

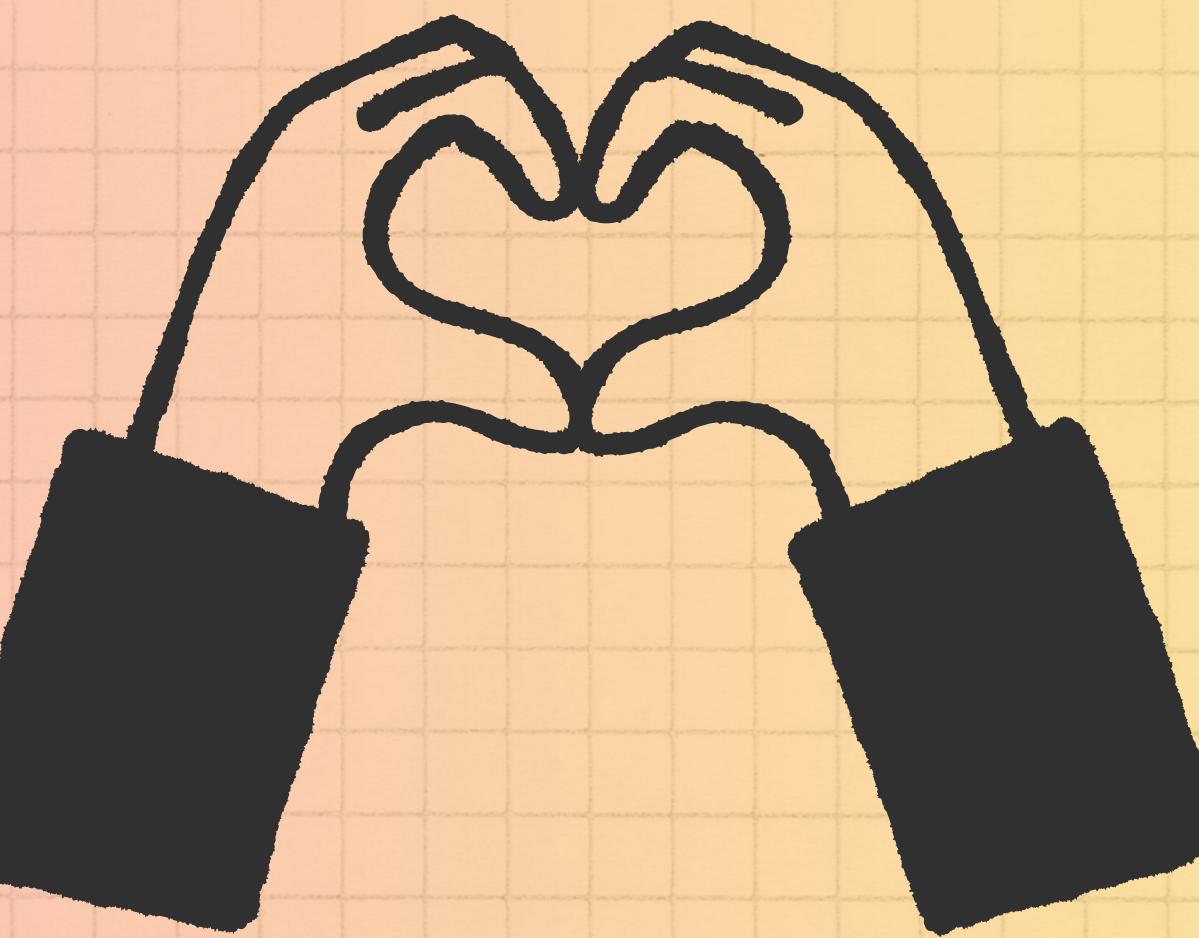
PROJECT PRESENTATION

By Mohit Singh



INTRODUCTION: OVERALL PROJECT OBJECTIVE

This project aims to clean, standardize, and analyze a dataset containing product information, including discount offers and size options. The primary objectives are to ensure data consistency, fill in missing values, and derive meaningful insights that can aid in business decisions.





CONTEXT: BACKGROUND AND JUSTIFICATION

In today's competitive retail market, businesses rely on accurate data to optimize pricing strategies and meet customer preferences. Our dataset includes product discounts and sizes but faces issues such as inconsistent discount formats, missing values, and varied size descriptions, hindering reliable analysis. This project focuses on cleaning, standardizing, and analyzing the dataset to ensure data integrity and usability. By addressing these challenges, we aim to derive actionable insights that support informed business decisions and strategic planning, ultimately enhancing customer satisfaction and driving sales.



OBJECTIVES

MAIN OBJECTIVE

- DATA CLEANING AND PREPARATION:
- IDENTIFY AND REMOVE DUPLICATE ENTRIES.
 - STANDARDIZE DISCOUNT OFFERS TO A UNIFORM PERCENTAGE FORMAT.
 - HANDLE MISSING VALUES IN THE "DISCOUNTPRICE" AND "SIZEOPTION" COLUMNS.

SECONDARY OBJECTIVES

- DATA ANALYSIS:
- CALCULATE THE OVERALL AVERAGE ORIGINAL PRICE FOR PRODUCTS WITH HIGH RATINGS.
 - COUNT PRODUCTS WITH SIGNIFICANT DISCOUNT OFFERS.
 - IDENTIFY THE AVAILABILITY OF PRODUCTS IN SPECIFIC SIZES.

- DATA RETRIEVAL AND LOOKUP:
- USE EXCEL FUNCTIONS TO RETRIEVE SPECIFIC PRODUCT DETAILS BASED ON PRODUCT IDs.



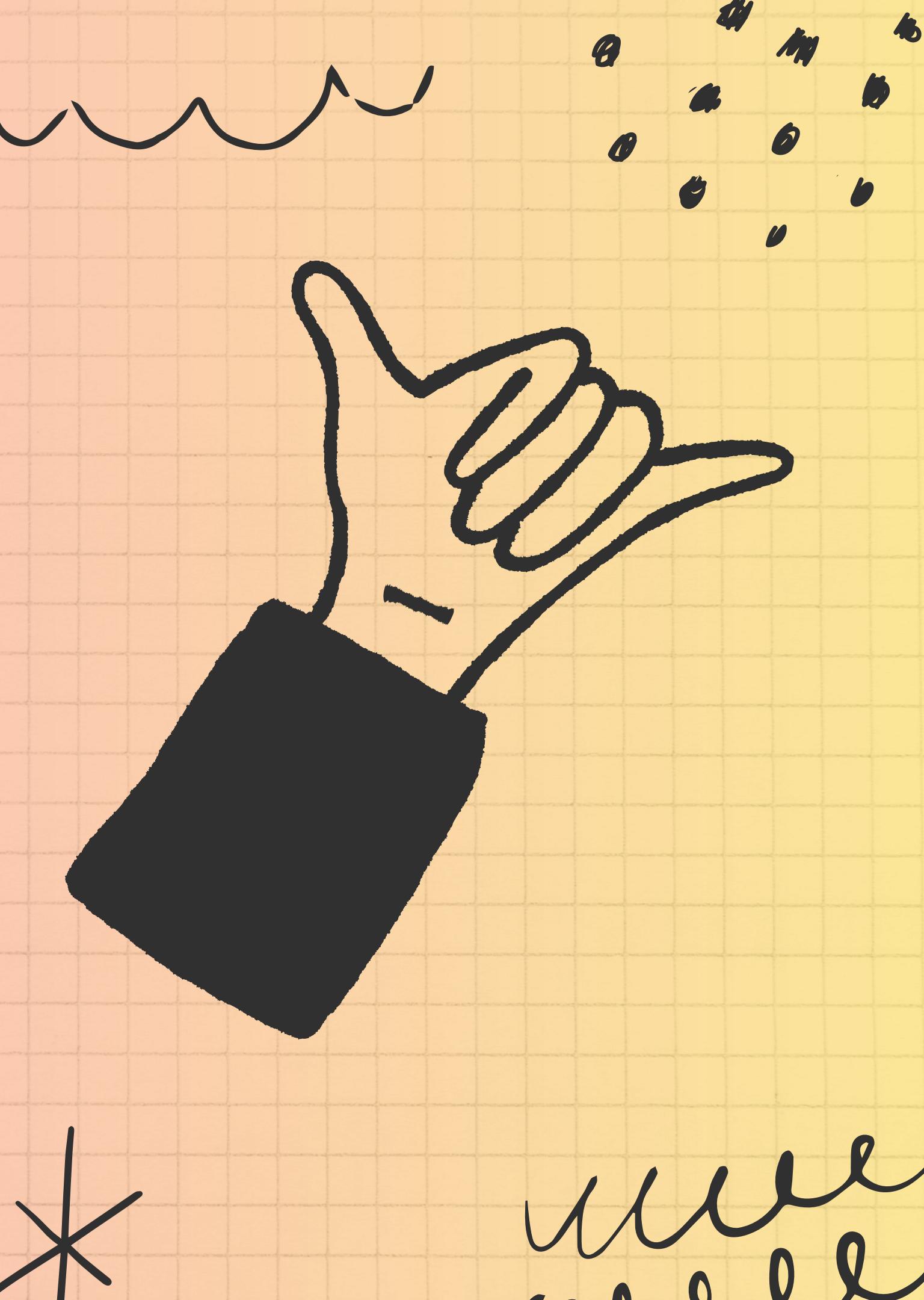
METHODOLOGY: PROCEDURES AND TOOLS

Procedure:

1. Data Cleaning: Remove duplicates and standardize discount formats using Excel formulas.
2. Handle Missing Values: Fill missing "DiscountPrice" based on category averages and replace null "SizeOption" values with "Not Available."
3. Data Analysis: Calculate average original prices for highly-rated products, count significant discount offers, and assess product size availability.
4. Data Retrieval: Use VLOOKUP, XLOOKUP, INDEX, and MATCH functions for specific product details.

Tools:

- Microsoft Excel for data cleaning, preparation, and analysis.
- Excel formulas and functions for efficient data processing and retrieval

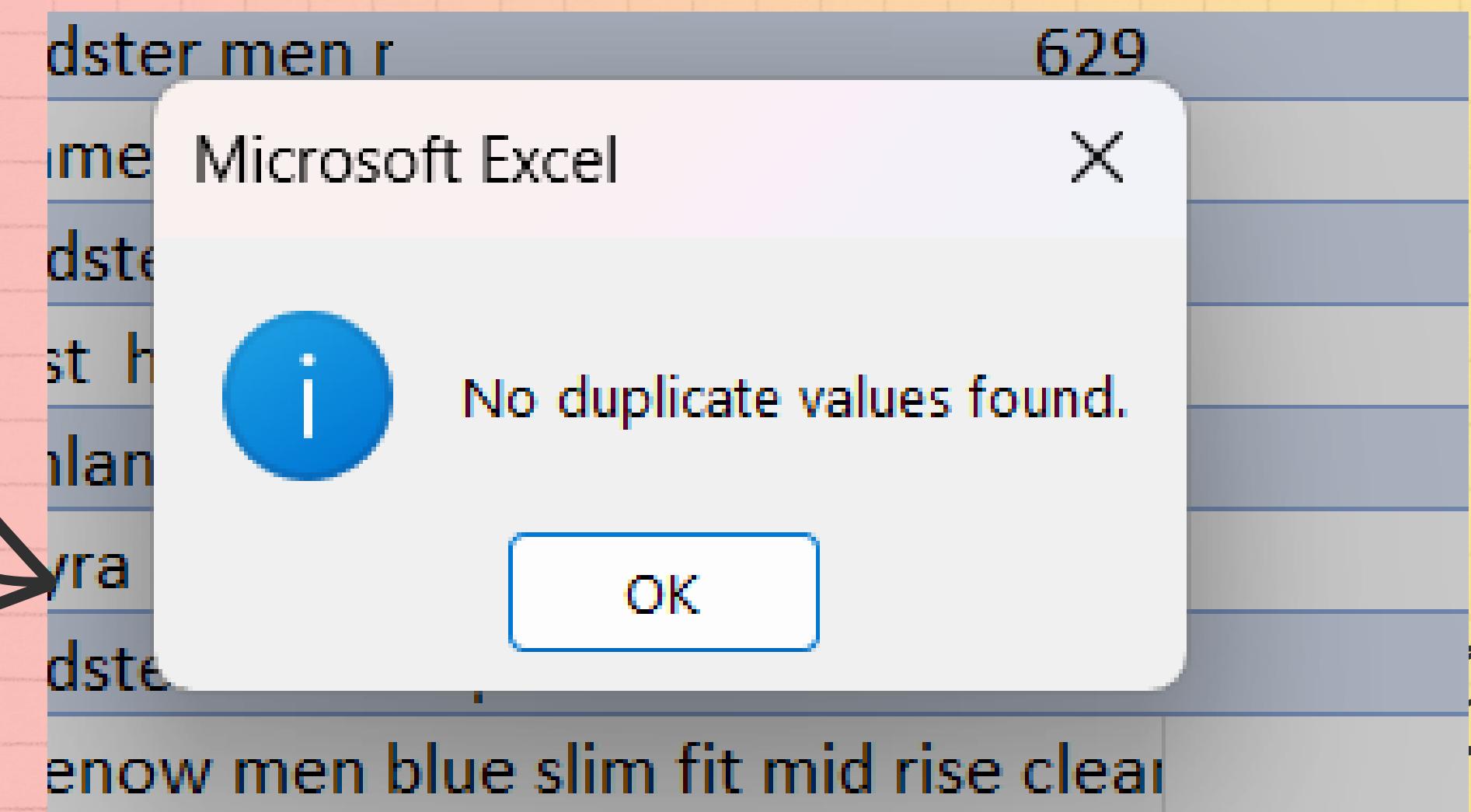
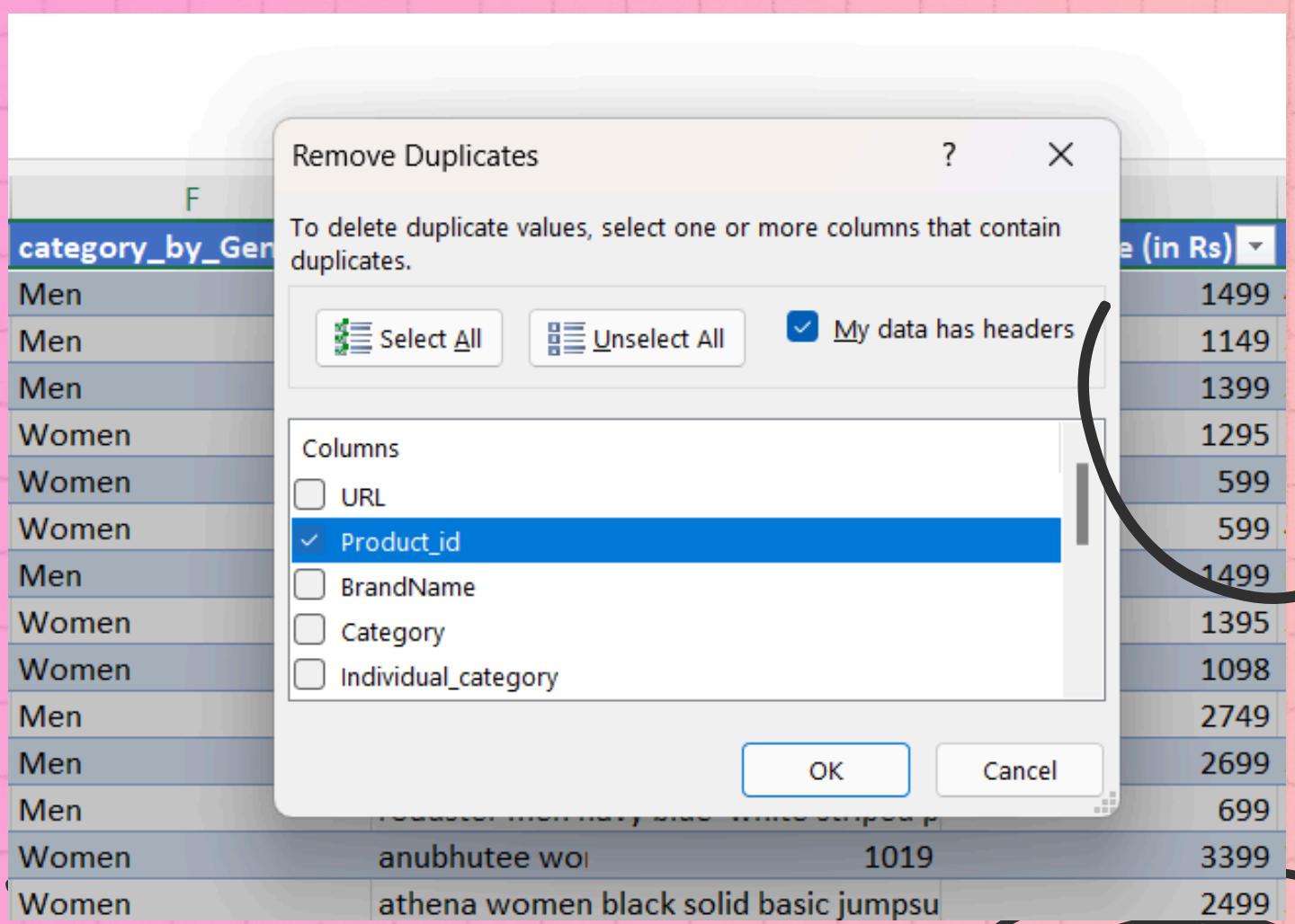




PROJECT QUESTIONS

A. DATA CLEANING AND PREPARATION

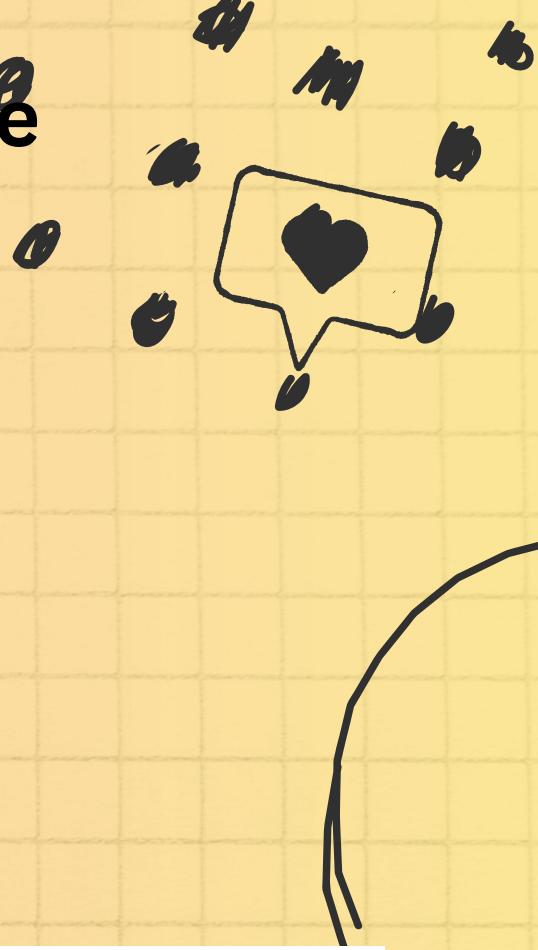
1. Check for duplicate values in your dataset and remove them.



1.2. STANDARDIZE THE "DISCOUNTOFFER" COLUMN TO A SINGLE FORMAT, ENSURING ALL VALUES ARE UNIFORM.

	N	O	P	Q	R	S
1	Column					
999	=TRIM(SUBSTITUTE([DiscountOffer],"Rs.",""))					
999						
999						
999						

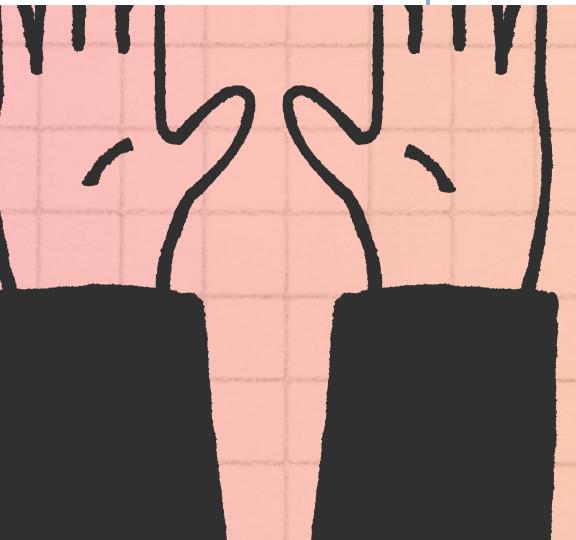
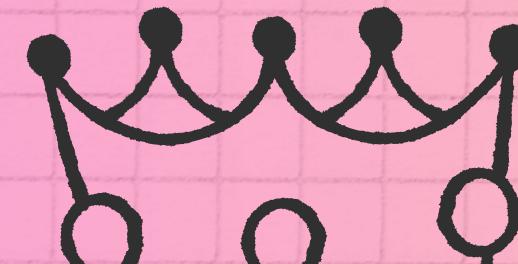
1.3. Identify rows where both "DiscountPrice" and "DiscountOffer" are null and fill the "DiscountPrice" with the average discount price of the respective category.



	Q	R	S	T	U	V	W	X	Y	Z
Column										
45	=IF([@Discount%] = "", AVERAGEIFS([Discount%], [Category], [@Category]), [@Discount%])									
55	IF(logical_test, [value_if_true], [value_if_false])									
55	55									
31	31									



O	P	Q	R	S	T	U
Discounted Price						
=FLOOR.MATH([@OriginalPrice (in Rs)]) -[@Discount amount])						
FLOOR.MATH(number, [significance], [mode])						
029						
893						
389						



B. Data Analysis

1. Calculate the overall average original price for products with ratings greater than 4.

```
=AVERAGEIFS(Table1[OriginalPrice (in Rs)],Table1[Ratings],">4")
```

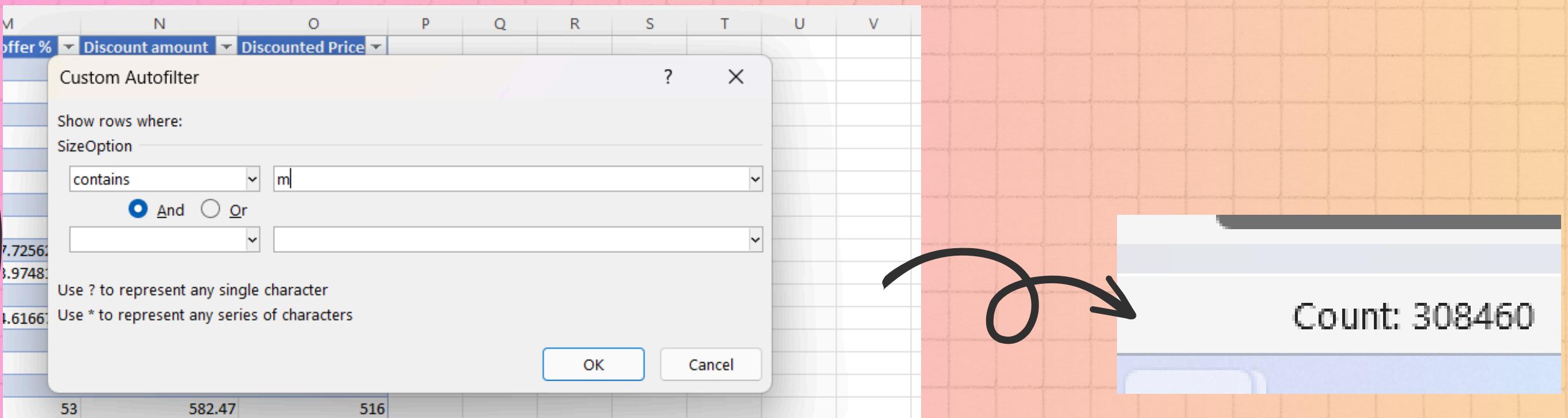
1.2.Count the number of products with a discount offer greater than 50% OFF.

=COUNTIF(Table1[Discountoffer %],">=50")

COUNTIF(range, criteria)

1966.667

1.3.Count the number of products available in size "M."



1.4. Create a new column to label the products as "High Discount" if the discount offer is greater than 50% OFF, otherwise label them as "Low Discount."

Column
=IF([@[Discountoffer %]]>50,"High Discount","Low Discount")
IF(logical_test, [value_if_true], [value_if_false])

P	Q
Column1	
Low Discount	
High Discount	
High Discount	
Low Discount	
Low Discount	
Low Discount	
High Discount	
High Discount	

C. Data Retrieval and Lookup

1. Use VLOOKUP/XLOOKUP to find the product brand, price, and rating of the product with Product_id "11226634".

Product_id	BrandName	Rating
17385142	Roadster	4.5
11226634	Maniac	4.0

1.2.Add Find the "DiscountPrice" for the product with the Product ID "6744434" using the INDEX and MATCH functions.
a little bit of body text

=INDEX(O:O,MATCH(6744434,Table1[Product_id],0))
INDEX(array, row_num, [column_num])
INDEX(reference, row_num, [column_num], [area_num])
359

**THANK YOU
VERY MUCH!**

