



# AMAZON SALES DATA ANALYSIS



## IE6600 COMPUTATION & VISUALIZATION

### GROUP 2:

~Brinda Raj Lingaraju

~Viraj Patil

~Aman Malawade

~Hardika Shroff

~Naga Sumanth Reddy Baredy





# TABLE OF CONTENTS

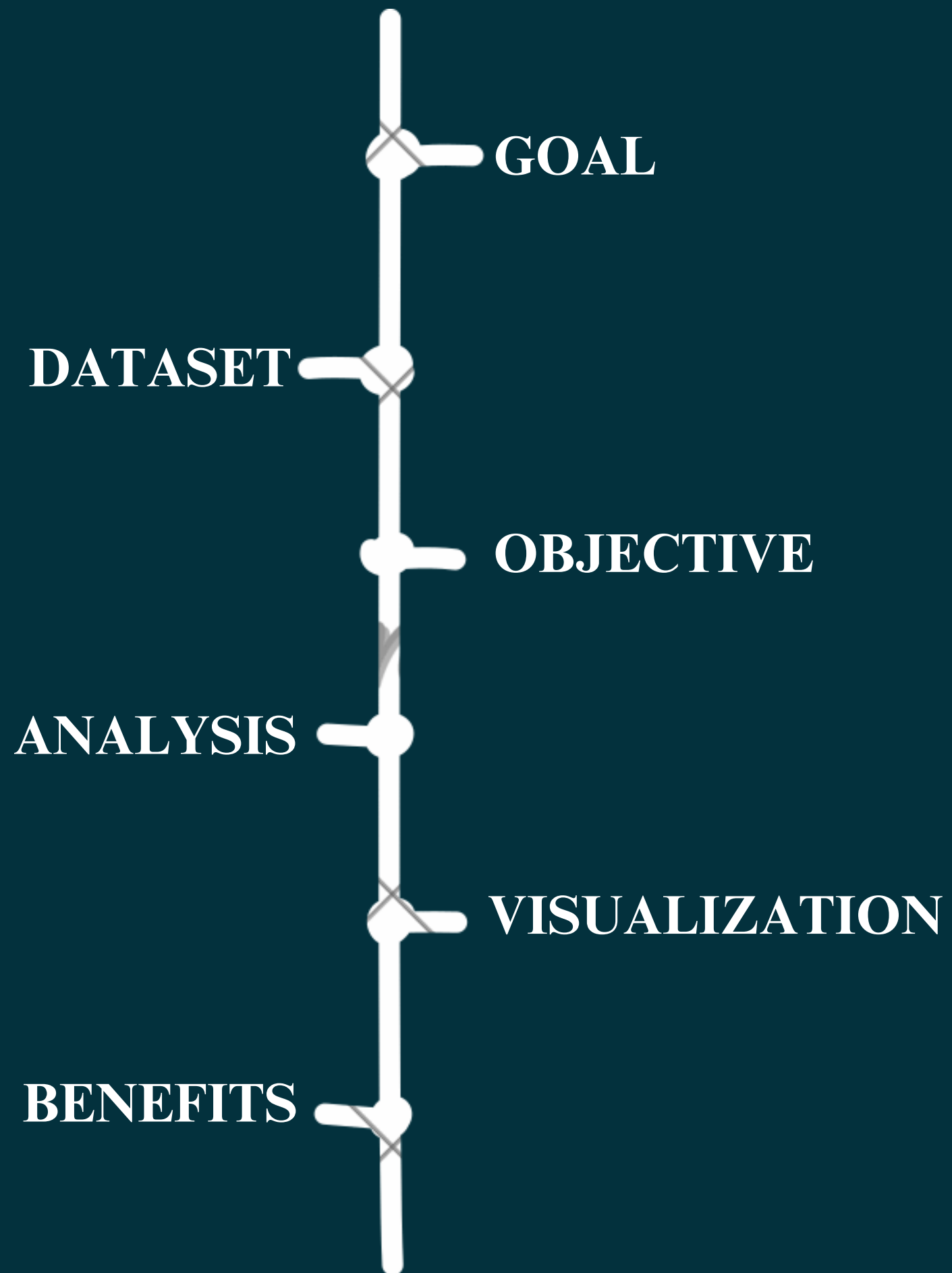
- Overview
- Problem Hypothesis
- Dataset
- Objectives and Methodology
- Data Wrangling and Preprocessing
- Exploratory Data Analysis
- Visualization using Python
- Visualization using Tableau
- Summary



# AMAZON SALES ANALYSIS



## PROJECT OVERVIEW





# PROBLEM HYPOTHESIS



*Analyzing the sales data of different product categories on Amazon, we can identify the top-performing categories and products, as well as the factors that influence their sales. This analysis can help Amazon and other businesses optimize their product offerings, marketing strategies, and inventory management to maximize their sales and improve customer satisfaction.*

# DATASET UTILIZED



Column Name	Description	Data Format	Validation
name	The name of the product	String	Required
product_category	The category that the product belongs to	String	Optional
image	The URL of the product image	String	Required
link	The URL of the product page	String	Required
ratings	The average rating of the product (out of 5)	Float	Optional
no_of_ratings	The number of ratings the product has received	Integer	Optional
discount_price	The discounted price of the product	Float	Optional
actual_price	The actual (original) price of the product	Float	Required
discount_percentage	The percentage of discount on the product (if applicable)	Float	Optional

# OBJECTIVES AND METHODOLOGY



- *Collated 7 different datasets into one to obtain the main dataset*
- *Dealt with missing values and inconsistencies in the dataset*
- *Conducted Exploratory Data Analysis on the cleaned dataset*
- *Conducted in-depth analysis and visualized insights using graphs*
- *Analyzed the influence of consumer interaction with the products to draw data-informed business insights*



# DATA WRANGLING, PREPROCESSING AND CLEANING



- *Collate data from multiple workbooks on popular categories of products to generate the final dataset*

```
amazondf[ 'product_category' ].value_counts( )
```

```
tv, audio & cameras      8988
appliances               8689
sports & fitness         2082
home & kitchen           1132
car & motorbike          1104
grocery & gourmet foods   746
pet supplies             541
Name: product_category, dtype: int64
```

# DATA WRANGLING, PREPROCESSING AND CLEANING



- *Identify and remove duplicates and handle missing values to improve data quality, increase analysis accuracy, and enhance the efficiency of data-driven decision making.*



## *#Checking Missing Values*

```
amazondf.isna().sum()
```

```
name          0
product_category  0
image         0
link          0
ratings       851
no_of_ratings  956
discount_price 1469
actual_price   196
discount_percentage 220
dtype: int64
```

```
amazondf['ratings'].value_counts()
```

```
4.1    2863
4      2722
4.2    2673
4.3    2551
3.9    2242
3.8    1826
4.4    1708
3.7    1316
4.5    1069
3.6    1046
3.5     749
5       700
3.4     545
4.6     500
3.3     391
3       271
3.2     270
4.7     226
3.1     220
4.8     125
2.9     116
Get     105
1       104
```



# EXPLORATORY DATA ANALYSIS



*Exploratory Data Analysis conducted on cleaned dataset:*

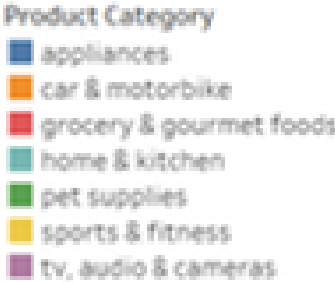
- *Identifying key factors influencing sales performance on Amazon*
- *Examined Interrelationship between different variables*
- *Analyzed distribution of variables*
- *Identify patterns and trends*
- *Identified outliers*
- *Used various statistical measures and visualizations to summarize data*



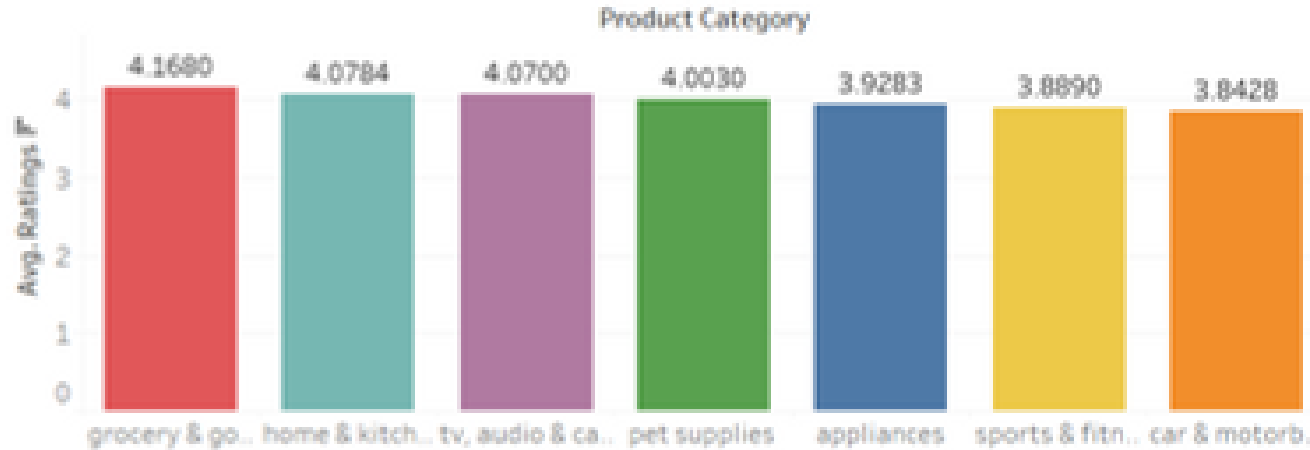
# TABLEAU INSIGHTS



## Amazon Sales Data Analysis



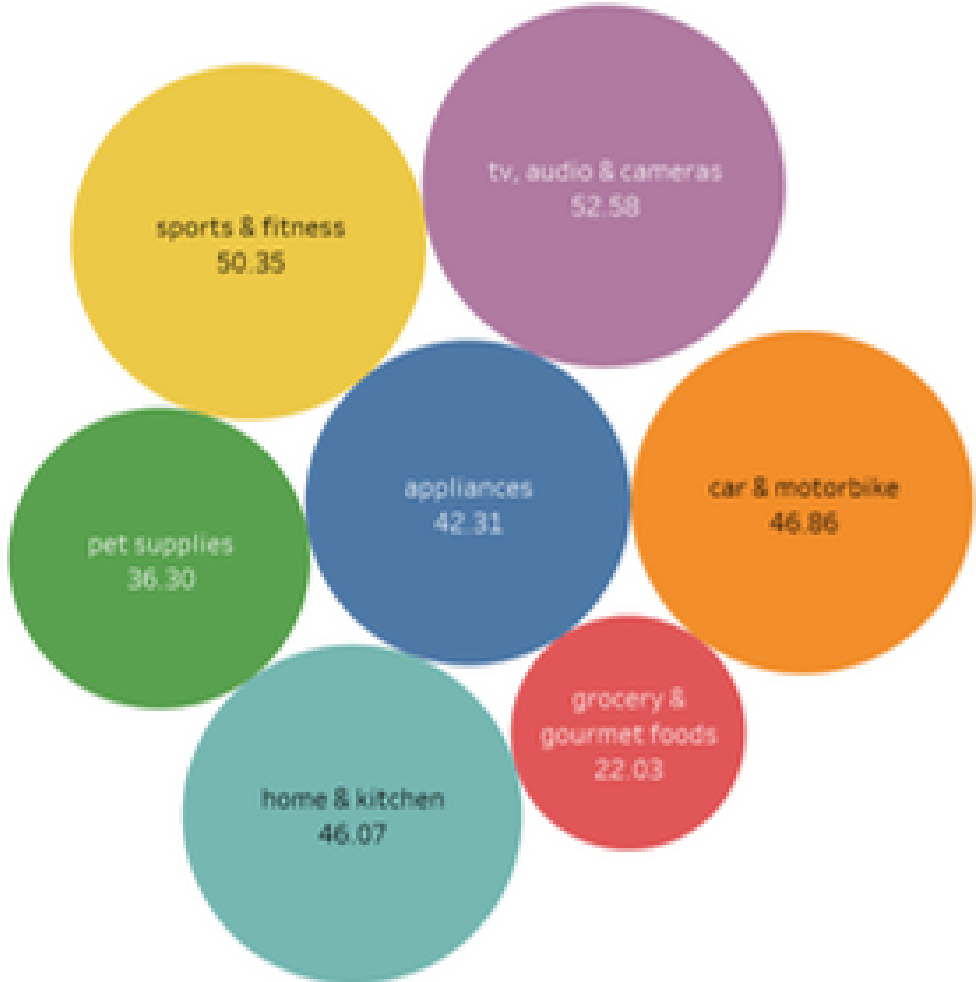
Average rating over product categories



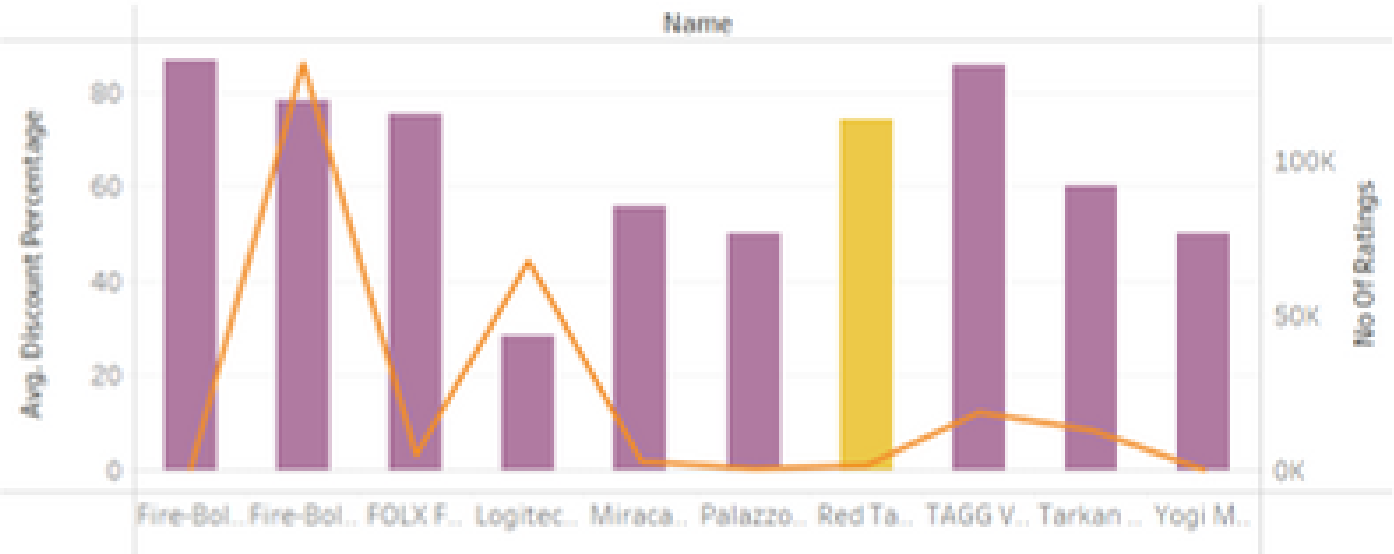
Average number of rating in each category



Average discount percentage in each category



Average discount percentage for each product

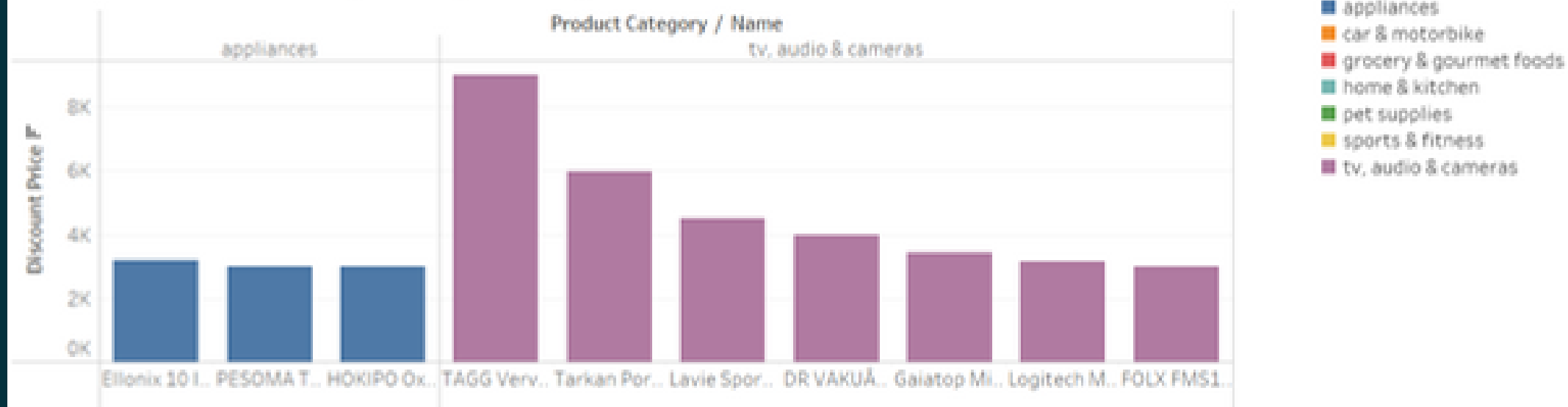


# TABLEAU INSIGHTS

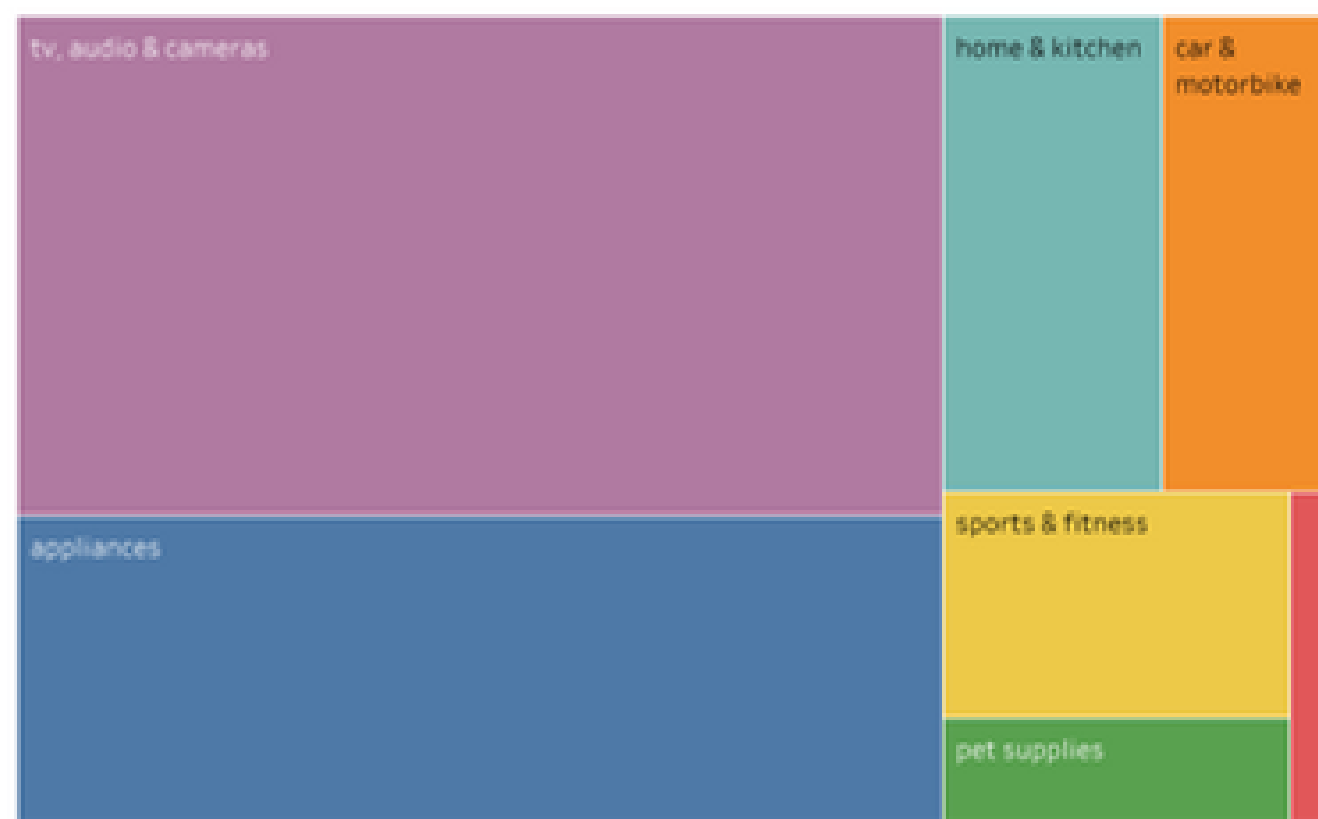


## Amazon Sales Data Analysis

Top 10 Expensive products after discount



Maximum number of ratings in each category



Discount percentage distribution





# SUMMARY AND INSIGHTS



- *The insights gained from this analysis can be used to optimize marketing strategies, improve customer engagement, and increase sales for Amazon sellers and marketers, providing a competitive edge in the market.*
- *Future insights could include identifying seasonal trends, improving supply chain management, and enhancing product listings to increase sales performance.*



THANK YOU

