visualization Steps
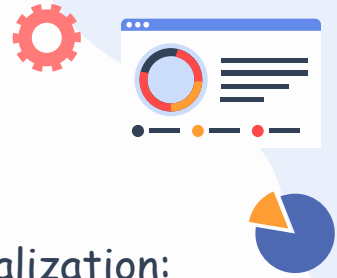
Here are some general steps you can follow when creating a data visualization:

I.   Identify the goal of your visualization:

II. Collect and prepare the data:

III. Choose the appropriate visual representation:

IV. Create the visualization:

V.   Customize the visualization.

VI. Interpret and analyze the visualization:

VII. Share the visualization:

I. Identify the goal of your visualization:

Identifying the goal of your data visualization is an <u>important first step</u>, as it will help guide the rest of the process and <u>ensure that the visualization effectively communicates the message you want to convey.</u>

EX:
- What is the purpose of the visualization? Are you trying to understand the data, communicate information to others, or both?

- Who is the audience for the visualization? What do they already know about the data, and what do they need to know?

II. Collect and prepare the data:

-Data Formats:
(CSV, TXT, Microsoft Excel format, Markup languages(HTML, XML), JSON)

-Data Sources:
• **Structured:**
CSV & Excel files (tables)

• **Semi-structured:**
JSON & XML formats

•**Unstructured:**
Images, videos

III. Choosing the Appropriate Visual Representation:

- **The type of data:**
For example, continuous numerical data is often best represented with line charts or scatter plots, while categorical data is often best represented with bar charts or pie charts.

- **The number of variables:**
If you are trying to visualize multiple variables, you may need to use a more complex visualization, such as a scatter plot or a multi-axis chart.

- **The message you want to convey:**
For example, a line chart can show trends over time, while a bar chart can show comparisons between categories.

- **The audience:** Consider the audience for your visualization and choose a visual representation that will be clear and easy for them to understand.

Guided Visualizations for Charts and Graphs
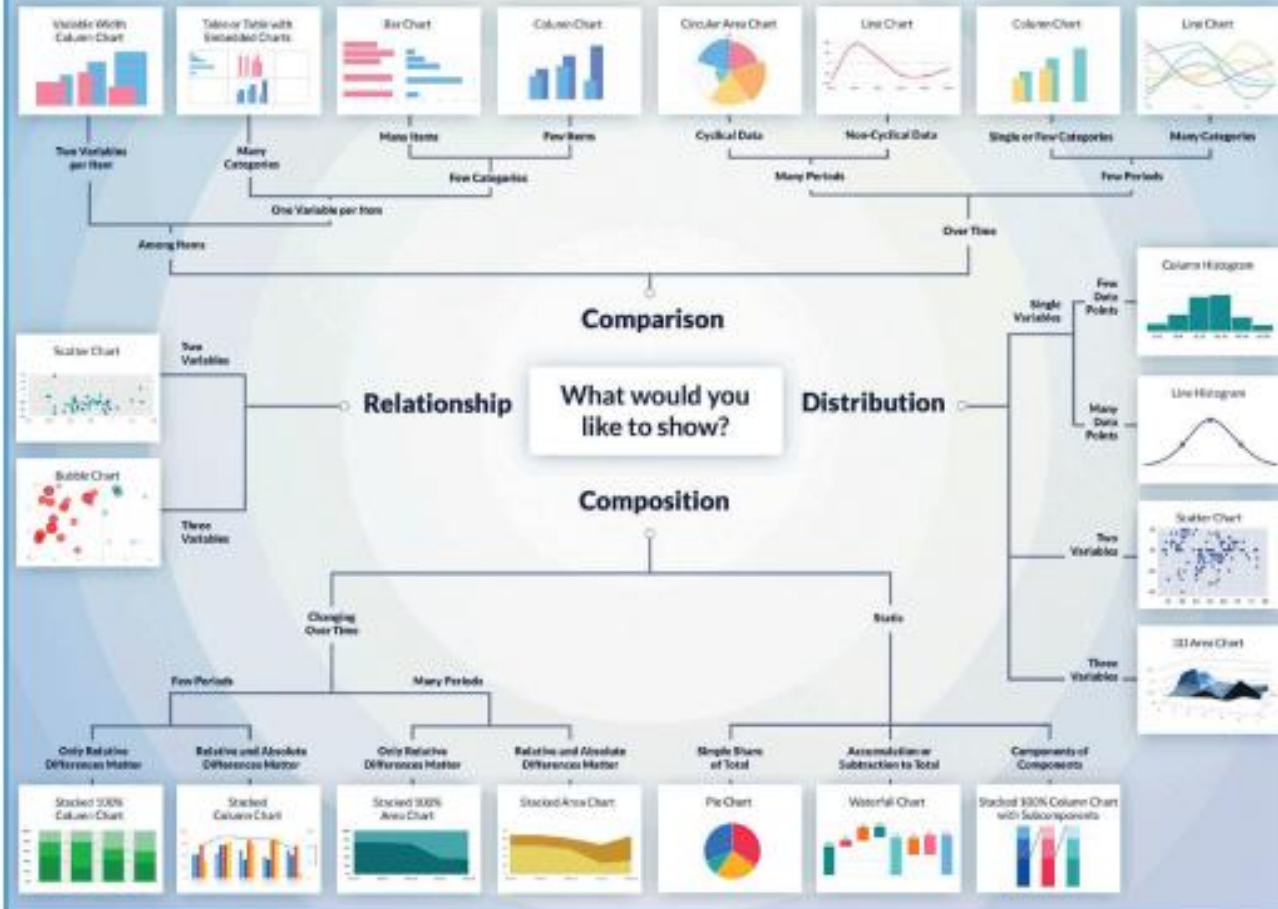
IV. Creating the Visualization :

you can create the visualization using a tool such as **Excel**, **Google Sheets**, or specialized data visualization software, such as **Microsoft Power BI** or **Tableau**
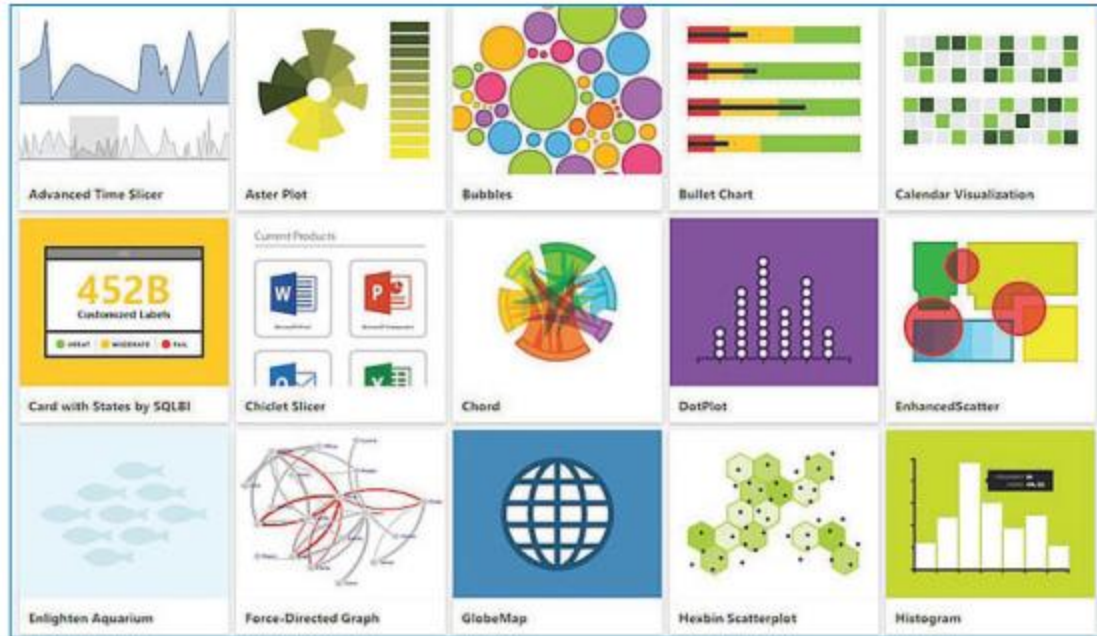
V. Customizing the Visualization:

Adding labels and annotations: Add labels to the axes of the visualization and to any data points or other important elements. You can also use annotations to highlight specific points or areas of the visualization.

- **Using colors effectively:** Use colors to differentiate between different data points or emphasize certain visualization elements. Avoid using too many colors, as this can be confusing.

Using appropriate scales: Make sure the scales on the axes of the visualization are appropriate for the data you are displaying. Using an inappropriate scale can distort the data and make it difficult to interpret.

- **Adding a title:** Add a title to the visualization to provide context and help the audience understand what they are looking at.

- **Adding a legend:** If you use multiple data series or different data types in the visualization, consider adding a legend to help the audience understand what each element represent

VI. Interpreting and Analyzing the Visualization :

-Review the goals of the visualization: Remind yourself of the goals and what you want to understand or communicate through them.

- Looking for patterns and trends: Look for patterns and trends in the data and try to identify any relationships or correlations.

- Checking for outliers: Look for any data points that are significantly different from the rest of the data. These may be errors or represent interesting trends that warrant further investigation.

- Considering the limitations of the visualization: Keep in mind that it is just one way of looking at the data and may not show the whole picture. Consider the limits of the visualization and think about other ways to analyze the data

VI. Interpreting and Analyzing the Visualization :

-Concluding: Use the visualization and your analysis to conclude the data. Be careful not to jump to conclusions based on limited data, and consider seeking additional data or context to confirm your conclusions.

VII. Sharing the Visualization :

There are a few different ways you can share a data visualization with others:

- Export the visualization as an image
- Embed the visualization in a report or presentation:
- Publish the visualization online:
- Share the data and visualization tool

Data Analysis Steps

Business Issue Understanding	Data Understanding	Data Preparation	Exploratory Analysis/ Modeling	Validation	Visualization & Presentation
• Business Objectives	• Initial Data collection	• Gather data from multiple sources	• Develop methodology	• Evaluate results	• Communicate results
• Information needed	• Data requirements	• Cleanse	• Determine Important variables	• Review process	• Determine best method/ graph to present insights based on analysis and audience.
• Type of analysis	• Data availability	• Format	• Build model	• Determine next steps:	
• Scope of work	• Data exploration and characteristics	• Blend	• Assess model	➢ OK	• Craft a compelling story
• Deliverables		• Sample		➢ NOT OK	• Make recommendations

Data Analysis Steps

1. **Business issue understanding:** the first step is <u>clearly defining the problem</u>, this helps ensure that the analysis is focused and relevant to the business context.

2. **Data understanding:** gather and understanding the relevant data (sources, collecting, quality, completeness).

3. **Data preparation:** after collecting data, it needs to be <u>cleaned</u>, integrated and transformated into format suitable for analysis.

Data Analysis Steps

4. **Modeling:** analysis patterns and relationships.

5. **Validation**: it's important to validate the results to ensure that thy are accurate (testing on different data sets)

6. **Visualization and presentation**: results need to be communicated to the relevant stakeholders.(charts, graphs).

7. **Sharing**: finally, insights and recommendation need to share with stakeholders.(presenting results in a report or dashboard)

Types of Analytics

Descriptive	Diagnostic	Predictive	Prescriptive
What happened?	Why did it happen?	What will happen?	What shoud I do?
• Descriptive Statistics • Data Clustering • Business Intelligence	• Business Intelligence • Sensitivity Analysis • Design of Experiments	• Linear and Logistic Regression • Neural Networks • Support Vector Machines	• Simulation such as Monte Carlo • Optimization such as Linear/Nonlinear Programming

Low ⟶ **Analytic Sophistication** ⟶ **High**

- **Descriptive Analytics:**
- Used to explain what is happening in a given situation.
e.g., Who are my customers? How many types of users do we have?

Types of Analytics

- **Diagnostic Analytics:**
- Helps you understand why certain things happened & the key drivers.
e.g., Why are dropped calls increasing? Why a drop in sales over the 3rd Qr.?

- **Predictive Analytics:**
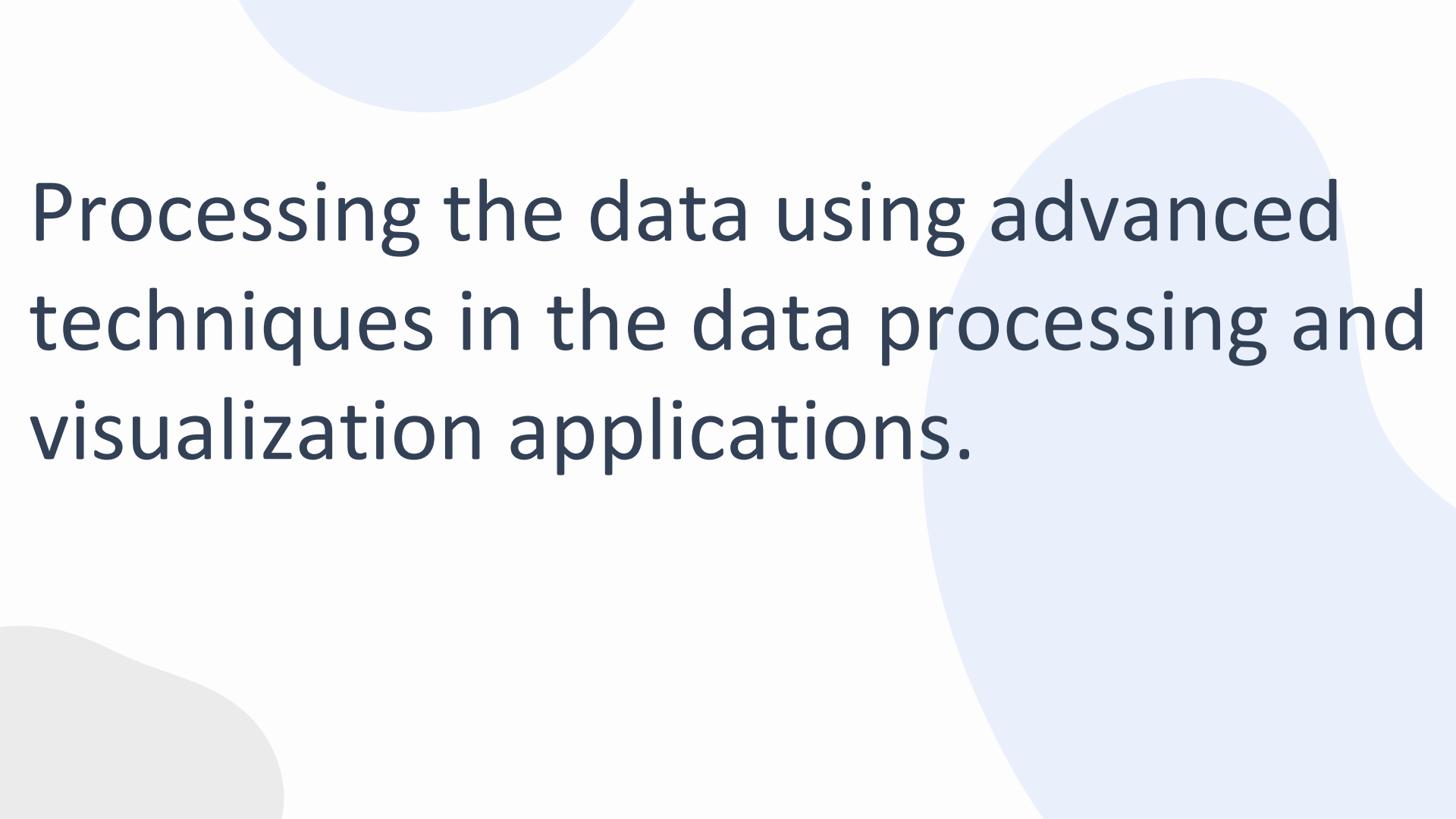- Helps to predict what will happen in the future. Use current & historical data.
e.g., Predict if a credit card transaction is fraudulent or not.

- **Prescriptive Analytics:**
- Necessary actions that are to be taken in case of a certain predicted event.
e.g., What measures should be taken to increase sales in the off season?

Processing the data using advanced techniques in the data processing and visualization applications.

Data Visualization Tools

1. Excel: A spreadsheet program that is commonly used to create simple charts and graphs.

2. Tableau: Data visualization software that allows users to create interactive dashboards and charts.

3. Google Charts: A free charting library that can be used to create a wide range of charts and graphs.

4. D3.js: A JavaScript library that can be used to create interactive visualizations using web standards

Data Visualization Tools

5. Matplotlib: A Python library that is commonly used to create static, animated, and interactive visualizations.

6. ggplot2: A data visualization package for the R programming language that is used to create a wide range of charts and graphs.

7. Plotly: A data visualization platform that allows users to create interactive charts and dashboards.

8. Infogram: A cloud-based data visualization platform that allows users to create a wide range of charts, graphs, and infographics.

9. FusionCharts: A data visualization library that allows users to create interactive charts and dashboards using JavaScript

Data Visualization Tools

10. Highcharts: A data visualization library that allows users to create interactive charts and dashboards using JavaScript.

11. Sisense: A data visualization and business intelligence platform that allows users to create interactive dashboards and charts.

12. Power BI: is a standalone Microsoft business intelligence product, which includes both desktop and web-based applications for loading, modeling, and visualizing data.

Tools of chart

| Tool	Cost	Ease of Use	Data Source Compatibility	Features	Rating	Mobile App
Excel	Paid	Easy	Limited	Basic visualizations, pivot tables, charting	3.5/5	Yes
Power BI	Free/ Paid	Easy	Wide range	Customizable visualizations, data modeling, collaboration	4.5/5	Yes
Tableau	Free/ Paid	Moderate	Wide range	Customizable visualizations, data blending, mapping	4.5/5	Yes
Google Charts	Free	Easy	Limited	Customizable visualizations, data visualization library	3/5	No
D3.js	Free	Difficult	Limited	Customizable visualizations, JavaScript library	4/5	No
Matplotlib	Free	Moderate	Limited	Customizable visualizations, data visualization library	3/5	No
ggplot2	Free	Difficult	Limited	Customizable visualizations, data visualization library	3/5	No
Plotly	Free/ Paid	Moderate	Wide range	Customizable visualizations, data visualization library	4/5	Yes
Infogram	Free/ Paid	Easy	Wide range	Customizable visualizations, data visualization library	3.5/5	Yes

Cleaning and Preparing the Data

Here are some general steps to clean and prepare the data:

1. Data acquisition: Collect and import data from various sources such as databases, spreadsheets, and text files.

2. Data inspection: Examine the data for any errors, inconsistencies, or missing values.

3. Data cleaning: Clean and preprocess the data to remove any errors, inconsistencies, or missing values. This may involve filling in missing values, removing duplicates, or transforming data to make it consistent.

4. Data transformation: Transform the data into the format required for visualization. This may involve aggregating data, creating new variables, or normalizing data

Data bases

a *database* is any logically modeled collection of information. A database does not necessarily have to be stored on a computer, and things like a stack of patient files in a hospital, a set of contacts in a rolodex, or file cabinet filled with old invoices all qualify as examples of databases.

For Example

create MySQL database
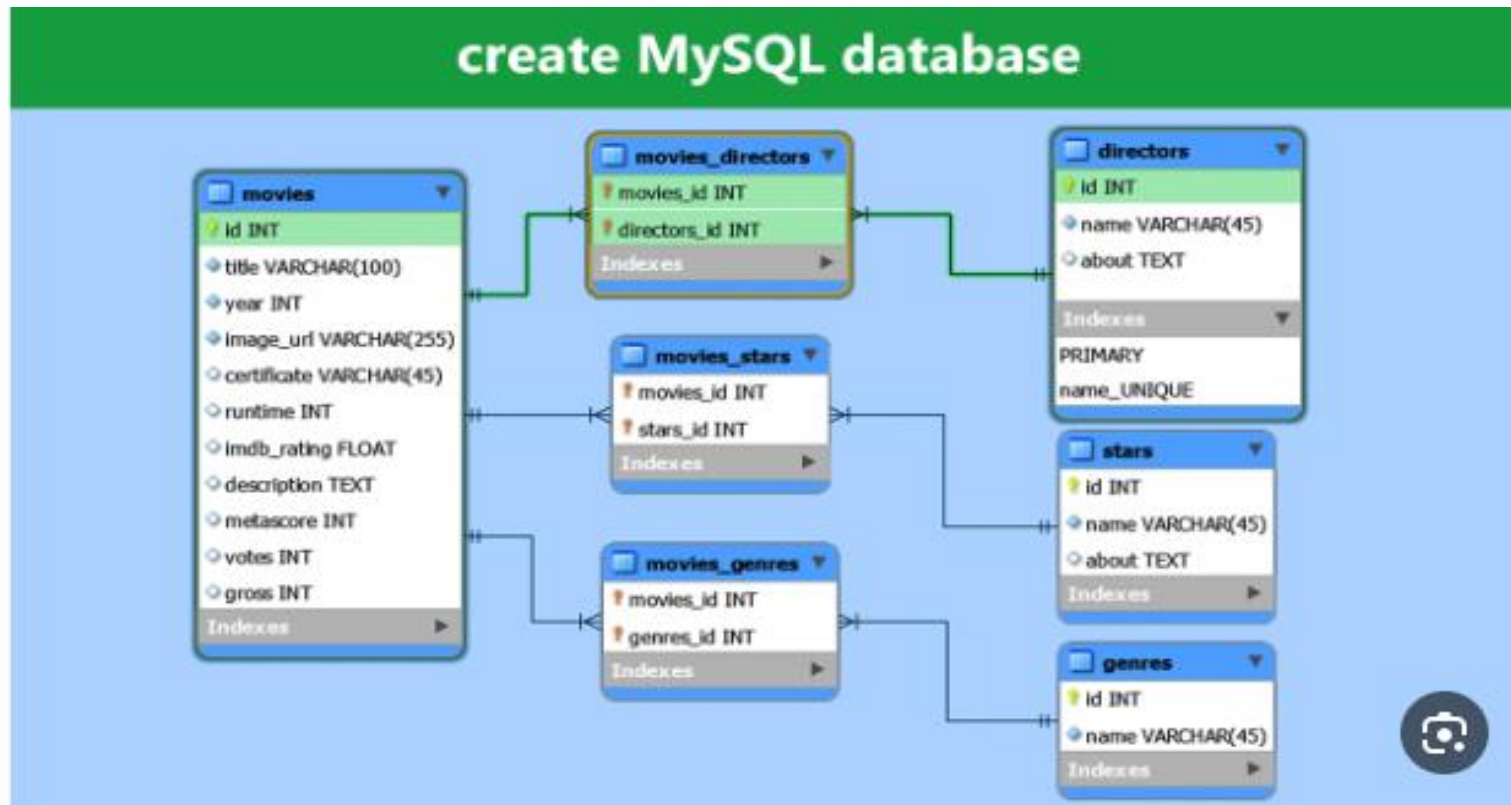
Ssis

Thanks!

Do you have any questions?

youremail@freepik.com

+34 654 321 432

yourwebsite.com