Q4 readme

Before conducting the analysis, we need to preprocess the data by removing missing values and retaining only the data where both ‘Quantity’ and ‘Price’ are greater than 0.

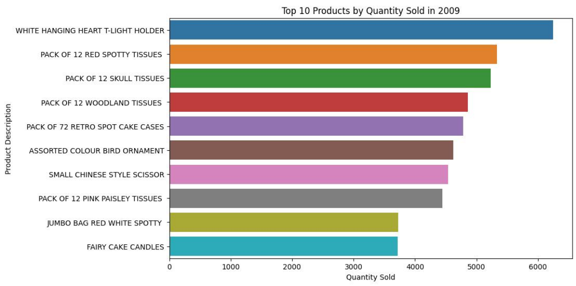
a. At least 10 pictures and at least 10 business insights

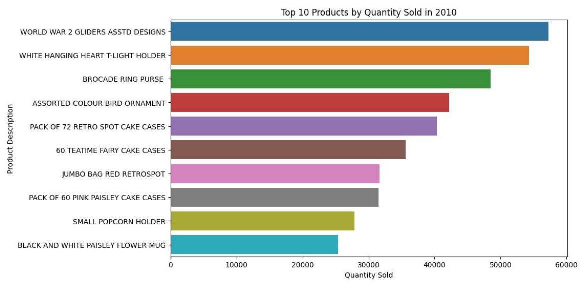
1) Total Sales Volume per Product per Year and totally

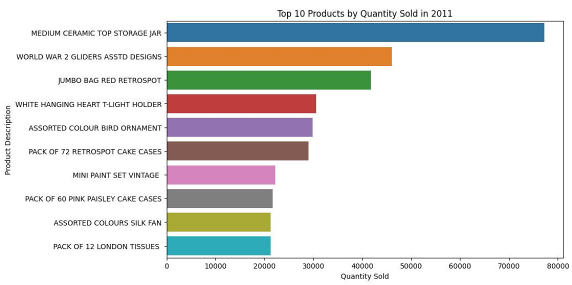
Description: The first graph displays the top 10 products with the highest quantities sold in 2009. However the second graph for 2010 suggests a different set of popular products compared to 2009, indicating changes in consumer preferences or stock levels. The third graph for 2011 shows that some products that appear consistently over the years, signifying steady demand. The fourth graph aggregates the data from the three years to highlight the overall top sellers.

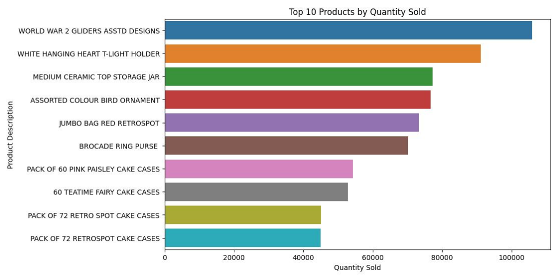
Business Insights:

- ✓ Product trends can shift year-to-year, indicating the need for flexible stock management.
- ✓ Certain items maintain popularity over time and should be kept well-stocked such as packs.
- ✓ Analyzing multi-year trends can guide inventory decisions and marketing strategies.







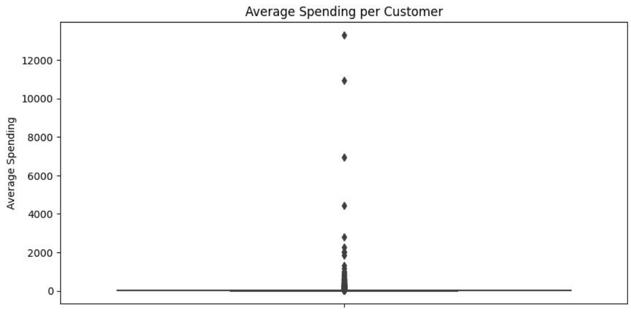


2) Average Spending per Customer

Description: There's a wide variance in customer spending. While most customers fall within a certain spending range, there are outliers who spend significantly more.

Business Insights:

- ✓ The supermarket owner could focus on these high-spend customers with tailored marketing strategies or loyalty programs to maintain their business, while also exploring ways to increase the average spend of the wider customer base.

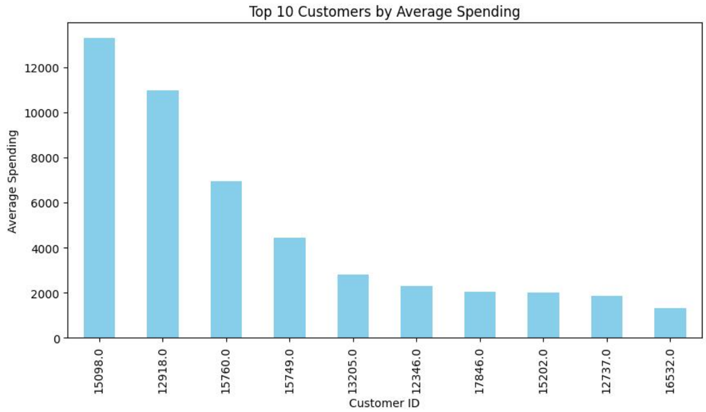


3) Top 10 Customers by Average Spending

Description: This bar chart displays the top 10 customers by average spending and we can identify which customer IDs have spent the most money.

Business Insights:

✓ These customer IDs demonstrate strong purchasing power, making them suitable targets for promoting high-end products.

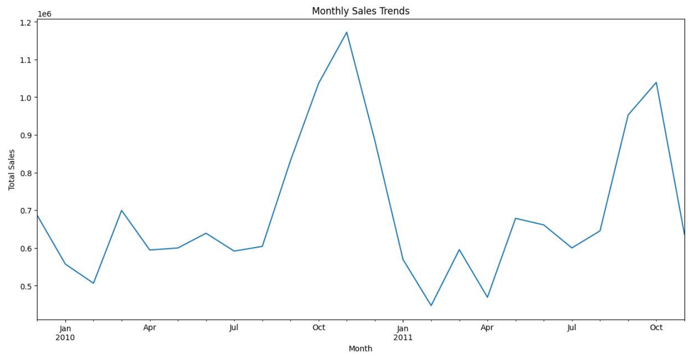


4) Monthly Sales Trends

Description: This line chart shows marked seasonal variations in monthly sales. It shows fluctuations in total sales across different months, with a notable peak suggesting a significant increase in sales during Oct., followed by a sharp decline.

Business Insights:

✓ The supermarket owner should investigate the cause of the peaks to capitalize on these trends, perhaps through targeted promotions or stock adjustments during expected high-sales periods.

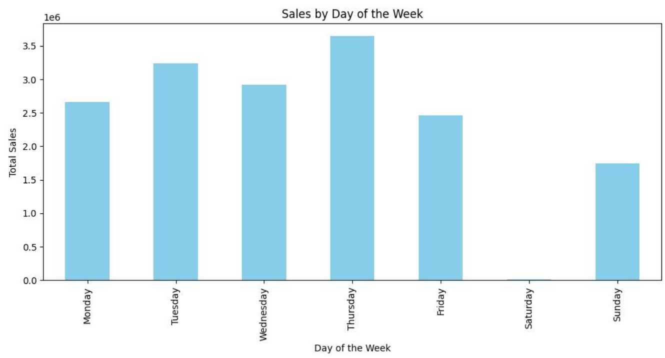


5) Day of the Week Sales

Description: The bar chart presents total sales for each weekday, indicating higher sales mid-week and a decline as the weekend approaches.

Business Insights:

✓ Mid-week sales peak suggests needing restocking, while lower weekend sales could benefit from targeted promotions or events.

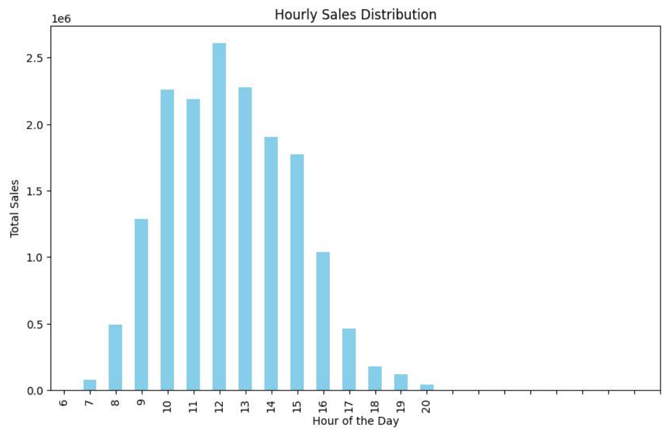


6) Hourly Sales Distribution

Description: The bar chart illustrates sales volume distributed by hour of the day, showing a significant surge in sales during the midday hours, with a sharp decrease after 3 pm.

Business Insights:

✓ Leveraging the sales peak around lunchtime by offering time-specific deals and promotions can sustain sales momentum throughout the afternoon and drive overall sales growth.

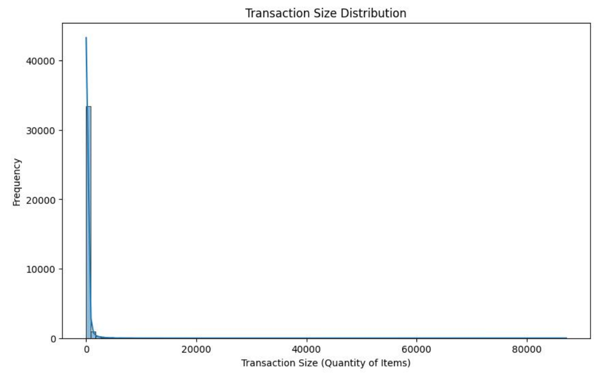


7) Transaction Quantity Distribution

Description: The chart illustrates most transactions involve a small number of items, with very few transactions containing larger quantities.

Business Insights:

✓ The concentration of transactions with fewer items could indicate a high frequency of small purchases, so there could be an opportunity for upselling and cross-selling to increase the transaction size.

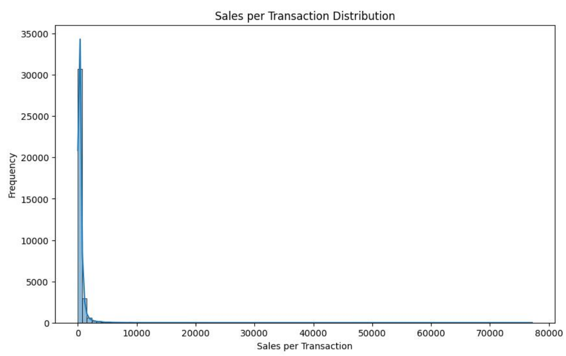


8) Transaction TotalPrice Distribution

Description: The chart illustrates most transactions involve a small number of items, with very few transactions containing larger quantities.

Business Insights:

✓ Smaller transactions are far more common than larger ones. Bundling products or offering discounts on multiple item purchases could be effective in increasing the average transaction value.

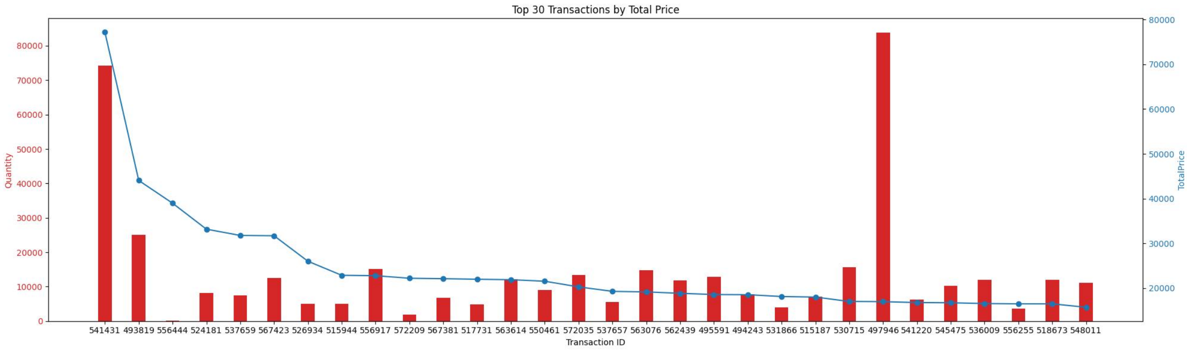


9) Quantity and TotalPrice of Top 30 Transactions by Total Price

Description: The dual-axis graph displays several top transactions by TotalPrice, bars indicate the quantity of items per transaction, and the line indicates the total price. There also exists a significantly higher total price than the others.

Business Insights:

✓ Most transactions have a consistent range of quantities.

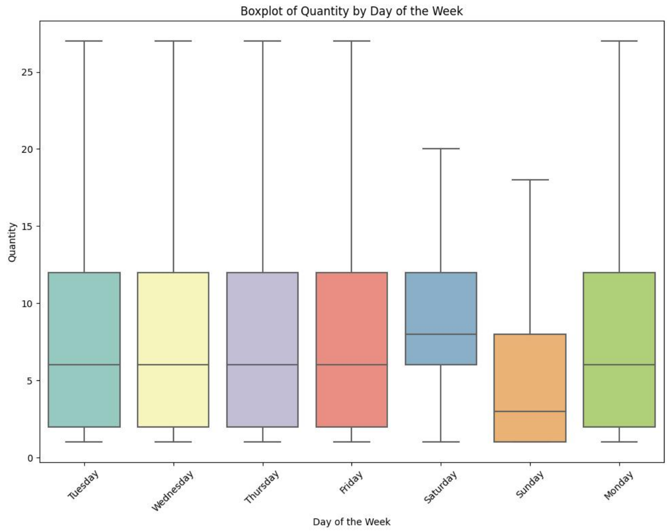


10) Boxplot of Quantity by Day of the Week

Description: The boxplot illustrates the distribution of quantities purchased across different days of the week, with weekdays showing wider variability in transaction sizes than weekends .

Business Insights:

✓ The varying spread of transaction sizes by day indicates inconsistent purchasing patterns, which could be used to tailor daily marketing efforts or specials to encourage more consistent or increased purchasing.



b. The algorithm details, process, and results of association rule analysis

To perform association rule analysis on transaction data, the following steps are followed: Firstly, the transaction data is transformed into a suitable format for association rule analysis. Then, the Apriori algorithm is applied to the transformed data to discover frequent itemsets. A minimum support threshold of 0.01 is set to filter out infrequent itemsets, resulting in the generation of frequent itemsets that meet the threshold requirement.

According to the research, the "metric" parameter can take the following values: ['support', 'confidence', 'lift', 'leverage', 'conviction'].

In this retail transactions we inference:

- Support: Commonly bought items identified by high support should be well-stocked.
- Confidence: High confidence suggests placing related items together to boost sales.
- Lift: Pairs of items with high lift values are prime candidates for bundling.

- Leverage: Items with high leverage can be promoted together to maximize profit.
- Conviction: Items with high conviction indicate a strong reliance on each other, suggesting marketing one to sell the other.

We now have frequent itemsets, "frequent_itemsets," obtained from the Apriori algorithm. Next, we proceed to find association rules (using the Support metric as an example). The following code snippet shows an example of the process along with output:

```
# Support metric
rules = association_rules(frequent_itemsets, metric="support", min_threshold=0.01)
top_rules = rules.nlargest(10, 'support')
for i, rule in top_rules.iterrows():
    print(f"Rule: {rule['antecedents']} -> {rule['consequents']}")
    print(f"Support: {rule['support']:.6f}\n")
```

```
Rule: frozenset({'21733'}) -> frozenset({'85123A'})  Support: 0.031160
Rule: frozenset({'85123A'}) -> frozenset({'21733'})  Support: 0.031160
Rule: frozenset({'22386'}) -> frozenset({'85099B'})  Support: 0.029584
Rule: frozenset({'85099B'}) -> frozenset({'22386'})  Support: 0.029584
Rule: frozenset({'20725'}) -> frozenset({'22384'})  Support: 0.028208
Rule: frozenset({'22384'}) -> frozenset({'20725'})  Support: 0.028208
Rule: frozenset({'20725'}) -> frozenset({'20727'})  Support: 0.027577
Rule: frozenset({'20727'}) -> frozenset({'20725'})  Support: 0.027577
Rule: frozenset({'82494L'}) -> frozenset({'82482'})  Support: 0.027147
Rule: frozenset({'82482'}) -> frozenset({'82494L'})  Support: 0.027147
```

We display the output for four additional metrics.

```
Rule: frozenset({'22699', '22698'}) -> frozenset({'22697'})
Confidence: 0.900175
Rule: frozenset({'22698', '22423'}) -> frozenset({'22697'})
Confidence: 0.891089
Rule: frozenset({'22746'}) -> frozenset({'22748'})
Confidence: 0.886105
Rule: frozenset({'21086', '21080'}) -> frozenset({'21094'})
Confidence: 0.879093
Rule: frozenset({'22698', '22423'}) -> frozenset({'22699'})
Confidence: 0.868812
Rule: frozenset({'22697', '22698'}) -> frozenset({'22699'})
Confidence: 0.859532
Rule: frozenset({'22745'}) -> frozenset({'22748'})
Confidence: 0.852273
Rule: frozenset({'22698'}) -> frozenset({'22697'})
Confidence: 0.847025
Rule: frozenset({'22697', '22423'}) -> frozenset({'22699'})
Confidence: 0.840000
Rule: frozenset({'22746'}) -> frozenset({'22745'})
Confidence: 0.833713
```

Confidence metric

```
Rule: frozenset({'22746'}) -> frozenset({'22745'})
Lift: 55.081901
Rule: frozenset({'22745'}) -> frozenset({'22746'})
Lift: 55.081901
Rule: frozenset({'22748'}) -> frozenset({'22746'})
Lift: 53.020376
Rule: frozenset({'22746'}) -> frozenset({'22748'})
Lift: 53.020376
Rule: frozenset({'22745'}) -> frozenset({'22748'})
Lift: 50.996024
Rule: frozenset({'22748'}) -> frozenset({'22745'})
Lift: 50.996024
Rule: frozenset({'21086', '21080'}) -> frozenset({'21094'})
Lift: 45.431537
Rule: frozenset({'21094'}) -> frozenset({'21086', '21080'})
Lift: 45.431537
Rule: frozenset({'21240'}) -> frozenset({'21239'})
Lift: 44.474702
Rule: frozenset({'21239'}) -> frozenset({'21240'})
Lift: 44.474702
```

Lift metric

```
Rule: frozenset({'22386'}) -> frozenset({'85099B'})
Leverage: 0.025191
Rule: frozenset({'85099B'}) -> frozenset({'22386'})
Leverage: 0.025191
Rule: frozenset({'21733'}) -> frozenset({'85123A'})
Leverage: 0.025188
Rule: frozenset({'85123A'}) -> frozenset({'21733'})
Leverage: 0.025188
Rule: frozenset({'82494L'}) -> frozenset({'82482'})
Leverage: 0.024934
Rule: frozenset({'82482'}) -> frozenset({'82494L'})
Leverage: 0.024934
Rule: frozenset({'20725'}) -> frozenset({'22384'})
Leverage: 0.024587
Rule: frozenset({'22384'}) -> frozenset({'20725'})
Leverage: 0.024587
Rule: frozenset({'20725'}) -> frozenset({'20727'})
Leverage: 0.023712
Rule: frozenset({'20727'}) -> frozenset({'20725'})
Leverage: 0.023712
```

Leverage metric

```
Rule: frozenset({'22699', '22698'}) -> frozenset({'22697'})
Conviction: 9.760242
Rule: frozenset({'22698', '22423'}) -> frozenset({'22697'})
Conviction: 8.945982
Rule: frozenset({'22746'}) -> frozenset({'22748'})
Conviction: 8.633264
Rule: frozenset({'21086', '21080'}) -> frozenset({'21094'})
Conviction: 8.110794
Rule: frozenset({'22698', '22423'}) -> frozenset({'22699'})
Conviction: 7.400850
Rule: frozenset({'22697', '22698'}) -> frozenset({'22699'})
Conviction: 6.911909
Rule: frozenset({'22745'}) -> frozenset({'22748'})
Conviction: 6.656100
Rule: frozenset({'22698'}) -> frozenset({'22697'})
Conviction: 6.369132
Rule: frozenset({'22697', '22423'}) -> frozenset({'22699'})
Conviction: 6.068147
Rule: frozenset({'22746'}) -> frozenset({'22745'})
Conviction: 5.922676
```

Conviction metric

Business Recommendation Suggestions:

- ✓ Stock Up on Items with High Support: Items that appear together in many transactions have high support and should be adequately stocked. These products are likely to be the staple goods that customers expect to find regularly.
  - Keep a robust inventory of products '21733' and '85123A' due to their frequent joint appearance in transactions, as indicated by a support of 0.031160. This suggests they are popular items often purchased together.
  - Similarly, ensure products '22386' and '85099B', with a support of 0.029584, are well-stocked. Products '20725' and '22384', and '20725' and '20727', both with support over 0.027, should also be readily available. Additionally, '82494L' and '82482' are paired products with a support of 0.027147, indicating they are regularly bought in conjunction.
- ✓ Strategically Place High Confidence Items: Items with strong confidence metrics should be placed in proximity within the store or advertised together online. This proximity can enhance the likelihood of these items being purchased together.
  - Place product '22699' next to '22698' to capitalize on their strong association, as indicated by a confidence of 90.0175%. Consider in-store signage or online banners that feature both items together.
  - Similarly, products '22698' and '22423' should be placed near '22697', following their confidence of 89.1089%. Continue this strategy with product '22746' next to '22748', and '21086'/'21080' alongside '21094', utilizing their high confidence levels to potentially increase sales of these paired items.
- ✓ Bundle Products with High Lift: Products that frequently appear together and have high lift values are prime for bundling. Consider special offers that package these items at a discounted rate to encourage increased sales.
  - Products '22746' and '22745' are frequently purchased together, as indicated by a lift of 55.081901. Consider creating a bundled offer for these products to encourage additional purchases.
  - Similarly, bundle '22748' with '22746' and '22745', both showing a lift over 50, to capitalize on their strong association. Also, '21086'/'21080' and '21094' have a lift of 45.431537, suggesting a bundled promotion could be beneficial. And pair '21240' with '21239', with a lift of 44.474702, to leverage their buying relationship.
- ✓ Promote Products with High Leverage: Highlight products with high leverage in marketing campaigns. Since these items are purchased together more often than expected by chance, promotions can further strengthen this relationship.
  - Highlight products '22386' and '85099B' in some cases, as they exhibit a leverage of 0.025191, indicating they are purchased together more often than expected.
  - Similarly, highlight the pairing of '21733' and '85123A', and '82494L' with '82482', both pairs showing significant leverage, to tap into their combined selling power. Also, consider promoting '20725' alongside '22384' and '20727', as their leverage scores suggest these products are frequently bought in conjunction.
- ✓ Incentivize Products with High Conviction: Items leading to high conviction metrics indicate a strong dependency. Offering discounts on the antecedent product or featuring it prominently could significantly increase the sale of the consequent product.
  - When products '22699' and '22698' are sold, customers are very likely to also purchase '22697', given the conviction of 9.760242. Consider offering a discount on '22699' and '22698' to increase sales of '22697'.
  - Similarly, when '22698' and '22423' are bought, '22697' is often purchased as well, as suggested by a conviction of 8.945982. Use this insight to create package deals or discounts that include these items. Additionally, '22746' has a strong likelihood of leading to the sale of '22748', indicated by a conviction of 8.633264, so consider pairing these items in promotions. The same approach applies to '21086' and '21080' with '21094', and '22745' with '22748', where offering a discount on the initial set can promote sales of the consequent item.

References：

[1] 关联规则----Apriori 算法以及代码实现:https://blog.csdn.net/qq_41230076/article/details/106094841

[2] 关联规则——关联分析 min_support 和 min_threshold: https://blog.csdn.net/h_jlwg6688/article/details/107793274