



# Data Visualization Using Python



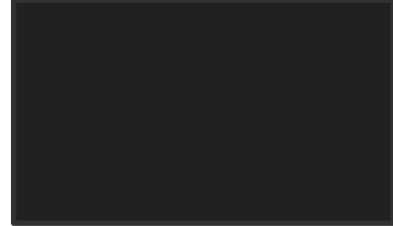
# Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.



By the end  
of this  
session,  
you will:

- Utilize interactive visualization libraries like Plotly and Bokeh to create dynamic plots and charts.
- Apply interactivity to enhance data exploration and presentation.
- Communicate insights effectively through well-designed visualizations to support data-driven decision-making processes.
- Gain an understanding of how to use specific chart types to best represent different types of libraries.



# Key Takeaways from This Session

---

**Visual Representation:** Data visualization provides a graphical representation of data, making complex information more accessible, understandable, and interpretable. It helps identify patterns, trends, and relationships that may not be apparent in raw data.

**Matplotlib and Seaborn:** Matplotlib and Seaborn are widely used Python libraries for data visualization. Matplotlib provides a flexible and customizable interface for creating a wide range of static plots, while Seaborn simplifies the creation of informative statistical visualizations.

**Interactive Visualizations:** Python libraries like Plotly and Bokeh offer interactive visualization capabilities, enabling users to create dynamic plots, charts, and dashboards. Interactivity enhances the exploration and presentation of data, allowing users to zoom, pan, and interact with the visualizations.

**Communication and Insights:** Data visualization serves as a powerful tool for communicating insights and findings to stakeholders.

## Pop Quiz

Q. Which of the following is incorrect regarding Data visualization?

- A. Data Visualization can be done using Matplotlib library in Python
- B. Visualizing large and complex data does not produce effective results
- C. Data Visualization is immensely useful in data analysis
- D. Decision makers use data visualization to understand business problems easily and build strategies



## Pop Quiz

Q. Which of the following is incorrect regarding Data visualization?

- A. Data Visualization can be done using Matplotlib library in Python
- B. Visualizing large and complex data does not produce effective results**
- C. Data Visualization is immensely useful in data analysis
- D. Decision makers use data visualization to understand business problems easily and build strategies





# Introduction to Plotly

# Importance of Data Visualization

---

## **Data Comprehension:**

Data visualization enables users to understand complex data sets quickly and easily. Visual representations, such as charts, graphs, and maps, provide a clear and concise summary of information, making patterns, trends, and outliers more apparent.

## **Insights and Decision-making:**

Effective data visualization facilitates the discovery of insights and patterns that may not be readily apparent in raw data. Visualizing data allows decision-makers to make more informed choices, identify correlations, and gain valuable insights to drive strategic decisions.



# Importance of Data Visualization

---

## **Communication and Engagement:**

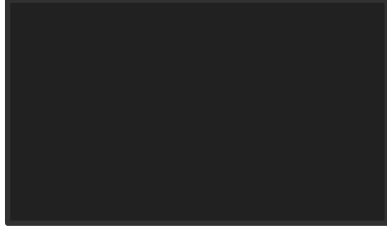
Visualizations are powerful tools for communication. They help convey complex information in a simplified and intuitive manner, making it easier for stakeholders to grasp and interpret data. Visualizations can engage audiences, facilitate discussions, and support storytelling.

## **Data Exploration and Discovery:**

Visualizations encourage exploration and experimentation with data. By interacting with visual representations, users can drill down into specific aspects, filter data, and uncover hidden relationships, leading to discoveries and a deeper understanding of the data.

# Plotly Fundamentals

---



- **Interactive:** Plotly creates interactive visualizations for data exploration.
- **Chart Variety:** Plotly offers diverse chart types like scatter plots, bar charts, and heatmaps.
- **Web-Based:** Plotly visualizations are web-based and can be easily shared.
- **Customizable:** Plotly allows customization of colors, fonts, and other visual elements.
- **Integration and Collaboration:** Plotly integrates with other libraries and supports collaborative work.

# Need for Plotly

---

- **Interactive:** Plotly enables interactive data visualization, allowing users to explore and interact with the data.
- **Visual Variety:** Plotly offers a wide range of chart types and visualizations to suit different data and analysis needs.
- **Web-Based Sharing:** Plotly visualizations can be easily shared online, making it convenient for collaboration and dissemination of insights.
- **Customization Options:** Plotly provides extensive customization options to tailor visualizations to specific preferences and branding requirements.
- **Integration Capabilities:** Plotly seamlessly integrates with other Python libraries and frameworks, enhancing its versatility and compatibility.

# Basic Concepts of Plotly

---

- **Figure:** A Figure in Plotly represents a complete visualization, which can include one or more traces or chart components.
- **Trace:** A Trace is a data object in Plotly that defines the data and visual properties of a specific chart element, such as a line, scatter point, or bar.
- **Layout:** The Layout object in Plotly defines the overall appearance and structure of the visualization, including axes, titles, annotations, and styling.
- **Annotations:** Annotations are text labels or markers that can be added to a Plotly visualization to provide additional context or information.
- **Offline and Online Modes:** Plotly can be used in both offline and online modes. In offline mode, visualizations are generated locally, while in online mode, visualizations can be stored, shared, and accessed online via Plotly's cloud platform.

## Pop Quiz

Q. Which of the following Plotly chart types is best for showing the distribution of a single variable?

- A. Line chart
- B. Bar chart
- C. Scatter plot
- D. Histogram



## Pop Quiz

Q. Which of the following Plotly chart types is best for showing the distribution of a single variable?

- A. Line chart
- B. Bar chart
- C. Scatter plot
- D. Histogram**





# Demo – Creating plots using Plotly

## Pop Quiz

Q. Which of the following Plotly chart types is best for comparing multiple data series?

- A. Line chart
- B. Bar chart
- C. Scatter plot
- D. Heat map





## Pop Quiz

Q. Which of the following Plotly chart types is best for comparing multiple data series?

- A. Line chart
- B. Bar chart**
- C. Scatter plot
- D. Heat map







# Bokeh Fundamentals

# Introduction to Bokeh

---

- **Interactive Visualization:** Bokeh is a Python library for creating interactive visualizations that can be displayed in web browsers. It allows users to create interactive plots, charts, and dashboards that respond to user interactions.
- **Modern Web-Based Graphics:** Bokeh leverages modern web technologies like HTML, CSS, and JavaScript to generate visually appealing and interactive plots. The visualizations are rendered using HTML5 canvas or WebGL for high-performance graphics.
- **Multiple Language Support:** Bokeh supports multiple programming languages, including Python, R, and Julia. This flexibility allows users to leverage the Bokeh library across different data analysis ecosystems.
- **Server-Based Applications:** Bokeh provides a server component called Bokeh Server, which allows the creation of server-based applications for real-time streaming, updating, and synchronization of data visualizations.
- **Seamless Integration:** Bokeh seamlessly integrates with other Python libraries, such as pandas and NumPy, making it convenient for data manipulation, transformation, and analysis prior to visualization.

# Need for Bokeh

---

- **Interactive Visualization:** Bokeh enables the creation of interactive visualizations, allowing users to explore and interact with data in a dynamic and engaging manner.
- **Web-Based Deployment:** Bokeh generates visualizations that can be deployed as web applications, making them easily accessible and shareable across different devices and platforms.
- **High-Performance Graphics:** Bokeh leverages modern web technologies to deliver high-performance graphics, enabling the visualization of large datasets with smooth interactivity.
- **Integration with Data Ecosystem:** Bokeh seamlessly integrates with other popular Python libraries, such as pandas and NumPy, facilitating the integration of data manipulation and analysis workflows with interactive visualizations.

# Basic Concepts of Bokeh

---

- **Glyphs:** Bokeh uses Glyphs to represent visual elements such as lines, bars, scatter points, or patches in a plot.
- **Plotting Interface:** Bokeh provides a high-level plotting interface that allows users to easily create visualizations using Python syntax.
- **Layouts:** Bokeh allows for arranging multiple plots or visual components into layouts, such as rows, columns, or grids.
- **Interactivity:** Bokeh supports interactive features like hover tooltips, zooming, panning, and selection tools to enhance user engagement with the visualizations.
- **Bokeh Server:** The Bokeh Server enables the creation of interactive applications and dashboards that can handle real-time updates and streaming data.

## Pop Quiz

Q. What is a Bokeh callback?

- A. A way to handle user interactions with a Bokeh widget
- B. A way to adjust the size of a plot
- C. A way to add annotations to a plot
- D. A way to organize multiple plots in a single figure



## Pop Quiz

Q. What is a Bokeh callback?

- A. A way to handle user interactions with a Bokeh widget**
- B. A way to adjust the size of a plot
- C. A way to add annotations to a plot
- D. A way to organize multiple plots in a single figure







# Demo – Creating plots using Bokeh

## Pop Quiz

Q. What is a Bokeh widget?

- A. A type of glyph used to represent data in a plot
- B. A way to organize multiple plots in a single figure
- C. A way to adjust the size of a plot
- D. A way to add interactive controls to a plot



## Pop Quiz

Q. What is a Bokeh widget?

- A. A type of glyph used to represent data in a plot
- B. A way to organize multiple plots in a single figure
- C. A way to adjust the size of a plot

**D. A way to add interactive controls to a plot**



# Comparison and Best Practices

| Library | Features                                  | Capabilities   | Use Cases  |
|---------|---|--|--|
| pandas  | - Data manipulation with DataFrames       | - Data cleaning, filtering, grouping, and reshaping data | - Data preprocessing                                 |
|         | - Series for one-dimensional data         | - Handling missing data effectively                      | - Data wrangling                                     |
| Seaborn | - High-level interface for visualizations | - Creating attractive statistical plots                  | - Visualizing statistical relationships and patterns |
|         | - Built on top of Matplotlib              | - Distribution plots, box plots, violin plots, etc.      | - Exploratory data analysis                          |

# Comparison and Best Practices

| Library | Features  | Capabilities                                      | Use Cases                              |
|---------|---|---|--|
| Plotly  | - Supports both static and interactive visualizations | - Wide range of chart types                       | - Interactive dashboards               |
|         | - Can be used with various programming languages      | - Zooming, panning, hover effects                 | - Web-based visualizations             |
| Bokeh   | - Interactive visualizations                          | - Smooth and responsive interactions              | - Interactive data applications        |
|         | - Supports large datasets                             | - Line plots, bar charts, scatter plots, etc.     | - Interactive data exploration tools   |
|         |   | - Customizable and works well with large datasets | - Visually appealing interactive plots |



# Demo – Working with pandas, Seaborn, Plotly and Bokeh

## Poll Time

Q. Which of the following is NOT a type of plot available in Seaborn?

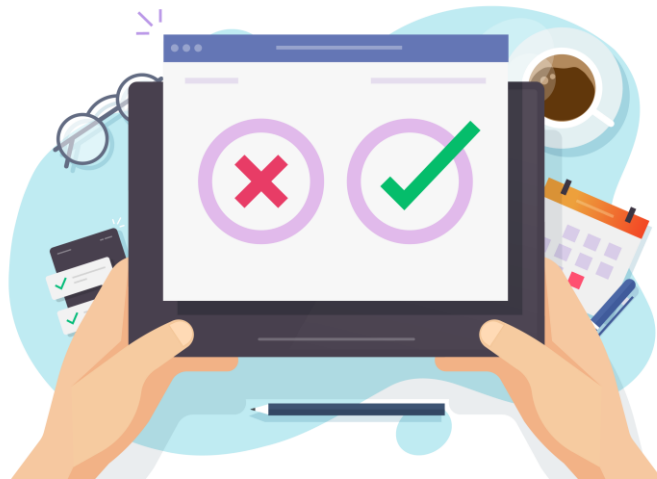
- A. Line plot
- B. Scatter plot
- C. Bar plot
- D. Tree plot



## Poll Time

Q. Which of the following is NOT a type of plot available in Seaborn?

- A. Line plot
- B. Scatter plot
- C. Bar plot
- D. Tree plot**







# Summary

---

- ✓ Developed proficiency in generating various types of visualizations.
- ✓ Utilized interactive visualization libraries, Plotly and Bokeh, to create dynamic plots and charts.
- ✓ Applied interactivity to enhance data exploration and presentation.
- ✓ Learned to communicate insights effectively through well-designed visualizations, supporting data-driven decision-making processes.

# Activity 1

---

## **Pre-requisites:**

- Basic understanding of Python programming.
- Familiarity with pandas library for data manipulation.

## **Scenario:**

- A company wants to analyze its sales data over the past year and visualize the monthly sales trends. They have provided a CSV file containing the sales data with columns for the date and sales amount.

## **Expected Outcome:**

Create a static line plot using Matplotlib or Seaborn that shows the monthly sales trends over the past year.

# Activity 1

---

## Steps:

- a. Import the necessary libraries: pandas, Matplotlib or Seaborn.
- b. Load the sales data from the provided CSV file into a Pandas DataFrame.
- c. Extract the month from the date column and create a new column for the month.
- d. Group the data by month and calculate the total sales for each month.
- e. Use Matplotlib or Seaborn to create a line plot showing the monthly sales trends.
- f. Label the axes and provide a title to the plot.
- g. Save the plot as an image or display it using the appropriate function.

## Activity 2

---

### **Pre-requisites:**

- Basic knowledge of Python programming and data manipulation with pandas.
- Understanding of static visualization with Matplotlib or Seaborn.

### **Scenario:**

An e-commerce company wants to explore the relationship between customer reviews and product ratings for a particular product category. They have a dataset with columns for product names, ratings, and customer reviews.

### **Expected Outcome:**

Create an interactive scatter plot using Plotly that allows users to hover over data points to view the product name, rating, and customer review.

## Activity 2

---

### Steps:

- a. Import the necessary libraries: Pandas, Plotly.
- b. Load the product data from the provided dataset into a Pandas DataFrame.
- c. Prepare the data by cleaning and organizing it if necessary.
- d. Create a scatter plot using Plotly, where the x-axis represents product ratings, and the y-axis represents the number of customer reviews.
- e. Customize the plot to include hover information displaying the product name and customer review when hovering over data points.
- f. Add labels to the axes and provide a title for the plot.
- g. Save the interactive plot as an HTML file or display it in the notebook using the appropriate function.

## Activity 3

---

### **Pre-requisites:**

- Knowledge of Python, understanding of interactive visualization libraries like Plotly or Bokeh.

### **Scenario:**

- You are working as a Data Scientist in a tech company. You are analyzing a large dataset of user behavior on your company's platform. Create an interactive dashboard using Plotly or Bokeh that includes different types of plots to display user metrics (like daily active users, time spent on the platform, and most visited sections).

### **Expected Outcome:**

- An interactive dashboard containing various dynamic plots, and charts that effectively present user metrics.

## Activity 3

---

### Steps:

- a. Import necessary libraries (pandas, plotly/bokeh).
- b. Load the user behavior dataset using pandas.
- c. Preprocess the data, handle missing values, and outliers if any.
- d. Create multiple plots to represent various user metrics.
- e. Combine these plots into a single dashboard.
- f. Test the interactivity features, such as zooming and panning.



## Next Session:

Data Visualization – Case Study

# THANK YOU

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.





# Data Visualization – Case Study





# Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.



# By the End of This Session, You Will Be Able To:

- Perform data visualization using Python libraries like Matplotlib, Seaborn, and pandas to gain insights from the smartphone sales dataset.
- Plot histograms to compare the distribution of screen height (sc\_h) and screen width (sc\_w) to identify any patterns or outliers.
- Learn how to filter mobile categories falling within a given price range and compare their internal storage using visualization techniques.
- Discover how to examine the dimensional nature of mobiles by visualizing various features and draw conclusions based on the findings.
- Understand how to create a joint plot to analyze the relationship between 3G and 4G devices and identify any correlations.

## Poll Time

Q. Which of the following type of chart is not valid?

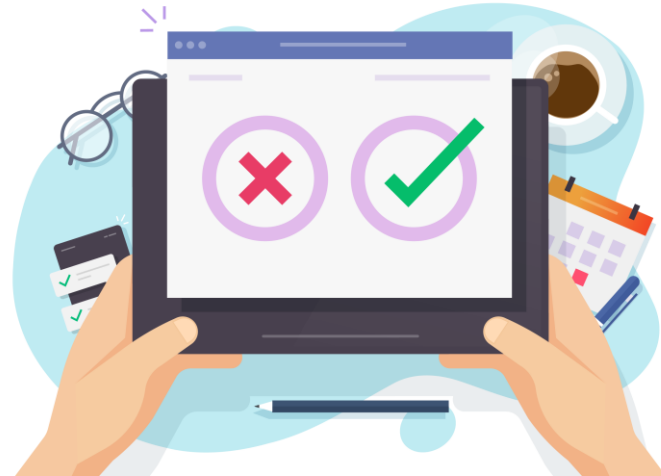
- A. Line
- B. Pie
- C. Curve
- D. Bar



## Poll Time

Q. Which of the following type of chart is not valid?

- A. Line
- B. Pie
- C. Curve**
- D. Bar





# Case Study – Problem Statement

# Overview of Dataset

| battery_power | blue | clock_speed | dual_sim | fc | four_g | int_memory | m_dep | mobile_wt | n_cores | pc | px_height | px_width | ram  | sc_h | sc_w | talk_time | three_g | touch_screen | wifi | price_range |
|---------------|------|-------------|----------|----|--------|------------|-------|-----------|---------|----|-----------|----------|------|------|------|-----------|---------|--------------|------|-------------|
| 842           | 0    | 2.2         | 0        | 1  | 0      | 7          | 0.6   | 188       | 2       | 2  | 20        | 756      | 2549 | 9    | 7    | 19        | 0       | 0            | 1    | 1           |
| 1021          | 1    | 0.5         | 1        | 0  | 1      | 53         | 0.7   | 136       | 3       | 6  | 905       | 1988     | 2631 | 17   | 3    | 7         | 1       | 1            | 0    | 2           |
| 563           | 1    | 0.5         | 1        | 2  | 1      | 41         | 0.9   | 145       | 5       | 6  | 1263      | 1716     | 2603 | 11   | 2    | 9         | 1       | 1            | 0    | 2           |
| 615           | 1    | 2.5         | 0        | 0  | 0      | 10         | 0.8   | 131       | 6       | 9  | 1216      | 1786     | 2769 | 16   | 8    | 11        | 1       | 0            | 0    | 2           |
| 1821          | 1    | 1.2         | 0        | 13 | 1      | 44         | 0.6   | 141       | 2       | 14 | 1208      | 1212     | 1411 | 8    | 2    | 15        | 1       | 1            | 0    | 1           |
| 1859          | 0    | 0.5         | 1        | 3  | 0      | 22         | 0.7   | 164       | 1       | 7  | 1004      | 1654     | 1067 | 17   | 1    | 10        | 1       | 0            | 0    | 1           |
| 1821          | 0    | 1.7         | 0        | 4  | 1      | 10         | 0.8   | 139       | 8       | 10 | 381       | 1018     | 3220 | 13   | 8    | 18        | 1       | 0            | 1    | 3           |
| 1954          | 0    | 0.5         | 1        | 0  | 0      | 24         | 0.8   | 187       | 4       | 0  | 512       | 1149     | 700  | 16   | 3    | 5         | 1       | 1            | 1    | 0           |
| 1445          | 1    | 0.5         | 0        | 0  | 0      | 53         | 0.7   | 174       | 7       | 14 | 386       | 836      | 1099 | 17   | 1    | 20        | 1       | 0            | 0    | 0           |
| 509           | 1    | 0.6         | 1        | 2  | 1      | 9          | 0.1   | 93        | 5       | 15 | 1137      | 1224     | 513  | 19   | 10   | 12        | 1       | 0            | 0    | 0           |
| 769           | 1    | 2.9         | 1        | 0  | 0      | 9          | 0.1   | 182       | 5       | 1  | 248       | 874      | 3946 | 5    | 2    | 7         | 0       | 0            | 0    | 3           |



# Overview of Problem Statement

---

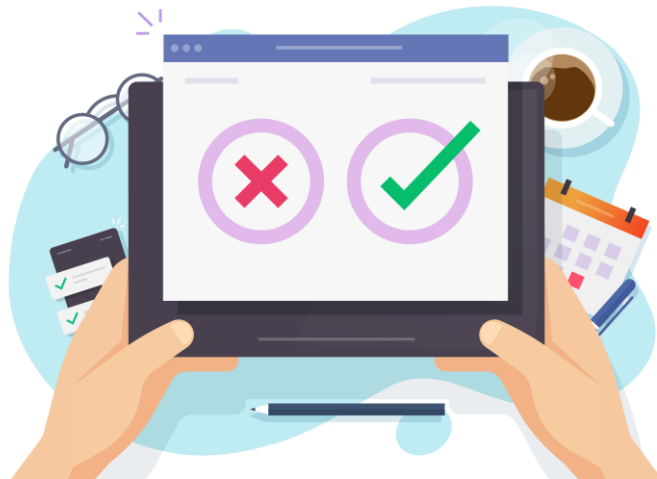
John is seeking a comprehensive understanding of the data points and distributions present in his dataset. Specifically, he is interested in identifying correlations between different features and their respective prices. He also wants to gauge the significance of each feature present in the dataset. To achieve this, you need to employ data visualization techniques using various plots that allow us to analyze these features.

By visualizing the data, you can identify any trends, outliers, or relationships that may exist between different variables. This will enable John to make more informed decisions based on the insights gained from the visualization of his data.

## Poll Time

Q. What is the Bokeh server used for?

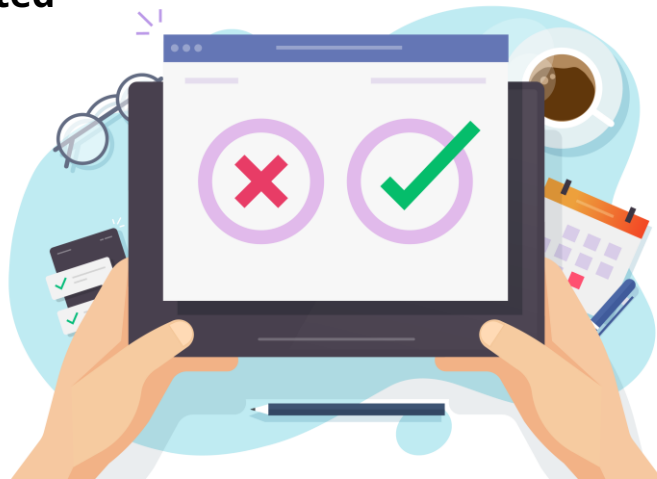
- A. To create static visualizations
- B. To build interactive web applications that are connected to Python code running on a server
- C. To create data pipelines
- D. To analyze data



## Poll Time

Q. What is the Bokeh server used for?

- A. To create static visualizations
- B. To build interactive web applications that are connected to Python code running on a server**
- C. To create data pipelines
- D. To analyze data



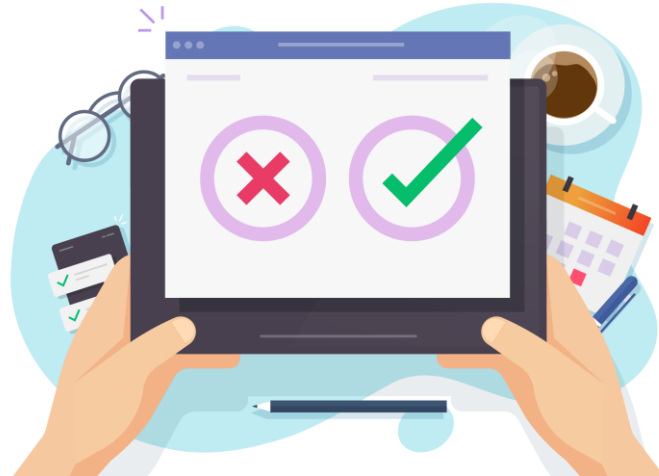


# Hands-on: Case Study Questions

## Poll Time

Q. Point out the correct combination with regards to the kind keyword for graph plotting.

- A. 'hist' for histogram
- B. 'box' for boxplot
- C. 'area' for area plots
- D. All of the listed



## Poll Time

Q. Point out the correct combination with regards to the kind keyword for graph plotting.

- A. 'hist' for histogram
- B. 'box' for boxplot
- C. 'area' for area plots
- D. All of the listed**







# How Do Things Work in the Real World?





How would You Go About with  
this Twist?

## Poll Time

Q. The \_\_\_\_\_ project builds on top of pandas and Matplotlib to provide easy plotting of data.

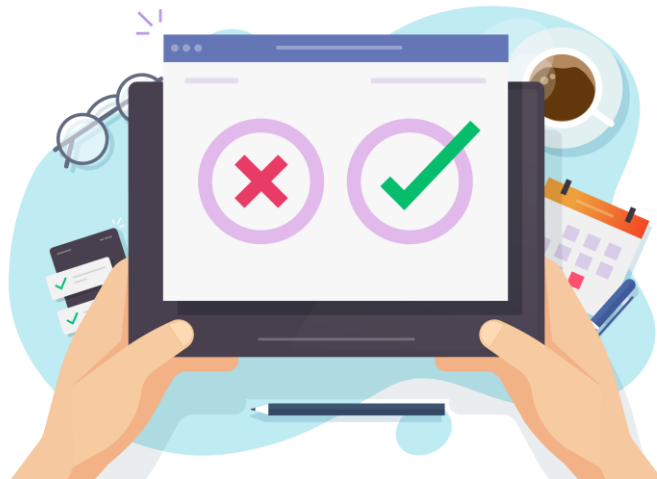
- A. Yhat
- B. Seaborn
- C. Vincent
- D. Pychart



## Poll Time

Q. The \_\_\_\_\_ project builds on top of pandas and Matplotlib to provide easy plotting of data.

- A. Yhat
- ☒ B. **Seaborn**
- C. Vincent
- D. Pychart



## Summary

---

- ✓ Understood the importance of data visualization in gaining insights from a dataset.
- ✓ Discovered how to determine user preferences for 3G by visualizing the count of 3G-supported devices.
- ✓ Acquired the skills to filter mobile categories falling within a given price range and compare their internal storage using visualizations.
- ✓ Learned how to examine the dimensional nature of mobiles by visualizing various features and draw conclusions from the findings.
- ✓ Got exposed to Python libraries like Matplotlib, Seaborn, and Plotly to visualize and analyze data effectively.

# Activity 1

---

Pre-requisites: Basic understanding of Python programming, pandas, and data visualization libraries like Matplotlib and Seaborn.

Scenario: John has recently expanded his electronics stores and collected data on the sales of various smartphones from the new locations. He wants to identify any trends or patterns in the data that could help him optimize the inventory and make informed business decisions.

Expected Outcome: Learners are expected to create a comprehensive data visualization report using Python's data visualization libraries to uncover insights from the new smartphone sales dataset.

# Activity 1

---

## Steps:

- a. Load the new dataset into a Pandas DataFrame.
- b. Explore the dataset to understand its structure and available features.
- c. Use Matplotlib and Seaborn to create visualizations that compare smartphone features like battery power, RAM, camera quality, and screen size with their corresponding prices.
- d. Plot histograms, scatter plots, and bar plots to identify any relationships between different variables and the price\_range.
- e. Create a heat map to visualize the correlation between different features.
- f. Identify any outliers or unusual patterns in the data.
- g. Summarize the key findings in a written report, explaining the insights gained from the visualizations.

## Activity 2

---

**Pre-requisites:** Familiarity with Python, pandas, and Bokeh library for interactive visualization.

**Scenario:** John wants to track the sales performance of different smartphone models over time. He wants an interactive data dashboard that allows him to view trends and patterns easily, helping him make data-driven decisions for his electronics stores.

**Expected Outcome:** Learners should create an interactive data dashboard using Bokeh that visualizes smartphone sales data, enabling John to monitor sales trends and make strategic business decisions.

## Activity 2

---

### Steps:

- a. Load the smartphone sales dataset into a pandas DataFrame.
- b. Clean and preprocess the data if necessary.
- c. Use Bokeh to create interactive visualizations like line charts or bar charts that show the sales performance of each smartphone model over different months or quarters.
- d. Include dropdown menus or sliders to allow John to filter the data based on specific features like RAM, camera quality, or price range and see how it impacts sales.
- e. Implement hover tools to show additional information like battery power or screen size when the user hovers over data points.
- f. Design the dashboard layout with appropriate titles, labels, and interactive elements for a user-friendly experience.
- g. Deploy the interactive data dashboard on a web server or share it as a standalone HTML file, allowing John to access and explore the data easily.





# Session Feedback



## Next Session:

Introduction to NoSQL

# THANK YOU

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

