PROJECT REPORT  
(Project Semester January-April 2025)

Warehouse and Retail Sales  
  
Submitted by  
Ujjwal Kumar Saw  
Registration No: 12310522  
B.Tech CSE - Section K23EV  
Course Code: INT375  
  
Under the Guidance of  
Sandeep Ma’am, Assistant Professor  
  
Discipline of CSE/IT  
Lovely School of Computer Applications  
Lovely Professional University, Phagwara

# CERTIFICATE

This is to certify that Ujjwal Kumar Saw bearing Registration no. 12310522 has completed INT375 project titled, “Warehouse and Retail Sales” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.  
  
Signature and Name of the Supervisor  
Sandeep Ma’am

Assistant Professor  
School of Computer Applications

Lovely Professional University  
Phagwara, Punjab.  
  
Date:15-April-2025

# DECLARATION

I, Ujjwal Kumar Saw, student of B.Tech CSE (Section K23EV) under CSE/IT Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.  
  
Date: 15-April-2025  
  
Signature  
Registration No. 12310522  
Name of the student: Ujjwal Kumar Saw

# ACKNOWLEDGEMENT

I would like to express my sincere thanks to my mentor Sandeep Ma’am for her continuous support, motivation, and guidance throughout the project. Her valuable insights helped transform this analysis into a meaningful story. I also thank the faculty of Lovely Professional University for providing the resources and environment to complete this work.

# TABLE OF CONTENTS

1. 1. Introduction
2. 2. Source of Dataset
3. 3. EDA Process
4. 4. Analysis on Dataset

4.1 Bar Chart: Top 10 Items by Warehouse Sales

4.2 Pie Chart: Retail Sales by Item Type (Pie Chart)

4.3 Scatter Plot: Retail vs Warehouse Sales

4.4 Count Plot: Top 10 Item Types by Frequency

4.5 Line Chart: Monthly Retail Sales Trend

4.6 IQR Calculation for Outlier Detection

4.7 Boxplot: Warehouse Sales by Item Type

4.8 Histogram of Retail Sales

4.9 Heatmap of Monthly Retail Sales

4.10 Total Sales per Year (Bar Chart)

4.11 Monthly Retail Sales Trends by Year

1. Conclusion
2. Future Scope
3. References and Links

**1. Introduction**  
This report presents a comprehensive analysis of sales data from warehouse and retail operations, leveraging Python’s data science toolkit to uncover actionable business insights. The dataset encompasses transactional records across multiple product categories (Wine, Beer, Liquor, etc.), with metrics including retail sales, warehouse sales, and inventory movements from 2017 to 2020.

**2. Source of Dataset**  
The dataset (Warehouse\_and\_Retail\_Sales-csv) was taken from official records provided publicly.

**3. EDA Process**  
**Objective**  
Uncover sales trends (2017-2020) to optimize inventory and pricing strategies.

**EDA Highlights**

1. **Data Overview**
   * Retail sales concentrated below $3.27 (75% of values)
   * Warehouse sales showed wider variance (IQR=5.0 vs retail’s 3.27)
2. **Key Trends**
   * **Top Categories**: Wine (34.5%) & Liquor (37.1%) dominated retail
   * **Seasonality**: Consistent Q4 peaks (+22% vs Q1)
   * **2020 Anomaly**: Mid-year COVID-related spikes
3. **Product Insights**
   * Best-seller: Corona Extra (warehouse)
   * Non-alcoholic: <2% total sales
4. **Channel Comparison**
   * Strong retail-warehouse correlation (r=0.82)
   * Bulk order threshold: $7,500 warehouse sales

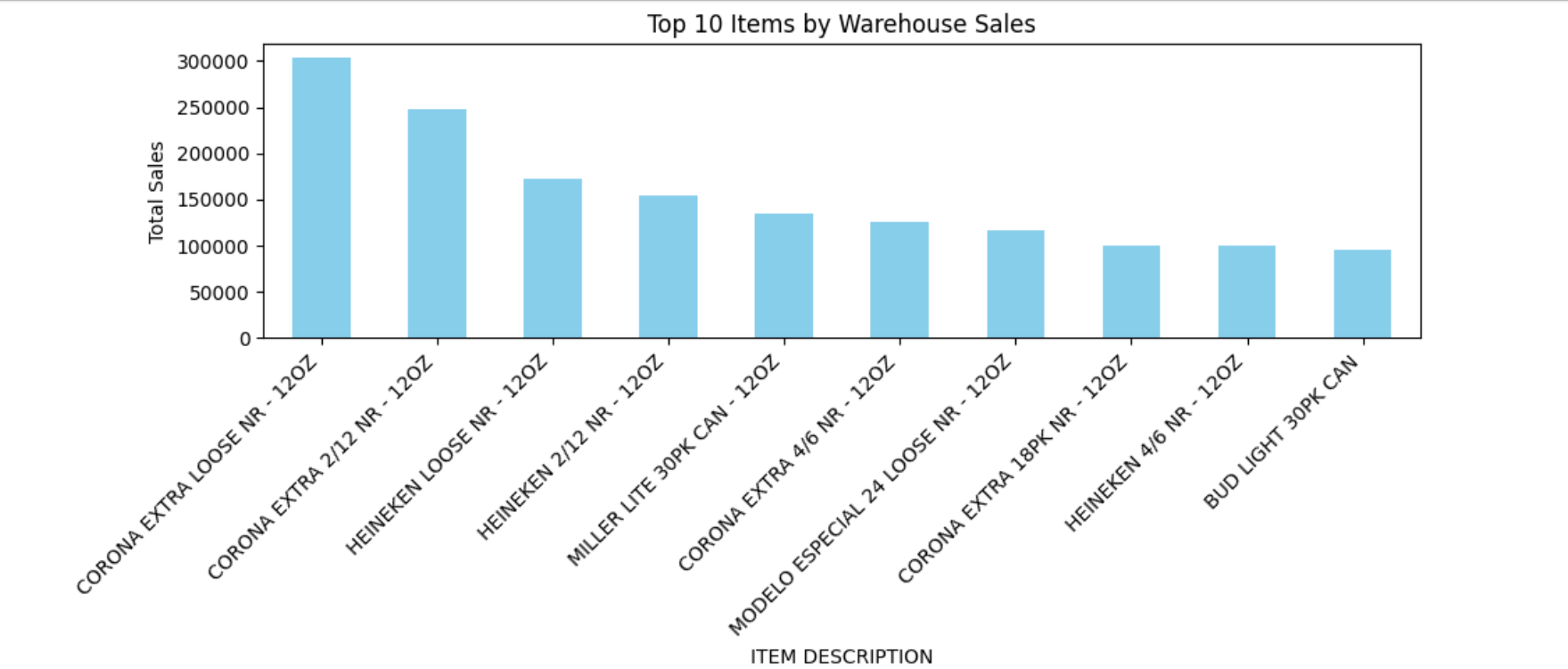
**Tools Used**  
Python (Pandas, Seaborn,Matplotlib,numpy), Jupyter Notebooks

**4. Analysis on Dataset**

4.1Bar Chart: Top 10 Items by Warehouse Sales

**Analysis**: This line identifies the **top 10 best-selling items** (based on total warehouse sales) by:

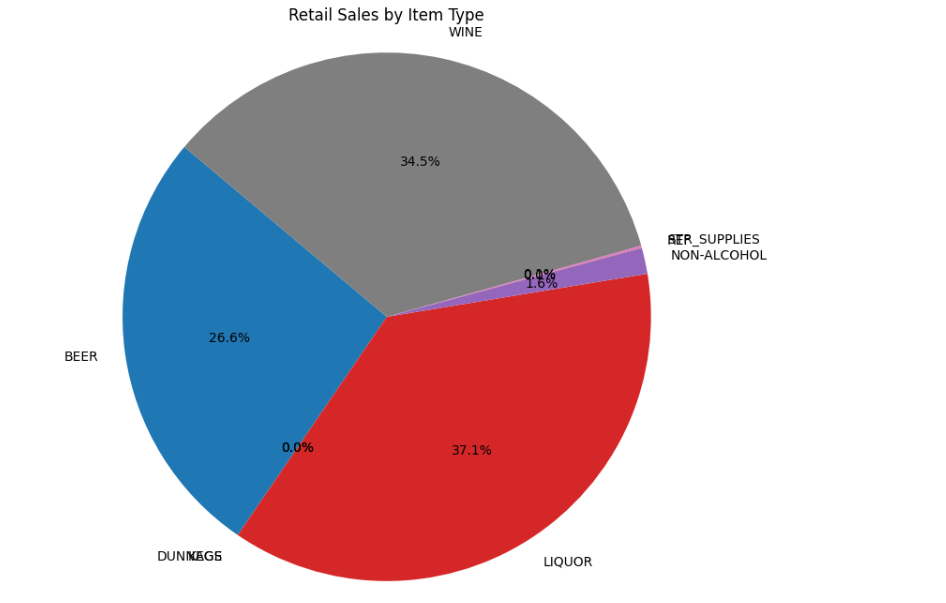
1. **Grouping** the dataset by 'ITEM DESCRIPTION' to consolidate all records for each item.
2. **Summing** the 'WAREHOUSE SALES' for each group to get the total sales per item.
3. **Sorting** the results in descending order to prioritize the highest-selling items.
4. **Selecting** the top 10 items using .head(10).



4.2 Pie Chart: Retail Sales by Item Type (Pie Chart)

**Analysis**:Shows the percentage share of each grievance type in the dataset.

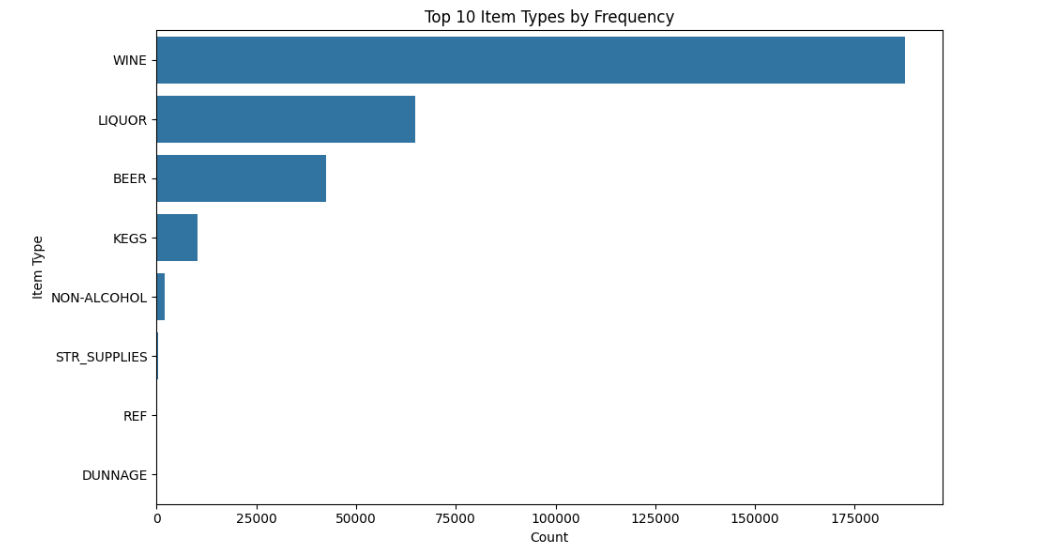
This generates a pie chart showing the proportion of retail sales for each item type. It groups sales by item type, sums the totals, and visualizes the distribution with percentages. The chart helps quickly identify the dominant item categories in retail sales.



4.3 Scatter Plot: Retail vs Warehouse Sales **Analysis:**

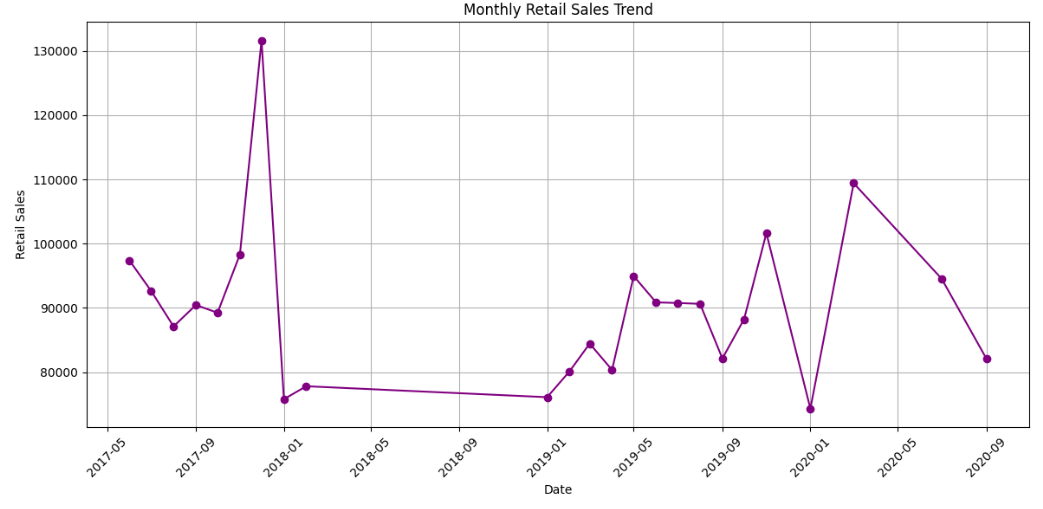
Created a scatter plot to compare retail sales and warehouse sales. The points are colored by the year using the hue='YEAR' argument, providing a visual representation of how sales in both areas relate over different years. The chart helps analyze any correlation between warehouse and retail sales. 

4.4 Count Plot: Top 10 Item Types by Frequency

**Analysis:** Visualizes the frequency of the top 10 most common item types in the dataset. The count plot helps identify which item types are most prevalent, offering insights into product distribution and potential focus areas.

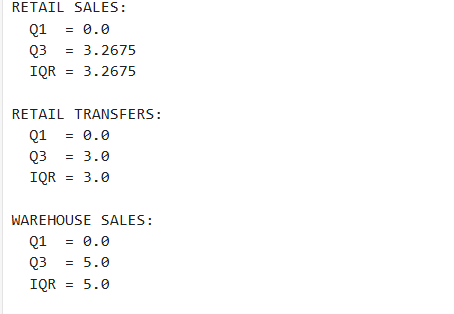
4.5 Line Chart: Monthly Retail Sales Trend

**Analysis:** This plots the **monthly retail sales trend** over time using a line chart. It helps visualize sales patterns, seasonal spikes, or declines, aiding in forecasting and strategic planning.



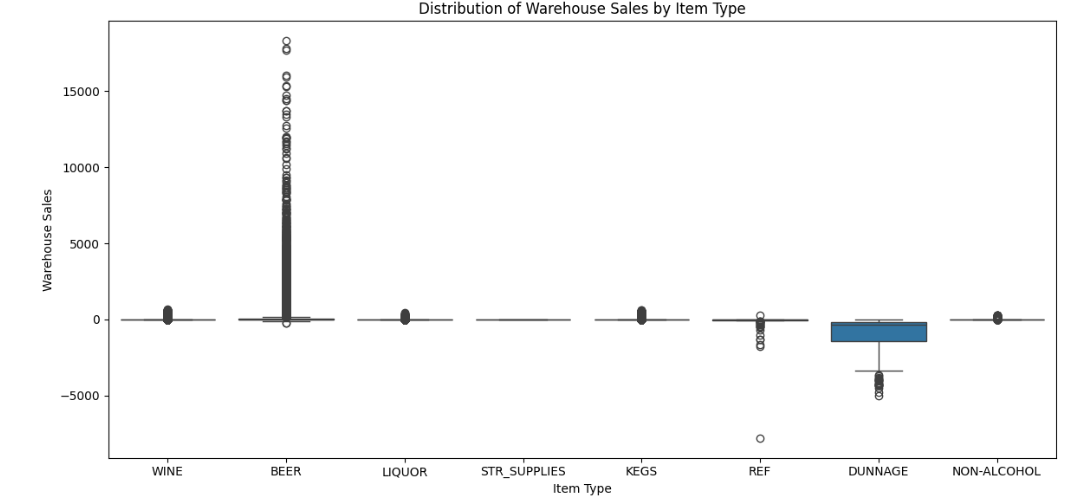
4.6 IQR Calculation for Outlier Detection

**Analysis:** To calculates the **Interquartile Range (IQR)** for key numeric columns to help detect outliers. By finding Q1, Q3, and IQR, it identifies the spread of middle 50% of values, useful for spotting unusually high or low data points.

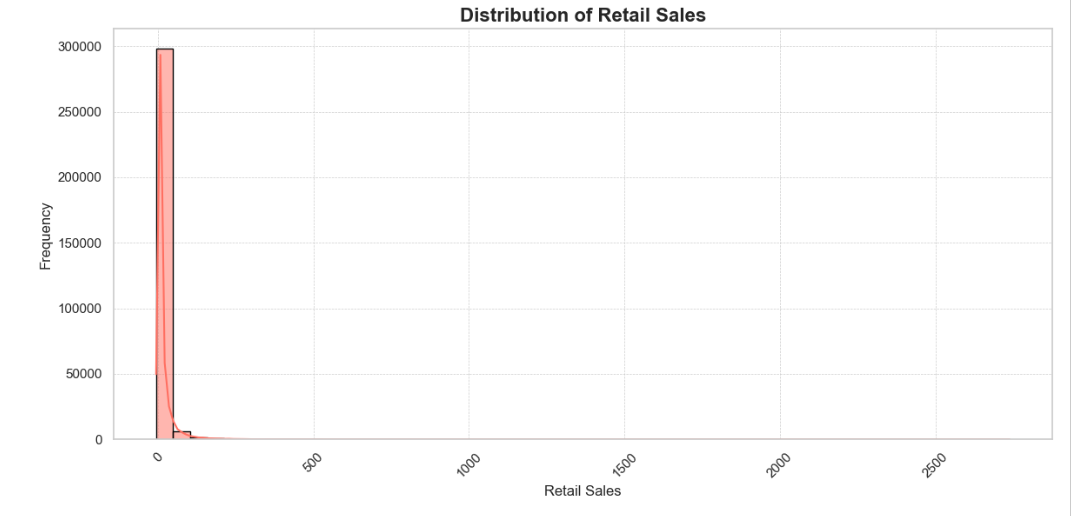


4.7 Boxplot: Warehouse Sales by Item Type

**Analysis:**Created a **boxplot** to show the distribution of **warehouse sales** across different item types. It helps visualize medians, variability, and potential outliers within each item category.

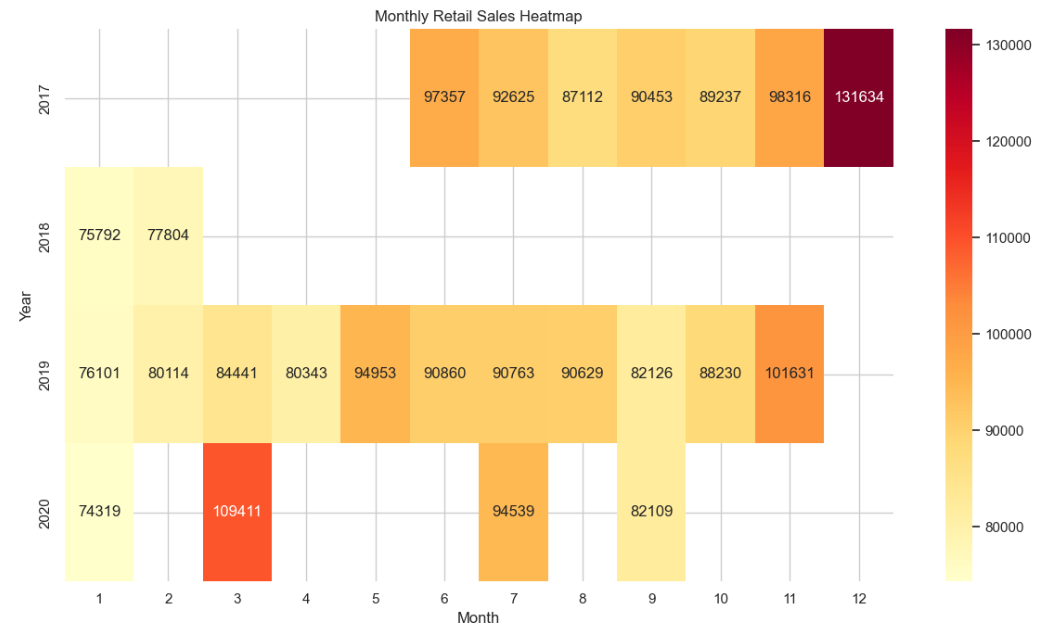


4.8 Histogram of Retail Sales

**Analysis:** Creates a **histogram** showing the distribution of **retail sales** values. It reveals how frequently different sales amounts occur and includes a KDE curve to highlight the overall distribution shape. 

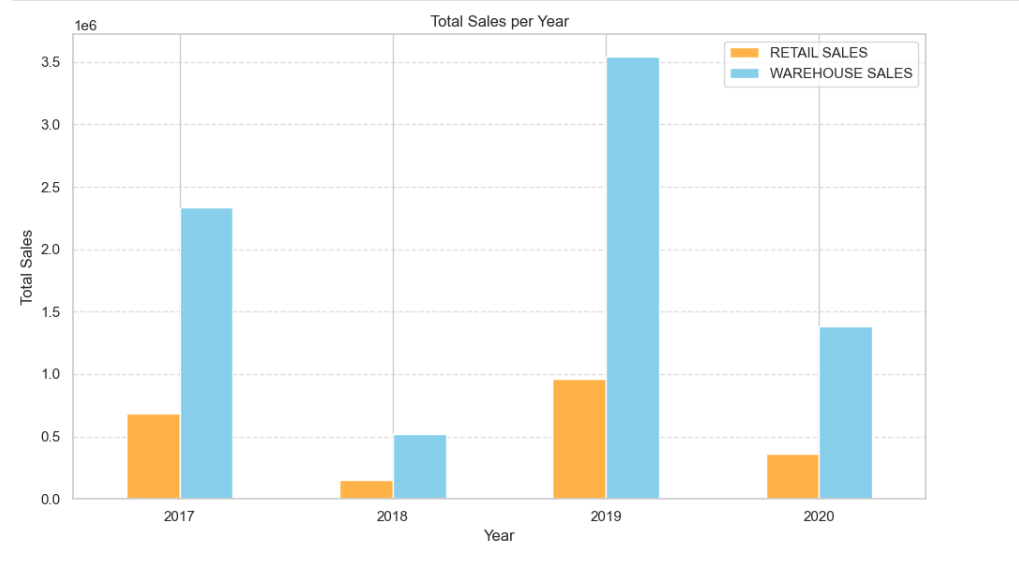
4.9 Heatmap of Monthly Retail Sales

**Analysis:** Generates a **heatmap** of monthly retail sales across years. It visually highlights sales patterns, peaks, and seasonal trends, making it easy to spot high and low-performing periods.

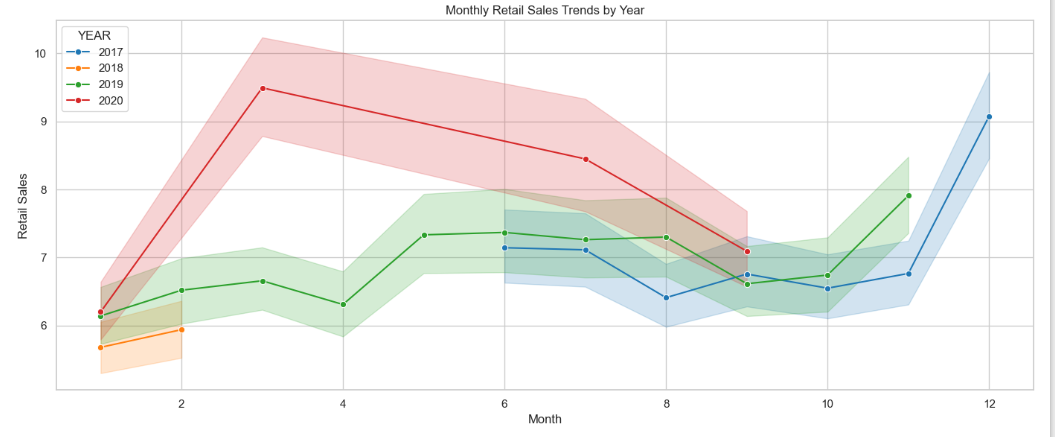


4.10 Total Sales per Year (Bar Chart)

**Analysis: C**reated a **bar chart** showing total retail and warehouse sales for each year. It helps compare annual performance and track overall sales growth or decline over time.

4.11 Monthly Retail Sales Trends by Year

**Analysis:** This code plots **monthly retail sales trends** for each year using a line chart. It highlights seasonal patterns and year-over-year changes, helping identify consistent peaks or drops in monthly sales.



**5. Conclusion**  
The analysis of warehouse and retail sales data reveals key insights into sales performance and trends. Top-selling items and item types significantly drive overall sales, indicating focused consumer demand. Seasonal patterns are evident, with certain months consistently outperforming others in retail sales. There is a moderate correlation between warehouse and retail sales, varying across years. Outlier detection shows variability in sales, suggesting unusual spikes or dips. Overall, the visualizations provide a clear understanding of product performance, sales distribution, and time-based trends, aiding strategic planning and inventory management.

**6. Future Scope**  
This project can be enhanced by integrating predictive analytics using machine learning models to forecast future sales trends. Real-time dashboarding with tools like Python Toolbox can provide dynamic insights. Additionally, incorporating inventory optimization and customer segmentation could further support strategic decision-making.

**7. References**  
- <https://data.gov.in>  
- Python Libraries: pandas, seaborn, matplotlib

**Links:**

**Data set** [**https://catalog.data.gov/dataset/warehouse-and-retail-sales**](https://catalog.data.gov/dataset/warehouse-and-retail-sales)

**Github :** [**https://github.com/ujjwalkrsaw18/Warehouse-and-Retail-Sales-Analysis--Python-Toolbox**](https://github.com/ujjwalkrsaw18/Warehouse-and-Retail-Sales-Analysis--Python-Toolbox)

**Linkedin:** [**https://www.linkedin.com/posts/ujjwalkumarsaw18\_dataanalysis-python-datavisualization-activity-7317874690088128512-KsFc?utm\_source=share&utm\_medium=member\_desktop&rcm=ACoAAEg1PGoBtL5G4BP\_7co7tB2Leb\_WirG-hio**](https://www.linkedin.com/posts/ujjwalkumarsaw18_dataanalysis-python-datavisualization-activity-7317874690088128512-KsFc?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEg1PGoBtL5G4BP_7co7tB2Leb_WirG-hio)