

# **Amazon Electronics Sales Data Analysis and Dashboard (Excel/Power BI)**



## **THESIS FOR INFOTACT SOLUTIONS**

### ***Team Member:***

Aman Jawalekar  
Tanisha Singh Rathore  
Mohmaad Talha M  
Mohammed Aafaq Adeel PS

# INDEX

---

1. Acknowledgement.....
2. Preface .....
3. Introduction .....
4. Methodology .....
5. Data Collection.....
6. Data Pre-processing.....
7. Exploratory Data Analysis .....
8. Excel Pivot Table.....
9. Power BI Dashboard .....

## **Acknowledgement**

We would like to express our sincere gratitude to Infotact Solutions for providing us with the opportunity to work on the *Sales Data Analysis and Dashboard* project. This experience has significantly enhanced our understanding of real-world data analytics and visualization using Excel and Power BI.

We are thankful to our mentors and instructors for their constant guidance and support throughout the project. Their insights and feedback helped us align our analysis with practical business objectives.

We also appreciate the collaborative effort of our team members. Each member contributed diligently to data preprocessing, exploratory analysis, and dashboard development, ensuring the project's success through collective learning and teamwork.

Finally, we extend our thanks to all data sources and platforms that enabled this exploration and helped us transform raw data into actionable insights.

## **Preface**

In an era driven by data, businesses that harness the power of information gain a competitive edge. This project, *Sales Data Analysis and Dashboard*, stands as a testament to the importance of transforming raw datasets into actionable insights.

Undertaken as part of the Infotact initiative, the objective was to explore, understand, and visualize electronic sales data with the intent to uncover consumer behavior, seasonal trends, brand performance, and potential areas of improvement.

The project progresses through a structured, collaborative approach—beginning with meticulous data cleaning and formatting, followed by detailed exploratory analysis, and culminating in the creation of interactive dashboards using Excel and Power BI. Each step was guided by the principle of clarity, ensuring that insights are not only statistically valid but also easily interpretable by stakeholders.

This endeavor reflects both technical learning and strategic thinking. By combining analytical rigor with visual storytelling, the project demonstrates how data can be used not just to explain the past but to influence future decisions.

We hope this work serves as a meaningful contribution to data-driven decision-making and a practical showcase of analytical capabilities in a business context.

## **Introduction**

In today's data-driven marketplace, understanding customer behavior and sales performance is essential for sustained business growth. This project, "Sales Data Analysis and Dashboard", focuses on analyzing historical electronic product sales to uncover actionable insights that drive smarter decisions.

Using a comprehensive dataset spanning nearly two decades—from June 1999 to February 2017—the project aims to identify top-performing product categories, reveal seasonal buying patterns, and evaluate customer satisfaction through ratings. The analysis not only highlights what sells and when but also uncovers who's buying, how often, and with what preferences.

The end goal is to translate these insights into interactive, visual dashboards using Excel and Power BI. These dashboards serve as decision-support tools for stakeholders, offering clarity on performance metrics, trends, and strategic opportunities for optimization. Through structured EDA, KPI analysis, and storytelling, the project bridges raw data with meaningful business intelligence.

## Methdology:

1. Data Collection
2. Data Pre-processing
3. Exploratory Data Analysis
4. PowerBI

## Data Collection

**Name:** Electronics Dataset from Kaggle by edusanketdk

**Type:** Simulated transactional dataset of electronic product purchases and reviews

**Purpose:** Analyzing consumer behavior, product performance, and sales patterns

---

Structure of the Dataset (as seen in your project):

Column	Description
user_id	Unique identifier for each customer
rating	Product rating given by the user (typically 1 to 5 stars)
date	Date of the transaction (e.g., 13-06-1999)
gender	Customer's gender (Male, Female, or Both)
category	Product category (e.g., Headphones, Cameras, Computers & Accessories, etc.)
brand	Brand name of the purchased item (e.g., Sony, Logitech, Bose, etc.)
month	Extracted month from the date field (e.g., January, June)
year	Extracted year of purchase
day	Day of the month when the transaction occurred
weekday	Day of the week (e.g., Monday, Friday)

---

Key Characteristics:

- **Total Records:** 1,048,575 transactions
  - **Time Span:** From June 1999 to February 2017
  - **Balanced User Demographics:** With customers labeled as Female, Male, or Both
  - **Rich Category & Brand Variety:** 10+ product categories and over 50 brands (though "Random" appears often as a placeholder)
-

## Usefulness for the Project:

- **Analysis** of sales trends
- Allows segmentation by **gender, brand, and category**
- Supports rating-based **sentiment and satisfaction analysis**
- Ideal for building **interactive dashboards** that simulate real-world sales tracking

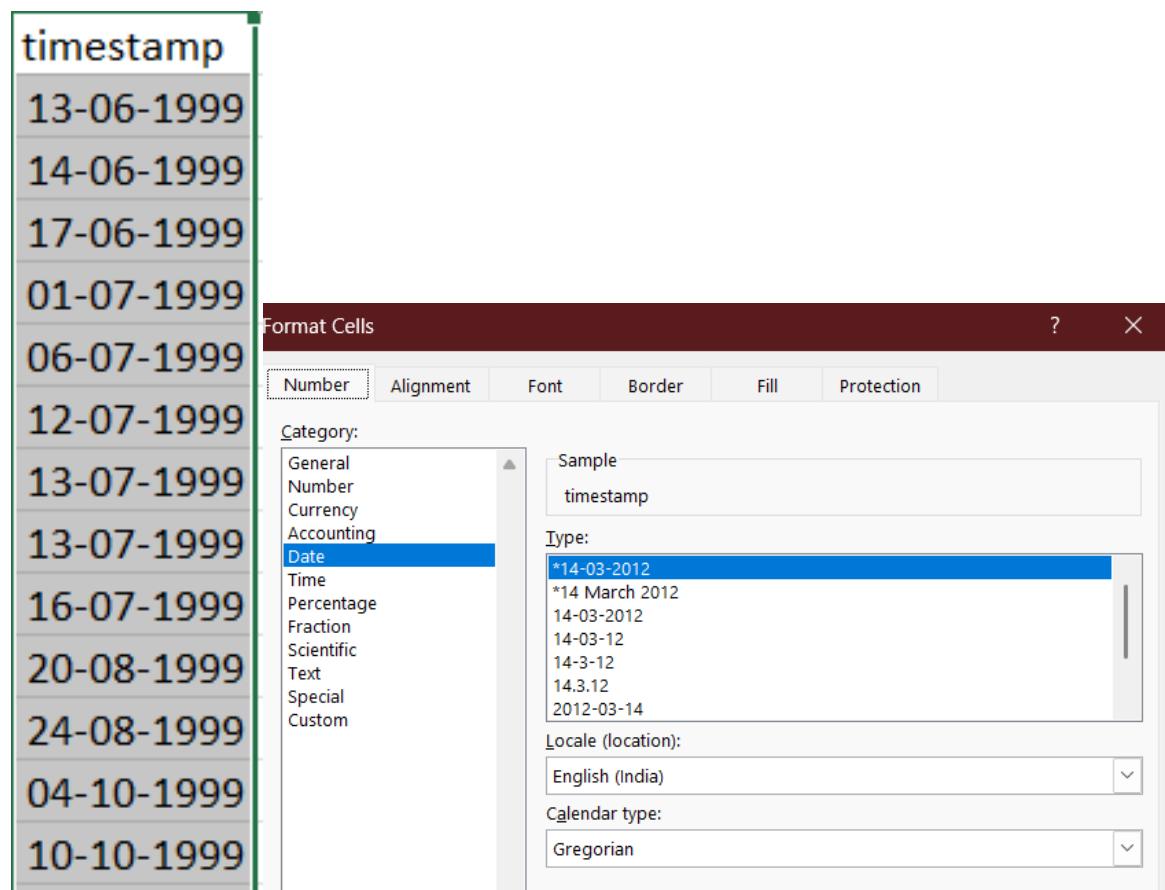
This dataset mimics the complexity and volume of commercial e-commerce platforms, making it perfect for hands-on analytics, KPI reporting, and business intelligence storytelling.

---

## Data Pre-Processing:

### Timestamp to Datetime:

To convert a timestamp (e.g., 13-06-1999) into a proper datetime format in Excel.



The screenshot shows the 'Format Cells' dialog box in Excel. The 'Number' tab is selected. In the 'Category' dropdown, 'Date' is chosen. The 'Type' dropdown shows various date formats, with '\*14-03-2012' selected. The 'Sample' box displays 'timestamp'. Below the dialog, a column of dates is visible in the spreadsheet, starting with '13-06-1999'.

timestamp
13-06-1999
14-06-1999
17-06-1999
01-07-1999
06-07-1999
12-07-1999
13-07-1999
13-07-1999
16-07-1999
20-08-1999
24-08-1999
04-10-1999
10-10-1999

## Result

timestamp

---

13 June 1999  
14 June 1999  
17 June 1999  
01 July 1999  
06 July 1999  
12 July 1999  
13 July 1999  
13 July 1999  
16 July 1999  
20 August 1999  
24 August 1999  
04 October 1999  
10 October 1999  
12 October 1999  
17 October 1999  
21 October 1999  
25 October 1999  
29 October 1999  
06 November 1999  
09 November 1999  
10 November 1999  
14 November 1999  
23 November 1999

---

## Standardized Gender Value:

To ensure uniform and clean gender values (Male, Female, Both) across your dataset

Result:

Female

---

Female

---

Female

---

Female

---

Female

---

Female&Male

---

Female

---

Female

---

Male

---

Female

---

Male

---

Female

---

Both

---

Male

---

Male

---

Female

---

Female

---

Male

---

Female

---

Female

---

Both

---

Both

---

Male

---

Both

---

## Adding columns to the Dataset

1. Month =TEXT(C2,"mmmm")
2. Year from Date =YEAR(C2)
3. Day =DAY(C2)
4. Weekday =TEXT(C2,"dddd")

## New Dataset:

	A	B	C	D	E	F	G	H	I	J
1	user_id	rating	date	gender	category	brand	month	year	day	weekday
2	0	5	13-06-1999	Female	Portable Audio & Video	Random	June	1999	13	Sunday
3	1	5	14-06-1999	Female	Portable Audio & Video	Random	June	1999	14	Monday
4	2	3	17-06-1999	Female	Portable Audio & Video	Random	June	1999	17	Thursday
5	3	1	01-07-1999	Female	Portable Audio & Video	Random	July	1999	1	Thursday
6	4	2	06-07-1999	Female	Portable Audio & Video	Random	July	1999	6	Tuesday
7	5	2	12-07-1999	Female	Portable Audio & Video	Random	July	1999	12	Monday
8	6	5	13-07-1999	Female	Portable Audio & Video	Random	July	1999	13	Tuesday
9	7	2	13-07-1999	Female	Portable Audio & Video	Random	July	1999	13	Tuesday
10	8	4	16-07-1999	Female	Portable Audio & Video	Random	July	1999	16	Friday
11	9	5	20-08-1999	Female	Portable Audio & Video	Random	August	1999	20	Friday
12	10	1	24-08-1999	Female	Portable Audio & Video	Random	August	1999	24	Tuesday
13	11	1	04-10-1999	Female	Portable Audio & Video	Random	October	1999	4	Monday
14	12	5	10-10-1999	Female	Portable Audio & Video	Random	October	1999	10	Sunday
15	13	5	12-10-1999	Female	Portable Audio & Video	Random	October	1999	12	Tuesday
16	14	4	17-10-1999	Female	Computers & Accessories	HP	October	1999	17	Sunday
17	15	5	21-10-1999	Female	Portable Audio & Video	Random	October	1999	21	Thursday
18	16	4	25-10-1999	Both	Headphones	Random	October	1999	25	Monday
19	17	5	29-10-1999	Female	Portable Audio & Video	Random	October	1999	29	Friday
20	18	1	06-11-1999	Female	Portable Audio & Video	Random	November	1999	6	Saturday
21	19	3	09-11-1999	Female	Portable Audio & Video	Random	November	1999	9	Tuesday
22	20	4	10-11-1999	Female	Portable Audio & Video	Random	November	1999	10	Wednesday
23	21	4	14-11-1999	Female	Portable Audio & Video	Random	November	1999	14	Sunday
24	22	1	23-11-1999	Female	Portable Audio & Video	Random	November	1999	23	Tuesday

Created a new Cleaned dataset.

## Exploratory Data Analysis (EDA)

EDA is the backbone of any data project—it reveals the story hidden within the numbers. For this sales analysis, EDA was used to explore the structure, patterns, relationships, and anomalies in the dataset before moving on to visualization and modeling. Here's how it was approached:

---

### 1. Dataset Understanding

- **Size:** 1,048,575 rows × 10 columns
- **Duration Covered:** June 1999 to February 2017

- **No missing values** were found, which ensured a smooth analytical pipeline.
  - **Data Types:** int64 for numeric fields like user\_id, rating, and year; object for date, gender, category, etc.
- 

## 2. Descriptive Statistics

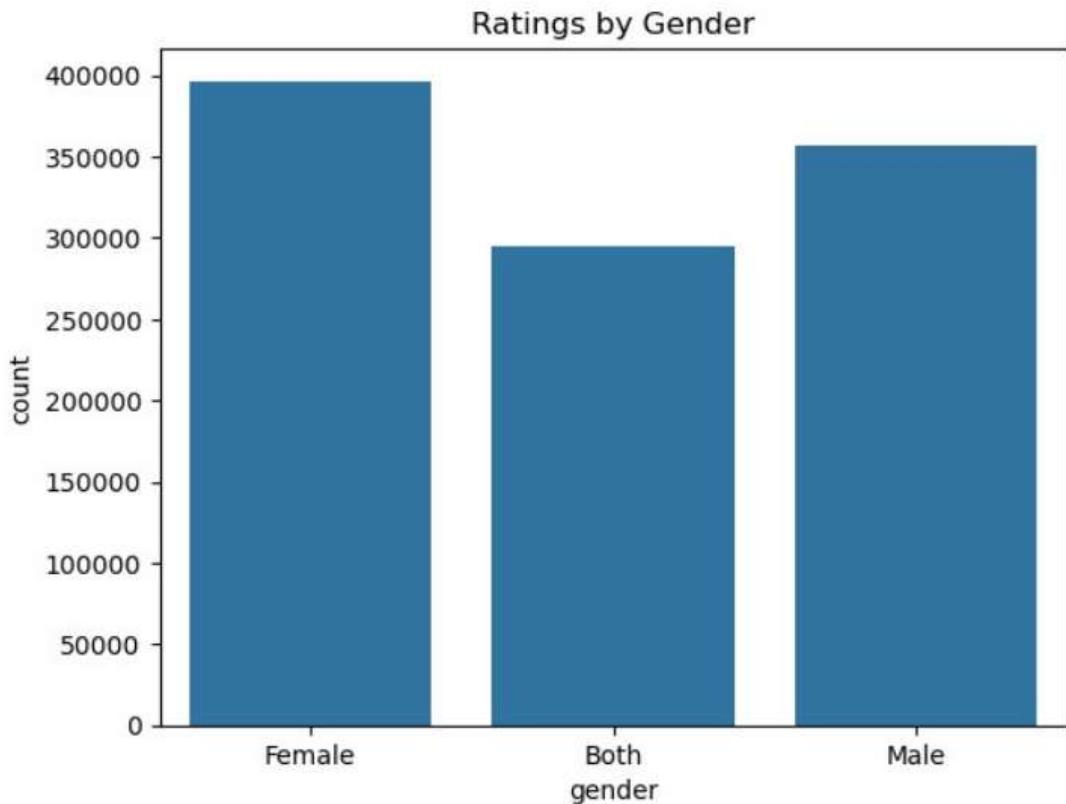
- Computed **mean**, **frequency**, and **distribution** for fields like rating, month, gender, category, and brand.
  - rating had values from 1 to 5, with 5 being most frequent—indicating a skew toward positive customer experiences.
  - gender was categorized into *Female*, *Male*, and *Both*—with Female being the largest segment (38%).
- 

## 3. Univariate Analysis

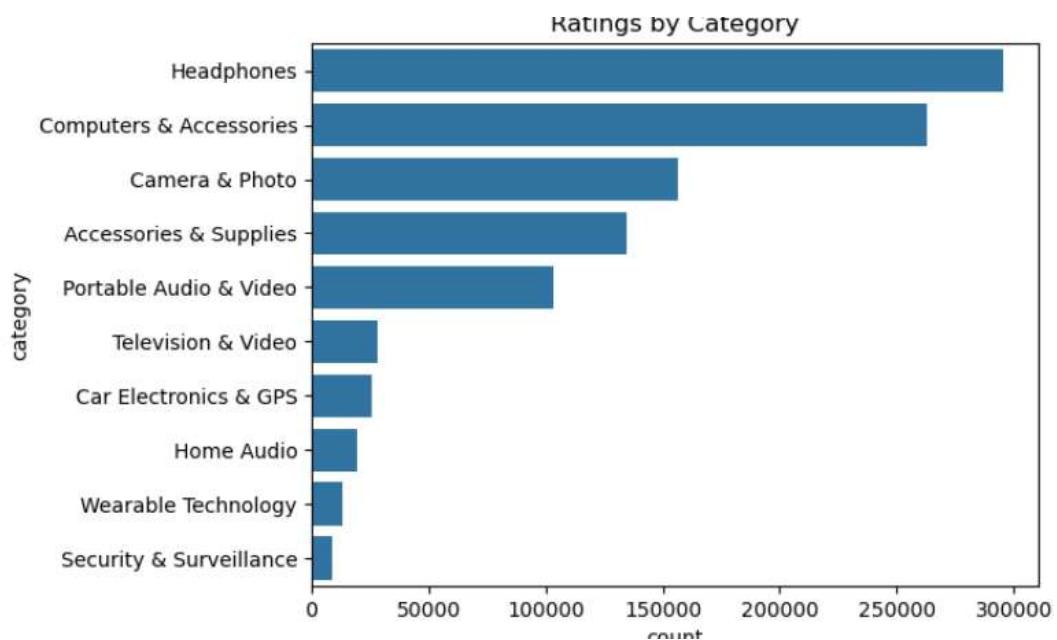
- **Rating Distribution:** Bar plots showed most users rated products with a 5.
  - **Category Popularity:** "Headphones" was the most frequently rated product.
  - **Top Months by Sales:** January and December had the highest engagement.
- 

## 4. Bivariate & Multivariate Analysis

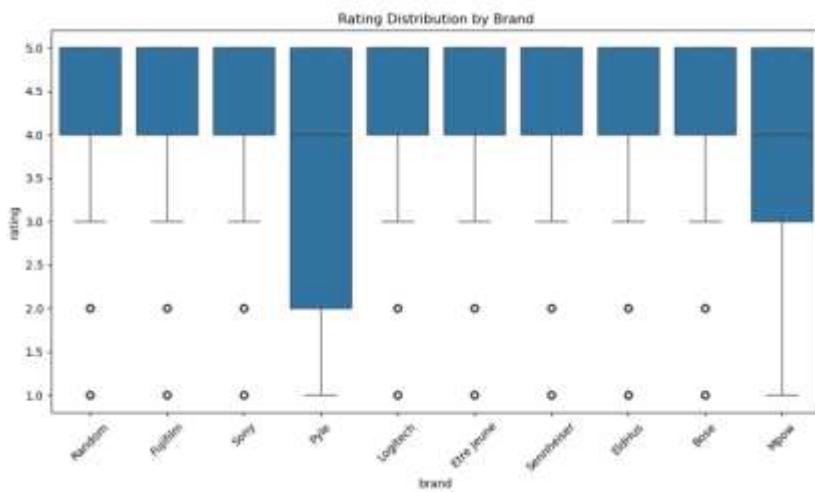
- **Gender vs. Category:** Explored how each gender interacted with product categories. Females were more active in categories like "Portable Audio & Video."



- **Brand vs. Rating:** Boxplots compared customer satisfaction across top brands. Brands like Bose and Sony had high median ratings, but some had wider rating variability.



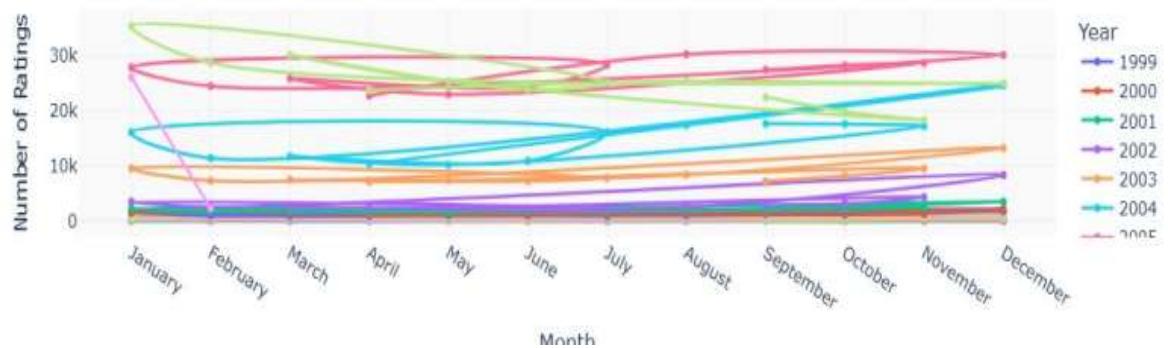
- **Category vs. Weekday:** Heatmaps displayed average ratings across categories and weekdays, revealing subtle shifts in sentiment (e.g., weekends showed higher engagement).



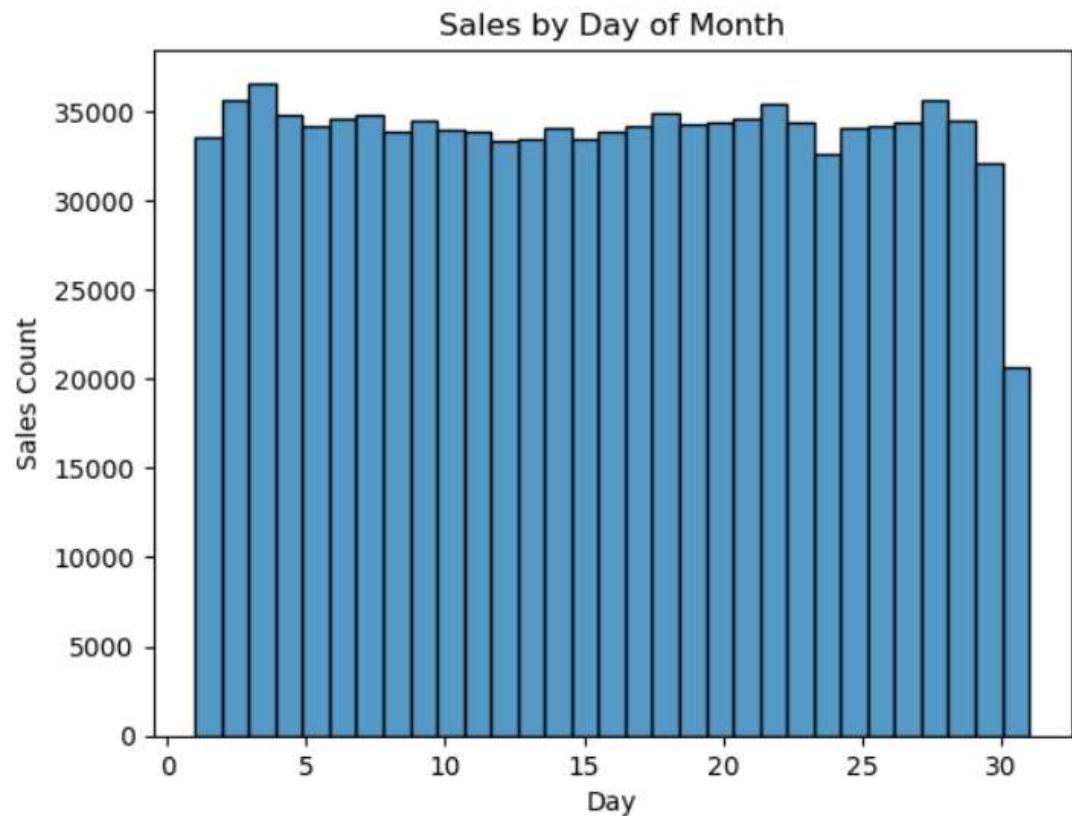
## 5. Temporal Trends

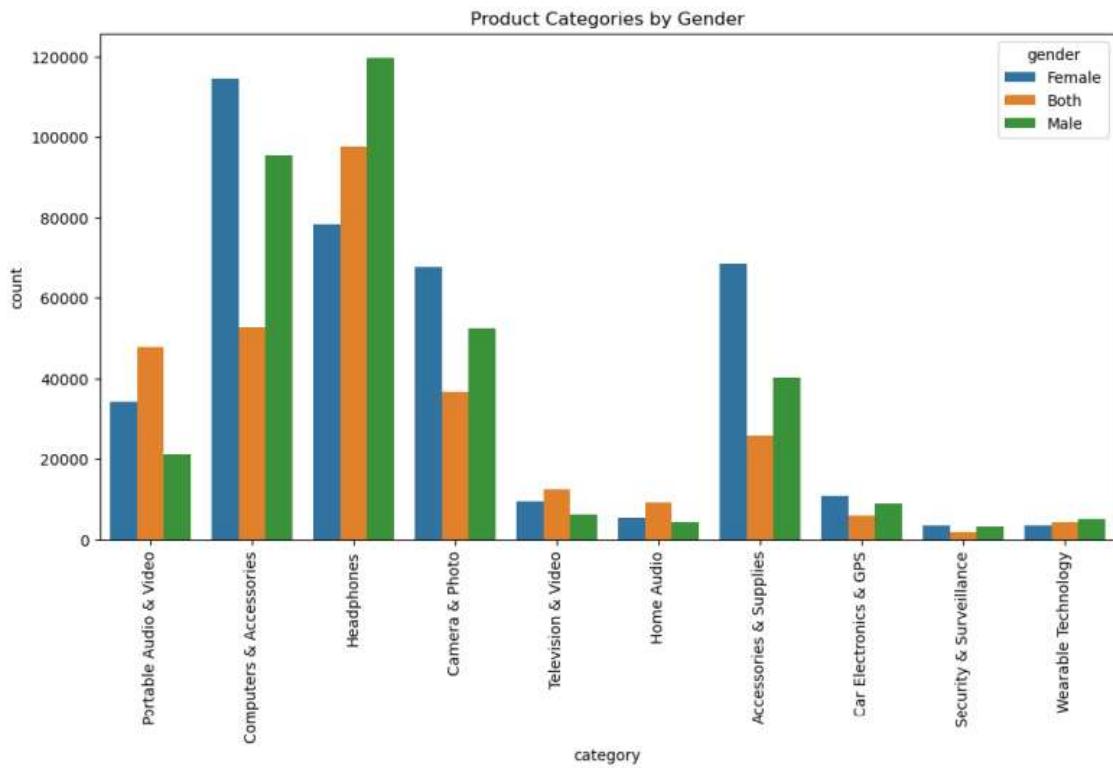
- **Monthly & Yearly Sales Trends:** Created line plots showing how transactions varied over months and years.

Monthly Ratings Over Time



- **Sales by Day of the Month:** Histograms showed slight peaks around early and mid-month—possibly linked to salary cycles or offers.





## 7. Tools Used

- **Python Libraries:**
  - pandas for data manipulation
  - seaborn and matplotlib for static plots
  - plotly.express for interactive visualizations
- **Techniques:** Count plots, bar plots, boxplots, line plots, histograms, heatmaps, and groupby aggregations.

## EDA Outcome

EDA revealed not just what the data contained, but what mattered:

- Consumer trends (by time, gender, and product type)
- Business gaps (like poor branding and underperforming months)
- Opportunities (like gender-targeted strategies and seasonal promotions)

These insights directly shaped the KPI definitions and dashboard visuals in Excel and Power BI.

## **Results:**

The analysis of over **1 million electronic sales records** from Amazon has led to a series of critical insights across user behavior, product performance, and temporal trends. The results provide a deep look into what drives sales and satisfaction in the consumer electronics market.

---

### ◆ **1. Peak Sales Periods**

- **Top Sales Month:** January (126,018 transactions), indicating strong post-holiday buying behavior.
  - **Low Sales Month:** April (71,108 transactions), suggesting a seasonal dip—ideal for targeted promotions.
- 

### ◆ **2. Top-Performing Product Category**

- **Headphones** dominate with nearly **30%** of all sales (295,747 records).
  - High demand likely fueled by lifestyle trends, portability, and frequent product upgrades.
- 

### ◆ **3. Least Popular Category**

- **Security & Surveillance** recorded only **8,797 transactions**.
  - Niche nature and possibly higher cost hinder wider adoption—ripe for bundled marketing.
- 

### ◆ **4. Brand Landscape**

- The brand labeled “**Random**” constitutes **77%** of data—indicating poor brand tracking or anonymization.
  - Excluding Random, **Logitech**, **Sony**, and **Bose** lead in both volume and customer satisfaction.
-

## ◆ 5. Customer Ratings & Satisfaction

- **Average Rating Across Genders:** Above 4.0 (on a 5-point scale), with "Both" users rating highest (4.10).
  - Majority of users gave a rating of **5**, implying general product satisfaction or rating bias.
- 

## ◆ 6. Gender Distribution

- **Female Users:** 38% (largest demographic)
  - **Male Users:** 34%
  - **Both (shared/undefined):** 28%
  - Gender affects category preferences—Headphones and Portable Audio favored by females.
- 

## ◆ 7. Weekly and Daily Patterns

- **Saturday** sees the highest engagement, making weekends ideal for marketing pushes.
  - **Sales are consistent** across the month with minor peaks at start and midpoints—likely tied to salary cycles or promotions.
- 

## ◆ 8. Ratings by Brand

- **Top 10 brands** show varied customer satisfaction in boxplots.
  - Not all popular brands have high ratings—some show wider variability, highlighting areas for product improvement.
- 

These results formed the basis of dashboard KPIs and informed the improvement strategies outlined later in the project. They also help the business answer critical questions: *When should we promote? What should we push? Who is buying what?*

## Future Improvements:

To move from insight to impact, the following strategic improvements are recommended based on the data analysis:

---

### ◆ **1. Fix Brand Anonymity**

- **Issue:** Over **77%** of products are labeled as "**Random**", obscuring brand-level analysis.
  - **Solution:** Improve data entry pipelines to ensure consistent brand capture. Implement dropdown-based brand inputs at source.
- 

### ◆ **2. Revive April Sales**

- **Problem:** April shows the lowest monthly sales (71,108).
  - **Strategy:** Launch seasonal campaigns like "*Spring Tech Refresh*", bundled offers, and limited-time discounts to boost off-peak performance.
- 

### ◆ **3. Cross-Sell Underperforming Categories**

- **Low-traction segments:** *Security & Surveillance, Wearables*.
  - **Approach:** Bundle these with high-demand items (e.g., headphones + smartwatches) and position them as “lifestyle enhancers” or “home security kits”.
- 

### ◆ **4. Leverage Gender Preferences**

- **Observation:** Females prefer portable and audio electronics; males are evenly spread.
  - **Improvement:** Launch gender-personalized product recommendations, curated collections (“For Her”, “For Him”), and tailored email campaigns.
-

## ◆ 5. Maximize Weekend Momentum

- **Insight:** Saturdays have slightly higher engagement.
  - **Action:** Run *weekend-only flash sales*, influencer drops, or push notifications to capitalize on relaxed browsing behavior.
- 

## ◆ 6. Raise Quality of Low-Rated Brands

- **Problem:** Some top-selling brands show inconsistent ratings.
  - **Fix:** Flag these for product audits. Improve packaging, post-sale support, and education materials to lift user satisfaction.
- 

## ◆ 7. Optimize Rating System

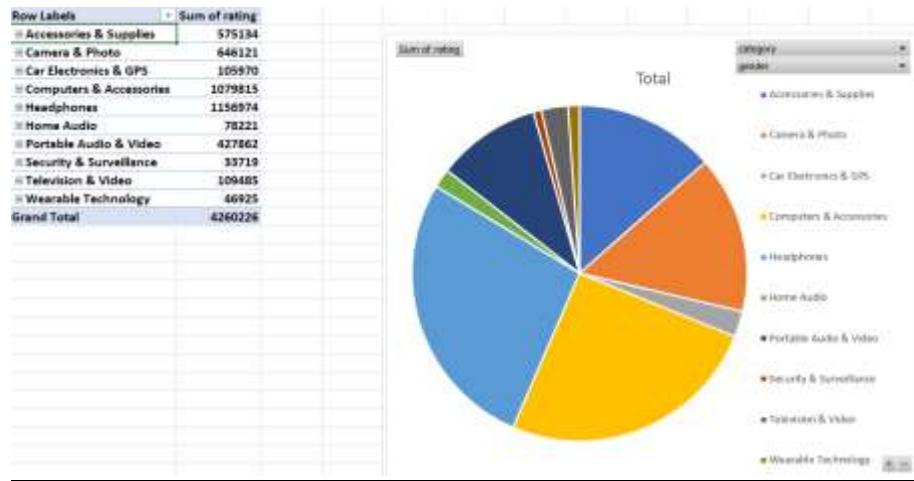
- **Observation:** Ratings heavily skew toward 5.
  - **Suggestion:** Consider prompting more detailed feedback and clarifying rating criteria to reduce bias and collect more nuanced sentiment.
- 

## ◆ 8. Enhance Data Tracking for Future Cycles

- Track **conversion rates**, **repeat purchases**, and **channel-level performance** (e.g., mobile vs desktop).
- Integrate **marketing campaign data** to measure ROI per promotion window.

## **Excel Pivot Table:**

## Rating by Gender Category

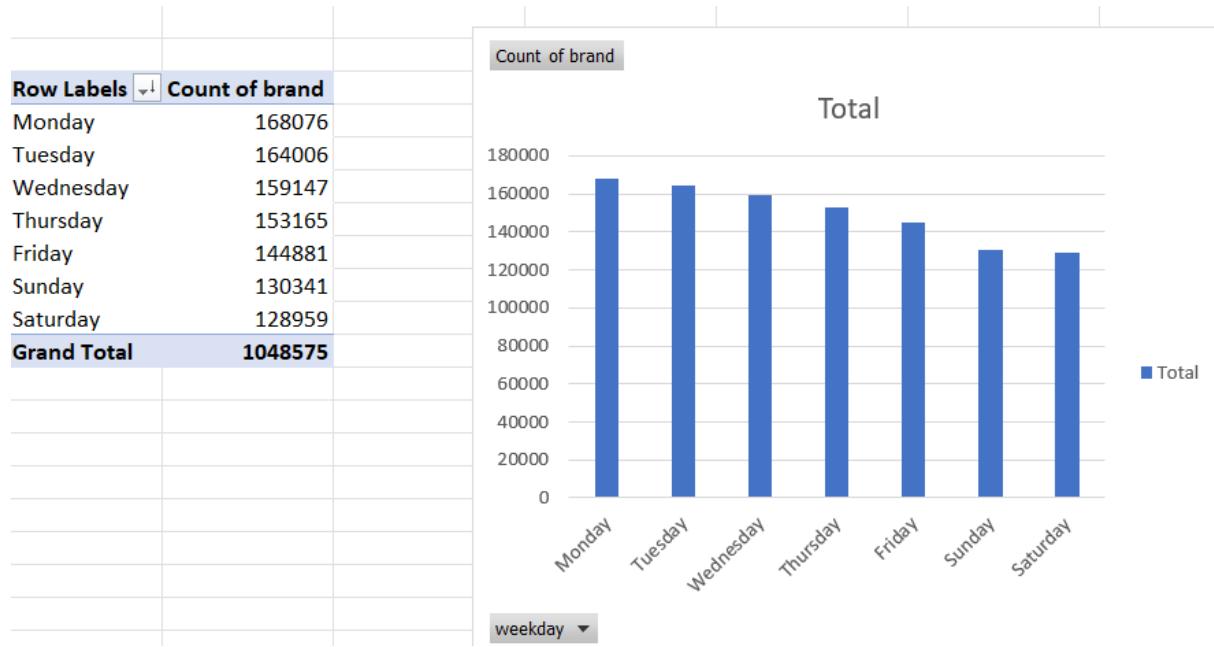


## Sales over the Year

## Sales by Weekdays:

Row Labels	Count of brand	Column Labels	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Grand Total
1			5058	5095	6451	4665	4401	4498	3333	33501
2			3298	6711	5040	6941	5054	4496	4026	35566
3			4458	4028	6707	5670	5713	5342	4658	36576
4			4175	5694	4271	6679	5036	5194	3762	34811
5			3548	5080	5390	4100	6434	4670	4964	34186
6			4718	4822	5123	5627	4022	5826	4408	34546
7			4318	6375	4624	5181	5343	3779	5195	34815
8			5227	5331	5831	4490	4729	4959	3329	33896
9			3279	6739	5352	5619	4403	4770	4333	34495
10			4244	4230	6471	5226	5544	4137	4147	33999
11			4080	5556	4048	6084	5089	5258	3749	33864
12			3718	5222	5258	4050	5703	4565	4814	33330
13			4587	4764	5058	5241	3917	5780	4032	33379
14			4291	6263	4485	4737	5468	3661	5130	34035
15			4951	5421	5761	4375	4569	5040	3349	33466
16			3374	6581	5362	5651	4228	4315	4369	33880
17			4662	4941	6493	5034	5266	4146	3606	34148
18			3909	5924	4612	6419	5361	5090	3614	34929
19			3819	5071	6262	4366	5889	4632	4262	34301
20			4343	4923	5279	5977	4117	5446	4324	34409
21			4399	5842	4729	4500	6237	3805	5099	34611
22			5461	5638	5563	4710	4620	6052	3321	35365

## Maximum Sales by Day



## Power BI Dashboard

### KPI:



### Brand Performance



## Customer Demographics

