

# **Retail Sales Analysis Documentation (Nuel Assessment - 3)**

## **Problem Definition**

RetailPulse, a retail analytics company, seeks to optimize sales performance and customer engagement across its stores. The primary challenge is to understand the factors influencing sales, uncover regional and customer behavior trends, and assess the impact of external influences such as weather and holidays. This analysis will guide strategic decisions to drive revenue growth, improve product placements, and enhance overall store efficiency.

## **Objectives**

### **Understanding Sales Drivers**

The analysis aims to identify the primary factors contributing to sales performance. This includes evaluating the influence of pricing, product categories, promotions, and store-specific attributes. Additionally, external elements like weather and holidays will be assessed for their impact on sales fluctuations.

### **Analyzing Customer Behavior**

Customer segmentation will provide insights into how demographics such as age and gender, along with behavioral patterns like spending habits and purchase frequency, vary across locations. This will help uncover trends and regional differences in customer engagement.

### **Evaluating External Factors**

External conditions, including weather patterns and holiday periods, will be analyzed to determine their effect on sales. Understanding these relationships will aid in tailoring marketing strategies and inventory planning.

### **Key Performance Indicators**

The success of the analysis will be measured using the following KPIs:

- **Revenue Trends:** Tracking growth and identifying seasonal patterns.
- **Category Performance:** Highlighting top-performing and underperforming product categories.
- **Customer Retention:** Measuring loyalty and repeat purchase frequency.
- **Location-Based Insights:** Identifying high- and low-performing cities or regions.

## **Description of the Datasets Used and Their Relevance**

The analysis leverages three key datasets sourced from **Kaggle**, each offering unique and complementary insights into sales performance, customer behavior, and store operations. These datasets provide the foundation for understanding the drivers of sales and identifying actionable insights.

### **1. Sales Data**

- **Description:**  
This dataset contains transaction-level details such as sales date, product category, price, and total transaction amounts.
- **Key Columns:**
  - *Date:* Sales transaction date.
  - *Product Category:* Type of product purchased.
  - *Price per Unit:* The unit price of the product.
  - *Quantity:* Number of units sold.
  - *Total Amount:* Total sales amount for the transaction.
- **Relevance:**  
This dataset is crucial for analyzing sales trends, identifying top-performing products, and evaluating the impact of promotions and discounts on sales.

**Dataset Link:** <https://www.kaggle.com/datasets/mohammadtalib786/retail-sales-dataset>

### **2. Customer Data**

- **Description:**  
This dataset captures customer-specific information, including demographics, purchase behavior, and preferences.

- **Key Columns:**

- *Customer ID*: Unique identifier for customers.
- *Age and Gender*: Demographic attributes of customers.
- *Location*: The city or region where the customer resides.
- *Frequency of Purchases*: Indicates the customer's loyalty and purchase patterns.

- **Relevance:**

This dataset helps segment customers into high spenders and frequent buyers, analyze demographic trends, and assess behavior across different locations.

**Dataset Link:**

<https://www.kaggle.com/datasets/bhadramohit/customer-shopping-latest-trends-dataset>

### 3. Store Data

- **Description:**

This dataset provides operational data for stores, including product categories, weather conditions, and seasonality details.

- **Key Columns:**

- *Store ID*: Unique identifier for stores.
- *Weather Condition*: External weather factors impacting sales.
- *Seasonality*: Seasonal patterns influencing product demand.
- *Category*: Types of products offered by the store.

- **Relevance:**

This dataset is essential for evaluating store performance, understanding the influence of weather and seasonality on sales, and identifying regional trends.

**Dataset Link:**

<https://www.kaggle.com/datasets/anirudhchauhan/retail-store-inventory-forecasting-dataset>

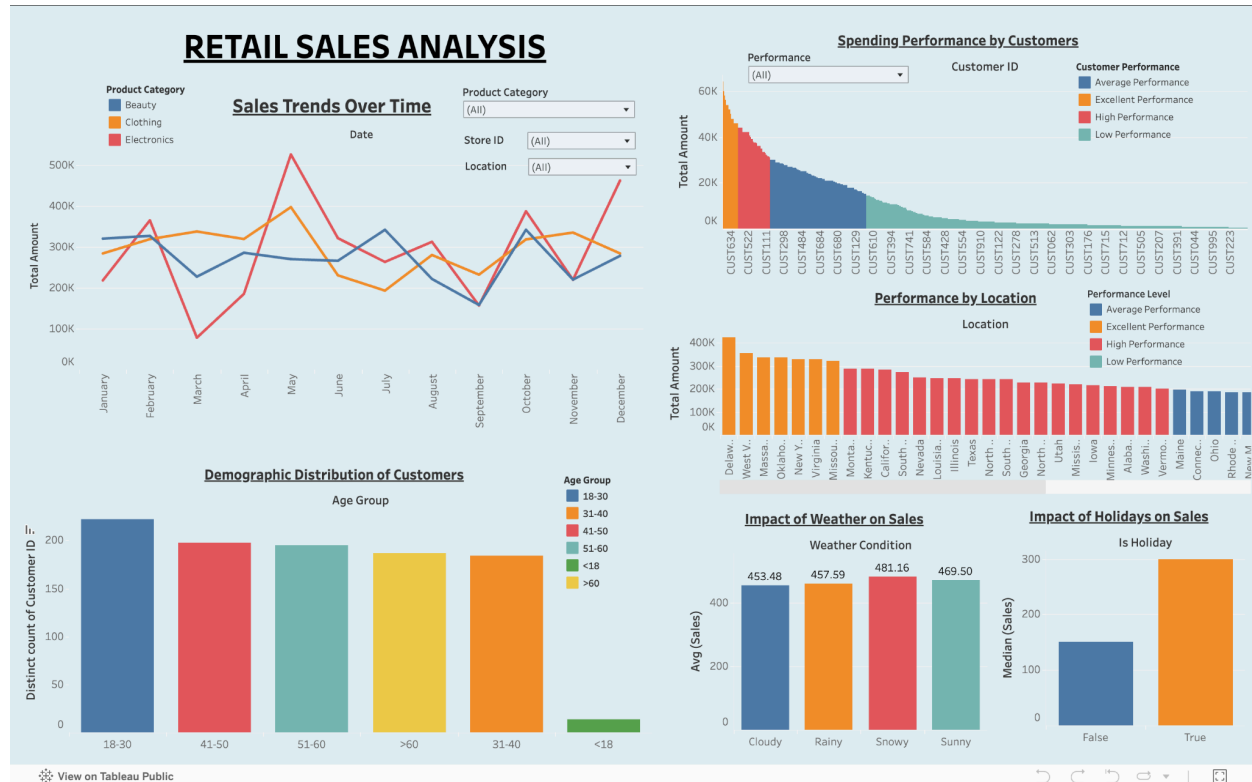
**Note:** All the datasets, python file, tableau workbook are compressed in the zip file.

## Tableau Dashboard

### Tableau Dashboard Link:

[https://public.tableau.com/views/RetailSalesAnalysisDashboard\\_17379234251960/RetailSalesAnalysis?:language=en-US&publish=yes&:sid=&:redirect=auth&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/RetailSalesAnalysisDashboard_17379234251960/RetailSalesAnalysis?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link)

### Dashboard Screenshot



## Key Findings

### *Sales Trends*

- **Electronics** sales were highly variable, peaking in **May** and seeing the lowest levels in **October** and **December**. There was a significant dip in **March**, and a gradual decline from **June** to **November**, with a slight recovery in **August**.
- **Clothing** sales had a clear peak in **May** but dropped to the lowest point in **July**, indicating seasonality and potential for targeted promotions during the peak period.

- **Beauty** products experienced steady sales up to **June**, with a spike in **July** and **October**. However, **September** saw the lowest sales, followed by a second-lowest dip in **November**.

### *Customer Insights*

- The largest groups of customers are between **18-30** and **41-50**, with over 200 and 198 customers respectively. These age groups make up a significant portion of the customer base and should be the focus of targeted marketing efforts.
- The **<18** age group is very small, with only 14 customers, indicating a potential opportunity to engage and attract younger buyers with tailored products or campaigns.

### *Location Performance*

- Top-performing locations include **Delaware**, **West Virginia**, **Massachusetts**, **Oklahoma**, and **New York**, each with sales exceeding \$300,000. These regions are key drivers of sales and should be prioritized for further expansion and marketing efforts.
- Low-performing locations such as **Kansas**, **New Hampshire**, **Arizona**, **New Jersey**, and **Idaho** have sales of less than \$100,000, suggesting a need for targeted strategies to boost performance in these regions.

### *Impact of external factors (weather and holidays)*

- **Weather conditions** have a noticeable impact on sales, with **snowy** weather yielding the **highest average sales** at 481.16, followed closely by **rainy** weather (457.59), suggesting that adverse weather conditions might drive consumers to spend more on indoor-related products.
- Sales during **holidays** show a significant boost, with the **median sales** on holidays being **three times higher** (300) compared to **non-holiday** sales (100), highlighting the importance of holiday promotions.

## **Challenges**

- **Data Integration Issues:** Differences in Customer ID formats across datasets caused challenges when merging data. For example, in the customer dataset, Customer IDs were prefixed with "CUST", while in the sales data, the IDs were just numbers. Additionally, variations in season names (e.g., "Fall" vs. "Autumn") required standardization to ensure consistency for accurate analysis.
- **Limited Geographic Details:** The absence of latitude and longitude in the datasets hindered the ability to create geographic heatmaps. As a result, we had to rely on city-level aggregation for analyzing location-based sales performance, which limited the granularity of insights.
- **API Integration Challenges:** Integrating weather data proved difficult due to limited free access and a cap on requests per day from various weather APIs. Attempts to use APIs like OpenWeather, Visual Crossing, Weatherbit, and Open-Meteo encountered restrictions, requiring additional adjustments to get the required data.
- **External Data Integration:** Integrating external datasets for holidays and weather from sources like Kaggle proved challenging due to the large size of the data and incompatible formats. Merging these datasets with sales data was difficult, especially due to the lack of a proper date match, which led to data alignment issues.
- **Missing Sales per Square Foot Metric:** The absence of a sales-per-square-foot column in the store dataset limited the ability to analyze store performance based on physical space, a key metric for evaluating the efficiency of each store.
- **Duplicate Columns:** Several datasets contained duplicate columns with similar or identical data, leading to unnecessary complexity in data cleaning. Identifying and removing these redundancies was a time-consuming process but crucial for accurate analysis.

## **Recommendations**

- **Electronics Sales Variability:** Electronics sales are unpredictable, with big peaks in May and drops in March and December. Therefore, launch special offers and new products around May to capitalize on the peak. For the slow months like March and December, plan targeted marketing to boost sales during these dips.
- **Clothing Sales Drop in July:** Clothing sales fall in July. Therefore, old inventory clearance events or offer mid-year discounts in July to keep sales up during the summer slowdown and prepare for the busy months ahead.
- **Targeting the Right Customer Groups:** The largest customer groups are in the 18-30 and 41-50 age ranges. Focus your marketing efforts on these groups. Use social media for the younger crowd and email marketing for the older group to engage them more effectively.
- **Improving Sales in Underperforming Locations:** Locations like Kansas and New Jersey have sales below \$100,000. Therefore, run local promotions and tailor your product offerings to better match these regions' needs to drive more sales.
- **Weather-Driven Sales:** Sales are higher during snowy and rainy weather. When the weather gets bad, focus on promoting indoor products like electronics or home goods to take advantage of the increase in consumer spending during such times.
- **Holiday Sales Surge:** Sales during the holidays are three times higher than non-holidays. Prepare special holiday promotions well in advance. Offering exclusive discounts and bundles will help you capture the full potential of these high-demand periods.