

Exploratory Data Analysis Report

Participant Name: Mohammed Shakawat Hossen

Competition Name: DataWar

Date of Submission: 25/12/2024

Exploratory Data Analysis Report

1. Data Overview

Key Insights:

- **Unique Products:** There are **20** unique products in the dataset.
- **Order Date Range:** The data spans from "2019-01-01 03:07:00" To "2020-01-01 05:13:00" indicating a comprehensive time frame for sales activity.
- **Missing or Null Values:** Missing values were detected in multiple columns, but these were handled by removing rows with missing values.
- **Duplicate Rows:** There were **618** duplicate rows in the dataset, which were removed to ensure the data's accuracy.

Recommendations:

- Ensure future data sources are consistently cleaned to prevent missing or duplicate values.
 - Maintain a record of cleaning steps to enhance reproducibility.
-

2. Basic Statistics

Key Insights:

- **Average Quantity Ordered:** The average quantity ordered per product was **1.0603304866645662**
- **Total Revenue:** The total revenue generated from all sales amounted to **34465537.94**
- **Price Statistics:**
 - Minimum Price: **2.99**
 - Maximum Price: **1700.0**
 - Average Price: **184.51925546352444**

Recommendations:

- Focus on products with higher quantities ordered and lower prices for potential promotions.
 - Monitor products with high prices and low quantities to identify whether they are underperforming.
-

3. Data Cleaning

Key Insights:

- The **Order Date** column was cleaned by converting non-datetime values to valid formats and removing invalid dates.
- The **Purchase Address** was split to extract city names for further analysis.

Recommendations:

- Regularly clean and validate data, especially in columns related to dates and addresses, to improve accuracy in reporting and analysis.
-

4. Sales Performance Analysis

Key Insights:

- **Best Month for Sales:** The month with the highest sales revenue was **Month 12**, with total revenue of **4608295.7**
- **Best Time for Advertisement:** The hour with the highest number of orders was around **19:00**, with **14451.0** orders.
- **Most Productive City:** The city with the highest sales revenue was **'San Francisco, CA'** generating **8254743.55** in revenue.

Recommendations:

- Increase targeted marketing efforts and advertisements in **month 12** and around **19:00** to maximize sales.
 - Focus promotional efforts on cities with the highest sales revenue, especially **'San Francisco, CA'**
-

5. Product-Specific Analysis

Key Insights:

- **Most Sold Products:** The top 5 most sold products by quantity are :
- **AAA Batteries (4-pack)** **30986.0**

- **AA Batteries (4-pack)** **27615.0**
- **USB-C Charging Cable** **23931.0**
- **Lightning Charging Cable** **23169.0**
- **Wired Headphones** **20524.0**

- **Product Pairings:** Popular product pairs sold together include **5**. These insights could lead to potential bundling strategies.

Recommendations:

- Bundle the most frequently sold products together to increase sales through cross-selling.
- Investigate the reasons behind the high sales of certain products, such as seasonal demand or pricing strategies.

6. Customer Insights

Key Insights:

- **Most Active Customers:** The top 10 most active customers based on revenue generated are **[top_10_customers]**.

	Purchase Address	total_purchases
• 89038	668 Park St, San Francisco, CA 94016	4.0
• 45621	391 1st St, Seattle, WA 98101	3.0
• 108753	795 1st St, Atlanta, GA 30301	3.0
• 98849	731 11th St, New York City, NY 10001	6.0
• 17178	208 Chestnut St, San Francisco, CA 94016	3.0
• 132927	949 Hickory St, New York City, NY 10001	3.0
• 80316	611 Wilson St, San Francisco, CA 94016	6.0
• 80070	610 14th St, Los Angeles, CA 90001	3.0
• 24670	256 Hill St, San Francisco, CA 94016	4.0
• 109078	797 Dogwood St, San Francisco, CA 94016	4.0

- **Top 10% Customer Revenue Contribution:** The top 10% of customers contribute **49.95%** of the total revenue, indicating a highly concentrated customer base.
 - This is determined by calculating the revenue of the top 10% customers and comparing it to the total revenue.
- **City-Wise Product Preferences:** Different cities show varied preferences for products, such as
- City-Wise Product Preferences:
- City Product total_quantity_sold total_revenue

- 5 Atlanta, GA AAA Batteries (4-pack) 2358.0 7050.42
- 4 Atlanta, GA AA Batteries (4-pack) 2193.0 8421.12
- 15 Atlanta, GA USB-C Charging Cable 1912.0 22848.40
- 12 Atlanta, GA Lightning Charging Cable 1874.0 28016.30
- 17 Atlanta, GA Wired Headphones 1576.0 18896.24
-
- 185 Seattle, WA ThinkPad Laptop 332.0 331996.68
- 171 Seattle, WA 20in Monitor 324.0 35636.76
- 187 Seattle, WA Vareebadd Phone 179.0 71600.00
- 182 Seattle, WA LG Washing Machine 58.0 34800.00
- 181 Seattle, WA LG Dryer 56.0 33600.00
-

Recommendations:

- Target the top 10% of customers with personalized offers to encourage loyalty and increase sales.
- Use city-specific product preferences to tailor marketing and product offerings to local tastes.

7. Time-Based Trends

Key Insights:

- **Weekly Sales Trends:** The sales trend over the weeks
- Weekly Sales Trends:
 - Week total_sales
 - 0 1 636646.16
 - 1 2 409389.43
 - 2 3 394921.11
 - 3 4 426020.17
 - 4 5 459570.62
 - 5 6 565300.05

- 6 7 568199.28
- 7 8 535207.51
- 8 9 578306.44
- 9 10 640496.38
- 10 11 595771.07
- 11 12 629060.55
- 12 13 670872.96
- 13 14 774492.97
- 14 15 761454.22
- 15 16 805297.45
- 16 17 808694.68
- 17 18 764462.62
- 18 19 729309.99
- 19 20 695631.88
- 20 21 709784.55
- 21 22 668240.16
- 22 23 616368.48
- 23 24 594043.83
- 24 25 611609.50
- 25 26 576724.03
- 26 27 603246.53
- 27 28 609980.59
- 28 29 615306.48
- 29 30 562283.15
- 30 31 543200.60
- 31 32 550560.31
- 32 33 492928.48
- 33 34 490216.33
- 34 35 487735.61

- 35 36 478627.99
- 36 37 537174.09
- 37 38 434557.29
- 38 39 504087.68
- 39 40 798896.68
- 40 41 826120.83
- 41 42 846519.54
- 42 43 847292.99
- 43 44 810632.55
- 44 45 746789.21
- 45 46 712967.26
- 46 47 775066.70
- 47 48 788596.13
- 48 49 1047118.24
- 49 50 1038000.94
- 50 51 1074422.49
- 51 52 1017333.16

- **Busiest Days of the Week:** The busiest days for sales are

- **Busiest Days of the Week for Sales:**

- **Weekday total_sales**

- 1 1 5086275.40
- 2 2 4986823.36
- 6 6 4927249.40
- 5 5 4900195.58
- 0 0 4877588.21
- 4 4 4853642.26
- 3 3 4833763.73

- **Recommendations:**
 - Consider running promotions on the busiest days of the week to maximize sales.
 - Plan weekly marketing campaigns based on sales patterns, focusing on the most lucrative weeks.
-

Conclusion

Actionable Recommendations:

1. **Marketing Strategy Optimization:** Focus advertising efforts during peak sales months and hours, such as **best_month** and **best_hour**
2. **Product Bundling:** Leverage product pair analysis to create effective bundles for cross-selling opportunities.
3. **Target High-Value Customers:** Prioritize personalized offers for top customers to increase repeat business and revenue.
4. **Localized Marketing:** Tailor promotions and product offerings based on city-specific preferences and trends.
5. **Optimize Sales Timing:** Focus on the busiest days of the week and peak sales periods to run special offers.
Here are additional actionable recommendations based on the analysis:
6. **Seasonal Promotions:** Leverage seasonal trends identified in the data to create targeted campaigns that align with high-demand periods, ensuring maximum impact during the best months for sales.
7. **Dynamic Pricing Strategy:** Analyze the price vs. quantity relationship to implement dynamic pricing strategies, where lower-priced items can be sold in bundles or at discounts to encourage higher order quantities.
8. **Customer Segmentation:** Use customer activity data to segment high-value customers by purchasing behavior, allowing for more targeted marketing efforts, such as exclusive deals, loyalty programs, and VIP offers.
9. **Inventory Optimization:** Use insights from the top-selling products to optimize inventory management, ensuring popular items are always stocked and avoiding stockouts during peak demand periods.
10. **Advertising Budget Allocation:** Allocate more of the advertising budget towards the hours of the day and days of the week with the highest order volumes to maximize the ROI on ad spend.

11. **Product Lifecycle Management:** Monitor and adapt the product catalog by analyzing product sales trends, discontinuing underperforming products, and introducing new ones based on market demands.
12. **Customer Retention Strategy:** Implement a loyalty program or a rewards system for the most active customers to encourage repeat purchases, and track customer preferences to create customized offers that keep them engaged.
13. **Geographical Expansion:** Identify cities with high revenue and explore expansion opportunities in those regions, including local marketing campaigns and distribution strategies tailored to local preferences.
14. **Cross-Platform Marketing:** Use insights from product preferences and sales data across different platforms (online, mobile) to align cross-platform marketing efforts, ensuring consistent promotions reach potential buyers at the right time and place.
15. These actionable recommendations are designed to maximize sales, improve customer engagement, and optimize operations based on the insights from your sales data.