SPREADSHEET AND STATISTICAL PROGRAMMING FOR BUSINESS

Analyzing Data Sets - Identifying Matched, Duplicated, and Unmatched Values

Team 1:

Maneesha Pusuluri

Vikram Srinivas Krishnamoorthy

Abhilash Kolapareddy

Gaurav Suresh Chide

Problem

- Identifying and analyzing unique, duplicated, and non-matching data points is crucial for various tasks in a defined dataset but challenges arise with duplicate values and differentiating them from unique and non-matching data points

Solution

This VBA code implementation for analyzing data sets, involves:

- identifying and highlighting matched values
- Counting occurrences of duplicate values
- Differentiating and separating non-matching values in the defined set

Benefits

- Improved data accuracy and reliability
- Enhanced data analysis and deeper insights
- Reduced processing time due to efficient implementation

This automated approach allows for efficient data cleaning and analysis, leading to more accurate insights and informed decision-making.

Methodology

- This project utilizes four macros:
- Matched Values(): Identifies and highlights unique values and their count in a data set
- Duplicate Values(): Detects and highlights duplicate values and their count excluding the first occurrence
- unmatched Values(): Identifies and highlights unmatched values and their count
- Pie chart(): the chart will be presented to visualize the distribution of data points into categories
- The macros automatically color the background of matched, duplicate, and unmatched values for easy visual identification

RESULTS

- The output of each macro, highlighting:
- List of identified matched values
- List of identified duplicate values with their occurrence counts
- List of identified unmatched values
- Total counts for each category will be displayed for easy comparison

Data Insights

- A pie chart will be presented to visualize the distribution of data points across matched, duplicated, and unmatched categories
- Data labels with percentages will provide detailed information about the proportion of each category
- This visual representation will assist in understanding the overall data structure

DEMONSTRATING RESULTS

Data set	Incoming data set
Alabama	California
Alaska	Texas
Arizona	New York
Arkansas	Florida
California	Ohio
Colorado	Arizona
Connecticut	Georgia
Delaware	Michigan
Florida	Ontario
Georgia	Quebec
Hawaii	British Columbia
Idaho	Alberta
Illinois	Manitoba
Indiana	Nova Scotia
Iowa	California
Kansas	Texas
Kentucky	New York
Louisiana	Florida
Maine	Ohio
Maryland	Colorado
Massachusetts	Ontario
Michigan	Quebec
Minnesota	British Columbia
Mississippi	Alberta
Missouri	Texas
Montana	New York
Nebraska	Florida
Nevada	Ohio
New Hampshire	
New Jersey	
New Mexico	
New York	
North Carolina	
North Dakota	
Ohio	
Oklahoma	
Oregon	
Pennsylvania	
Rhode Island	
South Carolina	
South Dakota	
Tennessee	
Texas	
Utah	
Vermont	
Virginia	
Washington	
West Virginia	
Wisconsin	
Wyoming	

Matched values	Duplicate Values	Unmatched Values
Arizona	California	Ontario
California	Texas	Quebec
Colorado	New York	British Columbia
Florida	Florida	Alberta
Georgia	Ohio	Manitoba
Michigan	Texas	Nova Scotia
New York	New York	Ontario
Ohio	Florida	Quebec
Texas	Ohio	British Columbia
		Alberta

	Total
Matched values	9
Duplicate Values	9
Unmatched Values	10
Input values	28
Output Values	28

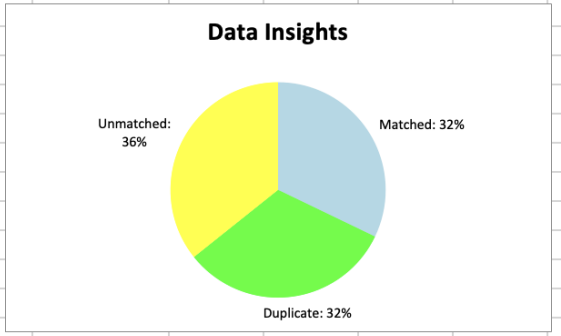
Matched values

Duplicated values

Unmatched values

Pie chart

Clear



WALKTHROUGH TIME

ANY QUESTIONS?

THANK YOU!