



INTERNSHIP PROJECT REPORT

TITLE :Amazon sales Analysis
NAME :Mohammed Shanofar S
BATCH :01

Table of Contents



01

About Project

02

Details of Data

03

Tools Used

04

Main KPI's

05

Screenshots

06

My Design

07

Conclusion

ABOUT PROJECT



Analyzing Amazon sales data is crucial for understanding customer preferences and optimizing product offerings. However, extracting meaningful insights from the dataset, including product details, user reviews, and pricing information, poses a challenge. The problem at hand is to develop a systematic approach for extracting valuable information from this dataset, enabling businesses to enhance decision-making, identify popular products, and improve overall customer satisfaction. The goal is to derive actionable insights that can inform marketing strategies, pricing decisions, and product improvements based on user feedback.

DETAILS OF DATA



EDA projects we will be performing amazon sales data analysis and it a interesting dataset to work on . Explore the popularity of products based on the number of user reviews and ratings.

TOOLS USED



- Pandas
- Matplotlib
- Seaborn
- Power bi
- Tableau

MAIN KPI's



Tasks

Identifying top 5 products based on rating

Analyze the distribution of product prices, including both discounted and actual prices.

Identify and list products that have the highest number of user reviews

Explore the popularity of products based on the number of user reviews and ratings.

SCREENSHOTS

Python Code



```
In [1]: #Importing Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: df = pd.read_csv('amazon.csv')
df
```

```
Out[2]:
```

	product_id	product_name	category	discounted_price	actual_price	discount_percentage	rating	rating_count
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories/Accessories&Peripherals...	₹399	₹1,099	64%	4.2	24,289
1	B09N56PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories/Accessories&Peripherals...	₹199	₹349	43%	4.0	43,994
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories/Accessories&Peripherals...	₹199	₹1,899	90%	3.9	7,928
3	B08HJ08NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories/Accessories&Peripherals...	₹329	₹699	53%	4.2	94,363
4	B08CF3B7N1	Portronics Kornect L 1.2M Fast Charging 3A/8 P...	Computers&Accessories/Accessories&Peripherals...	₹154	₹399	61%	4.2	16,905
...
1460	B08L7J3T31	Noir Aqua - Spas PP Spun Filter + 1 Spanner ...	Home&Kitchen/Kitchen&HomeAppliances/WaterPurif...	₹379	₹919	58%	4	1,090
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen/Kitchen&HomeAppliances/SmallKitch...	₹2,280	₹3,045	25%	4.1	4,118
1462	B009P2JL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen/Heating,Cooling&AirQuality/RoomHea...	₹2,219	₹3,080	28%	3.6	468
1463	B00J5OYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Plat...	Home&Kitchen/Heating,Cooling&AirQuality/Fans[E...	₹1,399	₹1,890	26%	4	8,031
1464	B01486F4G6	Borosi Jumbo 1000-Watt Grill Sandwich Maker (...)	Home&Kitchen/Kitchen&HomeAppliances/SmallKitch...	₹2,863	₹3,690	22%	4.3	6,987

1465 rows x 9 columns

```
In [3]: #view first 50 rows
df.head(6)
```

```
Out[3]:
```

	product_id	product_name	category	discounted_price	actual_price	discount_percentage	rating	rating_count
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories/Accessories&Peripherals...	₹399	₹1,099	64%	4.2	24,289
1	B09N56PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories/Accessories&Peripherals...	₹199	₹349	43%	4.0	43,994

```
In [4]: df.shape
Out[4]: (1465, 9)
```

```
In [5]: df.dtypes
Out[5]: product_id      object
product_name      object
category          object
discounted_price   object
actual_price       object
discount_percentage object
rating            object
rating_count      object
about_product     object
user_id          object
user_name         object
review_id         object
review_title      object
review_content    object
img_link          object
product_link      object
dtype: object
```

```
In [6]: #Replacing string as int in actual_price
df['actual_price'] = df['actual_price'].str.replace('₹','')
df['actual_price'] = df['actual_price'].str.replace(',','').astype('float64')
df.head(6)
```

```
Out[6]:
```

	product_id	product_name	category	discounted_price	actual_price	discount_percentage	rating	rating_count	about
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories/Accessories&Peripherals...	₹399	1099.0	64%	4.2	24,289	Cor (V
1	B09N56PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories/Accessories&Peripherals...	₹199	349.0	43%	4.0	43,994	(with de
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories/Accessories&Peripherals...	₹199	1899.0	90%	3.9	7,928	Cha & bul
3	B08HJ08NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories/Accessories&Peripherals...	₹329	699.0	53%	4.2	94,363	C
4	B08CF3B7N1	Portronics Kornect L 1.2M Fast Charging 3A/8 P...	Computers&Accessories/Accessories&Peripherals...	₹154	399.0	61%	4.2	16,905	[C
5	B08Y1TSP6	pTron Solero TB301 3A Type-C Data and Fast Cha...	Computers&Accessories/Accessories&Peripherals...	₹149	1000.0	85%	3.9	24,871	Fas & Sol

```
In [7]: #Replacing string as int discounted_price
df['discounted_price'] = df['discounted_price'].str.replace('₹','')
df['discounted_price'] = df['discounted_price'].str.replace(',','').astype('float64')
df.head(6)
```

```
Out[7]:
```

	product_id	product_name	category	discounted_price	actual_price	discount_percentage	rating	rating_count	about
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories/Accessories&Peripherals...	399.0	1099.0	64%	4.2	24,289	Cor (V
1	B09N56PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories/Accessories&Peripherals...	199.0	349.0	43%	4.0	43,994	(with de

SCREENSHOTS

Python Code



```
In [8]: #removing % from discount_percentage
df['discount_percentage'] = df['discount_percentage'].str.replace('%','').astype('float64')
df.head(6)
```

```
Out[8]:
```

	product_id	product_name	category	discounted_price	actual_price	discount_percentage	rating	rating_count	about
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories/Accessories&Peripherals...	399.0	1099.0	64.0	4.2	24,269	Cor (V
1	B098N56PVG	Ambrae Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories/Accessories&Peripherals...	199.0	349.0	43.0	4.0	43,994	(with de
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories/Accessories&Peripherals...	199.0	1899.0	90.0	3.9	7,928	Cha S bul
3	B08HDJ86NZ	boAt Device USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories/Accessories&Peripherals...	329.0	699.0	53.0	4.2	94,363	[
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories/Accessories&Peripherals...	154.0	399.0	61.0	4.2	16,905	Fl c
5	B08Y1TFSF6	pTron Sclero TB301 3A Type-C Data and Fast Cha...	Computers&Accessories/Accessories&Peripherals...	149.0	1000.0	85.0	3.9	24,871	Fast & Sol

```
In [9]: #removing , from rating_count
df['rating_count'] = df['rating_count'].str.replace(',','').astype('float64')
df.head(6)
```

```
Out[9]:
```

	product_id	product_name	category	discounted_price	actual_price	discount_percentage	rating	rating_count	about
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories/Accessories&Peripherals...	399.0	1099.0	64.0	4.2	24269.0	Cor (V
1	B098N56PVG	Ambrae Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories/Accessories&Peripherals...	199.0	349.0	43.0	4.0	43994.0	(with de
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories/Accessories&Peripherals...	199.0	1899.0	90.0	3.9	7928.0	Cha S bul
3	B08HDJ86NZ	boAt Device USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories/Accessories&Peripherals...	329.0	699.0	53.0	4.2	94363.0	[
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories/Accessories&Peripherals...	154.0	399.0	61.0	4.2	16905.0	Fl c
5	B08Y1TFSF6	pTron Sclero TB301 3A Type-C Data and Fast Cha...	Computers&Accessories/Accessories&Peripherals...	149.0	1000.0	85.0	3.9	24871.0	Fast & Sol

```
In [10]: #find the total values in --rating
df['rating'].value_counts()
```

```
Out[10]:
```

4.1	244
4.3	230
4.2	228
4.0	129
3.9	123
4.4	123
3.8	86
4.5	75
4	52
3.7	42
3.6	35
3.5	26
4.6	17
3.3	16
3.4	10
4.7	6
3.1	4
5.0	3
3.0	3
4.8	3
3.2	2
2.8	2
2.3	1
1	1
2	1
3	1
2.6	1
2.9	1

Name: rating, dtype: int64

```
In [11]: df.duplicated()
```

```
Out[11]:
```

0	False
1	False
2	False
3	False
4	False
...	...
1460	False
1461	False
1462	False
1463	False
1464	False

Length: 1465, dtype: bool

```
In [12]: #Replacing / to 3.5 from rating
df['rating'] = df['rating'].str.replace('/', '3.5').astype('float64')
df['rating'].value_counts()
```

C:\Users\ZIS\AppData\Local\Temp\ipykernel_8016\708494033.py:2: FutureWarning: The default value of regex will change from True to False in a future version. In addition, single character regular expressions will *not* be treated as literal strings when r regex=True.

```
df['rating'] = df['rating'].str.replace('/', '3.5').astype('float64')
```

```
Out[12]:
```

4.1	244
4.3	230
4.2	228
4.0	181
3.9	123
4.4	123
3.8	86
4.5	75
3.7	42
3.6	35
3.5	27
4.6	17
3.3	16
3.4	10
4.7	6
3.0	4
3.1	4
5.0	3
4.8	3
3.2	2

SCREENSHOTS

Python Code



```
In [14]: df.isnull().sum()
```

```
Out[14]: product_id      0
product_name    0
category        0
discounted_price 0
actual_price    0
discount_percentage 0
rating          2
rating_count    0
about_product   0
user_id        0
user_name      0
review_id      0
review_title   0
review_content  0
img_link       0
product_link   0
dtype: Int64
```

```
In [15]: cat_split=df['category'].str.split('|', expand = True)
cat_split
```

```
Out[15]:
```

	0	1	2	3	4	5	6
0	Computers&Accessories	Accessories&Peripherals	Cables&Accessories	Cables	USBCables	None	None
1	Computers&Accessories	Accessories&Peripherals	Cables&Accessories	Cables	USBCables	None	None
2	Computers&Accessories	Accessories&Peripherals	Cables&Accessories	Cables	USBCables	None	None
3	Computers&Accessories	Accessories&Peripherals	Cables&Accessories	Cables	USBCables	None	None
4	Computers&Accessories	Accessories&Peripherals	Cables&Accessories	Cables	USBCables	None	None
...
1460	Home&Kitchen	Kitchen&HomeAppliances	WaterPurifiers&Accessories	WaterPurifierAccessories	None	None	None
1461	Home&Kitchen	Kitchen&HomeAppliances	SmallKitchenAppliances	Rice&PastaCookers	None	None	None
1462	Home&Kitchen	Heating,Cooling&AirQuality	RoomHeaters	HeatConvector	None	None	None
1463	Home&Kitchen	Heating,Cooling&AirQuality	Fans	ExhaustFans	None	None	None
1464	Home&Kitchen	Kitchen&HomeAppliances	SmallKitchenAppliances	SandwichMakers	None	None	None

1465 rows x 7 columns

```
In [16]: cat_split = cat_split.rename(columns={0:'main category 1',1:'sub category 1',2:'sub category 2'})
cat_split
```

```
Out[16]:
```

	main category 1	sub category 1	sub category 2	3	4	5	6
0	Computers&Accessories	Accessories&Peripherals	Cables&Accessories	Cables	USBCables	None	None
1	Computers&Accessories	Accessories&Peripherals	Cables&Accessories	Cables	USBCables	None	None
2	Computers&Accessories	Accessories&Peripherals	Cables&Accessories	Cables	USBCables	None	None
3	Computers&Accessories	Accessories&Peripherals	Cables&Accessories	Cables	USBCables	None	None
4	Computers&Accessories	Accessories&Peripherals	Cables&Accessories	Cables	USBCables	None	None
...
1460	Home&Kitchen	Kitchen&HomeAppliances	WaterPurifiers&Accessories	WaterPurifierAccessories	None	None	None
1461	Home&Kitchen	Kitchen&HomeAppliances	SmallKitchenAppliances	Rice&PastaCookers	None	None	None
1462	Home&Kitchen	Heating,Cooling&AirQuality	RoomHeaters	HeatConvector	None	None	None
1463	Home&Kitchen	Heating,Cooling&AirQuality	Fans	ExhaustFans	None	None	None
1464	Home&Kitchen	Kitchen&HomeAppliances	SmallKitchenAppliances	SandwichMakers	None	None	None

1465 rows x 7 columns

```
In [19]: df.dropna(how='any')
```

```
Out[19]:
```

	product_id	product_name	discounted_price	actual_price	discount_percentage	rating	rating_count	about_product
0	B07JW9HJ41	Wayona Nylon Braided USB to Lightning Fast Cha...	399.0	1099.0	64.0	4.2	24269.0	High Compatibility: Compatible With iPhone 12...
1	B09NS6PVG	Ambrane Unbreakable 80W / 3A Fast Charging 1.5...	199.0	349.0	43.0	4.0	43994.0	Compatible with all Type C enabled devices, be...
2	B096MSW6CT	Sources Fast Phone Charging Cable & Data Sync U...	199.0	1899.0	90.0	3.9	7928.0	【 Fast Charge& Data Sync 】 -With built-in safet...
3	B09HJ96NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	329.0	699.0	53.0	4.2	94363.0	The boAt Deuce USB 300 2 in 1 cable is compati...
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A B P...	154.0	399.0	61.0	4.2	16905.0	[CHARGE & SYNC FUNCTION] -This cable comes wi...
...
1460	B08L7J3T31	Nair Aqua - Spca PP Span Filter + 1 Spanner ...	379.0	919.0	59.0	4.0	1090.0	SUPREME QUALITY 90 GRAM 3 LAYER THK PP SPUN F...
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	2280.0	3045.0	25.0	4.1	4118.0	230 Volts, 400 watts, 1 Year
1462	B006P2UL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	2219.0	3080.0	28.0	3.6	488.0	International design and styling for heat sett...
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Plat...	1399.0	1890.0	26.0	4.0	8031.0	Fan sweep area: 230 MM ; Noise level: (40 -45...
1464	B0148674G6	Borosil Jumbo 1000 Watt Grill Sandwich Maker (...)	2863.0	3690.0	22.0	4.3	6987.0	Brand Borosil, Specification 9C* 23V ~ 5Hz* 1 W...

1455 rows x 18 columns

```
In [20]: df['main category 1'].value_counts()
```

```
Out[20]: Electronics      526
Computers&Accessories    453
Home&Kitchen             448
OfficeProducts           31
MusicalInstruments       2
HomeImprovement          2
Toys&Games               1
Car&Motorbike            1
Health&PersonalCare      1
Name: main category 1, dtype: Int64
```

```
In [21]: df['main category 1'] = df['main category 1'].str.replace('8',' & ')
df['main category 1'] = df['main category 1'].str.replace('OfficeProducts','Office Products')
df['main category 1'] = df['main category 1'].str.replace('MusicalInstruments','Musical Instruments')
df['main category 1'] = df['main category 1'].str.replace('HomeImprovement','Home Improvement')
df
```

```
Out[21]:
```

	product_id	product_name	discounted_price	actual_price	discount_percentage	rating	rating_count	about_product
		Wayona Nylon						High

MY DESIGN Python Code



```
In [23]: df['sub_category 1']=df['sub_category 1'].str.replace(' & ', ' & ')
df['sub_category 1']=df['sub_category 1'].str.replace(' ', ',')
df['sub_category 1']=df['sub_category 1'].str.replace('WearableTechnology', 'Wearable Technology')
df['sub_category 1']=df['sub_category 1'].str.replace('NetworkingDevices', 'Networking Devices')
df['sub_category 1']=df['sub_category 1'].str.replace('OfficePaperProducts', 'Office Paper Products')
df['sub_category 1']=df['sub_category 1'].str.replace('ExternalDevices', 'External Devices')
df['sub_category 1']=df['sub_category 1'].str.replace('HomeStorage', 'Home Storage')
df['sub_category 1']=df['sub_category 1'].str.replace('HomeAudio', 'Home Audio')
df['sub_category 1']=df['sub_category 1'].str.replace('GeneralPurposeBatteries', 'General Purpose Batteries')
df['sub_category 1']=df['sub_category 1'].str.replace('BatteryChargers', 'Battery Chargers')
df['sub_category 1']=df['sub_category 1'].str.replace('CraftMaterials', 'Craft Materials')
df['sub_category 1']=df['sub_category 1'].str.replace('OfficeElectronics', 'Office Electronics')
df['sub_category 1']=df['sub_category 1'].str.replace('CarAccessories', 'Car Accessories')
df['sub_category 1']=df['sub_category 1'].str.replace('HomeMedicalSupplies', 'Home Medical Supplies')
df
```

Out[23]:

	product_id	product_name	discounted_price	actual_price	discount_percentage	rating	rating_count	about_product
	0	B07JW9H4U1	399.0	1099.0	64.0	4.2	24269.0	High Compatibility: Compatible With iPhone 12...
	1	B09N56PVG	199.0	349.0	43.0	4.0	43994.0	Compatible with all Type C enabled devices, be...
	2	B098MSW6CT	199.0	1899.0	90.0	3.9	7928.0	[Fast Charger & Data Sync] With built-in safety...
	3	B08HDJ88NZ	329.0	699.0	53.0	4.2	94363.0	The built-in Deuce USB 300 2 in 1 cable is compa...
	4	B08CF3B7N1	154.0	399.0	61.0	4.2	16905.0	[CHARGE & SYNC FUNCTION] The cable comes wit...
...
1460	B08L7J3T31	Nat Aqua-Spice PP Span Filter + 1 Spanner ...	379.0	919.0	59.0	4.0	1090.0	SUPREME QUALITY 90 GRAM 3 LAYER TRIP PP SPUN F...
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	2280.0	3045.0	25.0	4.1	4118.0	230 Volts, 400 watts, 1 Year
1462	B009P2LJL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	2219.0	3080.0	28.0	3.8	468.0	International design and styling Two heat set...
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Plat...	1399.0	1890.0	26.0	4.0	8031.0	Fan sweep area: 230 MM Noise level: (40 ~ 45...
1464	B01486F4G6	Borosi Jumbo 1000-Watt Grill Sandwich Maker (...)	2863.0	3690.0	22.0	4.3	6987.0	Brand Borosi, Specification 40° 23V ~ 5Hz 1 W...

1465 rows x 18 columns

```
In [24]: df['sub_category 2'].value_counts()

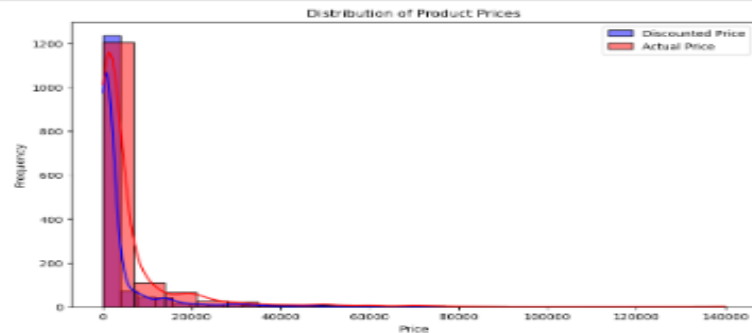
Out[24]:
```

Cables&Accessories	248
SmallKitchenAppliances	181
Accessories	97
MobileAccessories	84
Vacuum,Cleaning&Ironing	82

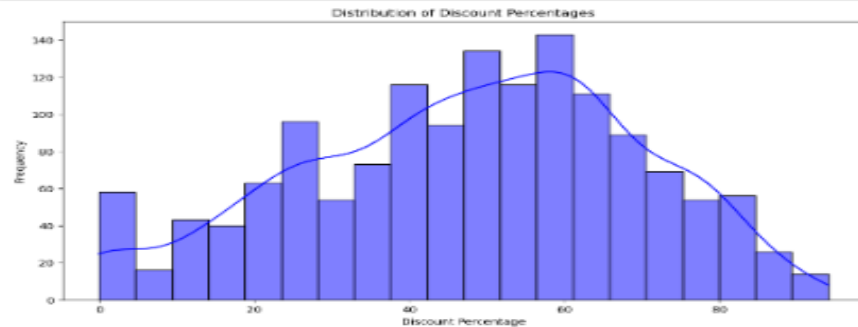
1. Top Products by Ratings:

Identifying top product by rating

```
In [33]: #Task: Identify and list the top-rated products based on user ratings.
plt.figure(figsize=(18, 6))
sns.histplot(df['discounted_price'], bins=20, kde=True, color='blue', label='Discounted Price')
sns.histplot(df['actual_price'], bins=20, kde=True, color='red', label='Actual Price')
plt.title('Distribution of Product Prices')
plt.xlabel('Price')
plt.ylabel('Frequency')
plt.legend()
plt.show()
```



```
In [34]: # Examine the distribution of discount percentages across different products.
plt.figure(figsize=(12, 6))
sns.histplot(df['discount_percentage'], bins=20, kde=True, color='blue')
plt.title('Distribution of Discount Percentages')
plt.xlabel('Discount Percentage')
plt.ylabel('Frequency')
plt.show()
```

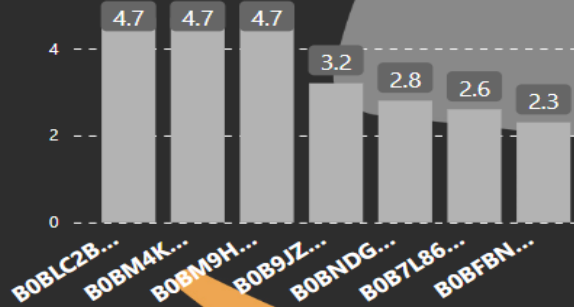




Top 5 Products ID By Rating

Sum of rating by product_id

Sum of rating

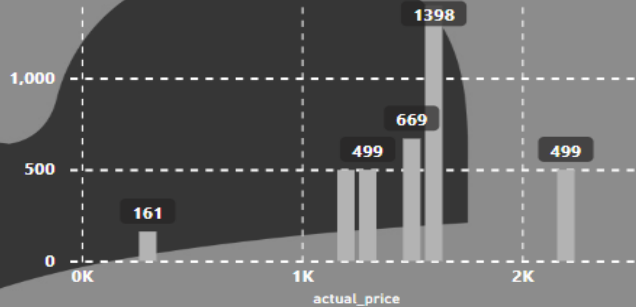


product_id

Price Distribution Based on Discount Price

Sum of discounted_price by actual_price

Sum of discounted_price



actual_price

sub category 2

Small Kitchen Appliances

59

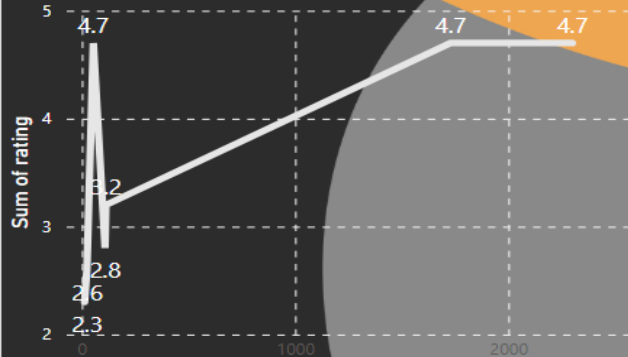
Average of discount_perc...

5

Count of rating

Popular Product

Sum of rating by rating_count and sub category 2



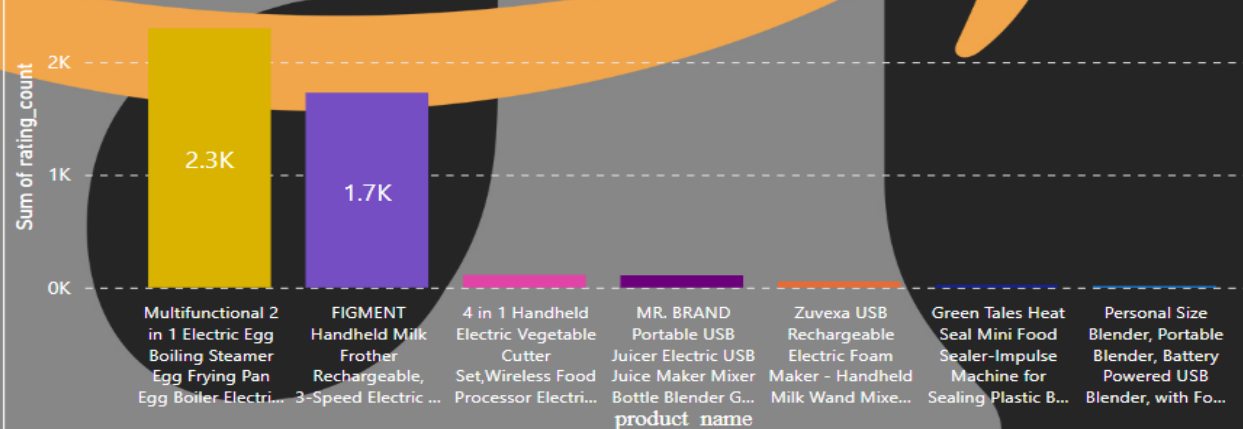
rating_count

● Small Kitchen Appliances

Most Discussed Product Based On Rating count

Sum of rating_count by product_name and rating_count

rating_count ● 13 ● 24 ● 54 ● 109 ● 113 ● 1729 ● 2300



CONCLUSION



1. **Continuous Growth:** Amazon's sales have continued to grow steadily over the years, demonstrating its strong position in the e-commerce market.
2. **Seasonal Trends:** Seasonal variations play a significant role in Amazon's sales, with peak periods during the holiday season and major shopping events like Prime Day. Understanding and capitalizing on these trends can help businesses maximize their sales on the platform.
3. **Product Categories:** Different product categories experience varying levels of demand on Amazon. Businesses should carefully select and manage their product offerings based on the data and trends within their specific niche.



SADHVI

SOFTWARE SOLUTIONS

Thank You!