Data Wrangling

Prepared to: Dr Amal  
by: Mohamed Tamer, Youssra Yasser, Iad Mohamed, Moustafa Mohamed, Ziad Mohamed

# data PreProcessing

customer\_shopping\_data\_expanded.csv  
their was 5% or of the code were missing values.

Percentage of missing values: 4.883 %

Missing values per column:

invoice\_no 960

customer\_id 951

gender 999

age 991

category 1023

quantity 972

price 991

payment\_method 956

invoice\_date 967

shopping\_mall 956

The initial dataset contained missing values, which were thoroughly addressed. Additionally, outliers were identified and subsequently removed to ensure data integrity.The outliers

A colorful pie chart with text

Description automatically generatedBefore and after.

A graph of a number of objects

Description automatically generated with medium confidence

# ensuring there is no Missing values

* A blue square with green and yellow text

  Description automatically generated with medium confidence

# tableu IMAGES

A graph of a number of people

Description automatically generated with medium confidence

A screen shot of a graph

Description automatically generated

# part tableu

* **Summary of Key Findings and Actionable Insights**
* Here’s a breakdown of the important findings from the Tableau visualization, along with recommendations for how to act on them:

1. **Country with the Highest Total Profit: USA**

**Conclusion:** The USA generates the highest total profit, making it the most profitable market in your data.

**Significance:** This suggests that focusing more marketing efforts and inventory planning on the USA market would be a smart strategy, as it shows the most potential for high profits.

1. **Categories with the Highest Revenue: Clothing and Books**

**Males: $8 million | Females: $5 million**

**Conclusion:** Both clothing and books are major revenue earners for both males and females, but there’s a noticeable difference in spending between the genders, with men contributing more to the total revenue.

**Significance:** There’s an opportunity to adjust your product offerings and marketing strategies to better meet the needs and preferences of both men and women. It may be worthwhile to explore what drives the higher revenue from males and try to apply similar strategies for females in these categories.

1. **Season with the Highest Profit: Winter**

**Conclusion:** Winter is the most profitable season, likely due to factors such as the holiday season and increased gift-buying.

**Significance:** To make the most of this profitable season, it’s important to prepare well in advance. Consider launching special promotions, offers, and ensuring that your supply chain is ready to meet the increased demand. Summer, on the other hand, shows steady sales but doesn’t have the same seasonal spikes.

1. **Payment Methods and Spending by Age Group:**

**Aged 36+: Mainly Pay by Cash**

**Conclusion:** People aged 36 and above tend to prefer paying with cash over other payment methods.

**Significance:** To engage this demographic more effectively, it would be beneficial to tailor marketing and loyalty programs to their preference for cash transactions. Offering incentives or special offers could encourage further spending within this group.

**Highest Normalized Spending (781.5):** The 36+ age group represents the highest spending demographic, showing that they are particularly valuable customers.

**Significance:** Since this group spends the most, it’s important to target them with focused marketing campaigns. Offering personalized deals or rewards could help increase their loyalty and encourage them to continue spending at high levels.

1. **3-Year Timeline:**

**Conclusion:** Using a 3-year timeline to analyze trends and patterns provides a more accurate and reliable picture of market behavior.

**Significance:** A multi-year analysis allows for a deeper understanding of long-term trends, making the insights more trustworthy and useful for strategic planning.

* **Actionable Insights:**
* **For the USA:** Given that the USA is the most profitable market, prioritize marketing efforts and tailor your inventory planning to this region. Consider customizing product offerings or creating targeted marketing campaigns to maximize revenue here.
* **For Clothing and Books:** Since clothing and books are high-revenue categories, look for ways to expand or introduce new products in these areas. Understanding why males spend more in these categories could help develop strategies to boost sales among females as well.
* **For Winter Sales:** Winter is the most profitable season, so it’s crucial to prepare well in advance. Launch seasonal promotions, create special offers, and optimize your supply chain to handle the increased demand during the winter months.
* **For Age Group 36+:** Focus on creating marketing campaigns and loyalty programs that cater to the 36+ age group, considering their preference for cash payments. Offering special incentives or deals could encourage them to continue spending at high levels.
* **For Highest Spending Group (781.5):** Since this group represents the highest spenders, prioritize them in your marketing efforts. Offering tailored deals or loyalty rewards can help keep them engaged and encourage further spending.