

Experiment-2

Aim:

Data Visualization/ Exploratory data Analysis using Matplotlib and Seaborn.

1. Create bar graph, contingency table using any 2 features.
2. Plot Scatter plot, box plot, Heatmap using seaborn.
3. Create histogram and normalized Histogram.
4. Describe what this graph and table indicates.
5. Handle outlier using box plot and Inter quartile range.

Introduction:

Exploratory Data Analysis (EDA), a concept pioneered by statistician John Tukey in the 1970s, represents a fundamental step in the data analysis process. EDA involves examining datasets in-depth through a combination of statistical summaries and visual representations to better understand the structure, patterns, and nuances within the data. By doing so, analysts can uncover hidden trends, spot anomalies or outliers, and gain crucial insights that guide further data preparation and model building. EDA acts as the bridge between raw data and the application of machine learning algorithms.

Why Perform EDA?

Exploratory Data Analysis is a critical phase in any data science or analytics workflow for several reasons:

- **Understanding the Data Structure:** It helps in recognizing the underlying distribution and characteristics of variables in the dataset.
- **Identifying Key Patterns and Trends:** Analysts can detect meaningful patterns, seasonality, or trends that may influence decision-making.
- **Detecting Relationships and Correlations:** EDA reveals how different features relate to each other, which can be vital for feature selection in predictive modeling.
- **Evaluating Data Quality:** It assists in identifying issues such as missing values, duplicate entries, and inconsistencies that could affect model performance.
- **Determining Preprocessing Requirements:** Through EDA, one can decide on necessary data cleaning or transformation steps like normalization, encoding, or imputation.

- **Facilitating Insightful Communication:** Visualizations generated during EDA help in conveying complex data insights clearly to stakeholders and team members.

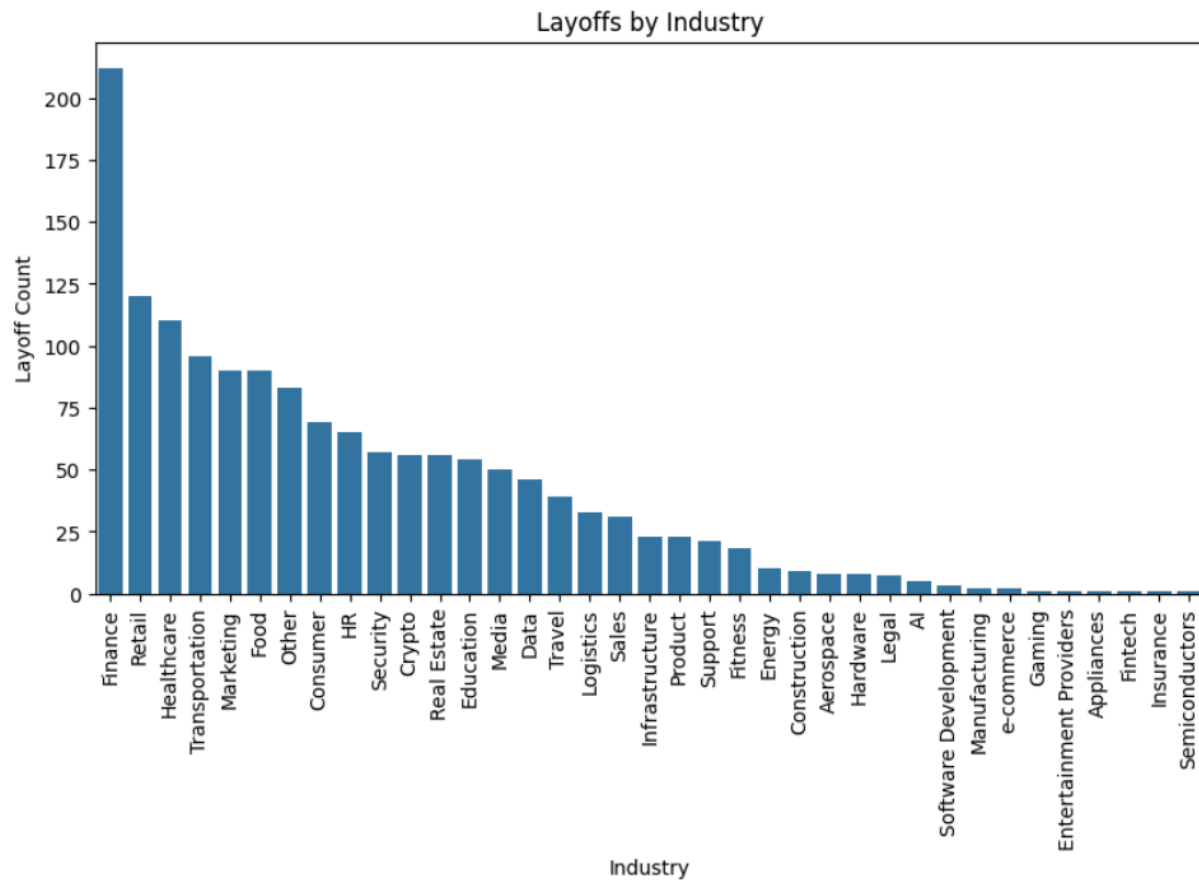
Common Techniques Used in EDA:

To effectively explore and interpret the dataset, a range of statistical and visualization tools are used:

- **Histograms and Frequency Distributions:** These help in understanding the distribution of numeric variables, such as whether the data is skewed or normally distributed.
- **Box Plots (Box-and-Whisker Plots):** Useful for detecting outliers and visualizing the spread and central tendency of data through quartiles.
- **Scatter Plots:** Ideal for examining the relationship or potential correlation between two continuous variables.
- **Heatmaps:** These provide a color-coded matrix that makes it easy to identify strong or weak correlations among multiple features in the dataset.
- **Bar Charts and Pie Charts:** These are effective for summarizing and comparing categorical data, offering a clear view of frequency and proportions across categories.

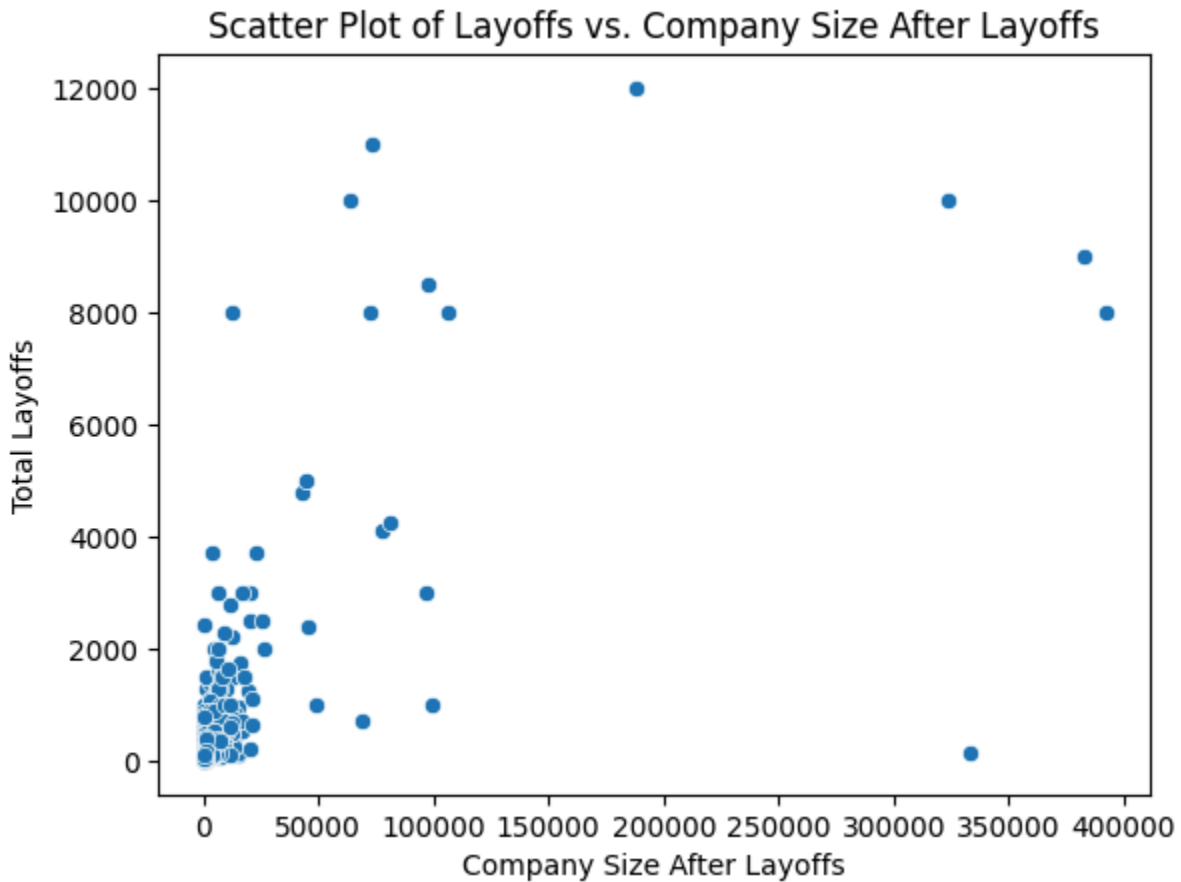
1)Bar Graph

```
plt.figure(figsize=(10, 5))
sns.barplot(x=df['Industry'].value_counts().index, y=df['Industry'].value_counts().values)
plt.xticks(rotation=90)
plt.xlabel('Industry')
plt.ylabel('Layoff Count')
plt.title('Layoffs by Industry')
plt.show()
```



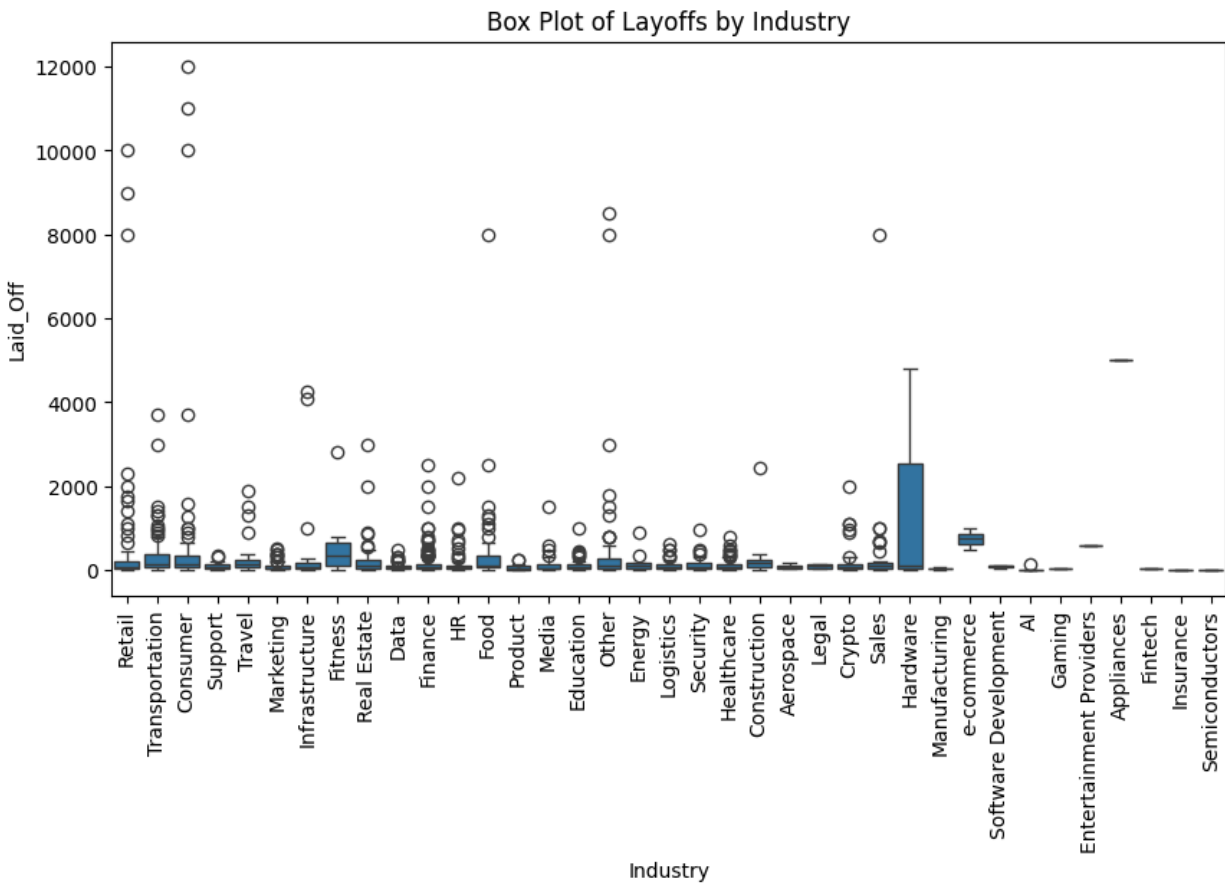
2)Scatter plot

```
▶ sns.scatterplot(x=df['Company_Size_after_layoffs'], y=df['Laid_Off'])  
plt.xlabel('Company Size After Layoffs')  
plt.ylabel('Total Layoffs')  
plt.title('Scatter Plot of Layoffs vs. Company Size After Layoffs')  
plt.show()
```



3)Box Plot

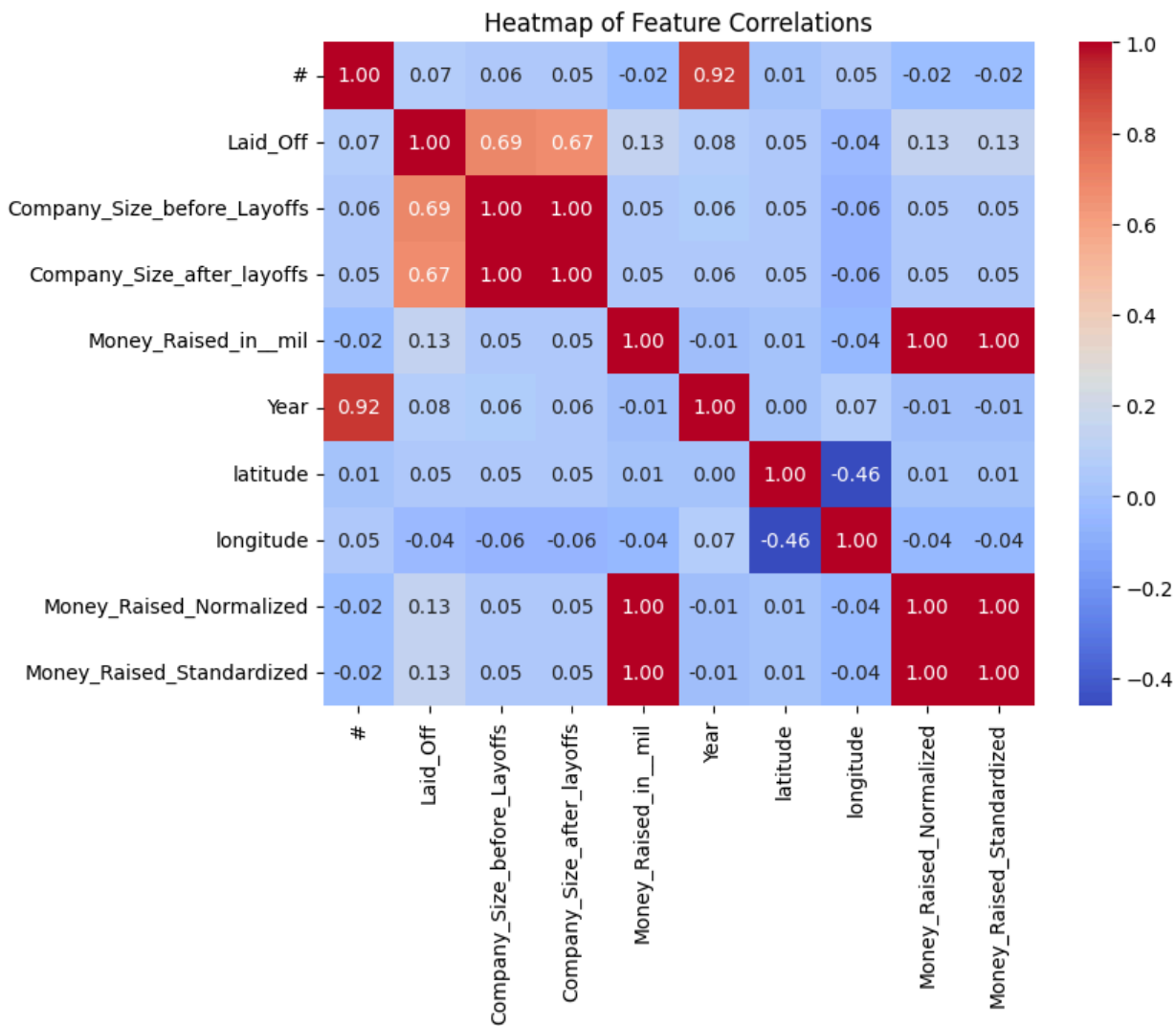
```
[ ] plt.figure(figsize=(10, 5))
sns.boxplot(x='Industry', y='Laid_Off', data=df)
plt.xticks(rotation=90)
plt.title('Box Plot of Layoffs by Industry')
plt.show()
```



4)Heat Map

```
numeric_df = df.select_dtypes(include=['number'])

plt.figure(figsize=(8, 6))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Heatmap of Feature Correlations")
plt.show()
```



CONCLUSION :

Exploratory Data Analysis (EDA) is a vital step in the data analysis process, acting as the bridge between raw data and meaningful insights. Through a combination of statistical summaries and visualizations, EDA allows analysts to understand the structure, distribution, and relationships within the data. It plays a crucial role in identifying trends, detecting anomalies, and assessing data quality. By thoroughly examining the data, EDA guides important decisions about preprocessing steps like handling missing values, encoding categorical data, or scaling numerical features—ultimately shaping the direction of further analysis or machine learning modeling.

Beyond technical preparation, EDA also facilitates effective communication of findings. The visual tools used in EDA—such as histograms, box plots, scatter plots, and heatmaps—translate complex data patterns into easily understandable formats for stakeholders and teams. This not only enhances transparency but also supports data-driven decision-making. In summary, EDA is more than just an analytical phase—it's a mindset that emphasizes curiosity, critical thinking, and a deep connection with the data. Mastering EDA equips analysts with the foundation needed to build accurate models and derive meaningful, actionable insights.