

INTERNSHIP PROJECT
ON
ANALYSING E-COMMERCE DATASETS
USING SQL

BY
KANDURI PADMA PRIYA
REDGNO:32206410024

AT:VYSHANAVI TECHNOLOGIES.
DURATION:29 OCT 2024 TO 25 DEC 2024

OBJECTIVE:

- **Determine the total revenue generated from each payment method.**
- **Identify the product categories with the highest average discount.**
- **Analyze the monthly revenue trend for the past year.**
- **Identify the top 10 customers based on their total spending.**
- **Calculate the average final price of products in each category.**

Approach for Analyzing E-Commerce Datasets Using MySQL

DATA COLLECTION AND PREPARATION

STEP1: DOWNLOADING DATA SETS FROM KAGGLE

STEP2:IMPORTING DATA SET INTO MY SQL USING MY SQL WORKBENCH

STEP3: CREATE NORMALIZED TABLE TO AVOID REDUNDANCY ENSURING CONSISTENCY

STEP4:PERFORM DATA CLEANING(i.e) REMOVING MISSING VALUES,DUPLICATES,NULLS USING SQL QUERIES

STEP5: ANALYZING GIVEN OBJECTIVES USING SQL QUERIES



PREPARING AND IMPORTING DATA

• PREPARING DATA

- a. **Visit [Kaggle's website](#) and log in to your account.**
- b. **Search for desired dataset(e-commerce)**
- c. **Click on download button and download .ZIP file**
- d. **Extract files from .ZIP to access .CSV file using WINRAR**

• IMPORTING DATA

1. **Navigate to the database where you want to import a .csv file**
2. **Right-click the desired table (or create a placeholder table) and select "[Table Data Import Wizard](#)".**
3. **Click [Browse](#) to locate your CSV file on your system.**
4. **Select the file and click [Next](#).**

ANALYSING DATA

- **Total Revenue by Payment Method:**
This analysis reveals the revenue contribution of different payment methods (e.g., credit cards, PayPal) to identify customer preferences and optimize payment options.
- **Highest Average Discount by Product Category:**
Identifies which product categories (e.g., electronics, fashion) received the highest average discount, providing insights into promotional strategies and customer appeal.
- **Monthly Revenue Trend for the Past Year:**
Tracks revenue variations month-over-month, helping to identify seasonal trends, peaks, and potential periods for strategic campaigns.
- **Top 10 Customers Based on Total Spend:**
Lists the customers with the highest cumulative spending, enabling targeted marketing and loyalty-building strategies for high-value customers.
- **Average Final Price of Products by Category:**
Calculates the average post-discount price for each product category to assess pricing strategies and understand customer purchase behavior.

CHALLENGES :

Extracting CSV File from a ZIP Archive

- Encountered difficulties in automating the extraction process for CSV files from compressed ZIP files.
- Required identifying efficient Python libraries (like zipfile) to streamline the process.

Data Verification for Cleanliness and Normalization

- Faced challenges in validating whether the dataset was properly cleaned and normalized.
- Included checking for missing values, inconsistent formats, and outliers, which required iterative preprocessing.

Handling Syntax Errors with Special Characters in Column Names

- Syntax errors occurred while working with column names containing special characters.
- Realized the importance of using backticks (`) in the syntax to reference such columns correctly in queries and code.

Iterative Debugging and Testing

- Debugging syntax issues consumed significant time, emphasizing the need for careful observation and documentation of coding stand

SOLUTIONS AND LEARNINGS

- Gained proficiency in handling ZIP files and managing special cases in data preprocessing.
- Learned the importance of robust error handling and adherence to coding best practices to mitigate such challenges in future projects.

CONCLUSION

- ❖ **Successfully analyzed e-commerce datasets to generate actionable insights, such as revenue trends, customer behavior patterns, and product performance.**
- ❖ **Overcame challenges in data extraction, cleaning, and syntax handling, enhancing technical and problem-solving skills.**
- ❖ **The project provided a deeper understanding of data analysis workflows, from preprocessing to visualization, ensuring data-driven decision-making.**
- ❖ **Key outcomes included a comprehensive report on sales trends, customer segmentation, and pricing strategies, which could support business optimization efforts.**
- ❖ **This experience reinforced the importance of meticulous data preparation, effective use of analytical tools, and a structured approach to problem-solving in real-world scenarios.**