BUSA8000:

Techniques in Business Analytics

Macquarie Business School

Macquarie University, Sydney

6th of November 2024

Group Project: LuminaTech Lighting

Prepared by Group 29:

Group Members	Student IDs
Isidora Gautier	48622613
Manuel Cabeza	48622605
Akash Bhatt	48596485
Pranjal Joshi	48614084

Word count = 6568

Contents

Section 1: Data Cleaning

1.1 Reading and Concatenating Data Bases

We were provided with two large datasets, 2012_Data and 2013_Data. While importing them into the Jupyter Notebook, we encountered a UnicodeDecodeError, indicating a mismatch in the file encoding. We changed the encoding from 'utf-8' to 'ISO-8859-1' to resolve this. Since both datasets share identical column headers, we decided to concatenate them into a single DataFrame, ignoring the original indices of the two DataFrames and reassigning a new, continuous index for the combined DataFrame. This step simplifies our data cleaning and analysis by allowing us to work with the combined data more efficiently.

1.2 Missing Values

a) The combined data frame has 41 columns and 1,988,382 rows. We identified that ‘item_source_class’ is composed entirely of null values and, thus, decided to drop this column.
b) For Currency there were 2 values with blank values (“ “), we checked the customer code and as that customer made its purchases in AUD, we replaced them with AUD.

1.3 Identifying Wrong Entries and Unique Values

a) As written in the Metadata document, some entries from the ‘order_type_code’ were labelled as ‘Do Not Use’; therefore, we transformed all those into “do not use” (‘EDS', 'OBS', 'PPD', 'PM0', 'PGS', 'SPL', 'ZOP', 'ZC2', '5TN', 'PPO', 'ZD3', 'CSO').
b) The currency column contained a ‘AUS’ values, which we replaced with “AUD” as it was a misspelling.

1.4 Standardize Currency

a) Initially the data set had the currencies “AUS”, “AUD”, “US”, NZD and 2 blank values. To identify the currency corresponding to the blank values, we identified the customers for those entries, and their common currency of use was selected to replace those blanks. As conducted in section 1.3, we replaced AUS with AUD.
b) To be able to analyse the data all currencies were converted to AUD using the average yearly rate found in the Reserve Bank of Australia. (Reserve Bank of Australia. (2024). Exchange Rates | Chart Pack., n.d.)

1.5 Correct data types

a) Initially, the date columns accounting_date, invoice_date, and order_date were stored as objects, causing potential inconsistencies in analysis. To standardize them, we converted these columns to string format for compatibility, then transformed them into a datetime format ('%Y%m%d'), using error handling to replace invalid dates with NaT. This ensures consistency and prepares the dates for analysis.

b) To maintain data integrity, we validated calendar_day and calendar_month ranges, confirming they fell within 1-31 and 1-12, respectively. We examined negative values in value_sales, value_cost, and value_quantity, uncovering significant negative counts that require further investigation. We also reviewed unique values in key categorical columns, including currency, order_type_code, and environment_group_code, to detect any inconsistencies and ensure uniformity for future analysis.

1.6 Strip Whitespace from Categorical Data

a) Initially, the environment_group_code column had extra whitespace that could affect analysis. We used the str.strip() method to remove this whitespace and verified the unique values afterward, ensuring they were properly formatted. This step simplifies comparisons and enhances data quality by eliminating unnecessary spaces from the entries.

b) We reviewed all categorical columns for uniformity by iterating through them and printing their unique values twice to spot any discrepancies. To further ensure consistency, we stripped trailing whitespace from a list of key categorical columns, including customer_code, item_code, and business_area_code. Additionally, we converted customer_order_number to a string format, preparing the data for reliable categorical analysis.

1.7 Analyse and Classify Negative Transactions

a) We identified transactions with negative value_sales, value_cost, and value_quantity, revealing significant negative values. We labelled transactions with negative values as 'Return/Adjustment' and the rest as 'Regular,' summarizing the overall distribution. We then broke down 'Return/Adjustment' transactions by order_type_code and calculated their total negative financial impact of -34,208,395.07 in value_sales.

1.8 Order Type Insights

To improve our understanding and usage of order types, we analyzed transaction types alongside company codes to identify consistent usage patterns.

• **NOR and NOH (Normal Orders and Head Office Sales)**

- **Regular Transactions**: Steady order volumes; NOH shows exceptionally high maximum values (up to 60,000), indicating large head office sales.
- **Returns/Adjustments**: Negative average values for returns, with most returns being small. Misuse of NOR and NOH for returns/adjustments was noted.

• **CRD (Credit Price Only) and CRR (Credit Goods Return)**

- **Returns/Adjustments**: CRR shows significant negative values, indicating bulk returns. CRD generally indicates price adjustments with no impact on physical stock.

• **CPR (Credit Project Goods Return) and CRP (Credit Project Price Only)**

- **Returns/Adjustments**: CPR shows moderate negative averages, suggesting use for project order adjustments or cancellations.

• **PMO (Project Order - Moonlighting)**

- **Regular Transactions**: Moderate average values, with high maximums suggesting bulk project orders.
- **Returns/Adjustments**: Notable one-time bulk return of -100 for Company 101.

• **ZCG (Credit Goods Return - Pnz Only) and ZCR (Credit Price - Pnz Only)**

- **Regular Transactions**: ZCR affects price only, not quantity.
- **Returns/Adjustments**: ZCG shows moderate returns for Company 950 (median of -5).

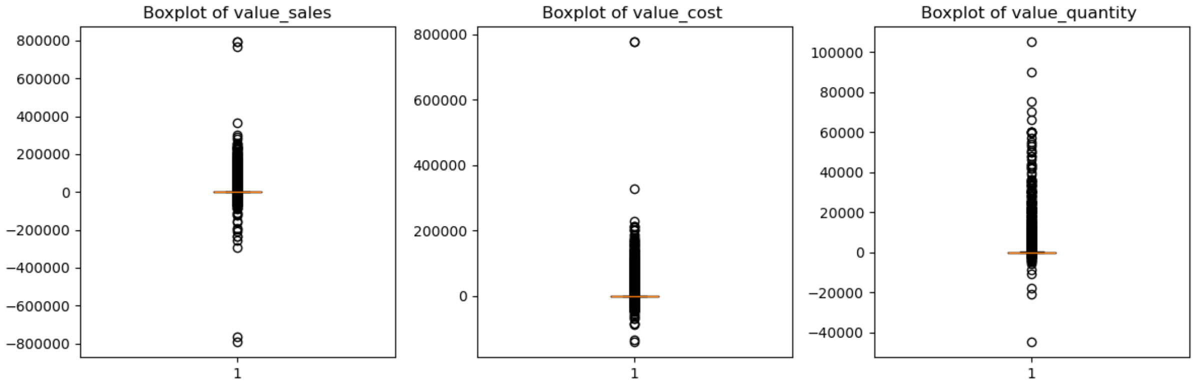
Additional Observations:

- **Export Orders (EXP)**: High average and peak quantities, especially for Company 100 (max 105,000), indicating significant international orders.
- **Specialty Orders (SPC)**: Positive values, with maximums around 200, indicating small custom orders for Company 120.

1.9 Outliers

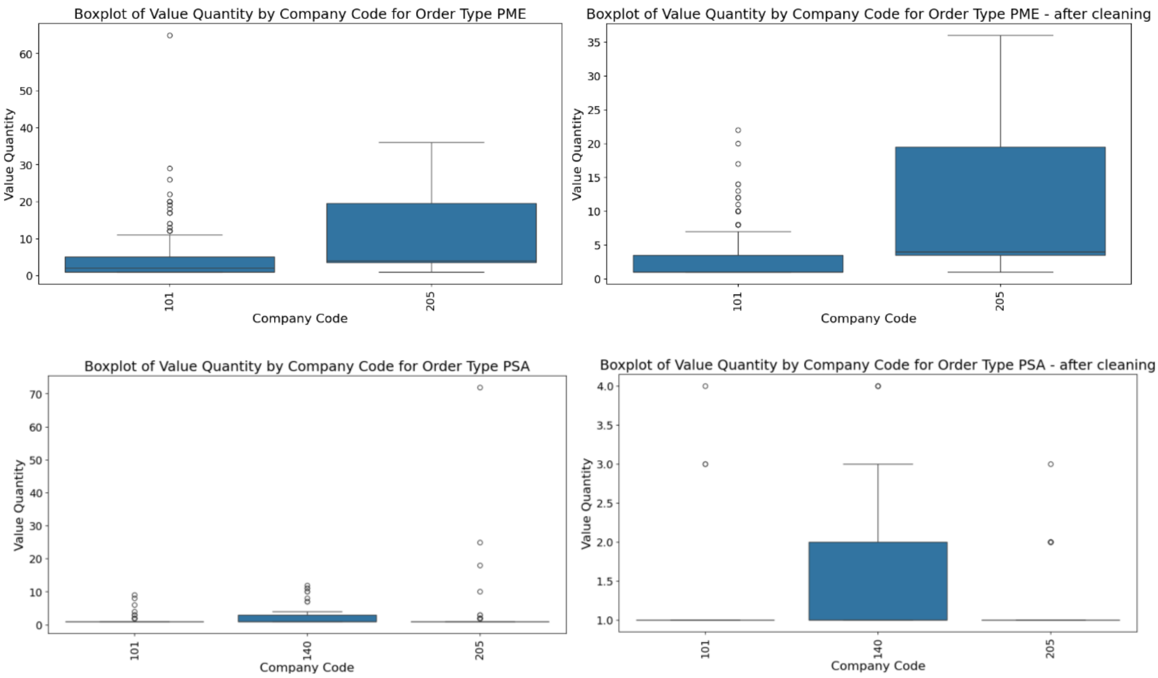
In Figure 1 below, the plots reveal a substantial number of outliers in the data, reflecting the wide variance across different items, ABC classifications, order types, purchasing patterns by company, and varying demand by district. To manage these outliers effectively, we will segment the data by order_type_code, ABC class volume, customer_district, and company_code. This approach enables more precise and customized outlier detection by accounting for both transaction type and unique company-specific patterns.

Figure 1: Outliers for total sales, cost and quantity



In Figure 2, we present box plots for value_quantity by Company Code, comparing Order Types PME and PSA both before and after the data cleaning process. Although some outliers persist in the box plots after cleaning, this can be attributed to the fact that these plots are two-dimensional representations based solely on two variables. The data cleaning was conducted on a broader subset involving four features. As a result, while the cleaning process effectively removed outliers based on this more comprehensive multidimensional analysis, some points may still appear as outliers when viewed through the lens of only these two dimensions.

Figure 2: Quantity Box Plot by Company Code for Order Type PME and PSA before and after cleaning



Next, we applied the Interquartile Range (IQR) method to eliminate outliers from each subgroup. For example, we specifically examined the outliers in value_quantity for the NOR code type, focusing on ABC Class G, Company Code 100, and Customer District Code 720.

Section 2: Exploratory Insights

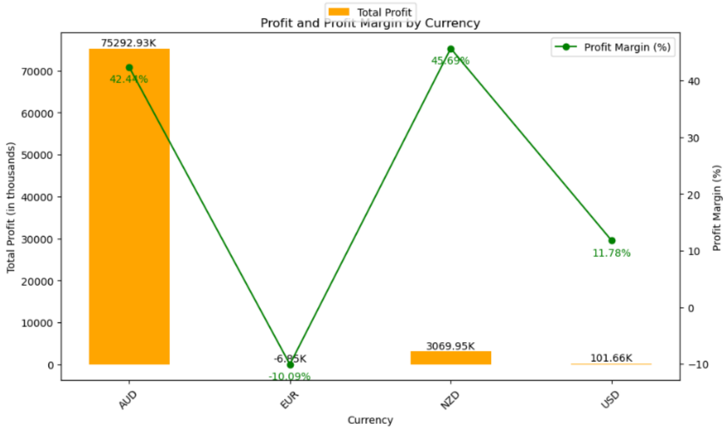
2.1 Profit Margin by Currency

2.1.1 Observation: The analysis of total sales and profit by currency reveals distinct patterns. AUD transactions contribute the highest total profit, demonstrating strong sales volume and effective cost management. The profit margin for AUD is substantial at 42.44%, indicating that

Figure 4 Sales and Profit by currency

Figure 3: profit and profit margin by currency

operations in this currency are highly profitable. In contrast, EUR transactions show a concerning trend with a total profit close to zero and a negative profit margin of -10.09%, suggesting inefficiencies or high costs relative to sales. NZD transactions have a high profit margin of 45.69% but a lower total profit, reflecting limited sales volume. USD, with a moderate 11.78% profit margin, indicates potential but lower transaction volume.



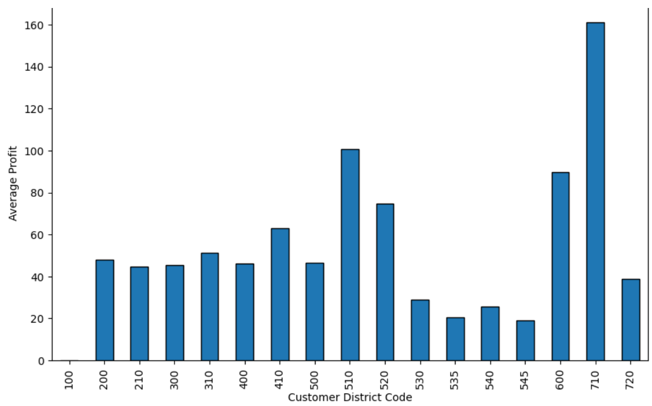
2.1.2 Value for Management

This analysis is critical for strategic decision-making. The high profitability in AUD underscores the importance of focusing resources on expanding operations in regions using AUD. The negative margin in EUR necessitates an urgent review of cost structures or pricing strategies to prevent financial losses. Despite NZD's limited volume, the high margins suggest an opportunity for growth with targeted marketing or product expansion. Enhancing sales volume in USD regions could yield further profitability, given the stable margins. These insights enable management to allocate resources efficiently, optimize pricing strategies, and identify areas for operational improvement based on concrete financial metrics.

2.2 Average Profit by Customer District Code

2.2.1 Insight: The analysis of average profit across customer district codes shows substantial variability. Districts 510, 600, and 710 have notably higher average profits, indicating these areas are key contributors to overall profitability. In contrast, districts such as 545 and 535 report much lower average profits, highlighting potential inefficiencies or challenges that may be affecting profitability in these regions.

Figure 5 Average profit by customer distict code



2.2.2 Value for Management

This analysis provides essential data-driven insights for strategic planning. By identifying the most profitable districts (e.g., 510, 600, and 710), management can focus efforts on maximizing returns through enhanced service or targeted investments in these areas. Conversely, addressing the low-profit districts (e.g., 545 and 535) could involve re-evaluating local strategies, adjusting operational costs, or refining marketing initiatives. These findings enable precise, region-specific decisions that align resources effectively, ensuring sustained profitability and performance optimization across all districts.

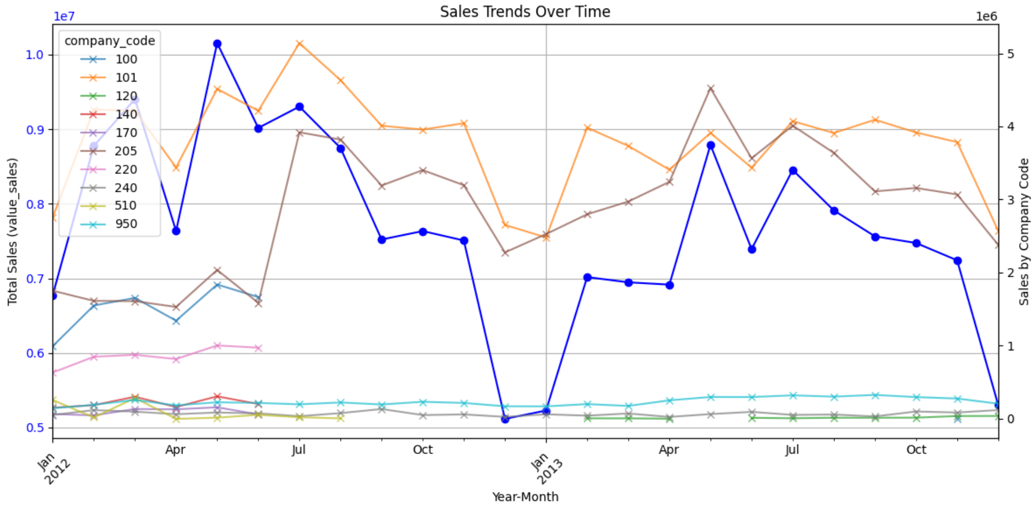
2.3 Sales trend over time

To analyse sales trends over time, we utilized the value_sales_aud feature to ensure all sales figures are expressed in the same currency. We then plotted the total value of sales and segmented it by company_code to assess the contributions of each company to overall sales performance.

2.3.1 Insight and Value: In Figure 3, the line chart depicts monthly fluctuations in total sales (value_sales_aud), represented by the blue line with dots. This visualization highlights peaks, troughs, and observable seasonal patterns. Notable dips and spikes occur in specific months, likely influenced by external factors such as seasonal demand or internal factors like marketing campaigns. Additionally, including sales data by company_code offers insights into individual company behaviours and their contributions to overall sales trends.

1. Top Performer: Company 101 consistently achieves the highest sales throughout the period, with only two months where Company 205 surpassed it.
2. Market Shifts: Five companies exited the market after August 2012, during which Company 205 absorbed much of the customer base, leading to nearly doubling its sales.
3. Seasonal Vulnerability: Company 101 experiences significant sales drops every January, potentially due to the post-holiday season, indicating that this period disproportionately impacts Company 101 compared to others.
4. Peak Season: The data reveals a consistent seasonal spike in May each year, marking it as the peak sales month.
5. Steady Performers: Companies 240 and 950 maintain low but stable revenue, suggesting a consistent, albeit smaller, customer base or product line.

Figure 6 Total Sales over time and by company



Understanding these sales patterns provides critical insights for demand forecasting and inventory management. The observed seasonal peak in May and decline in January allows the management team to:

1. Optimize Inventory: Adjust stock levels to align with peak and low sales months, reducing holding costs while ensuring sufficient inventory during high-demand periods.
2. Strategize Marketing Efforts: Launch targeted marketing campaigns for companies like 101 in January to mitigate sales dips, potentially leveraging promotions to retain customer engagement.
3. Evaluate Market Strategy: Investigate the shift in market share following the exit of certain companies after August 2012, assessing how company 205 effectively captured this demand and if there are opportunities for further consolidation.
4. Plan for Growth: For stable companies (240 and 950), consider strategies to grow their customer base, such as cross-selling or expanding their product range, given their reliable but modest performance.t also profitability.

2.4 Revenue and Profit Margin by Light Source (2012 vs 2013)

To evaluate how the different light sources performed we first the need to filter the data for each year and the calculate the profit margin (sales – cost / sales).

2.4.1 Insight and Value:

Figure 7 Revenue and Profit Margin by Light Source

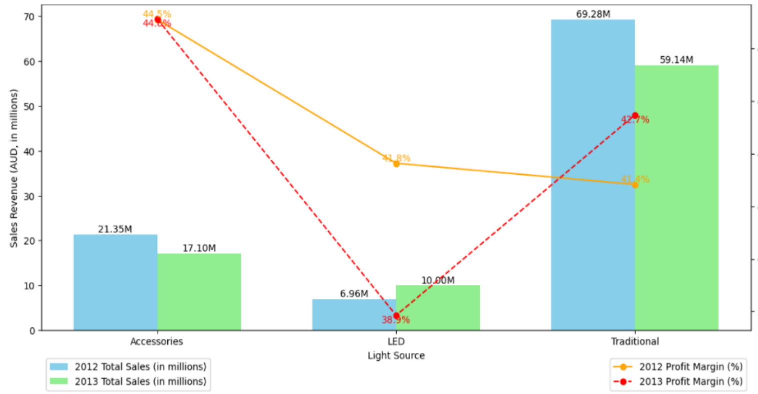
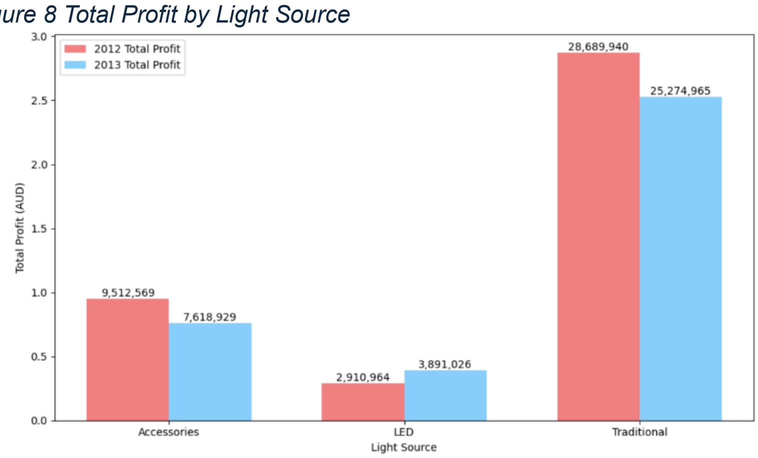


Figure 8 Total Profit by Light Source



In 2012, "Traditional" lighting led in revenue (Figure 7), indicating that traditional light sources were the primary sales driver. Profit margins across categories were steady, ranging from 41-45%, with LEDs showing growth potential due to comparable margins despite lower revenue. By 2013, LED sales revenue rose by 43%, pointing to increased demand for sustainable lighting. Although traditional lighting remains strong, a slight revenue decline suggests a consumer shift toward LEDs.

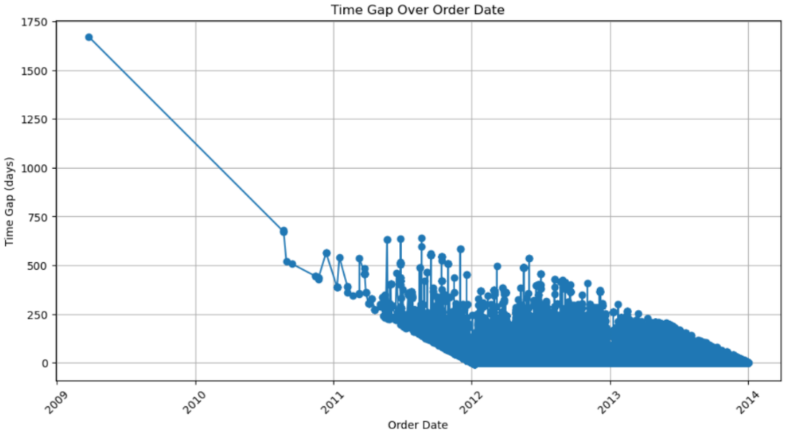
The profit margin for LEDs dropped in 2013, possibly due to scaling challenges or pricing pressures, even as its total profit increased by 33% (Figure 8). Meanwhile, total profits for Traditional and Accessories lighting decreased by 12% and 20%, respectively, indicating potential loss of competitiveness.

2.4.2 Recommendation: Given these trends, sustaining revenue from Traditional lighting is important, but expanding LED growth is key. Enhancing operational efficiency or negotiating supplier agreements could help improve LED margins as demand grows. Figure 8 shows that while Traditional and Accessories lighting saw a decline in total profit (-12% and -20% respectively), LED experienced a 33% increase, even with a slight 2.9% margin drop. This suggests Traditional and Accessories lighting may be losing competitiveness. Allocating resources toward LEDs while maintaining minimal investment in Traditional and Accessories could help the company stay aligned with emerging market trends without over-investing in declining categories.

<u>2.5 Time gap between order date and invoice date, as per warehouse and customer regions</u>

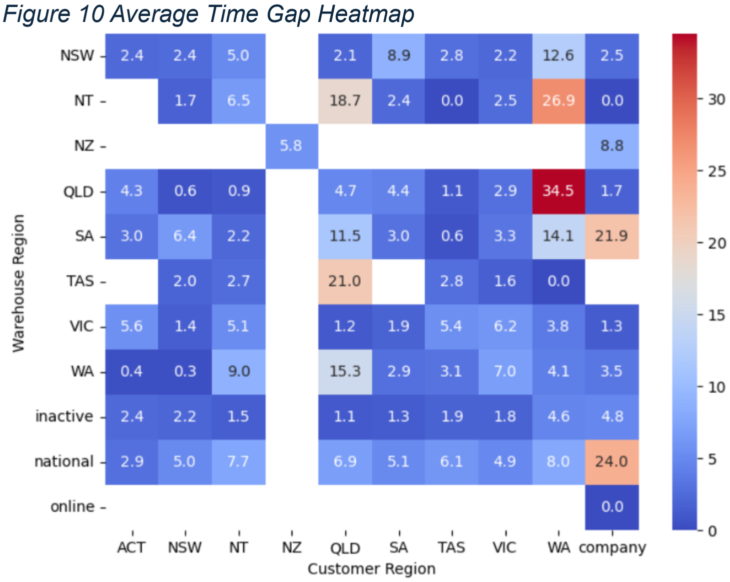
To analyze the time gap between order and invoice dates across warehouse and customer regions, we created a new column called time_gap to capture the difference between

Figure 9 Time Gap Over Order Date



invoice_date and order_date. We mapped this out in a time series chart, which showed the typical intervals between order placement and invoicing, along with some significant outliers. For example, there were delays of up to 1750 days (around 5 years) and 700 days (~2 years). Since these extreme cases were rare, we decided to treat the largest gap as an outlier and removed it. We also identified 22 rows with negative time gaps, falling below the IQR's lower bound, and chose to drop these entries to keep our dataset clean and accurate. Next, we calculated the average time_gap between different warehouse_region and customer_region locations. Creating these columns allowed us to group and cross-analyze processing times, giving us deeper insights into how invoicing times varied across different warehouse and customer regions.

Figure 10 Average Time Gap Heatmap



2.5.1 Insights:

1. Significant delays at warehouses: We found that QLD warehouses face a 34.5-day delay in processing invoices for VIC customers, which is unusually high compared to other regions served by QLD. This may point to logistical issues or specific invoice processing delays. Additionally, the national warehouses (Pierlite National) experience a 24-day delay for 'company' region customers, such as Intercompany and Head Office Sales. This delay could be due to internal processes like varied invoice cycles, unique audit policies, or lower priority for internal sales. Some Intercompany and Head Office sales are also in USD, which might contribute to further invoicing delays.

2. Challenges with WA customer region: WA's distance from QLD and NT warehouses causes extended shipping times, which in turn delays both order fulfillment and invoicing (as invoices are processed post-delivery). Time zone differences also contribute to invoicing delays, with WA orders potentially falling into later processing batches due to eastern time zone cutoffs.

3. Geographical bottlenecks: Warehouses in TAS experience significant delays when servicing QLD customers, primarily due to the distance and need for multiple courier transits (air and land). This highlights inefficiencies in supply allocation for TAS-stocked items that could be better positioned in warehouses closer to QLD customers.

2.5.2 Recommendations:

1. **Investigate High-Delay Routes:** Focus on the QLD-VIC routes to identify specific causes of delays and streamline processes for these connections. Develop better inventory allocation in VIC warehouses to reduce transportation delays and effectively serve VIC customers.
2. **Enhance Regional Logistics for Distant Warehouses:** For warehouses like NT and WA that serve distant customer regions, consider establishing partnerships with local logistics providers or implementing route-specific efficiency measures to reduce transportation delays.
3. **Replicate Best Practices from Efficient Warehouses:** I We recommend studying key processes from high-performing warehouses (e.g., NSW and VIC) and applying them in regions with inconsistent performance, such as NT, QLD, and SA. Standardizing the order and invoice processing systems across all warehouses could also help minimize human errors and speed up invoice processing.
4. **Reallocate Inventory:** By analysing high-demand items among QLD customers currently served by TAS warehouses, we can identify inventory needs and increase these items in QLD warehouses, reducing delivery and invoicing delays.

Section 3: Test Sub Sample Differences

3.1 Is there a significant difference between the average sales for Company 101 (company with the highest sales) between 2012 and 2013?

We conducted a two-sample t-test to determine if there was a significant difference in the average sales for Company 101, the company with the highest sales (totalling $91,077,433.19), between 2012 and 2013. The significance level used was 5%.

- Null Hypothesis (H0): There is no difference in average sales for company 101 between 2012 and 2013.
- Alternative Hypothesis (H1): There is a significant difference in n average sales for company 101 between 2012 and 2013.

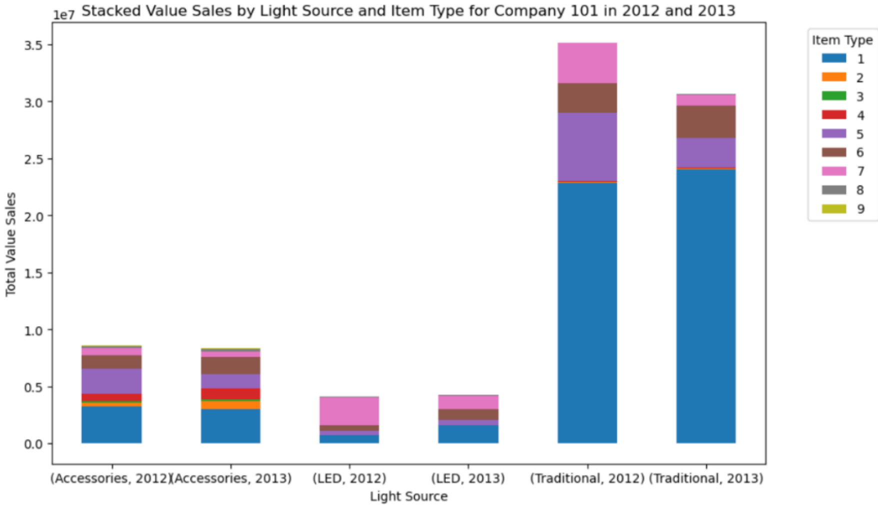
```
  company_code  value_sales_2012  value_sales_2013
1          101       47790398.50       43287034.69
T-statistic: -47.824220202635225
P-value: 0.0
```

The Results:

The T-Statistic: -47.82 is a very high in absolute terms and the negative sign indicates that the average sales in 2013 were lower than in 2012. This large absolute value indicates a significant difference between the two-sample means - the average sales of Company 101 in 2012 and 2013.

The P-Value: 0.00 is below the level of significance of 0.05, indicating that there is a statistically significant difference between the average sales for Company 101 in 2012 and in 2013; therefore, we reject the null hypothesis. As seen in section 2.3 above, Company 101 sales started dropping in Q3 2012 by reaching an all-time low in January 2013.

Figure 11: Company 101's Sales by Light Source stacked by item type



Analysis of Sales Drop

Breaking down sales by light_source and item_type revealed that Company 101's decline was mainly due to decreased sales of items 5 and 7, which are key products within its traditional light sources. Although sales for item 1 increased, it wasn't enough to offset the drop from items 5 and 7.

Recommendations:

1. Conduct a Sales Audit: Identify factors contributing to the decline in items 5 and 7, such as supply chain disruptions, increased market competition, or pricing and consumer trends.

2. Conduct Customer Surveys: Host focus groups or feedback interviews with major customers who purchase items 5 and 7 and assess product quality and satisfaction levels.

3. Investigate items 1 and 6: These two items are the only ones in the traditional light sources that increased sales from 2012 to 2013. Look at what lead to their sales increase and evaluate if similar factors can be implemented to boost sales for items 5 and 7.

3.2 Is There a Difference in Average Quantity Sold/Sales Value Between Trade and Professional Bonus Groups?

We want to assess whether there is a significant difference in the average quantity sold between the Trade and Professional bonus groups so we will perform a two-sample t-test to compare the means of value_quantity between the two groups using a 5% significance level.

- Null Hypothesis (H0): There is no difference in average quantity sold between the Trade and Professional bonus groups.

- Alternative Hypothesis (H1): There is a difference in average quantity sold between the Trade and Professional bonus groups.

The results are: Mean Quantity Sold (Trade): 9.23, **Mean Quantity Sold (Professional):** 10.73, **T-statistic:** -8.02, **P-value:** 1.02e-15

The same steps were followed to determine if there was a statistically significant difference in the average sales value between the two bonus groups obtaining the following **results: Mean Sales Value (Trade):** 71.91, **Mean Sales Value (Professional):** 376.64, **T-statistic:** -307.93, **P-value:** 0.0

3.2.1 Can the difference in performance be explained by the types of orders in each bonus group?

Figure 12 Professional Summary

Professional Summary:

[78]:

	bonus_group_code	order_type_code	order_count	total_revenue	total_quantity
11	Professional	NOR	161679	39895572.20	1261029.85
15	Professional	PRD	12205	16698491.01	101348.00
16	Professional	PRO	16630	14290048.78	94119.74

Figure 13 Trade Summary

Trade Summary:

[79]:

	bonus_group_code	order_type_code	order_count	total_revenue	total_quantity
33	Trade	NOR	1056663	71201197.56	10470539.00
29	Trade	EDI	91700	10392653.68	856517.00
32	Trade	NOH	1284	2809249.40	107293.00
38	Trade	PRO	6100	2671003.62	55063.00
37	Trade	PRD	4115	2344866.23	48948.00

3.2.2 Conclusion and recommendations: The analysis reveals significant performance differences between the Professional and Trade groups. The Professional group shows higher average quantities sold (10.73 vs. 9.23) and a much higher average sales value ($376.64 vs. $71.91), with both differences statistically significant (p-values < 0.05). The substantial performance differences between the Professional and Trade groups can be attributed to the different order types each group tends to focus on. The Professional group generates more revenue from fewer transactions, particularly in order types NOR and PRD, with impressive revenues of AUD 39,895,572.20 and AUD 16,698,491.01 (figure 10), respectively. On the other hand, the Trade group has a higher transaction volume but lower average sales value per order, indicating a volume-driven approach (figure 11).

Given the Professional group's superior performance, it is recommended that management explore and adopt some of the strategies used by this group. Enhancing the Trade group's approach through training or restructuring incentives to align with higher revenue goals could foster improvements and create a more balanced and competitive environment.

Section 4: Inference

4.1 Which factors are correlated with the average unit cost for “SUR” business area code?

4.1.1 Steps for the model creation:

1. *Filter Data for the Surface Product Group:* We created a new DataFrame containing only the rows where the product belongs to the "SUR" group, which represents Surface type products. This group is the second most purchased category in our dataset.
2. *Remove Rows with Zero Quantity:* Rows with quantity = 0 were eliminated because these represent revenue adjustments and do not contribute meaningful information for calculating the average unit cost.
3. **Calculate Unit Cost:**A new feature called unit_cost was created by dividing the value_cost_aud by value_quantity. This provides a per-unit cost for each transaction.
4. **Feature Selection for Analysis:** We selected features that could potentially be correlated with the unit cost of the "SUR" products. These features include: 'calendar_month', 'calendar_day', 'company_code', 'customer_district_code', 'item_class_code', 'bonus_group_code', 'environment_group_code', 'warehouse_code', 'abc_class_code', 'business_chain_l1_name', 'order_type_code', and 'currency'.
5. **Group Data to Calculate Mean Unit Cost:** We grouped the data by all the unique combinations of the selected features. For each unique combination, we calculated the mean unit cost, resulting in a summarized dataset that represents the average cost per subgroup.
6. **Convert Categorical Variables with One-Hot Encoding:** Categorical features were converted into numerical format using one-hot encoding, which transforms each categorical value into a binary (0 or 1) column. This step is necessary for modelling purposes.
7. **Check for Multicollinearity Using VIF:** We calculated the Variance Inflation Factor (VIF) for all features to check for multicollinearity, which occurs when independent variables are highly correlated with each other.
8. **Eliminate Features with Infinite VIF:** Features with VIF = infinity were eliminated iteratively. During this process, we discovered that 'environment_group_code' was correlated with 'warehouse_code', so all related features were removed. Similarly, we found multicollinearity issues with 'company_code', which led to its removal as well.

9. **Remove Features with High VIF:** We continued to eliminate features iteratively until all VIF values were below 10. Eventually, we achieved VIF < 5 for all features, ensuring minimal multicollinearity.

10. **Run Scaler for numerical values**

Since most of the variables are binary (with values of 0 or 1), we applied scaling to the numerical values to bring them to a similar range. This normalization helps ensure that all features contribute equally to the model, preventing any particular numerical feature from dominating due to scale differences.

11. **Run OLS Model to Determine Feature Significance:** We ran an Ordinary Least Squares (OLS) model with unit_price as the dependent variable and the selected features as independent variables. This allowed us to obtain the p-values for each feature, indicating their significance in explaining the variability in the unit cost. We selected as factors correlated with mean_unit_price all the features whose p-value was below 0.05 significance level.

12. *Selected Features for Analysis:* The features selected through this analysis are shown in Figure 11, grouped by their higher-level categories ("parent features"). A detailed list of all features used in the analysis is provided in Appendix 1.

Figure 14 Features groups correlated with unit price

```
Feature Group Counts:
            Feature Group  Count
0          calendar_month      1
1            calendar_day      1
2           mean_quantity      1
3  customer_district_code      4
4         item_class_code     15
5          warehouse_code     26
6          abc_class_code      8
7   business_chain_l1_name      8
8         order_type_code      4
9                currency      1
```

4.1.2 Explanation of the Method Used:

The method used here is Ordinary Least Squares (OLS) regression, which is part of multiple regression analysis. The goal of this analysis is to examine the relationship between independent variables (e.g., warehouse, item class code, ABC classification) and the dependent variable (mean unit price). OLS minimizes the sum of squared residuals (differences between observed and predicted values) to find the best-fitting line through the data and indicates which features are statistically significant correlated with the dependent variable.

The model includes 69 predictors (shown in the coefficients table). Significant predictors (P-value < 0.05) include ABC class codes, item class codes, customer district codes, warehouses, order types, and other factors. All these coefficients are statistically significant, as indicated by the p-values, suggesting that these predictors have a meaningful impact on the mean unit price.

4.1.3 Coefficient Insights:

1. A**BC Class Code J Increases Cost:** If the product has an ABC class code J, the cost increases by 11.73 units. This is intuitive, as products in class J are Make to Order, which implies additional customization or production requirements, thus resulting in higher costs.
2. **ABC Class Code B and C Comparison:** The ABC class code B has a coefficient of -2.94, which indicates a lower cost compared to class C (coefficient = 1.67). This makes sense since B-class items are typically mid-level sellers that may benefit from better economies of scale, while C-class items are lower-tier, often leading to lower volume discounts and thus higher relative costs.
3. **Currency EUR Increases Cost:** The use of EUR as the currency leads to an increase in cost by 701.68 units. This substantial increase could explain the lower profit margins observed for products sold in euros, potentially due to exchange rate effects, higher tariffs, or operational costs in European markets. This insight is valuable for the management team, as it highlights areas where pricing strategies need to be adjusted, or currency hedging should be considered to manage profitability.
4. **Warehouse EN0 Has Higher Costs:** Products sold from Warehouse EN0 have an increase in cost by 122.87 units, indicating this warehouse has significantly higher operational or logistical expenses. Management should investigate the cost structure of this warehouse to understand why it has higher associated costs, which might be due to labor costs, transportation expenses, or inefficiencies.
5. **Item Class Codes Impact:**

   Several item class codes significantly impact costs. For example:
   a. SUR80 has a coefficient of 64.01, indicating a higher unit price, which might be due to this item type having premium features or materials.
   b. Conversely, SUR02 has a strong negative effect (-31.96), suggesting this type of item is significantly cheaper. Understanding these variations helps the management team adjust pricing strategies or promotions effectively.
6. **Order Type Code Impact:** Products ordered with the EXP (export) code see an increase of 3.11 units in mean cost, while orders via EDI have a negative coefficient of -5.69, meaning they are less costly. This suggests that express orders are driving costs up, likely due to expedited processing, whereas Electronic Data Interchange (EDI) orders are more cost-effective, perhaps due to automation. Management could encourage customers to use EDI for bulk or regular orders to reduce costs.

7. **Customer District Code Impact:** Certain customer district codes (e.g., district 300 and 310) are associated with higher costs, possibly due to delivery logistics or specific regional conditions. Understanding this allows management to explore ways to mitigate these costs, such as optimizing delivery routes or working on local distribution partnerships.
8. **Seasonality Effects (Calendar Variables):** The calendar month and calendar day both have positive coefficients (0.2156 and 0.0342, respectively), indicating some seasonality in costs. It implies that average costs increase over time, possibly due to demand fluctuations, seasonal price adjustments, or supply chain variations. This is valuable for management to anticipate and plan for cost variations during specific periods of the year.

4.1.4 Why these Insights Are Valuable for the Management Team:

1. **Strategic Pricing and Cost Management:** The analysis provides insights into which factors drive up costs (e.g., item class, currency, warehouses), enabling management to adjust pricing strategies or focus on cost-cutting initiatives. **Operational Efficiency:** By identifying warehouses or regions with higher costs, management can target specific areas for process optimization or logistical improvements to reduce costs.
2. **Customer Segmentation and Targeting:** The insights around ABC class codes and customer districts provide valuable information for customer segmentation. Management can use this to develop targeted campaigns, offering discounts for higher-cost items or optimizing product assortments by region.
3. **Encouraging Efficient Order Types:** Promoting EDI orders can help reduce transaction costs, and this information can be used to create incentive programs for customers to switch to more cost-effective order channels.

4.2 Which factors define the quantity or sales?

4.2.1 Steps for the Model Creation:

1. **Identify Product Group with High Sales**:

Identify the business area generating the highest sales revenue grouped the data by business_area_code and calculated the total value_sales for each group. The results were sorted in descending order to highlight the top-performing area, which was identified as "SUR" (Surface) as the business area with the highest total sales.

2. **Data Filtering**:

Created a subset of the data for the "SUR" business area, named model_sales, to focus on the most significant segment. Filtered out rows where value_quantity was zero, removing records that did not contribute to sales.

3. **Feature Selection and Encoding**:

First, excluded features with extreme variability or irrelevance, such as warehouse_code, market_segment, and salesperson_code. Then, transformed calendar_year into ordinal values for modeling purposes. Finally, Handled missing values in categorical features by replacing them with "Missing" and applied one-hot encoding to convert categorical variables into a numerical format suitable for modeling.

4. **Multicollinearity Analysis Using Variance Inflation Factor (VIF)**:

Calculated VIF for each feature to detect multicollinearity, then identify and drop features with VIF values exceeding 10 to reduce multicollinearity, resulting in a refined dataset with all VIF values below the acceptable threshold.

5. **Run OLS Model for Feature Significance**:

First, we ran an Ordinary Least Squares (OLS) regression model with value_sales as the dependent variable. Then, dropped features with p-values greater than 0.05, indicating statistical insignificance, to retain only meaningful predictors. Finally, rechecked VIF to ensure multicollinearity was minimal.

6. **Check P values for coefficients from OLS results**

Significant predictors, their coefficients, and p-values were analyzed to understand their impact on sales.

4.2.2 Explanation of the Method Used: The method utilized is Ordinary Least Squares (OLS) regression, a statistical technique used to model the relationship between independent variables and a dependent variable (sales in this case. It allows identification of predictors that significantly influence sales, highlighted by their p-values and coefficients.

4.2.3 Insights:

1. **High Positive Impact on Sales**:

- *Business Chain Name*: CetraPro Distributors has the highest positive coefficient (5335.96), indicating a substantial contribution to sales.
- *Company Code 240*: Contributes 4668.70 units, highlighting a key area for focusing sales efforts.
- *Currency - USD*: Positively influences sales by 3121.64 units, suggesting the importance of USD-based transactions.
- *Customer District Codes:* 410, 510, 600, and 710 each have a notable positive effect, pointing to key geographic areas driving sales.

2. **High Negative Impact on Sales**:

- *Order Type Code*: CPR has the most negative influence, reducing sales by 989.98 units. This may indicate inefficiencies or challenges with this order type.
- *Currency - EUR*: Decreases sales by 541.00 units, suggesting a strategic need to manage transactions in this currency.
- *Bonus Group Code - Trade*: Reduces sales by 312.90 units, which could be indicative of ineffective bonus structures.

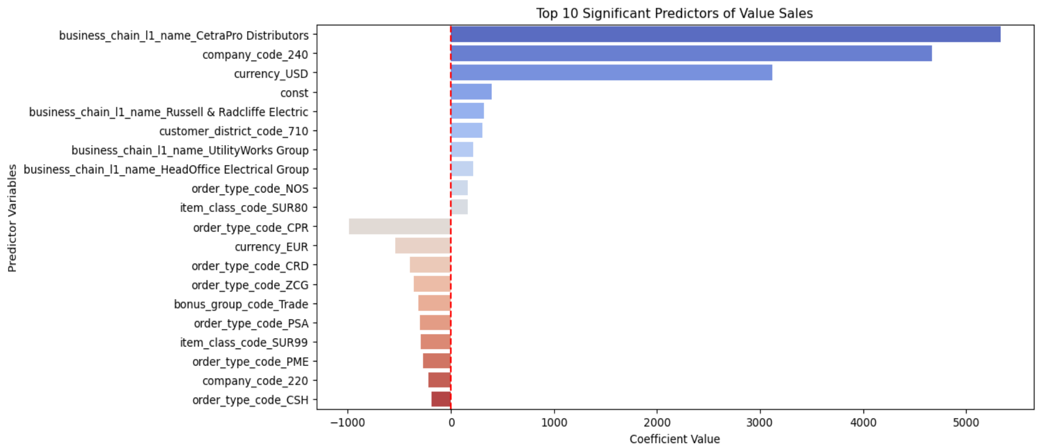
3. **Item Class Codes and Sales**:

- *SUR80* (160.94 units) and *SUR09* (133.79 units) are associated with higher sales, while classes like *SUR99* (-289.20 units) and *SUR90* (-151.20 units) have a strong negative impact.This variation suggests the need for differentiated marketing strategies for different item classes.

4. **Order Types and Sales**:

Certain order types, like NOS (167.20 units), boost sales, while others, such as CPR and PSA, have a negative impact, highlighting areas to optimize or reevaluate.

Figure 15 Top 10 Significant Predictors of Value Sales



4.2.4 Why These Insights Are Valuable for the Management Team:

1. **Sales Strategy Optimization**: By focusing on factors that drive sales, like high-performing business chains and favourable currencies, the management team can prioritize and refine sales strategies.
2. **Geographical Targeting**: Understanding the impact of customer district codes allows for better-targeted marketing campaigns and resource allocation.
3. **Order Type Management**: Identifying order types that negatively impact sales provides opportunities to improve processes or offer alternative solutions to customers.

4. **Product Mix Adjustments**: Insights into item class performance can guide inventory management and promotional strategies, ensuring that high-demand products are prioritized.

Section 5: Prediction Model

5.1 Steps for Model Creation:

1. **Independent Variable Selection:** To ensure the model's predictive capabilities, we carefully selected a set of features relevant to the sales price. The chosen independent variables include: 'calendar_year', 'calendar_month', 'calendar_day', 'company_code','customer_district_code', 'item_code', 'item_class_code', 'bonus_group_code','environment_group_code', 'abc_class_code', 'business_chain_l1_name','order_type_code', 'currency', 'value_cost', and 'value_quantity'. The 'value_sales' feature was excluded from this selection to prevent data leakage, as it directly correlates to the target variable we aim to predict.

2. **Data Splitting:** The dataset was split into training and validation subsets, with the training set comprising data from 2012 and early 2013, while the validation set included data from late 2013. This temporal separation allows for robust model evaluation and ensures that the model's performance reflects its ability to generalize to unseen data.

3. **Target Encoding and Scaling:** Categorical features underwent target encoding to convert them into numerical representations while preserving their relationship with the target variable. This encoding enhances the model's understanding of these features. Numerical features were then scaled to ensure they are on a similar scale, thus allowing the Random Forest model to leverage these features effectively.

4. **Model Training:** A Random Forest Regressor was chosen for its effectiveness in capturing complex relationships within the data. The model was trained using 100 estimators to balance bias and variance, optimizing predictive performance.

5. **Model Evaluation:** The model's performance was assessed using several metrics:

   - **Mean Absolute Error (MAE):** 25.53, indicating the average absolute difference between predicted and actual sales prices.

   - **Mean Squared Error (MSE):** 56973.43, reflecting the average squared difference, which emphasizes larger errors.

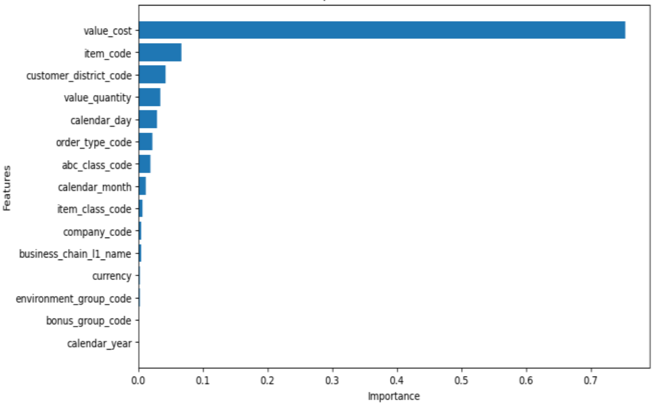
   - **R-squared:** 0.7569, suggesting that approximately 75.69% of the variability in sales prices can be explained by the model.

6. **Feature Importance Analysis:**

The analysis revealed that 'value_cost' had the highest importance score, reinforcing the understanding that item cost is a critical factor in determining sales prices.

'item_code' and 'customer_district_code' also played significant roles, suggesting that specific product and geographical information significantly influence pricing strategies.

Figure 16: Feature Importance from Random Forest Model



7. **Preparing for 2014 Predictions:**

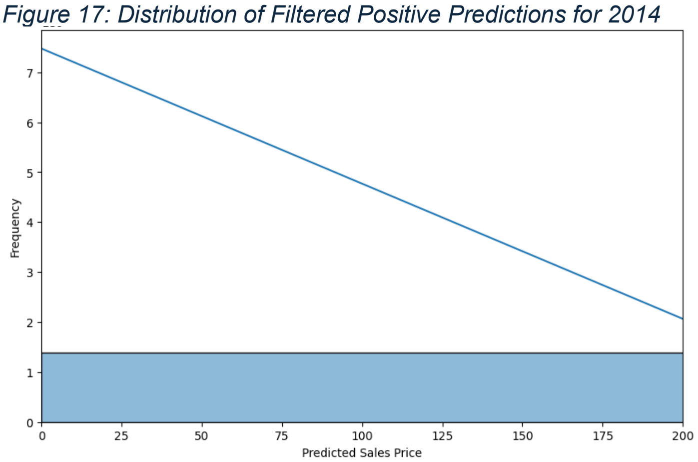
A new dataset for predicting sales prices in 2014 was created by modifying the 'calendar_year' to 2014 and retaining the selected features. This step ensures that the model can make predictions in a relevant context.

Sales Price Prediction: The model was then utilized to predict sales prices for 2014, yielding results that included: **Mean Predicted Sales Price:** 120.18; **Max Predicted Sales Price:** 51101.30

A total of 90292 negative predictions were identified and filtered out, primarily attributed to returns and adjustments, leaving a refined set of predictions with a mean positive sales price of: **Mean Positive Predicted Sales Price:** 133.53

8. **Distribution Analysis:** A distribution plot was generated for the filtered positive predictions, illustrating the range and frequency of predicted sales prices. This visualization helps in understanding how sales prices are expected to vary in 2014.

Figure 17: Distribution of Filtered Positive Predictions for 2014



5.2 Key Observations:

- **Value Cost:** The model highlights 'value_cost' as the most important predictor of sales prices, indicating a direct relationship where increases in cost correspond to increases in sales price. This insight suggests that effective cost management could enhance pricing strategies.

- **Item Code Influence:** The significance of 'item_code' in predictions points to the necessity of analyzing specific product categories to determine which items yield higher sales prices, potentially guiding inventory management and marketing efforts.
- **Geographical Factors:** The importance of 'customer_district_code' underlines regional pricing strategies. Certain districts may be associated with higher prices, indicating opportunities for targeted marketing campaigns based on geographical insights.
- **Negative Predictions:** The removal of 90292 negative predictions emphasizes the importance of data quality and accurate representation in pricing models, suggesting that attention must be paid to the implications of returns and adjustments on sales forecasting.

5.3 Conclusion: The prediction model developed effectively estimated sales prices for 2014 based on historical data. The insights drawn from the model emphasize the critical role of cost in determining sales prices and identify specific features that contribute to pricing predictions. The mean positive predicted sales price of 133.53 provides a realistic expectation for management, allowing for informed decision-making regarding pricing strategies and inventory management for the upcoming year.

Section 6: Higher Likelihood of Losing Customers

6.1 Method explanation:

To identify features associated with a higher likelihood of customer churn, we focused on the Metro Electrical Distributors business chain, which had the highest total sales value. To streamline the analysis, we filtered the data by specific parameters: bonus_group_code='trade', technology_group_code='SYLV', and business_area_code='LMP'.

For better understanding, we focused on a limited period. We defined January 1st, 2013, as the starting point and calculated the average purchase time between customers in the dataset. We then set the analysis period as three times the average purchase time. All customers with an average purchase time greater than this period were excluded from the analysis.

Our feature engineering process included calculating the purchase frequency and quantity for each customer, organizing purchase dates chronologically, and determining metrics like days between purchases and average buying intervals. We also aggregated all purchases made by a customer in a day into a single row for better data consistency.

We set a churn threshold as the average buying interval plus twice the standard deviation for each customer and flagged customers exceeding this threshold as churned. Based on these criteria, the resulting churn rate was 44.68%.

Following our approach based on previous insights, we identified a primary set of features that could impact customer churn: 'customer_district_code', 'item_type', 'value_sales_aud', 'days_between_orders', 'avg_days_between_orders', 'days_since_last_order', 'purchase_frequency', 'calendar_day', 'light_source', 'abc_class_code', 'value_quantity', and 'transaction_type'.

Categorical features were one-hot encoded, and the dataset was scaled to ensure comparability. We also calculated the Variance Inflation Factor (VIF) to mitigate multicollinearity, which led to the exclusion of 'abc_class_code_J' and 'abc_class_code_U'.

To identify churn-related features, we used logistic regression with churn as the dependent variable and all other features as independent variables.

Figure 18: Logistic Regression Result:

Dep. Variable:	churn	No. Observations:	1746
Model:	Logit	Df Residuals:	1724
Method:	MLE	Df Model:	21
Date:	Mon, 04 Nov 2024	Pseudo R-squ.:	0.7610
Time:	19:18:19	Log-Likelihood:	-268.32
converged:	True	LL-Null:	-1122.8
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-4.0415	6.31e+06	-6.4e-07	1.000	-1.24e+07	1.24e+07
customer_district_code_210	-0.0637	0.125	-0.507	0.612	-0.310	0.182
customer_district_code_300	-0.1666	0.139	-1.202	0.229	-0.438	0.105
customer_district_code_310	0.3921	0.125	3.131	0.002	0.147	0.637
customer_district_code_400	-0.5435	0.144	-3.769	0.000	-0.826	-0.261
customer_district_code_410	0.1926	0.123	1.567	0.117	-0.048	0.434
customer_district_code_500	-0.0444	0.123	-0.360	0.719	-0.286	0.197
customer_district_code_510	0.2521	0.162	1.560	0.119	-0.065	0.569
customer_district_code_600	0.3382	0.133	2.541	0.011	0.077	0.599
light_source_LED	0.0659	0.240	0.275	0.783	-0.404	0.536
light_source_Traditional	0.1113	0.151	0.735	0.462	-0.185	0.408
abc_class_code_B	0.5710	0.163	3.499	0.000	0.251	0.891
abc_class_code_C	-6.9090	9.3e+07	-7.43e-08	1.000	-1.82e+08	1.82e+08
abc_class_code_D	-0.1063	0.115	-0.927	0.354	-0.331	0.118
abc_class_code_I	0.0225	0.194	0.116	0.908	-0.358	0.404
transaction_type_Return/Adjustment	-0.3149	0.124	-2.546	0.011	-0.557	-0.072
value_sales_aud	-0.1782	0.136	-1.309	0.191	-0.445	0.089
days_between_orders	-0.2312	0.158	-1.468	0.142	-0.540	0.078
avg_days_between_orders	-4.4246	0.477	-9.267	0.000	-5.360	-3.489
days_since_last_order	5.1796	0.331	15.669	0.000	4.532	5.828
purchase_frequency	0.6963	0.235	2.968	0.003	0.237	1.156
value_quantity	0.1708	0.105	1.633	0.103	-0.034	0.376

After running the model, the statistically significant features at a 0.05 significance level were:

'customer_district_code_310', 'customer_district_code_400', 'customer_district_code_600', 'abc_class_code_B', 'transaction_type_Return/Adjustment', 'avg_days_between_orders', 'days_since_last_order', and 'purchase_frequency'.

6.2 Interpretation of the significant features from the logistic regression analysis:

1. 'customer_district_code_310' (coef = 0.3921, p-value = 0.002) and 'customer_district_code_600' (coef = 0.3382, p-value = 0.011):

Customers in these districts are more likely to churn compared to the baseline.

2. 'customer_district_code_400' (coef = -0.5435, p-value = 0.000):

Customers in district 400 are less likely to churn compared to the baseline. The negative coefficient indicates that these customers tend to remain more loyal.

3. 'abc_class_code_B' (coef = 0.5710, p-value = 0.000):

Products categorized as class B are significantly associated with higher churn rates compared to the baseline class. This means customers who purchase products in class B are more likely to stop purchasing in the future, possibly suggesting dissatisfaction or that these products do not drive customer loyalty.

4. 'transaction_type_Return/Adjustment' (coef = -0.3149, p-value = 0.011):

Customers with transactions marked as returns or adjustments are less likely to churn. This negative coefficient might suggest that the ability to return or adjust purchases leads to higher customer satisfaction and lower churn rates, perhaps due to increased trust or flexibility in transactions.

5. 'avg_days_between_orders' (coef = -4.4246, p-value = 0.000):

This negative coefficient indicates that as the average time between orders increases, the likelihood of churn significantly decreases. In other words, customers who have longer intervals between orders (i.e., they order less frequently) are less likely to churn, possibly because they maintain a consistent, if infrequent, purchasing pattern. It might reflect that these customers still see value in the products, even if their frequency is low.

6. 'days_since_last_order' (coef = 5.1796, p-value = 0.000):

The positive coefficient suggests that customers who have not placed an order for a long time are significantly more likely to churn. The longer the gap since the last order, the higher the churn risk, which is intuitive: as time without engagement increases, the probability of a customer not returning also increases.

7. 'purchase_frequency' (coef = 0.6963, p-value = 0.003):

Customers with higher purchase frequencies are more likely to churn. This positive coefficient might indicate a situation where customers make frequent purchases initially but then stop, suggesting possible dissatisfaction or reaching a saturation point. This pattern is sometimes seen when initial enthusiasm wears off, or customers have stockpiled enough of a product and decide not to buy more.

6.3 Summary

- Geographical Insights: Certain districts (310 and 600) show a higher likelihood of churn, while others (like 400) show loyalty. This can help in geographically targeted campaigns to reduce churn.
- Customer Experience (Returns/Adjustments): Customers who made returns or adjustments are less likely to churn, suggesting that providing easy return policies can positively impact customer retention.
- Customers who have a longer time since their last order are at high risk of churning, suggesting the importance of re-engagement campaigns to shorten this interval.
- However, customers with longer average intervals between orders might still be loyal in a non-frequent purchasing pattern.
- Frequent Buyers' Churn Risk: High purchasing frequency correlates with higher churn, which may indicate a need to maintain customer engagement and satisfaction for frequent buyers to avoid churn after initial enthusiasm fades.

6.4 Recommendations for Management

- Geographical Targeting: Focus retention efforts on districts with high churn risk (310 and 600), possibly by offering location-specific promotions or discounts.
- Product Strategy: Investigate the reasons behind higher churn for ‘abc_class_code_B’ products. Consider improving the quality or marketing message for these products.
- Return Policies: Emphasize and advertise flexible return policies, as they seem to build trust and reduce churn.
- Customer Engagement Campaigns: Develop targeted campaigns to re-engage customers before too much time passes since their last order. Offering promotions or reminders might be effective.
- Loyalty for Frequent Buyers: Implement loyalty programs or personalized offers to keep frequent buyers engaged and prevent churn after an initial spike in purchases.

Appendix

1. Coefficients for remaining features with p-value < 0.05:

Feature	Coefficient	Std Error	t-value	P-value	CI Lower 0.025	CI Upper 0.975
const	35.39	0.34	104.61	-	34.73	36.06
calendar_month	0.22	0.02	11.07	-	0.18	0.25
calendar_day	0.03	0.01	4.88	-	0.02	0.05
mean_quantity	-0.15	0.01	-22.32	-	-0.16	-0.14
customer_district_code_300	3.00	0.20	15.20	-	2.61	3.39
customer_district_code_310	3.94	0.44	8.90	-	3.07	4.80
customer_district_code_500	1.08	0.37	2.89	0.00	0.35	1.81
customer_district_code_510	3.47	0.54	6.40	-	2.41	4.53
item_class_code_SUR02	-31.96	0.30	-105.36	-	-32.56	-31.37
item_class_code_SUR03	-12.04	0.58	-20.88	-	-13.17	-10.91
item_class_code_SUR04	-18.85	0.28	-66.82	-	-19.41	-18.30

item_class_code_SUR05	- 21.32	0.32	- 65.97	-	- 21.95	- 20.69
item_class_code_SUR06	- 18.72	0.31	- 60.11	-	- 19.34	- 18.11
item_class_code_SUR07	10.94	0.34	32.05	-	10.27	11.61
item_class_code_SUR08	- 7.26	0.32	- 22.39	-	- 7.89	- 6.62
item_class_code_SUR09	22.22	0.55	40.72	-	21.15	23.29
item_class_code_SUR10	- 26.95	0.38	- 70.53	-	- 27.70	- 26.21
item_class_code_SUR11	- 31.59	0.37	- 86.10	-	- 32.31	- 30.87
item_class_code_SUR12	- 25.28	0.39	- 65.14	-	- 26.04	- 24.52
item_class_code_SUR14	- 32.45	0.61	- 52.92	-	- 33.65	- 31.25
item_class_code_SUR80	64.01	0.40	160.63	-	63.23	64.79
item_class_code_SUR90	- 29.98	0.38	- 79.76	-	- 30.72	- 29.25
item_class_code_SUR99	- 37.52	0.59	- 63.29	-	- 38.68	- 36.36
warehouse_code_1N1	33.60	0.33	101.21	-	32.95	34.25
warehouse_code_1Q0	32.58	0.42	77.54	-	31.75	33.40
warehouse_code_1Q1	28.37	0.74	38.47	-	26.92	29.82
warehouse_code_1S0	27.63	0.84	33.01	-	25.99	29.27
warehouse_code_1S1	30.75	1.24	24.75	-	28.31	33.18
warehouse_code_1T1	- 29.64	6.66	- 4.45	-	- 42.70	- 16.59
warehouse_code_1W0	34.10	0.42	80.74	-	33.27	34.93
warehouse_code_5N1	- 9.61	0.40	- 24.22	-	- 10.39	- 8.83
warehouse_code_5S0	- 2.48	0.47	- 5.32	-	- 3.40	- 1.57
warehouse_code_5S1	- 5.62	0.87	- 6.46	-	- 7.33	- 3.92
warehouse_code_5T0	- 4.10	1.02	- 4.01	-	- 6.10	- 2.09
warehouse_code_5V0	- 4.00	0.36	- 11.06	-	- 4.71	- 3.30

warehouse_code_5W0	-2.31	0.28	-8.19	-	-2.86	-1.76
warehouse_code_CS0	-4.78	0.65	-7.41	-	-6.05	-3.52
warehouse_code_CW0	-4.11	0.48	-8.52	-	-5.05	-3.16
warehouse_code_EN0	122.87	14.85	8.27	-	93.77	151.98
warehouse_code_GN0	44.28	1.65	26.92	-	41.06	47.50
warehouse_code_LN9	14.90	1.71	8.72	-	11.55	18.25
warehouse_code_LQ0	10.46	2.26	4.64	0.00	6.04	14.89
warehouse_code_LV0	7.01	2.27	3.09	-	2.56	11.46
warehouse_code_N0	7.79	0.43	18.12	-	6.94	8.63
warehouse_code_Q0	8.65	0.45	19.37	-	7.77	9.53
warehouse_code_Q1	7.20	0.67	10.82	-	5.90	8.51
warehouse_code_S0	4.80	0.75	6.36	-	3.32	6.27
warehouse_code_V0	5.13	0.48	10.72	-	4.19	6.07
warehouse_code_W0	10.71	0.49	21.69	-	9.75	11.68
abc_class_code_B	-2.94	0.23	-12.65	-	-3.40	-2.48
abc_class_code_C	1.67	0.26	6.55	0.03	1.17	2.17
abc_class_code_D	-0.55	0.26	-2.16	-	-1.05	-0.05
abc_class_code_E	-2.92	0.36	-8.08	-	-3.63	-2.21
abc_class_code_F	15.17	1.31	11.56	-	12.60	17.74
abc_class_code_G	-6.46	0.27	-23.89	-	-6.99	-5.93
abc_class_code_I	-2.67	0.32	-8.37	0.02	-3.29	-2.04
abc_class_code_J	11.73	0.21	55.92	-	11.32	12.14
business_chain_l1_name_Aussie Energy Group	0.77	0.32	2.37	-	0.13	1.40
business_chain_l1_name_Constructor Supplies	7.49	1.48	5.08	-	4.60	10.38
business_chain_l1_name_Corporate Builders Consortium	3.42	0.93	3.68	0.01	1.60	5.24

business_chain_l1_name_Corporate HQ Solutions	36.16	9.09	3.98	-	18.33	53.98
business_chain_l1_name_Global Electric Wholesalers	1.40	0.54	2.60	-	0.35	2.46
business_chain_l1_name_Inlite Solutions	32.23	3.45	9.35	-	25.47	38.99
business_chain_l1_name_PowerTools Direct	3.55	0.71	4.99	-	2.15	4.94
business_chain_l1_name_Russell & Radcliffe Electric	49.09	11.51	4.27	-	26.54	71.63
order_type_code_CRR	- 2.08	0.32	- 6.57	-	- 2.71	- 1.46
order_type_code_EDI	- 5.69	0.23	- 24.46	-	- 6.14	- 5.23

Bibliography

Reserve Bank of Australia. (2024). Exchange Rates | Chart Pack. (n.d.). Retrieved from https://www.rba.gov.au/chart-pack/exchange-rates.html