

# Assignment 1

---



**Name : Akash A Barage**

**PRN : 202401120075**

**Class : CS8**

**Roll No : CS8-66**

**DATASET : Grocery**

**DATASET Link :**

**[https://drive.google.com/file/d/1av25qLhBwsgggiMybxn5\\_QG3owT\\_6lX-/view?usp=drivesdk](https://drive.google.com/file/d/1av25qLhBwsgggiMybxn5_QG3owT_6lX-/view?usp=drivesdk)**

**Problem Statements:**

---

```

1 import pandas as pd
2 import numpy as np
3
4 # Load the dataset
5 df = pd.read_csv("Plant.csv")
6
7 # Strip extra spaces in column names
8 df.columns = df.columns.str.strip()
9
10 # Print all column names
11 print("Column names:", df.columns.tolist())
12
13 # Handle numeric conversion safely
14 for col in ['Price', 'DiscountedPrice', 'Review_Count', 'Rating']:
15     if col in df.columns:
16         df[col] = pd.to_numeric(df[col], errors='coerce')
17
18 # Drop rows with missing Price
19 if 'Price' in df.columns:
20     df.dropna(subset=['Price'], inplace=True)
21
22 # 1. Total number of products
23 print("\n1. Total number of products:", len(df))
24
25 # 2. Unique categories
26 if 'Category' in df.columns:
27     print("\n2. Unique categories:", df['Category'].unique())
28     print("Category names:", df['Category'].unique())
29
30 # 3. Average price
31 if 'Price' in df.columns:
32     print("\n3. Average product price: ₹", round(df['Price'].mean(), 2))
33
34 # 4 & 5. Most and least expensive products
35 print("\n4. Most expensive product:\n", df[df['Price'] == df['Price'].max()])
36 print("\n5. Least expensive product:\n", df[df['Price'] == df['Price'].min()])
37

```

```

38 # 6. Average price by category
39 if 'Price' in df.columns and 'Category' in df.columns:
40     print("\n6. Average price by category:\n", df.groupby('Category')['Price'].mean())
41
42 # 8. Count of products per category
43 if 'Category' in df.columns:
44     print("\n8. Product count per category:\n", df['Category'].value_counts())
45
46 # 9. Top 5 most reviewed products
47 if 'Review_Count' in df.columns:
48     print("\n9. Top 5 most reviewed products:\n", df.sort_values('Review_Count', ascending=False).head(5))
49
50 # 14. Total reviews per category
51 if 'Review_Count' in df.columns and 'Category' in df.columns:
52     print("\n14. Total reviews per category:\n", df.groupby('Category')['Review_Count'].sum())
53
54 # 15. Products priced between 100 and 300
55 if 'Price' in df.columns:
56     print("\n15. Products priced between ₹100 and ₹300:\n", df[(df['Price'] >= 100) & (df['Price'] <= 300)])
57
58 # 16. Products containing "Apple"
59 if 'Name' in df.columns:
60     print("\n16. Products containing 'Apple':\n", df[df['Name'].str.contains("Apple", case=False, na=False)])
61
62 # 17. Std deviation of price
63 if 'Price' in df.columns:
64     print("\n17. Price standard deviation:", round(df['Price'].std(), 2))
65
66 # 18. Median price
67 if 'Price' in df.columns:
68     print("\n18. Median price:", df['Price'].median())
69
70 # 19. 5 Cheapest products
71 if 'Price' in df.columns:
72     print("\n19. 5 Cheapest products:\n", df.sort_values('Price').head(5))
73
74 # 20. Add price category
75 if 'Price' in df.columns:
76     df['Price_Category'] = pd.cut(df['Price'], bins=[0, 100, 500, np.inf], labels=['Budget', 'Mid-Range', 'Premium'])
77

```

# Solution :

---

```
Column names: ['Name', 'Brand', 'Price', 'DiscountedPrice', 'Category', 'SubCategory', 'Quantity', 'Description', 'BreadCrumbs']
```

1. Total number of products: 5188

2. Unique categories: 29

```
Category names: ['Grocery' 'Personal Care' 'Packaged Food' 'DMart Grocery'
'Dairy & Beverages' 'Fruits & Vegetables' 'Home & Kitchen' 'Specials'
'Kitchen Aprons' 'Appliances' 'Wonderchef'
'Water Purifiers & Accessories' 'Tableware' 'Butterfly' nan
'Joyo Plastics' 'Geep' 'Zebronics' 'Beauty & Cosmetics' 'Syska' 'Pigeon'
'Clothing & Accessories' 'School Supplies' 'Backpacks'
'Computer Accessories' 'Plant Container' 'Motorbike Helmets'
'Smartwatches' 'Raincoat' 'Baby & Kids']
```

3. Average product price: ₹ 344.03

4. Most expensive product:

	Name	...	BreadCrums
2796	Philips UV-C Disinfection System	...	Home & Kitchen > Disinfectants

[1 rows x 9 columns]

5. Least expensive product:

	Name	...	BreadCrums
2230	Kopiko Cappuccino Coffee Candies	...	Packaged Food > Sweets

[1 rows x 9 columns]

6. Average price by category:

Category	
Appliances	2469.207547
Baby & Kids	224.750000
Backpacks	2188.384615
Beauty & Cosmetics	399.739130
Butterfly	4248.500000
Clothing & Accessories	372.134021
Computer Accessories	349.000000
DMart Grocery	343.612903
Dairy & Beverages	225.002657
Fruits & Vegetables	79.615385
Geep	288.000000
Grocery	311.383333
Home & Kitchen	476.566330
Joyo Plastics	907.500000
Kitchen Aprons	285.000000
Motorbike Helmets	1110.750000
Packaged Food	150.768889
Personal Care	296.426196
Pigeon	3595.000000
Plant Container	116.782609
Raincoat	937.363636
School Supplies	215.202899
Smartwatches	5999.000000
Specials	774.002674
Syska	799.000000
Tableware	96.944444
Water Purifiers & Accessories	875.000000
Wonderchef	5300.000000
Zebronics	801.000000

Name: Price, dtype: float64

#### 8. Product count per category:

```
Category
Personal Care          1233
Packaged Food          1125
Home & Kitchen         880
Grocery                840
Dairy & Beverages      429
Specials               187
Clothing & Accessories  97
Fruits & Vegetables     78
School Supplies        69
Appliances             53
Beauty & Cosmetics      46
Raincoat               33
DMart Grocery          31
Plant Container         23
Tableware              18
Backpacks              13
Zebronics              5
Motorbike Helmets      4
Baby & Kids             4
Kitchen Aprons         3
Geep                   3
Computer Accessories   2
Butterfly              2
Joyo Plastics          2
Wonderchef             1
Water Purifiers & Accessories 1
Syska                  1
Pigeon                 1
Smartwatches           1
Name: count, dtype: int64
```

15. Products priced between ₹100 and ₹300:

		Name	...	
BreadCrumbs				
1	Premia Badam (Almonds)	...	Grocery > Gro	
cery/Dry Fruits				
2	Premia Badam (Almonds)	...	Grocery > Gro	
cery/Dry Fruits				
7	Rajma White	...	G	
rocery > Pulses				
8	Rajma Kashmiri Red	...	G	
rocery > Pulses				
11	Tata Sampann Chana Dal	...		
Grocery > Dals				
...		...	...	
...				
5178	Glue Stick	...	School Supplies	
> School Needs				
5180	Camel Water Colour Paint Cake Set - 24 Shades	...	School Supplies	
> School Needs				
5181	Future Books Colouring Book	...	School Supplies	
> School Needs				
5183	Navneet Youva 6 Subject Spiral Book - Assorted	...	School Supplies	
> School Needs				
5187	Navneet Youva Drawing Book - Assorted	...	School Supplies	
> School Needs				

[2152 rows x 9 columns]

## 16. Products containing 'Apple':

	Name	...	
BreadCrumbs			
1175	Budweiser Green Apple Non-Alcoholic Beer Can	...	Dairy & B
everages > Beverages			
1183	Cravova Green Apple Mojito	...	Dairy & B
everages > Beverages			
1223	Real Fruit Power Apple	...	Dairy & B
everages > Beverages			
1231	Tropicana Apple Delight	...	Dairy & B
everages > Beverages			
1268	Cipla Prolyte ORS - Apple Flavour	...	Dairy & B
everages > Beverages			
1293	Monin Green Apple Flavoured Syrup	...	Dairy & B
everages > Beverages			
1979	Borges Apple Cider Vinegar With The Mother	...	Packaged Fo
od > Ketchup & Sauce			
1980	American Garden Apple Cider Vinegar	...	Packaged Fo
od > Ketchup & Sauce			
2234	Tic Tac Apple Treat	...	Pa
ckaged Food > Sweets			
2397	Apple Royal Gala (Seb)	...	
Fruits & Vegetables			
2402	Apple Red Delicious	...	
Fruits & Vegetables			
2405	Fresh Apple Green	...	
Fruits & Vegetables			
2412	Fresh Apple (Seb) Washington	...	
Fruits & Vegetables			
2417	Apple Shimla (Seb)	...	
Fruits & Vegetables			
3204	Nayasa Pineapple PET Bottle - Assorted	...	Home & Kitchen >
Cookware & Serveware			
3792	WOW Apple Cider Vinegar Face Wash	...	Perso
hal Care > Face Care			
4030	WOW Apple Cider Vinegar Shampoo	...	Perso
hal Care > Hair Care			
5080	Britannia Pineapple Cake	...	Pa
ckaged Food > Bakery			
5139	Nestle Cerelac Wheat Apple Cherry	...	
Baby & Kids			

[19 rows x 9 columns]

17. Price standard deviation: 601.03

18. Median price: 175.0

19. 5 Cheapest products:

	Name	...
	BreadCrumbs	
2230	Kopiko Cappuccino Coffee Candies	...
	Packaged Food > Sweets	Pac
1344	Parle-G Original Gluco Biscuits	...
	Biscuits & Cookies	Packaged Food >
2368	Knorr Hot & Sour Veg Cup-A-Soup	...
	Packaged Food	
1745	4700 BC Instant Barbeque Popcorn	...
	Packaged Food > Ready To Cook	
	> Snacks & Farsans	
2369	Ching's Manchow Instant Soup	...
	Packaged Food	

[5 rows x 9 columns]

20. Products with Price Category:

	Name	Price	Price_Category
0	Premia Badam (Almonds)	451.0	Mid-Range
1	Premia Badam (Almonds)	109.0	Mid-Range
2	Premia Badam (Almonds)	202.0	Mid-Range
3	Nutraaj California Almonds (Badam)	599.0	Premium
4	Nutraaj California Almonds (Badam)	1549.0	Premium