Amazon E-Commerce Sales Insights Dashboard

Tools Used: Python (Pandas, NumPy), SQL (MySQL), Tableau / Power BI
Project Type: End-to-End Data Analytics & Visualization
Duration: 2 Weeks
Difficulty: Intermediate

---

1. Project Overview

The **Amazon E-Commerce Sales Insights Dashboard** project is an end-to-end data analytics initiative that transforms raw transactional data into meaningful business insights.

The workflow involves **data cleaning and transformation in Python**, **data querying and analysis using MySQL**, and **visual storytelling in Tableau or Power BI**.
The dashboard empowers decision-makers to understand revenue trends, regional performance, customer satisfaction, and delivery efficiency.

---

2. Objectives

- Identify top-performing products and categories by sales and ratings.
- Evaluate discount impact on revenue and profit margins.
- Measure regional sales performance and delivery efficiency.
- Understand customer satisfaction through ratings and reviews.
- Enable interactive data-driven exploration for business users.

---

3. Dataset Description

Dataset: `Amazon_Ecommerce_Sales.csv`

Column Name	Description
Order_ID	Unique identifier for each transaction
Customer_Name, Email	Customer identity and contact
Product_Name, Category, Sub_Category	Product information
Price, Quantity, Discount(%)	Transactional details
Order_Date, Delivery_Date	Purchase and delivery timestamps
Region, Warehouse_ID	Geographic and logistics info
Payment_Method, Shipping_Status	Payment and delivery channel data
Rating, Review_Text	Customer satisfaction
Currency, Delivery_Delay_Days	Operational and logistics data

Total Records: ~25,000
Columns: 19
Data Source: Simulated Amazon E-Commerce dataset

---

4. Data Cleaning & Preparation (Python)

🔹 Libraries Used

`pandas`, `numpy`, `datetime`

🔹 Key Cleaning Steps

Handle Missing Values

```
df['Price'].fillna(df['Price'].median(), inplace=True)
df['Region'].fillna('Unknown', inplace=True)
df['Customer_Name'].fillna('Anonymous', inplace=True)
```

1.

Fix Date Relationships

```
df['Order_Date'] = pd.to_datetime(df['Order_Date'], errors='coerce')
df['Delivery_Date'] = pd.to_datetime(df['Delivery_Date'],
errors='coerce')
mask = df['Delivery_Date'] < df['Order_Date']
df.loc[mask, 'Delivery_Date'] = df['Order_Date']
```

2.

Normalize Text Columns

```
text_cols = ['Region', 'Category', 'Sub_Category', 'Payment_Method']
for col in text_cols:
    df[col] = df[col].str.strip().str.title()
```

3.

Handle Outliers (Capping at 99th Percentile)

```
numeric_cols = ['Price', 'Quantity', 'Discount(%)', 'Rating',
'Delivery_Delay_Days']
for col in numeric_cols:
    cap = df[col].quantile(0.99)
    df[col] = np.where(df[col] > cap, cap, df[col])
```

4.

Export Clean Data to MySQL

```
import mysql.connector
from sqlalchemy import create_engine

engine =
create_engine("mysql+mysqlconnector://root:password@localhost:3306/ama
zon")
df.to_sql("sales", engine, if_exists="replace", index=False)
```

5.

---

5. Data Analysis (MySQL)

🔹 Key Analytical Queries

1️⃣ Total Revenue by Region

```
SELECT region, SUM(price * quantity) AS total_revenue
FROM sales
GROUP BY region
ORDER BY total_revenue DESC;
```

2️⃣ Average Delivery Delay by Region

```
SELECT region, AVG(delivery_delay_days) AS avg_delay
FROM sales
GROUP BY region
ORDER BY avg_delay;
```

3️⃣ Top 5 Highest Rated Products per Category (CTE + Window Function)

```
WITH top_rated AS (
    SELECT category, product_name, rating,
           RANK() OVER (PARTITION BY category ORDER BY rating DESC) AS
rnk
    FROM sales
)
SELECT category, product_name, rating
FROM top_rated
WHERE rnk <= 5;
```

4️⃣ Daily Sales Trend with LAG

```
SELECT
    order_date,
    SUM(price * quantity) AS daily_sales,
    SUM(price * quantity) - LAG(SUM(price * quantity)) OVER (ORDER BY
order_date) AS sales_change
FROM sales
```

```
GROUP BY order_date
ORDER BY order_date;
```

5️⃣ **Customer Retention Insight (Repeat Customers)**

```
SELECT customer_name, COUNT(order_id) AS total_orders
FROM sales
GROUP BY customer_name
HAVING total_orders > 1
ORDER BY total_orders DESC;
```

---

6. Visualization (Power BI)

📊 Dashboard Features

KPIs (Cards):

- Total Revenue
- Total Orders
- Average Rating

Charts & Visuals:

- **Regional Revenue Map:** Revenue by region with hover tooltips.
- **Category Sales Breakdown:** Horizontal bar chart for category-level sales.
- **Top Products by Quantity:** Tree map for high-demand products.
- **Discount vs. Sales Volume:** Scatter plot comparing discounts and revenue.
- **Customer Ratings Over Time:** Line chart tracking satisfaction trend.
- **Payment Method Share:** Donut chart of transaction modes.

Filters/Interactivity:

- Region
- Category
- Date Range
- Payment Method

---

7. Business Insights

✅ **Product Insights:** Electronics and Home categories drive over 60% of total revenue.
✅ **Customer Trends:** Repeat purchase behavior evident among top 15% customers.
✅ **Regional Insight:** Southern and Western regions show highest order volumes.
✅ **Delivery Insight:** Average delivery delay reduced by 12% post optimization.
✅ **Discount Analysis:** Discounts above 25% don't significantly increase sales volume.

---

8. Key Learnings

- Gained hands-on experience integrating **Python → MySQL → Tableau**.
- Strengthened understanding of **SQL window functions and CTEs**.
- Built an end-to-end **ETL + BI pipeline** ready for real-world data analytics.
- Developed proficiency in **data storytelling and dashboard design**.

---

9. Project Impact

The Amazon E-Commerce Sales Insights Dashboard showcases the complete data lifecycle — from data cleaning and database integration to advanced SQL querying and visualization.
This project highlights skills in **data engineering, analytics, and business intelligence**, making it ideal for **Data Analyst or BI Developer** portfolio presentations.