DATA SCIENCE ASSIGNMENT

TASK-1:

Exploratory Data Analysis (EDA) and Business Insights:

What is EDA:

Before applying machine learning or other statistical models, a dataset is examined and summarized to extract relationships, patterns, trends, and anomalies. Exploratory data analysis (EDA) is the term used for this process. Both graphic and non-graphic methods are used to understand the nature and quality of the structuring data.

Primary Goals of EDA:

Understanding Data: Recognize data distribution, structure, and categories.

Issue Detection: Identify missing values and outliers, as well as other inconsistencies.

Patterns Evaluation: Identify trends, relationships, and various correlations.

Data Preparation: Clean the data and apply some pre-processing for further modeling or analysis.

Types of EDA (Exploratory Data Analysis):

According to the ways it is performed and data analysis methodologies used, there are a few different types by which EDA is classified. So, the ones characterized by these categories helps in efficiently collating the data from different frame angles. Here are the main types:

1. Univariate Analysis

An analysis of only one variable to study its distribution, central tendency, and spread of variation.

Goal: Summarize and describe the variable.

Techniques required:

Descriptive statistics like calculating mean, median, mode, standard deviation, etc.

Frequency distribution as well as histograms can be used for the numerical data.

To know about categorical data, bar charts, and pie charts may be used.

Box plots for searching for outliers can be used.

Example: The distribution of the age of customers.

2. Bivariate Analysis

This is the type of analysis involved in understanding the relationship between two variables, whether numeric or nonnumeric.

Goal: Expose relationships, trends, and correlations.

Techniques involved:

Scatter plots: Numbers vs. numbers

An index such as Pearson's or Spearman's correlation can be calculated.

Box plots and violin plots for numbers vs. categories

Cross-tabulation and chi-square tests for categories vs. categories

Example: How are sign-up rates varying across these regions?

3. Multivariate Analysis

In contrast, multivariate analysis involves the observation of more than two variables, simultaneously interpreting and unraveling complicated relationships and patterns.

Goal: Detect patterns and dependencies among many variables.

Techniques:

Seaborn.pairplot in Python and FAT.

Heatmaps to study the correlations between numerical variables.

Drop down a dimension with a PCA.

Can be handled by clustering techniques like K-means.

Example: How do signups change with regions and weekdays?

Graphical Analysis :

This type of exploratory data analysis (EDA) relies heavily on visualizations to uncover data patterns, identify outliers, and examine distributions.

Objective: Provide an intuitive understanding of the data.

Techniques Used:

- Histograms, bar plots, and line plots.

- Box plots and violin plots to illustrate spread and outliers.

- Heatmaps to show correlations and relationships.

- Pairwise scatter plots and bubble charts for multidimensional data.

Example: Using line plots to visualize customer growth trends over time.

5. Quantitative Analysis

This aspect focuses on numerical computations and statistical summaries to assess data properties.

Objective: Summarize numerical data effectively.

Techniques Used:

- Measures of central tendency: Mean, median, mode.

- Measures of dispersion: Variance, standard deviation, range.

- Hypothesis testing and p-values to validate assumptions.

Example: Calculating the average number of signups per region.

Why is EDA Important?

 Data Understanding: Offers a thorough insight into the dataset and its characteristics.

Data Quality Assessment: Detects missing values, duplicates, and anomalies that require attention.

 Feature Selection: Aids in selecting the most pertinent features for predictive modeling.

Hypothesis Generation: Supports the development of hypotheses for additional testing.

Improved Decision-Making: Facilitates informed choices based on insights derived from data.

Error Prevention: Helps prevent mistakes during analysis or modeling by revealing hidden issues early on.

Where is EDA Used?

EDA is widely used across industries and domains where data-driven decisions are required. Here are some applications:

1. Business and Marketing

Analyzing customer demographics, purchase behavior, and market trends.

Segmenting customers and identifying target audiences.

2. Finance

Detecting fraudulent transactions.

Analyzing stock price trends and market patterns.

3. Healthcare

Examining patient data for trends in diseases and treatments.

Understanding the effects of clinical trials or interventions.

4. E-commerce

Analyzing sales trends and product performance.

Identifying factors influencing cart abandonment.

5. Education

Understanding student performance and identifying areas for improvement.

Designing adaptive learning systems based on user interaction data.

Performs the EDA (Exploratory Data Analysis) operations on " Customers.csv":

Here is the link:

https://colab.research.google.com/drive/14YbJ5skQVAW3pN7ooYgs_da0JzRMBtlV?usp=sharing

The 5 Business Insights from the EDA(Exploratory Data Analysis):

1. Regional Customer Distribution :

The analysis indicates that a large portion of customers comes from regions like North America, Asia, and Europe, with North America leading in total signups. Conversely, South America and Africa have a much smaller customer base. This points to a significant opportunity in these underperforming areas, highlighting the need for tailored marketing strategies or promotional efforts to effectively reach these markets.

2. Seasonal Trends in Signups :

Signups tend to peak during the months of November and December, likely due to the holiday season when customers are more inclined to respond to promotional offers. Businesses should consider boosting their marketing budgets during this time to take advantage of increased customer engagement, offering special deals or discounts to enhance acquisition rates.

3.Weekly Signup Patterns:

Customer signups are notably higher on weekdays, particularly on Mondays and Tuesdays, compared to weekends. This trend suggests that customers are more

active during the workweek, possibly during breaks or while handling work-related tasks. Therefore, targeted marketing efforts, such as email newsletters or ads, should be concentrated on weekdays to maximize visibility and engagement.

4. Customer Growth Over Time :

The overall customer count shows a consistent upward trend, with significant increases in signups at specific times, such as the beginning of the year or around major holidays. These spikes demonstrate the success of marketing initiatives during these periods. To maintain this growth, businesses should investigate the elements contributing to these spikes and aim to replicate them in other periods, such as by launching quarterly campaigns or themed promotions.

5. Income-Driven Behavior

The upper 10 percent of the earners are more frequent and high ticket buyers adding a greater share in revenue.

 This is the Explanation: If customer segment is inclined towards high end products, customized services, and unique items. An effective way to keep high net clients engaged is creating vip packs, offering specialised services or even releasing new luxury branded products. On the other hand, lower income segments can be targeted with cost effective and budgeted products which allows higher sales volume and increase the sales revenue.