2024 Amazon Best Sellers: Top Valentine Gifts 🎯 💘

Overview:

Dive into the heart of Valentine's Day shopping trends with our curated dataset, "2024 Amazon Best Sellers: Top Valentine Gifts." This dataset offers a snapshot of the most popular gifts purchased from Amazon for Valentine's Day in 2024, providing unique insights into consumer preferences and market dynamics during the season of love.

Dataset Details:

The dataset is structured into key columns, detailed as follows:

- title: Name of the product.
- brand: Brand associated with the product.
- description: Brief description of the product.
- starsBreakdown: Percentage distribution of ratings (3-star, 4-star, 5-star).
- reviewsCount: Total number of reviews.
- price: Listed price of the product in USD.
- categoryPageData: Product's positioning data on Amazon's category pages.

Import the neccesarry modules

In [1]:
```python
# This Python 3 environment comes with many helpful analytics libraries
installed
# It is defined by the kaggle/python Docker image: https://github.com/ka
ggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt #for data visualization
import seaborn as sb #an improvement of matplotlib


# to suppress warnings
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```python
# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) wi
ll list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
/kaggle/input/2024-amazon-best-sellers-top-valentine-gifts/amazon_20
24_valentines_best_sellers.csv
```

In [3]:
```python
#here you use the pandas library to load the data for analysis
df = pd.read_csv('/kaggle/input/2024-amazon-best-sellers-top-valentine-
gifts/amazon_2024_valentines_best_sellers.csv')
```

Understanding of our dataset

In [4]:
```python
# see the number of rows and columns
df.shape
```

Out[4]:
```
(218, 11)
```

In [5]:
```python
# see the name of the columns
df.columns
```

Out[5]:
```
Index(['title', 'brand', 'description', 'starsBreakdown/3star',
       'starsBreakdown/4star', 'starsBreakdown/5star', 'reviewsCoun
t', 'price',
       'price/currency', 'price/value', 'categoryPageData/productPos
ition'],
      dtype='object')
```

In [6]:
```python
# see a brief statistical decription on each column
df.describe()
```

Out[6]:

	starsBreakdown/3star	starsBreakdown/4star	starsBreakdown/5star	reviewsCount	p
count	218.000000	218.000000	218.000000	216.000000	0
mean	0.049220	0.105275	0.767936	2863.759259	N
std	0.041966	0.066756	0.151997	7104.953282	N
min	0.000000	0.000000	0.000000	1.000000	N
25%	0.030000	0.072500	0.720000	54.750000	N
50%	0.040000	0.100000	0.790000	387.500000	N
75%	0.070000	0.130000	0.850000	1581.500000	N
max	0.380000	0.490000	1.000000	54895.000000	N

In [7]:

```python
# see the top 5 rows
df.head()
```

Out[7]:

	title	brand	description	starsBreakdo...
0	Ferrero Rocher, 24 Count, Premium Milk Chocola...	Ferrero Rocher	Ferrero Rocher's milk chocolate gift box offer...	0.02
1	HERSHEY'S NUGGETS Assorted Chocolate, Valentin...	HERSHEY'S	This HERSHEY'S NUGGETS candy assortment is fil...	0.03
2	LEGO Icons Flower Bouquet Building Decoration ...	LEGO	Giving and receiving beautiful flowers is such...	0.01
3	BodyRefresh Shower Steamers Aromatherapy - 8 P...	BodyRefresh	NaN	0.07
4	JoJowell Shower Steamers Aromatherapy - 21Pcs ...	JoJowell	NaN	0.10

Data Cleaning

In [8]:
```python
# see the total number of missing values in each row
df.isnull().sum()
```

Out[8]:

```
title                             0
brand                             1
description                     172
starsBreakdown/3star              0
starsBreakdown/4star              0
starsBreakdown/5star              0
reviewsCount                      2
price                           218
price/currency                   49
price/value                      49
categoryPageData/productPosition  0
dtype: int64
```

In [9]:
```python
# Delete the 'description' and 'price/currency' columns
df.drop(['description', 'price/currency'], axis=1, inplace=True)

# Display the modified DataFrame
df.head()
```

Out[9]:

	title	brand	starsBreakdown/3star	sta
0	Ferrero Rocher, 24 Count, Premium Milk Chocola...	Ferrero Rocher	0.02	0.0
1	HERSHEY'S NUGGETS Assorted Chocolate, Valentin...	HERSHEY'S	0.03	0.1
2	LEGO Icons Flower Bouquet Building Decoration ...	LEGO	0.01	0.0
3	BodyRefresh Shower Steamers Aromatherapy - 8 P...	BodyRefresh	0.07	0.1
4	JoJowell Shower Steamers Aromatherapy - 21Pcs ...	JoJowell	0.10	0.1

In [10]:
```python
# see how your new dataset looks with dropped columns
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 218 entries, 0 to 217
Data columns (total 9 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   title                          218 non-null    object
 1   brand                          217 non-null    object
 2   starsBreakdown/3star           218 non-null    float64
 3   starsBreakdown/4star           218 non-null    float64
 4   starsBreakdown/5star           218 non-null    float64
 5   reviewsCount                   216 non-null    float64
 6   price                          0 non-null      float64
 7   price/value                    169 non-null    float64
 8   categoryPageData/productPosition  218 non-null    int64
dtypes: float64(6), int64(1), object(2)
memory usage: 15.5+ KB
```

In [11]:
```python
#see remining columns with missing values
df.isnull().sum()
```

Out[11]:
```
title                             0
brand                             1
starsBreakdown/3star              0
starsBreakdown/4star              0
starsBreakdown/5star              0
reviewsCount                      2
price                           218
price/value                      49
categoryPageData/productPosition  0
dtype: int64
```

In [13]:
```python
# replace each column of missing values with its respective mean

mean_review = df['reviewsCount'].mean()
mean_currency = df['price/value'].mean()


df['price/value'].fillna(mean_currency, inplace=True)
```

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

Out[14]:
```
title                              0
brand                              1
starsBreakdown/3star               0
starsBreakdown/4star               0
starsBreakdown/5star               0
reviewsCount                       2
price                            218
price/value                        0
categoryPageData/productPosition   0
dtype: int64
```

In [15]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 218 entries, 0 to 217
Data columns (total 9 columns):
 #   Column                           Non-Null Count  Dtype
---  ------                           --------------  -----
 0   title                            218 non-null    object
 1   brand                            217 non-null    object
 2   starsBreakdown/3star             218 non-null    float64
 3   starsBreakdown/4star             218 non-null    float64
 4   starsBreakdown/5star             218 non-null    float64
 5   reviewsCount                     216 non-null    float64
 6   price                            0 non-null      float64
 7   price/value                      218 non-null    float64
 8   categoryPageData/productPosition 218 non-null    int64
dtypes: float64(6), int64(1), object(2)
memory usage: 15.5+ KB
```

In [16]:
```python
#drop brand row with missing data
df.dropna(subset=['brand'], axis=0, inplace=True)
```

In [17]:
```python
df.isnull().sum()
```

Out[17]:
```
title                              0
brand                              0
starsBreakdown/3star               0
starsBreakdown/4star               0
starsBreakdown/5star               0
reviewsCount                       2
price                            217
price/value                        0
categoryPageData/productPosition   0
dtype: int64
```

Explaratory Data Analysis

In [18]:
```python
df.dtypes
```

Out[18]:
```
title                              object
brand                              object
starsBreakdown/3star               float64
starsBreakdown/4star               float64
starsBreakdown/5star               float64
reviewsCount                       float64
price                              float64
price/value                        float64
categoryPageData/productPosition   int64
dtype: object
```

In [19]:

```python
df.head()
```

Out[19]:

	title	brand	starsBreakdown/3star	sta
0	Ferrero Rocher, 24 Count, Premium Milk Chocola...	Ferrero Rocher	0.02	0.0
1	HERSHEY'S NUGGETS Assorted Chocolate, Valentin...	HERSHEY'S	0.03	0.1
2	LEGO Icons Flower Bouquet Building Decoration ...	LEGO	0.01	0.0
3	BodyRefresh Shower Steamers Aromatherapy - 8 P...	BodyRefresh	0.07	0.1
4	JoJowell Shower Steamers Aromatherapy - 21Pcs ...	JoJowell	0.10	0.1

Questions to be answered

1.Which products and brands have the best & worst reviews?

In [20]:

```python
best_reviews_product = df.loc[df['reviewsCount'].idxmax(), 'title']
worst_reviews_product = df.loc[df['reviewsCount'].idxmin(), 'title']
best_reviews_brand = df.loc[df['reviewsCount'].idxmax(), 'brand']
worst_reviews_brand = df.loc[df['reviewsCount'].idxmin(), 'brand']


# Display Results
print("Products and Brands with the Best and Worst Reviews:")
print("\nBest Reviews Product:", best_reviews_product)
print("\nWorst Reviews Product:", worst_reviews_product)
print("\nBest Reviews Brand:", best_reviews_brand)
print("\nWorst Reviews Brand:", worst_reviews_brand)
```

```
Products and Brands with the Best and Worst Reviews:

Best Reviews Product: BAIMEI Jade Roller & Gua Sha, Face Roller Redn
ess Reducing Skin Care Tools, Massager for Face, Eyes, Neck, Relieve
Wrinkles, Self Care Gift for Men Women, Valentine's Day Gifts - Rose
Quartz

Worst Reviews Product: 32pcs Valentines Day Gifts for Kids - Valenti
nes with Mini Pop Fidget Toys Bulk - Valentine Exchange for Girls Bo
ys Classroom

Best Reviews Brand: BAIMEI

Worst Reviews Brand: Jetrvat
```

2. What brands are on the best seller list most often?

```
In [21]:    # Brands on the best seller list most often
            best_seller_brands = df['brand'].value_counts().idxmax()


            # Display Result
            print("\nBrands on the Best Seller List Most Often:", best_seller_brand
            s)
```

```
Brands on the Best Seller List Most Often: Ferrero Rocher
```

3. Are there any trends between prices and ratings?

```
In [22]:    # Trends between prices and ratings
            average_rating_by_price = df.groupby(pd.cut(df['price/value'], bins=5))
            ['starsBreakdown/5star'].mean()


            # Display Result
            print("\nAverage Rating by Price:")
            print(average_rating_by_price)
```

```
Average Rating by Price:
price/value
(1.912, 17.582]      0.752500
(17.582, 33.174]     0.781887
(33.174, 48.766]     0.780000
(48.766, 64.358]     0.836667
(64.358, 79.95]      0.790000
Name: starsBreakdown/5star, dtype: float64
```
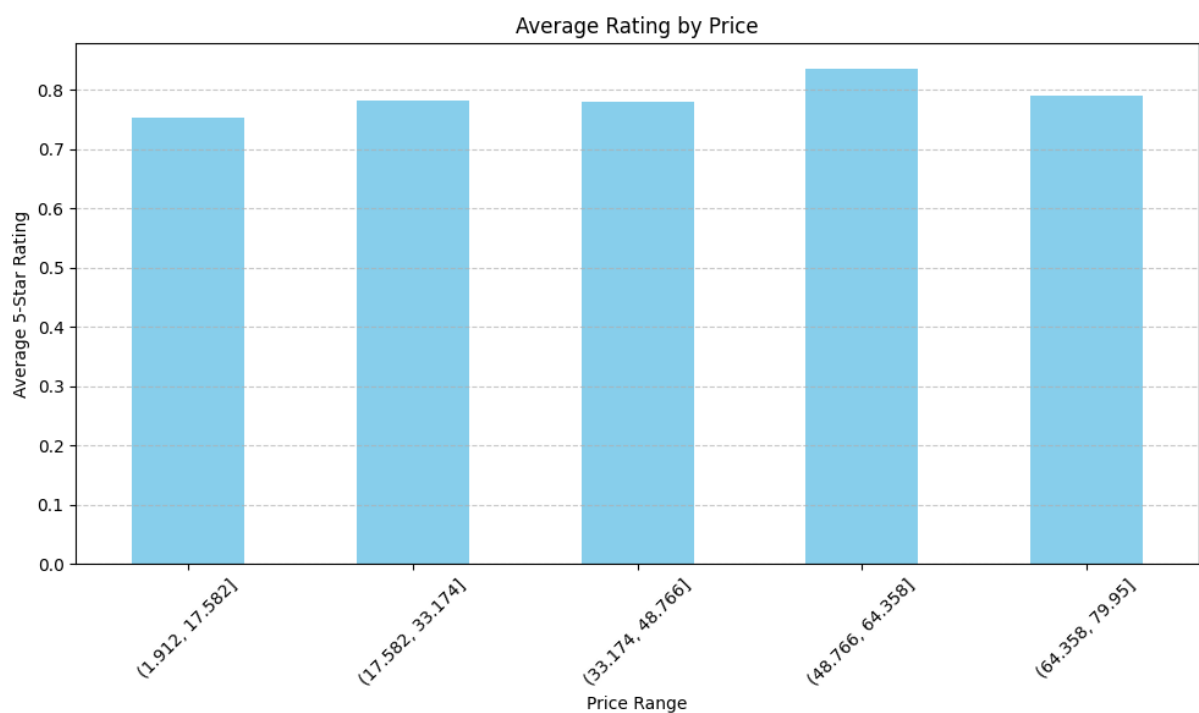
In [23]:
```python
# Visualizations
# Plotting average rating by price
plt.figure(figsize=(10, 6))
average_rating_by_price.plot(kind='bar', color='skyblue')
plt.title('Average Rating by Price')
plt.xlabel('Price Range')
plt.ylabel('Average 5-Star Rating')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```


Average Rating by Price

4. Which products are above average for 5-star ratings?

In [24]:

```python
# Products above average for 5-star ratings
above_average_5_star = df[df['starsBreakdown/5star'] > df['starsBreakdown/5star'].mean()]['title']



# Display Result
print("\nProducts Above Average for 5-Star Ratings:")
print(above_average_5_star)
```

```
Products Above Average for 5-Star Ratings:
0        Ferrero Rocher, 24 Count, Premium Milk Chocola...
1        HERSHEY'S NUGGETS Assorted Chocolate, Valentin...
2        LEGO Icons Flower Bouquet Building Decoration ...
6        LEGO Cherry Blossoms Gift for Valentine's Day,...
9        JOYIN 24 PCS Valentine's Day Heart Stress Ball...
                               ...
209      LWBDD 13" Valentines Day Gift Bags 2 Pcs with ...
210      J'FLORU Valentine's Day Gifts,6 Pack Valentine...
212      Cheerin Valentines Day Card with Envelope - Fu...
214      THEMEROL Natural Gemstone Bracelet Gifts for D...
217      Valentine's Day Gifts For Her - Rose in Glass ...
Name: title, Length: 138, dtype: object
```