Sales Data Analysis Report

1. Introduction

This report covers the preprocessing techniques applied to the dataset and the analysis conducted to answer business-related questions based on sales data.

2. Preprocessing Techniques

Preprocessing Summary

The dataset underwent several preprocessing steps to ensure its quality and suitability for forecasting and modeling:

- **Missing Value Handling:**
  - Filled missing oil prices (*dcoilwtico*) using forward fill and backward fill methods.
  - Filled missing promotion and transaction values with zero where appropriate.
  - Addressed inconsistencies from the holiday dataset merge by assigning missing entries a default "Normal Day" or "None" for locale-related fields.
  - Replaced missing values in the *transferred* field with *False* to indicate no transfer.
- **Feature Engineering:**
  - **Crisis Indicator:** Created a binary *is_crisis* feature identifying crisis-related dates (e.g., the Manabi earthquake).
  - **Lag Features:** Generated a *sales_lag_7* feature, representing sales from 7 days prior for each store-family combination.
  - **Rolling Features:** Created a *rolling_mean_7* feature, computing a 7-day rolling average of sales (excluding the current day).
  - **Weekend Indicator:** Added an *is_weekend* feature to distinguish weekends from weekdays.
  - **Holiday Indicator:** Added an *is_holiday* feature to flag days associated with any holiday type.
  - **Promotion History:** Computed the *promo_last_7_days* feature, counting promotions in the previous 7 days.
- **Handling Feature-Engineered Missing Values:**
  - Filled missing lag and rolling mean values using backward filling to maintain continuity.
  - Filled missing promotion history values with zero, assuming no promotions in the early data.
- **Column Removal:**
  - Dropped unnecessary columns such as *id*, *locale_name*, and *description* to reduce dataset dimensionality and redundancy.

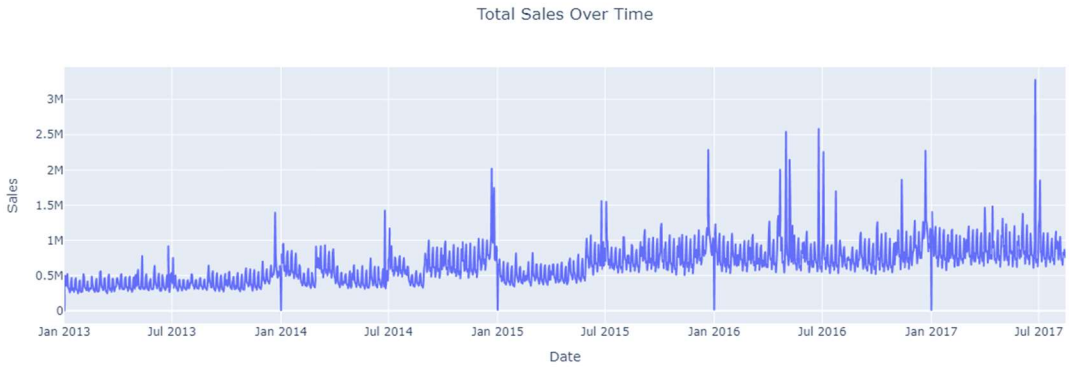
Okay, here is the edited and consolidated version of the business analysis report section, aiming for consistency in tone, structure, and summarization where appropriate.

---

3. Business Questions and Analysis

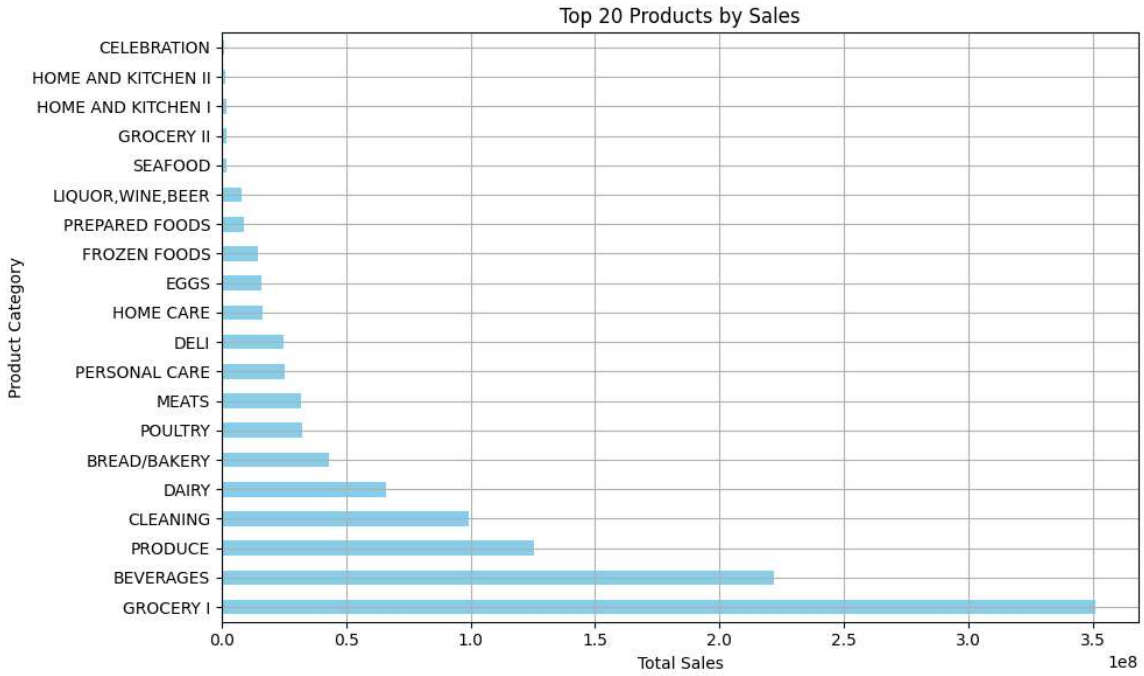
3.1 How Have Total Sales Evolved Over Time?

- **Findings:** *Total daily sales exhibit significant variability, ranging from approximately $2.5K to over $860K. The overall trend shows a clear upward trajectory over time, though marked by fluctuations likely due to seasonal factors.*
- **Implication:** *While growth is evident, understanding seasonality is key to interpreting performance variations.*



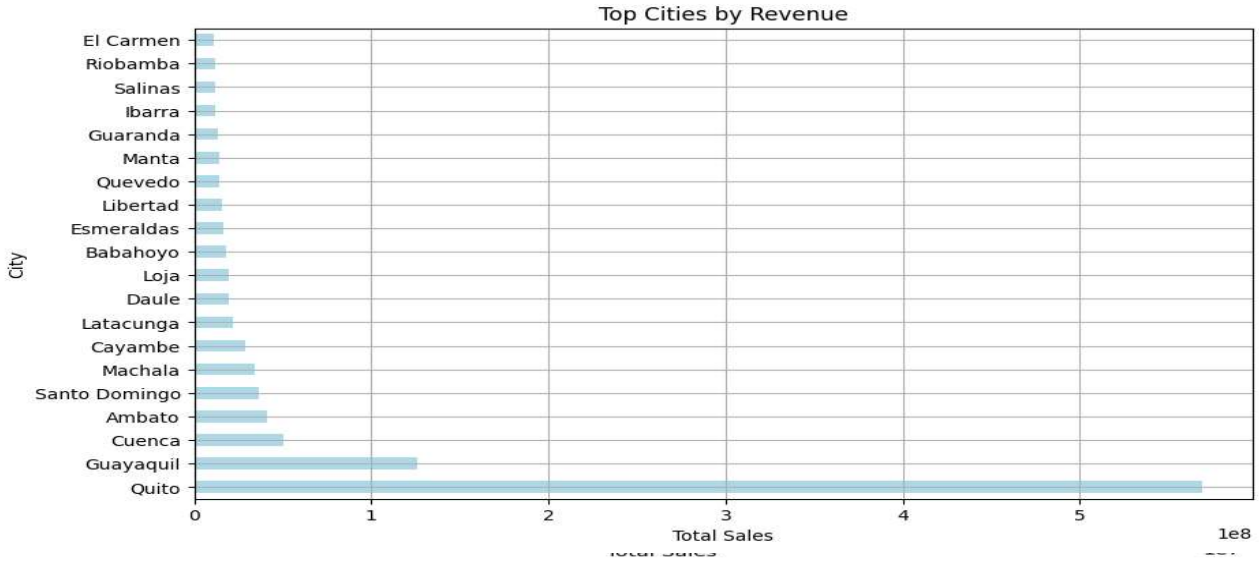
3.2 Which Products or Categories Contribute the Most to Total Revenue?

- **Findings:** *The top 5 revenue-generating categories are:*
    1. *GROCERY I: $350,827,298*
    2. *BEVERAGES: $221,663,540*
    3. *PRODUCE: $125,447,968*
    4. *CLEANING: $99,421,019*
    5. *DAIRY: $65,823,605*
- **Implication:** *Essential categories (Grocery, Beverages, Produce) dominate revenue, reflecting consistent consumer demand. GROCERY I is the clear leader.*



3.3 Which Stores, Cities, or States Are the Top Performers in Terms of Revenue?

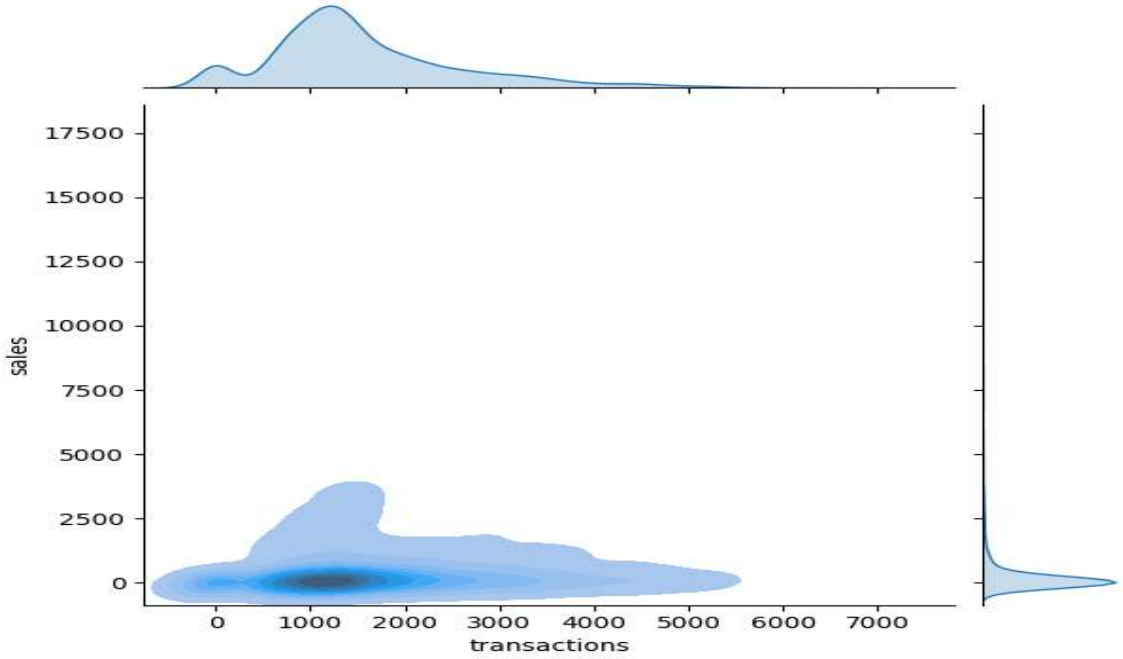
- **Findings:**
  - **Top Stores:** *Store 44 ($63.4M), Store 45 ($55.7M), Store 47 ($52.0M), Store 3 ($51.5M), Store 49 ($44.3M).*
  - **Top Cities:** *Quito ($568.7M), Guayaquil ($125.6M), Cuenca ($50.2M), Ambato ($41.2M), Santo Domingo ($36.6M).*
  - **Top States:** *Pichincha ($597.6M), Guayas ($168.6M), Azuay ($50.2M), Tungurahua ($41.2M), Santo Domingo de los Tsáchilas ($36.6M).*



- **Implication:** *Store 44, Quito, and Pichincha are the highest revenue generators, highlighting key geographical and specific store locations driving performance.*

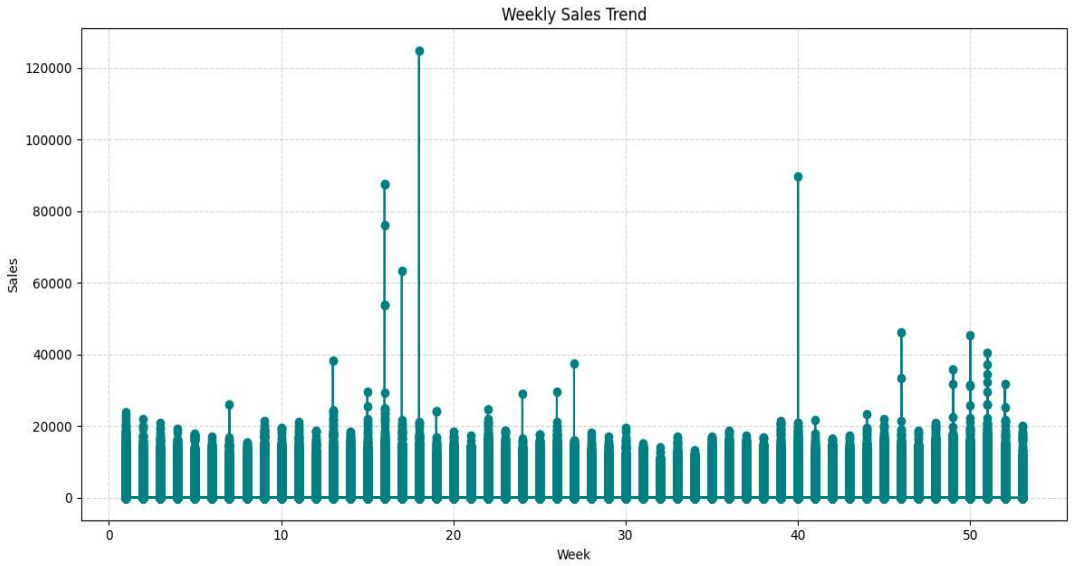
3.4 What is the Average Order Size Across Stores, Regions, and Categories?

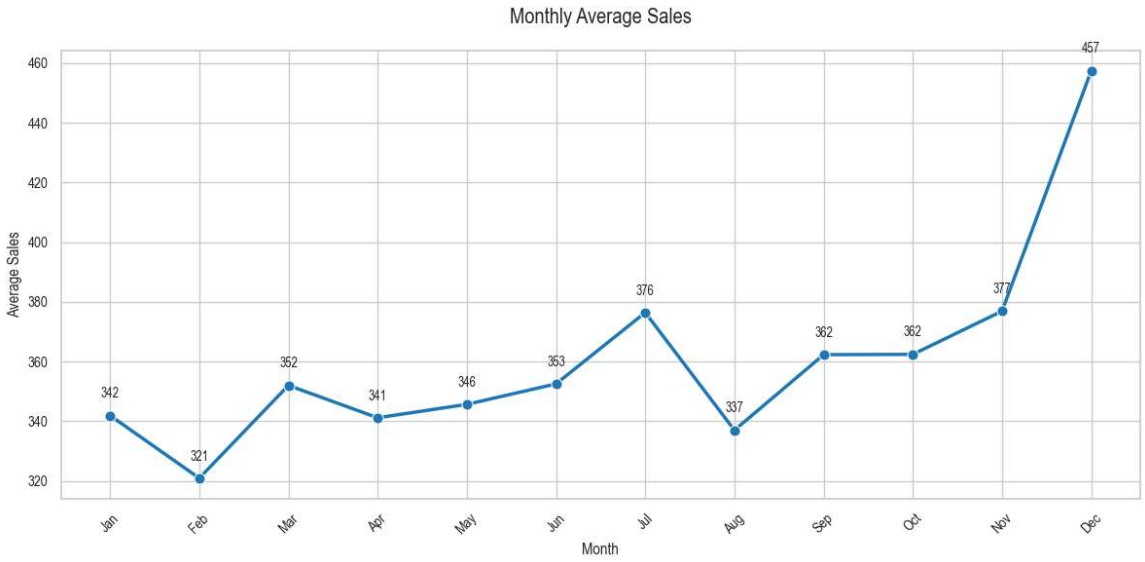
- **Findings:** *Further analysis is required to determine the average order size across different segments (stores, regions, categories).*
- **Implication:** *Understanding average order size is crucial for identifying high-value customer segments and optimizing sales strategies.*

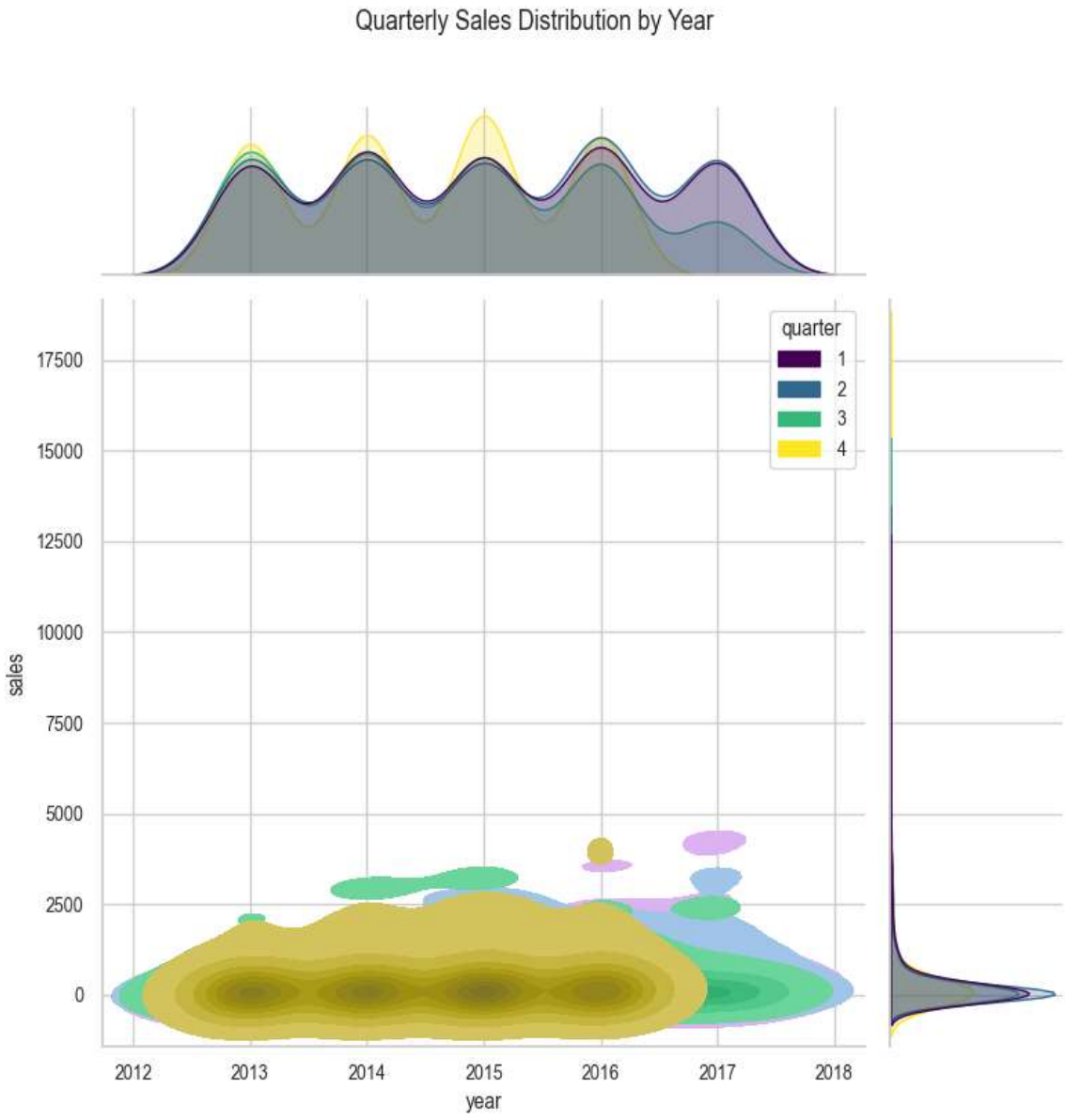


3.5 Are There Noticeable Weekly, Monthly, or Quarterly Seasonality Patterns in Sales?

- **Findings:** *Clear seasonal patterns exist:*
  - **Weekly:** *Sales peak in the final weeks of the year (Weeks 51, 52) and are lowest mid-year (Week 34).*
  - **Monthly:** *December shows the highest sales (holiday season), with July also showing strength. February consistently records the lowest sales.*
  - **Quarterly:** *Sales grew consistently from 2013-2017. Q3 and Q4 consistently demonstrate strong performance, particularly in later years.*
- **Implication:** *Sales peak significantly around the end-of-year holidays and show mid-year strength, suggesting opportunities to align marketing and inventory with these periods.*


Weekly Sales Trend
Sales
Week
0
20000
40000
60000
80000
100000
120000
0
10
20
30
40
50


Monthly Average Sales
Average Sales
Month
320
340
360
380
400
420
440
460
342
321
352
341
346
353
376
337
362
362
377
457
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

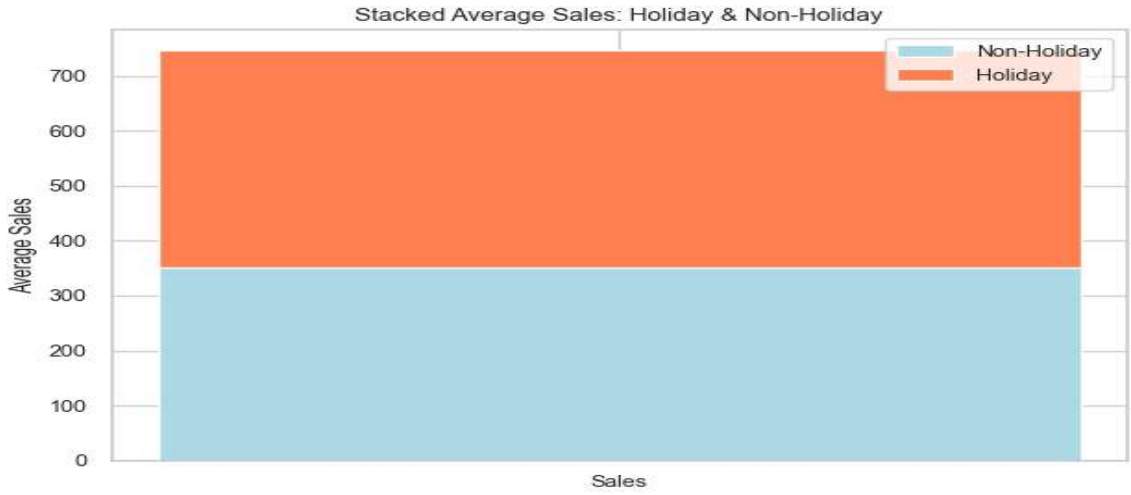


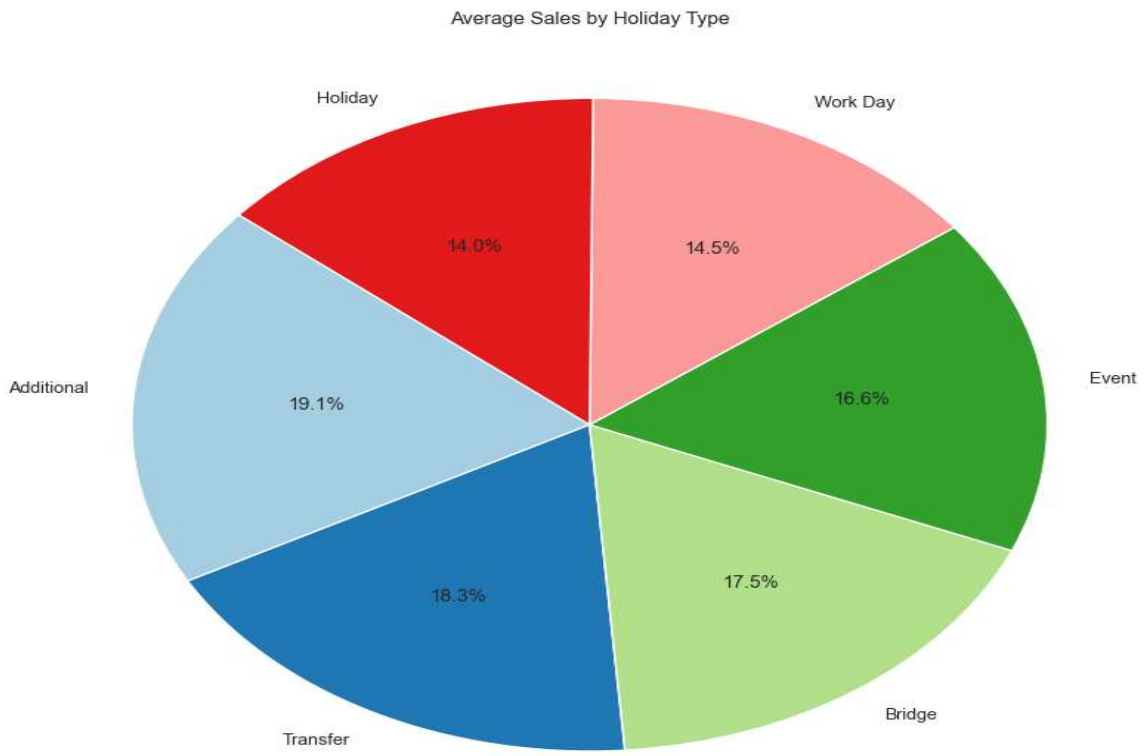
3.6 How Do Sales Differ on Weekdays Versus Weekends?

- **Findings:** *Weekends exhibit higher average daily sales compared to weekdays. However, weekdays contribute more to total sales volume due to having more days.*
- **Implication:** *Consumer spending intensity appears higher during weekends, possibly due to increased foot traffic or larger shopping trips.*

3.7 Are Sales Peaking During Certain Months, Holidays, or Quarters of the Year?

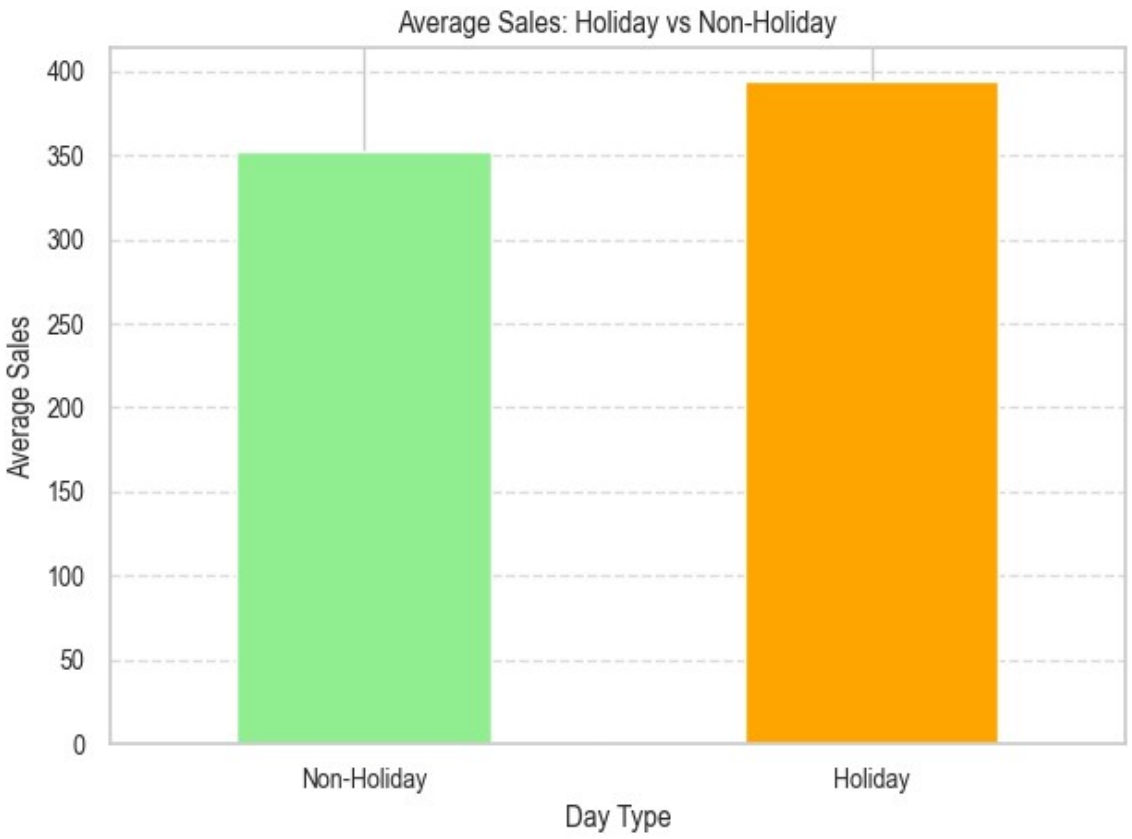
- **Findings:** *Sales peak strongly in December, aligning with the holiday season. November and July also show elevated sales. February consistently registers the lowest average sales. (This reinforces findings in 3.5).*
- **Implication:** *End-of-year holidays are the primary driver of peak sales, with other potential seasonal boosts in July and November.*





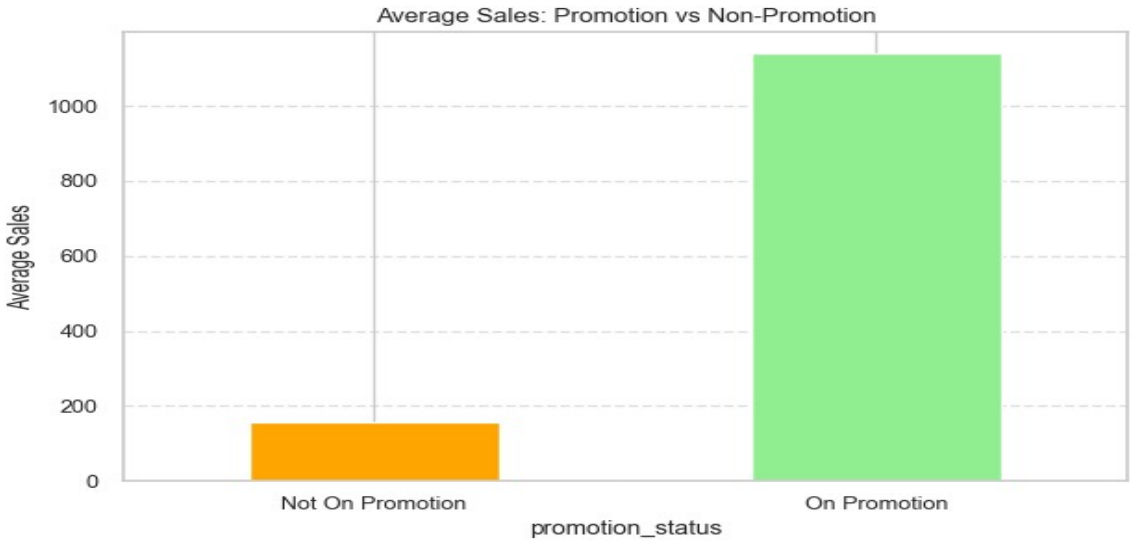
3.8 What Impact Do Holidays Have on Sales?

- **Findings:**
  - *Overall, holidays result in higher average sales (394) compared to non-holidays (352), an increase of approximately 11.9%.*
  - *The impact varies significantly by holiday type:*
    - *Highest Impact: Additional Holidays (Avg Sales: 488), Transfer Holidays (468), Bridge Holidays (447).*
    - *Moderate Impact: Event Holidays (426).*
    - *Lowest Impact: Work Day Holidays (372), Regular Holidays (358).*
- **Implication:** *Holidays provide a positive sales lift, but flexible holidays (creating long weekends) have a much stronger impact than fixed-date or standard holidays. Strategies should differ based on holiday type.*



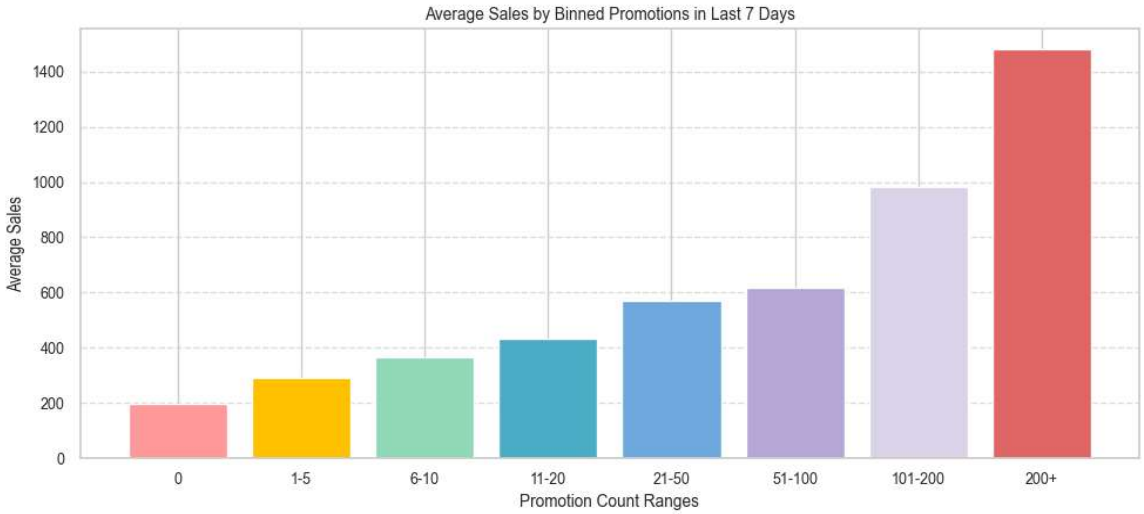
3.9 What Impact Do Promotions Have on Sales Volume?

- **Findings:** *Products on promotion have significantly higher average sales (1,140) compared to those not on promotion (158).*
- **Implication:** *Promotions are extremely effective at driving sales volume, highlighting the impact of marketing efforts like discounts and special offers.*



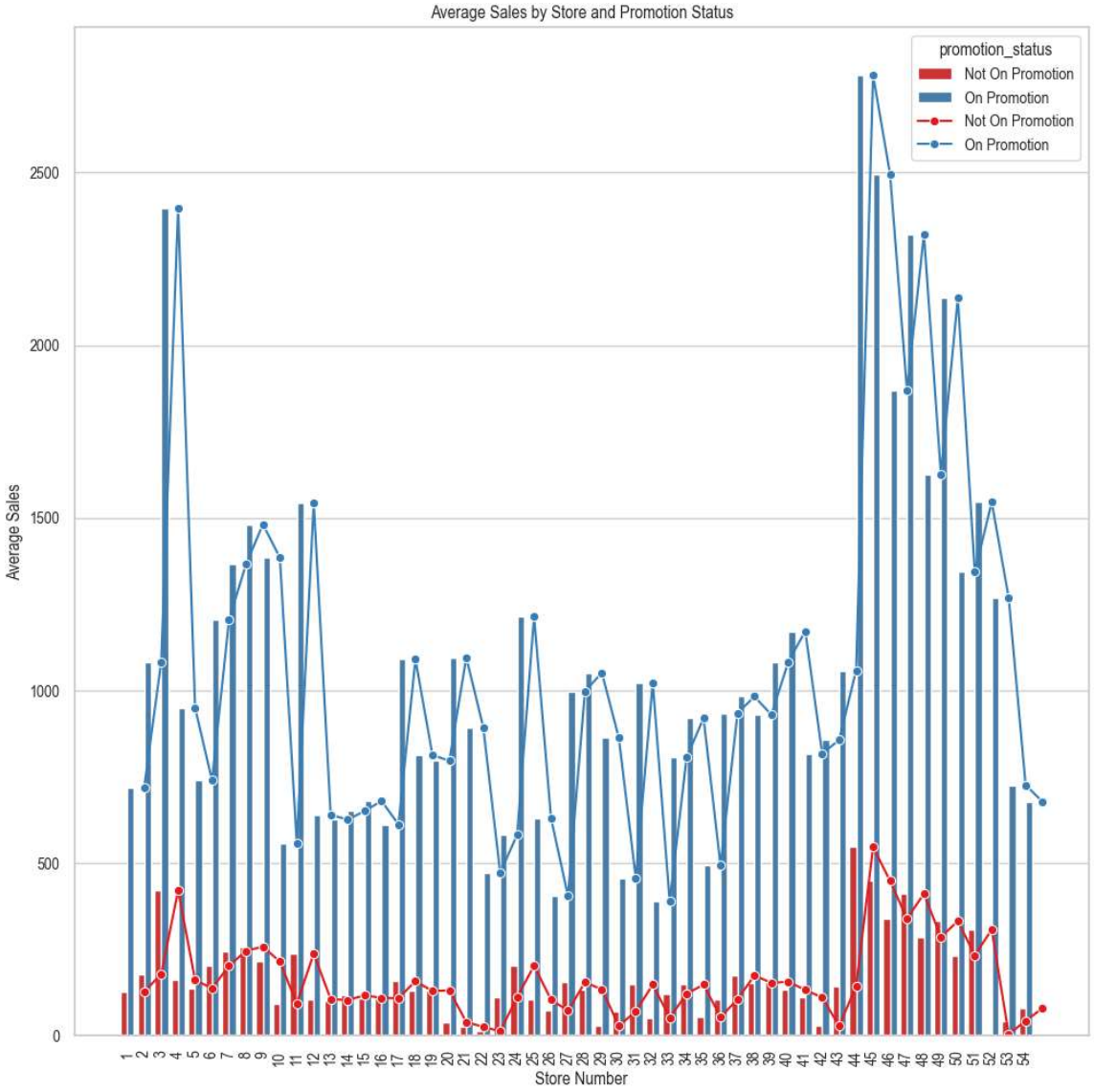
3.10 Is There a Cumulative Effect of Promotions (e.g., Last 7 Days of Promo)?

- **Findings:** *Sales periods following a promotion within the last 7 days show significantly higher average sales (490.60) compared to periods without recent promotions (198.35).*
- **Implication:** *Recent promotional activity has a clear positive carry-over effect on sales, suggesting sustained benefits from marketing campaigns.*



3.11 Are There Specific Families or Stores Where Promotions Are More Effective?

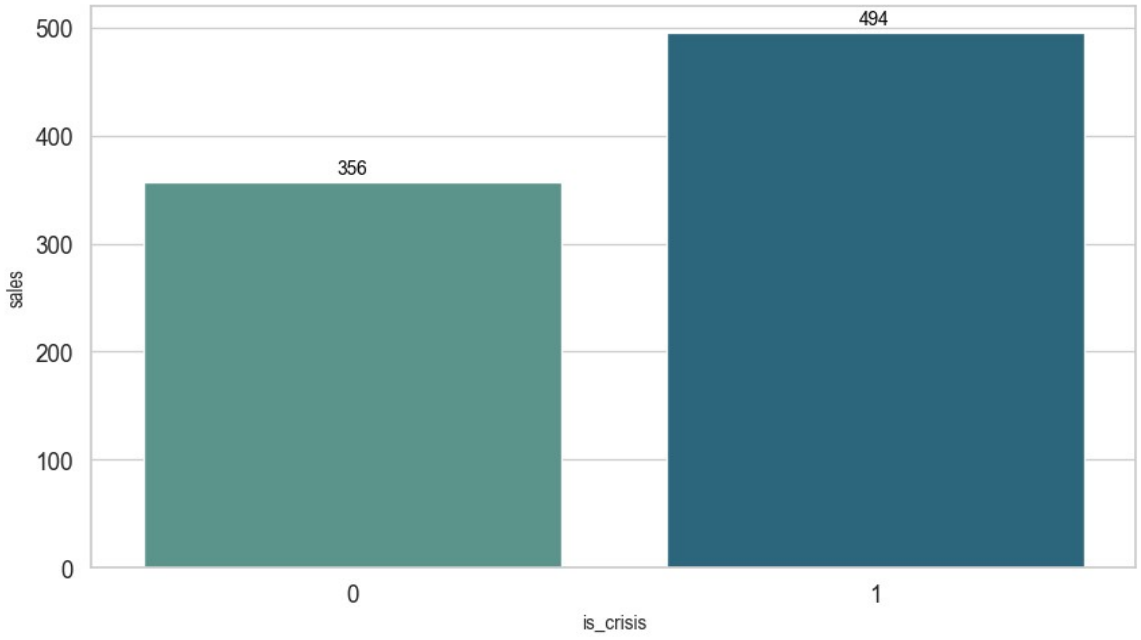
- **Findings:**
  - *Promotions are particularly effective for high-demand product families like GROCERY I and BEVERAGES.*
  - *Certain stores, such as Store 44 and Store 45, show a significantly higher sales boost from promotions compared to others (e.g., Store 52 shows low sales without promotions).*
- **Implication:** *Promotional effectiveness varies by product category and store, allowing for targeted strategies towards the most responsive segments.*



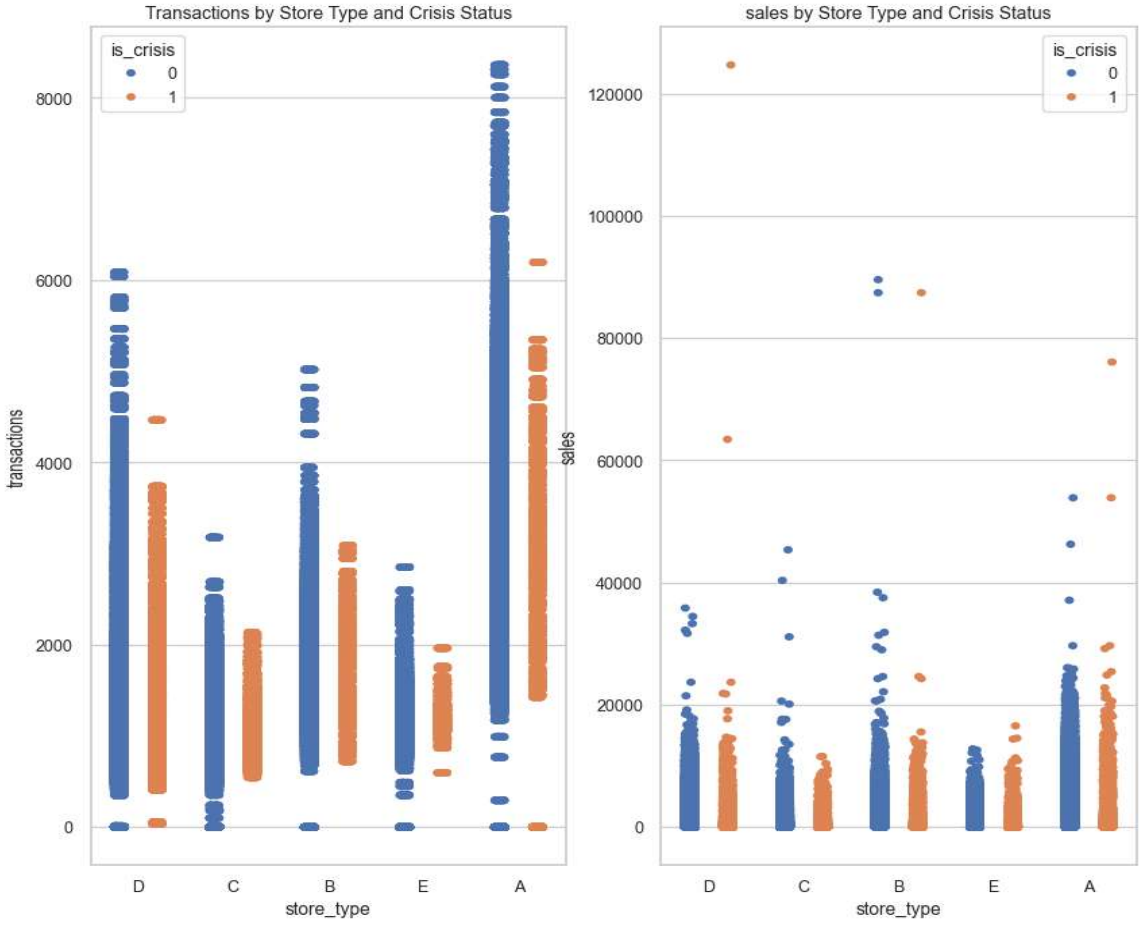
3.12 Crisis Impact on Sales and Transactions

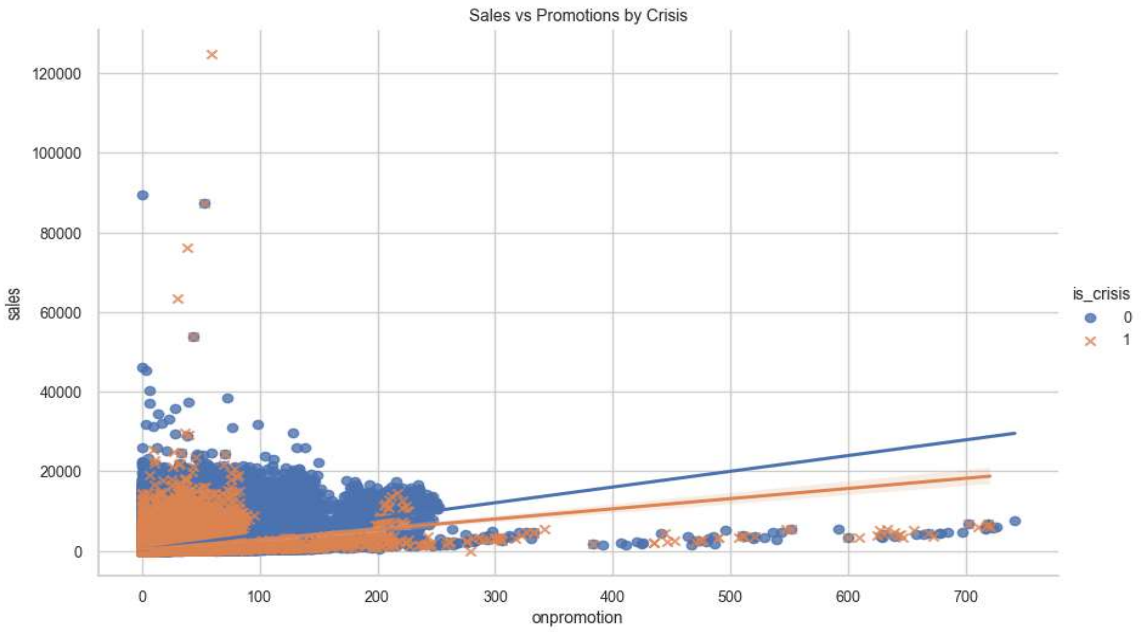
- **Summary of Findings:**
  - **Overall Impact:** *Crisis periods drive a substantial sales increase (+38.7%), primarily through larger basket sizes, while transaction volume increases more moderately (+5.9%).*
  - **Store Performance:** *Mid-tier (B, D) and Budget (C, E) stores show strong percentage growth, while Premium stores (A) see higher absolute sales but smaller relative increases.*
  - **Promotions:** *Extremely effective during crises, especially mid-to-high tier promotions, amplifying consumer demand.*
  - **Product Categories:** *Essentials (grocery, produce, cleaning, pharmacy) surge; non-essentials show minimal change.*

    - **Holidays:** *Crisis impact tends to amplify or override typical holiday sales patterns.*
    - **Regional Variations:** *Urban centers (Quito, Guayaquil) show large absolute growth; smaller cities/certain states (e.g., Manabi, Amazonian states) show dramatic percentage increases. Some areas (e.g., Salinas) may see negative growth.*
- **Strategic Recommendations:**
    - **Inventory:** *Prioritize stocking essentials, especially in urban centers, and prepare for disproportionate demand surges in smaller cities/regions.*
    - **Promotions:** *Leverage mid-to-high level promotions aggressively during crises.*
    - **Operations:** *Prepare supply chains and staffing for immediate ramp-up and sustained high volume (larger baskets).*
    - **Monitoring:** *Use rolling averages to track trends and adjust KPIs for the elevated baseline.*



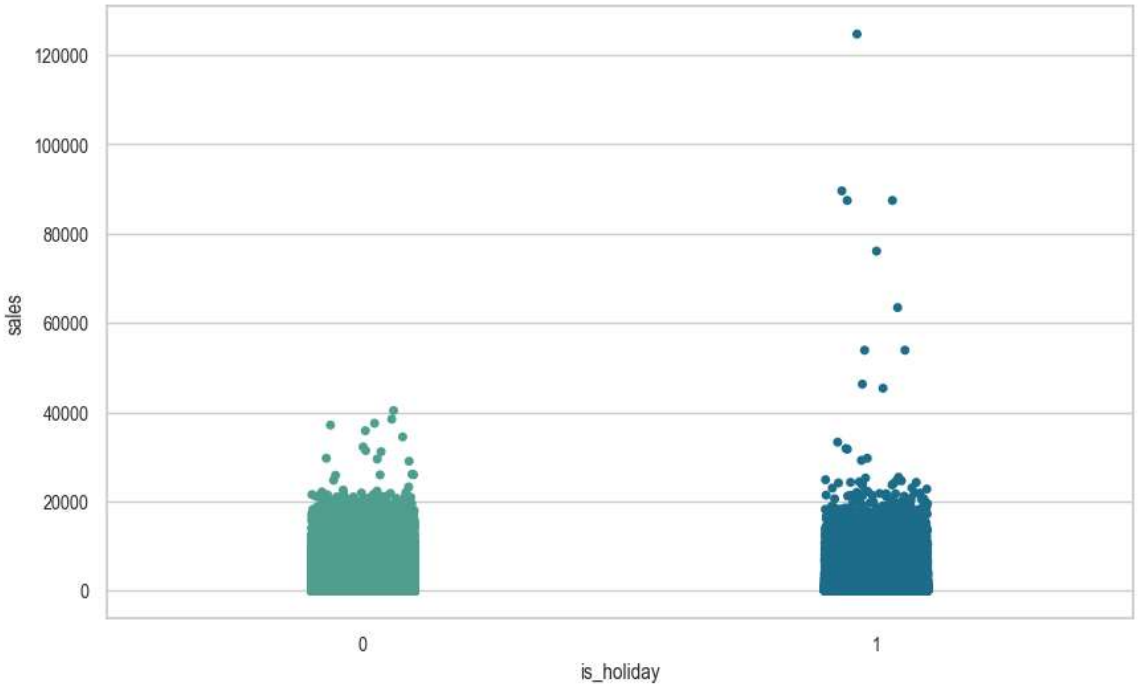
Transactions and sales by Store Type and Crisis Status





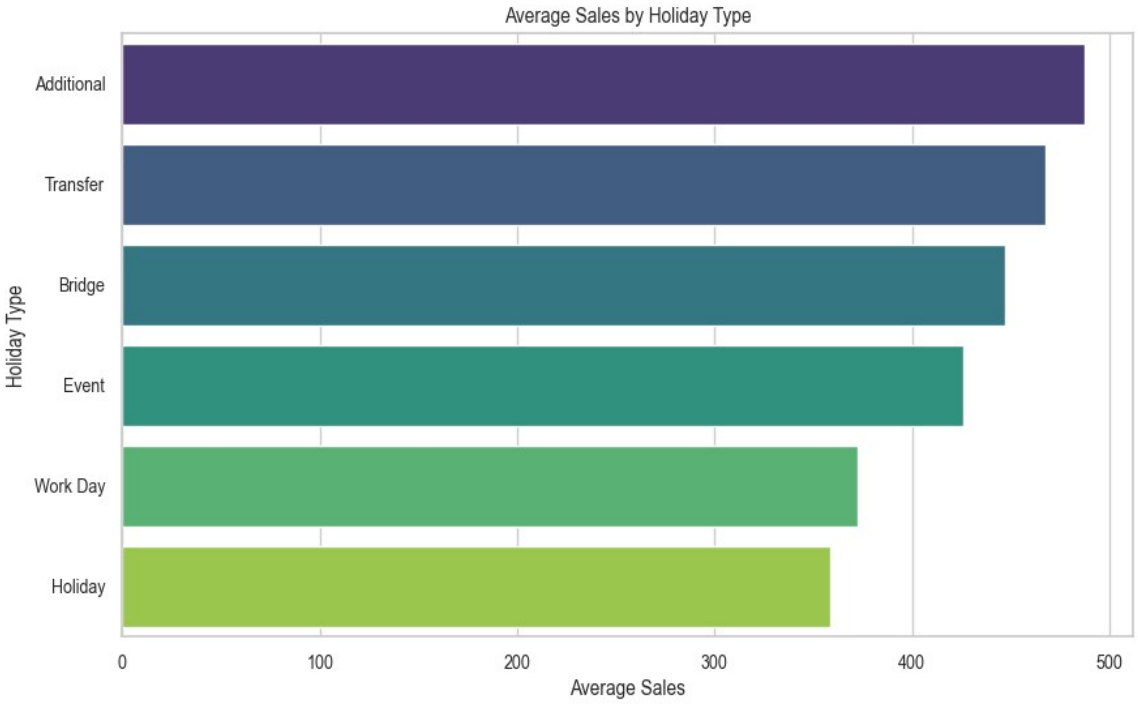
3.13 How do sales differ on holidays vs. non-holidays overall? (Consolidated View)

- **Findings:** *Holidays provide a moderate average sales lift of +11.9% (352 to 394). This boost is significant but considerably smaller than the impact of crisis periods (+38.7%).*
- **Behavioral Context:** *Holiday shopping likely involves more discretionary/gift items, while crisis shopping focuses on necessities and stockpiling.*
- **Strategic Recommendations:** *Plan for moderate inventory increases (+10-15%) focused on seasonal/gift items for holidays. Schedule slightly increased staffing (+15%). Launch targeted holiday campaigns in advance.*



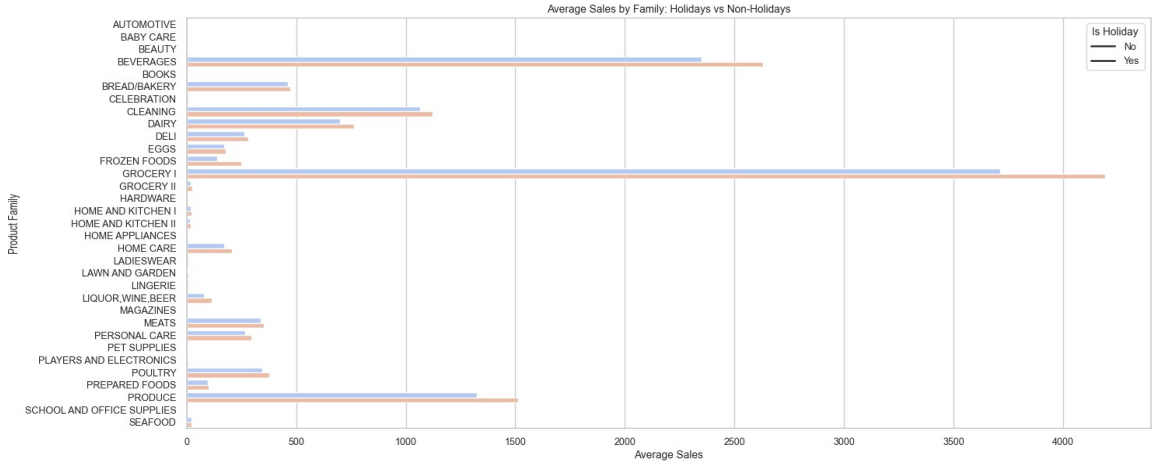
3.14 Which type of holiday (national, regional, local) drives the highest sales? (Consolidated View)

- **Findings:** *Flexible-date holidays that create long weekends (Additional, Transfer, Bridge types) yield the highest sales lift (+27% to +38%). These outperform fixed-date national holidays (+2%) and local/event-based holidays (+21%). Workday holidays show minimal impact (+6%).*
- **Implication:** *Date flexibility and the creation of extended weekends are more impactful for sales than the specific cultural relevance or fixed date of a holiday.*
- **Strategic Recommendations:** *Prioritize inventory, staffing, and marketing efforts heavily around Additional, Transfer, and Bridge holidays. Use a localized approach for Event holidays. Reduce special preparations for standard fixed-date holidays.*



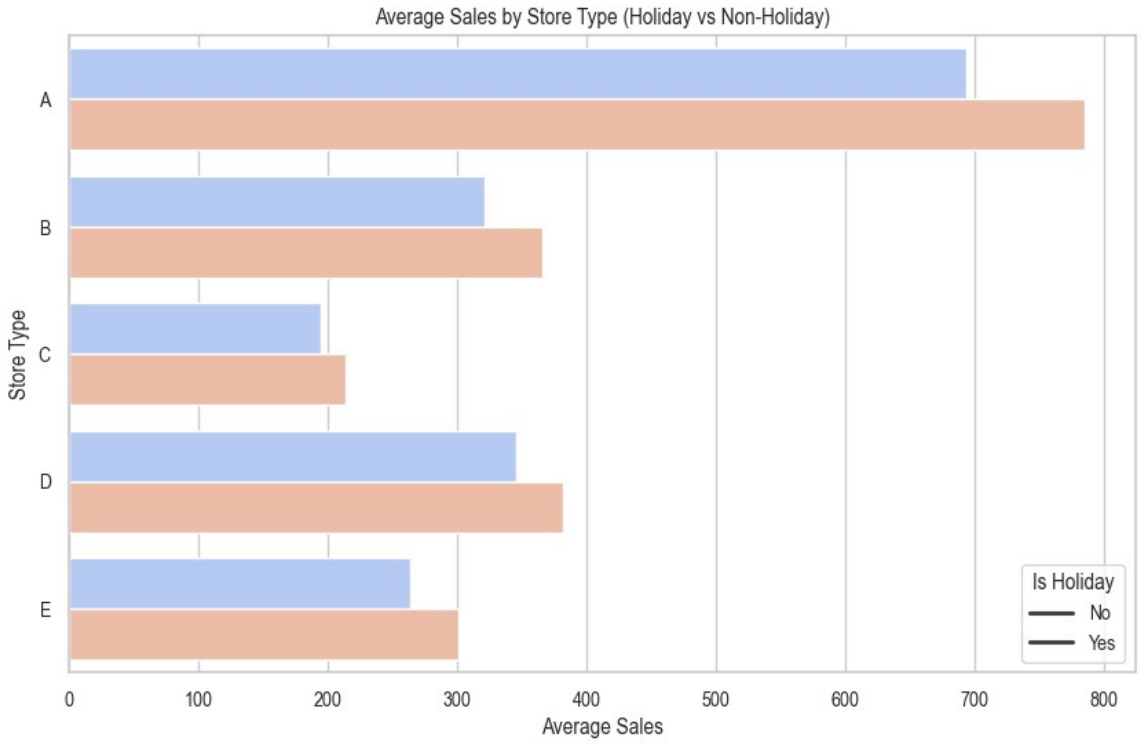
3.15 Which product families see the biggest sales boost during holidays?

- **Findings:** *The largest holiday sales increases are seen in:*
  - *Produce (+30%)*
  - *Deli (+25%)*
  - *Meats (+24%)*
  - *Beverages (+20%) Families like Baby Care, Seafood, and School/Office Supplies show minimal change.*
- **Implication:** *Holiday demand significantly boosts fresh food, ready-to-eat items, and beverages, likely for meals and gatherings. Focus promotions and inventory on these categories during holiday periods.*



3.16 Are certain stores or store types more sensitive to holiday sales spikes?

- **Findings:** *Store Types B and E show the highest relative increase (+14%) during holidays. Store Type A (high base sales) also sees a strong increase (+13.3%). Store Type C is the least sensitive (+9.2%).*
- **Implication:** *Large format (A) and mid-tier stores (B, E) benefit most from holiday demand. Target holiday efforts towards these store types.*

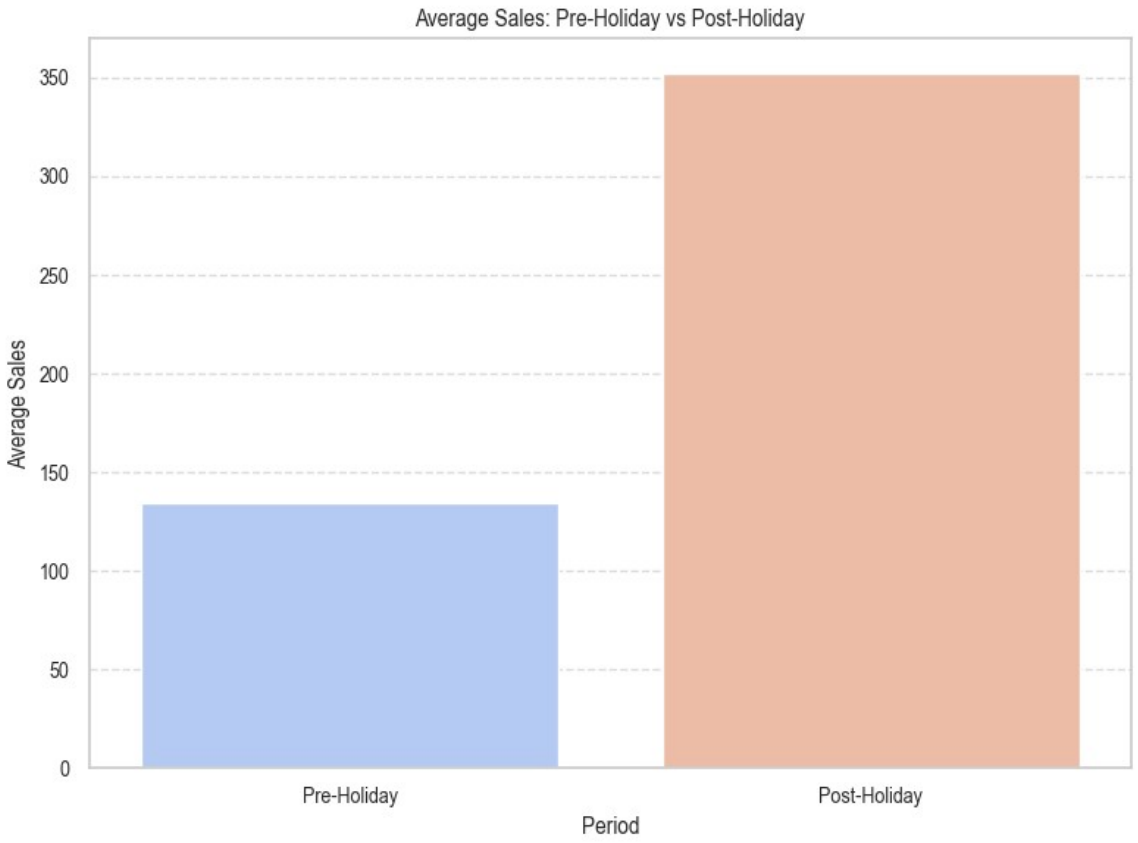


3.17 How many days before a holiday does sales start increasing?

- **Findings:** *Analysis of the 7 days immediately preceding holidays does not show a sales increase compared to non-holiday days within that same window. Average sales on non-holiday days in this period were higher.*
- **Implication:** *This suggests significant pre-holiday shopping likely occurs more than 7 days in advance, or sales peak sharply right on the holiday itself, rather than building gradually in the final week.*
- **Recommendation:** *Analyze a broader time window (e.g., 14-21 days prior) to identify the actual start of the pre-holiday sales lift.*

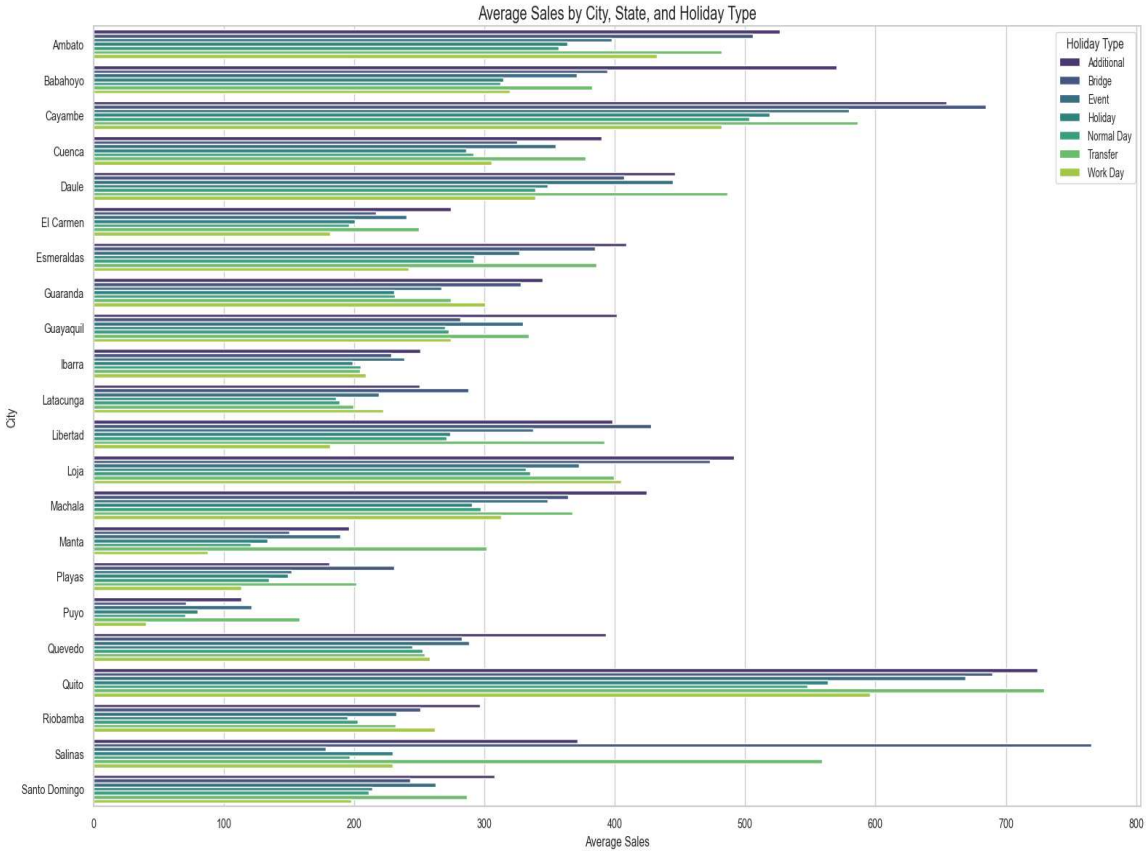
3.18 Do sales drop after holidays ("post-holiday effect")?

- **Findings:** *Contrary to expectations of a slump, average sales increase in the period following holidays compared to the pre-holiday period analyzed (Post-Holiday Avg Sales reported as 2.6x higher than Pre-Holiday Avg Sales in the analyzed slice).*
- **Implication:** *The data does not support a post-holiday drop; instead, it suggests a recovery, rebound, or continued momentum, possibly fueled by promotions or restocking.*
- **Recommendation:** *Consider maintaining promotions or ensuring stock availability immediately after holidays to capture sustained consumer interest.*



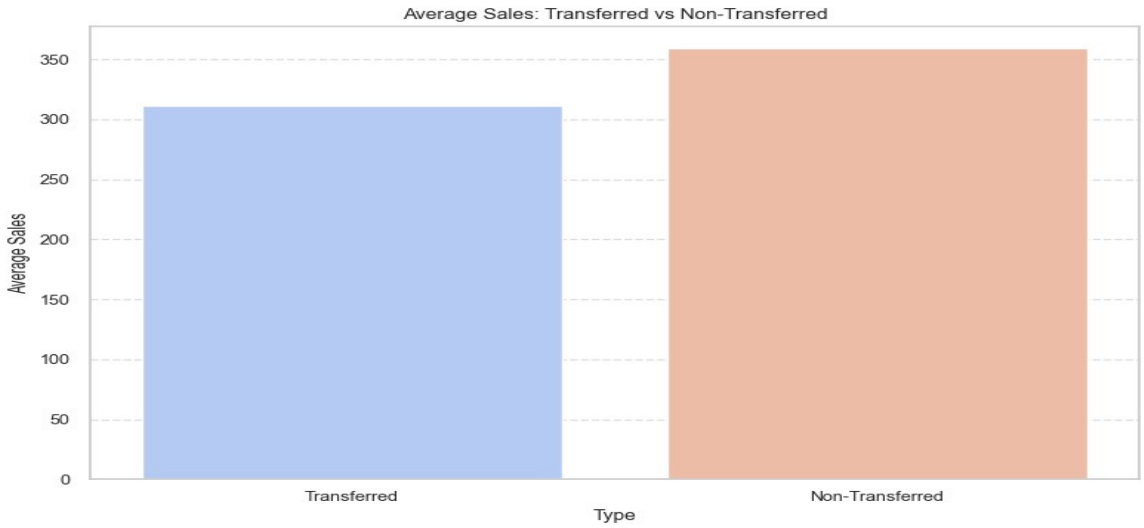
3.19 Which city or state benefits the most from local holidays?

- **Findings:** *Holiday impact varies geographically. Ambato (Tungurahua state) shows particularly strong performance during Additional and Bridge holidays compared to its baseline. Santo Domingo shows moderate lifts during Transfer holidays.*
- **Implication:** *Local context matters significantly. Regions like Tungurahua (especially Ambato) appear highly responsive to certain holiday types and should be targeted with localized promotions and inventory planning.*



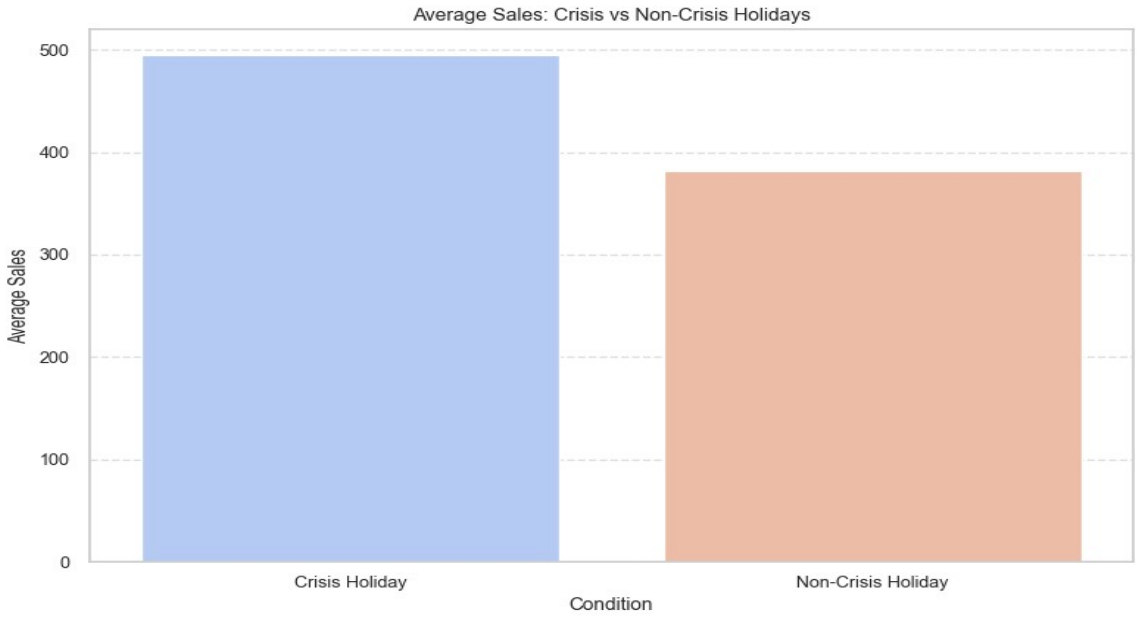
3.20 Is there a difference in sales between transferred holidays and non-transferred holidays?

- **Findings:** Average sales during transferred holidays are approximately *13%* lower than during their non-transferred counterparts.
- **Interpretation:** Moving a holiday's observed date might diminish its commercial impact. This could be due to reduced consumer anticipation, misalignment with traditional shopping patterns, or general confusion.
- **Implication:** Non-transferred (fixed-date or traditionally observed) holidays appear to present a stronger, more predictable opportunity for sales and promotions compared to dates adjusted via transfer. Forecasts should account for this potential reduction for transferred holidays



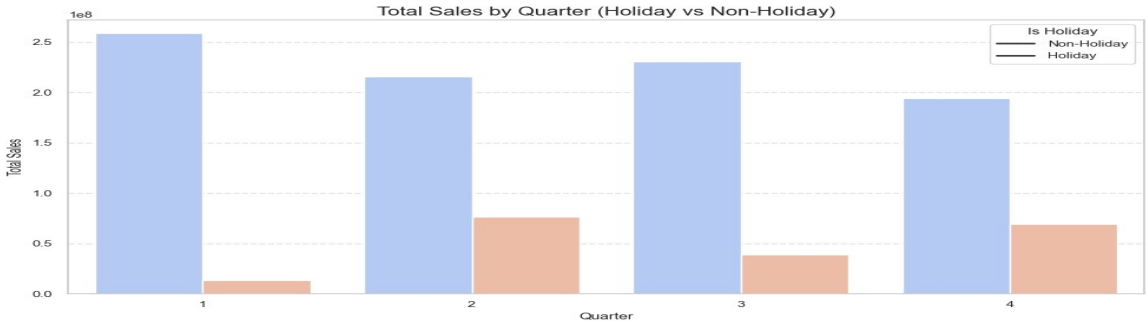
3.21 Do crisis periods reduce the usual holiday sales spike?

- **Findings:** Contrary to expectation, average holiday sales are observed to be nearly *30%* higher when occurring during crisis periods compared to holidays outside of crises.
- **Interpretation:** Potential drivers include amplified stockpiling behavior extending into holiday purchases, retailers using more aggressive promotions during crises to maintain volume, or consumers prioritizing holiday spending for morale despite adverse conditions.
- **Implication:** Crisis conditions do not necessarily dampen holiday sales; they can potentially amplify them. Businesses need to remain adaptable, as typical holiday patterns might be altered during crises.



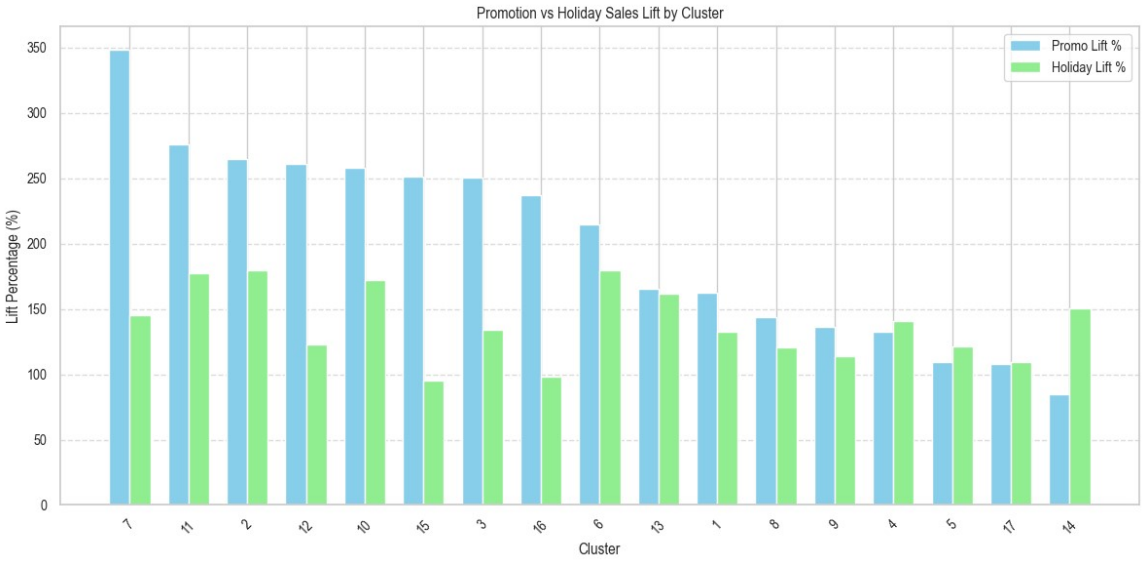
3.22 Which quarter has the highest number of holidays and how does that affect total sales?

- **Findings:**
  - Quarter 1 (Q1) exhibits the highest *overall* sales volume, including both holiday and non-holiday periods.
  - However, Quarter 2 (Q2) shows the most significant *sales spike specifically during holidays*. Holiday sales in Q2 reached *$76.57M*, substantially higher than Q1 (*$13.6M*), Q3 (*$38.57M*), and Q4 (*$69.18M*).
- **Implication:** While Q1 drives the largest portion of annual sales, the holidays occurring within Q2 appear to have a disproportionately strong positive impact on sales during that quarter, making them a key period for targeted promotions.



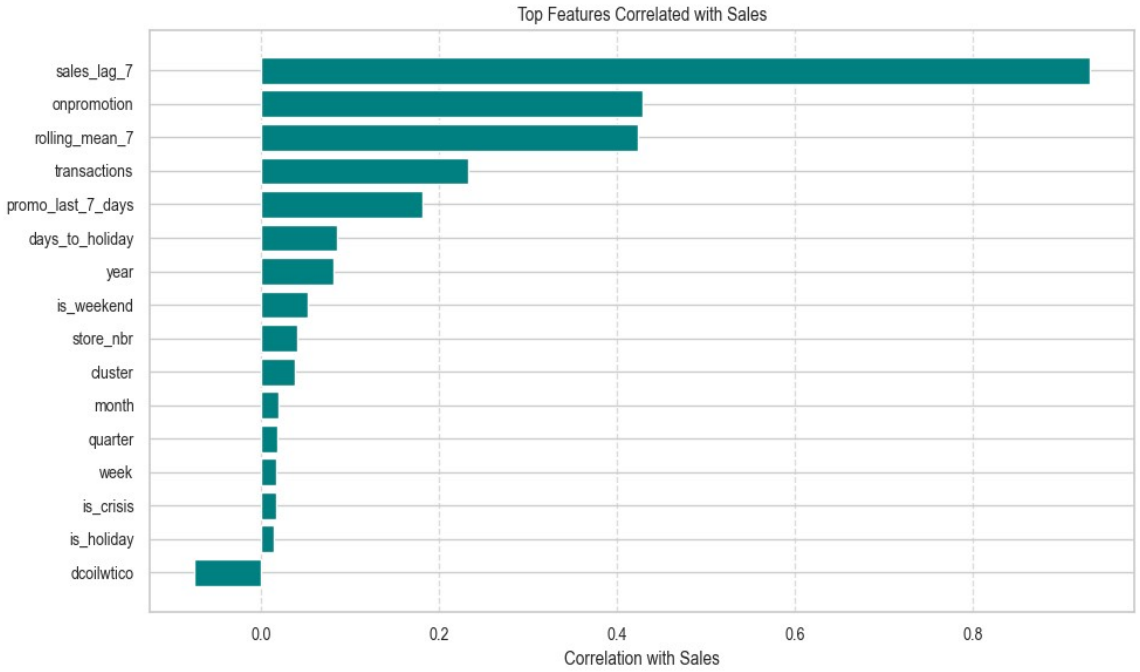
3.23 Are certain clusters more sensitive to promotions or holidays?

- **Findings:** (Based on Promo Lift % = Sales increase during promotion vs. normal; Holiday Lift % = Sales increase during holiday vs. normal)
  - **Promotion Sensitivity:** Cluster 7 shows the highest sensitivity (*+349%* lift). Clusters 11 (*+276%*) and 2 (*+265%*) are also highly responsive to promotions.
  - **Holiday Sensitivity:** Cluster 2 demonstrates high sensitivity to holidays (*+180%* lift), alongside its strong promotion response. Cluster 14 is notably more sensitive to holidays than promotions.
  - **Balanced Sensitivity:** Clusters like 5 and 13 show more moderate lifts for both promotions and holidays.
  - **Overall Trend:** For most clusters, promotions tend to generate a larger percentage increase in sales compared to holidays.
- **Implication:** Different customer/store clusters react differently to promotions versus holidays. This allows for targeted marketing strategies and inventory planning based on the dominant sensitivity of each cluster.



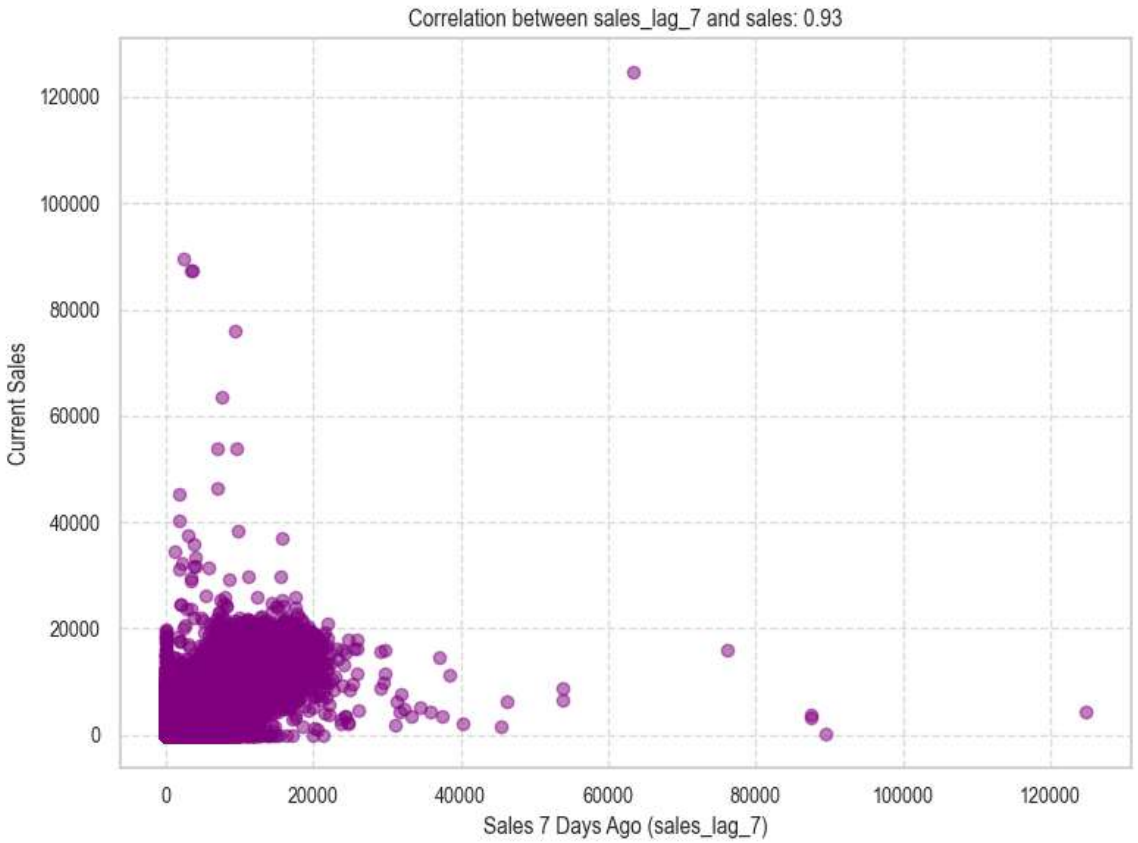
3.24 Which features are most correlated with sales?

- **Findings:** Correlation analysis with 'sales' reveals:
  - **Strong Positive:** *sales_lag_7* (0.93) - Sales in the previous 7 days.
  - **Moderate Positive:** *onpromotion* (0.43), *rolling_mean_7* (0.42) - Current promotion status and 7-day rolling average sales.
  - **Weak Positive:** *transactions* (0.23), *promo_last_7_days* (0.18) - Number of transactions and promotion status in the last 7 days.
  - **Minimal/Negligible:** *days_to_holiday* (0.09), *year* (0.08), *is_weekend* (0.05), *store_nbr* (0.04).
- **Implication:** Recent historical sales (*sales_lag_7*) are by far the strongest predictor of current sales. Promotions and recent sales trends (*rolling_mean_7*) also have a meaningful positive association. Factors like day of the week or specific store number show very little direct correlation with overall sales volume in this analysis. This guides feature importance for forecasting models.



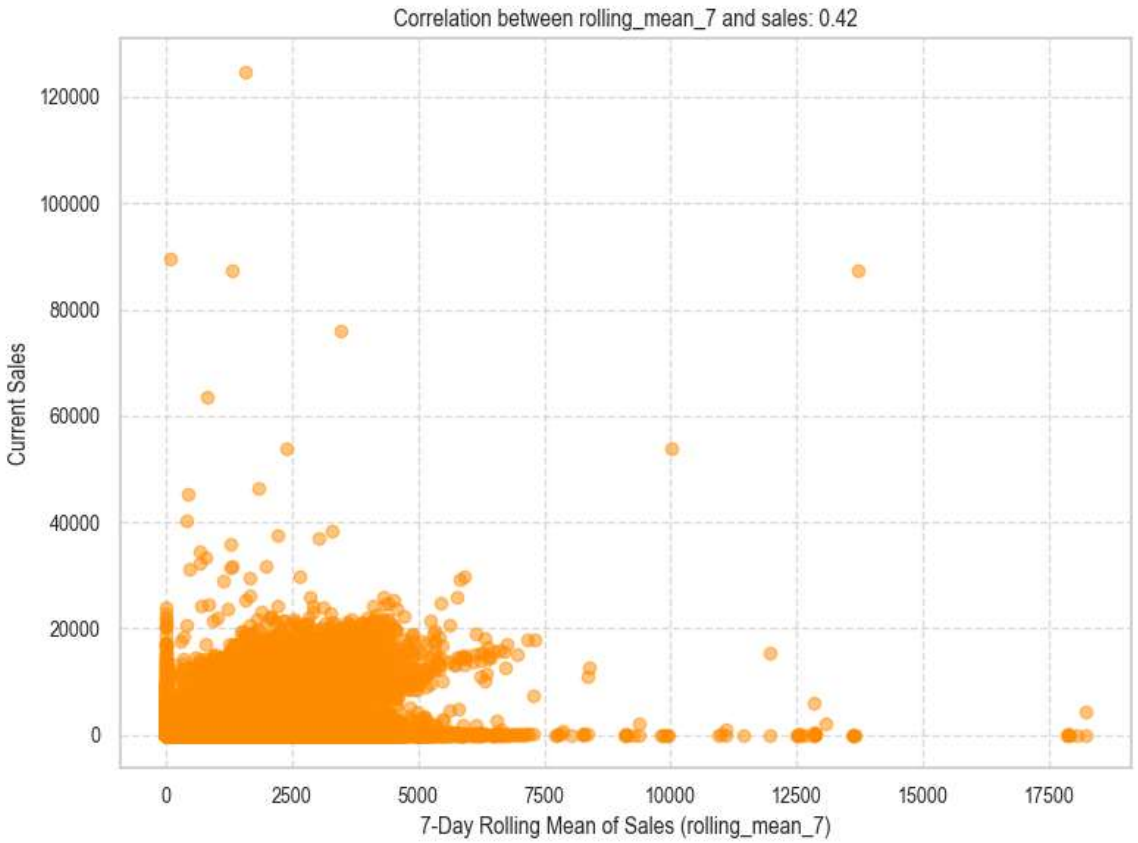
3.25 How do 7-day lag features (sales_lag_7) correlate with current sales?

- **Findings:** The correlation between *sales_lag_7* (sales from the previous 7 days) and current sales is *0.93*.
- **Interpretation:** This indicates a very strong positive linear relationship. Sales levels in the immediate past are highly indicative of current sales levels.
- **Implication:** Historical sales data, particularly recent lags, is extremely valuable for accurate sales forecasting. [1] This feature should be central to predictive modeling efforts.



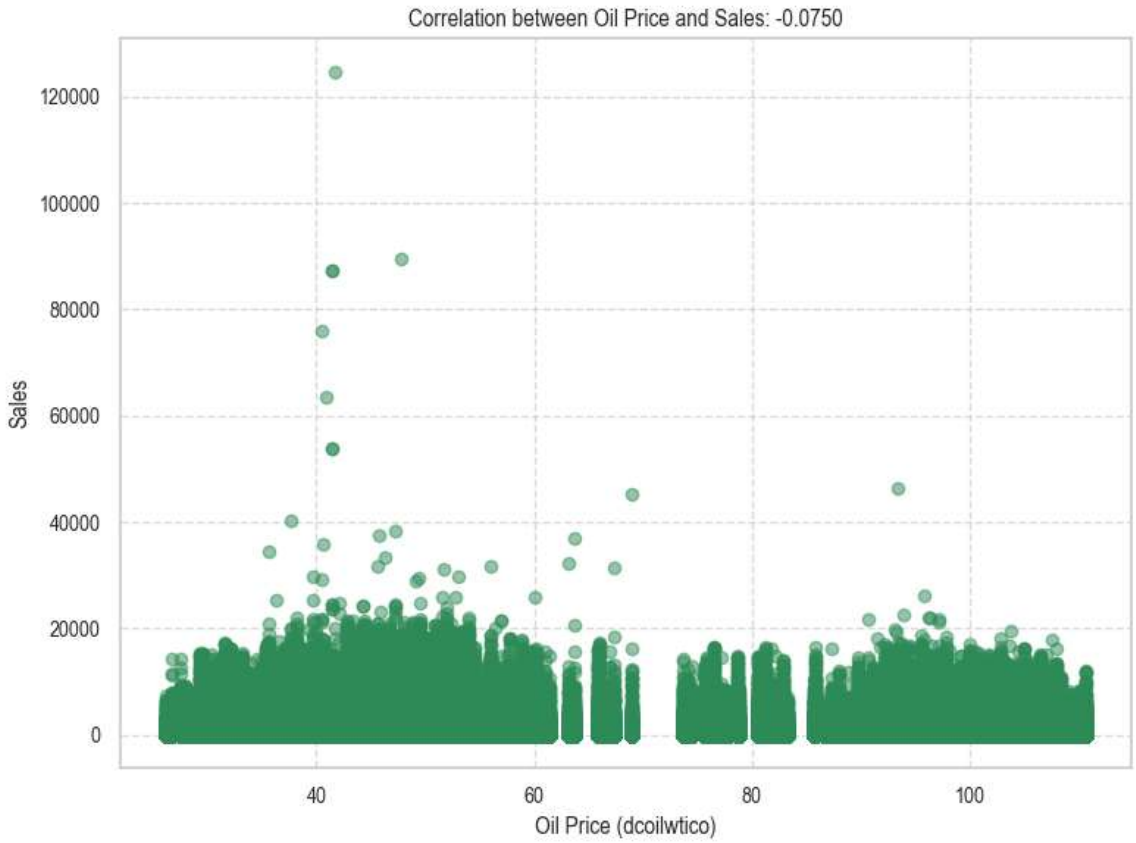
3.26 What does the rolling mean of sales tell us about seasonality or stability?

- **Findings:** The 7-day rolling mean of sales (*rolling_mean_7*) has a moderate positive correlation (*0.42*) with daily sales.
- **Interpretation:** The rolling mean effectively smooths short-term noise and captures underlying trends or seasonality to some extent. However, the moderate correlation suggests that significant daily variations exist beyond what the 7-day average captures.
- **Implication:** While useful for identifying broader sales stability and trends, the rolling mean alone isn't sufficient to explain all daily sales fluctuations, indicating the influence of other short-term factors (like specific events, daily promotions, etc.).



3.27 Is there a correlation between oil prices (dcoilwtico) and sales behavior?

- **Findings:** The correlation between daily oil prices (*dcoilwtico*) and sales is *-0.075*.
- **Interpretation:** This value is very close to zero, indicating a negligible linear relationship. The negative sign suggests a very slight tendency for sales to decrease as oil prices increase (or vice-versa), but the effect is minimal.
- **Implication:** Based on this data, daily fluctuations in oil prices do not appear to be a significant direct driver of overall sales behavior. Other economic or market factors likely have a much stronger influence.



4. Conclusion

The preprocessing techniques and detailed business analysis provide deep insights into sales behavior, seasonal effects, promotion effectiveness, and crisis impact. These findings will guide further modeling efforts and strategic business decisions.