**Retail Sales Data Analysis Project**

**🔹 1. Introduction**

This project analyzes **12,575 retail transactions** from a sample store dataset (source: Kaggle). The dataset includes transaction-level details such as customer ID, category, items, price, quantity, payment method, location, and discounts.

**Tools used:**

* SQL (data cleaning, handling missing values)
* Power BI (dashboard creation, visualization, analysis)

**Objective:**  
To explore customer behavior, sales trends, category performance, and discount effectiveness, and to create an executive-level dashboard for decision-making.

**🔹 2. Data Cleaning Process**

* Removed inconsistencies in column names and standardized formats.
* Replaced or flagged missing values (e.g., blank discount values).
* Verified numeric fields (price, quantity, total spent).
* Ensured unique transaction identifiers.

**🔹 3. KPIs (Key Metrics)**

The following metrics were created to guide analysis:

* **Total Revenue – 1.55 million**
* **Total Transactions – 12.58 thousand**
* **Average Spend per Transaction – 190**
* **% Transactions, Discount – 4219(with), 4157(without), 4219(undocumented)**

**🔹 4. Customer Analysis**

* Top 10 customers contribute ~40% of total revenue → high concentration risk.
* Average spend per customer ≈ **127.01**
* Repeat customers drive most sales; one-time buyers spend less.

A comparison of blue bars

AI-generated content may be incorrect.

5 highest spending customers vs 5 lowest spending customers

**🔹 5. Category & Item Analysis**

* Butchers and Electrical Household Essentials are the highest revenue categories.
* Some singular items have the most revenue (e.g. item\_25\_FUR, item\_25\_EHE)

*A blue rectangular object with text

AI-generated content may be incorrect.*

**🔹 6. Time Trends**

* Sales are steady across the dataset timeframe with noticeable weekend spikes.
* Average basket size remains stable but increases slightly during discount periods.

*(Insert screenshot: Line chart of revenue over time)*

**🔹 7. Location & Payment Insights**

* **Location:** Online revenue came up to 791.40K whereas in-store revenue came to 760.67K
* **Payment:** Cash remains the most used for transactions by a small margin while digital payments and credit cards remain at similar usage

A graph showing a number of lines

AI-generated content may be incorrect.

**🔹 8. Discount Analysis**

* Non-discounted sales generated $1.03M in revenue (≈50% of total).
* Discounted sales generated $524K in revenue (≈25% of total).
* An additional $512K (≈25% of revenue) has missing discount data.
* Discounts contribute meaningfully but are not the primary driver of sales.
* Data quality issues (missing discount values) limit clear conclusions on discount effectiveness.

*A blue bar graph with black text

AI-generated content may be incorrect.*

**🔹 9. Executive Summary Dashboard**

Final dashboard consolidates:

* KPI cards for total revenue, transactions, average spend, discount %.
* Revenue by category.
* Revenue over time.
* Top 5 customers.

A close-up of a number

AI-generated content may be incorrect.A screenshot of a phone

AI-generated content may be incorrect.

**🔹 10. Conclusion**

This project demonstrates the ability to:

* Clean and prepare raw retail data in SQL.
* Build interactive dashboards in Power BI.
* Extract actionable insights for business strategy (customer loyalty, discount effectiveness, payment preferences).

The analysis highlights how retailers can monitor performance and optimize discounts, marketing, and customer retention strategies.