PROJECT REPORT: Sales Data Analysis and Reporting for a Retail Chain

Date: February 20, 2026

Course: Internship Studio Data Analyst Internship Final Project

1. EXECUTIVE SUMMARY

The aim of this project was to leverage Python, SQL, and Excel to analyze retail sales data and generate meaningful reports to support business decision-making. By moving through four distinct phases—from database setup to automated reporting—the project successfully identified key sales trends and performance metrics.

2. DATA SOURCE & DESCRIPTION

The dataset used for this analysis was sourced from Kaggle and includes the following primary fields:

- **TransactionID:** Unique identifier for each transaction.
- **TransactionTime:** Timestamp of the purchase.
- **ItemCode & Description:** Identification and details of the product.
- **NumberOfItemsPurchased:** Quantity of items per transaction.
- **CostPerItem:** The price per unit.
- **Country:** The geographic location of the sale.

3. PHASE 1: DATABASE SETUP

- **Data Collection:** The raw data was downloaded as a CSV file and stored in the local project directory.
- **SQL Setup:** A dedicated SQL database was established to manage the retail data.
- **Schema Design:** Tables were designed and implemented using SQL DDL (Data Definition Language) commands to ensure data structure and integrity.

4. PHASE 2: DATA CLEANING & PREPARATION

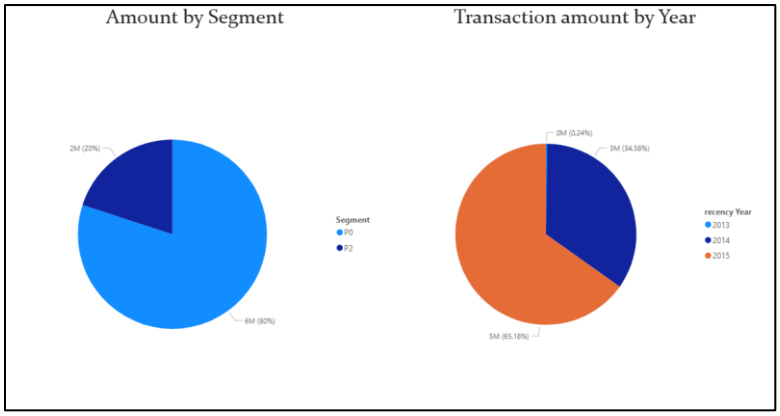
- **Data Cleaning:** Used a combination of SQL queries and Python (Pandas) to handle missing values, resolve inconsistencies, and identify outliers.
- **Data Preparation:** Transformed raw data by creating calculated fields, such as "Total Sales Value".
- **Time-Based Analysis:** Extracted Month and Year fields from the `TransactionTime` column to facilitate temporal trend analysis.

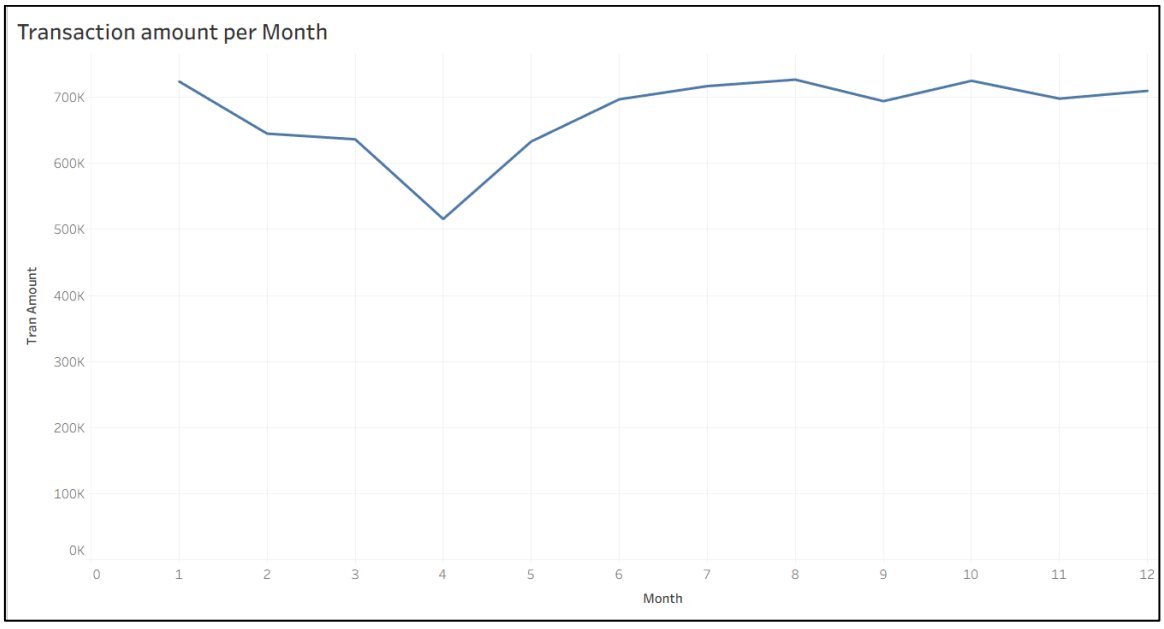
5. PHASE 3: DATA ANALYSIS

- **Exploratory Data Analysis (EDA):** Leveraged Python libraries (Pandas, Matplotlib, Seaborn) and SQL to identify overarching patterns and trends in the retail data.
- **Advanced Analytics:** Conducted complex analysis including:
    - **Time Series Analysis:** To visualize and predict sales trends.
    - **Cohort Analysis:** To evaluate customer behavior and purchasing cycles.

6. PHASE 4: REPORTING & PRESENTATION

- **Tabular Reports:** Created summary tables in Excel showcasing sales by product, store, and month.
- **Visual Reports:** Developed charts and graphs using Python's visualization libraries, which were then integrated into Excel for executive presentation.
- **Automation:** Implemented Python scripts to automate report generation, ensuring dashboards remain updated with minimal manual intervention.
- **Presentation:** Final findings were compiled into an Excel dashboard and a PowerPoint presentation for stakeholders.





7. SUMMARY OF DATA INSIGHTS

The following insights were derived from the data analysis and visualization phase:

A. Monthly Transaction Trends

Analysis of transaction amounts throughout the year reveals significant fluctuations in retail activity:

- **Peak Performance:** The highest transaction volume occurs in **Month 1** (January) and **Month 8** (August), both exceeding $700,000$.
- **Significant Dip:** There is a notable decline in **Month 4** (April), where transaction amounts drop to their lowest point of approximately $500,000$.
- **Recovery Phase:** Post-April, there is a steady recovery trend leading into the mid-year peak.

B. Customer Segmentation

Using the "Amount by Segment" analysis, the customer base was divided into two primary groups:

- **Segment P0:** This is the dominant group, contributing **80%** of the total transaction amount (6M).
- **Segment P2:** This segment accounts for the remaining **20%** (2M).
- **Strategic Impact:** Business efforts should prioritize the retention of Segment P0, as they represent the vast majority of revenue.

C. Annual Revenue Growth

The "Transaction Amount by Year" visualization highlights the retail chain's rapid growth over a three-year period:

- **2013:** Represented the baseline with a minimal share of total historical transactions (0.24%).
- **2014:** Showed significant growth, accounting for **34.58%** (3M) of total volume.
- **2015:** Recorded the highest performance to date, contributing **65.18%** (5M) of total transactions.
- **Conclusion:** The data indicates a strong upward trajectory, with 2015 nearly doubling the performance of the previous year.

8. UPDATED RECOMMENDATIONS

- **Seasonal Planning:** Investigate the cause of the April slump (e.g., end of holiday periods or tax season) and implement promotional "spring sales" to bridge the revenue gap.
- **Segment Prioritization:** Develop a "VIP Loyalty Program" specifically for **Segment P0** to safeguard the 80% revenue stream they provide.
- **Scale for Growth:** Given that 2015 saw a massive 65% share of total historical sales, the retail chain should prepare for increased operational and inventory demands in 2016.