Report

Course 2 : Data Management and Analysis with MS Excel

Project Title : KPMG Data Analysis using Excel.

Submitted By : Supriya Kore

---

1. Introduction

In today's data-driven world, organizations rely heavily on accurate data analysis to guide business decisions.
This project was conducted as part of the KPMG Data Analytics Virtual Internship Program. It simulates a real-world business scenario using Microsoft Excel for transforming, analyzing, and generating actionable insights from customer data.

The project was divided into 6 main tasks—ranging from raw data quality assessment to customer segmentation, CLV calculation, and final recommendations. It showcases how powerful tools in Excel (like Pivot Tables, VLOOKUP, and charts) can uncover meaningful patterns.

2. Objectives

The main objectives of this project are:

- To clean and format raw datasets for analysis
- To segment customers by demographic and behavioural categories
- To calculate Customer Lifetime Value (CLV) and analyse it
- To identify business patterns and insights

- To provide actionable recommendations based on findings
- To practice Excel-based business data analytics in a practical setup

3. Tools and Techniques Used

All tasks were completed using **Microsoft Excel**, utilizing the following tools and formulas:

- **VLOOKUP & IFERROR** – to merge datasets and handle missing data
- **Pivot Tables** – for summarizing and comparing segment data
- **COUNTIF, COUNTA, AVERAGEIF** – for custom logic and calculations
- **Date functions** – to extract year, month, and quarter
- **Charts** – to visualize trends and comparisons (bar, line, pie)
- **CLV Formula:**

CLV=APV×Purchase Frequency×Tenure\text{CLV} = \text{APV} \times \text{Purchase Frequency} \times \text{Tenure}CLV=APV×Purchase Frequency×Tenure

4. Data Sources and Preparation

There were **two primary datasets** used:

1. **Customer Demographic Dataset:**
    - Contained fields like customer ID, gender, wealth segment, job industry category, and tenure.
2. **Transaction Dataset:**

- Included customer ID, transaction date, and list price.

Data Preparation Steps:

- Columns were cleaned and renamed for clarity
- Merged the two datasets using VLOOKUP on customer ID
- Derived new fields like APV, tenure, and total revenue
- Removed duplicates and blank entries using filters

5. Task-wise Breakdown and Analysis

- **Task 1: Data Cleaning**
- Identified missing gender, job industry, and tenure values
- Highlighted blanks using conditional formatting
- Cleaned entries to ensure consistency (e.g., uniform date formats)

- **Task 2: Data Transformation**
- Extracted Year, Month, and Quarter from transaction dates
- Added calculated fields like:
  - **Total Revenue** (Sum of list price)
  - **Number of Purchases** (Count of transactions)
  - **Average Purchase Value (APV)**
- Created a structured and usable merged sheet

- **Task 3: Customer Segmentation**

Using Pivot Tables, segmentation was done on:

- **Wealth Segment:** Mass Customer, Affluent, High Net Worth
- **Gender:** Male, Female
- **Job Industry:** Retail, Health, Financial Services, etc.

Findings:

- High Net Worth customers had the highest APV
- Most customers were in the “Mass” segment, but with lower CLV
- Health & Finance customers spent more frequently

- **Task 4: Demographic Analysis**
- **Gender Analysis:**
    - Males showed slightly higher average CLV than females
    - Female customers were dominant in Mass & Affluent segments
- **Job Industry Analysis:**
    - Highest CLV observed in Health & Financial Services
    - Retail customers had lower CLV

These demographics helped understand the value contribution from various customer groups.

- **Task 5: Customer Lifetime Value (CLV) Calculation**

CLV Formula:

CLV=APV×PF×Tenure\text{CLV} = \text{APV} \times \text{PF} \times \text{Tenure}CLV=APV×PF×Tenure

- **APV:** Revenue / Number of Purchases
- **Purchase Frequency (PF):** Total Transactions / Unique Customers
- **Tenure:** Looked up from demographic sheet

Segment-Wise CLV:

- HNW → ₹21,000+
- Affluent → ₹17,000+
- Mass → ₹14,000+

CLV was visualized using bar charts & segment comparison.

- **Task 6: Executive Summary & Recommendations**

Key Insights:

- HNW customers give long-term value — low in number but high in revenue
- Female customers show good loyalty but need personalized targeting
- Health sector professionals have the highest average CLV
- Mass segment has potential if tenure is improved

Business Recommendations:

- Offer **exclusive retention plans** to HNW customers
- Improve **APV** for mass customers with bundled offers
- Focus on **urban cities and health/finance sectors** for expansion
- Personalized email campaigns based on job industry and past purchase

6. Conclusion

This project helped develop real-world business analytics skills using Excel. Each step — from cleaning, merging, analyzing, and visualizing data — contributed to meaningful customer insights.
We also learned how to make data-driven recommendations based on CLV and segmentation analysis.

Microsoft Excel, when used with the right logic and structure, can become a powerful tool for business insight generation — even without code!