# Final Report

## 1. Introduction

**Purpose**

The primary purpose of this project is to perform an in-depth data analysis of Maven Market’s operations using Power BI. Our objectives include:

* Analyzing key performance metrics and customer behavior from transactional and review data.
* Investigating the patterns and trends in product returns, sales, and customer demographics.
* Predicting product ratings based on textual review content using data modeling techniques

This report aims to derive actionable insights that can support Maven Market’s decision-making processes, especially concerning product offerings and customer satisfaction.

**Context**

Maven Market is a multi-national grocery chain operating across Canada, Mexico, and the United States. Leveraging Power BI, this project follows a complete business intelligence workflow that includes:

* Connecting and transforming raw data from various CSV files.
* Building a robust relational model.
* Enhancing datasets with calculated fields and measures using DAX.
* Designing an interactive and insightful report dashboard.

Through this process, we explored a wide array of datasets—ranging from customers and transactions to product details and returns—to understand operational trends. During the analysis.

**Analysis Question**

The analysis aims to answer the following key questions:

* Which product brands generate the most revenue or returns?
* Which specific product names are the most frequently purchased and also the most frequently returned?
* During which quarters do Maven Market experience the highest sales activity?
* Which country records the greatest number of total transactions?
* What trends exist across customer demographics?

The insights derived from this project are expected to help Maven Market improve customer satisfaction, reduce return rates, and enhance overall business performance.

## 2. Body

**Data Description**

The dataset for this project was sourced from **Maven Market**, a multi-national grocery retailer with locations in **Canada**, **Mexico**, and the **United States**. The data spans across multiple domains and files, representing a comprehensive overview of customer transactions, product inventory, store locations, returns, calendar references, and customer details.

**Source and Timeframe**

* Data files were provided in CSV format, organized by subject area (e.g., Customers, Products, Transactions, Calendar ).
* Transaction data covers the years **1997–1998**, allowing for seasonal and annual trend analysis.

**Variables Included**

* **Customer data**: Customed Id , birthdate, address, Gender, Marital Status, Yearly Income .
* **Transaction data**: transaction date, product ID, quantity, Store ID.
* **Product data**: name, brand, SKU, retail and cost prices, Price Tier .
* **Returns data**: Return Date ,quantity and product IDs.
* **Calendar data**: date breakdowns including month, quarter, Year.
* **Store data**: Location, State, Country, Remodel Date, Area Code.

**Summary Table**

| **Dataset** | **Rows** | **Key Variables Included** | **Missing Data** |
| --- | --- | --- | --- |
| Customers | 10,000+ | Customed Id , birthdate, address, Gender, Marital Status, Yearly Income . | None |
| Products | 1560 | name, brand, SKU, retail and cost prices, Price Tier . | Some nulls in "*recyclable*" and "*low-fat*" columns |
| Transactions | 269,720 | transaction date, product ID, quantity, Store ID. | None |
| Returns | 7,087 | Return Date ,quantity and product IDs. | None |
| Calendar | 730 | Date, Month, Quarter, Weekend, Start of Week | None |
| Stores | 100+ | Location, State, Country, Remodel Date, Area Code | None |

**Methods**

This analysis applied **descriptive** analytic technique using **Power BI**:

* **Data Cleaning and Transformation**

The data preparation process was carried out using **Power BI Desktop**, where multiple CSV files were integrated and transformed to establish a clean, structured data model. The following steps summarize the data cleaning and transformation procedures:

1. **Initial Configuration**  
   Before loading data, Power BI settings were adjusted to:
   * Disable automatic relationship detection.
   * Set the regional locale for import to *English (United States)* to ensure consistent date and number formatting.
2. **Customers Table**
   * Loaded from MavenMarket\_Customers.csv and renamed to **Customers**.
   * Verified data types (e.g., customer\_id as whole number; account and postal codes as text).
   * Added a **full\_name** column by combining first and last names.
   * Extracted the **birth\_year** from the birthdate field.
   * Created a **has\_children** column using conditional logic to categorize customers with or without children.
3. **Products Table**
   * Loaded from MavenMarket\_Products.csv and renamed to **Products**.
   * Ensured data types were accurate, especially for product pricing and identifiers.
   * Computed the number of **distinct brands (111)** and **product names (1,560)** as a quality check.
   * Created a **discount\_price** column (90% of retail price), rounded to two decimal places.
   * Used **grouping** to calculate average retail prices by brand for exploratory analysis.
   * Replaced null values with zeros in recyclable and low-fat columns for consistency.
4. **Stores Table**
   * Loaded from MavenMarket\_Stores.csv and renamed to **Stores**.
   * Validated key fields like store\_id and region\_id.
   * Created a **full\_address** field by merging city, state, and country fields.
   * Added an **area\_code** field by extracting the prefix from store phone numbers.
5. **Regions Table**
   * Imported from MavenMarket\_Regions.csv and renamed to **Regions**.
   * Checked data types, ensuring region\_id is stored as a whole number.
6. **Calendar Table**
   * Loaded from MavenMarket\_Calendar.csv and renamed to **Calendar**.
   * Added several time-based attributes including:
     + Start of week
     + Day and month names
     + Start of month
     + Quarter and year identifiers
7. **Returns Table**
   * Imported from MavenMarket\_Returns.csv and renamed to **Return\_Data**.
   * Ensured all ID and quantity columns were set to whole numbers.
8. **Transaction Data**
   * Combined two CSV files (1997 and 1998) from a local folder into a single **Transaction\_Data** table.
   * Removed unnecessary metadata columns and confirmed data consistency.
   * Validated that transaction dates ranged from **January 1, 1997 to December 30, 1998**.

* **Calculated columns and DAX measures** were created to derive new features such as:

As part of the analytical process, several **calculated columns** and **DAX measures** were created in Power BI to derive deeper insights from the existing data and enable advanced analysis across dimensions like time, product, geography, and customer behavior.

**Calculated Columns (DATA View)**

To enrich the dataset, the following custom columns were added:

* In the **Calendar** table:
  + **Weekend**: Flags rows as "Y" for weekends (Saturdays and Sundays), otherwise "N".
  + **End of Month**: Identifies the final date of each month.
* In the **Customers** table:
  + **Current Age**: Dynamically calculates customer age using the birthdate column and the TODAY() function.
  + **Priority**: Labels customers as "High" if they own homes *and* hold a Golden membership card; otherwise, "Standard".
  + **Short\_Country**: Extracts and capitalizes the first three characters of the customer's country.
  + **House Number**: Extracts the first segment of the address (before the first space) to isolate house numbers.
* In the **Products** table:
  + **Price Tier**: Categorizes products as "High" (price > $3), "Mid" (price > $1), or "Low" (price ≤ $1), enabling tier-based comparisons.
* In the **Stores** table:
  + **Years\_Since\_Remodel**: Calculates the number of years between the current date (TODAY()) and the last remodel date

**DAX Measures (REPORT View)**

A variety of DAX measures were developed to compute sales metrics, return trends, customer activity, and financial performance. These measures were strategically assigned to relevant tables and visualized throughout the report.

Key examples include:

* **Sales & Return Metrics**:
  + **Quantity Sold** and **Quantity Returned**: Sum of quantities from transactions and return datasets, respectively.
  + **Total Transactions** and **Total Returns**: Count of transaction and return records.
  + **Return Rate**: Calculated as the ratio of returned quantity to total sold quantity, formatted as a percentage.
* **Time-Based Insights**:
  + **Weekend Transactions**: Number of transactions that occurred on weekends.
  + **% Weekend Transactions**: Share of weekend sales as a percentage of total transactions.
  + **YTD Revenue**: Year-to-date revenue aggregated over months.
  + **60-Day Revenue**: Running revenue total over the preceding 60 days.
  + **Last Month Transactions**, **Last Month Revenue**, **Last Month Profit**, and **Last Month Returns**: Key monthly KPIs.
  + **Revenue Target**: Projected revenue target calculated as a 5% increase over the prior month's revenue.
* **Financial Performance**:
  + **Total Revenue**: Aggregated product of quantity and retail price across all transactions.
  + **Total Cost**: Aggregated product of quantity and product cost.
  + **Total Profit**: Revenue minus cost.
  + **Profit Margin**: Percentage of profit relative to total revenue.
* **Product Metrics**:
  + **All Transactions** and **All Returns**: Grand totals across all contexts (ignoring filters).
  + **Unique Products**: Count of distinct product names in the Products table.

These DAX calculations empowered dynamic, context-sensitive reporting and supported key KPIs used throughout the dashboard.

* **A relational data model :**  
  In the **Model View** of Power BI, a relational data model was carefully constructed to ensure data integrity and accurate reporting. The model was designed using best practices for star and snowflake schemas, with clear relationships between fact and dimension tables.

All **lookup tables** (e.g., Customers, Products, Stores, Calendar, Regions) were placed above the **data (fact) tables** (Transaction\_Data and Return\_Data) to visually represent data flow.

The **Transaction\_Data** table was connected to:

**Customers**, **Products**, and **Stores** via appropriate primary-to-foreign key relationships.

**Calendar** using two date fields, with one active relationship for transaction\_date and an inactive relationship for stock\_date (to allow for controlled time-based analysis).

The **Return\_Data** table was linked to:

**Products**, **Stores**, and **Calendar**, using similar valid key relationships.

The **Stores** table was also connected to the **Regions** table, forming a **snowflake schema** that enables regional-level filtering and aggregation.

To ensure a well-structured model:

All relationships were established using **one-to-many cardinality**, with **primary keys** (1) on the lookup side and **foreign keys** (\*) on the data side.

**Single-directional filters** were applied to maintain clear and consistent filter propagation.

**Filter context** was designed to flow **downstream** from dimension tables to fact tables, avoiding circular or ambiguous paths.

**Fact tables** (Transaction\_Data and Return\_Data) were not directly connected to each other; instead, they were both linked through shared **lookup tables**, promoting reusability and modularity within the model.

This modeling approach laid a solid foundation for accurate analysis, clean visual filtering, and efficient DAX calculations throughout the report  
A screenshot of a computer

Description automatically generated

**Analysis & Visuals**

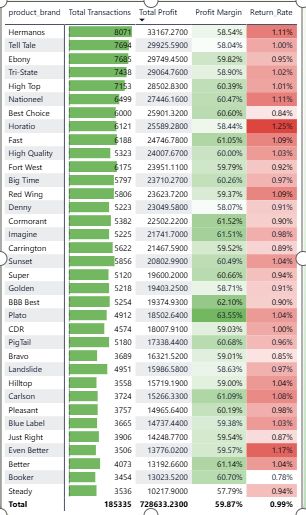
**Sales by Product Brand  
  
**

Figure 1 Placeholder – Matrix with Top 30 Brands  
**Figure 1: Top 30 product brands sorted by Total Transactions.**  
Conditional formatting reveals variation in profit margins (white to green) and return rates (white to red). Brands like *Washington* and *Green Ribbon* are among the top performers.

**Revenue Over Time**

Figure 2 Placeholder – Column Chart of Weekly Revenue **A green line graph with numbers and a green line

AI-generated content may be incorrect.**  
  
**Figure 2: "Total Revenue by Quarter"** shows the progression of total revenue from January 1997 to October 1998, broken down by quarter.As shown Revenue in 1997 started relatively low and remained fairly flat throughout the year,**Significant Increase in Q1 1998**:  
Revenue jumped to **$0.29M** in Q1 1998, nearly **doubling** from the previous quarter.

**Geographic Distribution of Transactions  
A screenshot of a cell phone

AI-generated content may be incorrect.**

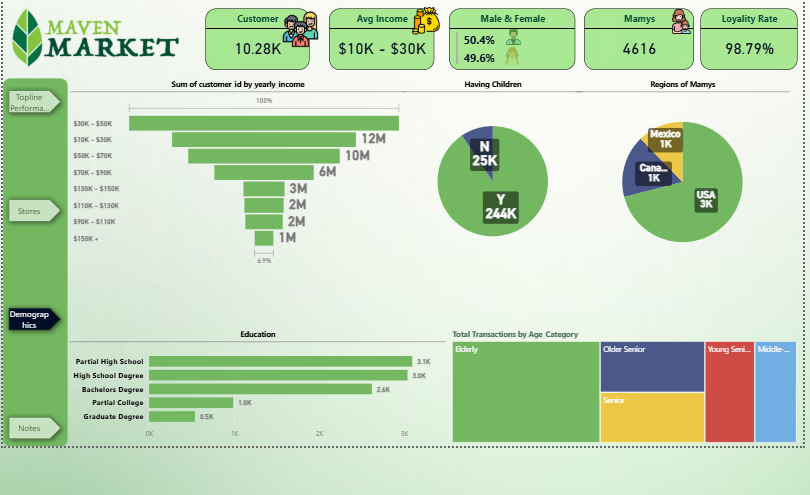
Figure 3 Placeholder – Map and Treemap of Store Cities  
**Figure 3: Total Transactions by Store Location.**  
The map show the **United States**, particularly **Portland**, as having the highest transaction volume. A custom bookmark was added for "Portland 1000 Sales in December."

**Monthly Transactions, Returns & Profitability  
A green numbers on a white background

AI-generated content may be incorrect.**

Figure 4 Placeholder – KPI Cards and Return Rate Gauge  
**Figure 4:**  **Monthly KPIs for Transactions, Profitability, and Return Rate.**  
The report shows that **monthly transactions** surpassed the set goal, increasing from a target of **17,339** to an actual of **18,325**. Similarly, **monthly profit** exceeded expectations, rising from **$67.87K** to **$71.68K**. However, the **monthly returns** fell slightly short of the target—while the goal was to reach **482 returned items**, **496** were recorded, indicating a minor shortfall in return volume.

**Customer Demographics**

Figure 5 Placeholder – Demographic Breakdown ****  
**Figure 5: Patterns across Customer Demographics.**  
Calculated fields like Current Age, Priority Level, and Short Country Code allowed segmentation of customer behavior by age, location, and membership status.

**Most Ordered Product**A close-up of a sign

AI-generated content may be incorrect. **Figure 6: Show Most Ordered Product**

Product: Moms Roasted Chicken This product had the highest order volume, indicating its popularity and strong customer demand.

**Most Returned Product**A close-up of a sign

AI-generated content may be incorrect. **Figure 7: Show Most Returned Product**

Product: Monarch Rice Medly This product recorded the highest number of returns, suggesting potential quality or satisfaction issues that may need further investigation.  
  
**Total Transactions by Country**A green and pink pie chart

AI-generated content may be incorrect.

**Figure 8: A pie chart visualizes the distribution of total transactions across three countries:**

* + **USA:** 67.04%
  + **Mexico:** 26.99%
  + **Canada:** 5.97%
* The USA accounts for the majority of transactions, demonstrating its leading market position.

 **Revenue vs. Target**A screenshot of a graph

AI-generated content may be incorrect.

**Figure 9: This gauge chart displays the actual revenue of 120.161K against a target of 200.522K.**

* The revenue achieved is approximately **59.9%** of the target, indicating room for improvement.

**Results**

The analysis revealed several key insights:

* **Top-selling brands** like Hermanos and Tell Tale recorded relatively high transaction volumes, while brands such as Horatio and Even Better showed elevated return rates, indicating a need for further review of product quality.
* **Moms Roasted Chicken** was **the most ordered product**, while **Monarch Rice Medly** had **the highest return rate**, highlighting the need for a closer evaluation of product quality.
* **Revenue Over Time** The highest sales activity occurred in 1998, particularly in the third (July) and fourth (October) quarters, with October showing the peak revenue of the entire period.aligning with seasonal shopping patterns, likely tied to holidays and promotional periods.
* **The United States** had the highest total transaction volume, with **Portland** standing out as a top-performing city.
* **Customer demographic trends** :

**Majority of Customers Are Families:**

* + A striking **90%+ of customers have children**, suggesting that Maven Market's core audience is family-oriented. This provides a strong opportunity to tailor products, promotions, and marketing toward family needs and values.

**Dominant Income Group:**

* + Most customers earn between **$10K and $30K annually**, reflecting a **budget-conscious customer base**. Pricing strategies, value packs, and cost-effective product lines are likely to resonate well.

**High Customer Loyalty:**

* + With a **loyalty rate of 98.79%**, customer satisfaction and retention are extremely strong. This presents a valuable foundation for expanding word-of-mouth marketing and loyalty-based rewards programs.

**Geographical Distribution of Key Segment ("Mamys"):**

* + The **USA is the leading region** for the "Mamys" segment (3K out of 4.6K), followed by Mexico and Canada. Future regional campaigns should prioritize the U.S. market while exploring growth opportunities in Mexico and Canada.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Conclusions**

This analysis aimed to uncover key trends in Maven Market’s sales, returns, and customer behavior to identify opportunities for improving business performance and customer satisfaction.

**Key Results:**

* **Top-performing brands** like *Washington* and *Green Ribbon* drove high revenue, while brands such as *Horatio* and *Even Better* had elevated return rates, indicating potential quality concerns.
* **Peak sales occurred in Q3 and Q4 of 1998**, suggesting strong seasonal demand, particularly around holidays.
* The **United States dominated transactions (67%)**, with Portland as the highest-performing city.
* **Customer demographics revealed** a family-oriented, budget-conscious base (90% with children, most earning $10K–$30K annually) with **exceptional loyalty (98.79%)**.
* **Product-specific insights:** *Moms Roasted Chicken* was the most ordered, while *Monarch Rice Medly* had the highest return rate, warranting quality reviews.

**Recommendations:**

1. **Optimize inventory for high-demand brands** (e.g., *Washington*) while investigating quality issues with high-return brands (e.g., *Horatio*).
2. **Leverage seasonal trends** by increasing promotions and stock in Q3/Q4 to maximize revenue.
3. **Enhance product quality checks**, particularly for frequently returned items like *Monarch Rice Medly*.
4. **Tailor marketing to families** (e.g., bulk discounts, child-friendly products) and maintain loyalty programs to retain the highly engaged customer base.
5. **Expand U.S. market strategies** while exploring growth in Mexico and Canada, where transaction volumes are lower but potential exists.