[Data Cleaning Proccess]

	index	Product Name	Product URL	Brand	Sale Price	Mrp	Discount Percentage	Number Of Ratings	Number Of Reviews	Upc	Star Rating	Ram
0	0	APPLE iPhone 8 Plus (Gold, 64 GB)	https://www.flipkart.com/apple-iphone-8-plus-g...	Apple	49900	49900	0	3431	356	MOBEXRGV7EHHTGUH	4.6	2 GB
1	1	APPLE iPhone 8 Plus (Space Grey, 256 GB)	https://www.flipkart.com/apple-iphone-8-plus-s...	Apple	84900	84900	0	3431	356	MOBEXRGVAC6TJT4F	4.6	2 GB
2	2	APPLE iPhone 8 Plus (Silver, 256 GB)	https://www.flipkart.com/apple-iphone-8-plus-s...	Apple	84900	84900	0	3431	356	MOBEXRGVGETABXWZ	4.6	2 GB
3	3	APPLE iPhone 8 (Silver, 256 GB)	https://www.flipkart.com/apple-iphone-8-silver...	Apple	77000	77000	0	11202	794	MOBEXRGVMZWUHCBA	4.5	2 GB

To analyze datasets effectively, we can explore various relationships and insights that can help understand the patterns and trends related to Datasets
Here's a structured approach to analyze the data:

Data Understanding and Preprocessing

Before diving into analysis, it's essential to understand and preprocess the data to ensure accuracy and reliability.

a. Data Exploration

Overview: Get a general overview of the dataset by checking the number of rows and columns, data types, and sample entries.

Missing Values: Check for any missing or null values in the dataset and decide on methods to handle them (e.g., removal, imputation).

Duplicate Entries: Identify and remove any duplicate rows that may skew analysis.
Outliers Detection: Detect any outliers in numerical columns (e.g., extremely high or low prices) that might affect the analysis.

b. Feature Extraction

Storage Capacity: Extract storage capacity (e.g., 64 GB, 256 GB) from the 'Product Name' column for detailed analysis.

Color Variants: Extract color information from the 'Product Name' column.

Model Identification: Identify different iPhone models (e.g., iPhone 8, iPhone 11, iPhone SE) for comparative analysis.

c. Data Transformation

Price Formatting: Ensure that 'Sale Price' and 'MRP' are in appropriate numerical formats for calculations.

Discount Calculation: Verify the 'Discount Percentage' or calculate it if necessary using 'Sale Price' and 'MRP'.

Date Fields: If available, parse and format any date-related information for time-series analysis.

a. Data Exploration

————————————————————————————————————x————————————————————————————————————

1.General Overview:
This step involves checking the number of rows and columns, data types, and a few sample entries.
2. Missing Values:
This step involves checking for any missing or null values in the dataset.
3. Duplicate Entries:
This step identifies and removes any duplicate rows that may skew the analysis.
4. Outliers Detection:
This step detects outliers in numerical columns (e.g., extremely high or low prices).

b. Feature Extraction :
————————————————————————————————————x————————————————————————————————————

1. Extract Storage Capacity
2. Extract Color Variants

1. Extract Storage Capacity:
You can use regular expressions to extract the storage capacity (e.g., 64 GB, 256 GB) from the 'Product Name' column.
2. Extract Color Variants :
Similarly, extract color information from the 'Product Name' column.

c. Data Transformation

—---x—--

1) Price Formatting:
Ensure that 'Sale Price' and 'MRP' are in appropriate numerical formats for calculations.

2) Discount Calculation:
Verify the 'Discount Percentage' or calculate it if necessary using 'Sale Price' and 'MRP'.

3) Date Fields:
If available, parse and format any date-related information for time-series analysis.