

EDS THEORY ACTIVITY NO 1

Name: Jay Ankush Galande

Roll No:-CS2-59

Batch:-CS23

PRN:-202401040371

- Importing Amazon Product Dataset

```

[19] import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

df = pd.read_csv("/content/sample_data/marketing_sample_for_amazon_com-ecommerce_20200101_20200131_10k_data.csv")

[30] # Clean price columns
for col in ['Selling Price', 'List Price']:
    df[col] = df[col].replace(['\$', ' ', regex=True)
    df[col] = pd.to_numeric(df[col], errors='coerce')
# Clean price columns: Remove dollar signs, commas, and convert to float
for col in ['Selling Price', 'List Price']:
    df[col] = df[col].astype(str).str.replace(['\$', ' ', regex=True)
    df[col] = pd.to_numeric(df[col], errors='coerce')

```

1. Display first 5 rows of the dataset

```

print(df.head())

```

```

      Uniq Id \
0  4c69b61db1fc16e7013b43fc926e502d
1  66d49bbcd043f5be260fa9f7fbff5957
2  2c55cae269aebf53838484b0d7dd931a
3  18018b6bc416dab347b1b7db79994afa
4  e04b990e95bf73bbe6a3fa09785d7cd0

      Product Name  Brand Name  Asin \
0  DB Longboards CoreFlex Crossbow 41" Bamboo Fib...  NaN  NaN
1  Electronic Snap Circuits Mini Kits Classpack, ...  NaN  NaN
2  3Doodler Create Flexy 3D Printing Filament Ref...  NaN  NaN
3  Guillow Airplane Design Studio with Travel Cas...  NaN  NaN
4  Woodstock- Collage 500 pc Puzzle  NaN  NaN

      Category  Upc  Ean  Code  List Price \
0  Sports & Outdoors | Outdoor Recreation | Skate...  NaN  NaN
1  Toys & Games | Learning & Education | Science ...  NaN  NaN
2  Toys & Games | Arts & Crafts | Craft Kits  NaN  NaN
3  Toys & Games | Hobbies | Models & Model Kits |...  NaN  NaN
4  Toys & Games | Puzzles | Jigsaw Puzzles  NaN  NaN

      Selling Price  Quantity  Model Number  ... \
0  $237.68  NaN  NaN  ...
1  $99.95  NaN  55324  ...
2  $34.99  NaN  NaN  ...
3  $28.91  NaN  142  ...
4  $17.49  NaN  62151  ...

      Product Url  Stock  Product Details \
0  https://www.amazon.com/DB-longboards-CoreFlex-...  NaN  NaN
1  https://www.amazon.com/Electronic-Circuits-Cla...  NaN  NaN
2  https://www.amazon.com/3Doodler-Plastic-Innova...  NaN  NaN
3  https://www.amazon.com/Guillow-Airplane-Design...  NaN  NaN
4  https://www.amazon.com/Woodstock-Collage-500-p-...  NaN  NaN


      Dimensions  Color  Ingredients  Direction To Use  Is Amazon Seller \
0  NaN  NaN  NaN  NaN  Y
1  NaN  NaN  NaN  NaN  Y
2  NaN  NaN  NaN  NaN  Y
3  NaN  NaN  NaN  NaN  Y
4  NaN  NaN  NaN  NaN  Y

      Size  Quantity  Variant  Product Description
0  NaN  NaN  NaN
1  NaN  NaN  NaN
2  NaN  NaN  NaN
3  NaN  NaN  NaN
4  NaN  NaN  NaN

```


[5 rows x 28 columns]

2. Total unique products:

```
✓ 0s  unique_products = df['Uniq Id'].nunique()  
print("1. Total unique products:", unique_products)
```

```
⇒ 1. Total unique products: 10002
```

3. Unique brands

```
✓ 0s  unique_brands = df['Brand Name'].dropna().unique()  
print("2. Number of unique brands:", len(unique_brands))
```


```
⇒ 2. Number of unique brands: 0
```

4. Average selling price

```
✓ 0s  avg_selling_price = df['Selling Price'].mean()  
print("3. Average selling price:", round(avg_selling_price, 2))
```

```
⇒ 3. Average selling price: 39.75
```

5. Products listed by Amazon

```
✓ 0s  amazon_products_count = df[df['Is Amazon Seller'] == 'Y'].shape[0]  
print("4. Number of products listed by Amazon:", amazon_products_count)
```

```
⇒ 4. Number of products listed by Amazon: 9723
```

6. Product with highest selling price

```
most_expensive_product = df.sort_values(by='Selling Price', ascending=False).iloc[0]
print("5. Most expensive product:")
print(most_expensive_product)
```

5. Most expensive product:

Uniq Id	bb96646206fa6a273374f57fd1d91d34
Product Name	3B Scientific U100061 Ultrasonic Continuous Wa...
Brand Name	NaN
Asin	NaN
Category	Toys & Games Learning & Education Science ...
Upc Ean Code	NaN
List Price	NaN
Selling Price	5332.0
Quantity	NaN
Model Number	U100061
About Product	NaN
Product Specification	Shipping Weight: 11 pounds (View shipping rate...
Technical Details	Other Documentation User Manual Ultrasonic c...
Shipping Weight	11 pounds
Product Dimensions	NaN
Image	https://images-na.ssl-images-amazon.com/images...
Variants	NaN
Sku	NaN
Product Url	https://www.amazon.com/3B-Scientific-U100061-U...
Stock	NaN
Product Details	NaN
Dimensions	NaN
Color	NaN
Ingredients	NaN
Direction To Use	NaN
Is Amazon Seller	Y
Size Quantity Variant	NaN
Product Description	NaN

Name: 5625, dtype: object

7. Top 5 product categories

```
top_categories = df['Category'].value_counts().head(5)
print("6. Top 5 product categories:")
print(top_categories)
```

6. Top 5 product categories:

Category	
Toys & Games Games & Accessories Board Games	284
Toys & Games Puzzles Jigsaw Puzzles	274
Toys & Games Stuffed Animals & Plush Toys Stuffed Animals & Teddy Bears	252
Toys & Games Toy Figures & Playsets Action Figures	235
Toys & Games Dolls & Accessories Dolls	193

Name: count, dtype: int64

8. Products in 'Toys & Games'

```
toys_games_count = df['Category'].str.contains('Toys & Games', na=False).sum()
print("7. Number of products in 'Toys & Games':", toys_games_count)
```

7. Number of products in 'Toys & Games': 6662

9. Avg. selling price per category

```
avg_price_category = df.groupby('Category')['Selling Price'].mean()
print("8. Average selling price per category:")
print(avg_price_category.head())
```

8. Average selling price per category:

Category	
Arts, Crafts & Sewing Beading & Jewelry Making Beading Supplies Fuse & Perler Beads	11.445000
Arts, Crafts & Sewing Beading & Jewelry Making Beads & Bead Assortments	14.030000
Arts, Crafts & Sewing Crafting Craft Supplies	11.623333
Arts, Crafts & Sewing Crafting Craft Supplies Adhesives Art Glues & Pastes	26.315000
Arts, Crafts & Sewing Crafting Craft Supplies Cutting Tools Art Knives & Blades	6.675000

Name: Selling Price, dtype: float64

10. Categories with more than 100 products

```
categories_gt_100 = df['Category'].value_counts()[lambda x: x > 100]
print("9. Categories with more than 100 products:")
print(categories_gt_100)
```

9. Categories with more than 100 products:

Category	
Toys & Games Games & Accessories Board Games	284
Toys & Games Puzzles Jigsaw Puzzles	274
Toys & Games Stuffed Animals & Plush Toys Stuffed Animals & Teddy Bears	252
Toys & Games Toy Figures & Playsets Action Figures	235
Toys & Games Dolls & Accessories Dolls	193
Toys & Games Party Supplies	193
Toys & Games Hobbies Remote & App Controlled Vehicles & Parts Remote & App Controlled Vehicle Parts	178
Toys & Games Dress Up & Pretend Play Costumes	158
Toys & Games Novelty & Gag Toys	141
Toys & Games Arts & Crafts	140
Toys & Games Building Toys Building Sets	122
Toys & Games Learning & Education Science Kits & Toys	121
Clothing, Shoes & Jewelry Costumes & Accessories Kids & Baby Girls Costumes	120
Clothing, Shoes & Jewelry Costumes & Accessories Women Costumes & Cosplay Apparel Costumes	116
Toys & Games Games & Accessories	110
Toys & Games Games & Accessories Card Games	109

Name: count, dtype: int64

11. Category with lowest average price

```
lowest_price_category = avg_price_category.idxmin()
print("10. Category with the lowest average price:", lowest_price_category)
```

10. Category with the lowest average price: Sports & Outdoors | Sports & Fitness | Team Sports | Ice Hockey | Accessories | Skate Accessories

12. Missing values per column

```
missing_values = df.isnull().sum()
print("11. Missing values in each column:")
print(missing_values)
```

```
11. Missing values in each column:
Uniq Id          0
Product Name     0
Brand Name      10002
Asin             10002
Category         830
Upc Ean Code     9968
List Price       10002
Selling Price    492
Quantity         10002
Model Number     1772
About Product    273
Product Specification 1632
Technical Details 790
Shipping Weight  1138
Product Dimensions 9523
Image            0
Variants         7524
Sku              10002
Product Url      0
Stock            10002
Product Details  10002
Dimensions       10002
Color            10002
Ingredients       10002
Direction To Use 10002
Is Amazon Seller 0
Size Quantity Variant 10002
Product Description 10002
dtype: int64
```

13. Discount = List Price - Selling Price

```
df['Discount'] = df['List Price'] - df['Selling Price']
print("14. Added 'Discount' column.")
```

```
14. Added 'Discount' column.
```

14. Remove rows with missing/zero selling price

```
df = df[df['Selling Price'] > 0]
print("15. Removed products with zero or missing selling price.")
```

```
15. Removed products with zero or missing selling price.
```

15. Classify products as 'Expensive' or 'Affordable'

```
df['Price Category'] = np.where(df['Selling Price'] > 100, 'Expensive', 'Affordable')
print("16. Added 'Price Category' column (Expensive/Affordable)")
```

```
16. Added 'Price Category' column (Expensive/Affordable)
```

16. % of products with non-null dimensions

```
✓ 0s  dimensions_percentage = df['Dimensions'].notnull().mean() * 100
print("17. Percentage of products with dimensions info:", round(dimensions_percentage, 2), "%")
```

17. Percentage of products with dimensions info: 0.0 %

17. Most common model number

```
✓ 0s  common_model_number = df['Model Number'].value_counts().idxmax()
print("18. Most common model number:", common_model_number)
```

18. Most common model number: -

18. Number of products with a discount greater than \$50

```
✓ 0s  discount_gt_50 = df[df['Discount'] > 50].shape[0]
print("21. Number of products with discount > $50:", discount_gt_50)
```

21. Number of products with discount > \$50: 0

19. Count of affordable vs expensive products

```
✓ 0s  price_category_counts = df['Price Category'].value_counts()
print("22. Count of products by Price Category:")
print(price_category_counts)
```

22. Count of products by Price Category:

Price Category	
Affordable	8912
Expensive	598

Name: count, dtype: int64

20. Average discount per category

```
✓ 0s  avg_discount_per_category = df.groupby('Category')['Discount'].mean().sort_values(ascending=False).head()
print("24. Average discount per category:")
print(avg_discount_per_category)
```

24. Average discount per category:

Category				
Arts, Crafts & Sewing	Beading & Jewelry Making	Beaded Supplies	Fuse & Perler Beads	NaN
Arts, Crafts & Sewing	Beaded & Jewelry Making	Beads & Bead Assortments		NaN
Arts, Crafts & Sewing	Crafting	Craft Supplies		NaN
Arts, Crafts & Sewing	Crafting	Craft Supplies	Adhesives Art Glues & Pastes	NaN
Arts, Crafts & Sewing	Crafting	Craft Supplies	Cutting Tools Art Knives & Blades	NaN

Name: Discount, dtype: float64