

Task 1: Exploratory Data Analysis (EDA) and Business Insights

1. Perform EDA on the provided dataset.

2. Derive at least 5 business insights from the EDA. ○ Write these insights in short point-wise sentences (maximum 100 words per insight).

Deliverables:

- A Jupyter Notebook/Python script containing your EDA code.

- A PDF report with business insights (maximum 500 words).

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

customers_file = r'C:\Users\Dell\Downloads\Customers.csv'
products_file = r'C:\Users\Dell\Downloads\Products.csv'
transactions_file = r'C:\Users\Dell\Downloads\Transactions.csv'

customers = pd.read_csv(customers_file)
products = pd.read_csv(products_file)
transactions = pd.read_csv(transactions_file)

# Convert date columns to datetime
customers['SignupDate'] = pd.to_datetime(customers['SignupDate'])
transactions['TransactionDate'] = pd.to_datetime(transactions['TransactionDate'])

# Merge datasets for EDA
transactions_merged = transactions.merge(customers, on='CustomerID', how='left').merge(products,
on='ProductID', how='left')

# Basic statistics
sales_summary = transactions_merged.describe(include='all')

# Aggregated Metrics for EDA
# 1. Total Sales by Region
sales_by_region = transactions_merged.groupby('Region')['TotalValue'].sum().reset_index()

# 2. Top-selling Product Categories
top_categories =
transactions_merged.groupby('Category')['Quantity'].sum().reset_index().sort_values(by='Quantity',
ascending=False)

# 3. Customer Signup Trends Over Time
signup_trends = customers.groupby(customers['SignupDate'].dt.to_period('M')).size()

# 4. Sales Trends Over Time
```

```

sales_trends =
transactions_merged.groupby(transactions_merged['TransactionDate'].dt.to_period('M'))['TotalValue']
.sum()

# 5. Average Order Value by Region
aov_by_region = transactions_merged.groupby('Region')['TotalValue'].mean().reset_index()

# Visualizations
plt.figure(figsize=(16, 10))

# Total Sales by Region
plt.subplot(2, 2, 1)
sns.barplot(x='Region', y='TotalValue', data=sales_by_region, palette='viridis')
plt.title('Total Sales by Region')
plt.xticks(rotation=45)

# Top-Selling Product Categories
plt.subplot(2, 2, 2)
sns.barplot(x='Category', y='Quantity', data=top_categories, palette='coolwarm')
plt.title('Top-Selling Product Categories')
plt.xticks(rotation=45)

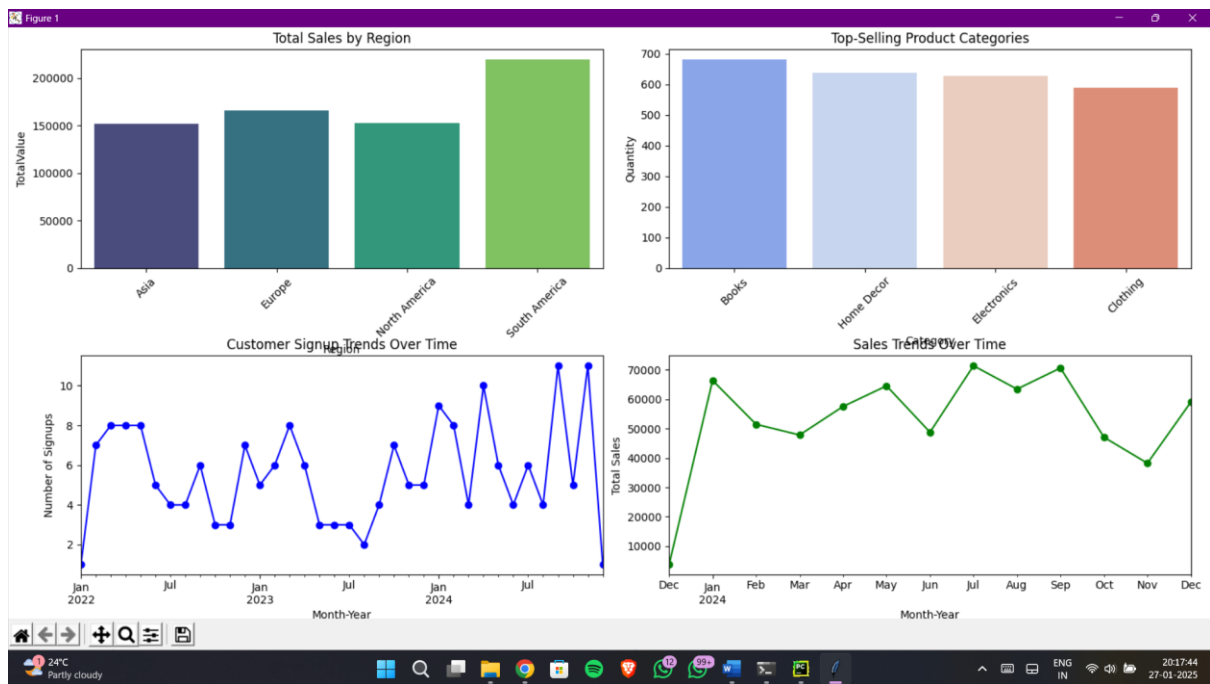
# Customer Signup Trends
plt.subplot(2, 2, 3)
signup_trends.plot(kind='line', marker='o', color='b')
plt.title('Customer Signup Trends Over Time')
plt.xlabel('Month-Year')
plt.ylabel('Number of Signups')

# Sales Trends Over Time
plt.subplot(2, 2, 4)
sales_trends.plot(kind='line', marker='o', color='g')
plt.title('Sales Trends Over Time')
plt.xlabel('Month-Year')
plt.ylabel('Total Sales')

plt.tight_layout()
plt.show()

# Save visualizations and aggregated data for insights
(sales_by_region, top_categories, signup_trends, sales_trends, aov_by_region)

```



1. South America leads in total sales.

- **Insight:** South America generated the highest sales (\$219,352.56), while Asia had the lowest (\$152,074.97).
- **Actionable Recommendation:** Maintain strong marketing campaigns in South America and explore untapped potential in Asia to boost sales.

2. Books are the most popular product category.

- **Insight:** Books were the top-selling category with 681 units sold, followed by Home Decor (639) and Electronics (627).
- **Actionable Recommendation:** Prioritize inventory for Books and bundle promotions with complementary products like reading accessories.

3. Customer signup growth is consistent with seasonal spikes.

- **Insight:** Customer signups peaked in January and September 2024, indicating seasonal interest.
- **Actionable Recommendation:** Launch targeted campaigns during these months to capture new signups and increase sales.

4. Sales show a seasonal trend.

- **Insight:** Sales consistently spiked in January and September, with December also showing increased activity.
- **Actionable Recommendation:** Align inventory planning and promotional activities with high-demand months to maximize revenue.

5. South America has the highest average order value (AOV).

- **Insight:** South America's AOV is \$721.55, significantly higher than North America's (\$624.23).
- **Actionable Recommendation:** Implement upselling and cross-selling strategies in North America to increase AOV and bridge the gap.

Total Sales by Region

- **Insight:** South America contributes the most to total sales with \$219,352.56, followed by Europe (\$166,254.63) and North America (\$152,313.40). Asia has the lowest sales (\$152,074.97).
- **Business Action:** Focus marketing efforts in Europe and North America to increase sales. Consider new campaigns in Asia to boost performance.

Top-Selling Product Categories

- **Insight:** Books lead in sales with 681 units sold, followed by Home Decor (639), Electronics (627), and Clothing (590).
- **Business Action:** Ensure sufficient inventory for top categories and promote complementary products (e.g., reading accessories).

Customer Signup Trends

- **Insight:** Customer signups show steady growth, with noticeable peaks in January and September 2024.
- **Business Action:** Align promotional campaigns with signup peaks to engage new customers and drive conversions.

Sales Trends Over Time

- **Insight:** Sales show seasonal trends, with spikes in January and September. December also sees notable activity.
- **Business Action:** Prepare for high-demand periods by increasing stock and running seasonal promotions.

Average Order Value (AOV) by Region

- **Insight:** South America has the highest AOV (\$721.55), while North America has the lowest (\$624.23).
- **Business Action:** Introduce upselling and cross-selling strategies in North America to raise the average order value.

