

# Analysis of Amazon sales dataset

---

## Introduction

The rise of e-commerce giants like Amazon has revolutionized retail, offering vast product selections and convenience to consumers worldwide. This project analyzes a dataset of Amazon sales to uncover insights into consumer behavior, product performance, and market trends. Through data-driven analysis, we aim to identify key patterns that can inform strategic decisions in online retail, such as optimizing product offerings and enhancing customer experience.

## Problem Statements

- 1. Count of Orders per Month:**
  - The analysis aims to determine the monthly order volume on Amazon. By aggregating and visualizing the count of orders placed each month, we seek to identify seasonal trends and peak order periods throughout the year.
- 2. Sum of Sales per Category:**
  - This analysis focuses on calculating the total sales revenue generated by each product category available on Amazon. By aggregating sales data across categories, we aim to identify the top-performing product categories in terms of revenue contribution.
- 3. Count of Orders per State in the United States:**
  - Understanding geographic patterns of Amazon orders within the United States is crucial. By counting orders per state, we aim to identify regions with high customer activity and potentially uncover regional preferences or market dynamics.
- 4. Sum of Profit per Year:**
  - Analyzing the profitability of Amazon over the years is essential for assessing financial performance. This analysis will focus on calculating and comparing annual profits to understand trends and fluctuations in Amazon's financial health.

## Dataset descriptions:

This Data is a Amazon Product Sales. This Dataset about Amazon Sales Contain **3204** Rows and **9** Columns.

### *Column Description*

**Order Date** – Order Date.

**Ship Date** - Shipping Date.

**Email ID** – Email ID of Users

**Geography** - Location of Orders by Users.

**Category** - Product Category

**Product Name** - Product Name of Amazon

**Sales** - Amazon Product Sales

**Quantity** - how many units of a particular product are available.

**Profit** - Amazon Sales Profit

## Data Cleaning and Processing

The initial phase of this project involved preparing the Amazon sales dataset for analysis. This process included several steps to ensure data accuracy and relevance:

**1. Column Selection and Removal:**

- The dataset was initially reviewed to identify and remove unnecessary columns that do not contribute to the analysis objectives. This step helps streamline the dataset and focus on relevant data points.

**2. Removing Duplicates:**

- The 'Remove Duplicates' function in Excel was utilized to eliminate any duplicate entries within the dataset. This ensures that each transaction or record is unique and prevents data skewing in subsequent analyses.

**3. Handling Missing Values:**

- Data was filtered and reviewed to identify any missing values across the dataset. Using Excel's filtering capabilities, missing values were visually inspected and appropriate actions were taken, such as imputation or exclusion, depending on the analysis requirements.

**4. Extracting Year and Month from Order Date:**

- The Year() and Month( ) functions in Excel were employed to extract the year and month information, respectively, from the 'Order Date' column. This transformation enabled temporal analysis of sales trends by year and month.

**5. Separating State from Geography Column:**

- To isolate the state information from the 'Geography' column, Excel functions such as Substitute(), Find(), Len(), and Right() were utilized. These functions helped extract the state information into a separate column, facilitating geographical analysis of Amazon orders within the United States.

**6. Data Verification and Overview:**

- Throughout the cleaning and processing stages, an overview of the dataset was continuously maintained to verify the accuracy and completeness of transformations. This step ensures that the dataset is prepared for subsequent analytical tasks without discrepancies.

M2										
	A	B	C	D	E	F	G	H	I	J
1	Order ID	order year	order month	Order Date	Geography	state	Category	Sales	Quantity	Profit
2	CA-2013-138688	2013	6	6/13/2013	United States, Los Angeles, California	California	Labels	14.62	2.00	6.87
3	CA-2011-115812	2011	6	6/9/2011	United States, Los Angeles, California	California	Furnishings	48.86	7.00	14.17
4	CA-2011-115812	2011	6	6/9/2011	United States, Los Angeles, California	California	Art	7.28	4.00	1.97
5	CA-2011-115812	2011	6	6/9/2011	United States, Los Angeles, California	California	Phones	907.15	4.00	90.72
6	CA-2011-115812	2011	6	6/9/2011	United States, Los Angeles, California	California	Binders	18.50	3.00	5.78
7	CA-2011-115812	2011	6	6/9/2011	United States, Los Angeles, California	California	Appliances	114.90	5.00	34.47
8	CA-2011-115812	2011	6	6/9/2011	United States, Los Angeles, California	California	Tables	1706.18	9.00	85.31
9	CA-2011-115812	2011	6	6/9/2011	United States, Seattle, Washington	California	Phones	911.42	4.00	68.36
10	CA-2013-161389	2013	12	12/6/2013	United States, Seattle, Washington	Washington	Binders	407.98	3.00	132.59
11	CA-2011-167164	2011	5	5/13/2011	United States, West Jordan, Utah	Utah	Storage	55.50	2.00	9.99
12	CA-2011-143336	2011	8	8/27/2011	United States, San Francisco, California	California	Art	8.56	2.00	2.48
13	CA-2011-143336	2011	8	8/27/2011	United States, San Francisco, California	California	Phones	213.48	3.00	16.01
14	CA-2011-143336	2011	8	8/27/2011	United States, San Francisco, California	California	Binders	22.72	4.00	7.38
15	CA-2012-106320	2012	9	9/25/2012	United States, Orem, Utah	Utah	Tables	1044.63	3.00	240.26
16	CA-2013-121755	2013	1	1/16/2013	United States, Los Angeles, California	California	Binders	11.65	2.00	4.22
17	CA-2013-121755	2013	1	1/16/2013	United States, Los Angeles, California	California	Accessories	90.57	3.00	11.77
18	CA-2013-101343	2013	7	7/18/2013	United States, Los Angeles, California	California	Storage	77.88	2.00	3.89
19	CA-2012-135545	2012	11	11/24/2012	United States, Los Angeles, California	California	Accessories	13.98	2.00	6.15
20	CA-2012-135545	2012	11	11/24/2012	United States, Los Angeles, California	California	Binders	25.82	6.00	9.36

## Analysis and Modeling

After completing the data cleaning process, the Amazon sales dataset was analyzed using pivot tables in Excel to extract key insights and trends. The following sections outline the findings and methodologies used for each analysis objective:

### 1. Number of Orders per Month

To analyze the monthly order volume on Amazon, a pivot table was constructed using the following steps:

- **Pivot Table Setup:** Created a pivot table with 'Order Month' as rows and 'Order ID' (counted) as values.
- **Insights:** Identified peak months for orders to understand seasonal trends and customer buying patterns.

### 2. Sum of Sales per Category

For understanding sales performance across product categories, the analysis proceeded as follows:

- **Pivot Table Configuration:** Constructed a pivot table with 'Category' as rows and 'Sales Amount' (summed) as values.
- **Key Findings:** Identified top-performing categories in terms of revenue generation, guiding marketing and inventory strategies.

### 3. Number of Orders per State in the United States

To analyze regional sales distribution within the United States:

- **Pivot Table Construction:** Created a pivot table with 'State' as rows and 'Order ID' (counted) as values.
- **Insights:** Highlighted states with the highest order volumes, informing localized marketing efforts and logistics planning.

### 4. Sum of Profit per Year

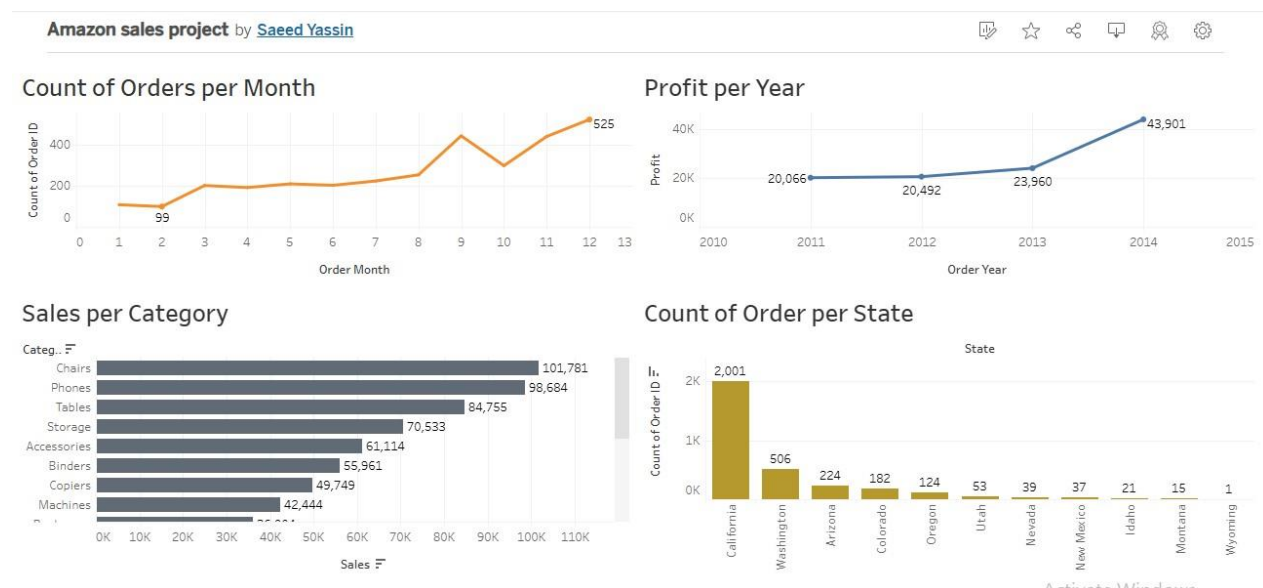
Analyzing Amazon's profitability over time involved the following steps:

- **Yearly Aggregation:** Created a pivot table with 'Order Year' as rows and 'Profit' (summed) as values.
- **Strategic Insights:** Evaluated profit trends over years to identify growth periods and areas for cost optimization.

	A	B	C	D	E	F	G	H	I	J	K
1											
2	Row Labels	Count of Order ID per month		Row Labels	Sum of Sales		Row Labels	Count of Order ID per state		Row Labels	Sum of Profit per year
3	1	108		Accessories	61114.116		Arizona	224		2011	20065.6912
4	2	99		Appliances	30236.336		California	2001		2012	20492.1947
5	3	202		Art	9212.066		Colorado	182		2013	23959.9374
6	4	192		Binders	55961.113		Idaho	21		2014	43900.6256
7	5	210		Bookcases	36004.1235		Montana	15		Grand Total	108418.4489
8	6	203		Chairs	101781.328		Nevada	39			
9	7	224		Copiers	49749.242		New Mexico	37			
10	8	254		Envelopes	4118.1		Oregon	124			
11	9	445		Fasteners	923.216		Utah	53			
12	10	299		Furnishings	30072.73		Washington	506			
13	11	442		Labels	5078.726		Wyoming	1			
14	12	525		Machines	42444.122		Grand Total	3203			
15	Grand Total	3203		Paper	26663.718						
16				Phones	98684.352						
17				Storage	70532.852						
18				Supplies	18127.122						
19				Tables	84754.562						
20				Grand Total	725457.8245						

## Visualizations with Tableau:

Tableau visualizations complemented the pivot table analyses by providing dynamic and interactive representations of key metrics and trends in the Amazon sales dataset. These visual insights are instrumental in understanding customer behavior, optimizing operational efficiency, and driving strategic business decisions. The combination of Excel for initial data preparation and Tableau for visualization ensures a comprehensive approach to data-driven analysis, enabling actionable insights for sustainable business growth.



## Result:

### 1. Seasonal Analysis of Orders

The analysis reveals significant variations in order volumes across different months. Specifically, months such as December, November, October, and September consistently exhibit higher order frequencies compared to months like January, February, March, and April. This seasonal trend suggests that certain times of the year witness heightened consumer activity and purchasing behavior on Amazon.

### 2. Top Performing Product Categories

Several product categories stand out in terms of total sales volume. Categories including chairs, phones, tables, storages, and accessories demonstrate notably high sums of sales throughout the dataset period. These categories appear to attract substantial consumer interest and contribute significantly to the overall sales figures observed in the dataset.

### **3. Regional Analysis of Orders**

Geographically, specific states such as California, Washington, Arizona, and Colorado emerge as key contributors to the total number of orders placed. These states consistently record a high volume of transactions, indicating a strong market presence and consumer engagement with Amazon's offerings in these regions.

### **4. Profit Growth Over Time**

Analysis of profit trends over the years from 2011 to 2014 reveals a positive trajectory in total profit. The dataset shows a noticeable growth in the sum of profit generated during this period, underscoring Amazon's increasing profitability over the analyzed years.

## **Conclusion**

In summary, the analysis highlights seasonal patterns in order volumes, identifies top-performing product categories, pinpoints key states with high order frequencies, and demonstrates a positive trend in profit growth over the studied period. These insights provide valuable understanding of consumer behavior, market dynamics, and financial performance within the Amazon sales dataset.

## **Appendix**

the link for the dataset: <https://www.kaggle.com/datasets/anandshaw2001/amazon-sales-dataset>

the link for the tableau visualization:

[https://public.tableau.com/app/profile/saeed.yassin/viz/Amazonsalesproject\\_17199461819870/Dashboard1](https://public.tableau.com/app/profile/saeed.yassin/viz/Amazonsalesproject_17199461819870/Dashboard1)

**Data Analyst: Saeed Yassin**

**Phone: +96899379304**

**Address: Muscat – Oman**

