INTRODUCTION

After completing my SQL projects, I decided to work on a full end-to-end project using Microsoft Excel. While searching for a suitable dataset, I came across a YouTube tutorial that provided access to a comprehensive sales dataset. The dataset contained six tables, but for this analysis, I focused on four - **Customer**, **Product**, **Sales Territory**, and **Fact_Internet_Sales**.

The **Customer** table contained information about customer demographics, the **Product** table described the products sold, the **Territory** table provided location data, and the **Fact_Internet_Sales** table included transactional details such as order dates, sales amounts, and foreign keys linking to the other tables.

After selecting these tables, I loaded them into **Power Query** for data preparation. I selected only the necessary columns, removed unwanted ones, and merged all four tables using their respective primary and foreign keys. Once the merge was complete, I cleaned the dataset by removing duplicates and trimming text fields.

The final dataset consisted of **22 columns** and over **60,000 rows** of transactional data, which I then loaded into Excel for analysis. Upon reviewing the dataset, I discovered it was a **transactional sales dataset** containing both sales and customer demographic details. Since the original business name was not stated, I named it **OmniMart**, as the product categories resembled those typically sold in a supermarket.

ABOUT OmniMart

OmniMart is a global retail supermarket chain that offers a wide range of consumer products from groceries and household essentials to electronics. With a strong presence across multiple countries, OmniMart serves a diverse customer base, attracting shoppers from different countries through its wide range of retail products.

As part of its commitment to data-driven decision-making, OmniMart collects and analyzes customer transaction and demographic data to better understand purchasing behaviour, optimize marketing strategies, and improve customer retention.

This dataset captures key information such as **sales transactions, product details, customer demographics, and engagement patterns**, enabling a deeper analysis of revenue trends, customer value, and churn behaviour.

DATA PREPARATION AND TRANSFORMATIONS

Dataset Overview

The dataset contained both transactional and customer demographic information, including product details, sales dates, order quantities, unit prices, customer attributes (such as gender, education, income, age, and marital status), and sales territory data. The transactions covered a period between **2005 and 2008**.

After merging the four selected tables namely: *Customer*, *Product*, *Sales Territory*, and *Fact_Internet_Sales*, the final dataset consisted of **22 columns** and over **60,000 rows** of data.

To enhance the quality and depth of analysis, several new columns were created to generate additional insights. The table below outlines the new columns and their purposes.

New Columns and Adjustments Made

1. **Full Name** - This column was created by combining the *First Name* and *Last Name* fields using the CONCATENATE function to uniquely identify each customer.

   Formula: Full Name = CONCATENATE ("First_name", " ", "Last_name")

2. **Total Revenue** - This column was created to calculate the total amount generated from each transaction.
   Formula: Total Revenue = Order Quantity * Unit Price

3. **Total Transaction Cost** - This column determines the total cost incurred for each transaction.
   Formula: Total Transaction Cost = Total Cost per Product * Order Quantity

4. **Profit** - This key metric represents the profit earned from each transaction.
   Formula: Profit = Total Revenue - Total Transaction Cost

5. **Days Since Last Purchase** - This column was introduced to measure how long it has been since a customer's last purchase.

   - The reference date was set as the most recent order date in the dataset (*31/07/2008*).
   - Using the MAXIFS function, the latest order date per customer was retrieved: Customer_Last_Order_Date = MAXIFS (Order_Date_Range, Full_Name_Range, Customer_Name)
   - The difference between the reference date and each customer's last purchase date was then calculated as:
     =DAYS(reference_order_date, customer_last_order_date)

6. **Customer Status** - This column categorized customers into three engagement levels: Active, At Risk, and About to Churn, based on the Days Since Last Purchase metric.
   Percentile analysis was used to determine this range. Below is the excel formula used:

   =PERCENTILE (Days_Since_Last_Purchase, 0.25) = 73 days
   =PERCENTILE (Days_Since_Last_Purchase, 0.75) = 241 days

Classification:

The above calculation means that customers whose days since last purchase is

1. ≤ (Less than or equal to) 73 is considered Active
2. 73 < Days ≤ 241 (above 73 days but less than or equal to 241) is considered At Risk
3. >(greater than) 241 is considered About to Churn

This segmentation helped identify customer engagement patterns and potential churn risks.

7. **Age Group** - This column grouped customers by age category to understand which segments contribute more to total sales.

   - First, the Year was extracted from the BirthDate column, then subtracted from 2025 to determine each customer's current age.

- The excel formula below is used to group the ages:
  =IF([@Age]<=29, "20–29", IF([@Age]<=39, "30–39", IF([@Age]<=49, "40–49", IF([@Age]<=59, "50–59", "60+"))))

The formula means:

Customer less than or equal 29 are within “20 - 29” age group

Customer less than or equal 39 are within “30 - 39” age group

Customer less than or equal 49 are withing “40 - 49” age group

Customer less than or equal 59 are withing “50 - 59” age group,

and finally, customers who do not fall within this age are considered 60+

This grouping provided insight into age groups that were most active or profitable.

8. **Salary Category** - To assess income-based purchasing behaviour, the Yearly Income column was grouped into three categories using a nested IF statement:
   =IF([@YearlyIncome] <=64000, "Low Income", IF(AND([@YearlyIncome]>64000, [@YearlyIncome] <=117000), "Mid Income", "High Income"))

The formula means:

Customers earning less than or equal to $64,000 are Low Income earners

Customers earning above $64,000 but less than $117,000 are Mid Income earners

Customers whose salary are above the above are High Income earners.

This classification allowed revenue analysis across Low, Mid, and High-Income customer segments.

DATA ANALYSIS USING PIVOT TABLES

Overview

After preparing and enhancing the dataset, Excel Pivot Tables were used to summarize, analyze, and visualize key insights. The Pivot Tables made it easy to transform large volumes of data into meaningful summaries such as **total revenue by product**, **sales by country**, **monthly performance trends**, and **customer churn patterns**.

The results from these analyses were used to develop two interactive dashboards:

1. **Sales Dashboard** - focused on revenue, product, and regional performance.
2. **Customer Dashboard** - focused on customer demographics, behaviour, and retention.

Key Analyses Performed

1. Sales Performance Analysis
Pivot tables were used to assess OmniMart’s revenue and profit trends. Metrics such as **Total Revenue**, **Total Cost**, **Total Profit**, and **Total Quantity Sold** were aggregated to measure overall business performance.

- **Monthly and Yearly Revenue Trends** were visualized using line charts to show sales progression over time.
- **Revenue by Country and Product** was analyzed through bar charts to highlight top-performing markets and best-selling products.
- **Top and Bottom Customers** were identified to recognize VIP buyers and customers who may need incentives to improve engagement.

2. Customer Demographics and Retention Analysis
Customer-related pivot tables explored behavioral and demographic patterns.

- **Revenue by Gender, Education Level, and Age Group** were analyzed to determine which segments contributed most to total revenue.
- **Churn Distribution** was visualized using a doughnut chart to display customer segmentation into Active, At Risk, and About to Churn.
- **Customer Count by Country** provided insight into where most customers were located and how this affected total revenue.

Finally, to make the dashboards more dynamic and user-friendly, **Slicers** were added. These allow users to filter data by attributes such as **Gender**, **Education Level**, **Income Category**, **Age Group**, and **Country**, making it easier to explore specific insights or compare customer segments.

After the dashboard design, I decided to give a full insight and exploration of the dataset. Below, I have written a full Report showing Problem Context, Problem Statement, Insights, Issues to solve, Recommendation, Metrics to tracks and finally conclusion.

PROBLEM BACKGROUND

Between **2005 and 2008**, OmniMart, a global retail store, faced a growing risk of losing 49% of its customers to churn despite recording a 1,023% increase in customer growth during the same period. OmniMart, which deals in a wide range of products from groceries and household essentials to electronics, also experienced a 205% increase in revenue.

However, despite this impressive growth in both revenue and customer base, only 21% of its customers were active, representing less than one-third of the total customer count. While 49% were at risk of churning, 30% were already about to churn.

If OmniMart can increase its active customer base from 21% to 50%, overall revenue is projected to grow by an average of 25% within the next 12 months. This can be achieved by re-engaging at least 40% of the about-to-churn customers while gradually reducing the at-risk group by 20% through sustained customer retention strategies.

PROBLEM STATEMENT

How can OmniMart design and implement a sustainable customer retention strategy that reduces the at-risk customer segment by 20% and increase the active customer base from 21% to 50% within the next 12 Months?

This Problem Statement is **SMART** because

1. It is **Specific** - To increase the active customer base and reduce the at-risk customer segment through a sustainable customer retention strategy.

2. It is **Measurable** - Success can be measured by tracking the increase in active customers from **21% to 50%** and quantifying a **20% reduction** in at-risk customers.

3. It is **Action-Oriented** - **Design and implementation** of a customer retention strategy.

4. It is **Realistic and Relevant** - The targets are achievable within the business context and directly address OmniMart's customer churn challenge, making the objective both realistic and relevant.

5. **Time-Bound** - Timeframe of **12 months**.

INSIGHT

The sales analysis revealed key revenue drivers and regional trends that highlight OmniMart's growth areas and operational performance between 2005 and 2008.

1. **Overall Performance:** Between 2005 and 2008, OmniMart generated a total revenue of **$307.1 million** and a **profit of $126.3 million**.

2. **Top Market:** Customers from the **United States** were the highest contributors, accounting for **31% of total revenue**.

3. **Best-Selling Product:** The **Mountain-200** series was the top-selling product line, contributing **22% of total revenue** during the period.

4. **Lowest Market:** Customers from **Canada** generated the least revenue, contributing **6.8% of total sales**.

5. **Regional Concentration:** Combined, customers from the **United States and Australia** accounted for **62% of OmniMart's total revenue**, indicating strong regional sales dominance.

6. **Quarterly Trends: Quarter 2** recorded the highest revenue of **$94.7 million**, while **Quarter 1** achieved the highest profit of **$78.4 million**, suggesting better cost efficiency early in the year.

While for the customer analysis, it revealed strong growth over the years but also highlighted variations in revenue performance across demographics, education levels, and income categories.

1. **Customer Growth:** Between 2005 and 2008, OmniMart's customer base grew by **1,023%**, with **2008 alone accounting for 61%** (over 11,000 customers) of the total.

2. **Churn Distribution: 2007** recorded the highest percentage of **"About-to-Churn" customers (51%)**, while **2008** had **no customers** in this category, suggesting improved engagement in that year.

3. **Education and Revenue:** Customers with a **Bachelor's degree** generated the highest **Average Revenue per Customer (ARPC)** at **$19.2K**, while those with **Partial High School education** had the lowest at **$11.2K**.

4. **Age Group Performance:** The **30–39 age group** contributed the highest revenue, accounting for **26% ($80.0M)** of total sales.

5. **Marital Status: Married customers** generated **2.48% ($7.6M)** more revenue than **single customers**, indicating slightly higher purchasing power among married individuals.

6. **Income Segmentation: Low-income customers** contributed the largest share of revenue at **52%**, followed by **mid-income (37.9%)** and **high-income (10.1%)** groups.

ISSUES TO SOLVE

Using Issue tree, I will like to structure the problem to ensure it is Mutually Exclusive, Collectively Exhaustive (MECE)

Issue A: How can OmniMart increase engagement among existing active customers?

Sub-Issue A1 - How can purchase frequency among active customers be increased?

Sub-Issue A2 - How can customer value and loyalty be strengthened?

Issue B: How can OmniMart reduce the number of customers who are about to churn?

Sub-Issue B1. How can OmniMart identify customers most likely to churn early?

Sub-Issue B2. How can OmniMart re-engage about-to-churn customers effectively?

Issue C: How can OmniMart convert at-risk customers into active customers?

Sub-Issue C1. How can at-risk customers be identified and segmented for intervention?

Sub-Issue C2. How can targeted strategies move at-risk customers back to active status?

After evaluating all six sub-issues, based on **impact** and **feasibility**, two sub-issues were identified as priorities for immediate action: A2 (strengthening customer loyalty) and B2 (re-engaging about-to-churn customers), are considered **high impact** and **high feasibility**.

If **Sub-Issue A2** is addressed, OmniMart will quickly improve revenue and stability, as retaining customers is cheaper than acquiring them, and loyalty programs lead to increased repeat purchase frequency. This solution is highly feasible because it leverages the existing customer base to roll out programs like discounts and point systems without requiring new infrastructure.

Simultaneously, **Sub-Issue B2** must be tackled to prevent immediate financial loss. This is a high-impact task because it protects revenue from a critical, about-to-churn segment. It is highly feasible because with the dashboard designed, OmniMart can identify these customers, enabling quick wins through timely, targeted interventions such as personalized re-engagement offers.

These sub-issues were selected because they directly address the core objective of improving retention, can be implemented within 12 months, and are supported by existing data and dashboard insights.

RECOMMENDATION

1. OmniMart should launch a Customer Loyalty and Reward Program to reward repeat purchases. The goal is to increase engagement among **active customers**, strengthen loyalty, and reduce the rate at which they become inactive.

2. Use the existing churn dashboard to identify and target about-to-churn customers with personalized offers (promotions, discounts, new product highlights). The purpose is to make them feel valued, encourage an immediate return, and reduce instant churn.

3. OmniMart can create targeted campaigns which include exclusive coupons, bundle discounts, reward multipliers specifically for at-risk customers. This aims to convert them back into active purchasers, increasing overall engagement and retention.

4. Establish a Customer Feedback and Engagement System. Channels like post-purchase surveys and feedback forms to collect direct customer insights. This system helps OmniMart understand satisfaction, uncover service gaps, refine retention strategies, and build trust by making customers feel heard.

METRIC TO MEASEURE

To monitor the effectiveness of OmniMart's customer retention strategy, the following key metrics were identified:

1. **Active Customer Percentage** - This will help measure progress toward increasing active customers from 21% to 50%.
2. **Churn Rate** - To track the percentage of customers lost over a given period.
3. **Reactivation Rate** - Evaluates the success of re-engagement campaigns targeting About-to-Churn customers.
4. **Customer Retention Rate (CRR)** - Monitors the overall percentage of customers retained annually.

CONCLUSION

This project provided an in-depth analysis of OmniMart's sales and customer data from 2005 to 2008 to better understand customer behavior, revenue patterns, and churn trends. The analysis revealed that while the company recorded strong growth in both customer base and revenue, customer retention remained a major challenge, with only 21% of customers being active.

Through the use of data-driven techniques such as performance, behavioural, and churn analysis, the study identified key opportunities to improve retention and engagement. The proposed strategies, including loyalty and reward programs, personalized re-engagement campaigns, and targeted win-

back offers, are projected to increase the active customer base to 50% and reduce churn within 12 months.

By continuously tracking key performance metrics such as active customer percentage, retention rate, churn rate, and reactivation rate, OmniMart can strengthen its customer relationships, improve overall satisfaction, and sustain long-term business growth.