Key Metrics and Insights: 1.Sales Trends(Month-wise, Year-wise, Year-Month-wise) to track overall business growth. 2.Top Performing Item Types to identify the best-selling categories. 3.Regional and Channel Analysis to optimize sales strategies in various markets and platforms. 4.Sales Performance by Country: Identify top 10 countries contributing the most revenue. **5.Profitability Analysis**Compare regions not only by total revenue but also by profit margins. 6.Seasonality Analysis: Find out if there are any seasonal trends affecting sales and profitability. 7.Price per Unit Sold Analysis: Add a new metric Price per Unit Sold and analyze it by item type. 8.Outliers Detection: Detect and visualize outliers in sales and cost data using boxplots. In [1]: import pandas as pd import numpy as np import matplotlib.pyplot as plt import seaborn as sns import plotly.express as px import plotly.graph_objects as go import warnings # Suppress warnings warnings.simplefilter(action='ignore', category=FutureWarning) # Load the dataset file_path = 'C:/Users/User/Downloads/Amazon Sales data.csv' amazon_sales_data = pd.read_csv(file_path) # Replace inf values with NaN in the entire DataFrame amazon_sales_data.replace([np.inf, -np.inf], np.nan, inplace=True) # Convert 'Order Date' and 'Ship Date' columns to datetime format amazon_sales_data['Order Date'] = pd.to_datetime(amazon_sales_data['Order Date'], errors='coerce') amazon_sales_data['Ship Date'] = pd.to_datetime(amazon_sales_data['Ship Date'], errors='coerce') # Extract year, month, and year-month from 'Order Date' amazon_sales_data['Order Year'] = amazon_sales_data['Order Date'].dt.year amazon_sales_data['Order Month'] = amazon_sales_data['Order Date'].dt.month amazon_sales_data['Year_Month'] = amazon_sales_data['Order Date'].dt.to_period('M') # Handle missing values in key columns amazon_sales_data.dropna(subset=['Total Revenue'], inplace=True) # Basic Analysis # Group by month-wise, year-wise, and year-month-wise for sales trends month_wise_sales = amazon_sales_data.groupby('Order Month')['Total Revenue'].sum() year_wise_sales = amazon_sales_data.groupby('Order Year')['Total Revenue'].sum() year_month_wise_sales = amazon_sales_data.groupby('Year_Month')['Total Revenue'].sum() # Group by Item Type to analyze the relationship with Total Revenue and Profit item_type_metrics = amazon_sales_data.groupby('Item Type').agg({ 'Total Revenue': 'sum', 'Total Profit': 'sum', 'Units Sold': 'sum' }).sort_values(by='Total Revenue', ascending=False) # Group by Region and Sales Channel to analyze their impact on Total Revenue region_sales_channel_metrics = amazon_sales_data.groupby(['Region', 'Sales Channel']).agg({ 'Total Revenue': 'sum', 'Total Profit': 'sum', 'Units Sold': 'sum' }).sort_values(by='Total Revenue', ascending=False) # Visualization 1: Month-wise Sales Trend plt.figure(figsize=(10,6)) sns.barplot(x=month_wise_sales.index, y=month_wise_sales.values, palette="Blues_d") plt.title('Month-wise Sales Trend', fontsize=16) plt.xlabel('Month', fontsize=12) plt.ylabel('Total Revenue', fontsize=12) plt.show() # Visualization 2: Year-wise Sales Trend plt.figure(figsize=(10,6)) sns.barplot(x=year_wise_sales.index, y=year_wise_sales.values, palette="Greens_d") plt.title('Year-wise Sales Trend', fontsize=16) plt.xlabel('Year', fontsize=12) plt.ylabel('Total Revenue', fontsize=12) plt.show() # Visualization 3: Top 5 Item Types by Total Revenue top_5_items = item_type_metrics.head(5) plt.figure(figsize=(10,6)) sns.barplot(x=top_5_items.index, y=top_5_items['Total Revenue'], palette="Oranges_d") plt.title('Top 5 Item Types by Total Revenue', fontsize=16) plt.xlabel('Item Type', fontsize=12) plt.ylabel('Total Revenue', fontsize=12) plt.show() # Visualization 4: Region and Sales Channel-wise Revenue plt.figure(figsize=(12,6)) region_sales = region_sales_channel_metrics.reset_index() sns.barplot(x='Region', y='Total Revenue', hue='Sales Channel', data=region_sales, palette="coolwarm") plt.title('Region and Sales Channel-wise Revenue', fontsize=16) plt.xlabel('Region', fontsize=12) plt.ylabel('Total Revenue', fontsize=12) plt.legend(title='Sales Channel') plt.show() # Visualization 5: Donut Chart for Item Type Distribution labels = amazon_sales_data['Item Type'].value_counts().index values = amazon_sales_data['Item Type'].value_counts().values fig = go.Figure(data=[go.Pie(labels=labels, values=values, hole=.4)]) fig.update_layout(title_text="Amazon Sales Product Distribution (Donut Chart)") fig.show() # Visualization 6: Bubble Chart for Revenue vs Cost by Item Type fig = px.scatter(amazon_sales_data, x="Total Revenue", y="Total Cost", size="Units Sold", color="Item Type", hover_name="Item Type", title="Bubble Chart: Revenue vs. Cost by Item Type") fig.show() # Visualization 7: Heatmap of Total Revenue by Region and Item Type plt.figure(figsize=(12,8)) heatmap_data = amazon_sales_data.pivot_table(values='Total Revenue', index='Region', columns='Item Type', aggfunc='sum') sns.heatmap(heatmap_data, annot=True, cmap='YlGnBu', fmt='.1f') plt.title('Heatmap of Total Revenue by Region and Item Type', fontsize=16) plt.xlabel('Item Type', fontsize=12) plt.ylabel('Region', fontsize=12) plt.show() # Visualization 8: Area Chart for Revenue over Time by Item Type fig = px.area(amazon_sales_data, x='Order Date', y='Total Revenue', color='Item Type', title="Area Chart: Revenue over Time by Item Type") iig.show() # Visualization 9: Choropleth Map for Total Revenue by Country fig = px.choropleth(amazon_sales_data, locations="Country", locationmode="country names", color="Total Revenue", hover_name="Country", color_continuous_scale=px.colors.sequential.Plasma, title="Choropleth Map: Total Revenue by Country") fig.show() # Visualization 10: Correlation Matrix of Numerical Variables plt.figure(figsize=(10,6)) correlation_matrix = amazon_sales_data[['Units Sold', 'Total Revenue', 'Total Cost', 'Total Profit']].corr() sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', linewidths=0.5) plt.title('Correlation Matrix of Numerical Variables', fontsize=16) plt.show() # Additional Insights # 1. Top 10 Countries by Total Revenue top_10_countries = amazon_sales_data.groupby('Country')['Total Revenue'].sum().nlargest(10) plt.figure(figsize=(10,6)) sns.barplot(x=top_10_countries.index, y=top_10_countries.values, palette='Blues_d') plt.title('Top 10 Countries by Total Revenue', fontsize=16) plt.xlabel('Country', fontsize=12) plt.ylabel('Total Revenue', fontsize=12) plt.xticks(rotation=45) plt.tight_layout() plt.show() # 2. Profitability Analysis - Profit Margin by Region amazon_sales_data['Profit Margin'] = (amazon_sales_data['Total Profit'] / amazon_sales_data['Total Revenue']) * 100 region_profit_margin = amazon_sales_data.groupby('Region')['Profit Margin'].mean().sort_values(ascending=False) plt.figure(figsize=(10,6)) sns.barplot(x=region_profit_margin.index, y=region_profit_margin.values, palette='Greens_d') plt.title('Average Profit Margin by Region', fontsize=16) plt.xlabel('Region', fontsize=12) plt.ylabel('Profit Margin (%)', fontsize=12) plt.xticks(rotation=45) plt.tight_layout() plt.show() # 3. Seasonality Analysis - Month-wise Average Revenue and Profit seasonality_data = amazon_sales_data.groupby('Order Month').agg({ 'Total Revenue': 'mean', 'Total Profit': 'mean' }).reset_index() plt.figure(figsize=(10,6)) sns.lineplot(x='Order Month', y='Total Revenue', data=seasonality_data, label='Average Revenue', color='blue', marker='o') sns.lineplot(x='Order Month', y='Total Profit', data=seasonality_data, label='Average Profit', color='green', marker='o') plt.title('Seasonality: Month-wise Average Revenue and Profit', fontsize=16) plt.xlabel('Month', fontsize=12) plt.ylabel('Amount (in USD)', fontsize=12) plt.legend() plt.tight_layout() plt.show() # 4. Price per Unit Sold Analysis by Item Type amazon_sales_data['Price per Unit Sold'] = amazon_sales_data['Total Revenue'] / amazon_sales_data['Units Sold'] price_per_unit_sold = amazon_sales_data.groupby('Item Type')['Price per Unit Sold'].mean().sort_values(ascending=False) plt.figure(figsize=(10,6)) sns.barplot(x=price_per_unit_sold.index, y=price_per_unit_sold.values, palette='Oranges_d') plt.title('Average Price per Unit Sold by Item Type', fontsize=16) plt.xlabel('Item Type', fontsize=12) plt.ylabel('Price per Unit Sold (in USD)', fontsize=12) plt.xticks(rotation=45) plt.tight_layout() plt.show() # 5. Outlier Detection - Boxplot for Total Revenue and Total Cost plt.figure(figsize=(12,6)) plt.subplot(1, 2, 1) sns.boxplot(y=amazon_sales_data['Total Revenue'], color='lightblue') plt.title('Boxplot of Total Revenue', fontsize=14) plt.ylabel('Total Revenue (in USD)', fontsize=12) plt.subplot(1, 2, 2) sns.boxplot(y=amazon_sales_data['Total Cost'], color='lightgreen') plt.title('Boxplot of Total Cost', fontsize=14) plt.ylabel('Total Cost (in USD)', fontsize=12) plt.tight_layout() plt.show() Month-wise Sales Trend 1e7 2.5 2.0 Total Revenue 0.5 5 8 10 9 11 Month Year-wise Sales Trend 1e7 3.0 2.5 1.0 -0.5 2011 2012 2013 2014 2015 2010 2016 2017 Year Top 5 Item Types by Total Revenue 1e7 3.5 3.0 2.5 1.0 0.5 0.0 Baby Food Household Office Supplies Clothes Cosmetics Item Type Region and Sales Channel-wise Revenue 1e7 Sales Channel 2.5 Offline Online 2.0 Total Revenue 0.5 Europe Australia and Octetatolike East and Octeontthal Affricerica and the Caribboerain America Sub-Saharan Africa Asia Region O iiii Amazon Sales Product Distribution (Donut Chart) Clothes Cosmetics Office Supplies Fruits Personal Care Household Baby Food Cereal Vegetables 7% Snacks 10% Meat 10% Bubble Chart: Revenue vs. Cost by Item Type 5M Item Type Baby Food Cereal 4M Office Supplies Fruits Household ЗМ Vegetables **Total Cost** Personal Care Clothes Cosmetics Beverages Meat Snacks 1M 2M Total Revenue Heatmap of Total Revenue by Region and Item Type 1e7 1551011**3**039414.458471.1 072701 630878.<mark>400558.7</mark> 594055.4 Asia -- 1.2 Australia and Oceania 3292856.890541.6140287.4082418.4220728.870768.0 24925261B04138.0 - 1.0 Central America and the Caribbean -5997055.0 523807.6339490.5 387002.2576782.8600821.4745426.0 - 0.8 Region Europe 5638369.464250.8 1<mark>287755.</mark>3159720 3204354.6 151454.<mark>436683.4</mark> 26344.3 - 0.6 Middle East and North Africa -835759.1529920.0324478.46454.1 1316095.4 - 0.4 North America -4647149.6 996207.0 - 0.2 Sub-Saharan Africa 1419101. 949000.0770069 16735366 5111742. 4290788. 129684512011149. 0582813. 623648 12741243 20468657.4 Cosmetics Snacks Clothes Item Type Area Chart: Revenue over Time by Item Type Item Type 6M Baby Food Cereal 5M Office Supplies Fruits ---- Household 4M Vegetables Total Revenue Personal Care Clothes Cosmetics Beverages ---- Meat 2M —— Snacks 2014 2017 Order Date Choropleth Map: Total Revenue by Country Total Revenue Correlation Matrix of Numerical Variables 0.45 0.37 0.56 - 0.9 - 0.8 0.45 0.98 - 0.7 0.37 0.98 0.8 - 0.6 - 0.5 0.56 0.8 Total Profit Total Cost Units Sold Total Revenue Top 10 Countries by Total Revenue 1e6 Total Revenue Country Average Profit Margin by Region 40 35 Profit Margin (%)
25
15 10 5 -Region Seasonality: Month-wise Average Revenue and Profit −●− Average Profit 2.0 Amount (in USD) 0.5 10 12 Month Average Price per Unit Sold by Item Type 700 600 Price per Unit Sold (in USD) 100 Item Type Boxplot of Total Revenue **Boxplot of Total Cost** 1e6 0 0 0 Total Revenue (in USD) Total Cost (in USD) ™ 0 -0 -Conclusion and Insights from Amazon Sales Data Analysis 1. Overall Sales Trends Month-wise and Year-wise Growth: Significant fluctuations in sales were observed, with peak sales in February and April, likely due to seasonal promotions. The overall revenue shows a consistent upward trend year-over-year, indicating growth in Amazon's e-commerce business. Seasonal Trends: Higher sales during November and December suggest the impact of holiday shopping, highlighting the importance of planning marketing strategies around these periods.

2.Item Type Performance

3.Regional Insights

4.Profitability Analysis

5.Customer Behavior Insights

8. Strategic Recommendations

Conclusion

6. Visualization Insights

Top Selling Items: Analysis reveals Cosmetics as the leading category in total revenue and profit, indicating strong demand and potentially high margins. Office Supplies and Household Items also contribute significantly to overall sales.

Sales by Region: Sales performance varies across different regions, with Sub-Saharan Africa (offline) and Europe (offline) generating substantial revenue. This indicates a preference for traditional retail in areas where e-commerce infrastructure is less developed.

Heatmaps and Correlation Matrices: Visual representations, such as heatmaps of total revenue by region and item type, reveal performance dynamics. The correlation matrix highlights relationships between units sold, total revenue, total cost, and total profit.

The analysis highlights critical sales trends, regional preferences, and item type performance within Amazon's dataset. Notable opportunities for growth exist in specific item categories and regions, particularly where offline channels dominate. Understanding these patterns will aid in

Profit Contribution: Some item types generate high revenue but have low-profit margins, emphasizing the need for a detailed evaluation of profitability by category.

Profit Margins: Average profit margins differ significantly by region, suggesting that regions with higher margins may employ better cost management or premium pricing strategies.

Revenue vs. Profit Analysis: Examining the relationship between total revenue and total profit helps identify areas for cost optimization and potential adjustments in pricing strategies.

Sales Channel Effectiveness: The effectiveness of different sales channels varies by region, informing targeted marketing and sales strategies.

Price Sensitivity: Analysis of average price per unit sold across item types reveals varying customer willingness to pay, guiding future pricing strategies.

Revenue and Cost Outliers: Boxplots identify outliers in total revenue and total cost, indicating potential errors or unique circumstances requiring further investigation..

Marketing and Sales Strategies: Tailored marketing campaigns can be developed for peak seasons or underperforming regions based on seasonal trends and regional performance.

Units Sold: A strong positive correlation exists between units sold and total revenue, indicating that higher unit sales directly drive revenue

Product Development: Focus on expanding inventory for top-selling item types while re-evaluating those with low performance.

refining sales strategies, optimizing marketing efforts, and enhancing distribution channels, ultimately driving sustained growth for Amazon.

Cost Management: Identify and mitigate areas with low-profit margins to enhance overall profitability.

Analyzing Amazon Sales data

Problem Statement: Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce costs and increase profits. Sales management today is the most important

function in a commercial and business enterprise. The objective of this project is to analyze sales trends, top-selling items, regional performance, and sales channels to gain insights that can improve decision-making.