

Exploratory Data Analysis (EDA) and Business Insights

VASU GAUR

25/01/2025

1.0 Steps

Import Libraries:

Import necessary libraries: pandas, numpy, matplotlib, and seaborn.

Load Data:

Load the datasets for customers, products, and transactions using `pd.read_csv()`.

Convert Date Columns:

Convert the `SignupDate` and `TransactionDate` columns to datetime format for proper handling of dates.

Merge DataFrames:

Merge the datasets (`customers_df`, `products_df`, `transactions_df`) on `CustomerID` and `ProductID` to create a single merged dataset (`merged_df`).

Check for Missing Data:

Use `isna().sum()` to check for any missing values in the merged dataset.

Verify Data Merging:

Display a sample of the merged dataset's `TransactionDate` and `SignupDate` columns to ensure the merge was successful.

Category vs Region Analysis:

Group data by `Category` and `Region`, count the number of transactions, and plot a stacked bar chart to visualize category orders by region.

Category vs Average Quantity:

Group by `Category` and calculate the mean quantity per category (rounded), then visualize it using a **bar plot**.

Region vs Average Total Value:

Group by `Region` and calculate the average transaction total value (`TotalValue`), then visualize it using a strip plot for precision.

Category vs Average Total Value:

Group by Category and calculate the average transaction value per category, then visualize it with a horizontal bar chart.

Category vs Average Price:

Group by Category from the products_df dataset and calculate the average price per category, then visualize it with a bar plot.

Price Distribution:

Use a boxplot to show the distribution of prices within each product category from products_df.

Recalculate and Display Insights:

Recalculate the same metrics (e.g., transactions per category by region, average quantity, average total value by region and category, average price per category).

Print out the results for review.

2.0 Business Insights

```
Number of Transactions per Category by Region:
Region      Asia  Europe  North America  South America
Category
Books       51    64      65              90
Clothing    58    47      60              63
Electronics 57    58      60              79
Home Decor  52    65      59              72
```

```
Average Quantity per Product Category (Rounded):
Category
Books       3.0
Clothing    3.0
Home Decor  3.0
Electronics 2.0
Name: Quantity, dtype: float64
```

```
Average Total Value of Transactions by Region:
Region
South America    721.554474
Europe           710.489872
Asia             697.591606
North America    624.235246
Name: TotalValue, dtype: float64
```

```
Average Total Value per Product Category:
Category
Clothing    728.818684
Electronics 711.746063
Books       711.657296
Home Decor  608.443266
Name: TotalValue, dtype: float64
```

```
Average Price per Product Category:
Category
Books       292.212308
Electronics 275.353462
Clothing    263.720000
Home Decor  235.020000
Name: Price, dtype: float64
```

1) Transactions per Category by Region:

Books had the highest number of transactions in South America (90), while Clothing had the least in Europe (47).

2) Average Quantity per Category:

All categories (Books, Clothing, Home Decor) had an average quantity of around 3, while electronics had a lower average of 2.

3) Average Total Value of Transactions by Region:

South America led with the highest average total value per transaction (\$721.55), followed closely by Europe (\$710.49).

4) Average Total Value per Category:

Clothing had the highest average total value per transaction (\$728.82), while Home Decor had the lowest (\$608.44).

5) Average Price per Category:

Books had the highest average price per item (\$292.21), followed by Electronics (\$275.35). Home Decor had the lowest average price (\$235.02).

