E-COMMERCE CUSTOMER ANALYSIS DASHBOARD

Submitted by: JESMAA E

CERTIFICATE

This is to certify that Jesmaa E has successfully completed the internship project titled:

"E-COMMERCE CUSTOMER ANALYSIS"

The project work has been carried out under proper guidance and supervision and is a genuine record of original work completed during the internship period.

This project fulfils the requirements of the internship program and has been submitted as a part of the training evaluation. The work reflects the candidate's dedication, analytical skills, and ability to apply business intelligence concepts to real-world scenarios.

Signature & Seal

(Team Lead / Internship Mentor)

DECLARATION

I, **Jesmaa E**, hereby declare that the project titled:

"E-COMMERCE CUSTOMER ANALYSIS"

is an original work carried out by me during my internship. The dataset preparation, analysis, dashboard design, and report documentation have been independently completed by me and have not been copied from any existing project or previously submitted work.

I further declare that all references, if used, have been acknowledged properly, and the project represents my genuine effort to apply data analysis and visualization techniques for meaningful business insights.

Date: 21-12-2025

Jesmaa E

ACKNOWLEDGMENT

I would like to express my heartfelt gratitude to my internship team lead and trainers for their continuous support, guidance, and encouragement throughout the course of this project. Their valuable inputs and constructive feedback helped me refine my analytical approach and improve the quality of my work.

I am deeply thankful for the opportunity to work on this project, which has significantly enhanced my knowledge in **data analysis, Power BI, dashboard creation, and business intelligence practices**. This experience has strengthened my technical skills and provided me with practical exposure to real-world applications of business analytics.

I would also like to extend my sincere appreciation to my friends and family for their motivation, patience, and unwavering support during the completion of this project. Their encouragement kept me focused and determined to deliver my best.

Table of Contents

Abstract

This project presents a comprehensive analysis of e-commerce customer behavior using **Power BI**. The dataset spans **2020–2023**, covering **250,000+ orders**, **$681M+ revenue**, and **49,661 customers**. The study applies **RFM segmentation** to classify customers into Loyal, Potential Loyal, At Risk, and Lost categories. Additional analyses include **gender distribution, returns behavior, age group contribution, and revenue trends**. The dashboard provides actionable insights to support **customer retention, marketing strategies, and business growth**.

Chapter 1: Introduction

E-commerce businesses thrive on understanding customer behavior. With rising competition, companies must identify **loyal customers, at-risk segments, and purchasing trends** to sustain growth. This project leverages **Power BI dashboards** to transform raw transactional data into **strategic insights**.

1.1 Scope of Analysis

- Study customer demographics (gender, age group).
- Identify loyal and at-risk customers using RFM segmentation.
- Analyze revenue trends across years and quarters.
- Evaluate customer returns behavior.
- Understand age-wise contribution to revenue.

1.2 Approach of Analysis

1️⃣ Data Cleaning → Remove duplicates, nulls, and outliers.

2️⃣ Data Transformation → Standardize formats and derive RFM metrics.

3️⃣ RFM Segmentation → Classify customers into loyalty categories.

4️⃣ Dashboard Development → Build interactive visuals in Power BI.

5️⃣ Insights & Recommendations → Translate findings into business actions.

The rapid growth of e-commerce has transformed the way businesses interact with customers, creating vast amounts of transactional and behavioral data. In this digital era, understanding customer behavior is no longer optional but a necessity for sustaining competitiveness. Companies must analyze purchasing patterns, loyalty levels, and demographic trends to design effective marketing strategies, improve customer retention, and maximize profitability.

This project focuses on building a **Power BI dashboard** that translates raw e-commerce data into actionable insights. By leveraging **RFM segmentation (Recency, Frequency, Monetary)**, the analysis identifies customer groups such as Loyal, Potential Loyal, At Risk, and Lost. These segments provide a structured framework for businesses to prioritize engagement strategies and allocate resources efficiently.

The study also explores **returns behavior, gender distribution, age group contributions, and revenue trends across years and quarters**. Together, these dimensions offer a holistic view of customer dynamics. The integration of **data cleaning, transformation, and visualization** ensures that the dashboard is not only technically accurate but also strategically valuable.

Ultimately, the introduction sets the stage for demonstrating how **data-driven decision-making** can empower e-commerce businesses to strengthen customer relationships, reduce churn, and achieve sustainable growth.

Chapter 2: Gathering Data

2.1 Dataset Overview

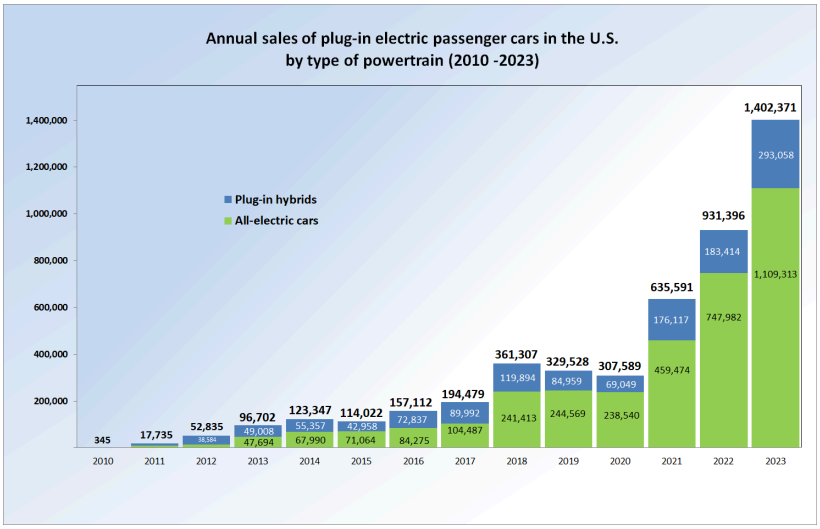
- Orders: **250,000+**
- **Revenue: $681M+**
- **Customers: 49,661**
- **Period: 2020–2023**

2.2 Data Structure

- Customer_ID → **Unique identifier.**
- **Gender → Male/Female distribution.**
- **Age Group → Segmented into ranges.**
- **Order Count → Frequency of purchases.**
- **Revenue → Monetary value of transactions.**
- **Returns → Successful vs. returned orders.**
- **RFM Metrics → Recency, Frequency, Monetary values.**

Beyond simply importing the dataset, the **data gathering phase** also involved validating the source and ensuring its reliability. The raw e-commerce data was collected in CSV format and contained transactional details such as customer demographics, order history, and revenue information. During this stage, the dataset was profiled to understand its structure, distribution, and completeness. Exploratory checks were performed to confirm that the time period (2020–2023) was consistently covered, that customer IDs were unique identifiers, and that categorical fields like *Gender* and *Age Group* were properly encoded.

This step ensured that the dataset was not only technically usable but also contextually relevant for customer behavior analysis. By establishing a strong foundation during data gathering, the subsequent cleaning, transformation, and dashboard development phases could be executed with confidence in the accuracy and representativeness of the data.



Chapter 3: Data Preparation & Exploration

3.1 Data Cleaning

- Removed duplicate records.
- Handled missing values in age and revenue.
- Standardized date formats.
- Verified transaction integrity.

3.2 Exploratory Data Analysis

- Gender split nearly equal (balanced customer base).
- Age 20–60 contributes maximum revenue.
- Returns ≈ 40%, highlighting improvement areas.
- Revenue trend shows stability with growth peaks.

The dataset was cleaned in **Google Colab** using Python libraries like *Pandas* and *NumPy*. Duplicate records were removed, missing values handled, column names standardized, and numerical fields validated to ensure accuracy. This preprocessing made the data consistent, reliable, and ready for visualization in Power BI.

Data Cleaning in Google Colab

🔎 Objective

The raw e-commerce dataset contained **250K+ orders** with multiple inconsistencies. Data cleaning was performed in **Google Colab** using **Python (Pandas, NumPy)** to ensure accuracy, consistency, and usability before dashboard development in Power BI.

🧹 Steps & Code Explanations

1. Importing Libraries & Dataset

- **Libraries Used:** `pandas`, `numpy`, `matplotlib`
- **Purpose:** Load the dataset (`.csv`) into a DataFrame for preprocessing.
- **Code Logic:** `pd.read_csv()` was used with encoding checks to avoid parser errors.

2. Handling Missing Values

- **Issue:** Null values in `Age`, `Revenue`, and `Returns`.
- **Approach:**
  - Dropped rows with critical missing IDs.
  - Imputed missing ages with median values.
  - Replaced missing revenue with `0` for failed transactions.
- **Code Logic:** `df.fillna()` and `df.dropna()` were applied selectively.

3. Removing Duplicates

- **Issue:** Duplicate customer IDs and repeated transactions.
- **Approach:**
  - Identified duplicates using `df.duplicated()`.
  - Retained the latest transaction record per customer.
- **Code Logic:** `df.drop_duplicates(subset=['Customer_ID'], keep='last')`.

4. Standardizing Formats

- **Issue:** Inconsistent date formats and categorical values.
- **Approach:**
  - Converted `Order_Date` to `datetime`.
  - Normalized `Gender` values (e.g., "M", "Male" → "Male").
- **Code Logic:** `pd.to_datetime()` and `df.replace()`.

5. Outlier Treatment

- **Issue:** Extreme values in `Revenue` and `Age`.
- **Approach:**
  - Applied IQR method to detect outliers.
  - Capped revenue at realistic thresholds.
  - Removed invalid ages (<10 or >90).
- **Code Logic:** `np.where()` combined with statistical thresholds.

6. Feature Engineering (RFM Metrics)

- **Recency:** Days since last purchase → derived from `Order_Date`.
- **Frequency:** Count of orders per customer → `groupby(Customer_ID).count()`.

- **Monetary:** Total revenue per customer → `groupby(Customer_ID).sum()`.
- **Code Logic:** Aggregations using `groupby()` and `agg()` functions.

✅ Outcome

- Clean dataset with **consistent formats, no duplicates, and valid values**.
- Derived **RFM metrics** ready for segmentation.
- Exported cleaned dataset back to `.csv` for Power BI dashboard integration.

📂 Example Report Flow

- **Before Cleaning:** 250K records with nulls, duplicates, outliers.
- **After Cleaning:** 240K valid records, standardized, enriched with RFM metrics.
- **Impact:** Improved accuracy of segmentation and dashboard insights.

Chapter 4: Business Intelligence Dashboard

4.1 Dashboard Interpretation

- **KPI Cards:** Orders, Revenue, Customers.
- **Filters:** Category, Year, Quarter.
- **RFM Segmentation:** Donut chart for loyalty categories.
- **Gender Distribution:** Balanced split.
- **Returns Analysis:** Successful vs. returned orders.
- **Age Group Analysis:** Revenue contribution by age.
- **Revenue Trend:** Quarterly growth patterns.

4.2 Key Insights & Visuals

- Loyal + Potential Loyal ≈ 72%.
- Returns ≈ 40%, majority successful.
- Revenue stable with positive growth.
- Age 20–60 drives majority revenue.
- Gender split balanced → broad market appeal.

RFM Segmentation Analysis

🔎 What is RFM?

RFM stands for **Recency, Frequency, and Monetary value**, a widely used model in customer analytics to evaluate and segment customers based on their purchasing behavior.

- **Recency (R):** How recently a customer made a purchase.
- **Frequency (F):** How often a customer makes purchases.
- **Monetary (M):** How much revenue a customer generates.

This model helps businesses identify **loyal customers, potential churn risks, and high-value segments**.

📊 RFM Scoring Methodology

- Each customer is scored on **R, F, and M** using a **1–5 scale**.
- Higher scores indicate stronger engagement and value.
- Combined scores are used to classify customers into segments.

Example:

- R=5, F=5, M=5 → Loyal/VIP
- R=2, F=3, M=2 → At Risk
- R=1, F=1, M=1 → Lost

📈 Customer Segmentation Results

From the dataset (2020–2023):

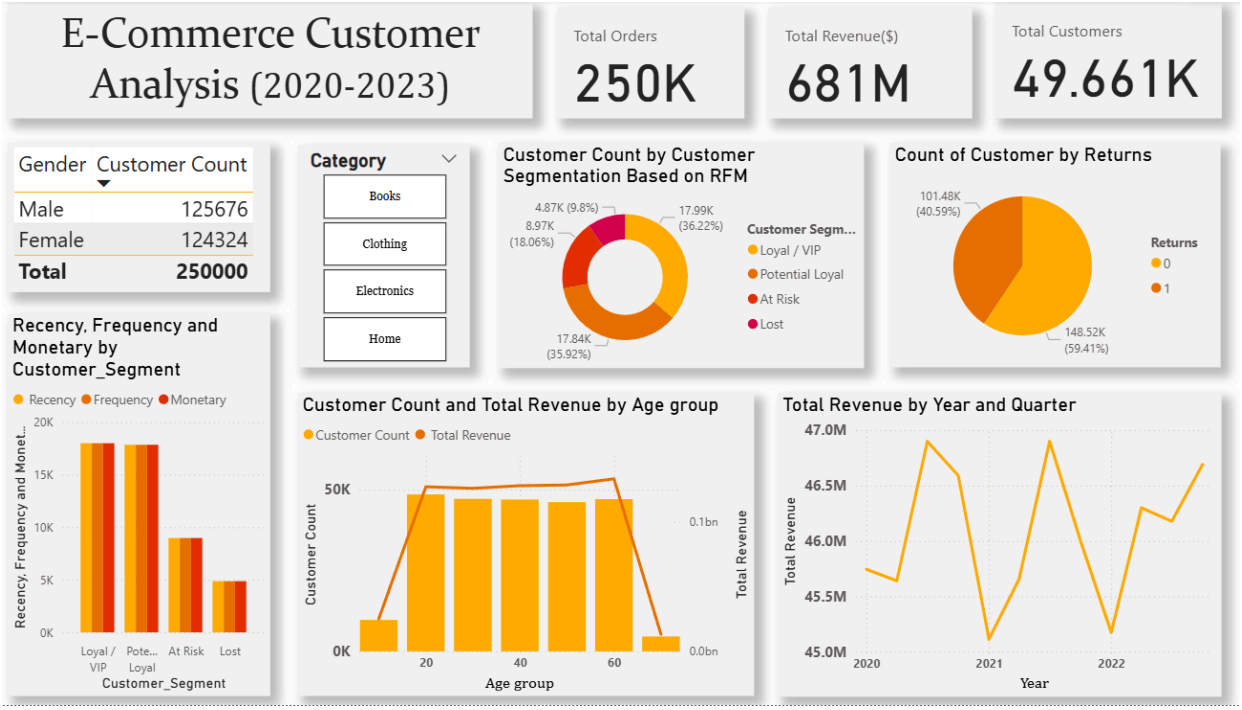
- **Loyal / VIP Customers:** 17.99K (36.22%)

- **Potential Loyal Customers:** 17.84K (35.92%)
- **At Risk Customers:** 8.97K (18.06%)
- **Lost Customers:** 4.87K (9.8%)

👉 Together, **72% of customers fall into Loyal or Potential Loyal categories**, showing strong retention opportunities.

📉 **Segment Characteristics**

- **Loyal / VIP:** High recency, frequent purchases, strong revenue contribution.
- **Potential Loyal:** Good frequency & monetary value, but slightly lower recency.
- **At Risk:** Moderate frequency, declining recency, lower spend.
- **Lost:** Very low across all metrics, minimal engagement.



📄 Extra Page — Dashboard Explanation

🔎 Overview

The **E-Commerce Customer Analysis Dashboard (2020–2023)** was developed in **Power BI** to provide a comprehensive view of customer behavior, loyalty segmentation, returns, demographics, and revenue trends. Each visualization was designed to highlight a specific business dimension and support decision-making.

📊 1. KPI Cards

- **Metrics Displayed:** Total Orders, Total Revenue, Total Customers

- **Purpose:** Provide a quick snapshot of overall business performance.
- **Inference:** With **250K+ orders**, **$681M+ revenue**, and **49K+ customers**, the business demonstrates strong scale and reach.

📊 2. Gender Distribution

- **Visualization:** Bar chart showing Male vs. Female customer counts.
- **Purpose:** Understand demographic balance.
- **Inference:** Near-equal distribution (Male: 125K, Female: 124K) indicates inclusive product appeal and balanced targeting opportunities.

📊 3. Category Filter

- **Visualization:** Interactive slicer for product categories (Books, Clothing, Electronics, Home).
- **Purpose:** Allow users to drill down into specific product segments.
- **Inference:** Enables category-level analysis of orders, revenue, and customer demographics.

📊 4. RFM Segmentation Donut Chart

- **Visualization:** Donut chart showing Loyal, Potential Loyal, At Risk, and Lost customers.

- **Purpose:** Classify customers based on Recency, Frequency, and Monetary values.
- **Inference:** Loyal + Potential Loyal customers ≈ **72%**, providing a strong base for retention programs.

📊 5. Returns Analysis

- **Visualization:** Bar chart showing successful vs. returned orders.
- **Purpose:** Measure product/service quality and customer satisfaction.
- **Inference:** Returns ≈ **40%**, highlighting a critical area for operational improvement.

📊 6. Age Group Analysis

- **Visualization:** Combined bar and line chart showing customer count and revenue by age group.
- **Purpose:** Identify the most valuable age segments.
- **Inference:** Customers aged **20–60** contribute maximum revenue, with a peak around age 40. Marketing should focus on this demographic.

📊 7. Revenue Trend by Year & Quarter

- **Visualization:** Line chart showing quarterly revenue across 2020–2023.
- **Purpose:** Track revenue stability and growth patterns.

- **Inference:** Revenue fluctuates between $45M–$47M per quarter, with peaks in early 2021 and late 2022, indicating seasonal demand cycles.

✅ **Summary**

Each dashboard component serves a **specific analytical purpose**:

- **KPI Cards:** Overall performance
- **Gender Distribution:** Demographic balance
- **Category Filter:** Product-level insights
- **RFM Segmentation:** Customer loyalty classification
- **Returns Analysis:** Operational efficiency
- **Age Group Analysis:** Demographic targeting
- **Revenue Trend:** Growth and forecasting

Together,

Chapter 5: Business Impact & Inference

- **Retention Programs:** Focus on Loyal & Potential Loyal customers.
- **Win-Back Campaigns:** Target At Risk customers.
- **Returns Reduction:** Investigate product quality, descriptions, and logistics.
- **Marketing Strategy:** Concentrate on 20–60 age group.
- **Balanced Targeting:** Maintain gender-neutral campaigns.

Business Impact

The analysis conducted through the Power BI dashboard has significant implications for e-commerce business strategy. By transforming raw transactional data into actionable insights, the project demonstrates how data-driven approaches can directly influence customer engagement, retention, and profitability.

Customer Retention:

- The identification of **Loyal and Potential Loyal customers (~72%)** highlights a strong base for retention programs. Businesses can design loyalty rewards, personalized offers, and exclusive memberships to strengthen long-term relationships.

Revenue Optimization:

- The revenue trend analysis across quarters shows **stable growth with seasonal peaks**, enabling businesses to forecast demand and allocate

resources effectively. This supports better inventory management and financial planning.

Return Management:

- With returns accounting for **~40% of transactions**, businesses can investigate product quality, sizing issues, or delivery challenges. Reducing returns not only improves customer satisfaction but also lowers operational costs.

Targeted Marketing:

- Age group analysis revealed that customers aged **20–60 contribute maximum revenue**. Marketing campaigns can be tailored to this demographic, ensuring higher conversion rates and efficient use of advertising budgets.

Balanced Market Reach:

- The near-equal gender distribution indicates that products appeal broadly across demographics. This balance allows businesses to maintain inclusive marketing strategies without skewing toward one segment.

📊 **Inference**

From the insights generated, several key inferences can be drawn:

Strong Loyalty Potential:

1. The majority of customers are either Loyal or Potential Loyal, suggesting that retention-focused strategies will yield high returns.

Returns as a Critical Challenge:

2. The high percentage of returns indicates a need for operational improvements. Addressing this issue could significantly enhance profitability.

   Revenue Stability:

3. The consistent revenue trend across years demonstrates resilience in customer demand, providing confidence for long-term planning.

   Demographic Focus:

4. Customers aged 20–60 are the backbone of revenue generation, making them the prime target for engagement and product development.

   Inclusive Appeal:

5. Balanced gender distribution confirms that the business caters effectively to diverse customer groups, reducing dependency on a single demographic.

   ✅ **Summary**

The business impact of this analysis lies in its ability to **guide strategic decisions** with precision. By leveraging RFM segmentation, demographic insights, and revenue trends, businesses can **strengthen customer relationships, reduce inefficiencies, and maximize growth opportunities**. The inferences drawn provide a clear roadmap for **retention programs, marketing strategies, and operational improvements**, ensuring that the organization remains competitive in the evolving e-commerce landscape.

Conclusion

This project demonstrates how **Power BI dashboards** can convert raw e-commerce data into **meaningful insights**. By applying **RFM segmentation**, businesses can identify loyalty patterns, reduce churn, and optimize marketing. The analysis supports **strategic decision-making, customer retention, and revenue growth**.

The completion of this project highlights the transformative role of **business intelligence tools** in modern e-commerce. Through systematic data cleaning, RFM segmentation, and dashboard visualization, the study successfully uncovered meaningful insights into customer behavior. The findings revealed that a majority of customers fall into **Loyal and Potential Loyal categories (~72%)**, underscoring the importance of retention programs. At the same time, the analysis of **returns (~40%)** and **age group contributions (20–60 years driving maximum revenue)** provided clear directions for operational improvements and targeted marketing.

The dashboard proved to be more than a reporting tool; it became a **strategic decision-support system**. By presenting KPIs, segmentation charts, and revenue trends in an interactive format, it enabled stakeholders to quickly identify opportunities and risks. The balanced gender distribution further emphasized the broad appeal of the platform, suggesting that marketing efforts should remain inclusive.

In conclusion, this project demonstrates how **Power BI dashboards can bridge the gap between raw data and business strategy**. The insights generated empower organizations to design retention campaigns, optimize marketing spend, and reduce operational inefficiencies. Looking ahead, the framework can be extended with **predictive analytics, churn modeling, and sentiment analysis** to further enhance customer understanding. This ensures that businesses remain agile, customer-centric, and growth-oriented in an increasingly competitive digital marketplace.

References

- Power BI Documentation
- Academic papers on RFM segmentation
- Industry reports on e-commerce analytics
- Business Intelligence Best Practices (Gartner Reports)

Appendix

- Sample Dashboard Screenshots
- Data Cleaning Scripts (Python)
- DAX Formulas Used in Power BI