**Customer Purchase Behavior Analysis for Amazon**

Amazon, as one of the world’s largest e-commerce platforms, collects vast amounts of customer purchase data. The goal is to analyze customer purchase behavior, product performance, and customer demographics to optimize marketing strategies, improve product recommendations, and enhance the customer experience. The data is available in various formats: Excel for customer demographics, CSV for product information, and JSON for transaction details.

To address this business challenge, **Exploratory Data Analysis (EDA)** will be conducted to uncover trends, correlations, and patterns within the data. This EDA will serve as the foundation for decision-making, identifying areas where Amazon can maximize revenue, streamline operations, and improve customer satisfaction.

**Steps of EDA**

**1: Data Collection**

* **Action**: You begin by collecting data from different sources: Excel for customer demographics, CSV for product information, and JSON for transaction details. After loading the datasets, you merge them using common identifiers like Customer\_ID and Product\_ID.
* **Goal**: Consolidate all data into a unified format for analysis.

**2: Data Cleaning**

* **Action**: Upon inspection, you discover missing values and incorrect entries. You fill missing numerical values (e.g., Age, Purchase\_Amount) using median or mode imputation, and either drop or impute missing categorical data like Customer\_Name and Product\_Name.
* **Goal**: Ensure the dataset is complete, clean, and ready for further analysis.

**3: Data Transformation**

* **Action**: You transform categorical variables (e.g., Gender, Product\_Category) into numerical values using one-hot encoding and normalize continuous variables like Purchase\_Amount and Price to ensure they are on the same scale.
* **Goal**: Prepare the dataset for advanced analysis and ensure all data is properly encoded and scaled.

**4: Univariate Analysis**

* **Action**: You analyze individual features like Purchase Amount, Age, and Product Category. You generate histograms and box plots to visualize the distribution of these features and detect any outliers.
* **Goal**: Understand the spread and distribution of key variables.

**5: Bivariate Analysis**

* **Action**: You analyze the relationships between two features, such as comparing customer Age with Purchase Amount, or product Price with customer Rating. Scatter plots and correlation matrices are generated to examine relationships.
* **Goal**: Identify correlations and relationships between variables that affect customer purchasing behavior.

**6: Multivariate Analysis**

* **Action**: You examine how multiple variables interact, such as the effect of Discounts, Purchase Date, and Product Price on customer purchases. Through multivariate visualization techniques, you understand the influence of multiple factors on purchase decisions.
* **Goal**: Deepen your understanding of complex relationships and trends within the dataset.

**7: Feature Selection**

* **Action**: Using correlation analysis and statistical methods, you select the most important features (e.g., Age, Purchase Amount, Product Category) for deeper analysis. Unnecessary or redundant features are removed.
* **Goal**: Streamline the dataset and focus on features that have the most impact on customer purchase behavior.

**8: Visualization**

* **Action**: You create visualizations like heatmaps, bar charts, and line graphs to illustrate key findings, such as purchase trends over time or the impact of discounts on sales.
* **Goal**: Present your findings in a visually compelling way to make insights more accessible to stakeholders.

**9: Conclusion and Reporting**

* **Action**: After completing the analysis, you summarize the key insights. For example, high-spending customers may fall within a specific age range, or certain product categories might perform better in specific regions. Recommendations are provided based on the analysis, such as optimizing pricing strategies or targeting specific customer segments.
* **Goal**: Provide actionable insights to Amazon’s marketing, product, and sales teams to enhance decision-making.