amazon

Contents

Introduction

Getting sales insights is transformative for Amazon sellers, as it allows them to understand customer behavior, optimize product offerings, and make data-driven decisions that can significantly boost sales and profitability. Here's a breakdown of how sales insights impact an Amazon business:

Dataset Overview

Here we will have a quick overview on our data:

- Dataset Features and Observations
- Null Values
- Datatypes

	product_id	product_name	category	discounted_price	actual_price	discount_percentage	rating	rating_count	about_product
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories|Accessories&Peripherals|...	₹399	₹1,099	64%	4.2	24,269	High Compatibility Compatible With iPhone 12...
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories|Accessories&Peripherals|...	₹199	₹349	43%	4.0	43,994	Compatible with all Type C enabled devices, be...
2	B096MSW6CT	Sounce Fast Phone Charging Cable & Data Sync U...	Computers&Accessories|Accessories&Peripherals|...	₹199	₹1,899	90%	3.9	7,928	[Fast Charger& Data Sync] -With built-in safet...
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories|Accessories&Peripherals|...	₹329	₹699	53%	4.2	94,363	The boAt Deuce USB 300 2 in 1 cable is compati...
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories|Accessories&Peripherals|...	₹154	₹399	61%	4.2	16,905	[CHARGE & SYNC FUNCTION] This cab comer...

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1465 entries, 0 to 1464
Data columns (total 16 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   product_id           1465 non-null    object
 1   product_name         1465 non-null    object
 2   category             1465 non-null    object
 3   discounted_price      1465 non-null    object
 4   actual_price         1465 non-null    object
 5   discount_percentage  1465 non-null    object
 6   rating               1465 non-null    object
 7   rating_count         1463 non-null    object
 8   about_product        1465 non-null    object
 9   user_id              1465 non-null    object
 10  user_name            1465 non-null    object
 11  review_id            1465 non-null    object
 12  review_title         1465 non-null    object
 13  review_content       1465 non-null    object
 14  img_link             1465 non-null    object
 15  product_link         1465 non-null    object
dtypes: object(16)
memory usage: 183.3+ KB
```

Feature Engineering

We will find that there are some Categorical data need to be converted to Quantative data by removing symbols, split values, change Dtype , Removing NaN values.

Here we will go through some data cleaning by removing symbols and changing data types

Splitting the category column must be done to figure out the category and subcategory of Items

Feature Engineering

After the preparation of the dataset, we will find the new dataset shape

amazon_ds_update								
	product_id	product_name	discounted_price	actual_price	discount_percentage	rating	rating_count	about_product
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	399.0	1099.0	0.64	4.2	24269.0	High Compatibility : Compatible With iPhone 12...
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	199.0	349.0	0.43	4.0	43994.0	Compatible with all Type C enabled devices, be...

	user_id	user_name	sub_category	main_category
	AG3D6O4STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB...	Manav,Adarsh gupta,Sundeep,S.Sayeed Ahmed,jasp...	USBCables	Accessories&Peripherals
	AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX...	ArdKn,Nirbhay kumar,Sagar Viswanathan,Asp,Plac...	USBCables	Accessories&Peripherals

Removed all unnecessary symbols

We will find the new columns added
- Main_Category
- Sub_Category

```
amazon_ds_update.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1465 entries, 0 to 1464
Data columns (total 12 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   product_id           1465 non-null   object
 1   product_name         1465 non-null   object
 2   discounted_price     1465 non-null   float64
 3   actual_price         1465 non-null   float64
 4   discount_percentage  1465 non-null   float64
 5   rating               1464 non-null   float64
 6   rating_count         1465 non-null   float64
 7   about_product        1465 non-null   object
 8   user_id              1465 non-null   object
 9   user_name            1465 non-null   object
 10  sub_category         1465 non-null   object
 11  main_category        1465 non-null   object
dtypes: float64(5), object(7)
memory usage: 137.5+ KB
```
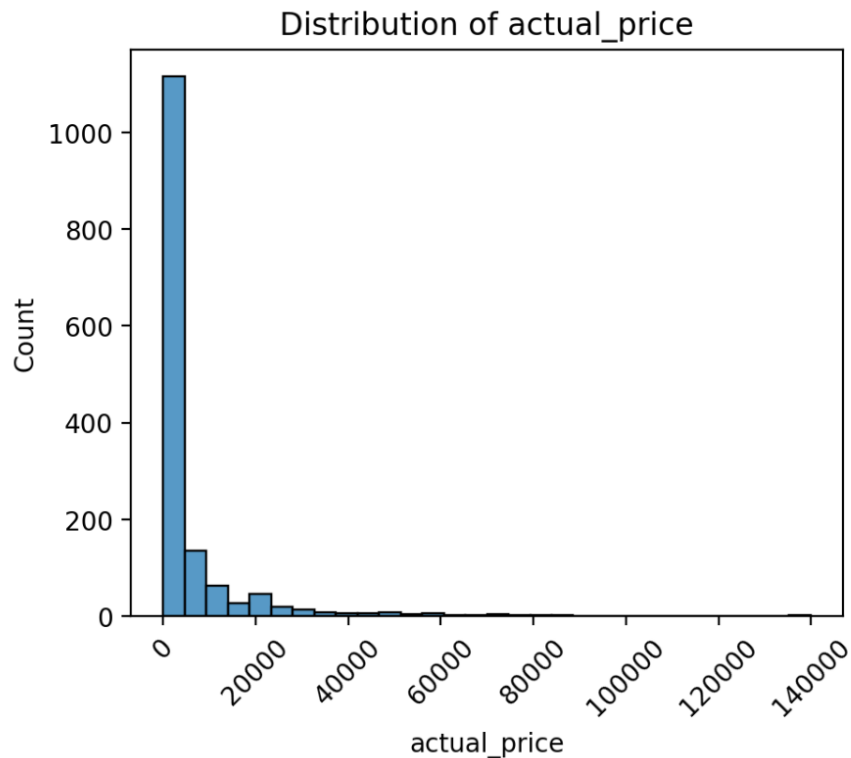
Updated necessary Dtypes

Exploratory Data Analysis (EDA)

Uni-Variate Analysis

Bi-Variate Analysis
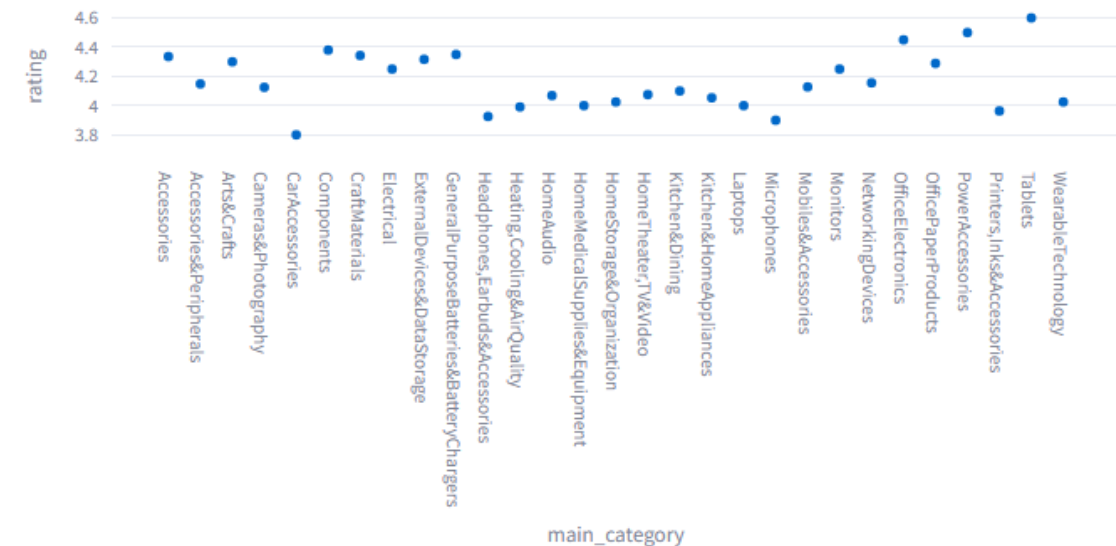
Application Overview
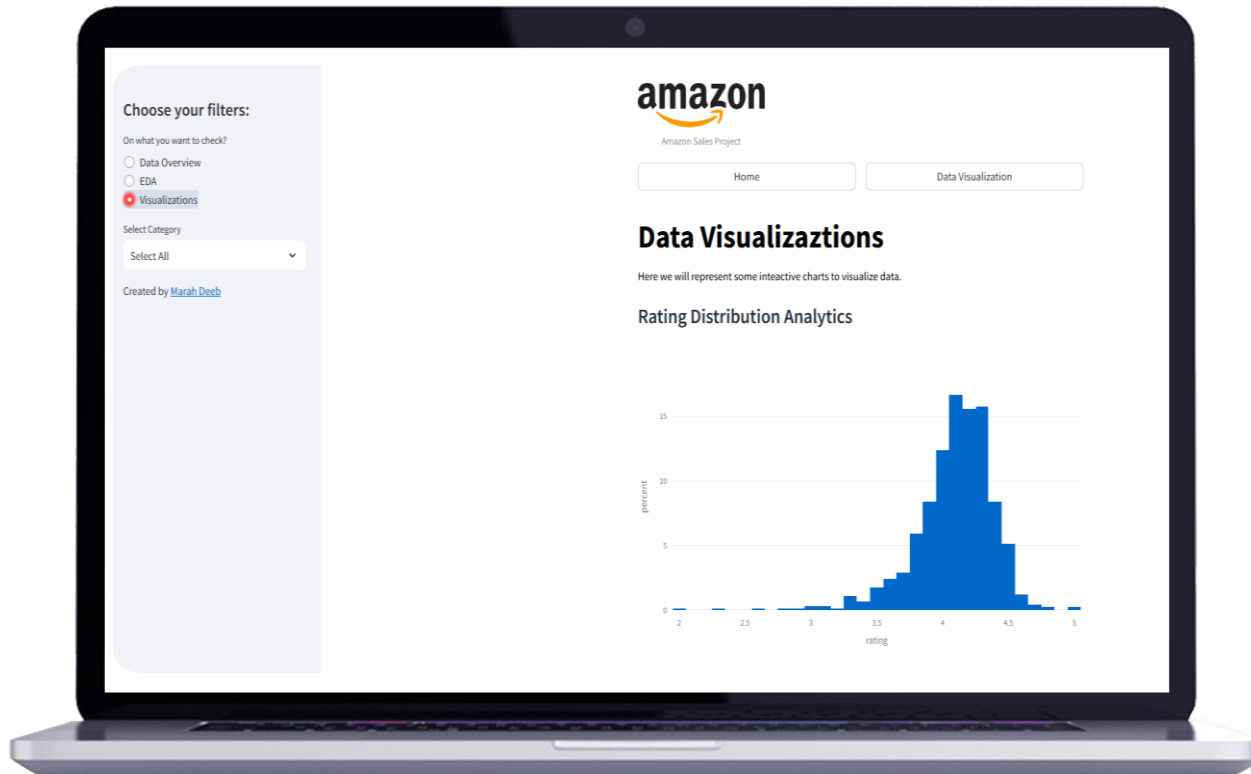
Let's Go through our application through the following URL:
Amazon App

Any Questions?

thanks!