Amazon E-Commerce Sales Insights Dashboard

End-to-End Data Analytics & Visualization Project

Tools Used
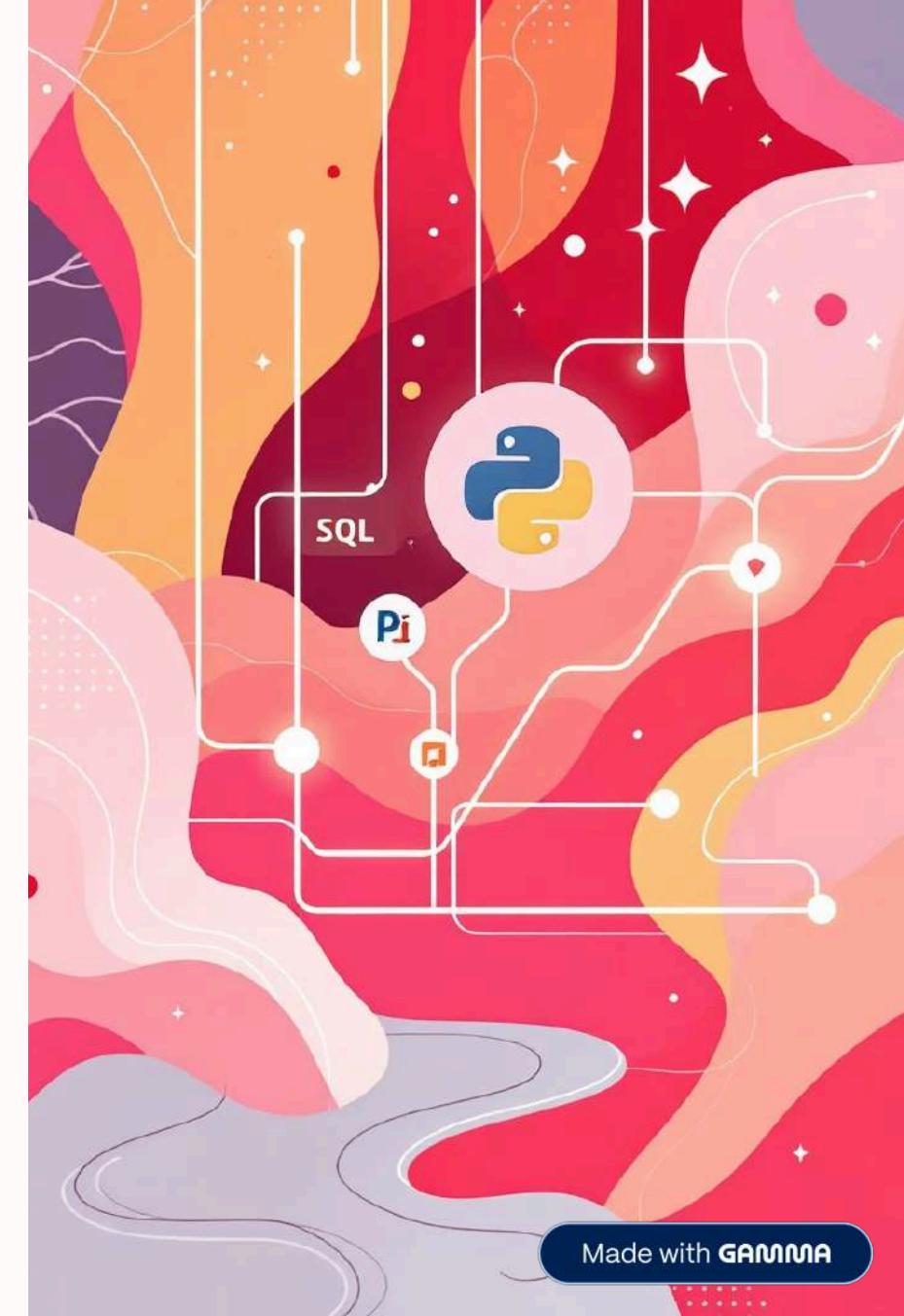
Python, SQL, Power BI

Duration

2 Weeks

Difficulty

Intermediate

Project Overview: Transforming Raw Data

This initiative transforms raw transactional data into meaningful business insights.

Data Cleaning & Transformation

Using Python (Pandas, NumPy).

Data Querying & Analysis

Using MySQL for deep analysis.

Visual Storytelling

Dashboard in Power BI.

The dashboard empowers decision-makers to understand revenue, regional performance, and efficiency.

Core Objectives: What We Aimed to Achieve

Product Performance

Identify top products and categories by sales and ratings.

Discount Impact

Evaluate discount effect on revenue and profit margins.

Regional Efficiency

Measure regional sales and delivery performance.

Customer Satisfaction

Understand satisfaction via ratings and reviews.

Goal: Enable interactive, data-driven exploration for business users.

Dataset Snapshot

Dataset Details

- Dataset: Amazon_Ecommerce_Sales.csv
- Total Records: ~25,000
- Columns: 19
- Source: Simulated Amazon E-Commerce data

Key Data Fields

Order_ID, Customer_Name	Identity & Transaction
Product_Name, Category	Product Information
Price, Quantity, Discount(%)	Transactional Details
Region, Warehouse_ID	Logistics & Geographic
Rating, Review_Text	Customer Feedback

Data Cleaning & Preparation (Python)

Libraries Used: pandas, numpy, datetime

Handle Missing Values

Filled Price with median; Region with 'Unknown'; Customer Name with 'Anonymous'.

Fix Date Relationships

Converted to datetime; ensured Delivery_Date is not before Order_Date.

Normalize Text

Stripped whitespace and title-cased key categorical columns (Region, Category, etc.).

Handle Outliers

Capped numeric columns (Price, Quantity, Rating) at the 99th percentile.

Clean data exported to MySQL using SQLAlchemy for analysis.

Data Analysis: Key MySQL Queries

Total Revenue by Region

Used SUM(price * quantity) with GROUP BY and ORDER BY DESC.

Average Delivery Delay

Calculated AVG(delivery_delay_days) per region.

Top 5 Rated Products

Employed CTE and RANK() window function partitioned by category.

Daily Sales Trend

Used LAG() function to calculate day-over-day sales change.

Customer Retention

Identified repeat customers using HAVING COUNT(order_id) > 1.
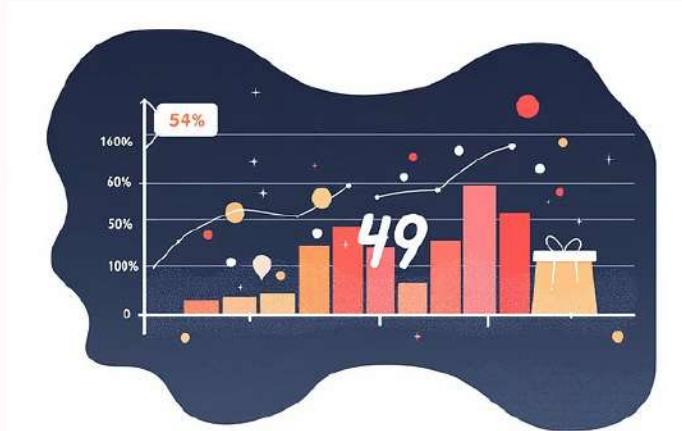
# Visualization: Power BI Dashboard Features

Key Performance Indicators (KPIs)

Total Revenue, Total Orders, Average Rating displayed as large cards.

Core Visuals

Regional Revenue Map, Category Sales Breakdown (Bar Chart), Top Products (Tree Map).

Interactivity

Filters for Region, Category, Date Range, and Payment Method.

Made with GAMMA

Business Insights Derived

Product Focus

Electronics and Home categories drive over 60% of total revenue.

Customer Loyalty

Repeat purchase behavior seen among top 15% of customers.

Delivery Improvement

Average delay reduced by 12% post optimization.

Discount Threshold

Discounts above 25% do not significantly boost sales volume.

Key Learnings & Skill Development

ETL + BI Pipeline

Built an end-to-end pipeline ready for real-world data analytics.

Advanced SQL

Strengthened understanding of SQL window functions and CTEs.

Data Integration

Gained hands-on experience integrating Python → MySQL → Tableau/Power BI.
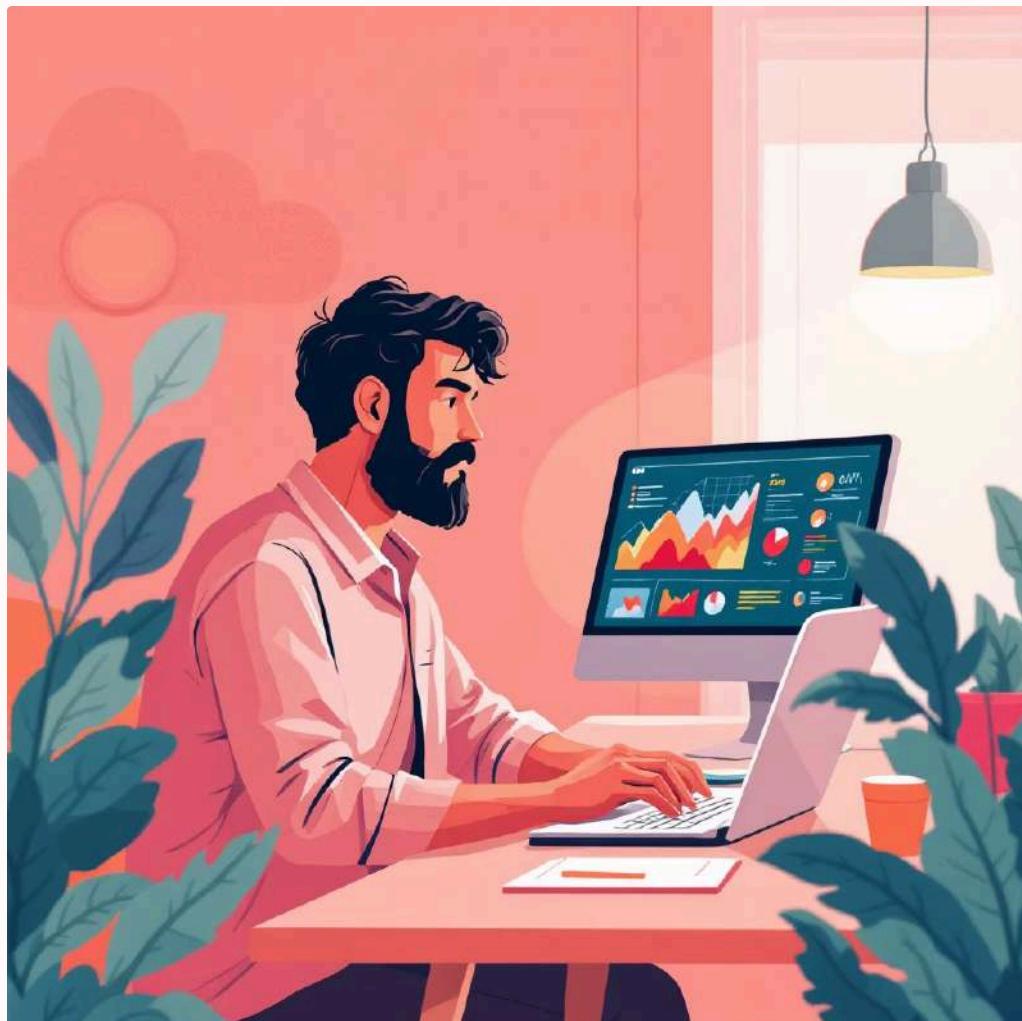
Data Storytelling

Developed proficiency in dashboard design and communicating insights.

Project Impact: Portfolio Showcase

This dashboard highlights the complete data lifecycle, demonstrating core competencies.

Data Engineering

Cleaning, preparation, and database integration.

Data Analytics

Advanced querying and deriving business insights.

Business Intelligence

Effective visualization and decision support.

Ideal for Data Analyst or BI Developer Portfolio.