

CRACKING THE CODE OF AMAZON SALES: A DATA-DRIVEN EXPLORATION

*A Deep Data Dive into Amazon Sales:
Unlocking Insights for Strategic Growth*



INTRODUCTION

Dataset Overview:

- ❖ **Product Information:** Categories, prices, Shipment details.
- ❖ **Sales Data:** Total sales, quantity sold, transaction details.
- ❖ **Customer Insights:** Region, demographics.
- ❖ **Temporal Data:** Sales trends over time (daily, monthly)

Reasons for Selection:

- ❖ **Diverse Data:** Combines numerical and categorical features for comprehensive analysis.
- ❖ **Business Relevance:** Helps optimize inventory, marketing, and customer strategies.
- ❖ **Exploratory Potential:** Offers insights into product performance, regional and sales trends.

DATA PREPROCESSING AND CLEANING

- ❖ **Handling Missing Data:** Example: Dropped unnecessary columns like 'New', 'PendingS', 'fulfilled-by', dropped rows with missing Currency and Amount etc.
- ❖ **Data Cleaning:** Converted Date to datetime format, converted postal code from float column to an integer column handled duplicate records, renamed Column name .
- ❖ **Feature Engineering :** Extracted **Month** from Date to analyze time-based patterns.

```
# dropping the null columns and columns which are superfluous  
df.drop(['New', 'PendingS', 'fulfilled-by'], axis=1, inplace=True)
```

```
df.dropna(axis=0, inplace=True)
```

```
# To handle the errors in the ship-state by using replace function  
df.replace({'NL': 'NAGALAND', 'PB': 'PUNJAB', 'AR': 'ARUNACHAL PRADESH',
```

```
# Changing the data type of postal code  
df['ship-postal-code'] = df['ship-postal-code'].astype('int64')
```

```
df['Date'] = pd.to_datetime(df['Date'])
```

```
df['Month Name'] = df['Date'].dt.strftime('%B')  
df['Month Name']
```

EXPLORATORY DATA ANALYSIS (EDA)

❖ Summary Statistics:

- ❖ Displaying basic statistics (mean, median, min, max, standard deviation) for numerical columns like Price, Quantity, Total Sales.

❖ Group-by and Aggregation:

- ❖ Grouped data by categories (e.g., product category, region) to identify sales trends and patterns.
- ❖ Data visualizations:
- ❖ Used **Matplotlib** and **Seaborn** to visualize sales trends by product category, region, and monthly order count, highlighting key patterns and seasonal variations.

```
df.describe()
```

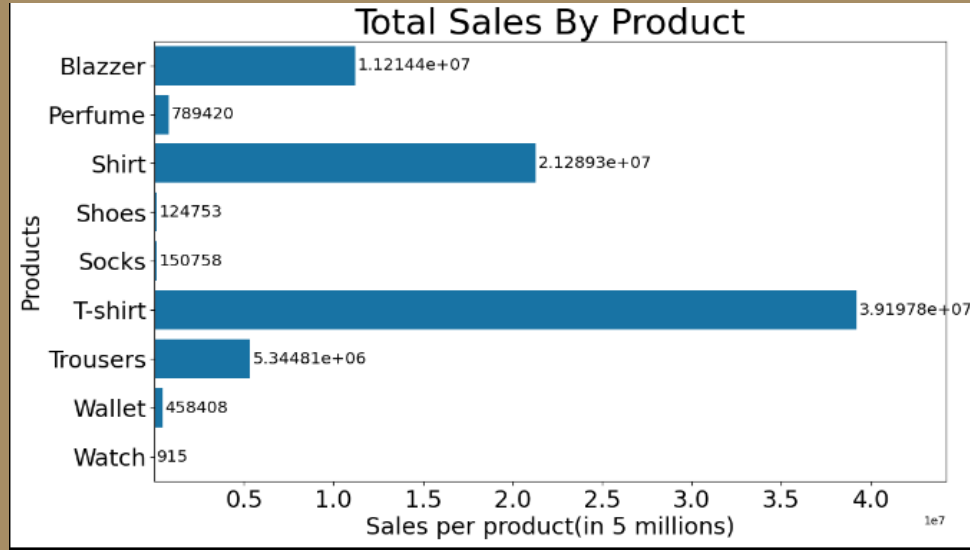
✓ 0.1s

	index	Qty	Amount	ship-postal-code
count	121143.000000	121143.000000	121143.000000	121143.000000
mean	64486.312655	0.961252	648.576874	463623.724507
std	37220.415404	0.214276	281.196896	191301.588170
min	0.000000	0.000000	0.000000	110001.000000
25%	32294.500000	1.000000	449.000000	382421.000000
50%	64477.000000	1.000000	605.000000	500032.000000
75%	96682.500000	1.000000	788.000000	600020.000000
max	128974.000000	8.000000	5584.000000	989898.000000

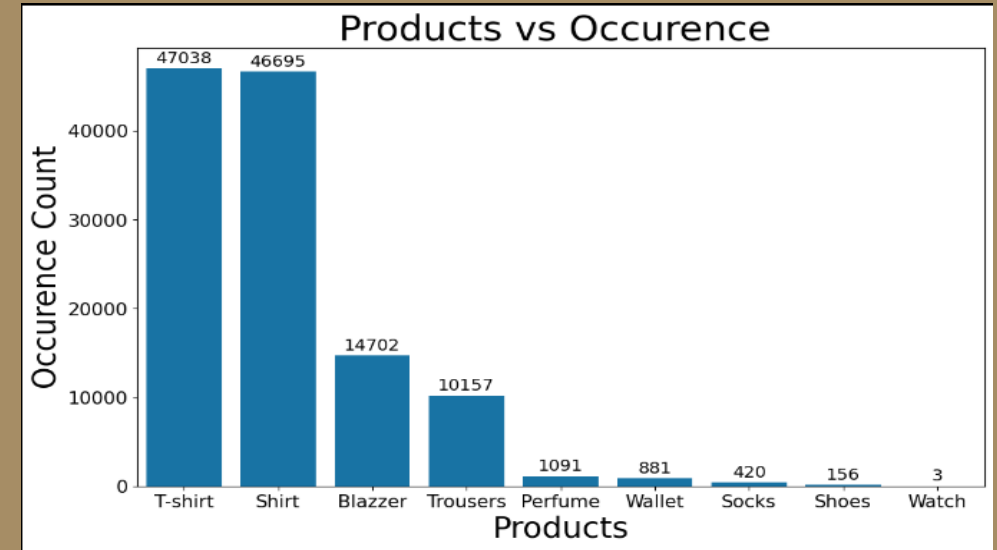
```
Sales_per_product=df.groupby(['Category'])['Amount'].sum()  
Sales_per_product
```

```
Courier_status= df['Courier Status'].value_counts()
```

SALES TRENDS AND PRODUCT PERFORMANCE ON AMAZON

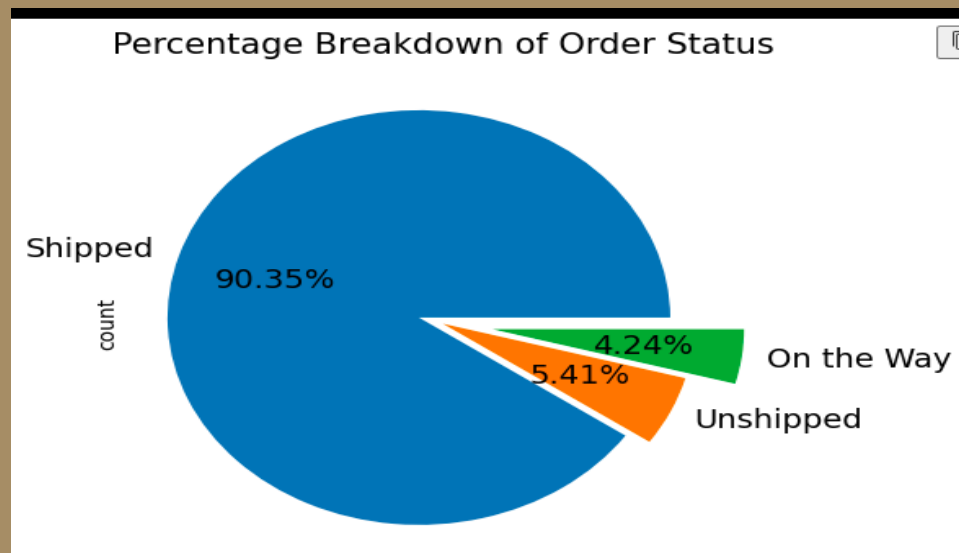


- ❖ The top-selling product on Amazon was **T-shirts**, with ₹40 million in sales, followed by **Shirts**, which generated ₹20 million in sales. On the other hand, **Watches** and **Shoes** had the lowest market share, with significantly lower sales.

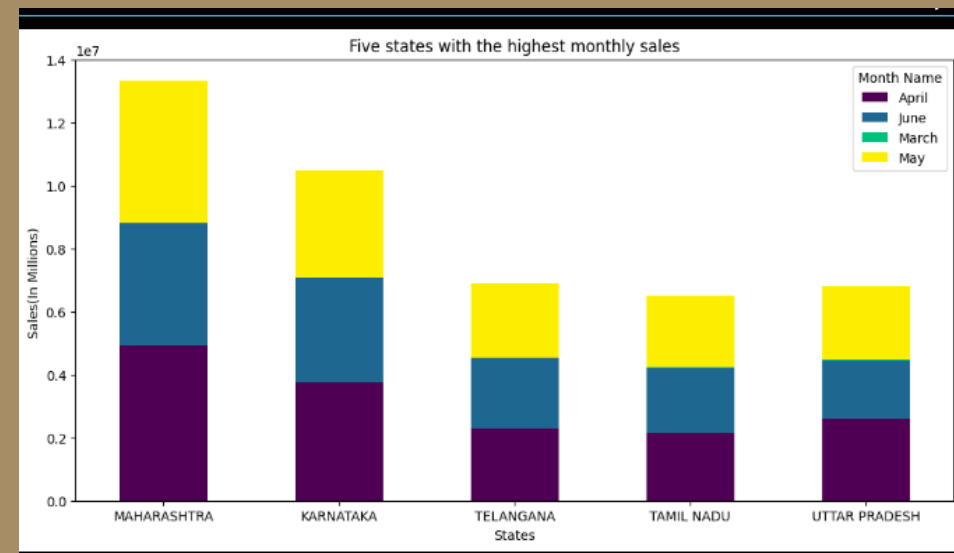


- ❖ A total of **47,000 units** of **T-shirts** were sold, making it the top-selling product, closely followed by **46,000 units** of **Shirts**. On the other hand, **Shoes** and **Watches** ranked at the bottom of the list with significantly lower sales figures.

ORDER STATUS BREAKDOWN AND TOP 5 STATES BY MONTHLY SALES

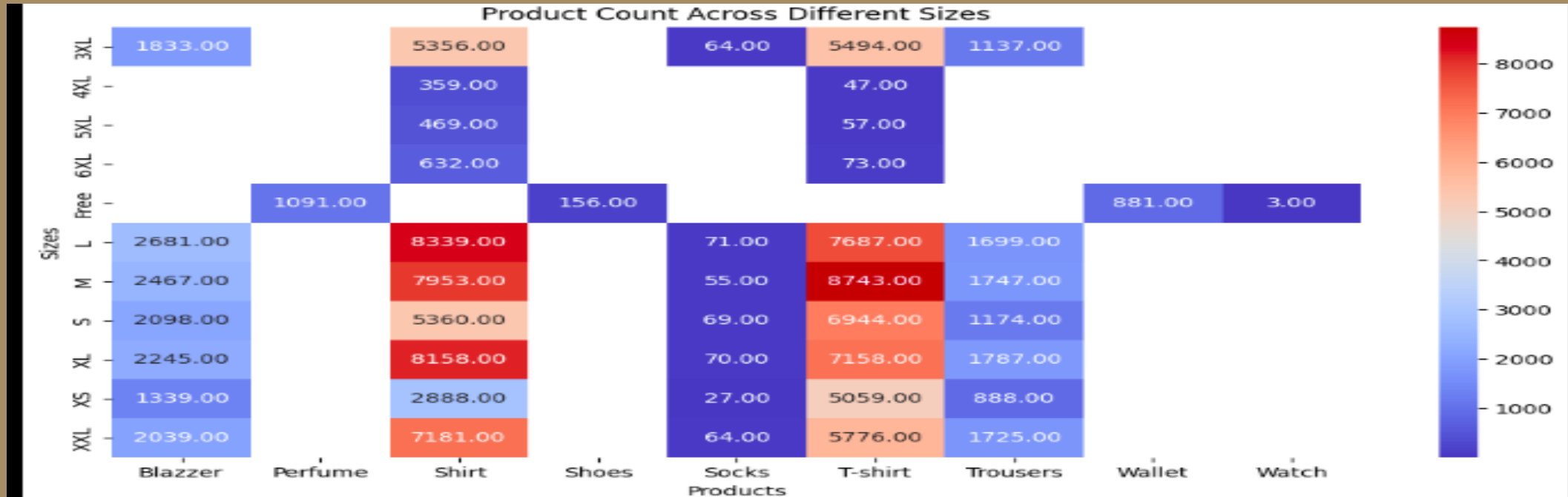


- ❖ Nearly **90%** of the orders were successfully shipped, while approximately **6%** remained unshipped. Additionally, about **4%** of the orders were still in transit to the buyers.



- ❖ The stacked bar graph illustrates the top 5 states with the highest sales over a span of 3 months, with **Maharashtra** leading the chart, followed by **Karnataka**.

EXPLORING SALES TRENDS ACROSS PRODUCT SIZES



- ❖ The heatmap illustrates the sales of products in different sizes, offering a quick overview of which sizes performed best. It provides valuable insights for Amazon to focus on the most popular product sizes based on sales

Conclusion

- ❖ The analysis reveals key insights into Amazon's sales:
- ❖ **Top-Selling Products:** **T-shirts** and **Shirts** lead in sales, while **Shoes** and **Watches** have lower demand.
- ❖ **Order Fulfilment:** **90%** of orders were shipped successfully, but **6%** remain unshipped and **4%** are still in transit.
- ❖ **Regional Insights:** **Maharashtra** and **Karnataka** are the top-performing states.
- ❖ **Product Size Preferences:** The heatmap reveals popular sizes, offering guidance on inventory focus.

Recommendations

- ❖ **Focus on Top Categories:** Prioritize **T-shirts** and **Shirts**, and boost sales strategies for **Shoes** and **Watches**.
- ❖ **Enhance Fulfilment:** Address **6% unshipped** and **4% in transit** orders to improve delivery speed.
- ❖ **Target Key Regions:** Focus marketing and inventory on **Maharashtra** and **Karnataka**.
- ❖ **Optimize Size Inventory:** Stock popular sizes more, adjust for less-demanded ones.
- ❖ **Prepare for Peaks:** Plan inventory and promotions for upcoming high-demand periods.

THANK YOU

