Presentation by: Shivani Nagar

Myntra Data Evaluation

Data Cleaning | Data Analysis | 2024

Overview

- Introduction
- Problem Statement
- Approach and solution.
- Screenshots for reference.
- Thank you.

Data Cleaning | Data Analysis | 2024

Introduction

Welcome to my data analysis project, which centers on a comprehensive dataset from Myntra. This dataset encompasses a variety of columns, including the URL of a product, product ID, category, discounts, offers, ratings, and other essential attributes.

Objective : To transform raw data into meaningful insights using the powerful features and formulas available in Microsoft Excel.

STEPS :

- **Data Cleaning :** Identifying and rectifying inconsistencies within the data.
- **Handling Null Values :** Addressing missing data to ensure a complete and accurate dataset.
- **Data Filling :** Imputing missing values to maintain data integrity.

After preparing the data, I applied various **Excel formulas and features to extract valuable insights** and answer pertinent questions related to the dataset.

In the following slides, I will present the problem statements along with screenshots to demonstrate the analysis process and the insights gained.

Part 1: Data Cleaning and Preparation

PROBLEM 1

Check for duplicate values in your dataset and remove them.

PROBLEM 2

Standardize the "DiscountOffer" column to a single format, ensuring all values are uniform.

PROBLEM 3

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

PROBLEM 4

Replace all null values in the "SizeOption" column with the text "Not Available

Approach :

01 Filtered out blank values using the unique product id column.

Convert data into table : Ctrl + T
Then, Using filters remove duplicates, if any.

02 To Standardize the Discount offer column.

- Remove Rs. from discount column.

```
=TRIM(SUBSTITUTE([@DiscountOffer],"Rs.",""))
```

- To Standardize Discount % column.

Convert % into Rupees.

```
=IF(ISNUMBER(SEARCH("%",[@[Discount %]])),LEFT([@[Discount %]],SEARCH("%",[@[Discount %]])-1)/100*[@[OriginalPrice (in Rs)]],[@[Discount %]])
```

Convert Ruppes into %.

```
=IF([@[Discount Rupee]]="","",[@[Discount Rupee]]/[@[OriginalPrice (in Rs)]]*100)
```

To remove OFF, convert formulas to text - > Ctrl +C, Ctrl + V(values).
Then, Text to column -> Using Delimited choose space.

Approach :

03

To fill the DiscountOffer(%) with Average discount price of that category.

```
=IF([@[Discount Offer]]="",AVERAGEIFS(N:N,D:D,[@Category]),[@[Discount Offer]])
```

To fill the Discount Price with the average of respective category (rupees).

```
=IF([@[DiscountPrice (in Rs)]]="",[@[Discount- Category]]*[@[OriginalPrice (in Rs)]]/100,[@[DiscountPrice (in Rs)]])
```

To fix values of DiscountedPrice and Discount% columns - > Ctrl +C, Ctrl + V(values).

04

To replace all null values in Size Option with Not available :

```
=IF(J2="","Not available",J2)
```

Part 2 : Data Analysis

PROBLEM 1

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

PROBLEM 2

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

PROBLEM 3

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

PROBLEM 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”.

Approach :

01 AverageIFS.

Av. original price for products with ratings greater than 4.
1890.640099

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

02 CountIF.

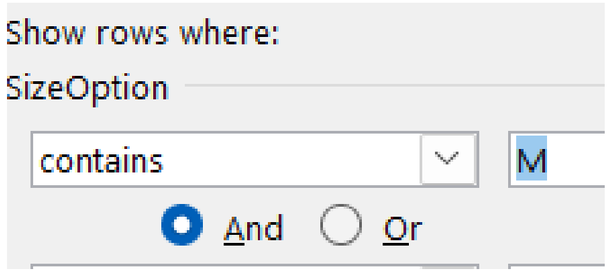
Count the No. of products with a discount greater than 50% OFF.
11301

```
=COUNTIF(Table1[[#All],[DiscountOffer]],"50")
```

Approach :

03 Count the number of products available in size "M"

Approach 1 : Filtering “Using Text Contains “



Count: 64998

Approach 2 : Searching “M”, converting to T/F, --converting to 1/0, adding.

```
=SUMPRODUCT(--(ISNUMBER(SEARCH("M",J:J))))
```

04 A new column : If Discount>50 -> High else Low Discount.

```
=IF([@DiscountOffer]>50,"High Discount","Low Discount")
```

Part 3 : Data Retrieval and Lookup

PROBLEM 1

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

PROBLEM 2

Find the "DiscountPrice" for the product with the Product ID "6744434" using the INDEX and MATCH functions

PROBLEM 3

Utilize nested xlookup to find any column’s detail of a product with it’s product id.

Approach :

01 Using XLOOKUP.

Product ID :		
11226634		
Brand	Price	Rating
Maniac	467	3.9

```
=XLOOKUP(S13,Table1[Product_id],Table1[BrandName],"No Match",0)
=XLOOKUP(S13,Table1[Product_id],Table1[DiscountPrice],"No match",0)
=XLOOKUP(S13,Table1[Product_id],Table1[Ratings],"No Match",0)
```

Using VLOOKUP.

```
=VLOOKUP(S13,B3:P100000,{2,10,14},0)
```

02 Index and Match.

Product id:
6744434
DiscountPrice
599

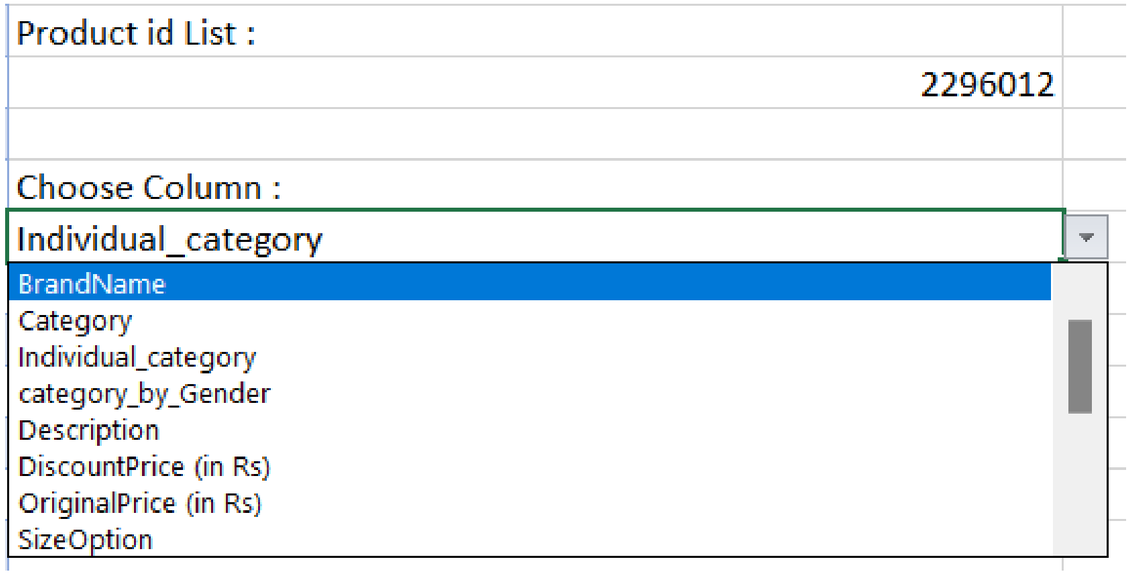
```
=INDEX(Table1[DiscountPrice],MATCH(T25,Table1[Product_id],0))
```

Approach :

01 Nested XLookup.

- Using **Data validation**, first create a list of product IDs.
- Then, Create another list of all other column headers.

Then, use nested Xloopup for extracting values of your choice.




```
=XLOOKUP(R25,Table1[Product_id],XLOOKUP(R28,Table1[[#Headers],[BrandName]:[DiscountType]],Table1[[BrandName]:[DiscountType]],"No Match",0),"No match",0)
```

Thank You

Presentation by: Shivani Nagar