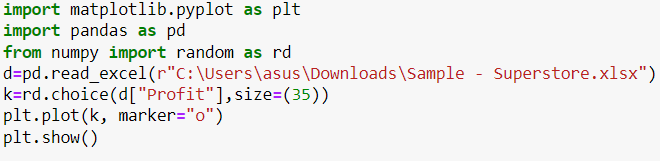
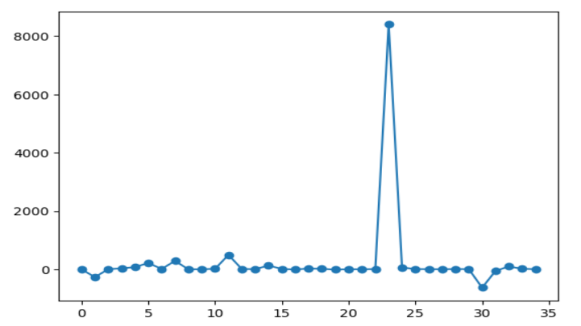
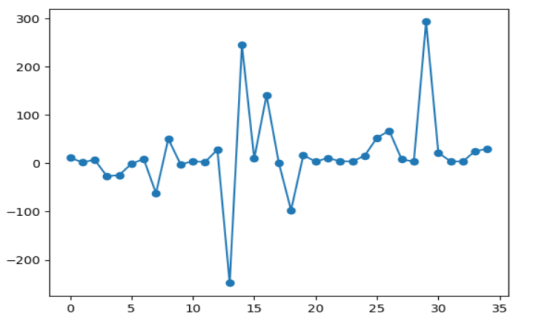
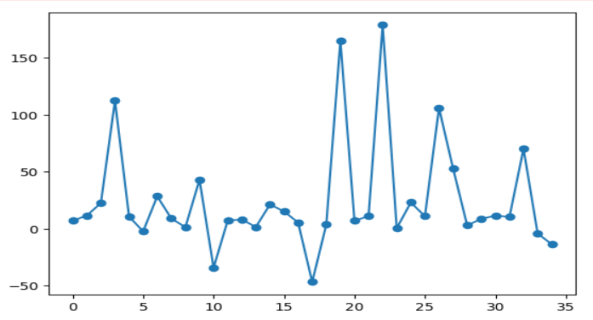
**Question:1.** Show three line graph on profit with random sampling of 35 rows?

**Ans:**

****

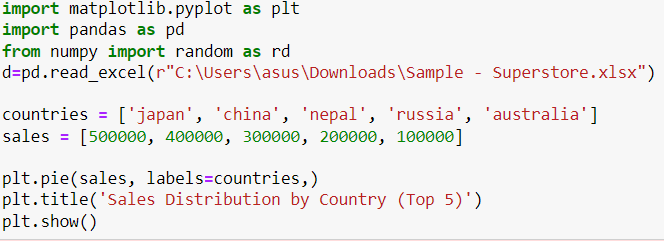
**Output:**

** **

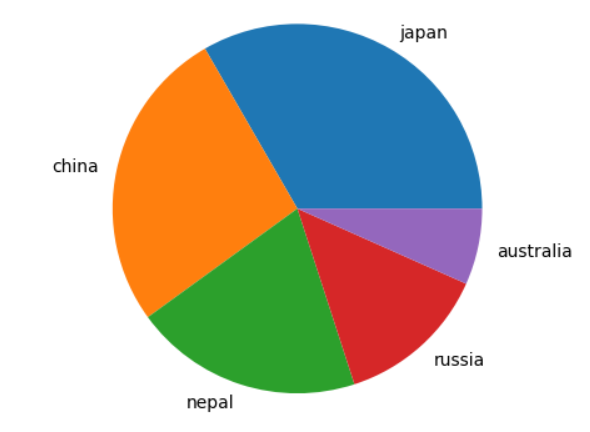


**Question:2.** Show the pie chart on sales with fixed sampling of top 5 country by sales.

**Ans:**

****

**Output:**

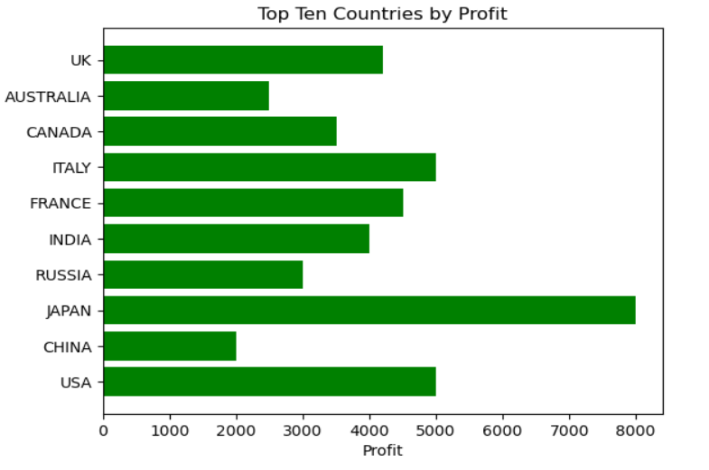
****

**Question:3. Plot top ten country by profit in horizontal bar chart.**

**Ans:**

****

**Output:**

****

**Question:4. Show a pie chart on top ten country by profit.**

**Ans:**

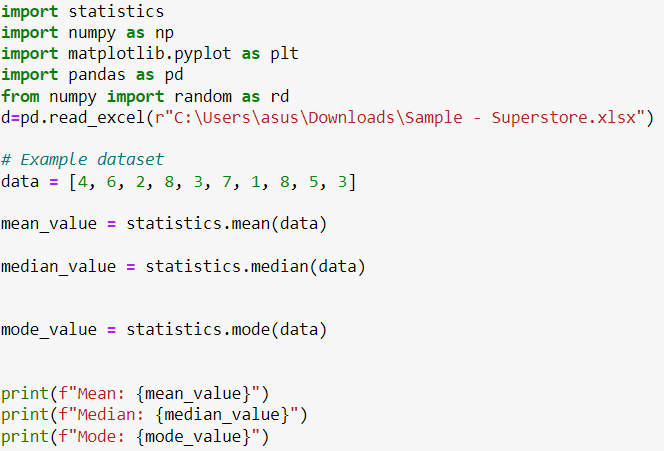
****

**Output:**

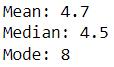
****

**Question:5. what do you mean by mean median mode give example code for all.**

**Ans:**

****

**Output:**

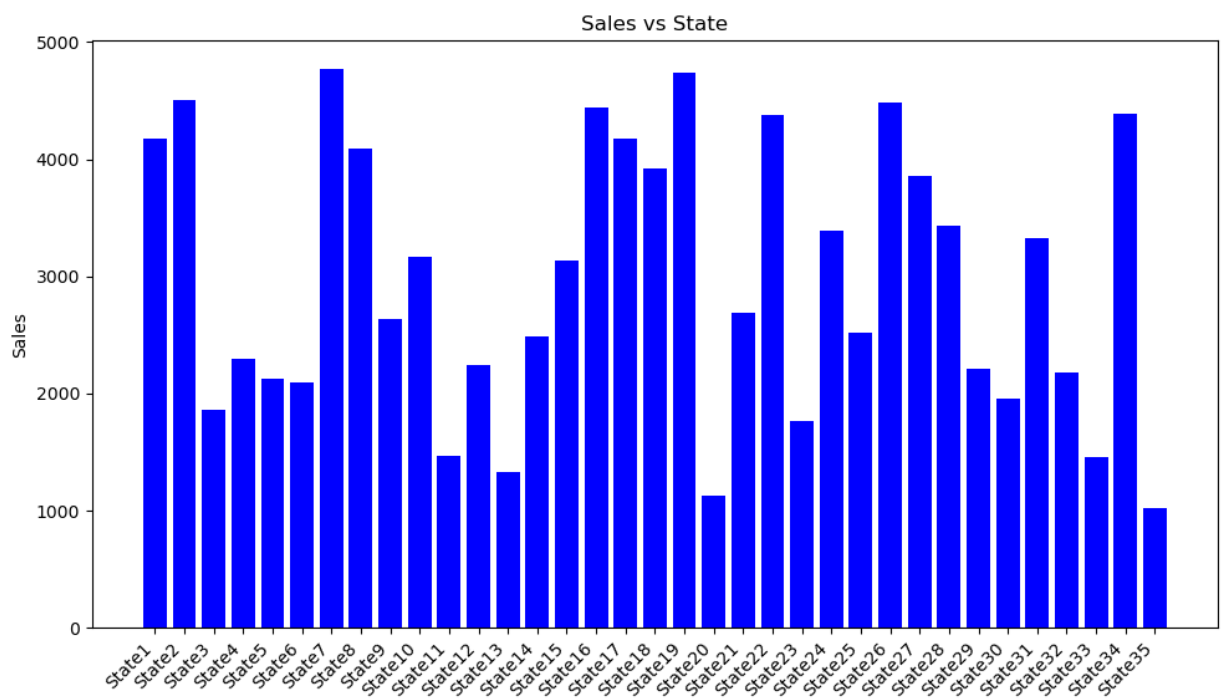
****

**Question:6. draw a bar chart sales vs state with random sampling of 35 rows.**

**Ans:**

****

**Output:**

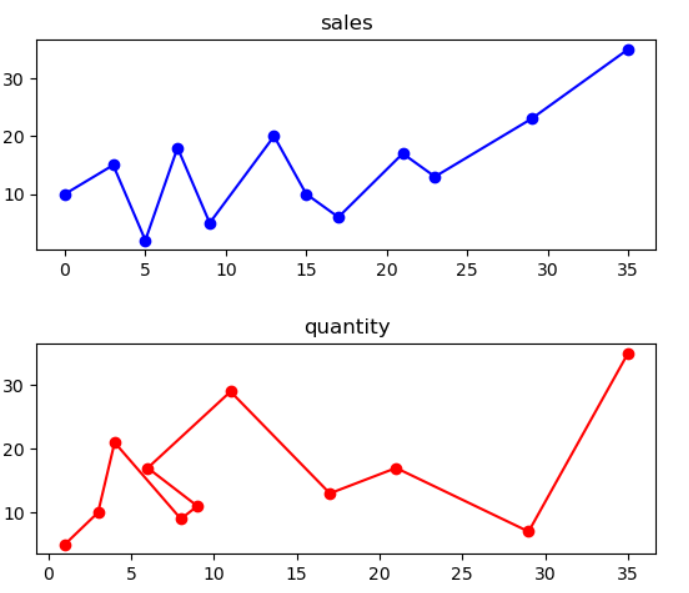
****

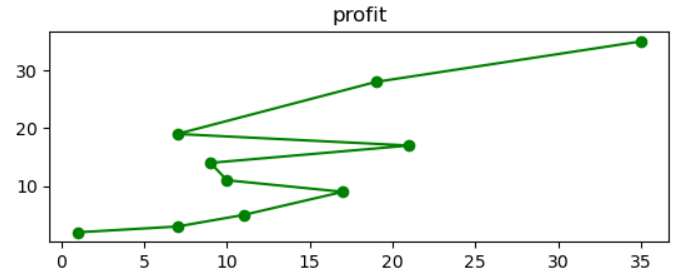
**Question.7: show a line graph comparison between sales ,quantity and profit with fixed sampling of 35 rows with sub plot.**

**Ans:**

****

**Output:**

****

****

**Question.8: Define data cleaning and explain its terminology.**

**Ans:** Data cleaning, also known as data cleansing or data scrubbing, is the process of identifying and correcting errors or inconsistencies in datasets to improve their quality. The goal of data cleaning is to enhance the accuracy, reliability, and usability of data for analysis and decision-making. Here are some key terminologies associated with data cleaning:

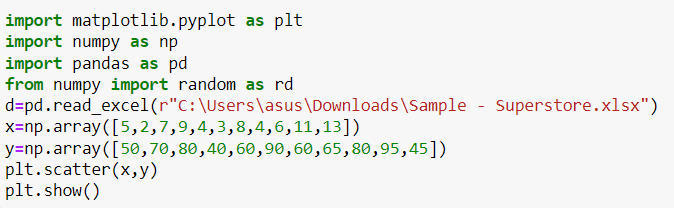
1. **Data Quality:**
   * **Accuracy:** The degree of closeness between the data values and their true or correct values. Inaccurate data may contain errors or discrepancies.
   * **Completeness:** The extent to which data is whole and contains all the necessary information. Incomplete data may have missing values.
   * **Consistency:** Ensures that data is uniform and follows predefined standards or rules. Inconsistent data may have conflicting information.
2. **Data Cleaning Processes:**
   * **Deduplication:** Identifying and removing duplicate records or entries within a dataset.
   * **Handling Missing Values:** Dealing with instances where data values are not recorded. This can involve imputation or removal of missing data.
   * **Outlier Detection:** Identifying and handling data points that deviate significantly from the overall pattern of the dataset.
   * **Normalization/Standardization:** Ensuring that data is in a consistent format or unit, making it easier to compare and analyze.
3. **Data Cleaning Techniques:**
   * **Imputation:** Filling in missing values with estimated or calculated values based on the surrounding data.
   * **Parsing:** Breaking down complex data fields into simpler components (e.g., extracting date or time information from a timestamp).
   * **Transformations:** Applying mathematical operations or functions to data to convert it into a more suitable form.
   * **Filtering:** Removing irrelevant or unnecessary data points from the dataset.
4. **Data Cleaning Tools:**
   * **OpenRefine:** An open-source tool for cleaning and transforming data.
   * **Trifacta:** A cloud-based data cleaning and preparation platform.
   * **Pandas (Python Library):** Widely used for data manipulation and cleaning in Python

Effective data cleaning is crucial for ensuring that the data used for analysis or machine learning models is reliable and produces meaningful insights. It is an essential step in the overall data preparation pipeline.

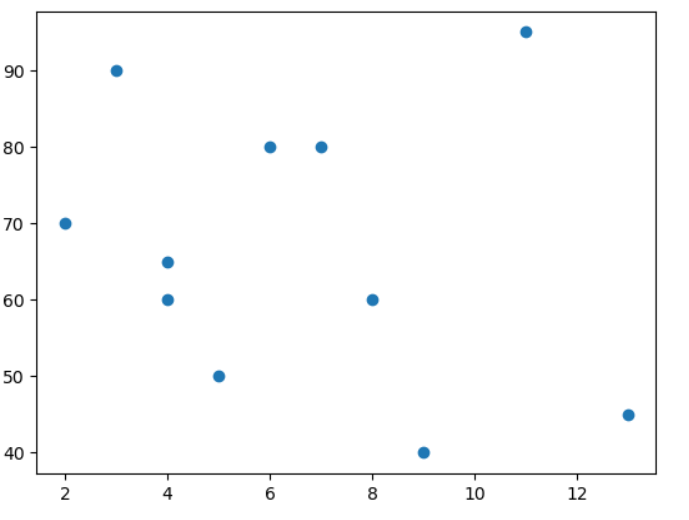
**Question:9. What is the scatter plot, pie chart, histogram, line chart, bar chart and how these are uses?**

**Ans: Scatter plot:** A scatter plot is a type of graph that displays two-dimensional information in a clear and organized method. A scatter plot consists of an x-axis that runs horizontally and a y-axis that runs vertically. The collected data appears on the graph with dots that indicate the corresponding points.

Uses: A scatter plot is a useful demonstration tool for when you have two variables within your data that you want to compare. After placing your dots on the graph, you can use the visualization to identify relationships between the two variables. instance, the following professions often use scatter plots: **Economists, Market research** **analyst, Census Bureau analyst , Educational Researchers**

****

**Output:**

****

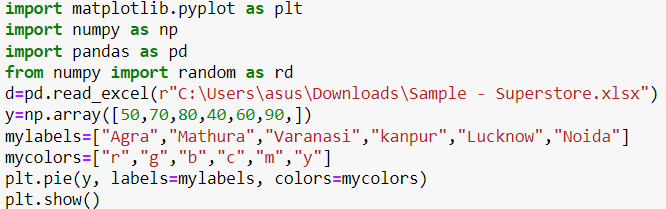
**Pie Chart:** A **pie chart** is a type of graph representing data in a circular form, with each slice of the circle representing a fraction or proportionate part of the whole. All slices of the pie add up to make the whole equaling 100 percent and 360 degrees. A pie chart is used to:

\* show parts-to-whole relationship.

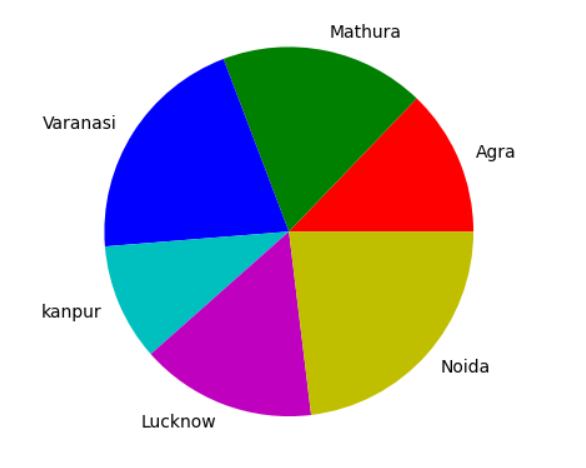
\* Compare the contribution of different categories to the whole.

\* To show the performance of projects, tools, strategies, or individuals

**Example:**

****

**Output:**

****

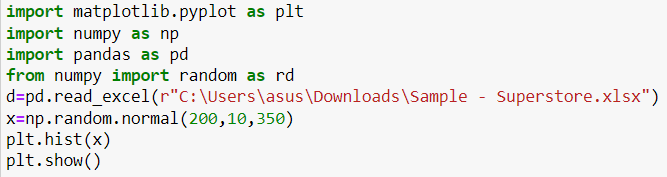
**Histogram:** A histogram is a graphical representation of the distribution of a dataset. It is a way to visualize the underlying frequency distribution of a continuous set of data. In a histogram, the data is divided into intervals, called bins, and the vertical axis represents the frequency or count of data points within each bin. The width of the bars corresponds to the range of values in each bin. Here's how to interpret and understand a histogram:

* **X-axis (horizontal):** Represents the range of values in the dataset, divided into bins.
* **Y-axis (vertical):** Represents the frequency or count of data points falling within each bin.

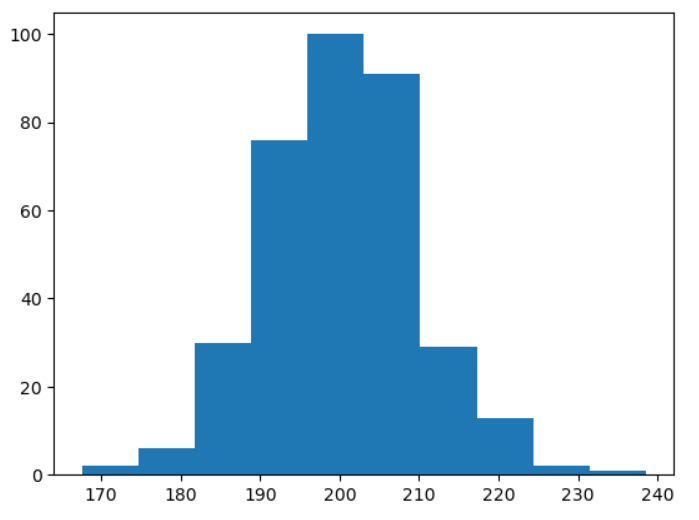
**Key features of a histogram:**

1. **Shape:**
   * **Symmetry/Asymmetry:** A histogram may be symmetric (bell-shaped, like a normal distribution) or asymmetric (skewed to the left or right).
   * **Modality:** The number of peaks in the distribution.
2. **Center:**
   * **Central Tendency:** The central value or values around which the data is centered.
3. **Spread:**
   * **Dispersion:** The extent to which data values are spread out.
4. **Outliers:**
   * **Outliers:** Data points that significantly deviate from the overall pattern.

Example:



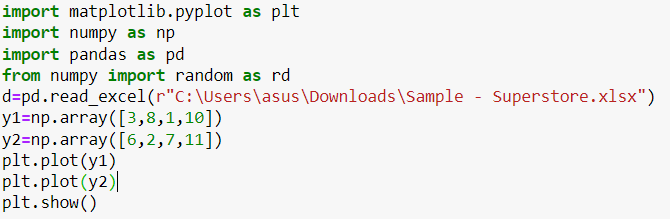
**Output:**

****

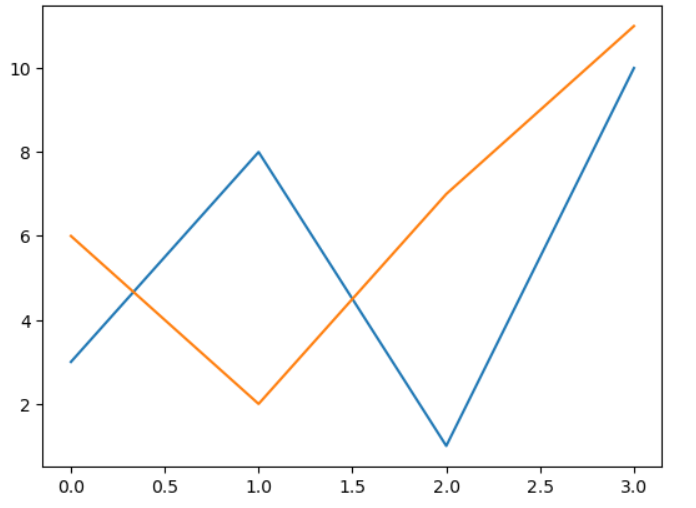
**Line chart:** A line chart is a type of data visualization that displays data points over a continuous interval or time span, connecting them with straight lines. The x-axis (horizontal) typically represents the independent variable, such as time or categories, while the y-axis (vertical) represents the dependent variable. Line charts are effective for showing trends and patterns in data over a continuous range. **Key features of line chart**

1. **Line Connections:**
   * Data points are connected with lines to show the trend or progression over the given interval.
2. **X and Y Axes:**
   * The x-axis represents the independent variable, and the y-axis represents the dependent variable.
3. **Data Points:**
   * Each data point is marked on the chart, making it easy to see the values at specific points.

**Example:**

****

**Output:**

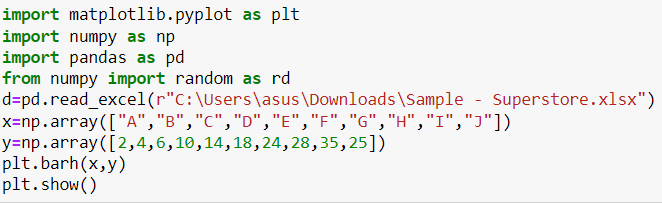
****

**Bar chart :** A bar chart is a graphical representation of data in which rectangular bars or columns are used to represent different categories or groups. The length or height of each bar corresponds to the value it represents. Bar charts are effective for comparing discrete categories or showing the distribution of data across different groups.

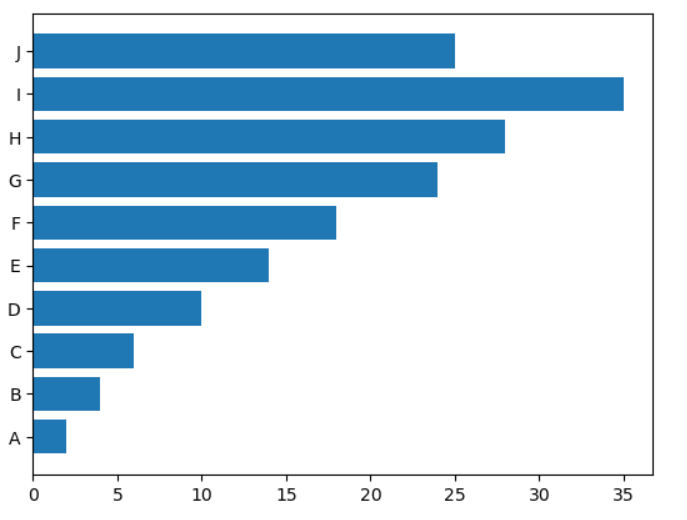
**Key features of bar charts:**

1. **Rectangular Bars:**
   * Each category or group is represented by a rectangular bar, with the length or height of the bar proportional to the value it represents.
2. **Categories on X-axis:**
   * The x-axis (horizontal axis) typically represents the different categories or groups being compared.
3. **Values on Y-axis:**
   * The y-axis (vertical axis) represents the values associated with each category.

**Example:**

****

**Output:**

****

**Example:**

**Question:10. what is the importance of data science in today’s business what do you mean by standard deviation?**

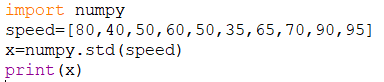
**Ans: Informed Decision-Making:**Data science helps businesses make informed decisions by analyzing large volumes of data. It provides insights into customer behavior, market trends, and internal processes, enabling strategic decision-making.

1. **Efficiency and Productivity:**
   * Businesses can enhance efficiency and productivity by using data science to automate repetitive tasks, optimize workflows, and identify areas for improvement.
2. **Customer Insights:**
   * Data science allows businesses to understand their customers better. By analyzing customer data, preferences, and feedback, companies can tailor their products and services to meet customer needs.
3. **Predictive Analytics:**
   * Predictive analytics, a subset of data science, enables businesses to forecast future trends and outcomes. This helps in proactive planning and mitigating risks.
4. **Competitive Advantage:**
   * Companies that leverage data science gain a competitive advantage. By analyzing market trends and consumer behavior, businesses can position themselves strategically in the market.
5. **Personalization:**
   * Data science enables personalized marketing and product recommendations. Businesses can create targeted campaigns based on individual preferences, improving customer engagement.
6. **Risk Management:**
   * Data science is crucial for risk management. It helps identify potential risks, fraud, and anomalies in financial transactions, ensuring the security of business operations.
7. **Cost Reduction:**
   * Through data-driven optimization, businesses can identify cost-saving opportunities, streamline operations, and allocate resources more efficiently.
8. **Innovation and Research:**
   * Data science fosters innovation by providing insights that drive the development of new products, services, and business models. It is particularly vital for industries like healthcare, finance, and technology.
9. **Supply Chain Optimization:**
   * Businesses can optimize their supply chains by using data science to forecast demand, manage inventory, and enhance overall logistics.

**Standard Deviation:**

Standard deviation is a measure of the amount of variation or dispersion in a set of values. It quantifies how much individual values in a dataset deviate from the mean (average). In other words, it provides a sense of the spread of data points around the mean.

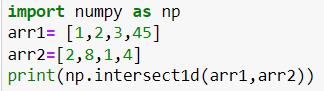
**Example:**

****

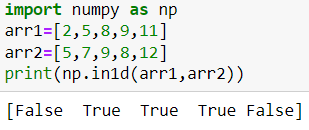
**Output:**

****

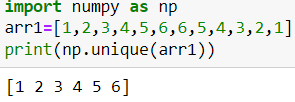
Question Write a NumPy program to find common values between two arrays.

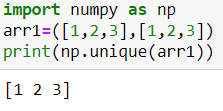
****

**Question:** Write a NumPy program to test whether each element of a 1-D array is also present in a second array.

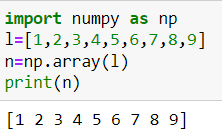
****

Question.:Write a NumPy program to get the unique elements of an array.

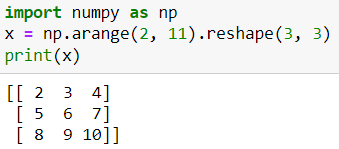




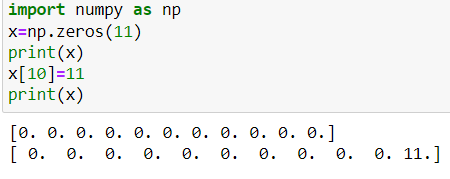
Question: write a Numpy program to covert a list of numeric values into a one-dimensional Numpy array.

Ans: 

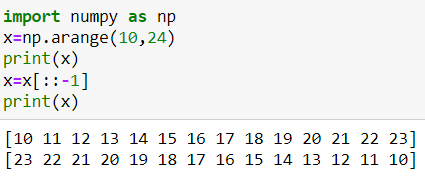
Question: write a Numpy program to create a 3\*3 matrix with values ranging from 2 to 10.



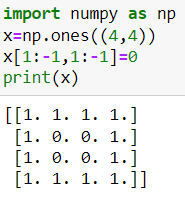
Question: write a Numpy program to create a null vector of size 10 and update the sixth value of 11.

Ans: 

Question: write a numpy program to revese any array (the first element becomes the last).

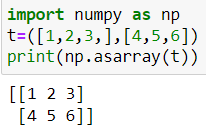


Question : write a numpy program to create a 2D array with 1 on the border and 0 inside.

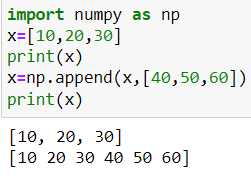


Question: write a numpy program to create an 8\*8 matrix and fill it with a checkerboard pattern.

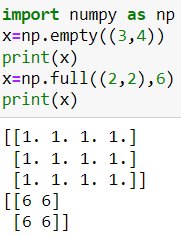
Question: write a Numpy program to convert a list and tuple into arrays.



Question: write a numpy program to append values to the end of an array.



Question: write a numpy program to create an empty and full array.



Question: write a numpy program to convert centigrade degree into Fahrenheit degrees. Centigrade values are stored in numpy array.

Question: write a numpy program to find the real and imaginary parts of an array of complex numbers.

Question: write a numpy program to find the number of elements in an array it also finds the length of one array element in byteys and the total bytes. Consumed by the elements.

Question: write a numpy program to test whether each element of a 1-D array is also present in second array.