

EDA Report

Introduction

The exploratory data analysis (EDA) focuses on understanding the dataset's structure, patterns, and potential insights to inform further analysis and business decisions.

Data Overview

1. Customers.csv

Column	Description	Example
CustomerID	Unique identifier for each customer	C001
CustomerName	Name of the customer	John Doe
Region	Region where the customer resides	North America
SignupDate	Date when the customer signed up	2021-01-01

2. Products.csv

Column	Description	Example
ProductID	Unique identifier for each product	P001
ProductName	Name of the product	Laptop
Category	Product category	Electronics
Price	Product price in USD	999.99

3. Transactions.csv

Column	Description	Example
TransactionID	Unique identifier for each transaction	T001
CustomerID	ID of the customer who made the transaction	C001
ProductID	ID of the product sold	P001
TransactionDate	Date of the transaction	2021-06-15
Quantity	Quantity of the product purchased	2

Column	Description	Example
TotalValue	Total value of the transaction	1999.98

Findings

1. Customer Distribution by Region:

- Most customers are from North America.
- Asia-Pacific has the second highest number of customers.

2. Popular Product Categories:

- "Electronics" and "Furniture" are the top-selling categories.
- "Clothing" has lower sales but high transaction frequency.

3. Transaction Trends:

- Transaction volumes peak during holiday seasons (e.g., December).
- Average transaction value increases significantly during sales events.

4. High-Value Customers:

- Identified top 10% of customers contributing to 60% of revenue.
- These customers frequently purchase high-priced electronics.

5. Product Performance:

- Certain products in "Furniture" underperform despite high price points.
 - Opportunity to promote underperforming products in targeted regions.
-

Visualizations

Customer Distribution by Region

```
sns.countplot(data=customers, x="Region")
plt.title("Customer Distribution by Region")
plt.show()
```

- **Insight:** North America dominates customer distribution.

Top Product Categories by Revenue

```
category_revenue = merged_data.groupby('Category')['TotalValue'].sum().reset_index()
```

```
sns.barplot(data=category_revenue, x='Category', y='TotalValue')
```

```
plt.title('Total Revenue by Category')
```

```
plt.show()
```

- **Insight:** Electronics contribute the most to total revenue.

Conclusion

The EDA reveals actionable insights, such as targeting high-value customers and promoting underperforming products. Seasonal trends and regional preferences should guide marketing and inventory strategies. Further modeling will refine these observations and help drive business outcomes.