Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, No. 4, July : 2023

DATA ANALYSIS AND AUTOMATION IN EXCEL

Keval Maheshbhai Ambasana, Intern, Mechanical Engineering Department, V. V. P. Engineering College, Rajkot, Gujarat, India.
Dr. Nirav P. Maniar, Associate Professor& Head, Mechanical Engineering Department, V. V. P. Engineering College, Rajkot, Gujarat, India.
Dr. Rupesh V. Ramani, Assistant Professor, Mechanical Engineering Department, V. V. P. Engineering College, Rajkot, Gujarat, India.
Prof. Anjana D. Saparia, Assistant Professor, Mechanical Engineering Department, V. V. P. Engineering College, Rajkot, Gujarat, India.

Abstract
The present volume of this research work explores the benefits of automation in MS Excel and provide an overview of the various tools and techniques available for automating tasks in Excel. These include the use of Queries, power pivot, pivot tables if statement, Macros, and add-ins. The paper will also discuss best practices for automation, including proper planning, documentation, testing, and maintenance. Overall, the goal of this paper is to demonstrate how automation in MS Excel can help users become more efficient and effective in their work and provide practical guidance on how to implement automation in Excel-based workflows.

Keywords: Data analysis, automation, Excel.

I. Introduction

Microsoft Excel is a widely used tool for managing and analysing data, and it provides users with a vast array of features that can be utilized to make data processing tasks more efficient. In recent years, the application of automation techniques in MS Excel has gained popularity, as it enables users to streamline repetitive tasks and focus on more value-adding activities. In this research paper, we will explore how MS Excel can be used as a powerful tool for automation, specifically focusing on the role of Power Query, macros, and if statements with conditional keywords in automating tasks. Power Query is a data connection tool that enables users to connect and gather data from various sources. With the use of Power Query, users can extract, transform, and load data into Excel for further analysis. By utilizing this tool, users can automate the process of gathering and analysing data, thereby saving time and improving the accuracy of data analysis. Macros are a powerful tool in Excel that enables users to automate repetitive tasks. Macros are recorded commands that can be replayed automatically, enabling users to save time and reduce errors associated with repetitive manual tasks. With the use of macros, users can automate tasks such as data entry, formatting, and report generation. Conditional statements in MS Excel are used to test a condition and execute an action based on the result of the test. If statements enable users to automate decision-making processes by testing a condition and executing an action based on the result of the test. By utilizing if statements, users can automate tasks such as data analysis, and create machine learning models for temporary work levels. In this research paper, we will explore the benefits of using Power Query, macros, and if statements with conditional keywords in automating tasks in MS Excel. We will provide practical guidance on how to implement these automation techniques, and outline best practices for planning, documenting, testing, and maintaining automated Excel-based workflows. Furthermore, we will give an overview of data analysis in MS Excel and demonstrate how automation techniques can be used to streamline the data analysis process.
Overall, the goal of this research paper is to demonstrate how MS Excel can be utilized as a powerful tool for automation and provide practical guidance on how to implement these techniques to make work easier, more efficient, and less error prone.

II. MS Excel – Powerful tool for Data Analysis and Automation

2.1 Problems working without automation in MS Excel

Some examples of the types of problems people faces when working without automation in MS Excel:

Repetitive manual tasks: Working without automation can mean that users need to perform repetitive manual tasks such as data entry, formatting, and report generation. These tasks can be time-consuming, error-prone, and tedious, leading to decreased productivity and job satisfaction.

Lack of efficiency: Without automation, users may need to spend significant amounts of time searching for and manipulating data to get it into a usable format. This can result in a lack of efficiency and lost productivity, especially when working with large and complex data sets.

Human error: Manual data entry and manipulation can result in errors, which can lead to incorrect analyses and decision-making. Human errors can also be time-consuming to correct, especially if they are not detected until later in the process.

Inconsistencies: When working without automation, users may struggle to maintain consistency in their data analysis and reporting. For example, different team members may use different methods to clean and analyze data, leading to inconsistencies in the results.

Limited scalability: Working without automation can limit the scalability of data analysis and reporting. For example, if a team is manually entering data into a spreadsheet, it can be difficult to scale the process to handle larger data sets or more complex analyses.

2.2 Power Query

Power Query is a data transformation and cleansing tool that is part of Microsoft Excel's suite of business intelligence tools. Power Query enables users to easily import and transform data from various sources such as databases, files, and online services, and then load the data into Excel for analysis and reporting. It provides a user-friendly interface for working with data and enables users to perform complex data transformations with just a few clicks. With Power Query, users can perform a range of transformations on their data, such as filtering, sorting, merging, pivoting, and grouping. It also includes a range of data cleaning and formatting features, such as removing duplicates, splitting columns, and changing data types. Power Query provides a graphical user interface that guides users through the process of importing and transforming data, making it easy for even non-technical users to work with complex data sets. It also includes a powerful query editor that allows users to create custom queries and formulas to perform more advanced transformations. Power Query can be used with a wide range of data sources, including SQL Server, Oracle, Access, SharePoint, and other data sources. Additionally, it can also connect to online services like Facebook, Google Analytics, and Salesforce. Overall, Power Query is a powerful tool that enables users to easily transform and cleanse data in Microsoft Excel, making it easier to work with large and complex data sets and improving the accuracy and efficiency of data analysis. Figure 1 shows display of power query.

2.3 Web Query and Folder Query

Web Query and Query using specific folder are two different functionalities in Microsoft Excel, so I will explain each of them separately:

1. Web Query: A web query is a feature in Excel that allows you to extract data from a website into an Excel worksheet. With a web query, you can easily pull data from websites that provide regularly updated information, such as stock quotes, weather forecasts, and sports scores.

The working of a web query involves the following steps:

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, No. 4, July : 2023

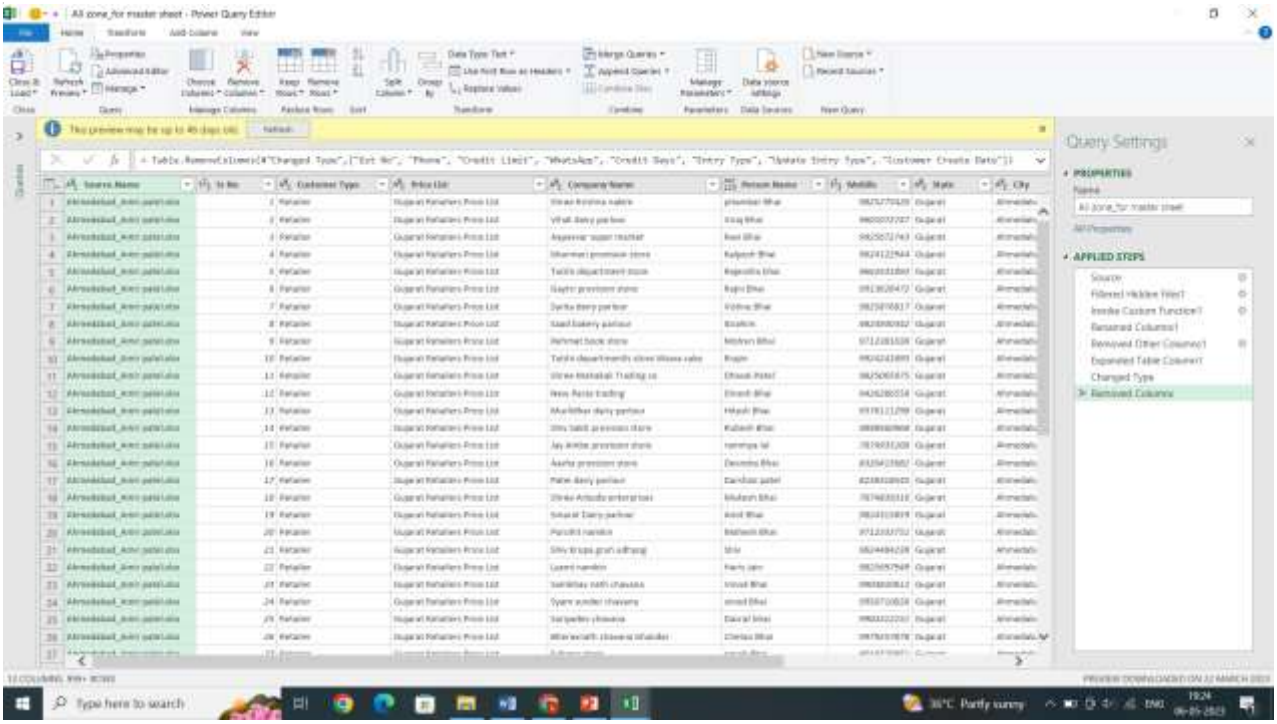


Figure 1: Display of Power Query

- Open a new or existing Excel worksheet.
- Select the cell where you want to insert the data.
- From the "Data" tab on the ribbon, click on "From Web".
- In the "New Web Query" dialog box, enter the URL of the website you want to extract data from and click "Go".
- Excel will display the website in the "Web Query" dialog box, where you can select the specific data you want to extract by clicking on the yellow arrow icons.
- Once you have selected the data you want to extract, click "Import" to insert the data into the worksheet.

2. Query using specific folder: A query using a specific folder is a feature in Excel that allows you to extract data from files located in a specific folder on your computer. This feature is useful when you have many files that contain data that you want to combine and analyze in Excel.

The working of a query using a specific folder involves the following steps:

- Open a new or existing Excel worksheet.
- From the "Data" tab on the ribbon, click on "Get Data" and then select "From Folder".
- In the "From Folder" dialog box, select the folder that contains the files you want to extract data from.
- Excel will display a list of all the files in the selected folder, where you can select the specific files you want to extract data from by clicking on the checkboxes.
- Once you have selected the files you want to extract data from, click "Edit" to open the Power Query Editor.
- In the Power Query Editor, you can perform various transformations on the data, such as filtering, sorting, and merging, and then load the data into the worksheet by clicking "Close & Load" from the "Home" tab on the ribbon.

Overall, both Web Query and Query using specific folder are useful features in Excel that enable users to extract data from various sources and perform data analysis and reporting. You can use Power Query in Excel to gather data from one folder. Power Query is a powerful data transformation and analysis tool that enables users to import, transform, and combine data from various sources.

Here are the steps to gather data from one folder using Power Query in Excel:

1. Open a new or existing Excel worksheet.
2. From the "Data" tab on the ribbon, click on "Get Data" and then select "From File" and then "From Folder".
3. In the "From Folder" dialog box, select the folder that contains the files you want to import.
4. Click "OK" to import all the files from the selected folder.
5. Excel will display a preview of the data in the "Power Query Editor" window.
6. Use the Power Query Editor to transform and clean the data as needed, using functions such as filtering, sorting, and merging.
7. Once you have completed the transformations, click "Close & Load" from the "Home" tab on the ribbon to import the data into the worksheet.

By using Power Query to gather data from one folder in Excel, you can easily combine data from multiple files into one table, perform data cleansing and transformation, and save time and effort in data processing.

2.4 Power pivot

Power Pivot is a powerful data analysis and visualization tool that is available as an add-in for Microsoft Excel. It enables users to create sophisticated data models, analyse large volumes of data, and generate dynamic reports and visualizations.

Here are some key features and benefits of Power Pivot:

1. Data modelling: Power Pivot allows users to create relationships between data tables, define calculated fields and measures, and create hierarchies to organize and summarize data.
2. Large data volumes: Power Pivot can handle large volumes of data, up to millions of rows, by compressing and aggregating data to optimize performance.
3. Advanced calculations: Power Pivot supports advanced calculations and functions such as DAX (Data Analysis Expressions) to create complex calculations and formulas.
4. Interactive reports and visualizations: Power Pivot allows users to create interactive reports and visualizations with features such as slicers, timelines, and charts.
5. Integration with other Microsoft tools: Power Pivot is integrated with other Microsoft tools such as Power BI and SharePoint, allowing users to share and collaborate on data models and reports.

To use Power Pivot in Excel, you need to enable the Power Pivot add-in and import data into the data model. You can then create pivot tables and charts based on the data model and use DAX formulas to perform calculations and analysis. Power Pivot offers a powerful and flexible way to analyse and visualize data and is particularly useful for large and complex datasets. Figure 2 shows data relationship in Power Pivot.

2.5 IF statement and conditional functions

In Microsoft Excel, the IF statement is a powerful tool for performing conditional logic and making decisions based on the value of a particular cell or range of cells. The IF statement uses conditional keywords such as "IF", "THEN", and "ELSE" to evaluate a logical test and return a value based on the outcome of that test.

The basic syntax of an IF statement in Excel is as follows:

=IF(logical test, [value_if_true], [value_if_false])

Where "logical test" is the condition or test to be evaluated, and "value_if_true" and "value_if_false" are the values to be returned if the test is true or false, respectively.

For example, suppose you have a cell containing a numeric value, and you want to display a message if the value is greater than 10. You can use the following IF statement:

=IF(A1>10, "Value is greater than 10", "")

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, No. 4, July : 2023

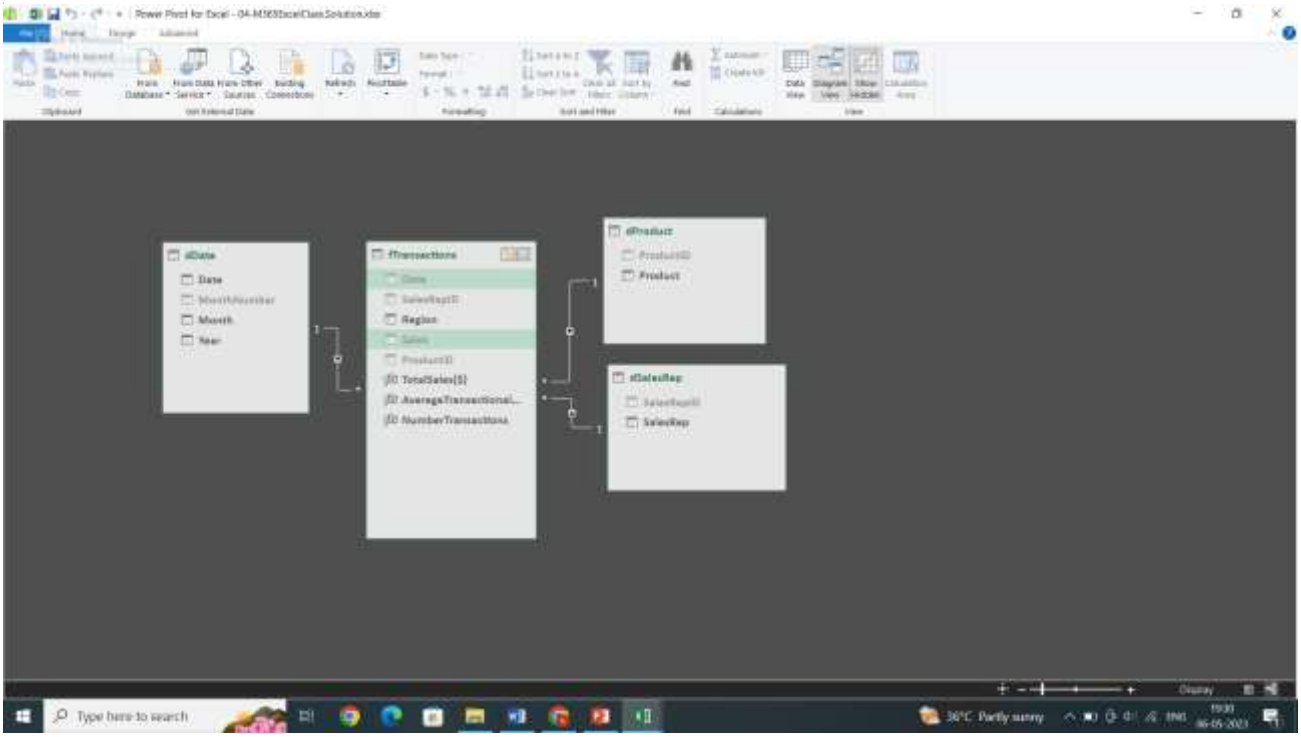


Figure 2: Data Relationship in Power Pivot

This formula will return the message "Value is greater than 10" if the value in cell A1 is greater than 10, and an empty cell otherwise.

In addition to the basic IF statement, Excel also supports several conditional keywords and functions that can be used in conjunction with the IF statement to perform more complex calculations and logic.

These include:

- AND: Returns TRUE if all of the arguments are true.
- OR: Returns TRUE if any of the arguments are true.
- NOT: Reverses the logic of a test.
- IFERROR: Returns a value if a formula results in an error.
- CHOOSE: Returns a value from a list of values based on an index number.

By using IF statements and conditional keywords in Excel, you can perform complex calculations and logic based on the values of your data, and automate repetitive tasks and decision-making processes.

2.6 Example of IF statement

This formula is an example of an IF statement in Excel that uses the OR function to test multiple conditions. The basic syntax of the formula is:

=IF(OR(condition1, condition2, ...), [value_if_true], [value_if_false])

In this formula, the logical test uses the OR function to test if the value in cell D4 is equal to any of the values "MUMBAI" or "PUNE", "RAJKOT", "BHUJ", "KUTCH", "BHAVNAGAR", "AHMEDABAD", "SURAT", "BARODA", or "JAIPUR", "BETUL", "PANIPAT", "JABALPUR", "NAGPUR", or "Orrisa".

If the test is true (i.e., if the value in cell D4 is equal to any of the specified values), the formula returns the value from the "Price" sheet in cell B4, B3, or B5, depending on which group the value in cell D4 belongs to.

If the test is false (i.e., if the value in cell D4 is not equal to any of the specified values), the formula returns an empty cell.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, No. 4, July : 2023

This type of formula can be useful for categorizing data based on specific criteria, and automating tasks that would otherwise require manual sorting or classification.

=IF(OR($D4={"MUMBAI","PUNE"}),Price!$B$4,

This is the first part of the formula. It begins with an IF statement, which allows you to perform a logical test and return one value if the test is true, and another value if the test is false. In this case, the logical test is an OR statement that checks if the value in cell D4 is equal to either "MUMBAI" or "PUNE".

If the test is true, the formula returns the value in cell B4 from a sheet named "Price".

IF(OR($D4={"RAJKOT","BHUJ","KUTCH","BHAVNAGAR","AHMEDABAD","SURAT","BARODA"}),Price!$B$3,

This is the second part of the formula. If the first logical test (i.e., checking for "MUMBAI" or "PUNE") is false, the formula moves on to the second logical test. This test checks if the value in cell D4 is equal to any of the values listed after the OR function (i.e., "RAJKOT", "BHUJ", etc.).

If the test is true, the formula returns the value in cell B3 from the "Price" sheet.

IF(OR($D4={"JAIPUR","BETUL","PANIPAT","JABALPUR","NAGPUR","Orrisa"}),Price!$B$5,))

This is the third and final part of the formula. If the first two logical tests are false, the formula moves on to this third and final test. This test checks if the value in cell D4 is equal to any of the values listed after the OR function (i.e., "JAIPUR", "BETUL", etc.).

If the test is true, the formula returns the value in cell B5 from the "Price" sheet. If the test is false (i.e., if the value in cell D4 does not match any of the values in the OR statements), the formula returns an empty cell.

So, in summary, this formula checks the value in cell D4 to see if it matches any of several different criteria. Depending on which criteria is met, the formula returns a different value from a specific cell in a sheet named "Price". This can be useful for categorizing or sorting data automatically based on specific criteria. Figure 3 shows use of nested IF - write name of city and it will give pricelist of it.

NO	REGION	ORDER												
		30g	RATE	VALUE	60g	RATE	VALUE	ROUND	RATE	VALUE	PAPAD	RATE	VALUE	TOTA
3	AHMEDABAD	0	1927.68				0.00	150	908.10	136215.00	0	1452.00	0.00	136215
4	RAJKOT	0	1927.68				0.00	90	908.10	81729.00	0	1452.00	0.00	81729
5	BARODA	75	1927.68	144576.00	0	0.00	0.00	114	908.10	103523.40	5	1452.00	7260.00	255359
6	PANIPAT	18	1745.92	31426.56	10	2184.96	21849.60	10	870.60	8706.00	0	1452.00	0.00	61982
7	PUNE		2403.84	0.00		2183.68	0.00	42	870.60	36565.20		1452.00	0.00	36565
8	JAIPUR	111	1745.92	193797.12	0	2184.96	0.00	0	870.60	0.00	0	1452.00	0.00	193797
9	SURAT	85	1927.68	163852.80	0	0.00	0.00	60	908.10	54486.00	40	1452.00	58080.00	276418
10	MUMBAI	0	2403.84	0.00	30	2183.68	65510.40	100	870.60	87060.00	3	1452.00	4356.00	156926
11	BHUJ	70	1927.68	134937.60	0	0.00	0.00	95	908.10	86269.50	10	1452.00	14520.00	235727
12	RAJKOT	97	1927.68	186984.96	0	0.00	0.00	25	908.10	22702.50	3	1452.00	4356.00	214043
13	JABALPUR	7	1745.92	12221.44	0	2184.96	0.00	7	870.60	6094.20	0	1452.00	0.00	18315
14	Orrisa		1745.92	0.00		2184.96	0.00	6	870.60	5223.60		1452.00	0.00	5223
15	AHMEDABAD	110	1927.68	212044.80	0	0.00	0.00	90	908.10	81729.00	0	1452.00	0.00	293773
16	PUNE		2403.84	0.00		2183.68	0.00	65	870.60	56589.00		1452.00	0.00	56589
17	Bhavnagar		1927.68	0.00		0.00	0.00	83	908.10	75372.30		1452.00	0.00	75372
18			0.00	0.00		0.00	0.00		0.00	0.00		0.00	0.00	0.0
19	RAJKOT	85	1927.68	163852.80	0	0.00	0.00	97	908.10	88085.70	3	1452.00	4356.00	256294
20	JAIPUR	35	1745.92	61107.20	0	2184.96	0.00	80	870.60	69648.00	0	1452.00	0.00	130755
21	PUNE		2403.84	0.00		2183.68	0.00	90	870.60	78354.00		1452.00	0.00	78354
22	BARODA	110	1927.68	212044.80	0	0.00	0.00	86	908.10	78096.60	0	1452.00	0.00	290141

Figure 3: Use of nested IF

2.6 Example of Macros

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, No. 4, July : 2023

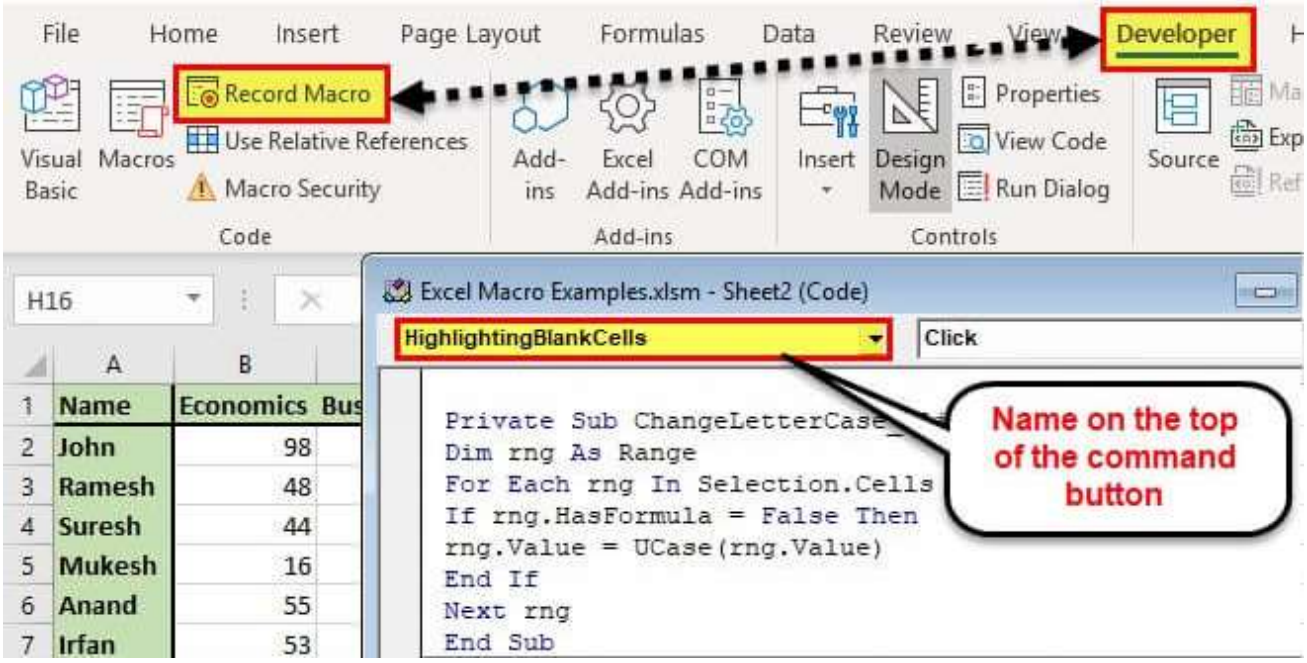
Let's say you have a large dataset containing employee information, including their names, addresses, phone numbers, and job titles. You need to create a report that lists all employees in a particular department, along with their phone numbers and job titles.

Normally, you would need to manually filter the data to display only the employees in the relevant department, and then copy and paste their names, phone numbers, and job titles into a new worksheet. This could take a lot of time and effort if you have a large dataset with many employees.

Instead, you could create a macro that automates this process. The macro could perform the following steps:

1. Open the original dataset and select all the data
2. Filter the data to display only the employees in the relevant department
3. Copy the employee names, phone numbers, and job titles
4. Open a new worksheet and paste the data into the appropriate cells
5. Format the report as desired (e.g., add headings, adjust column widths, etc.)

Once you've recorded the macro, you can run it whenever you need to generate the report. This can save you a significant amount of time and effort, especially if you need to create the report on a regular basis.



	A	B
1	Name	Economics
2	John	98
3	Ramesh	48
4	Suresh	44
5	Mukesh	16
6	Anand	55
7	Irfan	53

Figure 4: Macro in Excel

III. Conclusion

Automation in MS Excel refers to the process of using advanced features and programming techniques to automate repetitive tasks and streamline data processing workflows in Microsoft Excel. By automating tasks such as data entry, calculations, formatting, and report generation, Excel users can significantly increase their productivity, reduce errors, and save time. Excel is simply excellent.

References

[1] Campus, E. (n.d.). Excel IFERROR Function. Retrieved from Excel campus: https://www.excelcampus.com/functions/excel-iferror-function/.

[2] DataCamp. (n.d.). Introduction to power pivot in excel. Retrieved from datacamp: https://www.datacamp.com/community/tutorials/introduction-to-power-pivot-in-excel

[3] easy, E. (n.d.). web query in excel. Retrieved from Excel easy: https://www.excel-easy.com/examples/web-query.html

[4] Exceljet. (n.d.). Import multiple files into power query. Retrieved from Exceljet: https://exceljet.net/tips/how-to-import-multiple-files-into-power-query.
[5] Spreadsheeto. (n.d.). Excel AND Function -how to use. Retrieved from Spreadsheeto: //spreadsheeto.com/and-function/.
[6] support, M. (n.d.). what is power query. Retrieved from microsoft support: https://support.microsoft.com/en-us/office/what-is-power-query-e8a76cff-4ff5-432e-afe4-fde48b398c5d.