

WORLD MOVIES DATASET

PROJECT REPORT

BUSINESS INTELLIGENCE AND ANALYTICS

ECS51517

Submitted by

MYTHRESH VARMA R (22112064)

SAI CHARAN NAIDU P (22112381)

NAVEEN KRISHNA J (22112388)

Under the guidance of

Dr. P. RANJANA

Professor,CSE

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



**HINDUSTAN INSTITUTE OF TECHNOLOGY AND SCIENCE
CHENNAI – 603103**

DECEMBER 2024



HINDUSTAN

INSTITUTE OF TECHNOLOGY & SCIENCE
(DEEMED TO BE UNIVERSITY)

BONAFIDE CERTIFICATE

Certified that this project report "**WORLD MOVIES DATASET**" is the bonafide work of "**MYTHRESH VARMA R, SAI CHARAN NAIDU P, NAVEEN KRISHNA J**" who carried out the project work under my supervision during the academic year **2024-2025**.

DR.P.RANJANA , M.E. Ph.D
PROFESSOR,CSE
SUPERVISOR

**Hindustan Institute of Technology and Science,
Department of Computer Science and Engineering,
Chennai- 603103**

TABLE OF CONTENTS

CHAPTER NO		CHAPTER NAME	PAGE NUMBER
1	1.1	INTRODUCTION TO WEKA	5
2	2.1 2.2	Create a Marvel Movies Table with the help of Data Mining Tool Weka. Apply Pre-Processing techniques to the training data set of Marvel Movies Table.	11
3	3.1 3.2 3.3	Normalize the Marvel Movies Table data using Knowledge Flow. To construct a Decision Tree for top Marvel Movies data and classify it. Conclusion	17
4	4.1 4.2 4.3	INTRODUCTION TO MICROSOFT EXCEL THE SUM, AVERAGE, BOOLEAN OF THE DATASET AND THE PIVOT CHART Conclusion	24
5	5.1 5.2 5.3 5.4	INTRODUCTION TO POWER BI THE PARTS OF POWER BI THE FLOW OF WORK IN POWER BI INSTALL AND RUN POWER BI	27
6	6.1 6.2	INTRODUCTION TO POWER BI APPLY TRANSFORMATION TECHNIQUES	31
7	7.1	LOADING THE DATA AFTER TRANSFORMATION FOR VISUALIZATION	36
8	8.1	VISUALIZING USING VARIOUS CHARTS	37
9	9.1 9.2	ADDING TITLES TO THE GRAPHS AND ORGANIZING CONCLUSION	43

ABSTRACT

The "World Movies Dataset" project aims to analyze and extract insights from a comprehensive dataset encompassing the vast array of films within the Marvel Cinematic Universe (MCU). This dataset includes various attributes such as movie titles, release dates, box office earnings, ratings, directors, and cast members. The primary objective is to explore trends in movie performance, audience reception, and the influence of key factors such as cast and crew on the success of the films. Through data visualization, statistical analysis, and machine learning techniques, the project seeks to uncover patterns that contribute to the MCU's unprecedented success in the film industry. The findings from this analysis could provide valuable insights for future film production strategies and contribute to the broader field of entertainment analytics.

Attributes:

1. Film
2. Category
3. Worldwide gross \$m
4. % Budget recovered
5. Critics % score
6. Audience % score
7. Audience vs critics % deviance
8. Budget
9. Domestic gross \$m
10. International gross \$m
11. Opening weekend \$m
12. Second weekend \$m
13. 1st vs 2nd week dropoff
14. % gross from opening weekend
15. % gross from domestic

16. % gross from international

17. % budget opening weekend

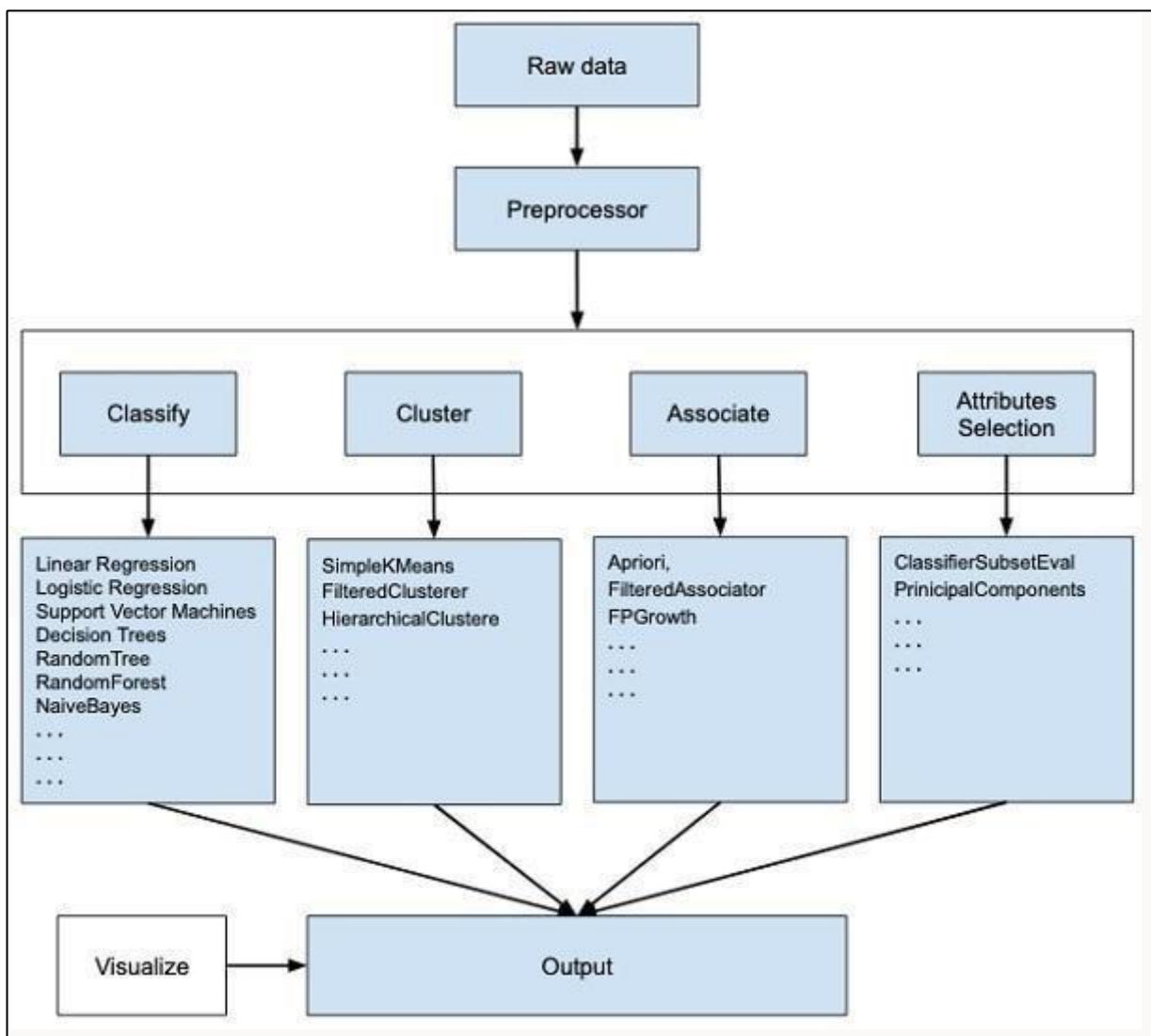
18. Year

19. Source

CHAPTER 1

1.1 INTRODUCTION TO WEKA

WEKA - an open source software provides tools for data preprocessing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problems. What WEKA offers is summarized in the following diagram –



If you observe the beginning of the flow of the image, you will understand that there are many stages in dealing with Big Data to make it suitable for machine learning –

First, you will start with the raw data collected from the field. This data may contain several null values and irrelevant fields. You use the data preprocessing tools provided in WEKA to cleanse the data.

Then, you would save the preprocessed data in your local storage for applying ML algorithms.

Next, depending on the kind of ML model that you are trying to develop you would select one of the options such as Classify, Cluster, or Associate. The Attributes Selection allows the automatic selection of features to create a reduced dataset.

Note that under each category, WEKA provides the implementation of several algorithms. You would select an algorithm of your choice, set the desired parameters and run it on the dataset.

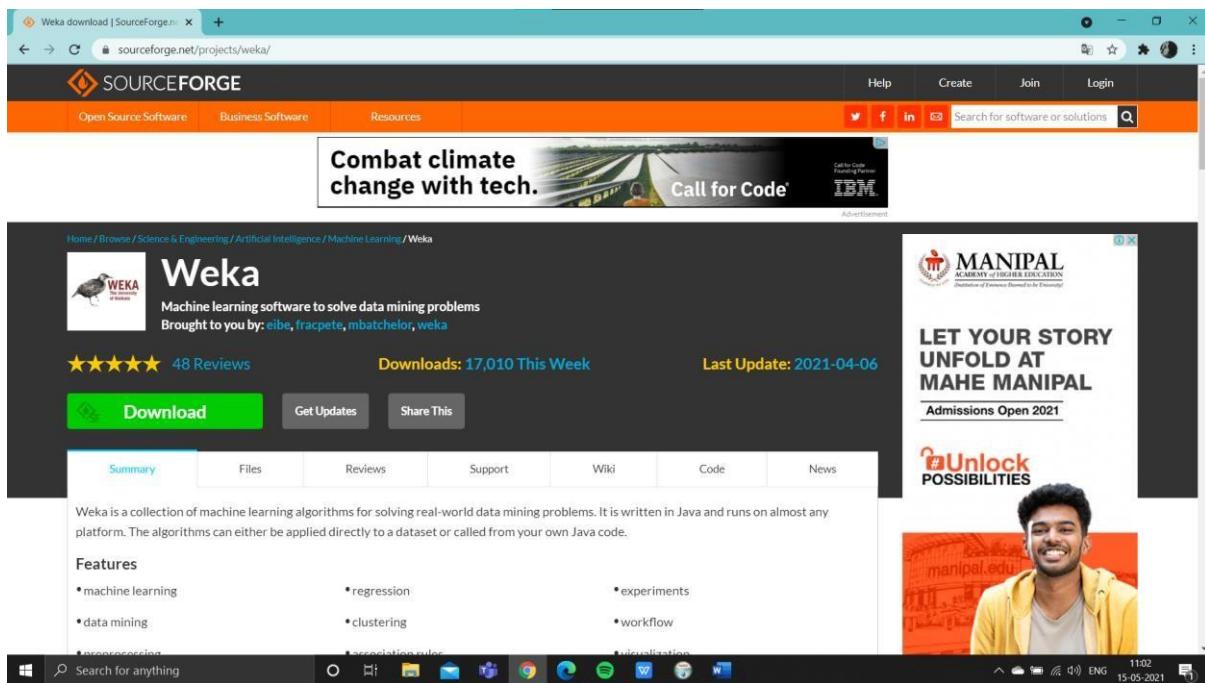
Then, WEKA would give you the statistical output of the model processing. It provides you a visualization tool to inspect the data.

The various models can be applied on the same dataset. You can then compare the outputs of different models and select the best that meets your purpose.

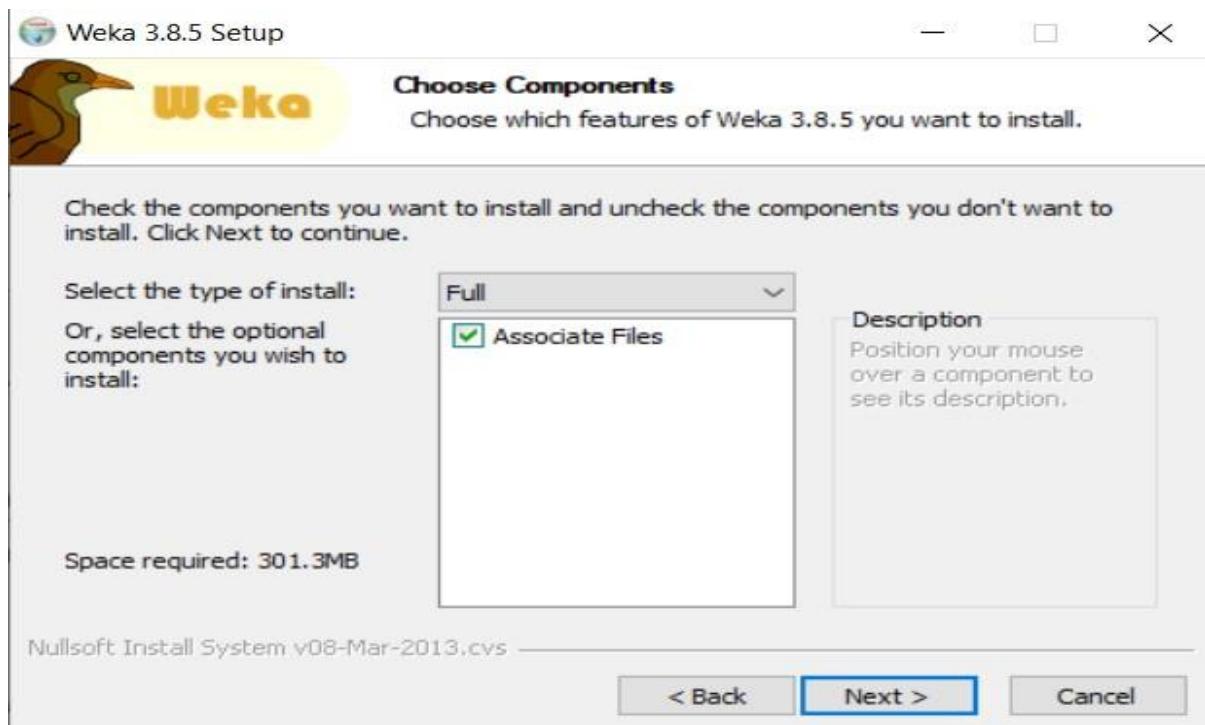
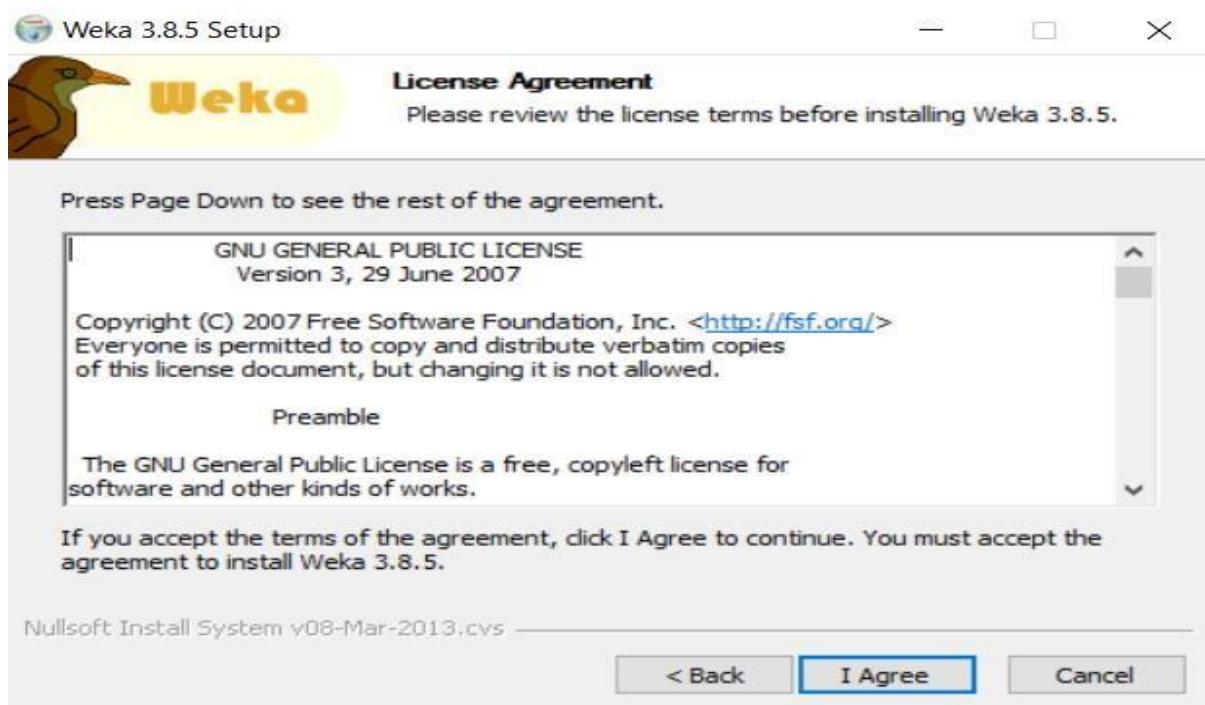
Thus, the use of WEKA results in a quicker development of machine learning models on the whole.

To install WEKA on your machine:

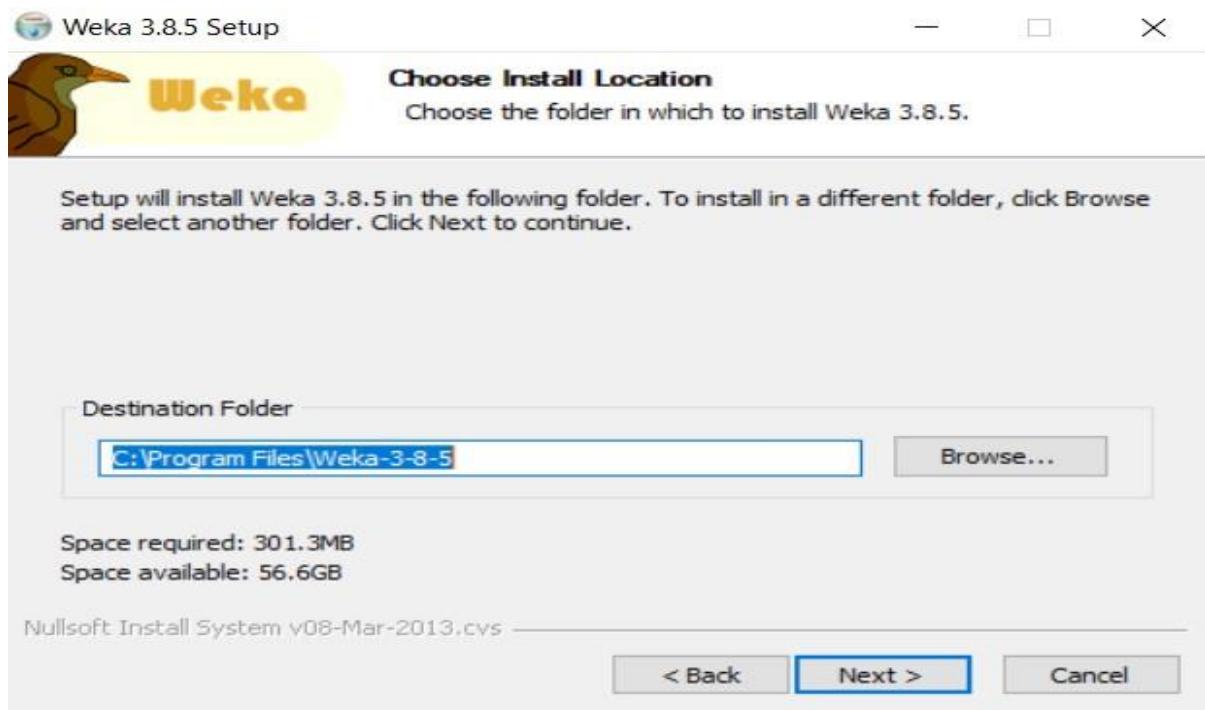
Step 1: Browse weka .org /io in google.



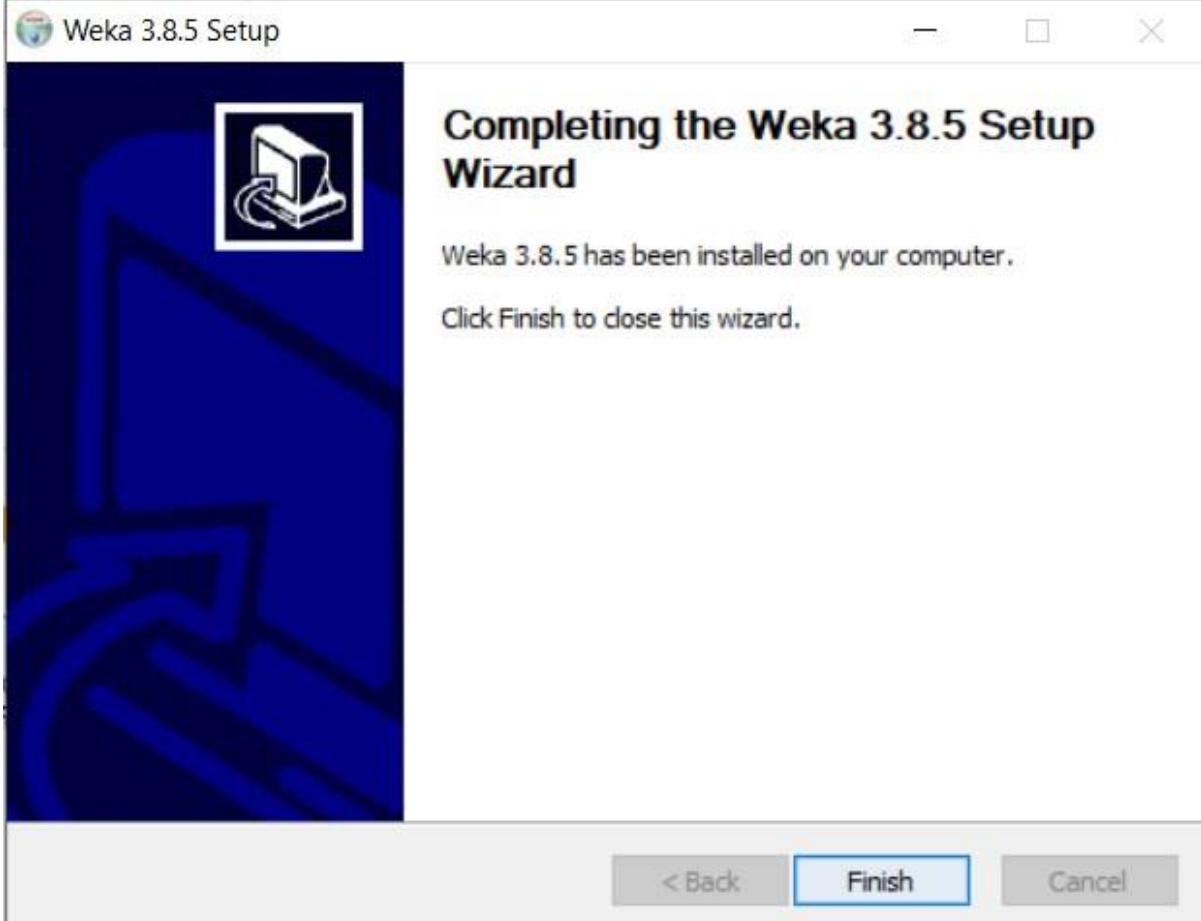
Step 2: Click on the download button for the installation of the weka tool in the pc.



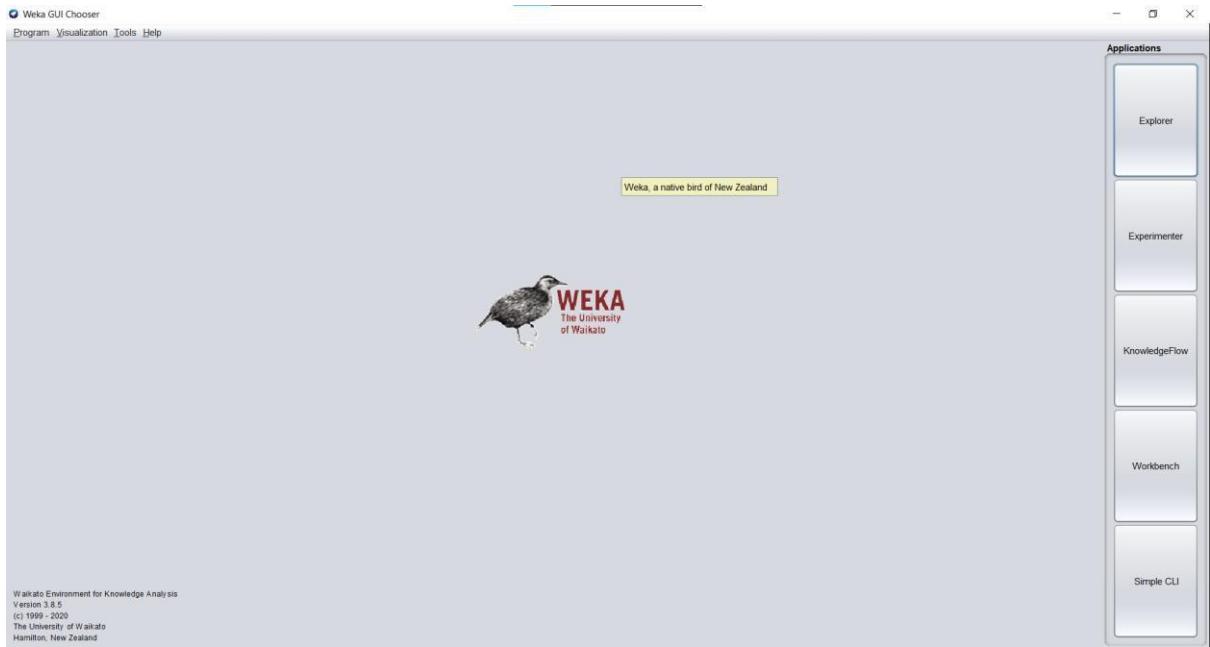
Step 3: Set the path to where the weka tool has to be downloaded and installed.



Step 4: Finally the installation is completed.



Step 5: Open the Weka Tool



The GUI Chooser application allows you to run five different types of applications as listed here –

- Explorer
- Experimenter
- KnowledgeFlow
- Workbench
- Simple CLI

CHAPTER 2

2.1 Create a Marvel Movies Table with the help of Data Mining Tool Weka.

[Weka](#) is a data mining visualization tool which contains a collection of machine learning algorithms for data mining tasks. It is an open source software issued under the GNU General Public License. It provides result information in the form of charts, trees, tables etc. Weka expects the data file to be in Attribute-Relation File Format (ARFF) file. So, first we have to convert any file into ARFF before we start mining with it in Weka.

Data Mining is a key process in analyzing Big Data. It is the computational process of unfolding patterns in the large raw data set which will support making right business decisions and designing strategies for organizational growth.

Raw Data in data mining process could be anything as below:

- CSV files (comma separated values)
- Data warehouse
- CRM
- Transactional Data ● Text, fact files, etc.

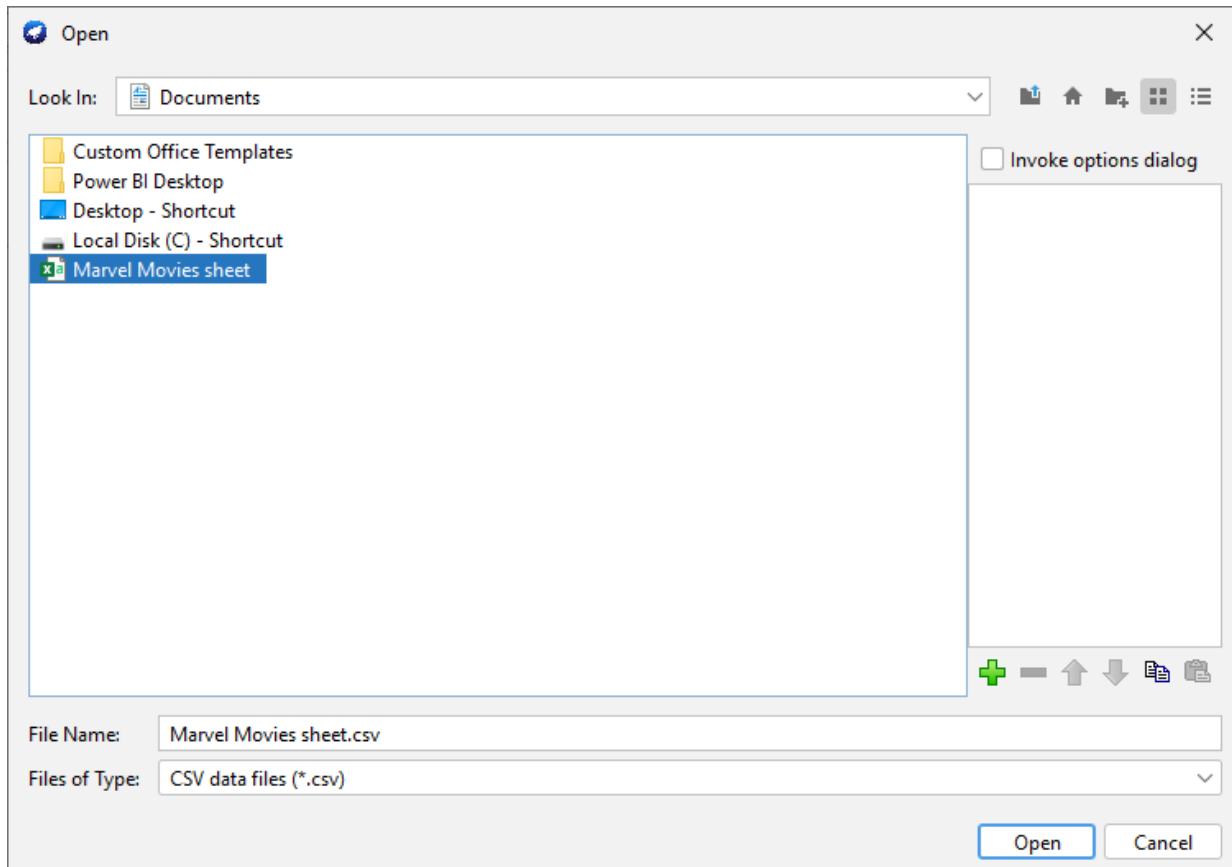
Data mining process is mainly applied on data warehouses (collection of large amounts of data) by using query methods to generate results. This process identifies relationships between input data, analyzes patterns and extracts information which gets transformed into user understandable format like, dashboard, tables, charts, reports, etc.

To create Marvel Movies Table -

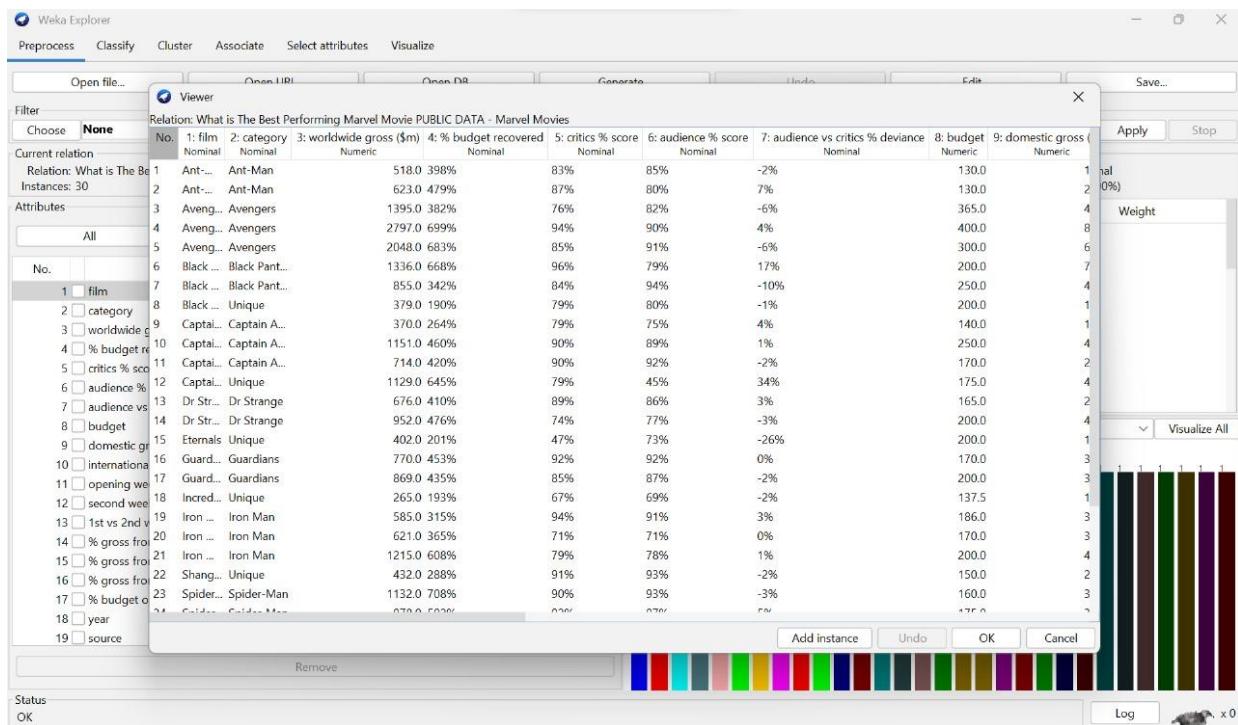
Step 1: Open the Explorer Window in the Weka tool.

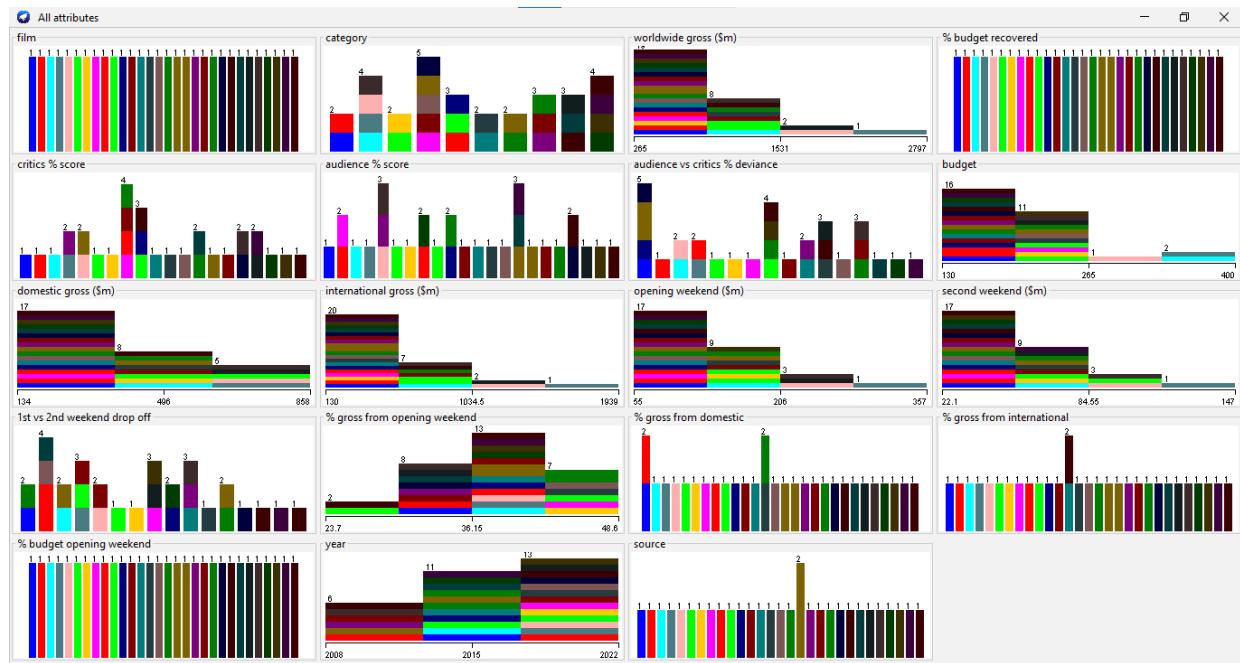
Step 2: Click the open file option which opens a tab with files

Step 3: Select the Marvel Movies.csv file and choose the file type as .csv in the open window



Step 4: Click on the edit button in the Explorer window which opens the required Marvel movies data



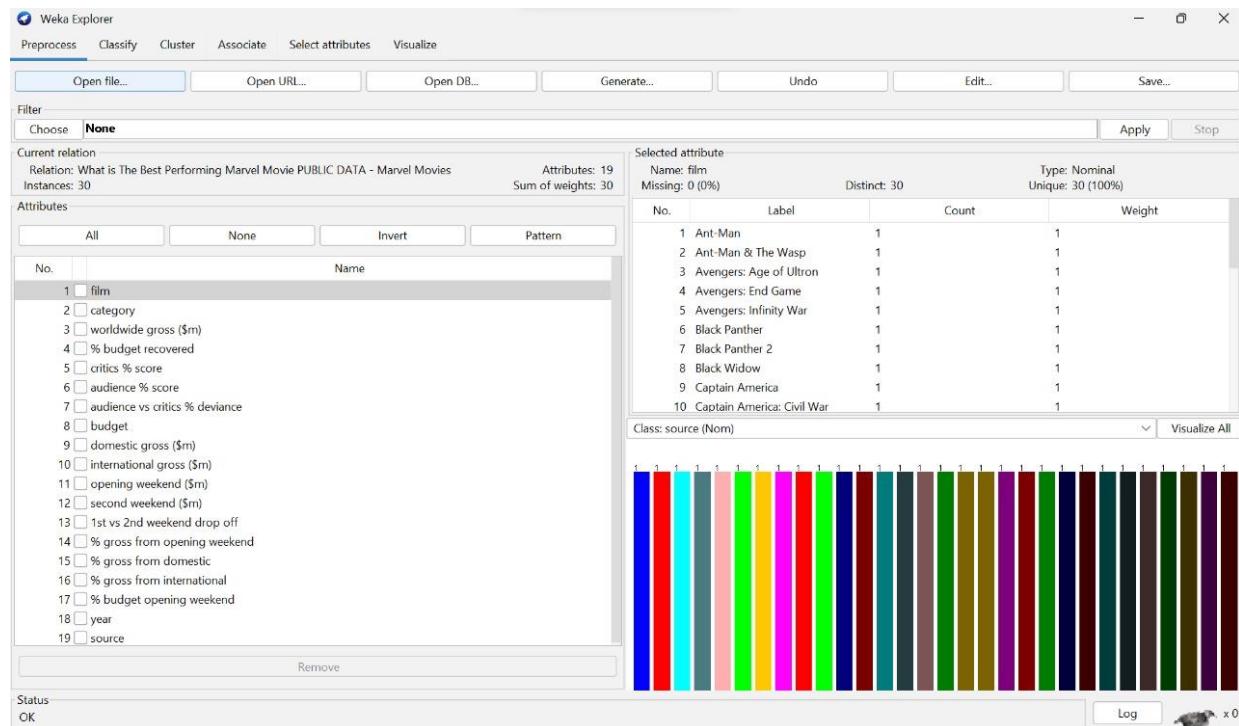


2.2 Apply Pre-Processing techniques to the training data set of Marvel Movies Table.

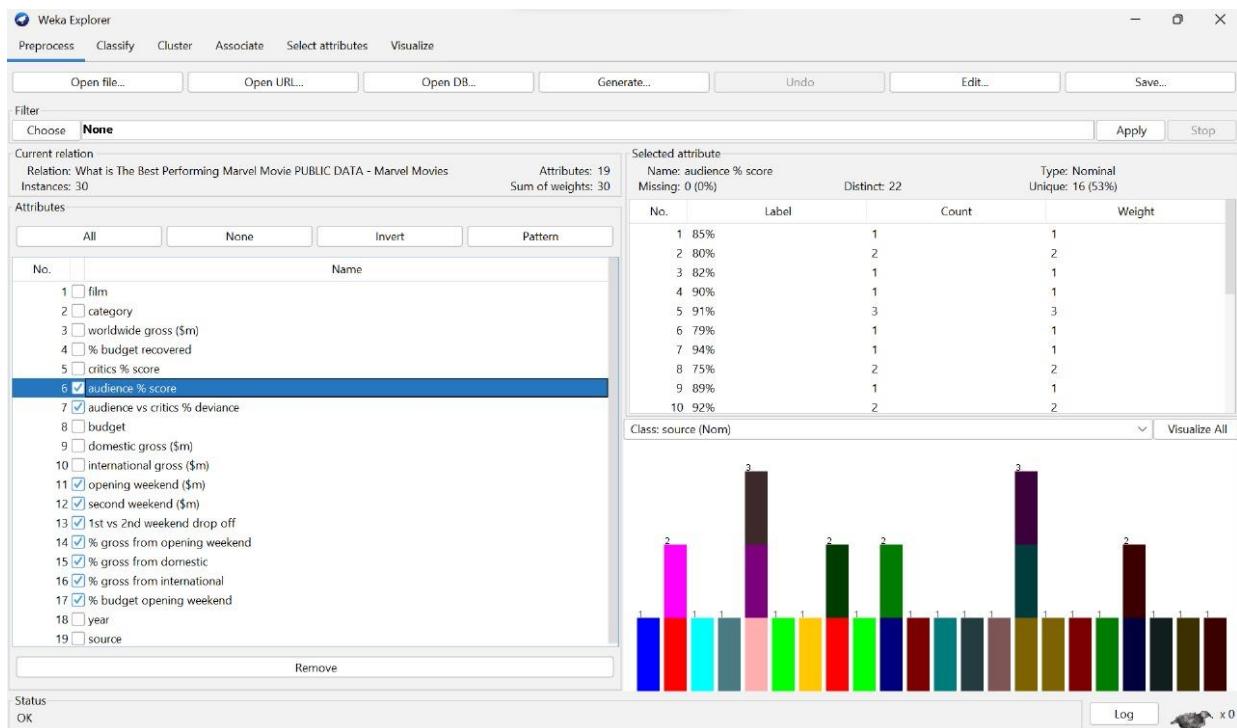
Real world databases are highly influenced by noise, missing and inconsistency due to their queue size so the data can be pre-processed to improve the quality of data and missing results and it also improves the efficiency.

Loading the Data

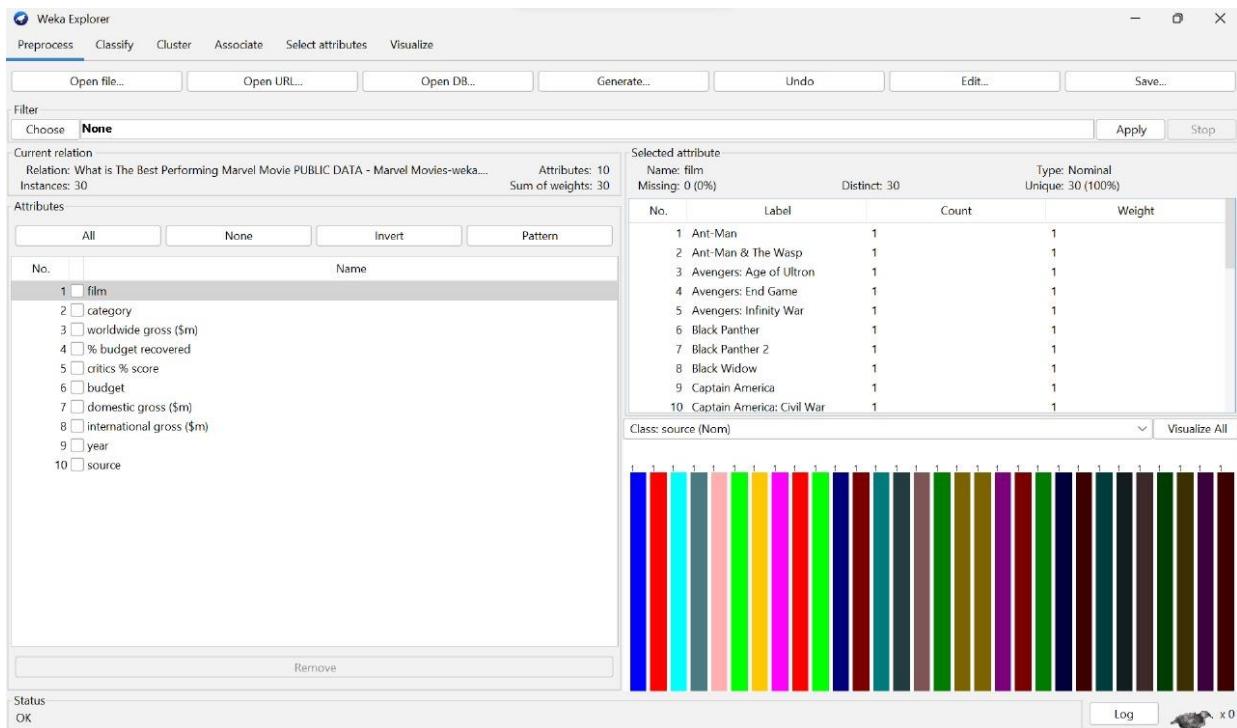
In addition to the native ARFF data file format, WEKA has the capability to read in ".csv" format files. This is fortunate since many databases or spreadsheet applications can save or export data into flat files in this format.



-> selecting and removing unnecessary attributes



-> needed attributes after removing



CHAPTER 3

3.1 Normalize the Marvel Movies Table data using KnowledgeFlow.

Data normalization is the process of rescaling one or more attributes to the range of 0 to 1.

This means that the largest value for each attribute is 1 and the smallest value is 0.

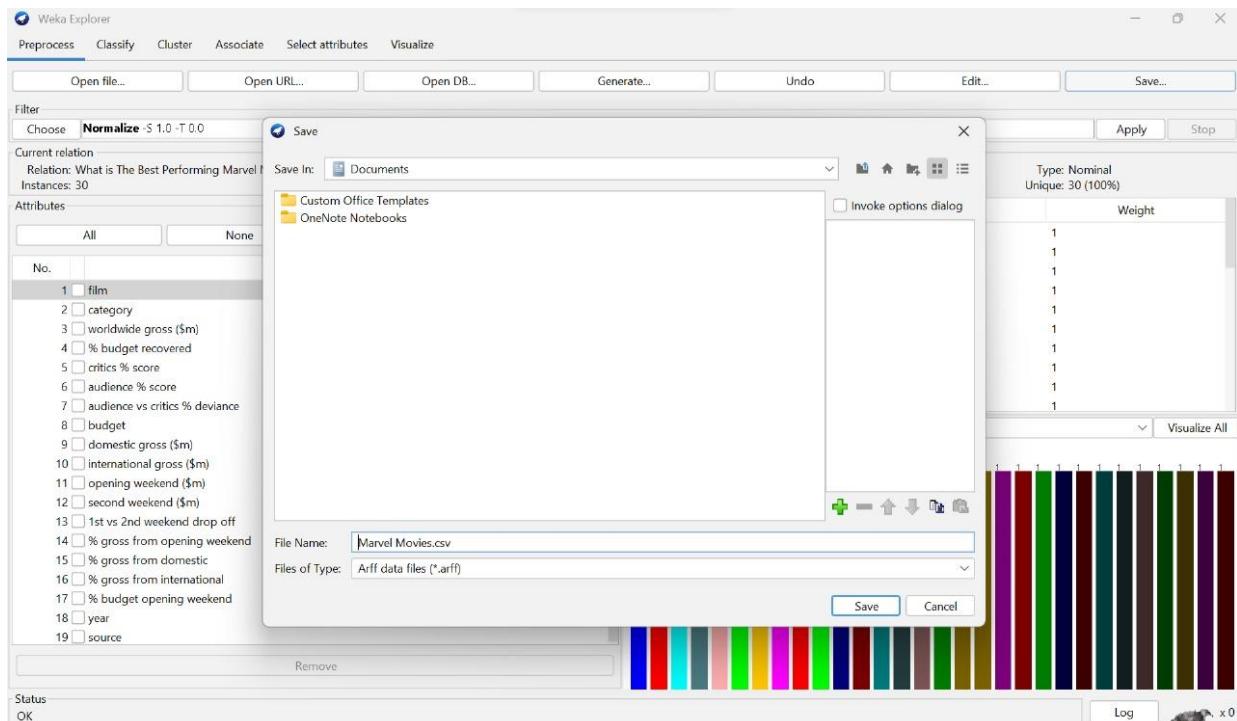
Normalization is a good technique to use when you do not know the distribution of your data or when you know the distribution is not Gaussian (a bell curve).

You can normalize all of the attributes in your dataset with Weka by choosing the Normalize filter and applying it to your dataset.

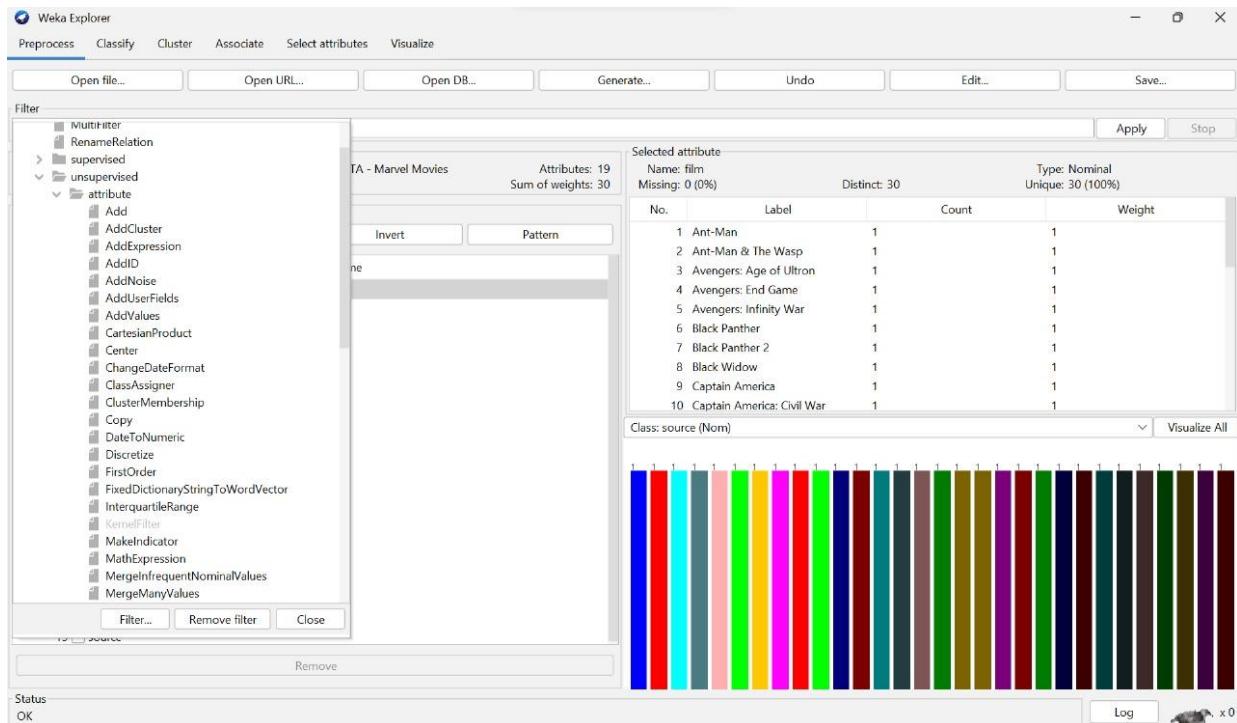
You can use the following recipe to normalize your dataset:

Step 1: Open the Weka Explorer.

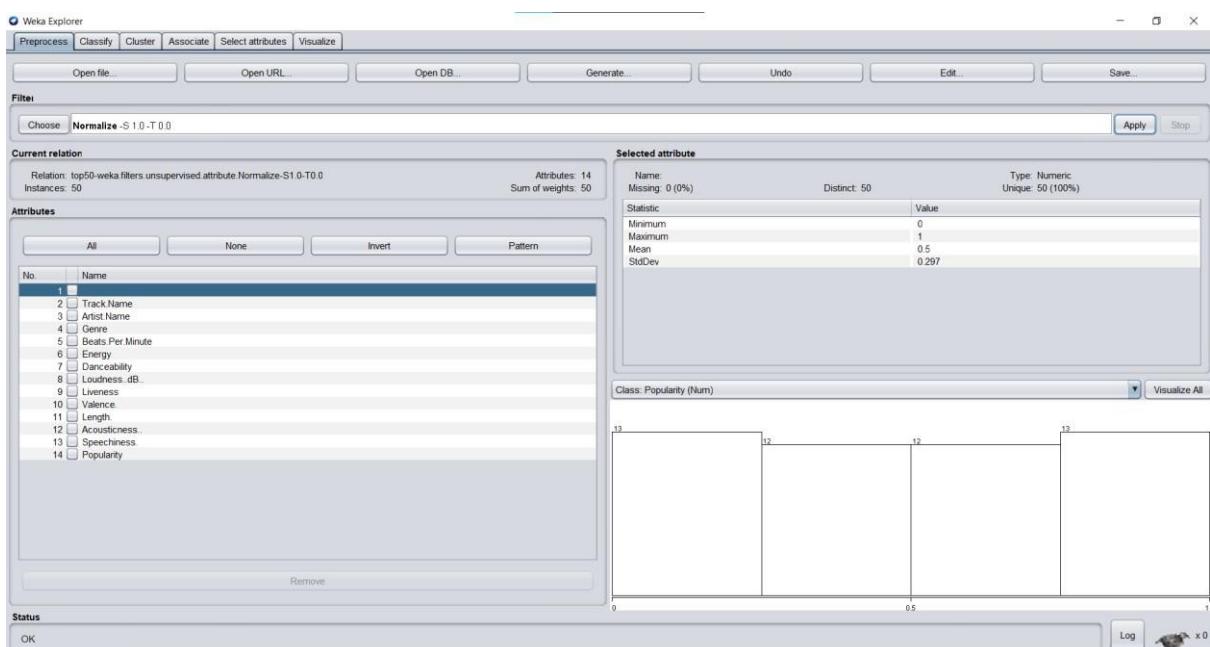
Step 2: Load your dataset.



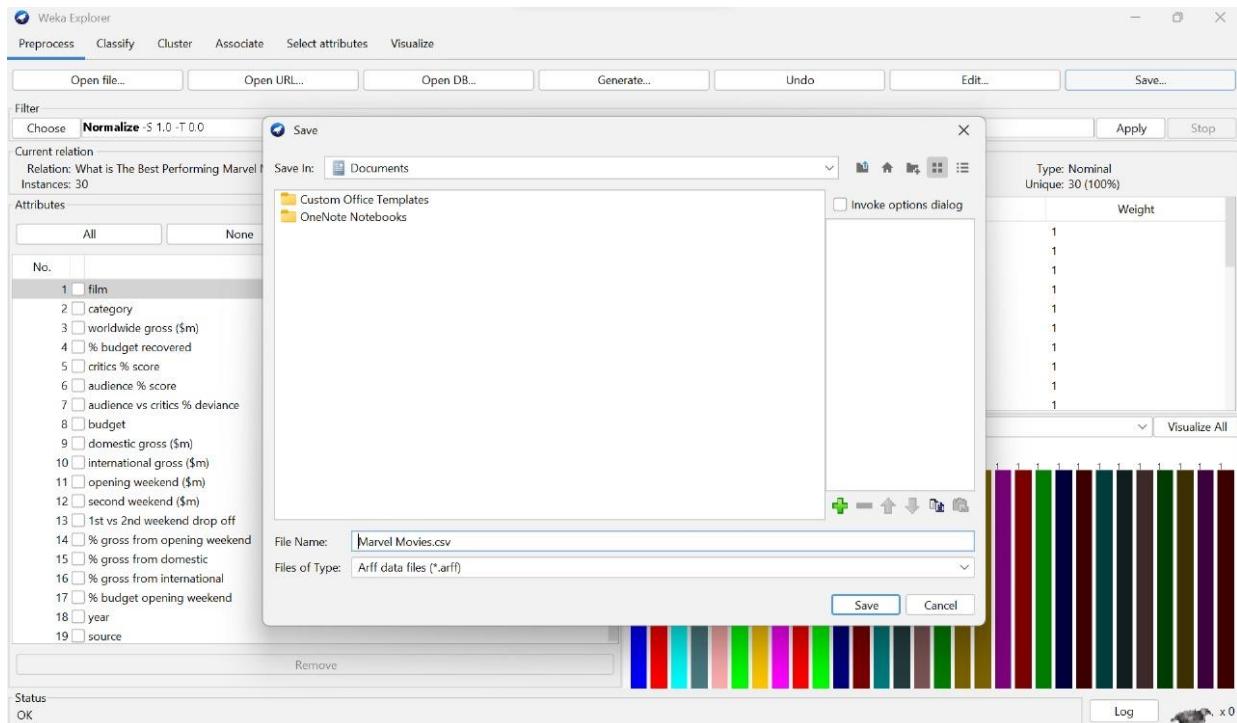
Step 3: Click the “Choose” button to select a Filter and select *unsupervised.attribute.Normalize*.



Step 4: Click the “Apply” button to normalize your dataset



Step 5: Click the “Save” button and type a filename to save the normalized copy of your dataset



Reviewing the details of each attribute in the “Selected attribute” window will give you confidence that the filter was successful and that each attribute was rescaled to the range of 0 to 1.

You can use other scales such as -1 to 1, which is useful when using support vector machines and adaboost.

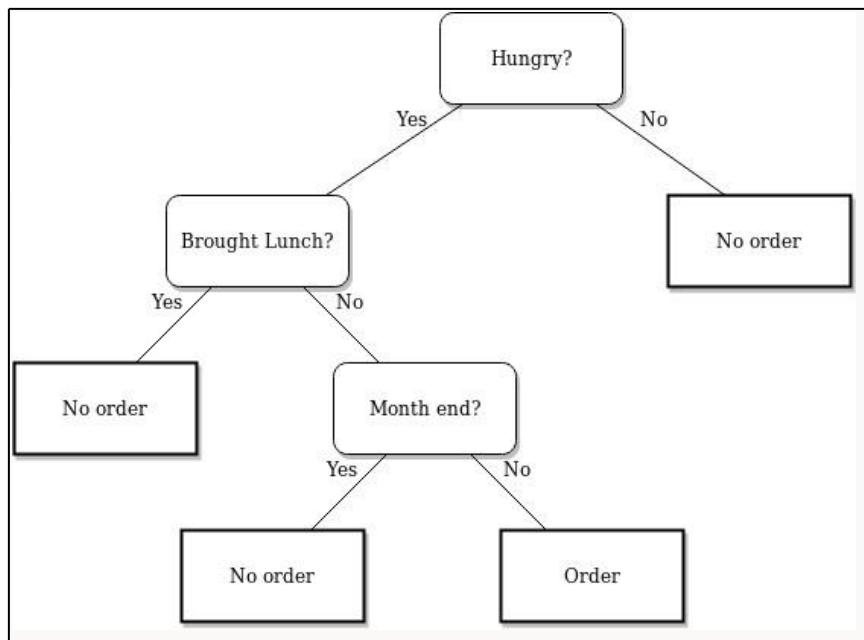
Normalization is useful when your data has varying scales and the algorithm you are using does not make assumptions about the distribution of your data, such as k-nearest neighbors and artificial neural networks.

3.2 To construct a Decision Tree for top Marvel Movies data and classify it.

Understanding Decision Trees

Decision trees are also known as **Classification And Regression Trees (CART)**. They work by learning answers to a hierarchy of if/else questions leading to a decision. These questions form a tree-like structure, and hence the name.

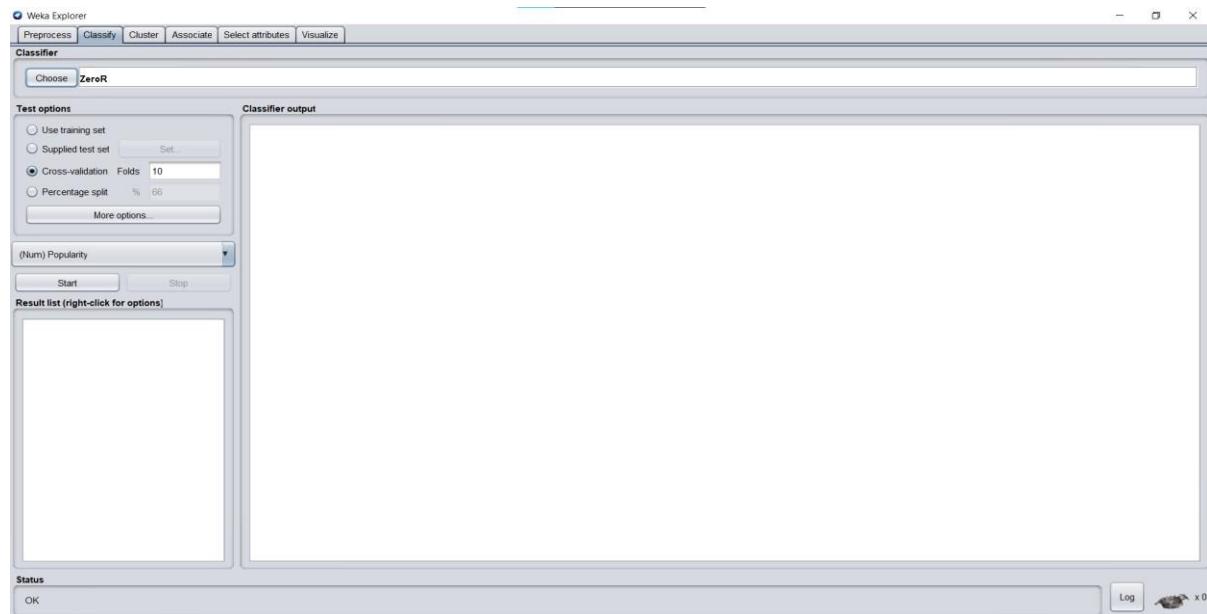
For example, let's say we want to predict whether a person will order food or not. We can visualize the following decision tree for this:



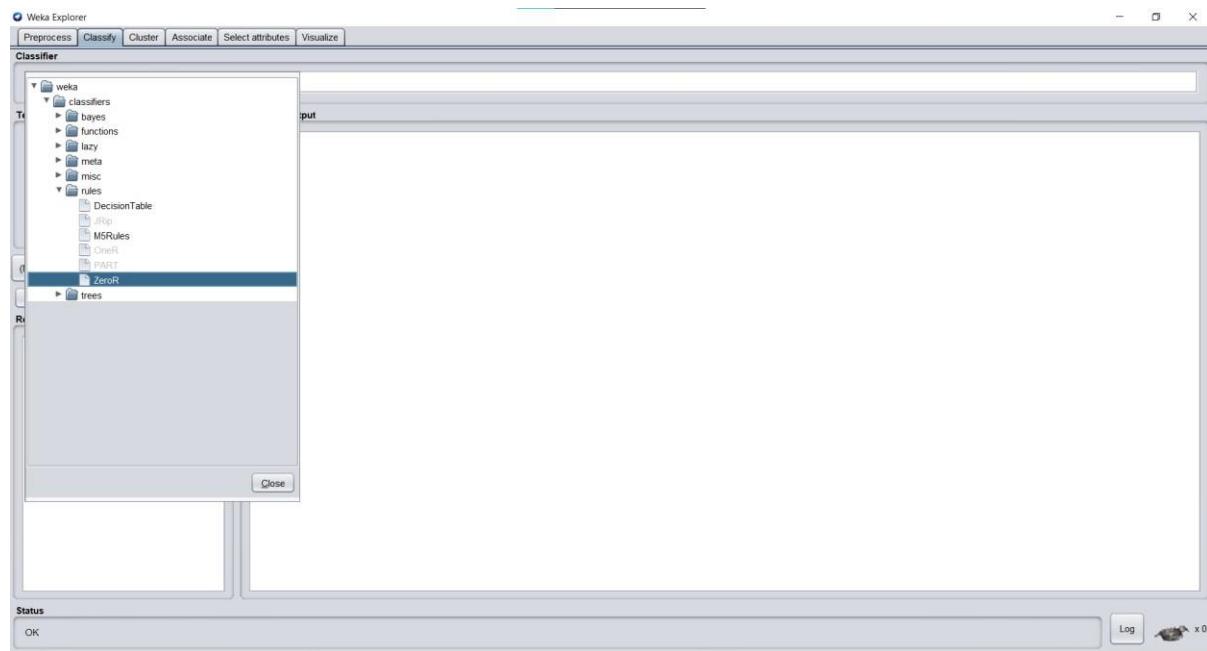
Each node in the tree represents a question derived from the features present in your dataset. Your dataset is split based on these questions until the maximum depth of the tree is reached. The last node does not ask a question but represents which class the value belongs to

Implementing a decision tree in Weka is pretty straightforward. Just complete the following steps:

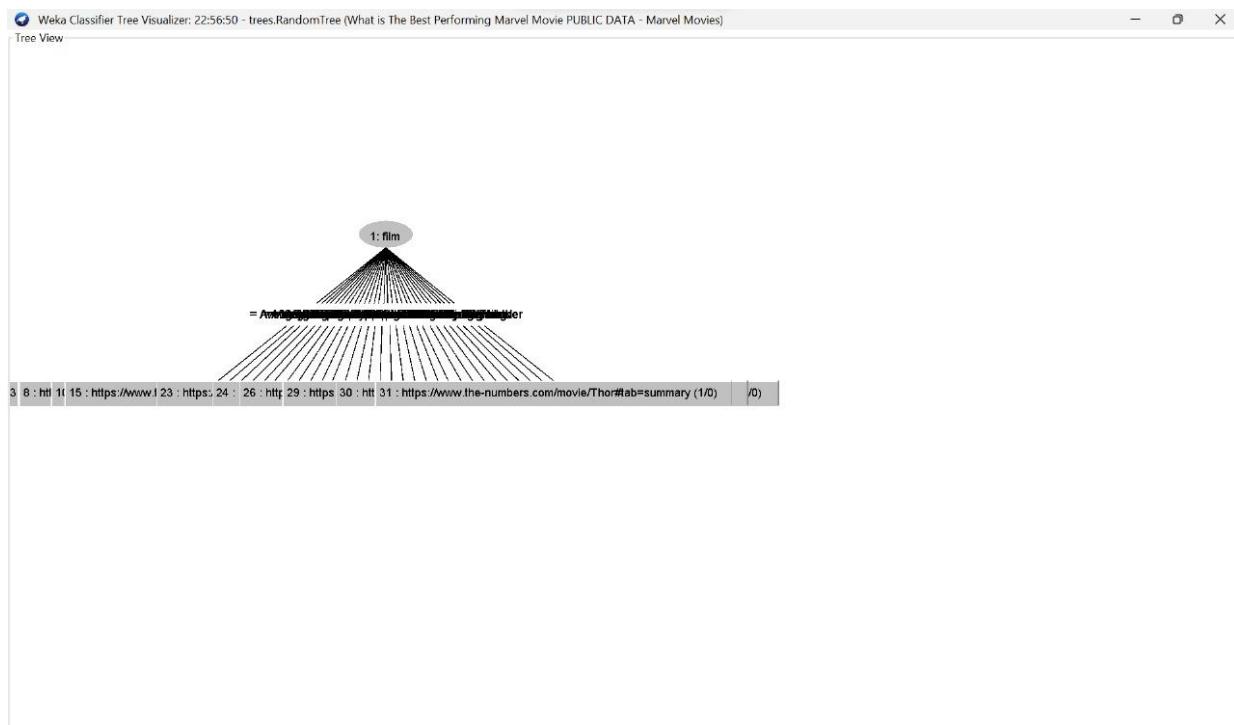
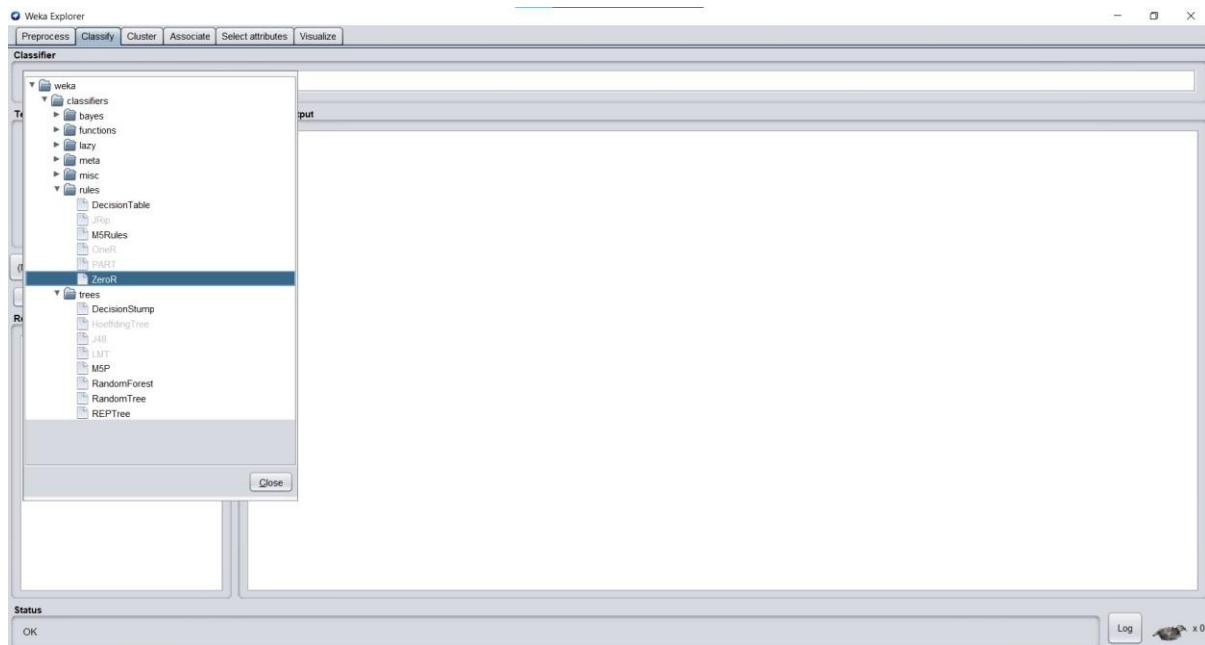
Step 1: Click on the “Classify” tab on the top



Step 2: Click the “Choose” button



Step 3: From the drop-down list, select “trees” which will open all the tree algorithms



```

Size of the tree : 31

Time taken to build model: 0 seconds

==== Stratified cross-validation ====
==== Summary ====

Correctly Classified Instances      0          0      %
Incorrectly Classified Instances   30         100    %
Kappa statistic                   -0.0638
Mean absolute error               0.0688
Root mean squared error           0.1889
Relative absolute error           101.69    %
Root relative squared error      101.828   %
Total Number of Instances         30


```

CONCLUSION

WEKA is an efficient data mining tool to perform many data mining tasks as well as experiment with new methods over datasets. WEKA has been developed by the Department of Computer Science, the University of Waikato in New Zealand.

Today's world is overwhelmed with data right from shopping in the supermarket to security cameras at our home. Data mining uses this raw data, converts it to information to make predictions. WEKA with the help of the Apriori Algorithm helps in mining association rules in the dataset. Apriori is a frequent pattern mining algorithm that counts the number of occurrences of an itemset in the transaction.

Cluster Analysis is a technique to find out clusters of data that represent similar characteristics. WEKA provides many algorithms to perform cluster analysis out of which simplekmeans are highly used.

Data Visualization in WEKA can be performed on all datasets in the WEKA directory. The raw dataset can be viewed as well as other resultant datasets of other algorithms such as classification, clustering, and association can be visualized using WEKA

CHAPTER 4

4.1 INTRODUCTION TO MICROSOFT EXCEL

Microsoft Excel is a software application designed for creating tables to input and organize data. It provides a user-friendly way to analyze and work with data. The image below provides a visual representation of what an Excel spreadsheet typically appears like.

1)TO PERFORM SUM OPERATION:

Type =SUM in a cell, followed by an opening parenthesis (). To enter the first formula range, which is called an argument (a piece of data the formula needs to run), type A2:A4 (or select cell A2 and drag through cell A6). Type a comma (,) to separate the first argument from the next.

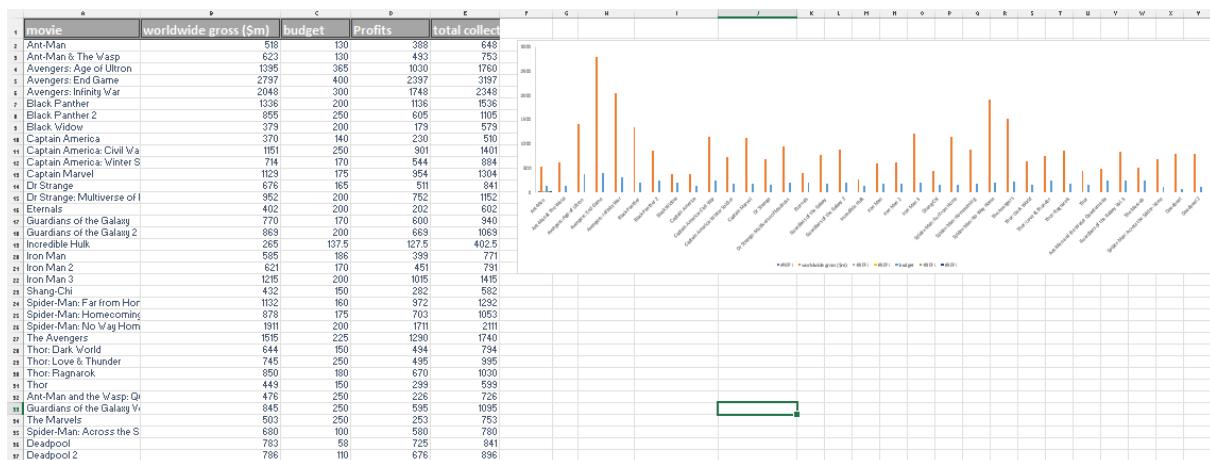
2)TO PERFORM AVERAGE:

The formula is displayed in the formula bar, =AVERAGE (A2:A7) if you're using the sample data. In the Formula Bar, select the content between the parentheses, which is A2:A7 if you're using the sample data and click the cells that you want to average, and then press ENTER.

3)TO PERFORM BOOLEAN:

A Boolean is a data type with only two possible values, TRUE or FALSE. You'll often see Boolean results or Boolean expressions in Excel. For example, if I enter the formula =B5>30 here, we'll get the Boolean result of TRUE. This is a Boolean expression—or logical statement—that returns either TRUE or FALSE.

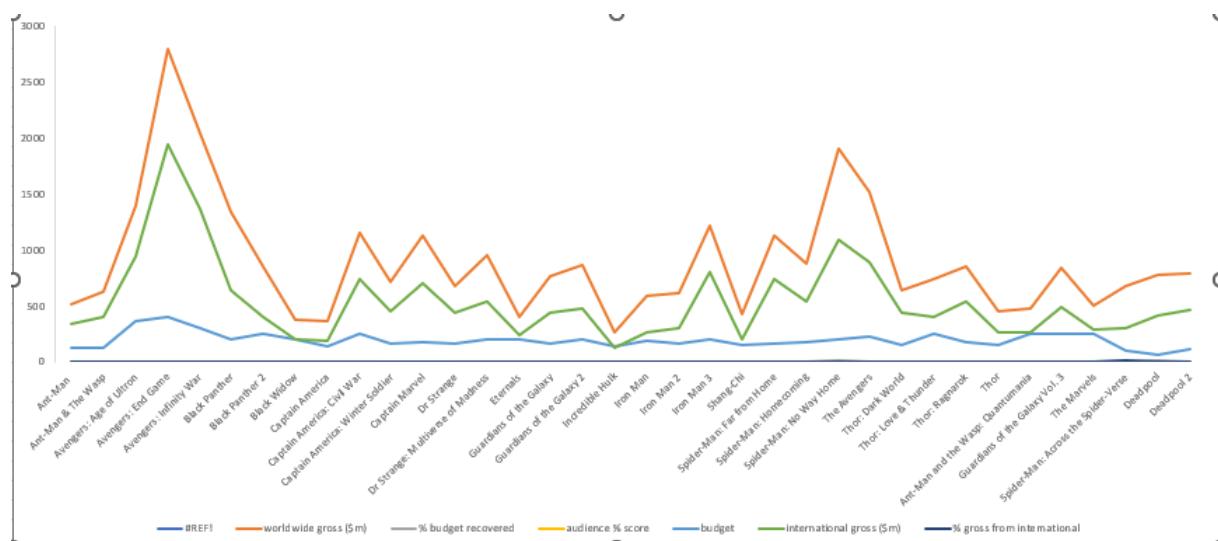
4.2 THE SUM, AVERAGE, BOOLEAN OF THE DATASET AND THE PIVOT CHART:

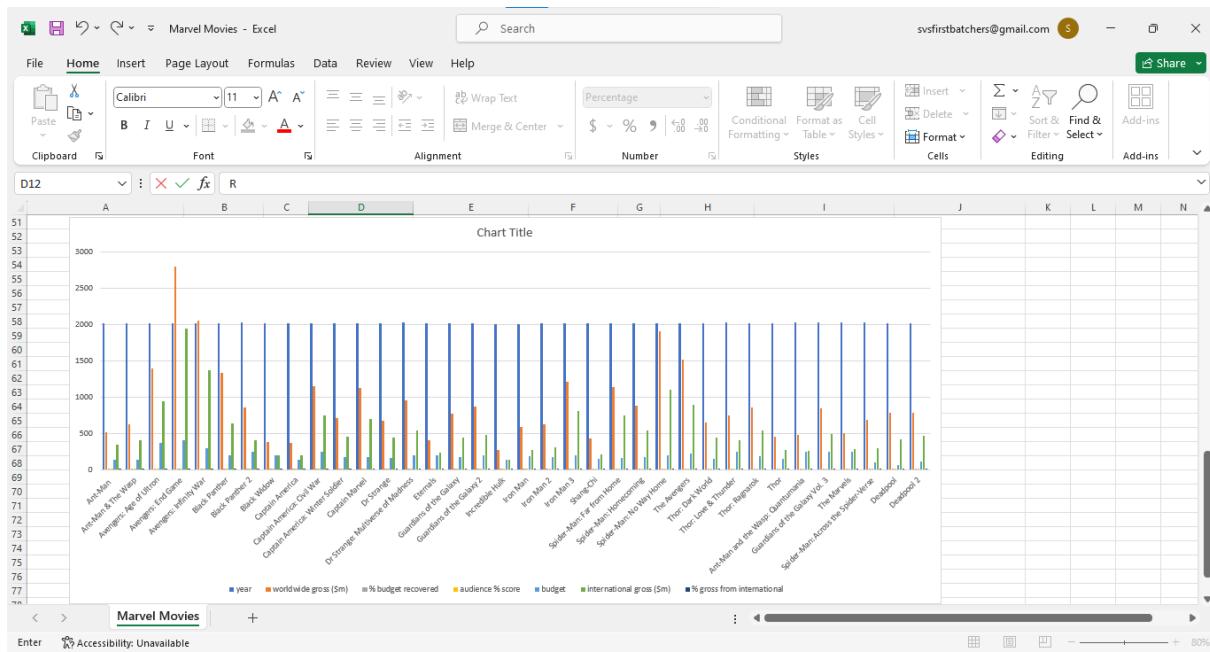


Here the sum, average and Boolean of the dataset has been displayed with the help of formulas.

THE BAR GRAPH AND THE LINE GRAPH OF THE DATASET HAS BEEN DISPLAYED

GO TO INSERT -> PIVOT CHARTS -> CHOOSE THE CHARTS AS PER REQUIREMENT.





CONCLUSION

Microsoft Excel is a powerful and versatile tool widely used for data analysis, financial modeling, reporting, and a variety of other applications. Its broad range of functions—such as mathematical computations, data visualization through charts and graphs, pivot tables for summarizing data, and its ability to automate repetitive tasks with macros—make it an essential tool for businesses, students, and professionals alike. Excel's flexibility in handling complex data sets, combined with its user-friendly interface, allows users of all skill levels to perform both simple and advanced tasks. Whether for personal finance, project management, or in-depth data analysis, Excel remains a valuable resource for efficient decision-making and problem-solving.

CHAPTER 5

5.1 INTRODUCTION TO POWER BI

Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights. Your data might be an Excel spreadsheet, or a collection of cloud-based and on-premises hybrid data warehouses. Power BI lets you easily connect to your data sources, visualize and discover what's important, and share that with anyone or everyone you want.

5.2 THE PARTS OF POWER BI

Power BI consists of several elements that all work together, starting with these three basics:

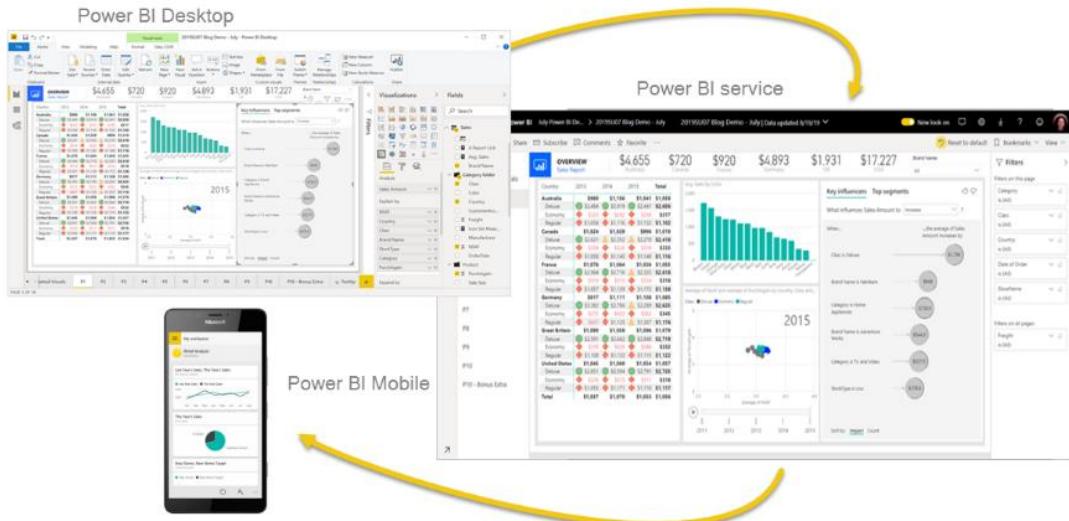
- A Windows desktop application called *Power BI Desktop*.
- An online software as a service (SaaS) service called the *Power BI service*.
- Power BI Mobile apps for Windows, iOS, and Android devices.

These three elements—Power BI Desktop, the service, and the mobile apps—are designed to let you create, share, and consume business insights in the way that serves you and your role most effectively.

Beyond those three, Power BI also features two other elements:

Power BI Report Builder, for creating paginated reports to share in the Power BI service. Read more about [paginated reports](#) later in this article.

Power BI Report Server, an on-premises report server where you can publish your Power BI reports, after creating them in Power BI Desktop. Read more about [Power BI Report Server](#) later in this article.

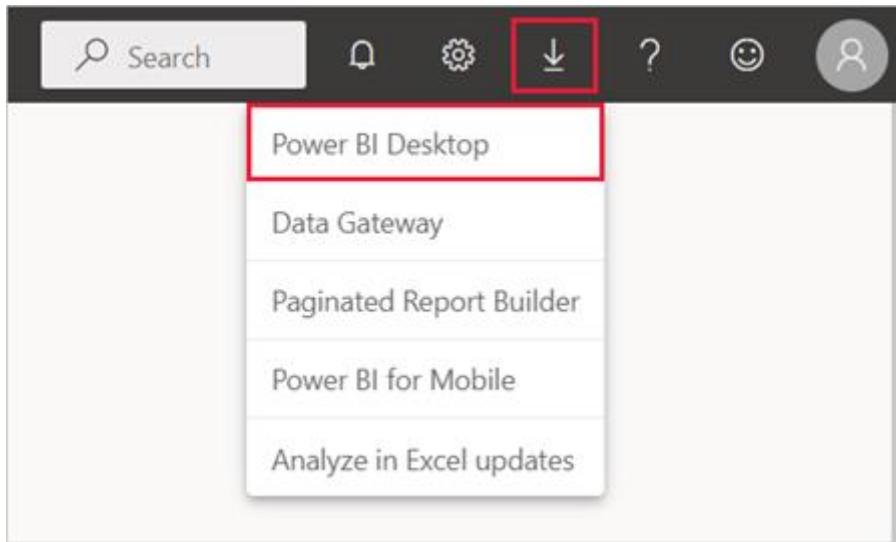


5.3 THE FLOW OF WORK IN POWER BI

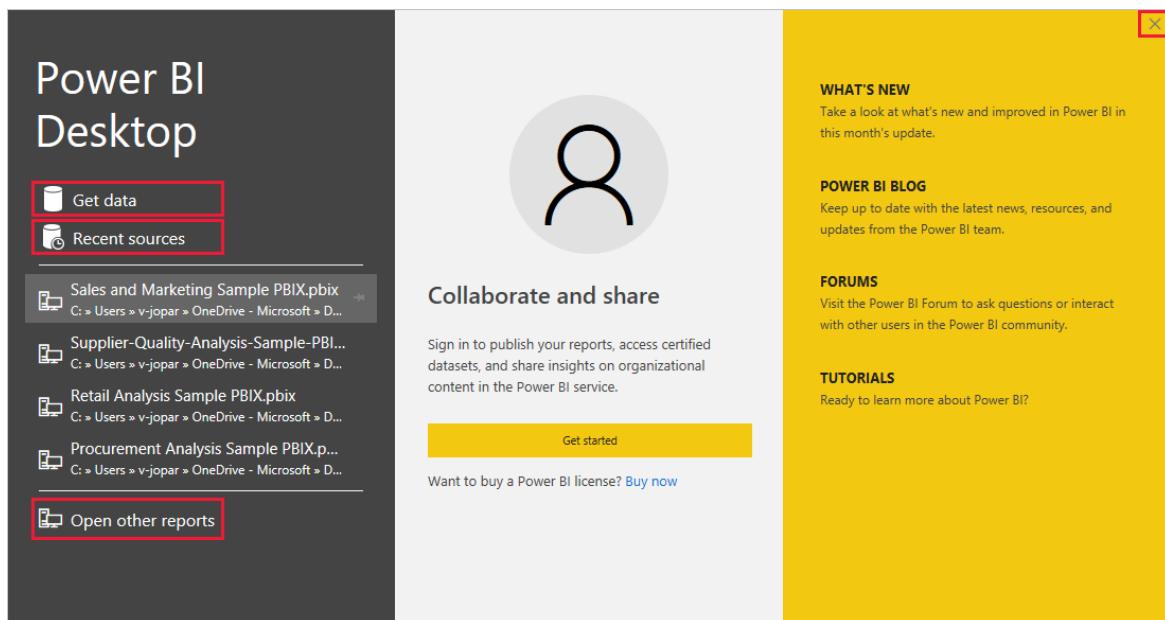
- One common workflow in Power BI begins by connecting to data sources in Power BI Desktop and building a report.
- You then publish that report from Power BI Desktop to the Power BI service, and share it so business users in the Power BI service and on mobile devices can view and interact with the report.
- This workflow is common, and shows how the three main Power BI elements complement one another.

5.4 INSTALL AND RUN POWER BI

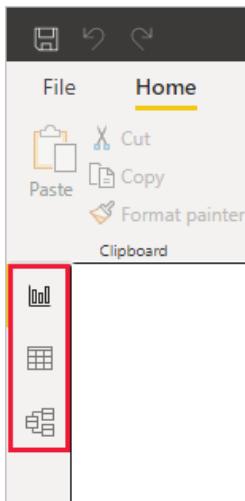
- To download Power BI Desktop, go to the [Power BI Desktop download page](#) and select **Download Free**. Or for download options, select [See download or language options](#).
- You can also download Power BI Desktop from the Power BI service. Select the **Download** icon in the top menu bar, and then select **Power BI Desktop**.



- On the Microsoft Store page, select **Get**, and follow the prompts to install Power BI Desktop on your computer. Start Power BI Desktop from the Windows **Start** menu or from the icon in the Windows taskbar.
- The first time Power BI Desktop starts, it displays the **Welcome** screen.
- From the **Welcome** screen, you can **Get data**, see **Recent sources**, open recent reports, **Open other reports**, or select other links. Select the close icon to close the **Welcome** screen.



- Along the left side of Power BI Desktop are icons for the three Power BI Desktop views: **Report**, **Data**, and **Model**, from top to bottom. The current view is indicated by the yellow bar along the left, and you can change views by selecting any of the icons.
- If you're using keyboard navigation, press Ctrl + F6 to move focus to that section of buttons in the window. To learn more about accessibility and Power BI, visit our [accessibility articles](#).



Report view is the default view.

The screenshot shows the Power BI Desktop interface with the Home tab selected. The ribbon menu includes File, Home, Insert, Modeling, View, and Help. The Home tab's ribbon bar contains various tools like Transform data, New visual, and Publish. On the left, there's a sidebar with navigation icons (grid, list, tree) and a main workspace. The workspace is divided into three main sections: Filters, Visualizations, and Fields. The Filters section shows 'Filters on this page' and 'Filters on all pages' with 'Add data fields here' buttons. The Visualizations section shows a grid of visualization icons. The Fields section shows a search bar and a message: 'You haven't loaded any data yet. Get data'. At the bottom, there are page navigation buttons (Page 1, +) and a status bar indicating 'Page 1 of 1'.

- Power BI Desktop also includes the **Power Query Editor**, which opens in a separate window.
- In **Power Query Editor**, you can build queries and transform data, then load the refined data model into Power BI Desktop to create reports.

CHAPTER 6

6.1 Power BI Desktop:

Power BI Desktop is a downloadable application that allows you to connect to various data sources, transform data, and create reports.

□Get Data:

- Click the **Home** tab, then click **Get Data**.
- A window will appear showing different data source options like **Excel**, **SQL Server**, **Web**, **CSV**, **JSON**, etc.

□Select the Data Source:

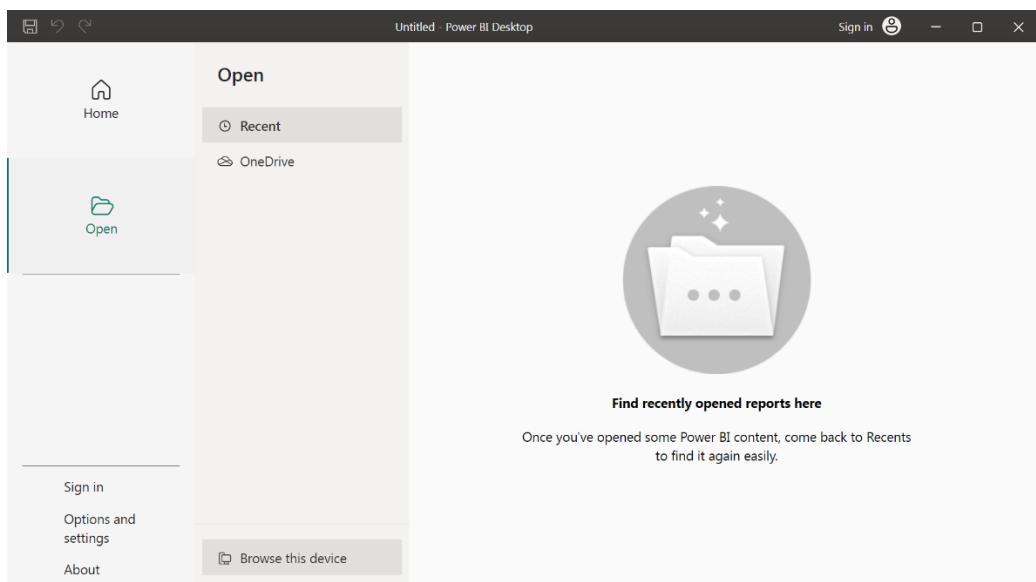
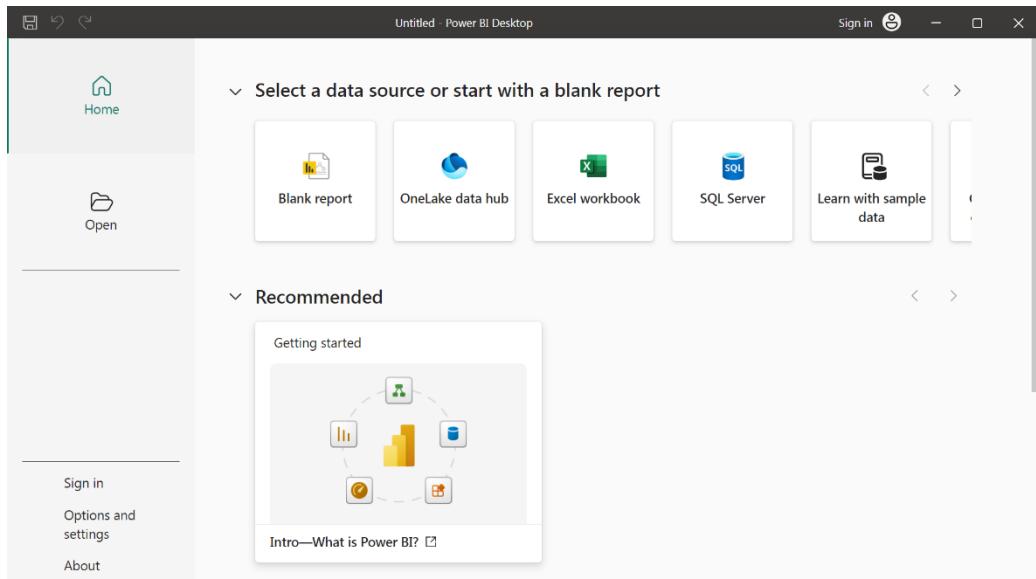
- Choose the type of data source you want to upload (e.g., **Excel** for a spreadsheet, **SQL Server** for a database, etc.).
- Click **Connect**.

□Load the Data:

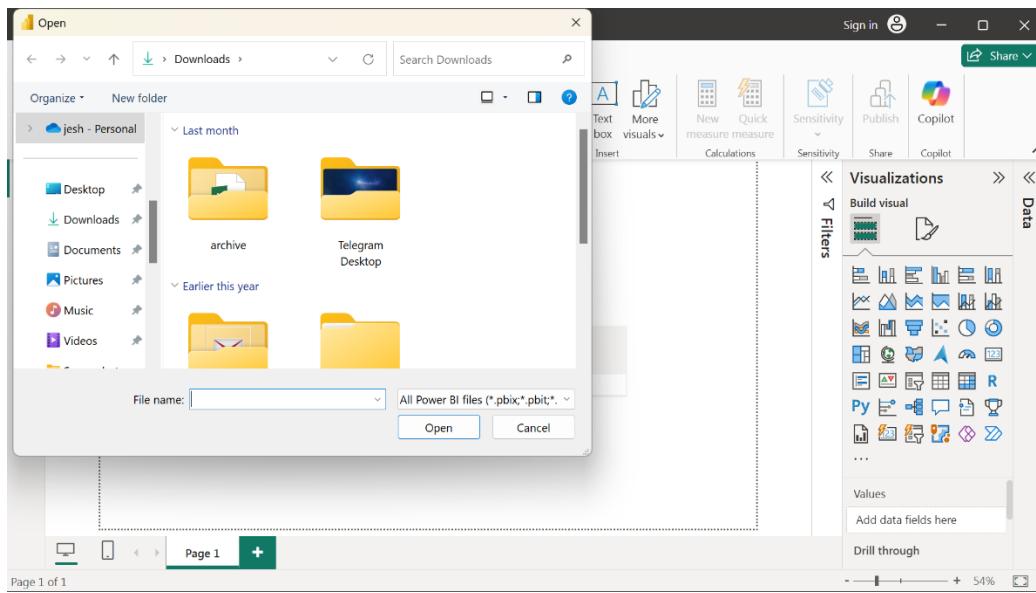
- Navigate to the data file on your computer or provide credentials for the data connection.
- After selecting the file or entering details, click **Load** or **Transform Data** if you need to clean or prepare it using **Power Query Editor**.

□Visualize the Data:

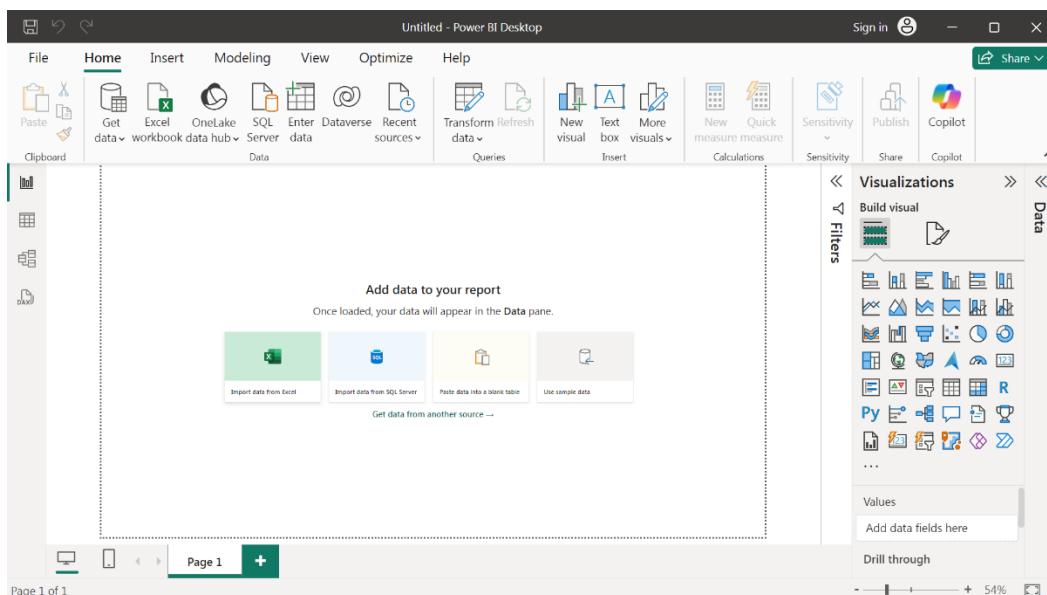
- Once the data is loaded, it will appear in the **Fields** pane.
- You can now drag fields to the report canvas and create visualizations.



- Choose the file type to be uploaded.



- Add the data to your report.



- After uploading the dataset, the data needs to be transformed as per user needs and then to be loaded to perform visualization using POWER BI.

6.2 APPLY TRANSFORMATION TECHNIQUES

- If you need to change a data type, select the column or columns to change.
- Hold down the **Shift** key to select several adjacent columns, or **Ctrl** to select non-adjacent columns.
- Either right-click a column header, select **Change Type**, and choose a new data type from the menu, or drop down the list next to **Data Type** in the **Transform** group of the **Home** tab, and select a new data type.

The screenshot shows the Power BI Desktop application interface. In the center, there is a table titled "Marvel Movies.csv" with the following data:

movie	category	year	worldwide gross (\$m)	% budget recovered	critics % score	audience % score	aud
Ant-Man	Ant-Man	2015	518	398.00%	83.00%	85.00%	
Ant-Man & The Wasp	Ant-Man	2018	623	479.00%	87.00%	80.00%	
Avengers: Age of Ultron	Avengers	2015	1395	382.00%	76.00%	82.00%	
Avengers: End Game	Avengers	2019	2797	699.00%	94.00%	90.00%	
Avengers: Infinity War	Avengers	2018	2048	683.00%	85.00%	91.00%	
Black Panther	Black Panther	2018	1336	668.00%	96.00%	79.00%	
Black Panther 2	Black Panther	2022	855	342.00%	84.00%	94.00%	
Black Widow	Unique	2021	379	190.00%	79.00%	80.00%	
Captain America	Captain America	2011	370	264.00%	79.00%	75.00%	
Captain America: Civil War	Captain America	2016	1151	460.00%	90.00%	89.00%	
Captain America: Winter Soldier	Captain America	2014	714	420.00%	90.00%	92.00%	
Captain Marvel	Unique	2019	1129	645.00%	79.00%	45.00%	
Dr Strange	Dr Strange	2016	676	410.00%	89.00%	86.00%	
Dr Strange: Multiverse of Madness	Dr Strange	2022	952	476.00%	74.00%	77.00%	
Eternals	Unique	2021	402	201.00%	47.00%	73.00%	
Guardians of the Galaxy	Guardians	2014	770	453.00%	92.00%	92.00%	
Guardians of the Galaxy 2	Guardians	2017	869	435.00%	85.00%	87.00%	
Incredible Hulk	Unique	2008	265	193.00%	67.00%	69.00%	
Iron Man	Iron Man	2008	585	315.00%	94.00%	91.00%	
Iron Man 2	Iron Man	2010	621	365.00%	71.00%	71.00%	

At the bottom of the clipboard interface, there are three buttons: "Extract Table Using Examples", "Load", and "Transform Data".

With the help of transform menu, the dataset can be altered as per the user requirement.

Queries [1] ✓ = Table.TransformColumnTypes(#"Promoted Headers", {{"movie", type text}, {"category", type text}, {"year", type text}, {"worldwide gross (\$M)", type text}}

	movie	category	year	worldwide gross (\$M)
1	Ant-Man	Ant-Man	2015	
2	Ant-Man & The Wasp	Ant-Man	2018	
3	Avengers: Age of Ultron	Avengers	2015	
4	Avengers: End Game	Avengers	2019	
5	Avengers: Infinity War	Avengers	2018	
6	Black Panther	Black Panther	2018	
7	Black Panther 2	Black Panther	2022	
8	Black Widow	Unique	2021	
9	Captain America	Captain America	2011	
10	Captain America: Civil War	Captain America	2016	
11	Captain America: Winter Soldier	Captain America	2014	
12	Captain Marvel	Unique	2019	
13	Dr Strange	Dr Strange	2016	
14	

18 COLUMNS, 36 ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 7:07 PM

Queries [1] ✓ = Table.TransformColumnTypes(#"Promoted Headers", {{"opening weekend", type percentage}, {"% gross from domestic", type percentage}, {"% gross from international", type percentage}, {"% budget opening weekend", type percentage}}

	opening weekend	% gross from domestic	% gross from international	% budget opening weekend
1	31.8	34.70%	65.30%	43.80%
2	35	34.70%	65.20%	58.30%
3	41.7	32.90%	67.10%	52.30%
4	41.6	30.70%	69.30%	89.30%
5	38	33.10%	66.80%	85.70%
6	28.9	52.40%	47.60%	101.00%
7	48.6	53.00%	46.90%	72.40%
8	43.8	48.30%	51.70%	40.20%
9	36.8	47.60%	52.20%	46.40%
10	43.9	35.40%	64.60%	71.60%
11	36.6	36.30%	63.60%	55.90%
12	35.9	37.70%	62.20%	87.40%
13	36.6	34.30%	65.50%	51.50%
14	45.6	42.20%	57.70%	88.20%

18 COLUMNS, 36 ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 7:07 PM

- Here the date attribute is transformed.
- It can be transformed using split by column method.

The screenshot shows the Power Query Editor interface. A table named "Marvel Movies" is selected. A context menu is open under the "Transform" tab, specifically the "Split Column" option. The "By Delimiter" section is active, displaying various splitting methods. The preview pane at the bottom shows the first 10 rows of the data.

- Split by using the delimiter method.
- Give the delimiter as required, here the delimiter “/” has been used to divide the date column into three,
- 1)Month
- 2)Date
- 3)Year

The screenshot shows the Power BI Desktop interface with the "Split Column by Delimiter" dialog box open. The "Delimited" tab is selected. Under "Select or enter delimiter", the "Custom" option is chosen. The "OK" button is highlighted. The background shows the Power BI ribbon and a table named "Marvel Movies".

Now we got the transformed data, once satisfied with the transformed data we can load the dataset for visualization. Transforming data in Power BI is crucial for creating better visualizations because it ensures that the data is clean, structured, and in a format that can be easily understood by both Power BI and the end-users. Power BI uses **Power Query Editor** for transformations, allowing you to perform tasks such as cleaning, reshaping, and enriching data before it's used in visualizations.

The screenshot shows the Power Query Editor interface with the following details:

- File Bar:** Home, Transform, Add Column, View, Tools, Help.
- Toolbars:** Data Type: Text, Merge Columns, Statistics, Trigonometry, Date, Time, Duration, Run R script, Run Python script.
- Queries List:** survey_lung_cancer_with... (selected), survey_lung_cancer_with...
- Table View:** A preview of the data table with columns: Month, Date, and Year and time.
- Query Settings Panel:**
 - Properties:** Name: survey_lung_cancer_with_state_and_date.
 - Applied Steps:** Shows the history of steps applied to the query, including "Renamed Columns".
- Bottom Status:** 20 COLUMNS, 309 ROWS, Column profiling based on top 1000 rows, PREVIEW DOWNLOADED AT 6:57 PM.

CHAPTER 7

7.1 LOADING THE DATA AFTER TRANSFORMATION FOR VISUALIZATION:

After transforming your data in Power BI, you need to load it into the model for visualization. Here's how you can do that in **Power BI Desktop**, which is the typical environment where transformations are done using Power Query:

Steps to Load Data after Transformation

1. Complete Transformations in Power Query:

- Open Power Query Editor by selecting Transform Data from the Home tab in Power BI Desktop.
- Apply all the necessary transformations like filtering, removing duplicates, creating calculated columns, etc.
- After you're done transforming the data, you'll want to load it into Power BI for visualizations.

2. Close and Apply Changes:

- Once you're satisfied with your transformed data, click the **Close & Apply** button on the top left of the Power Query Editor window.
- This will close Power Query Editor and apply your transformations to the dataset.

3. Loading the Data:

- Power BI will begin loading the transformed data into its model. You'll see a loading screen as Power BI processes and applies the changes. Depending on the size of the dataset, this might take a few moments.
- Once the data is loaded, it will appear in the Fields pane on the right-hand side of the Power BI interface.

4. Start Visualizing:

- Now that the data is loaded, you can begin creating visualizations:
- Drag and drop fields: Simply drag the relevant fields (columns) from the Fields pane

to the

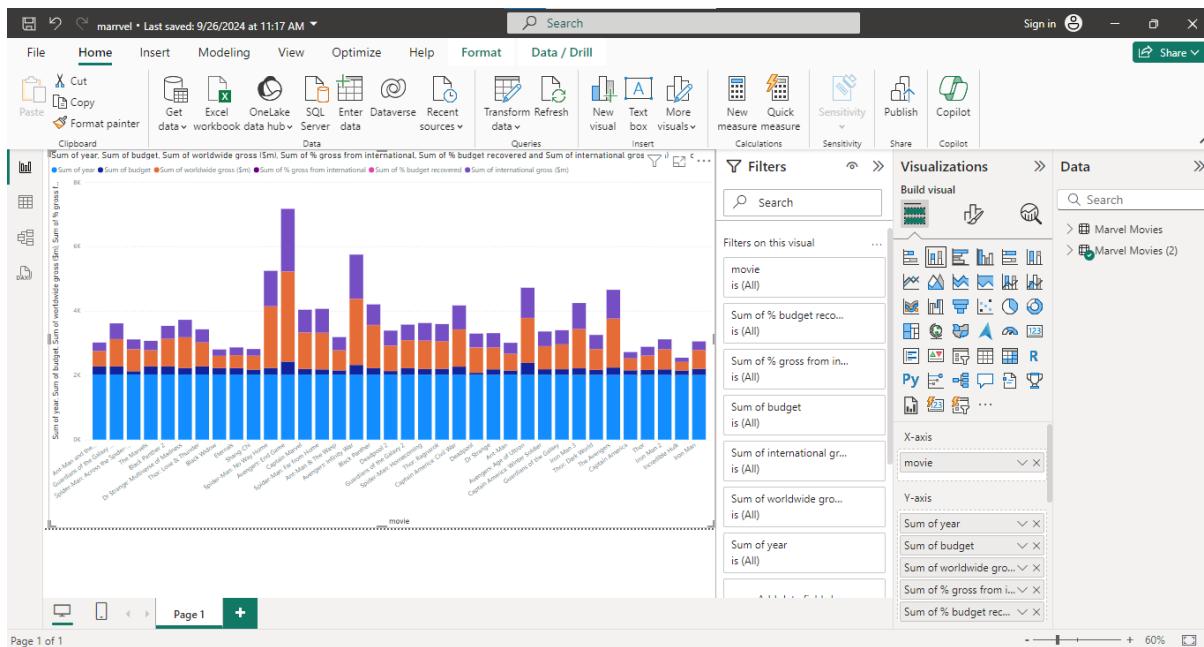
- Report Canvas to create charts, graphs, tables, and other visuals.
- Create visualizations: Power BI will automatically suggest default visualizations based on the type of data you select, but you can always change it using the Visualizations pane.
- Add filters: You can use slicers, filters, and other interactive elements to allow users to explore the data.

CHAPTER 8

8.1 VISUALIZING USING VARIOUS CHARTS:

i) STACKED COLUMN CHARTS:

A stacked column chart in Power BI is a type of bar chart where categories are represented as columns, and each column is divided into different segments, which represent sub-categories. It's useful for showing how different components contribute to the total across various categories.

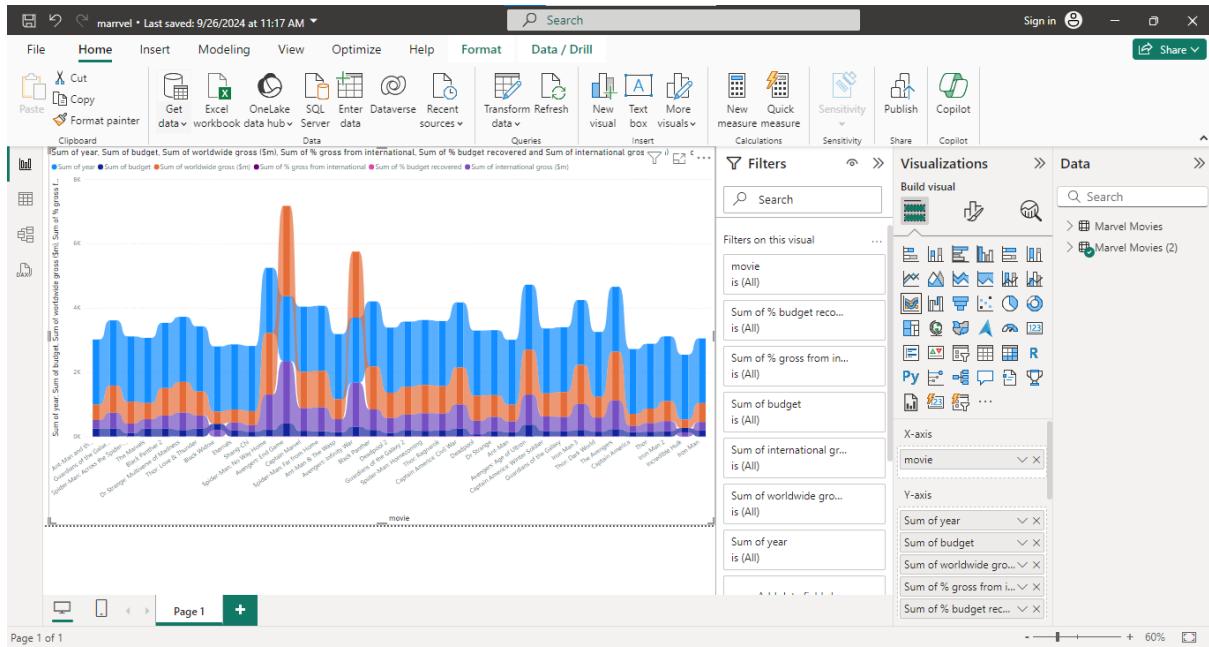


Here the attributes which has been used for the Stacked column chart are, age, alcohol consuming, chronic disease, date, gender and smoking.

ii) RIBBON CHART:

A **Ribbon Chart** in Power BI is a specialized chart that displays the rank of values across categories over time or another dimension. It's particularly useful for showing changes in rank or relative position over multiple categories, highlighting trends or fluctuations.

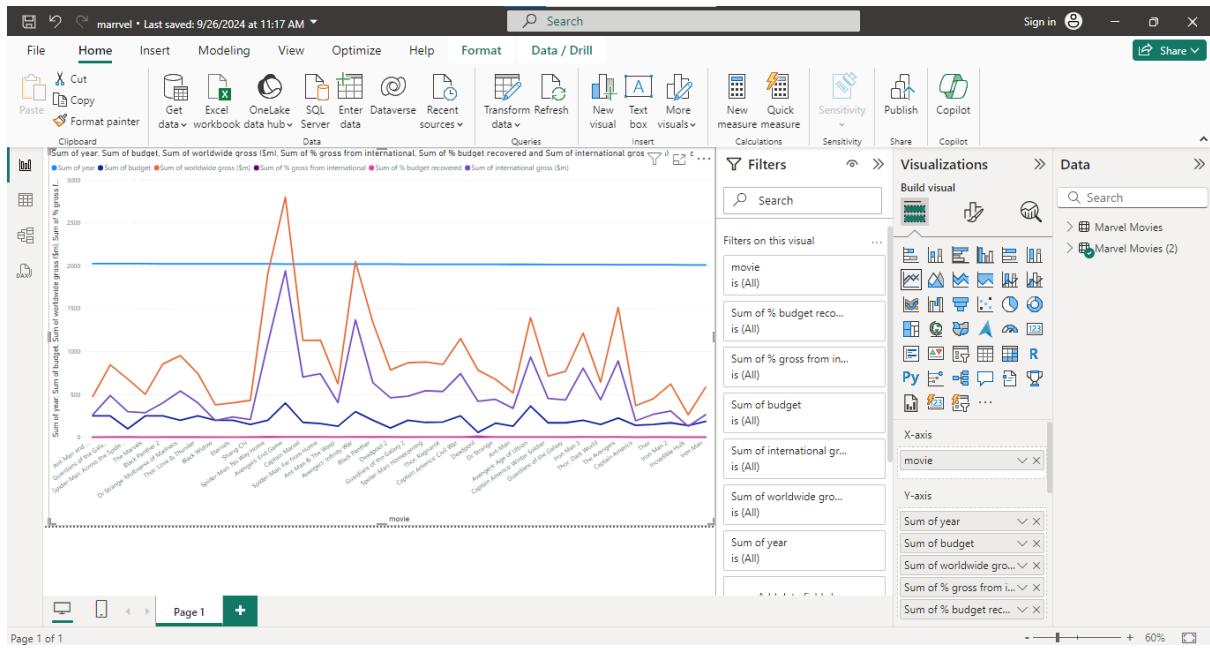
The chart is composed of ribbons that connect categories across the X-axis, with the width and position of each ribbon representing the value of that category at different points in time.



The above Ribbon chart visualization includes the attributes age, alcohol consuming, allergy, chest pain, date, gender, month, smoking and state.

iii) LINE CHART:

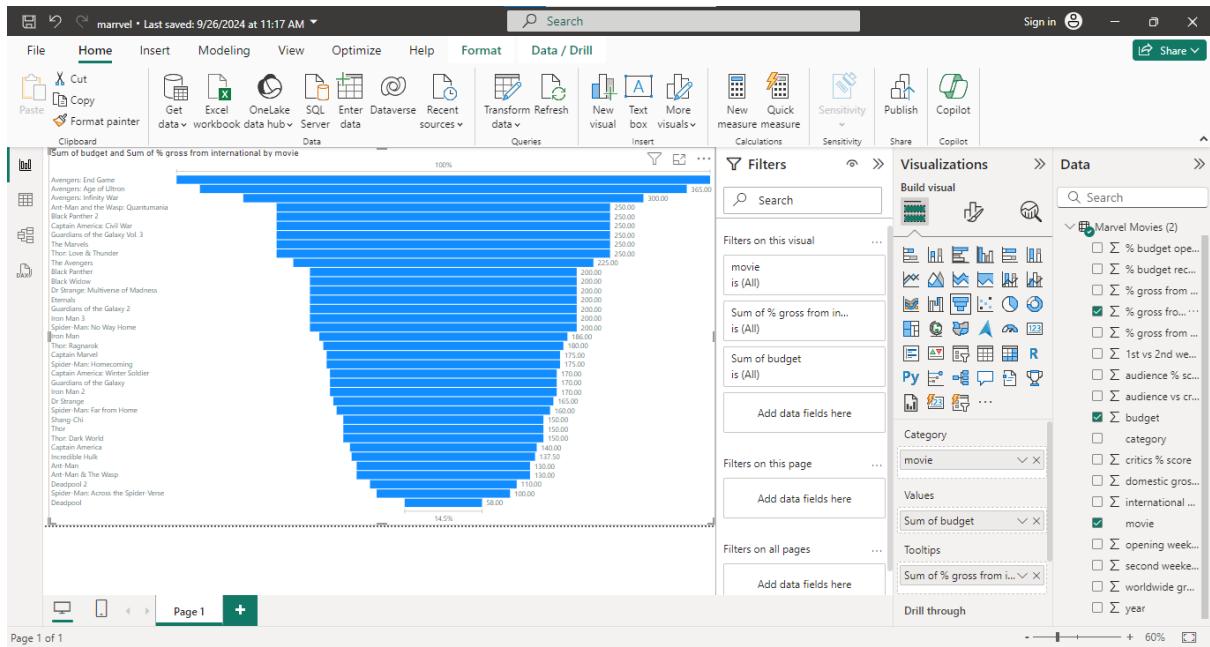
A **Line Chart** in Power BI is a common visualization used to display trends or changes over time. It's ideal for showing continuous data, such as monthly sales, stock prices, or any other time-series data. The chart connects individual data points with a line, making it easy to track changes and trends visually.



The above line chart includes attributes such as age, alcohol consuming, month, chest pain and smoking.

iv) FUNNEL:

A **Funnel Chart** in Power BI is used to visualize a linear process with sequential steps, typically representing stages in a business process, where values progressively decrease. It's ideal for visualizing data that represents stages in a funnel-like process (e.g., sales pipeline, website conversion, marketing campaigns). The funnel chart shows how data diminishes as it moves through stages, helping identify bottlenecks or drop-offs.

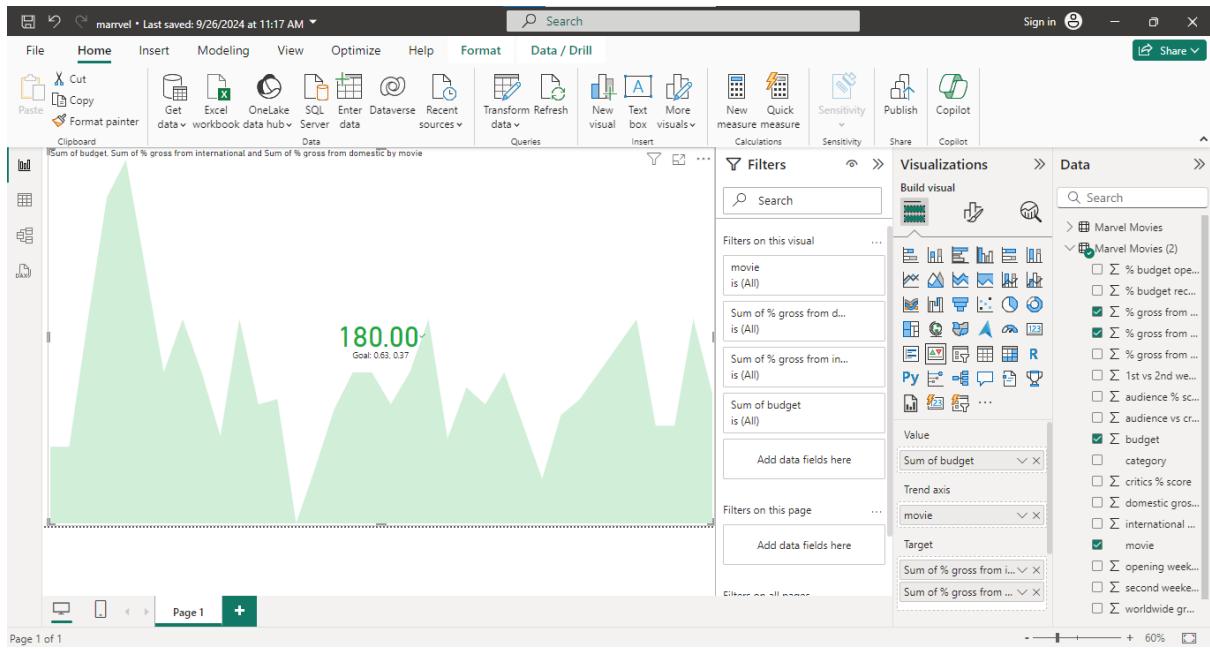


The attributes such as fatigue, date, age, anxiety and chronic disease has been used to visualize using the Funnel chart in POWER BI.

Here the Map is used to visualize the attributes such as State, Date, Year and time to show the states graphically.

v) KPI:

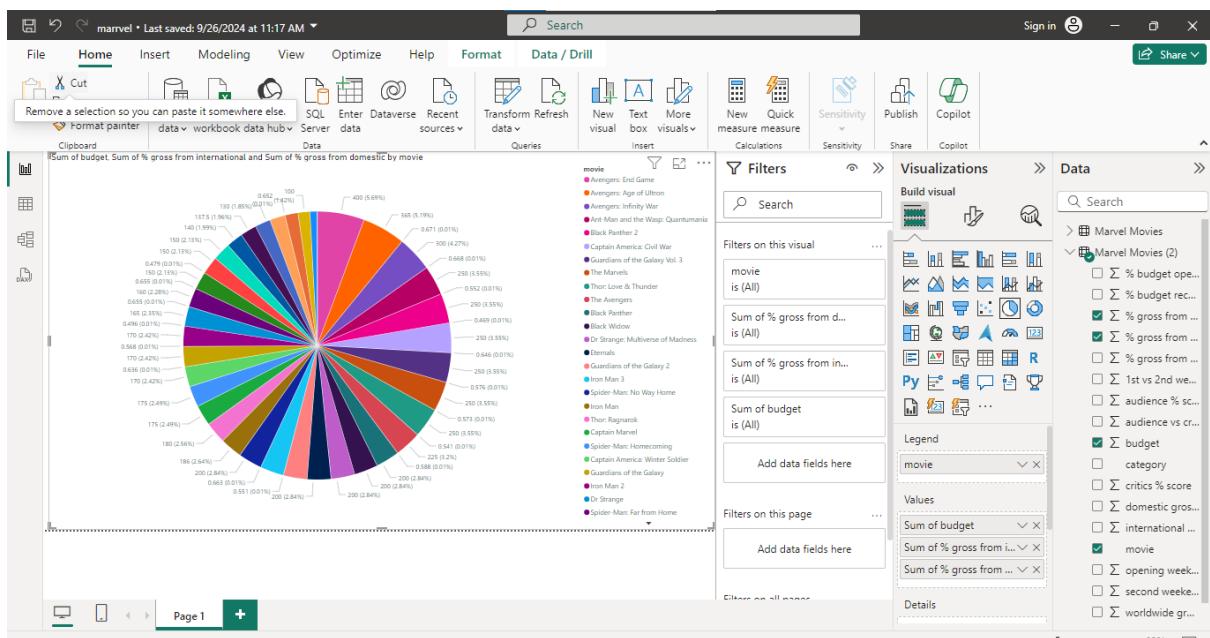
A **KPI (Key Performance Indicator)** in Power BI is a visualization that helps track the progress of a particular metric against a target. KPIs are commonly used to measure performance over time, such as sales goals, revenue targets, or customer satisfaction levels. The KPI visual provides a clear and concise view of whether performance is on track, ahead, or behind target, with visual cues like colors (green, red, or yellow) to indicate progress.



Here the attributes such as age, alcohol consuming, chronic diseases and date has been used to visualize the KPI.

vii) PIE CHART:

A **Pie Chart** in Power BI is a circular graph divided into slices to show the proportion or percentage of each category in a dataset. It is useful for displaying how individual parts contribute to a whole. Pie charts are most effective when comparing a small number of categories and when the differences between them are significant.



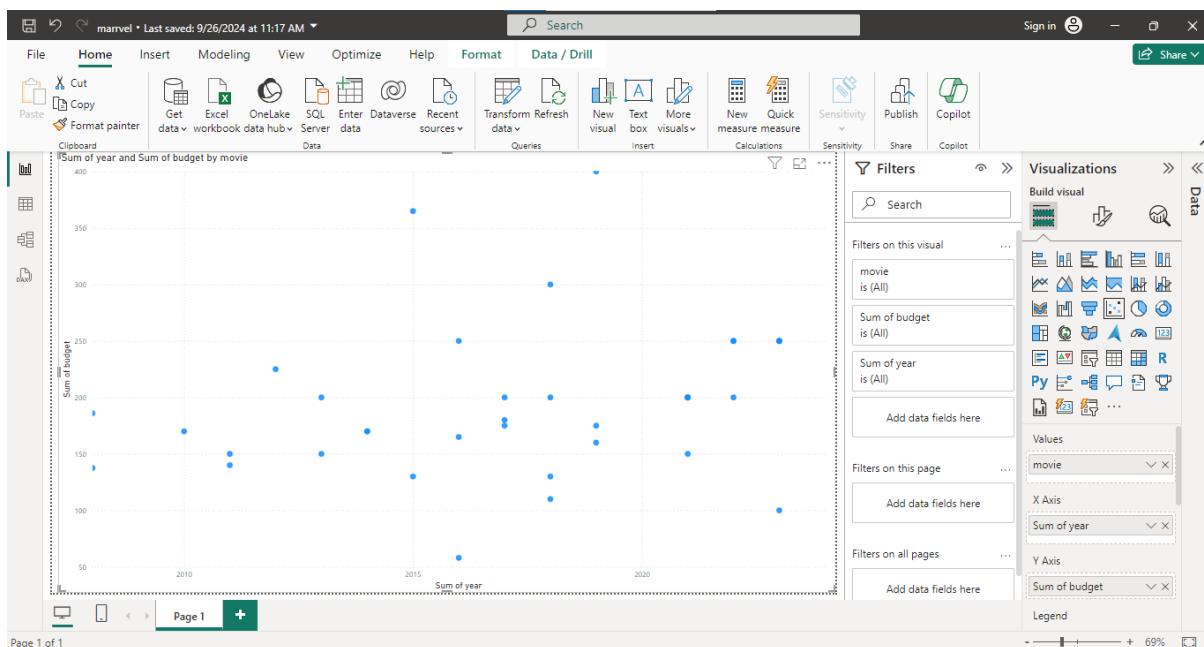
Here the attributes such as age, alcohol consuming, chest pain, coughing and smoking by gender has been visualized as PIE CHART with the help of Microsoft POWER BI.

viii) SCATTER CHART:

A **Scatter Chart** in Power BI is a type of visualization that displays relationships or correlations between two numerical values, typically represented on the X and Y axes. It's useful for exploring how different factors influence one another, spotting outliers, and identifying clusters or patterns in data. In Power BI, scatter charts can also be enhanced by adding a third variable using bubble size or color to provide more insights.

When to Use a Scatter Chart:

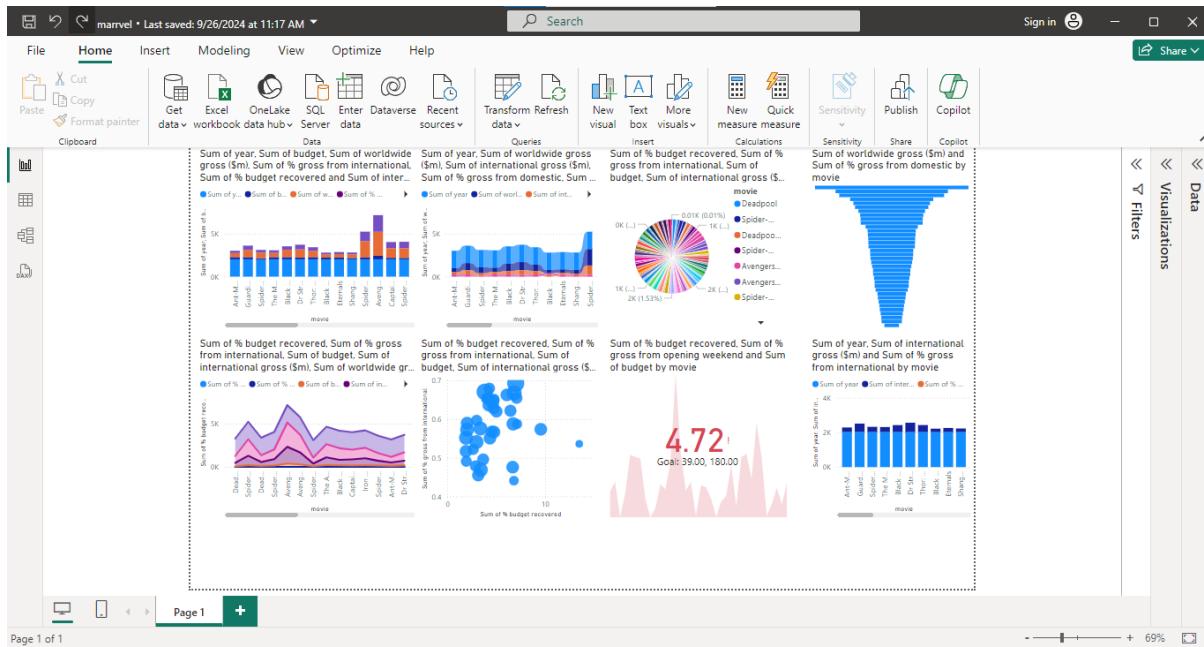
- To analyze the relationship between two numerical values.
- To spot trends, correlations, or clusters in the data.
- To detect outliers.
- To visualize changes over time or across categories with an additional dimension.



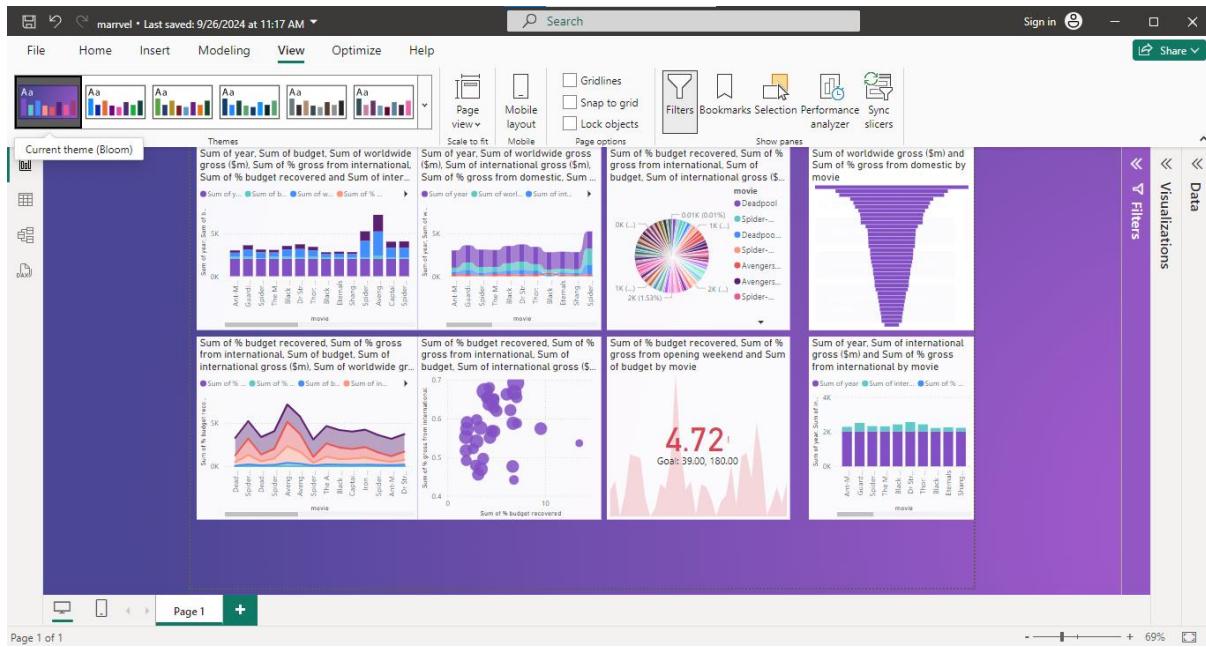
Here the attributes such as age, alcohol consuming, chest pain, gender, lung cancer, date and time has been used to visualize the data as the above viewed SCATTER CHART.

CHAPTER 9

9.1 A POWER BI DASHBOARD:



In the above images all the charts have been organized and title has been added to each graph for easy classification of the graphs.



Now with the help of view, the graph's visualization has been improved and the final output has been displayed.

9.2 CONCLUSION:

In conclusion, Power BI serves as a powerful business intelligence tool that enables organizations to transform raw data into meaningful insights through interactive and visually appealing dashboards. Its wide array of visualizations—such as bar charts, pie charts, KPIs, and scatter plots—allow for easy data analysis, helping businesses make data-driven decisions quickly and efficiently. The ability to integrate with various data sources, perform complex transformations, and create automated reports makes Power BI essential for organizations aiming to monitor performance, track KPIs, identify trends, and optimize operations.

Moreover, Power BI's intuitive drag-and-drop interface, coupled with advanced features like natural language queries, real-time data updates, and collaborative sharing, empowers users across all levels of expertise to explore data and gain actionable insights. As companies increasingly rely on data for competitive advantage, Power BI stands out as a versatile, scalable, and user-friendly platform for harnessing the power of analytics, making it a crucial tool in today's data-driven world.