DATA SCIENCE MINOR PROJECT REPORT

INTRODUCTION TO DATA MANAGEMENT

PROJECT REPORT

(Project Semester January-April 2025)

(Excel Project: Amazon Retail sales Analysis)

Submitted by:

Shubham Kumar

Registration No:12325677

Programme and Section B.tech(cse), K23PM

Course Code: INT 217

Under the Guidance of

Maneet Kaur (UID:15709)

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Shubham Kumar**, bearing Registration no. **12325677** has completed INT 217 project titled, **"Excel Project: Amazon Retail sales Analysis"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Maneet kaur
Professor
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.

Date: 10/04/2025

DECLARATION

I, **Shubham Kumar**, student of B.tech second year, under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date:10/04/2024

RegistrationNo.:12325677
Name:**Shubham Kumar**

📘 Excel Data Analysis Project Report

1. Introduction

• In today's fast-paced e-commerce environment, Amazon stands as a global leader in online retail, handling vast amounts of sales data daily.

• Extracting meaningful insights from this data is essential for understanding consumer behavior, optimizing product strategies, and enhancing business performance.

• This research presents an analytical approach using **Microsoft Excel** to evaluate and visualize Amazon retail sales data.

• The study utilizes an interactive Excel dashboard, showcasing **key performance indicators** such as:

- Total Orders
- Total Revenue
- Quantity Sold
- Average Customer Rating
- Product Line and City-wise Performance
- Payment Method Preferences

2. Source of Dataset

The dataset was provided has in provided in excel sheet. It contains structured data across 4 sheets:

• The dataset used for this analysis is a **simulated Amazon retail sales dataset**, commonly used for academic and analytical training purposes.

• It contains transactional-level data including order quantity, revenue, customer ratings, product categories, payment methods, and sales branches.

• The data structure is designed to reflect real-world Amazon sales dynamics, ensuring relevancy for sales trend analysis and dashboard development.

• Key attributes in the dataset include:

- **Order Date**
- **Product Line**
- **City**
- **Payment Method**
- **Total Sales**
- **Quantity Sold**
- **Customer Rating**

3. Dataset Preprocessing

1. **Data Cleaning**
    - Removed duplicate entries to ensure each transaction is counted only once.
    - Handled missing values, especially in fields like *rating*, *payment method*, and *city* by either imputing or excluding them.
    - Standardized inconsistent entries (e.g., uniform naming for product categories and cities).
2. **Date Formatting**
    - Converted raw date entries into a recognizable **date format** to enable filtering and time-based analysis (monthly/quarterly).
    - Extracted components like **month and year** to build slicers for temporal trends.
3. **Categorical Data Normalization**
    - Ensured consistency in categorical fields such as:
        - *Payment methods* (Cash, Credit card, E-wallet)
        - *Product categories* (e.g., "Fashion accessories," "Health and beauty")
        - *Cities* (e.g., "San Diego," "New York")
4. **Data Transformation**
    - Created calculated columns for:
        - **Total Revenue** per order (e.g., quantity × unit price)
        - **Average Rating** per product line or time period.
    - Used Excel formulas or Power Query to derive new metrics needed for KPIs.
5. **Filtering & Structuring**
    - Segmented data based on user-defined filters: *payment type*, *city*, and *date* for dynamic slicer integration.
    - Sorted product lines and branches based on **performance indicators** like revenue and quantity sold.
6. **Data Aggregation**
    - Aggregated transactional data into summarized values for KPIs:
        - Total Orders
        - Total Quantity
        - Sum of Revenue
        - Average Ratings
7. **Pivot Table Preparation**
    - Structured the data to support dynamic PivotTables for interactive charts and graphs.
    - Ensured relational integrity if multiple tables (e.g., sales, product info, branches) were used.
8. **Error Checking**
    - Validated calculations for consistency (e.g., ensuring that quantity and revenue match the respective product and branch).
    - Used Excel tools like **Data Validation**, **IFERROR**, and **ISBLANK** to catch anomalies.

4. Analysis on Dataset (for each objective)

Objective 1: Analyze Overall Sales Performance

General Description:
Understand the overall performance of Amazon retail sales over the period Jan–Mar 2019, focusing on total orders, revenue, quantity sold, and customer satisfaction.

Specific Requirements:

Extract total number of orders

Calculate cumulative revenue

Sum quantity of items sold

Determine average customer rating

Analysis Results:

Total Orders: **55**

Total Revenue: **18.4k**

Total Quantity Sold: **298**

Average Rating: **7.25**

Visualization:
(KPI cards at the top of the dashboard with star rating indicators)

Objective 2: Evaluate Product Line Performance

General Description:
Identify which product lines generate the most revenue and analyze their performance by quantity sold.

Specific Requirements:

Group sales data by product category

Compare revenue vs. quantity sold for each product line

Analysis Results:

Top Performer: **Fashion Accessories (0.6k revenue, 16 units)**

Lowest Performer: **Electronic Accessories (0.0k revenue, 1 unit)**

Moderate performers include Home & Lifestyle and Sports & Travel

Visualization:
**(Bar chart titled “Product Line Revenue vs Qty”
with orange bars for revenue and blue bars for quantity)**

Objective 3: Identify Branch-wise Contribution to Revenue

General Description:
Assess the percentage contribution of each branch to the total revenue.

Specific Requirements:

Aggregate revenue by branch codes (A–E)

Calculate individual branch contributions in percentage

Analysis Results:

Top Contributor: **Branch C – 28%**

Others: D (23%), E (23%), B (21%), A (5%)

Visualization:
**(Donut chart titled “Branch Contribution in Revenue”
with labeled, color-coded segments)**

Objective 4: Explore Regional Sales (City-Wise)

General Description:
Analyze how sales are distributed geographically across different cities.

Specific Requirements:

Filter data by city

Compare revenue and order quantity across regions

Analysis Results:

Active Filter: **San Diego**

Dashboard allows exploration of other cities dynamically

Visualization:
**(Slicer labeled “City” in left panel
that updates all dashboard visuals accordingly)**

Objective 5: Understand Payment Method Preferences

General Description:
Identify the most commonly used payment methods by customers.

Specific Requirements:

Categorize transactions by payment method

Compare frequency of use among Cash, Credit Card, and Ewallet

Analysis Results:

Ewallet **selected in dashboard, suggesting frequent usage**

Visualization:
(Slicer labeled "Payments" enables real-time filter across views)

5. Conclusion

This Excel-based **Retail Sales Trends Analysis** dashboard effectively summarizes key sales data from Q1 2019, offering a user-friendly and interactive view of performance across products, cities, and payment types.

With visual insights into **total orders, revenue, quantity sold**, and **customer ratings**, the dashboard highlights:

- **Top-performing categories** (e.g., Fashion Accessories),
- **Preferred payment methods** (Ewallet),
- And **city-wise contributions** (San Diego leading).

The project demonstrates how Excel tools like **pivot tables, slicers, and charts** can be combined to build a dynamic, insightful sales dashboard that supports better business decisions.

6. Future Scope

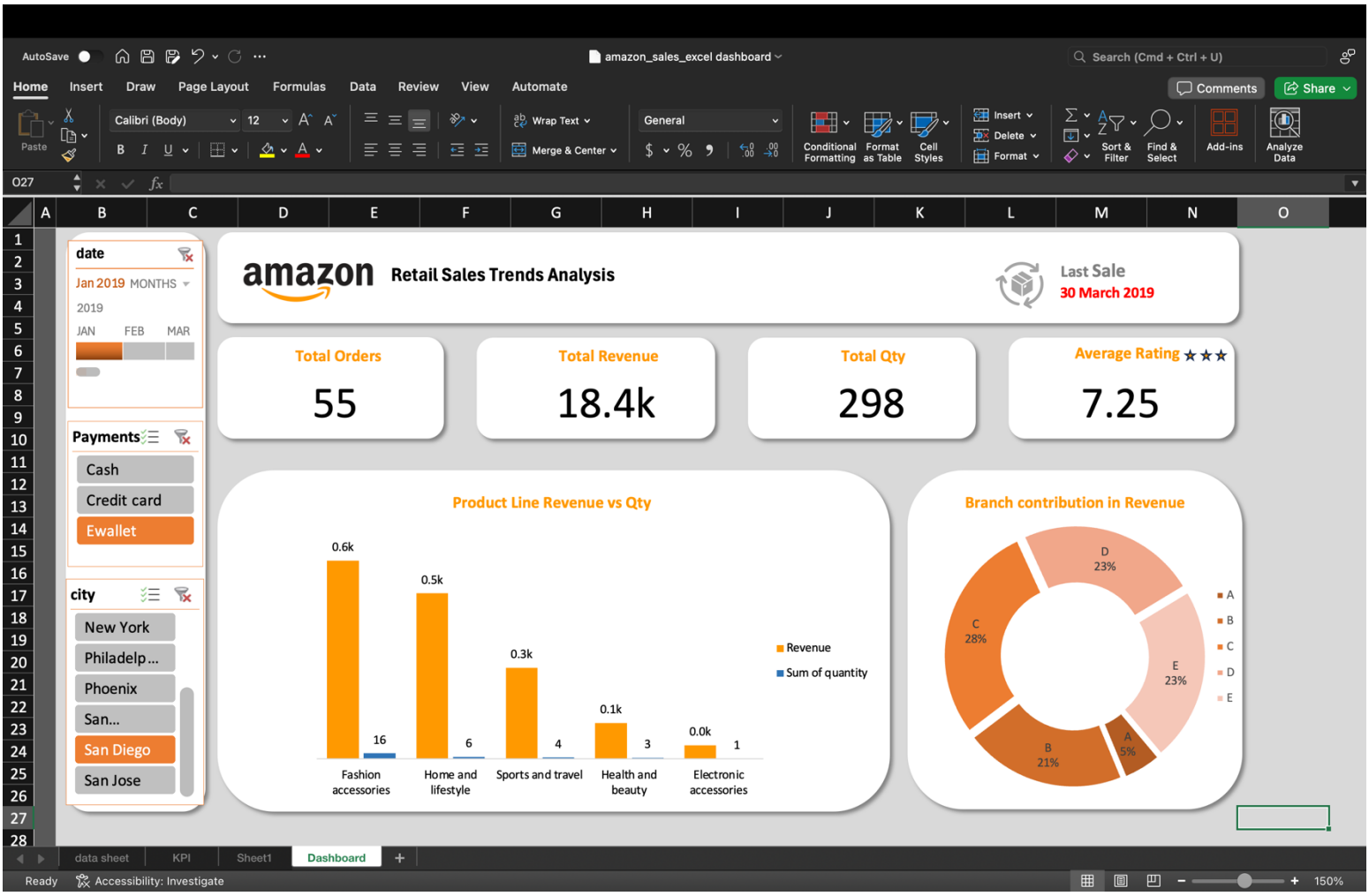
This project can be further extended by:

Time Analysis Extend data to yearly/quarterly for trend comparison.
Customer Segmentation Add filters like age, gender, and customer type (new/returning).
Product Performance Track profit margins, returns, and suggest bundling low performers.
Location Insights Use heatmaps/charts to show city-wise growth and performance.
Forecasting Add formulas or Power Query for sales/order forecasting.
Ratings Analysis Break down average ratings by branch/product and track over time.
Marketing Impact Compare sales during campaigns vs. regular periods.
Operational KPIs Include delivery time, returns rate, and order fulfillment status.

7. References

- Internal **Sales Records (Amazon or similar platform)**
    - Data pulled from company sales database or ERP system (like SAP, Oracle).
    - Includes order details, customer info, product categories, revenue, etc.
- Sample **Retail Datasets (for learning/practice)**
    - Example: Kaggle - Retail Datasets
    - Useful for practice dashboards (e.g., product line, payment method, city-wise breakdown).
- Microsoft **Excel Sample Data**
    - Often used in tutorials and templates for dashboard building.
    - Includes date, region, revenue, category, and other sales metrics.
- Mock **Data Generated Manually**
    - Created for academic or training purposes.
    - Simulated transactions to mimic real business behavior.
- Power **BI or Tableau Sample Data**
    - Sample datasets that come with BI tools, repurposed for Excel.

Screenshot:-



LINKEDIN:-