Name:Sanchit A.Ichole
PRN:202401040065
Roll.No:CS3-18

```
import numpy as np import
pandas as pd import seaborn as
sns import matplotlib.pyplot as
plt

from google.colab import drive
drive.mount('/content/drive')



# Adjust the path to where your CSV is located
df = pd.read_csv('/content/drive/MyDrive/amazon_products.csv')
df.head()
```

```
Mounted at /content/drive
```

{"type":"dataframe","variable_name":"df"}

```
# 1. Find the total number of products
total_products = df.shape[0]
print(total_products)
```

```
1426337
```

```
# 2. Find the average price of products
# Average price
avg_price = df['price'].mean()
print(avg_price)
```

```
43.37540368089727
```

```
# 3. Find the product with the highest price
highest_price_product = df.loc[df['price'].idxmax()]
print(highest_price_product)
```

```
asin                                                    B077BCV1JT
title                 Overland Storage Neoxl80 Storagelibrary Lto8 SAS
imgUrl               https://m.media-amazon.com/images/I/41ksAU5SLT...
productURL                      https://www.amazon.com/dp/B077BCV1JT
stars                                                            0.0
reviews                                                            0
price                                                        19731.81
listPrice                                                        0.0
category_id                                                       55
isBestSeller                                                   False
boughtInLastMonth                                                  0
Name: 594853, dtype: object
```

```python
# 4. Find the product with the lowest price
lowest_price_product = df.loc[df['price'].idxmin()]
print(lowest_price_product)
```

```
asin                                                    B0BH6N9WMT
title                Airconic Hardside Expandable Luggage with Spin...
imgUrl               https://m.media-amazon.com/images/I/81lQQrl0By...
productURL                        https://www.amazon.com/dp/B0BH6N9WMT
stars                                                          3.8
reviews                                                          0
price                                                          0.0
listPrice                                                      0.0
category_id                                                    104
isBestSeller                                                 False
boughtInLastMonth                                                0
Name: 177, dtype: object
```

```python
# 6. Find the average rating of products
average_rating = df['reviews'].mean()
print(average_rating)
```

```
180.7508197571822
```

```python
# 7. Find how many products have a rating above 4.0
highly_rated_products = df[df['reviews'] > 4.0].shape[0]
print(highly_rated_products)
```

```
263446
```

```python
# 8. Find how many stars have a rating above 4.0
highly_rated_products = df[df['stars'] > 4.0].shape[0]
print(highly_rated_products)
```

```
1105863
```

```python
# 8. Find the most reviewed product
most_reviewed_product = df.loc[df['reviews'].idxmax()]
print(most_reviewed_product)
```

```
asin                                                    B00T0C9XRK
title                essence | Lash Princess False Lash Effect Masc...
imgUrl               https://m.media-amazon.com/images/I/61K6cQhw4E...
productURL                        https://www.amazon.com/dp/B00T0C9XRK
stars                                                          4.3
reviews                                                     346563
price                                                         4.99
listPrice                                                      0.0
category_id                                                     48
isBestSeller                                                  True
boughtInLastMonth                                           100000
Name: 1070592, dtype: object
```

```python
#9. Find the less  reviewed product
less_reviewed_product = df.loc[df['reviews'].idxmin()]
print(less_reviewed_product)
```

```
asin                                                        B014TMV5YE
title                Sion Softside Expandable Roller Luggage, Black...
imgUrl               https://m.media-amazon.com/images/I/815dLQKYIY...
productURL                       https://www.amazon.com/dp/B014TMV5YE
stars                                                             4.5
reviews                                                             0
price                                                          139.99
listPrice                                                         0.0
category_id                                                       104
isBestSeller                                                    False
boughtInLastMonth                                                2000
Name: 0, dtype: object
```

```python
# 10. Find the average discount on products
average_discount = df['listPrice'].mean()
print(average_discount)
```

```
12.449159714709777
```

```python
# 11. Products with more than 500 reviews
products_500_reviews = df[df['reviews'] > 500].shape[0]
print(products_500_reviews)
```

```
73830
```

```python
# 12. Median price in 'Electronics'
median_price_electronics = df[df['category_id'] == '104']
['price'].median()
print(median_price_electronics)
```

```
nan
```

```python
#13. Find the number of products in each category
products_per_category = df['category_id'].value_counts()
print(products_per_category)
```

```
category_id
91     28619
84     24660
270    20846
114    19822
118    18994
...
195       76
186       50
185       42
102       40
```

```
194       22
Name: count, Length: 248, dtype: int64
# 14. Percentage of products that have a discount
discounted_products_percentage = (df['listPrice'].notna().mean()) *
100
print(discounted_products_percentage)
100.0
# 15. Product with maximum discount
max_discount_product = df.loc[df['listPrice'].idxmax()]
print(max_discount_product)
asin                                                        B07QXM2V6X
title                iRobot Roomba s9+ (9550) Self Emptying Robot V...
imgUrl               https://m.media-amazon.com/images/I/718mLGnc2-...
productURL                        https://www.amazon.com/dp/B07QXM2V6X
stars                                                              3.7
reviews                                                           6077
price                                                           699.99
listPrice                                                       999.99
category_id                                                        175
isBestSeller                                                     False
boughtInLastMonth                                                 1000
Name: 47276, dtype: object
# 15. Products priced between $100 and $500
products_in_range = df[(df['price'] >= 100) & (df['price'] <=
500)].shape[0]
print(products_in_range)
97120
# 16. Standard deviation of prices
price_std_dev = df['price'].std()
print(price_std_dev)
130.28929583109004
# 17. Group by category and find average ratings
average_ratings_per_category = df.groupby('category_id')
['reviews'].mean()
print(average_ratings_per_category)
category_id
1        57.457798
2        295.716384
3        208.895182
4        138.122951
5        0.000000
```

```
          ..
262      0.000000
263      0.000000
264      0.000000
265      0.000000
270    480.709585
Name: reviews, Length: 248, dtype: float64
# 18. Sort products by price descending
products_sorted_by_price = df.sort_values(by='price', ascending=False)
print(products_sorted_by_price)
               asin                                                title


 594853  B077BCV1JT   Overland Storage Neoxl80 Storagelibrary Lto8 SAS


\


108907  B0CJHXQS8R                                    replicas Marinas

994047  B0000516QJ  CISCO Systems 1 Port ATM Enhanced Oc12/Stm4 Si...

994048  B0000516Q1  Cisco Systems 7140 Router Dual 10/100 Fe Dual ...

868945  B08XMXFW1Q  PARTNERS BRAND Corrugated Trash Can Plain - 40...

...            ...                                                ...

483778  B08ZJSPQBG                        Unisex-Child Lorena Sneaker

808754  B08WJ5YFDT  Outsunny 6'x8' Outdoor Storage Shelter with Ro...

808749  B077V6QXLG                    Makita 198494-2 Steel Bucket
739915  B0CJR94MCW  PlayVital Fearlessness Custom Vinyl Skins for ...
235128  B09Y56XMV1  UOhost Hat Press 6.2 x 3.14 Inch Curved Elemen...
                                                      imgUrl  \
594853  https://m.media-amazon.com/images/I/41ksAU5SLT..
108907  https://m.media-amazon.com/images/I/81R-Xq0X+w..
994047  https://m.media-amazon.com/images/I/11Tihbi9s1..
994048  https://m.media-amazon.com/images/I/21rHMqCooA..
868945  https://m.media-amazon.com/images/I/21qsXigx3C..
...                                                   ..
483778  https://m.media-amazon.com/images/I/81baell+GI..
808754  https://m.media-amazon.com/images/I/613jw-xwki..
808749  https://m.media-amazon.com/images/I/51JbZxDRqO..
739915  https://m.media-amazon.com/images/I/61qGffyjE9..
235128  https://m.media-amazon.com/images/I/61lpFCpPQH..
                                productURL  stars  reviews     price
```

```
\
594853  https://www.amazon.com/dp/B077BCV1JT    0.0         0  19731.81

108907  https://www.amazon.com/dp/B0CJHXQS8R    0.0         0  19400.00

994047  https://www.amazon.com/dp/B0000516QJ    0.0         0  16468.70

994048  https://www.amazon.com/dp/B0000516Q1    0.0         0  12629.66

868945  https://www.amazon.com/dp/B08XMXFW1Q     0.0         0  12519.96

...                                       ...    ...       ...       ...

483778  https://www.amazon.com/dp/B08ZJSPQBG    4.4         0      0.00

808754  https://www.amazon.com/dp/B08WJ5YFDT    3.9         0      0.00

808749  https://www.amazon.com/dp/B077V6QXLG    4.1         0      0.00

739915  https://www.amazon.com/dp/B0CJR94MCW    0.0         0      0.00
235128  https://www.amazon.com/dp/B09Y56XMV1    2.0         6      0.00
        listPrice  category_id  isBestSeller  boughtInLastMonth
594853        0.0           55         False                  0
108907        0.0            2         False                  0
994047        0.0           54         False                  0
994048        0.0           54         False                  0
868945        0.0          160         False                  0
...           ...          ...           ...                ...
483778        0.0           97         False                  0
808754        0.0          173         False                  0
808749        0.0          173         False                  0
739915        0.0          262         False                  0
235128        0.0            4         False                  0
[1426337 rows x 11 columns]
# 19. Check missing values
missing_values = df.isnull().sum()
print(missing_values)
asin                0
title               1
imgUrl              0
productURL          0
stars               0
reviews             0
price               0
listPrice           0
category_id         0
isBestSeller        0
boughtInLastMonth   0
dtype: int64
```

```
# 20. Replace missing ratings with average rating
df['reviews'].fillna(df['reviews'].mean(), inplace=True)
```

```
<ipython-input-43-adcec3268bcb>:2: FutureWarning: A value is trying to
be set on a copy of a DataFrame or Series through chained assignment
using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never
work because the intermediate object on which we are setting values
always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
df[col].method(value) instead, to perform the operation inplace on the
original object.


  df['reviews'].fillna(df['reviews'].mean(), inplace=True)
```