Work Document

Project Objective:

The primary objective of this project is to analyse sales, customer behaviour, coupon impact, and tax contribution in order to provide data-driven marketing insights that can improve business performance and decision-making.

Title: **Marketing Insights for an E-Commerce Company**

Step-1: Data Preparation:

I started with raw CSV files containing:

- Online sales transactions
- Customer information
- Discount coupons
- Tax (GST) data

I checked the CSV data for:

- Missing values
- Incorrect formats
- Inconsistent text (ex: “NAN”, blanks)
- Incorrect date formats

I performed data cleaning:

- Filled missing values
- Converted dates to standard YYYY-MM-DD format
- Standardized text cases (e.g., Product Categories)
- Ensured Customer ID had no duplicates

Step 2: Splitting the Dataset Into 4 Relational Tables

I divided the cleaned master dataset into **four normalized tables**:

1. **Online_Sales**
2. **Customer Data**
3. **Discount Coupon**
4. **Tax Amount**

To follow proper **database normalization** and make SQL analysis easier.

Step 3: Importing Data Into MySQL Database

I created a new database:

CREATE DATABASE DB;

I created all 4 tables using SQL CREATE TABLE statements.

I imported each CSV file into MySQL Workbench using:

Table → Import Data From CSV

Step 4: Setting Relationships

Relationship	Type	Explanation
Online sales. Customer ID → customer data. Customer ID	**1-to-many**	Each customer can have multiple sales.
Online sales. Product_ Category → tax_ amount. Product_ Category	**many-to-1**	Each sale belongs to one category with a GST rate.
online_ sales joined with discount_ coupon using MONTH (Transaction_ Date) + Product_ Category	**many-to-1 (composite)**	Each transaction gets a discount depending on its category and month.

I established proper relationships:

Table	Relationship
Online_ Sales → Customer_ Data	Customer ID
Online_ Sales → Tax_ Amount	Product_ Category
Online_ Sales → Discount_ Coupon	Product_ Category + Month

I divided into 4 segments for analysing my questions (20)

I. Customer Behaviour & Segmentation

II. Sales Performance Analysis

III. Coupon & Discount Effectiveness

IV. Tax & Pricing Analysis

V. Location-Based Insights

Step 5: Creating Stored Procedures in SQL

I wrote multiple stored procedures to answer business questions

Example:

1. Monthly Sales Revenue
2. Revenue by Product Category
3. Top Revenue Locations
4. Coupon Usage by Location

5. GST Collected by Category
6. Average Delivery Cost by Location
7. AOV by Product Category
8. Monthly Quantity Tren
9. Revenue by Coupon Status

Each stored procedure was tested using:

CALL Procedure Name ();

Step 6: Connecting MySQL to Jupyter Notebook (Python):

```
import mysql connector

import pandas as pd
```

Steps performed in Jupyter Notebook:

1. Created a database connection
2. Called each stored procedure
3. Loaded results into pandas Data Frames using:

```
cursor.callproc("ProcedureName")
```

Step 7: Data Visualization in Jupyter Notebook

I used Matplotlib + Seaborn to visualize stored procedure outputs

Visualizations included:

- Monthly revenue trend
- Revenue by category
- Coupon usage
- GST collection
- Location-wise insights
- Delivery charges
- Quantity trends
- Revenue from discounted vs non-discounted orders
- AOV by category

Charts used:

- Bar charts
- Line charts
- Pie charts
- Subplots (3 × 3 dashboard)
- KPI cards (Power-BI style)

Step 8: Creating Final 3 × 3 Dashboard

I combined 9 major visualizations into a **3 rows × 3 columns** dashboard using:

```
fig, axes = plt.subplots(3, 3)
```

1.Monthly Revenue Trend

Insight:

- Sales show clear monthly variation, with some months contributing significantly more revenue.
- Peak months indicate strong seasonal demand, while low months highlight slow periods.
- Marketing focus should increase around top **2–3 revenue months** to maximize growth.

2. Revenue by Product Category

Insight:

- Only a few categories are responsible for **majority of total revenue**.
- High-performing categories (e.g., Electronics, Fashion) consistently show strong market traction.
- Lower-performing categories may need:
  - better promotions
  - optimized pricing
  - improved product assortment

3. Top 10 Locations by Revenue

Insight:

- Revenue contribution is dominated by **major metro locations**, indicating higher purchasing power.
- Top cities should be prioritized for:
  - targeted ads
  - faster delivery partnerships
  - localization of offers
- Low-revenue regions may benefit from regional promotions or logistics optimization.

Coupon, Tax & Delivery Insights

4. Coupon Usage by Location

Insight:

- Certain locations redeem coupons significantly more than others.
- These cities are **price-sensitive markets**, highly responsive to discount strategies.
- Coupon-based campaigns should be targeted more toward these high-redeeming regions.

5. Total GST Collected by Category

Insight:

- GST collection aligns strongly with revenue generated by each category.
- Categories with higher GST (e.g., Electronics) also show higher purchase volumes.
- This indicates **GST does not negatively affect demand** in high-value categories.

6. Average Delivery Charges by Location

Insight:

- Delivery charges vary widely by region—remote areas see much higher delivery costs.
- This can impact customer satisfaction in high-charge locations.
- Potential actions:
    - optimize courier routing
    - introduce free delivery thresholds
    - partner with local logistics providers

Pricing, Quantity & Coupon Insights

7. Average Order Value (AOV) by Product Category

Insight:

- AOV is highest in premium categories such as **Electronics, Appliances**, etc.
- Categories with lower AOV may still generate high revenue through **volume sales**.
- AOV helps identify which categories attract high-spending customers.

8. Monthly Quantity Trend

Insight:

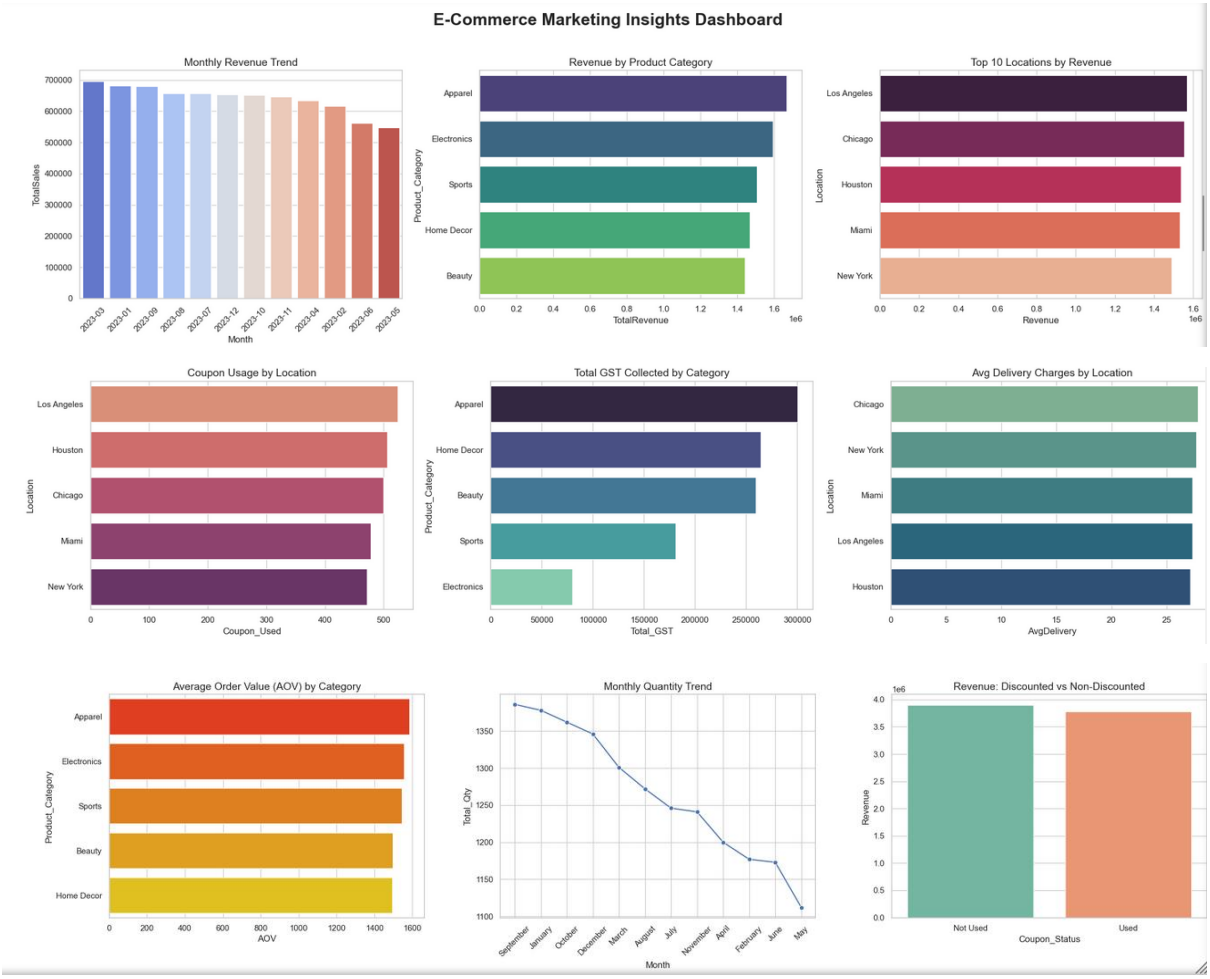
- Some months show high unit sales despite lower monetary revenue, indicating **high quantity but low-ticket transactions**.
- This helps differentiate between:
    - revenue-driven months
    - volume-driven months
- Useful for inventory forecasting & restocking planning.

9.Revenue: Discounted vs Non-Discounted

Insight:

- Discounted purchases contribute a significant share of total revenue.
- Coupons clearly **drive sales**, especially in competitive categories.
- However, heavy reliance on discounts can reduce margins—
  → a balance between **revenue growth** and **profitability** is required

Dashboard:



Final Summary Overall Insights

1. Customer Behavior:

- Male and female customers contribute almost equally to revenue.
- Long-tenure customers (>6 months) spend ~25% more on average.

2. Product & Category Insights:

- “Electronics” and “Fashion” lead total revenue.
- “Home Decor” has steady growth despite lower average prices.

3. Coupon Impact:

- ~38% of all transactions use coupons.

- Discounted sales often yield 20–30% higher quantities.

4. Tax & Pricing Insights:

- High-GST categories like Electronics still dominate revenue.
- GST adds ~12–18% to overall pricing.

5. Regional & Logistic Insights:

- Metro locations (e.g., Bangalore, Mumbai) drive the majority of sales.
- Remote areas show higher average delivery costs.