Part 2 Data Optimization

Prepared by : BERNICE LIM JING XUAN (A22EC0038)

Step 1 : Install and Import Libraries

```
In [ ]:  !pip install pandas
         import pandas as pd
         import re
         import time
         import tracemalloc
         import psutil
         import io
         #import os
```

Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.2)
Requirement already satisfied: numpy>=1.23.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)

Step 2 : Upload Execl Files

```
In [ ]:  from google.colab import files
         uploaded = files.upload()
```

Choose Files   No file chosen   Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

Saving pandas_cleaned_dataset.csv to pandas_cleaned_dataset.csv

Step 3 : Load and Display Dataset, Checking on Total Numbers of Rows and Columns

```
In [ ]:  %%time

         tracemalloc.start()
         start_time = time.perf_counter()
         total_rows = 0

         filename = list(uploaded.keys())[0]
         file_content = uploaded[filename].decode('utf-8')
         df_cleaned = pd.read_csv(io.StringIO(file_content), keep_default_na=False)

         total_rows = df_cleaned.shape[0]
         display(df_cleaned.head(10))
         print(f"Total rows: {df_cleaned.shape[0]}")
         print(f"Total columns: {df_cleaned.shape[1]}\n\n\n")

         current, peak = tracemalloc.get_traced_memory()
         end_time = time.perf_counter()
         tracemalloc.stop()

         execution_time = end_time - start_time
         throughput = total_rows / execution_time

         print("==================== Performance ====================\n")
         print(f"Total rows processed: {total_rows}")
         print(f"Code Execution time: {execution_time:.4f} seconds")
         print(f"Throughput: {throughput:.2f} rows per second")
         print(f"Current memory usage: {current / 10**6:.4f} MB")
         print(f"Peak memory usage: {peak / 10**6:.4f} MB")

         cpu_usage = psutil.cpu_percent(interval=1)
         print(f"CPU usage: {cpu_usage}%")

         print("=====================================================")

         print("\nTotal time for this cell (Including time to display the performance):")
```

	Product Name	Price	Location	Quantity Sold	Total Reviews	Category
0	❣️ SAKA GLOWING FOUNDATION SPF 50+ / FW FATIN W...	3.99	PENANG	55	9	BEAUTY & SKINCARE
1	BIO-ESSENCE BIO-GOLD 24K RADIANCE/WHITENING/B...	3.50	JOHOR	46	10	BEAUTY & SKINCARE
2	L'OCCITANE IMMORTELLE DIVINE FOAMING CLEANSING...	1.50	SELANGOR	13	1	BEAUTY & SKINCARE
3	YOUBUY FRECKLE CREAM EFFECTIVELY REMOVE MELASM...	5.45	CHINA	13	0	BEAUTY & SKINCARE
4	DT37 JOMTAM JOLYUM NICOTIMIDE AMINO ACID FACIA...	6.15	MELAKA	175	40	BEAUTY & SKINCARE
5	❤️ 88HOME ❤️ JOMTAM JOLYUM NICOTIMIDE AMINO ACID...	6.19	MELAKA	29	2	BEAUTY & SKINCARE
6	DT37 VEZE MEN'S VOLCANIC MUD FACIAL CLEANSER B...	5.44	MELAKA	129	46	BEAUTY & SKINCARE
7	NEVEA MEN ROLL ON 500ML	5.00	WP KUALA LUMPUR	25	4	BEAUTY & SKINCARE
8	❤️ 88HOME ❤️ VEZE MEN'S VOLCANIC MUD FACIAL CLEA...	5.48	MELAKA	41	11	BEAUTY & SKINCARE
9	DT37 HYMEY'S FACIAL CLEANSER CREAM WHITENING M...	1.53	MELAKA	1300	246	BEAUTY & SKINCARE

```
Total rows: 113596
Total columns: 6
```

```
==================== Performance ====================

Total rows processed: 113596
Code Execution time: 0.8587 seconds
Throughput: 132290.82 rows per second
Current memory usage: 87.7873 MB
Peak memory usage: 161.4432 MB
CPU usage: 15.7%
=====================================================

Total time for this cell (Including time to display the performance):
CPU times: user 734 ms, sys: 118 ms, total: 852 ms
Wall time: 1.89 s
```

Step 4 : Dividing Products into 4 Categories Based on Price using Pandas

```python
%%time

tracemalloc.start()
start_time = time.perf_counter()

total_rows = df_cleaned.shape[0]

df_price = df_cleaned[df_cleaned["Price"] > 0]

q1 = df_price["Price"].quantile(0.25)
q3 = df_price["Price"].quantile(0.75)
iqr = q3 - q1
lower_bound = q1 - 1.5 * iqr
upper_bound = q3 + 1.5 * iqr

df_cleared = df_price[(df_price["Price"] >= lower_bound) & (df_price["Price"] <= upper_bound)]

min_price = df_cleared["Price"].min()
max_price = df_cleared["Price"].max()
print(f"Min Price: {min_price}")
print(f"Max Price: {max_price}")

price_range = (max_price - min_price) / 4
bound1 = min_price + price_range
bound2 = min_price + 2 * price_range
bound3 = min_price + 3 * price_range

group1 = df_price[df_price["Price"] <= bound1].sort_values("Price")
group2 = df_price[(df_price["Price"] > bound1) & (df_price["Price"] <= bound2)].sort_values("Price")
group3 = df_price[(df_price["Price"] > bound2) & (df_price["Price"] <= bound3)].sort_values("Price")
group4 = df_price[df_price["Price"] > bound3].sort_values("Price")

print(f"\nGroup 1 (Budget Friendly Price): {group1.shape[0]} products")
print(f"Group 2 (Affordable Price): {group2.shape[0]} products")
print(f"Group 3 (Mid-Range Price): {group3.shape[0]} products")
print(f"Group 4 (Premium Price): {group4.shape[0]} products")

def print_category_counts(df_group):
    print(f" 📊 Category count:")
    category_counts = df_group["Category"].value_counts()
    for category, count in category_counts.items():
        print(f"- {category}: {count} products")

print("\nGroup 1 (Budget Friendly Price):")
display(group1.head(10))
```

```python
print(f"Total rows: {group1.shape[0]}")
print(f"Total columns: {group1.shape[1]}\n")
print_category_counts(group1)

print("\nGroup 2 (Affordable Price):")
display(group2.head(10))
print(f"Total rows: {group2.shape[0]}")
print(f"Total columns: {group2.shape[1]}\n")
print_category_counts(group2)

print("\nGroup 3 (Mid-Range Price):")
display(group3.head(10))
print(f"Total rows: {group3.shape[0]}")
print(f"Total columns: {group3.shape[1]}\n")
print_category_counts(group3)

print("\nGroup 4 (Premium Price):")
display(group4.head(10))
print(f"Total rows: {group4.shape[0]}")
print(f"Total columns: {group4.shape[1]}\n")
print_category_counts(group4)

current, peak = tracemalloc.get_traced_memory()
end_time = time.perf_counter()
tracemalloc.stop()

execution_time = end_time - start_time
throughput = total_rows / execution_time

print("\n==================== Performance ====================")
print(f"Total rows processed: {total_rows}")
print(f"Code Execution time: {execution_time:.4f} seconds")
print(f"Throughput: {throughput:.2f} rows per second")
print(f"Current memory usage: {current / 10**6:.4f} MB")
print(f"Peak memory usage: {peak / 10**6:.4f} MB")
print(f"CPU usage: {psutil.cpu_percent(interval=1)}%")
print("====================================================")

print("\nTotal time for this cell (Including time to display the performance):")
```

```
Min Price: 0.05
Max Price: 172.82

Group 1 (Budget Friendly Price): 65992 products
Group 2 (Affordable Price): 22626 products
Group 3 (Mid-Range Price): 10327 products
Group 4 (Premium Price): 14651 products

Group 1 (Budget Friendly Price):
```

	Product Name	Price	Location	Quantity Sold	Total Reviews	Category
54273	LIVE TRACKING #NOT FOR SALES	0.05	KELANTAN	0	1	HOME & LIVING
65102	【MALAYSIA 3PIN PLUG】3L ELECTRIC KETTLE CONSTAN...	0.05	SELANGOR	0	0	HOME APPLIANCES
81087	1PCS 0.5MM BALL GEL INK PEN MATA TAJAM NEEDLE ...	0.10	SELANGOR	33500	1354	STATIONERY
27877	HEALTH TREE REISSUE PRODUCT LINK, CONTACT SELL...	0.10	CHINA	10	3	HEALTH & WELLNESS
31656	BEFREE BESEEN PLUS VITAMIN EYE-BRAIN BOOSTER (...	0.10	JOHOR	0	0	HEALTH & WELLNESS
31640	SPOT BESEEN PLUS EYE CARE + BRAIN BOOSTER 干眼症 ...	0.10	JOHOR	0	0	HEALTH & WELLNESS
53688	READY STOCK MICKEY MOUSE MASCOT COSTUME MASKOT...	0.11	WP KUALA LUMPUR	49	1	HOME & LIVING
53809	✨ HARI RAYA HIASAN GANTUNG KRAF KAYU PERHIASAN...	0.12	WP KUALA LUMPUR	24	1	HOME & LIVING
30085	PENUTUP BOTTLE CAPS ARTEMIA,ARTEMIA BOTTLE CAP...	0.13	PENANG	7	0	HEALTH & WELLNESS
30239	《 MILD EXFOLIATION 》 SOAP FOAMING NET BAG MESH...	0.14	PERAK	922	20	HEALTH & WELLNESS

```
Total rows: 65992
Total columns: 6

📊 Category count:
- STATIONERY: 20460 products
- BEAUTY & SKINCARE: 17957 products
- HEALTH & WELLNESS: 10315 products
- HOME & LIVING: 8293 products
- WOMEN'S FASHION: 4653 products
- HOME APPLIANCES: 3837 products
- MOTHER & BABY: 477 products

Group 2 (Affordable Price):
```

	Product Name	Price	Location	Quantity Sold	Total Reviews	Category
17493	CERAVE ULTRA-LIGHT MOISTURIZING LOTION WITH SU…	43.25	JOHOR	0	0	BEAUTY & SKINCARE
108958	NATURAL HAIRLINE BANGS WIG PIECE WOMEN'S TOP K…	43.25	N/A	0	0	WOMEN'S FASHION
15752	【READY STOCK】FREEPLUS MILD SOAP CLEANSER (100G…	43.25	SELANGOR	0	0	BEAUTY & SKINCARE
18281	CERAVE PM FACIAL LOTION 89ML FOR ALL SKIN TYPE…	43.25	JOHOR	0	0	BEAUTY & SKINCARE
38716	BIOLEAF BOTANICAL BEVERAGE 300ML	43.26	SELANGOR	0	0	HEALTH & WELLNESS
36883	BERRYFULL ORIGINAL INAYAH BEAUTY	43.26	KELANTAN	0	0	HEALTH & WELLNESS
38847	ASH II DIGESTA FOR DIGESTIVE HEALTH (1.4G X 20…	43.26	WP KUALA LUMPUR	42	16	HEALTH & WELLNESS
68112	SMALL DESKTOP AIR PURIFIER NEGATIVE ION STERIL…	43.26	OVERSEAS	524	0	HOME APPLIANCES
36773	HIKARI PREMIUM JAPAN L-GLUTATHIONE & COLLAGEN …	43.26	SELANGOR	0	0	HEALTH & WELLNESS
37961	WHITENING SUPPLEMENT L-GLUTATHIONE PUTIHKAN KU…	43.26	SELANGOR	0	1	HEALTH & WELLNESS

```
Total rows: 22626
Total columns: 6
```

📊 Category count:
- HEALTH & WELLNESS: 6380 products
- BEAUTY & SKINCARE: 5373 products
- WOMEN'S FASHION: 3623 products
- HOME APPLIANCES: 3053 products
- HOME & LIVING: 2309 products
- STATIONERY: 1170 products
- MOTHER & BABY: 718 products

Group 3 (Mid-Range Price):

	Product Name	Price	Location	Quantity Sold	Total Reviews	Category
24233	URIAGE BARIEDERM CLEANSING CICA-GEL (CU-ZN+) 2…	86.44	NEGERI SEMBILAN	0	1	BEAUTY & SKINCARE
25707	JERGENS COCONUT MOISTURIZING BODY LOTION, DRY …	86.45	PENANG	0	1	BEAUTY & SKINCARE
47639	EXP:10/2026 NUTRITION CARE GUT RELIEF POWDER 1…	86.45	SELANGOR	383	109	HEALTH & WELLNESS
61086	[BEST DFFER] ONLINE EXCLUSIVE BATA CAROL WOME…	86.46	WP KUALA LUMPUR	0	0	HOME & LIVING
47791	[EROM] (FREE SHAKER INCLUDED) EATS GREAT MULTI…	86.46	OVERSEAS	51	18	HEALTH & WELLNESS
61393	[BEST DFFER] BATA HALIL WOMEN SANDALS - BLACK…	86.46	WP KUALA LUMPUR	0	1	HOME & LIVING
24064	⭐ ⭐READY STOCK⭐ ⭐ ☞ LUNA JELLY EYE CARE UNDER …	86.48	WP KUALA LUMPUR	0	0	BEAUTY & SKINCARE
44972	SWISSE BEAUTY BRIGHT SKIN BOOSTER (60 CAPSULES…	86.49	SELANGOR	85	30	HEALTH & WELLNESS
22676	SWISSE BEAUTY BRIGHT SKIN BOOSTER (60 CAPSULES…	86.49	SELANGOR	85	30	BEAUTY & SKINCARE
45798	GLUCOMATE COMPLETE BLOOD GLUCOSE TEST STRIP 50PCS	86.50	KELANTAN	38	12	HEALTH & WELLNESS

```
Total rows: 10327
Total columns: 6
```

📊 Category count:
- HEALTH & WELLNESS: 3866 products
- BEAUTY & SKINCARE: 2008 products
- HOME APPLIANCES: 1938 products
- HOME & LIVING: 977 products
- WOMEN'S FASHION: 774 products
- STATIONERY: 451 products
- MOTHER & BABY: 313 products

Group 4 (Premium Price):

	Product Name	Price	Location	Quantity Sold	Total Reviews	Category
63196	JKLIV JAPANESE-STYLE WABI-SABI THICK IMITATION...	129.66	CHINA	0	0	HOME & LIVING
76693	1PCS BLENDER CUP COVER SUITABLE FOR PANASONIC ...	129.66	CHINA	0	0	HOME APPLIANCES
76538	CREATIVE GIFT FOR NURSES DAY PRACTICAL HEALTH ...	129.70	N/A	207	5	HOME APPLIANCES
112774	YOUNG WOMEN'S PURPLE CHEONGSAM DRESS DAILY WEA...	129.70	N/A	600	0	WOMEN'S FASHION
49632	CREATIVE GIFT FOR NURSES DAY PRACTICAL HEALTH ...	129.70	N/A	195	5	HEALTH & WELLNESS
103895	PENTEL VICUNA EX3 - 2 COLOR (BLACK/RED) 0.7 MM...	129.70	SELANGOR	5	2	STATIONERY
61710	THICKENED XPE BABY CRAWLING MAT ENVIRONMENTALL...	129.70	CHINA	0	0	HOME & LIVING
51294	【LOCAL DELIVERY】XIAO.MI 1000ML 除湿机 NEW AIR DEH...	129.71	KEDAH	52	23	HEALTH & WELLNESS
23358	【HOT SALES】[VTCOSMETICS] FOAM CLEANSER 300ML L...	129.71	SELANGOR	0	0	BEAUTY & SKINCARE
23593	【HOT SALES】[VT COSMETICS] SUPER HYALON FOAM CL...	129.71	SELANGOR	0	0	BEAUTY & SKINCARE

```
Total rows: 14651
Total columns: 6

📊 Category count:
- HOME APPLIANCES: 4700 products
- HEALTH & WELLNESS: 3581 products
- BEAUTY & SKINCARE: 2388 products
- HOME & LIVING: 1642 products
- MOTHER & BABY: 923 products
- STATIONERY: 794 products
- WOMEN'S FASHION: 623 products

==================== Performance =====================
Total rows processed: 113596
Code Execution time: 0.6036 seconds
Throughput: 188187.82 rows per second
Current memory usage: 17.8138 MB
Peak memory usage: 19.2549 MB
CPU usage: 88.6%
======================================================

Total time for this cell (Including time to display the performance):
CPU times: user 479 ms, sys: 5.03 ms, total: 484 ms
Wall time: 1.61 s
```

Step 5 : Filtering Based on 'Total Reviews' to Determine Popularity of Products

```python
%%time

tracemalloc.start()
start_time = time.perf_counter()
total_rows = df_cleaned.shape[0]

Q1 = df_cleaned["Total Reviews"].quantile(0.25)
Q3 = df_cleaned["Total Reviews"].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

df_filtered = df_cleaned[(df_cleaned["Total Reviews"] >= lower_bound) &
                         (df_cleaned["Total Reviews"] <= upper_bound)]

min_ratings = df_filtered["Total Reviews"].min()
max_ratings = df_filtered["Total Reviews"].max()
print(f"Minimum Number of Total Reviews: {min_ratings}")
print(f"Maximum Number of Total Reviews: {max_ratings}")

rating_range = round((max_ratings - min_ratings) / 4)
bound1 = min_ratings + rating_range
bound2 = min_ratings + 2 * rating_range
bound3 = min_ratings + 3 * rating_range

group1 = df_filtered[df_filtered["Total Reviews"] <= bound1].sort_values("Total Reviews", ascending=False)
group2 = df_filtered[(df_filtered["Total Reviews"] > bound1) & (df_filtered["Total Reviews"] <= bound2)].sort_values("Total Reviews", ascending=False)
group3 = df_filtered[(df_filtered["Total Reviews"] > bound2) & (df_filtered["Total Reviews"] <= bound3)].sort_values("Total Reviews", ascending=False)
group4 = df_filtered[df_filtered["Total Reviews"] > bound3].sort_values("Total Reviews", ascending=False)

def print_category_counts(df_group):
    print(f"📊 Category count:")
    category_counts = df_group["Category"].value_counts()
    for category, count in category_counts.items():
        print(f"- {category}: {count} products")

for i, (group, name) in enumerate(zip([group1, group2, group3, group4],
                        ["Least popular", "Below Average Popularity", "Above Average Popularity", "Most popular"]), 1):
    print(f"\nGroup {i} ({name}):")
```

```
        display(group.head(10))
        print(f"Total rows: {group.shape[0]}")
        print(f"Total columns: {group.shape[1]}\n")
        print_category_counts(group)

current, peak = tracemalloc.get_traced_memory()
end_time = time.perf_counter()
tracemalloc.stop()

execution_time = end_time - start_time
throughput = total_rows / execution_time

print("\n==================== Performance ====================")
print(f"Total rows processed: {total_rows}")
print(f"Code Execution time: {execution_time:.4f} seconds")
print(f"Throughput: {throughput:.2f} rows per second")
print(f"Current memory usage: {current / 10**6:.4f} MB")
print(f"Peak memory usage: {peak / 10**6:.4f} MB")

cpu_usage = psutil.cpu_percent(interval=1)
print(f"CPU usage: {cpu_usage}%")
print("====================================================")
```

Minimum Number of Total Reviews: 0
Maximum Number of Total Reviews: 27

Group 1 (Least popular):

	Product Name	Price	Location	Quantity Sold	Total Reviews	Category
12471	BENTON DEEP GREEN TEA CLEANSING FOAM 25G / 120...	28.00	WP KUALA LUMPUR	20	7	BEAUTY & SKINCARE
42782	FLAVETTES EFFERVESCENT GLAMZ (30'S)	79.90	SELANGOR	29	7	HEALTH & WELLNESS
85652	CAPABALA ERASABLE PEN GOOD-LOOKING PERIPHERAL ...	3.81	N/A	266	7	STATIONERY
54996	VENICENIGHT WALL STICKER MOISTURE-PROOF DECORA...	10.40	CHINA	27	7	HOME & LIVING
100470	FELTON 10 TIER DRAWER FDD 176	76.79	SELANGOR	39	7	STATIONERY
2585	BODY MILK WHITENING BRIGHTEN SKIN TONE MOISTUR...	11.89	CHINA	14	7	BEAUTY & SKINCARE
92321	(0.7MM / 1.0MM) M&G EXPERT STICK GEL PEN / SIG...	12.50	PENANG	22	7	STATIONERY
100465	BEIGE WATER MILWAUKEE INKZALL BLACK FINE HEAD ...	53.00	NEGERI SEMBILAN	28	7	STATIONERY
12479	VIBRANT GLAMOUR SALICYLIC ACID ACNE TREATMENT ...	29.00	CHINA	27	7	BEAUTY & SKINCARE
33241	UTIX EFFERVESCENT GRANULES SACHET 4G X 28'S	21.90	PAHANG	45	7	HEALTH & WELLNESS

Total rows: 80729
Total columns: 6

📊 Category count:
- BEAUTY & SKINCARE: 18080 products
- HEALTH & WELLNESS: 16519 products
- STATIONERY: 15505 products
- HOME & LIVING: 10196 products
- HOME APPLIANCES: 9757 products
- WOMEN'S FASHION: 8441 products
- MOTHER & BABY: 2231 products

Group 2 (Below Average Popularity):

	Product Name	Price	Location	Quantity Sold	Total Reviews	Category
36	ROREC SADOER NICOTINAMIDE WHITENING FRECKLE MO...	2.90	WP KUALA LUMPUR	52	14	BEAUTY & SKINCARE
113420	EVENING DRESS WOMEN'S NEW BANQUET SEQUIN FISHT...	192.81	N/A	9	14	WOMEN'S FASHION
157	DRKR FACIAL CLEANSER PENCUCI MUKA JERAGAT DEGI...	8.99	SELANGOR	44	14	BEAUTY & SKINCARE
112656	ZY·HT BAJU PEREMPUAN KOREA FASHION SUMMER BLAC...	124.20	CHINA	24	14	WOMEN'S FASHION
112539	VELVET SLIM FIT LONG SLEEVE PARTY DRESS WOMEN'...	90.30	N/A	345	14	WOMEN'S FASHION
112268	NIKE WOMEN'S SPORTSWEAR COLLECTION ESSENTIALS ...	203.15	SELANGOR	67	14	WOMEN'S FASHION
33902	DR.ROS CLASSIC ENZYME 1400ML	25.00	WP KUALA LUMPUR	129	14	HEALTH & WELLNESS
34169	MALTOFER FOL 100MG CHEWABLE TABLET 30'S	29.80	SELANGOR	36	14	HEALTH & WELLNESS
34247	TAIWAN ORIGINAL POINT FEVER GINGER PLASTER OFF...	18.60	N/A	799	14	HEALTH & WELLNESS
33695	佛慈耳鸣左慈丸 FOCI TSO TZU PILLS (200PILLS X 0.2G)	18.18	PENANG	53	14	HEALTH & WELLNESS

```
Total rows: 7845
Total columns: 6
```

📊 Category count:
- BEAUTY & SKINCARE: 2205 products
- HEALTH & WELLNESS: 1784 products
- STATIONERY: 1752 products
- HOME APPLIANCES: 864 products
- HOME & LIVING: 791 products
- WOMEN'S FASHION: 408 products
- MOTHER & BABY: 41 products

Group 3 (Above Average Popularity):

	Product Name	Price	Location	Quantity Sold	Total Reviews	Category
110308	BAMBOO STEM NECKLACE CAT'S EYE STONE WOMEN'S S...	58.68	N/A	79	21	WOMEN'S FASHION
133	EELHOE GINSENG ESSENTIAL OIL SOAP REPAIRS SKIN...	10.82	WP KUALA LUMPUR	43	21	BEAUTY & SKINCARE
112759	CHINOISERIE WEDDING DRESS CHEONGSAM WEDDING GO...	138.01	N/A	1200	21	WOMEN'S FASHION
454	3PCS THAILAND K BROTHERS RICE MILK WHITENING S...	7.90	WP KUALA LUMPUR	151	21	BEAUTY & SKINCARE
210	EMAMI FAIR AND HANDSOME REFRESHING FACE WASH 5...	7.90	JOHOR	140	21	BEAUTY & SKINCARE
110553	2023 SUMMER NEW ARRIVAL K-STYLE WOMEN'S CLOTHI...	49.10	N/A	969	21	WOMEN'S FASHION
65063	OCDEE CONVI SHOEBOX BUNDLE OF 12 SHOE STORAGE ...	369.00	SELANGOR	79	21	HOME & LIVING
64242	HOME PRIME HP054 DESIGNER GRADE SIDE CABINET S...	199.24	PENANG	44	21	HOME & LIVING
64124	NEW CHINESE STYLE METAL WALL DECORATIONS IRON ...	185.30	CHINA	24	21	HOME & LIVING
63730	FULL-BODY MIRROR LED SMART DRESSING MIRROR NOR...	240.00	PENANG	41	21	HOME & LIVING

```
Total rows: 4653
Total columns: 6
```

📊 Category count:
- BEAUTY & SKINCARE: 1279 products
- STATIONERY: 1083 products
- HEALTH & WELLNESS: 1031 products
- HOME APPLIANCES: 537 products
- HOME & LIVING: 469 products
- WOMEN'S FASHION: 231 products
- MOTHER & BABY: 23 products

Group 4 (Most popular):

	Product Name	Price	Location	Quantity Sold	Total Reviews	Category
253	DR RASHEL ALOE VERA FACIAL CLEANSER ...	11.30	WP KUALA LUMPUR	57	27	BEAUTY & SKINCARE
315	POND'S FACIAL FOAM100G	8.50	JOHOR	138	27	BEAUTY & SKINCARE
109679	PREPPY STYLE SUMMER U-NECK SHORT SLEEVE BIG SI...	58.31	N/A	602	27	WOMEN'S FASHION
34	SET JRAGAT 3IN1(SABUN. CREAM MALAM DAN SIANG)	4.99	WP KUALA LUMPUR	264	27	BEAUTY & SKINCARE
493	SABUN YS (DAUN PUDINA) BUATAN TANGAN OLEH YS H...	10.00	SELANGOR	119	27	BEAUTY & SKINCARE
475	MAY GLOWING AND RADIANT LIKE HOTEL BAR SOAP (1...	8.69	SELANGOR	211	27	BEAUTY & SKINCARE
66411	2LITRE CHOPPER MEAT GRINDER BAWANG PUTIH	25.89	JOHOR	112	27	HOME APPLIANCES
66311	(NEW) PENSONIC JUG KETTLE PAB1718 (MINT)	42.90	KEDAH	93	27	HOME APPLIANCES
66210	PREMIER SUPER G KM501 MIXER GRINDER SWITCH KNO...	6.00	PENANG	117	27	HOME APPLIANCES
66013	SURUCHI BLENDER JAR BIG / SURUCHI MIXER GRINDE...	35.00	NEGERI SEMBILAN	87	27	HOME APPLIANCES

```
Total rows: 2783
Total columns: 6

📊 Category count:
- BEAUTY & SKINCARE: 758 products
- HEALTH & WELLNESS: 636 products
- STATIONERY: 609 products
- HOME APPLIANCES: 329 products
- HOME & LIVING: 287 products
- WOMEN'S FASHION: 142 products
- MOTHER & BABY: 22 products

=================== Performance ====================
Total rows processed: 113596
Code Execution time: 0.9005 seconds
Throughput: 126141.24 rows per second
Current memory usage: 10.8595 MB
Peak memory usage: 15.0830 MB
CPU usage: 86.9%
====================================================
CPU times: user 581 ms, sys: 2.85 ms, total: 584 ms
Wall time: 1.9 s
```

Step 6 : Ranking Location Based on Market Performance

```python
%%time
tracemalloc.start()
total_rows = df_cleaned.shape[0]

df_location_sales = df_price.groupby("Location").agg(
    Total_Quantity_Sold=('Quantity Sold', 'sum'),
    Average_Price=('Price', 'mean')
).reset_index()

df_location_sales['Market Performance'] = df_location_sales['Total_Quantity_Sold'] * df_location_sales['Average_Price']

df_location_sales = df_location_sales.sort_values("Market Performance", ascending=False)

display(df_location_sales.head(10))

current, peak = tracemalloc.get_traced_memory()
end_time = time.perf_counter()
tracemalloc.stop()
execution_time = end_time - start_time
throughput = total_rows / execution_time
cpu_usage = psutil.cpu_percent(interval=1)

print("\n=================== Performance ====================")
print(f"Total rows processed: {total_rows}")
print(f"Code Execution time: {execution_time:.4f} seconds")
print(f"Throughput: {throughput:.2f} rows per second")
print(f"Current memory usage: {current / 10**6:.4f} MB")
print(f"Peak memory usage: {peak / 10**6:.4f} MB")

cpu_usage = psutil.cpu_percent(interval=1)
print(f"CPU usage: {cpu_usage}%")
print("====================================================")
```

	Location	Total_Quantity_Sold	Average_Price	Market Performance
23	SELANGOR	10562173	100.849751	1.065193e+09
14	N/A	18524452	43.647313	8.085426e+08
4	CHINA	6341944	97.622898	6.191190e+08
29	WP KUALA LUMPUR	1227922	101.519881	1.246585e+08
8	JOHOR	1499826	78.073532	1.170967e+08
16	OVERSEAS	2083568	48.480572	1.010126e+08
18	PENANG	962481	100.382696	9.661644e+07
19	PERAK	1029478	80.170026	8.253328e+07
6	HONG KONG	376652	138.300619	5.209120e+07
9	KEDAH	595993	80.050847	4.770974e+07

```
==================== Performance ====================
Total rows processed: 113596
Code Execution time: 2.1653 seconds
Throughput: 52461.36 rows per second
Current memory usage: 0.0738 MB
Peak memory usage: 6.2297 MB
CPU usage: 95.5%
=====================================================
CPU times: user 158 ms, sys: 4.42 ms, total: 163 ms
Wall time: 2.25 s
```

End of Part 2 Data Optimization