Experiment – 1.1

Student Name: Prabhjot Kaur **UID: 24MCS10022**

Branch: ME-CSE **Section/Group: 24MCS-1/A**

Semester: 2nd **Date of Performance: 10/01/2025**

Subject: Business Analytics **Subject Code: 24CSH-661**

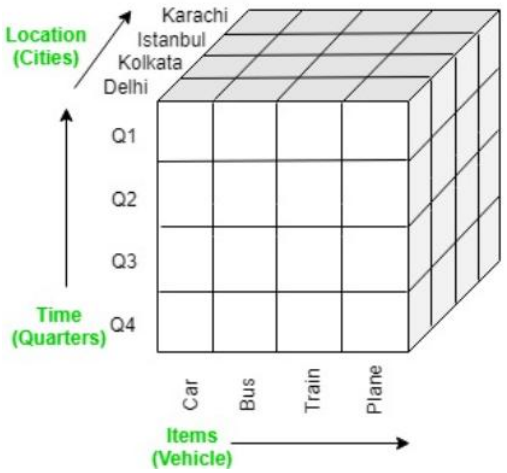
1. Aim : Revisiting basic data mining concepts like OLAP operations including slice, dice, roll-up, and drill down in detail.

2. Task to be done :

- Understand the fundamentals of data mining and its importance in data analysis.
- Explore OLAP operations like slice, dice, roll-up, and drill-down using practical examples.
- Perform hands-on exercises to apply these operations on a sample dataset.
- Analyze how OLAP operations help in summarizing and visualizing data effectively.

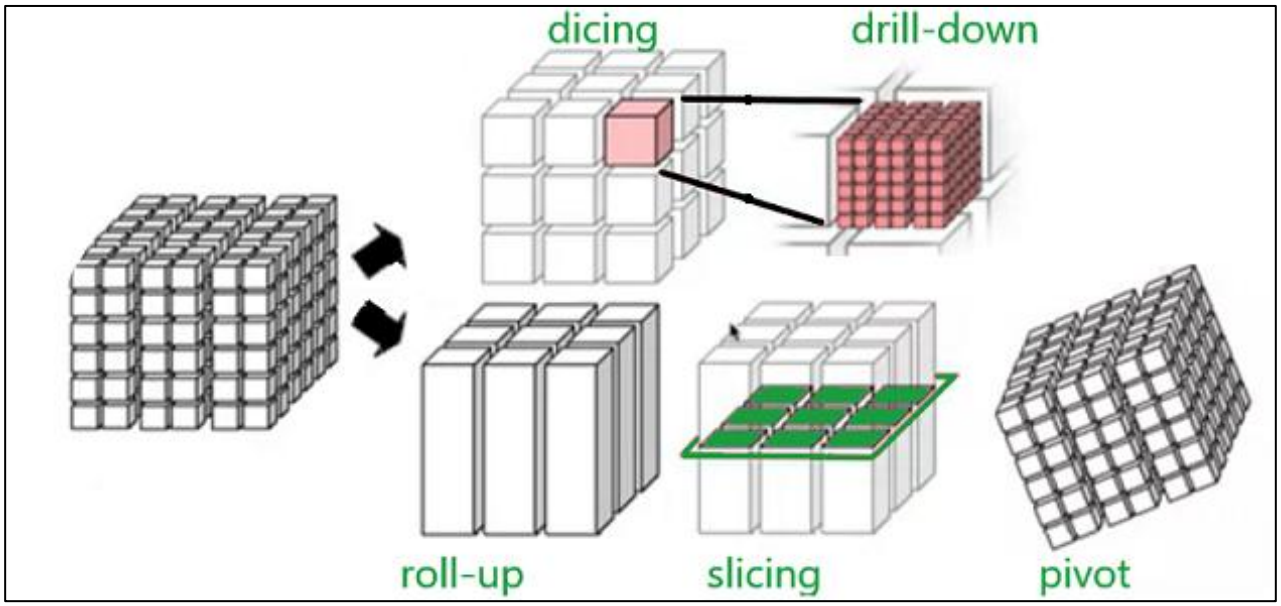
3. Description :

OLAP stands for ***Online Analytical Processing*** Server. It is a software technology that allows users to analyze information from multiple database systems at the same time. It is based on multidimensional data model and allows the user to query on multi-dimensional data (eg. Delhi -> 2018 -> Sales data). OLAP databases are divided into one or more cubes and these cubes are known as *Hyper-cubes*.



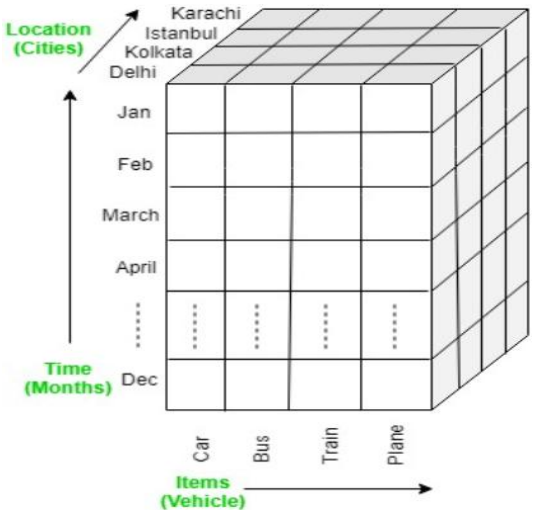
OLAP Operations

There are five basic analytical operations that can be performed on an OLAP cube:



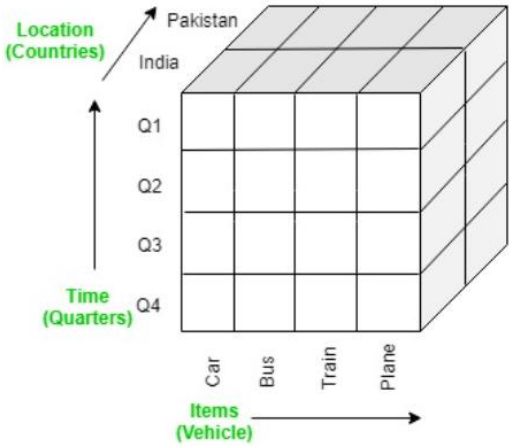
1. **Drill down:** In drill-down operation, the less detailed data is converted into highly detailed data. It can be done by:
   - Moving down in the concept hierarchy
   - Adding a new dimension

In the cube given in overview section, the drill down operation is performed by moving down in the concept hierarchy of *Time* dimension (Quarter -> Month).

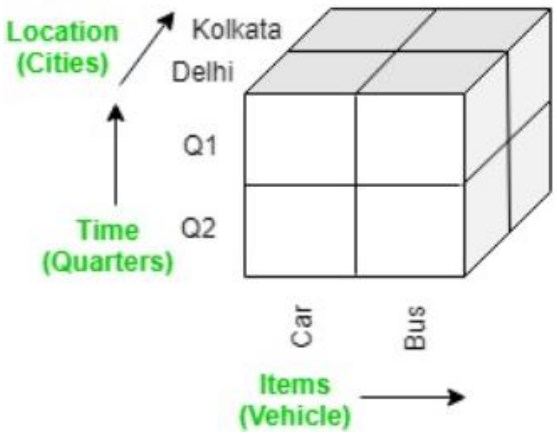


2. **Roll up:** It is just opposite of the drill-down operation. It performs aggregation on the OLAP cube. It can be done by:
   - Climbing up in the concept hierarchy
   - Reducing the dimensions

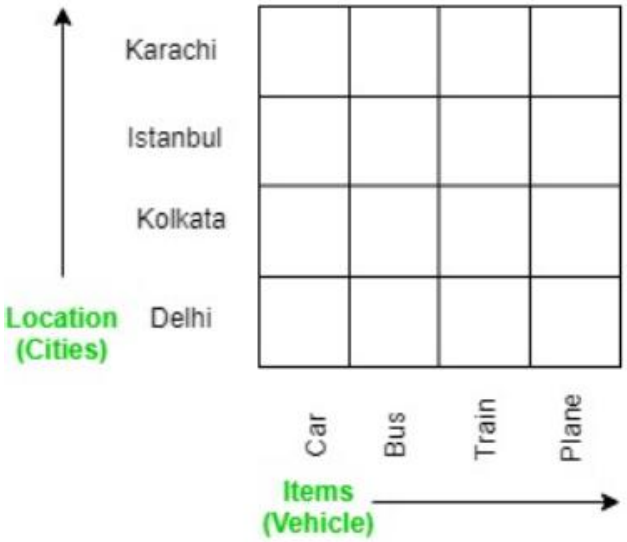
In the cube given in the overview section, the roll-up operation is performed by climbing up in the concept hierarchy of *Location* dimension (City -> Country).



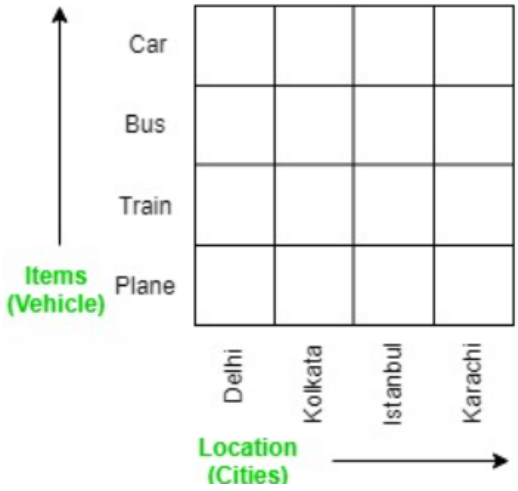
3. **Dice:** It selects a sub-cube from the OLAP cube by selecting two or more dimensions. In the cube given in the overview section, a sub-cube is selected by selecting following dimensions with criteria:
   - Location = “Delhi” or “Kolkata”
   - Time = “Q1” or “Q2”
   - Item = “Car” or “Bus”



4. **Slice:** It selects a single dimension from the OLAP cube which results in a new sub-cube creation. In the cube given in the overview section, Slice is performed on the dimension Time = "Q1".



5. **Pivot:** It is also known as *rotation* operation as it rotates the current view to get a new view of the representation. In the sub-cube obtained after the slice operation, performing pivot operation gives a new view of it.



4. Learning Outcomes :

- Gain a clear understanding of OLAP operations in data mining.
- Learn to manipulate data cubes for detailed and summarized analysis.
- Develop practical skills in implementing slice, dice, roll-up, and drill-down operations.
- Understand how these operations enhance decision-making through data exploration.

Experiment – 1.2

Student Name: Prabhjot Kaur **UID: 24MCS10022**

Branch: ME-CSE **Section/Group: 24MCS-1/A**

Semester: 2nd **Date of Performance: 17/01/2025**

Subject: Business Analytics **Subject Code: 24CSH-661**

1. Aim : Introduction of Power BI Desktop tool (or equivalent) installation with its features and strengths while visualizing data over multiple feature selections in OLAP operation.

2. Task to be done :

- Install Power BI Desktop from the Microsoft Store or Download Center.
- Explore key features like data import, visualizations, and automatic updates.
- Create visualizations such as bar charts, pie charts, and maps.
- Import and manipulate data from various sources for reports.

3. Description :

Introduction to Power BI Desktop

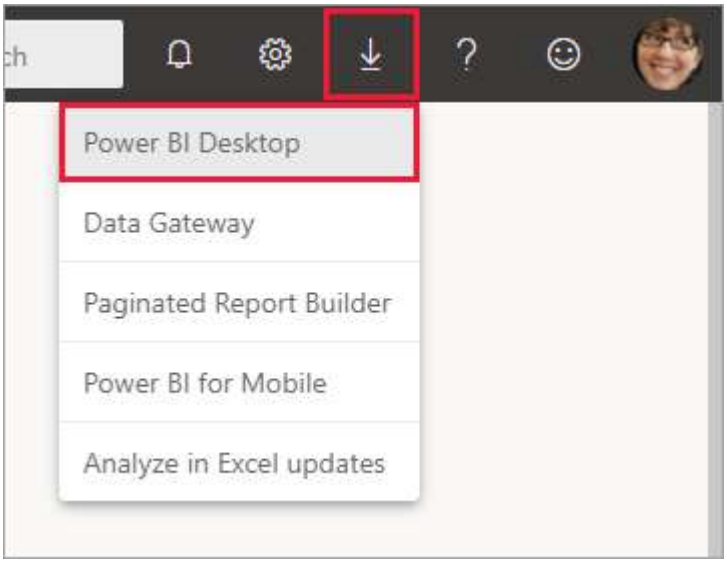
Power BI Desktop is a powerful business intelligence tool developed by Microsoft that allows users to connect to various data sources, transform data, and create interactive reports and dashboards. It provides a user-friendly interface for data visualization, analysis, and sharing insights across organizations. With built-in AI capabilities, robust data modeling tools, and seamless integration with cloud services, Power BI Desktop empowers users to make data-driven decisions efficiently. It is widely used for reporting, data analytics, and decision-making across industries.

Installation

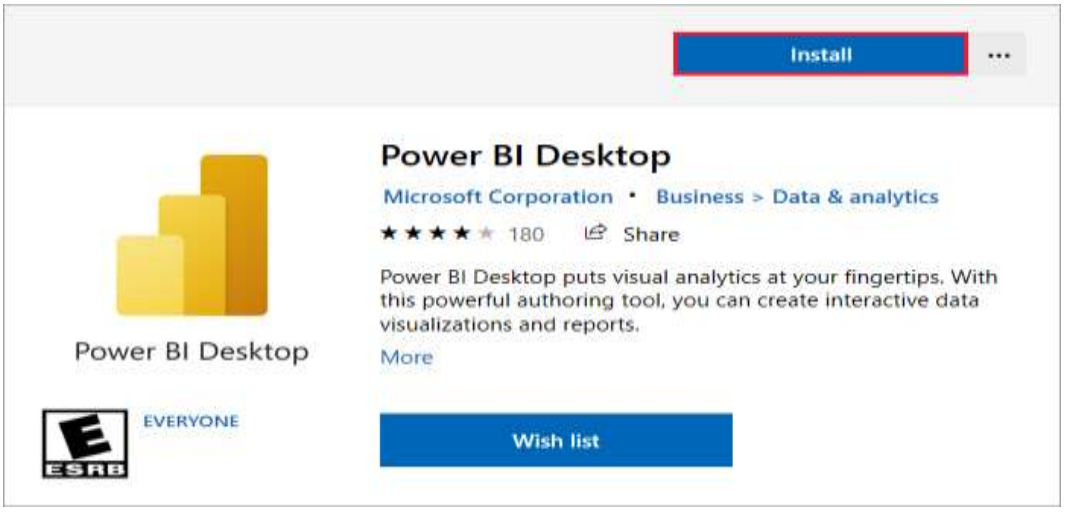
1. Use one of the following options to open the Power BI Desktop page of the Microsoft Store:

- Open a browser and go directly to the Power BI Desktop page of the Microsoft Store.
- From the Power BI service, in the upper right corner, select the Download icon and then choose Power BI Desktop.

- Go to the Power BI Desktop product page, and then select Download Free.



2. After you've landed on the Power BI Desktop page of the Microsoft Store, select Install.



Here are the advantages of getting **Power BI Desktop** from the **Microsoft Store**:

- **Automatic Updates**: The latest version is automatically downloaded and installed in the background, ensuring your software is always up to date.
- **Smaller Downloads**: Only the changed components are downloaded in each update, reducing the file size and saving bandwidth.
- **No Admin Privileges Required**: Unlike the direct download version, installing Power BI Desktop from the Microsoft Store does not require administrative rights.

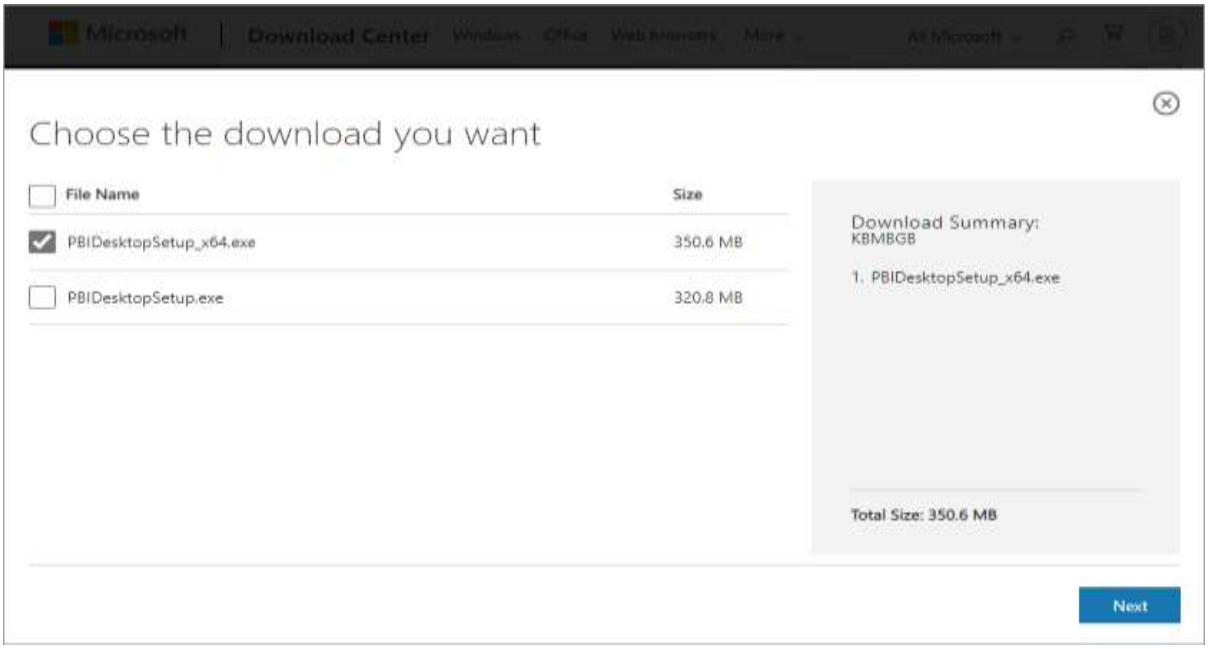
- **Easier IT Deployment**: Organizations can efficiently roll out Power BI Desktop across multiple users through the **Microsoft Store for Business**.
- **Built-in Language Detection**: The Microsoft Store version includes all supported languages and automatically matches the language settings of your system.

Here are the **considerations and limitations** when installing **Power BI Desktop** from the **Microsoft Store**:

- **SAP Connector Files**: If you use the SAP connector, you may need to manually move the SAP driver files to the Windows\System32 folder for proper functionality.
- **User Settings Not Transferred**: If switching from the .exe version to the Microsoft Store version, user settings, saved credentials, and connections are not automatically copied over. You will need to reconnect to data sources manually.
- **IT Control Restrictions**: Some organizations may have policies that restrict installations from the Microsoft Store, limiting access to Power BI Desktop.

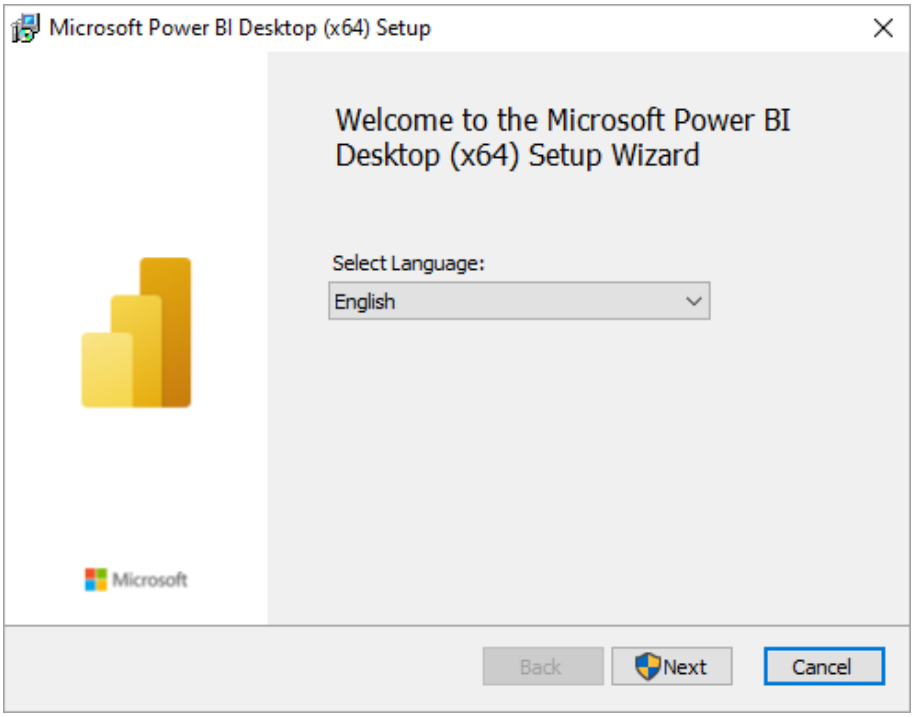
Download Power BI Desktop directly

To download the Power BI Desktop executable from the Download Center, select **Download** from the Download Center page. Then specify the 32-bit or 64-bit installation file to download.



Install Power BI Desktop after download

You're prompted to run the installation file after you finish downloading it. Power BI Desktop ships as a single *.exe* installation package that contains all supported languages, with separate *.exe* files for the 32-bit and 64-bit versions. The *.msi* packages are no longer available. You need the executable for installation. This approach makes distribution, updates, and installation much easier and more convenient, especially for administrators. You can also use command-line parameters to customize the installation process, as described in Using command-line options during installation.After you launch the installation package, Power BI Desktop installs as an application and runs on your desktop.



When you launch Power BI Desktop, a welcome screen appears

When you launch Power BI Desktop for the first time, if the installation isn't an upgrade, you're prompted to fill out a form or sign in to the Power BI service before you can continue.

After that, you can begin creating data models or reports, and share them with others on the Power BI service. Check out the Next steps section for links to guides to help you get started using Power BI Desktop.

Minimum System Requirements

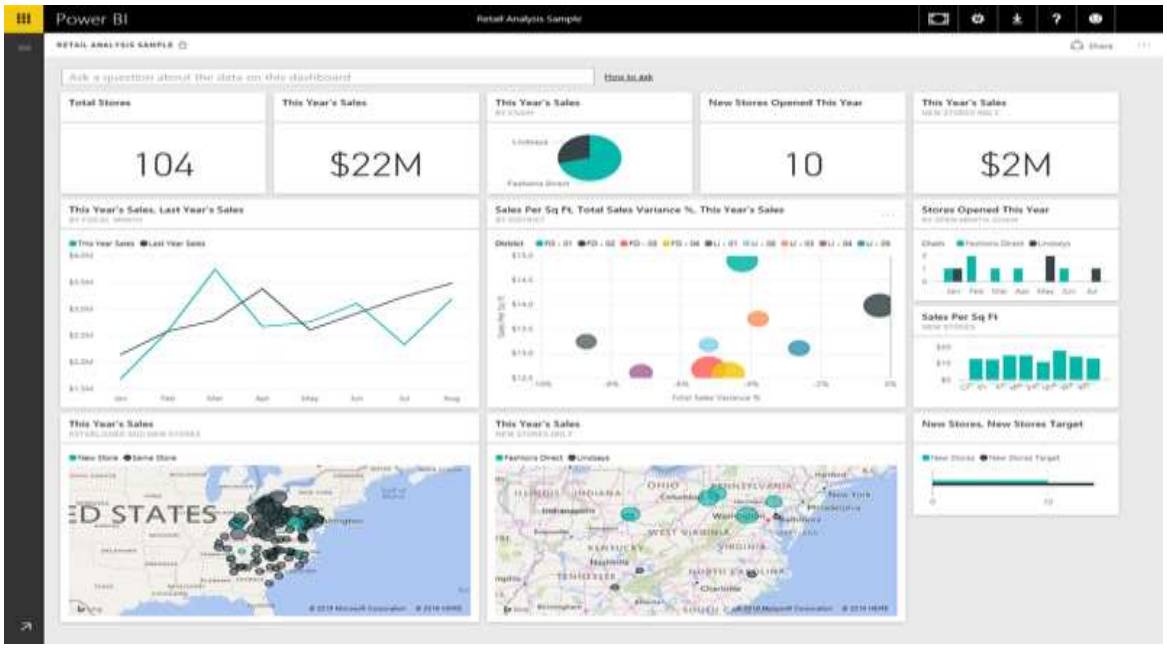
- Power BI Desktop is no longer supported on Windows 7.
- Windows 8.1 or Windows Server 2012 R2 or later.
- .NET 4.6.2 or later.
- Internet Explorer 11 or later.
- **Memory (RAM)**: At least 2 GB available, 4 GB or more recommended.
- **Display**: At least 1440x900 or 1600x900 (16:9) required. Lower resolutions such as 1024x768 or 1280x800 aren't supported because some controls (such as closing the startup screens) display beyond those resolutions.
- **Windows Display Settings**: If you set your display to change the size of text, apps, and other items to more than 100%, you won't see some dialogs that you must interact with to continue using Power BI Desktop. If you encounter this issue, check your display settings in Windows by going to **Settings > System > Display**, and use the slider to return display settings to 100%.
- **CPU**: 1 gigahertz (GHz) 64-bit (x64) processor or better recommended.

Features of Power BI

1. Range of Visualizations Power BI allows users to create detailed and visually appealing reports and dashboards. A variety of visualization options are available, including:

- Stacked bar chart
- Stacked column chart
- Clustered bar chart
- Clustered column chart

- 100% stacked column/bar chart
- Line chart
- Area chart
- Stacked area chart
- Ribbon chart
- Waterfall chart
- Scatter chart
- Pie chart
- Donut chart
- Treemap chart
- Map
- Filled map
- Funnel chart
- Gauge chart



2. Data Source Connectivity Power BI offers the **Get Data** feature, allowing users to import data from a range of sources, including:

- On-premise and cloud-based sources
- Structured and unstructured data sources
- Regular updates adding new data source compatibility

4. Learning Outcomes :

- Understand the steps to install Power BI Desktop from the Microsoft Store.
- Identify the advantages of using the Microsoft Store version over a direct download.
- Recognize the limitations and considerations when installing Power BI Desktop.
- Learn the minimum system requirements needed to run Power BI Desktop effectively.

Experiment – 1.3

Student Name: Prabhjot Kaur **UID: 24MCS10022**

Branch: ME-CSE **Section/Group: 24MCS-1/A**

Semester: 2nd **Date of Performance: 25/01/2025**

Subject: Business Analytics **Subject Code: 24CSH-661**

1. Aim : Introduction to UCI repository and select and download the data from the UCI repository (more than 1mb) of your interest for the analysis.

2. Task to be done :

- Select and download a dataset from the UCI Machine Learning Repository.
- Load the dataset into Google Colab or Jupyter Notebook using pandas.
- Clean and name the dataset columns based on attribute information.
- Perform basic statistical analysis and identify patterns in the data.

3. Description :

The UCI Machine Learning Repository provides numerous datasets for machine learning, covering a wide range of domains like healthcare, finance, biology, and more. These datasets can help in developing, testing, and validating machine learning models. In this experiment, we will explore how to work with .data files, which are commonly used in the repository but not as straightforward to handle as .csv files.

We will demonstrate how to:

- Download datasets from the UCI Repository,
- Convert .data files into a structured Pandas DataFrame,
- Assign proper column names,
- Handle missing values, and
- Perform basic exploratory data analysis (EDA).

The goal is to prepare the dataset for further machine learning applications.

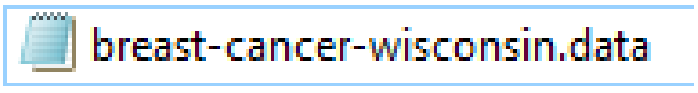
4. Procedure:

1. **Where to Find the Data**:

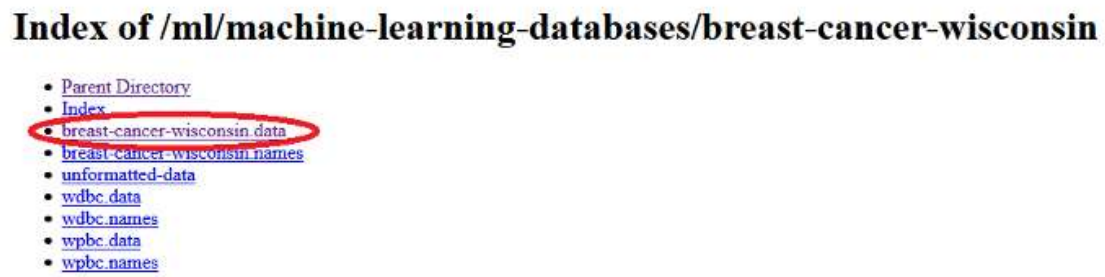
   The UCI Machine Learning Repository provides a wide variety of datasets. One popular dataset is the "Breast Cancer Wisconsin (Original) Data Set." Another resource for datasets is Kaggle, which primarily offers .csv files that are easily handled with Pandas. For this experiment, we will focus on .data files from the UCI Repository.

2. **Selecting and Downloading the Dataset**:

   - Go to the UCI Machine Learning Repository.
   - Browse through the datasets and select one that interests you and is larger than 1MB in size. For example, the Breast Cancer dataset is a good choice.

   - Once on the dataset page, click the "Data Folder" link. This will take you to a second page where you can download the dataset files.
   - Download the .data file, which can be opened with Notepad or Microsoft Excel.



3. **Loading the Dataset into Python**:

   To handle .data files, we can use the Pandas library in Python. The .data file does not have column names, so we will need to assign them manually. The steps are as follows:

   - Import Pandas and load the dataset using pd.read_csv() method.
   - If the dataset doesn't have column headers, we'll assign them ourselves.

4. **Preprocessing the Data**: Once the dataset is loaded into a DataFrame, we can clean and preprocess the data:

   - **Assign proper column names**: If the dataset lacks headers, you need to assign the column names manually.
   - **Handle missing values**: Missing values are common in real-world datasets. We can fill missing values with a specified value or drop rows with missing data.

5. **Exploratory Data Analysis (EDA)**:

   Now that the data is cleaned, perform basic exploratory data analysis:

   - **Shape of the dataset**: Check the number of rows and columns.
   - **Basic statistics**: Get a summary of numerical columns (mean, std, min, max, etc.).

5. Learning Outcomes :

- Learn to navigate the UCI Machine Learning Repository and select appropriate datasets.
- Convert .data files into structured DataFrames and preprocess them.
- Handle missing values and add column names during data preprocessing.
- Conduct basic EDA with Pandas to prepare datasets for machine learning tasks.

Experiment – 2.1

Student Name: Prabhjot Kaur **UID: 24MCS10022**

Branch: ME-CSE **Section/Group: 24MCS-1/A**

Semester: 2nd **Date of Performance: /02/2025**

Subject: Business Analytics **Subject Code: 24CSH-661**

1. Aim :
Compile and preprocess the datasheet for missing value and other minor imputation and prepare the dashboard with at least 7 major features in power bi desktop.

2. Task to be done :

- Import and preprocess the dataset by handling missing values and inconsistencies.
- Transform and model the data for structured analysis in Power BI.
- Design a dashboard with at least seven interactive features.
- Optimize and publish the report for insightful data visualization.

3. Description :

To ensure the dataset is clean and ready for analysis, we follow fundamental preprocessing steps, including handling missing values and performing necessary imputations.

The cleaned data is then used to develop an interactive dashboard in Power BI Desktop with at least seven major features. The key fundamentals covered in this process are:

1. **Data Collection and Importing**: The dataset is loaded into Power BI Desktop from sources like Excel, CSV, or databases.

2. **Data Cleaning and Preprocessing**: Missing values are handled through deletion, mean/mode/median imputation, or predictive techniques. Duplicate records are removed, and inconsistent formats are corrected.

3. **Data Transformation**: Features such as data type conversion, column splitting, merging, and calculated fields are applied to enhance data usability.

4. **Data Modeling**: Relationships between tables are established using primary and foreign keys, and measures are created for effective analysis.

5. **Visualization Development**: A Power BI dashboard is designed with at least seven key features, including charts, tables, KPIs, slicers, and maps for interactive insights.

6. **DAX (Data Analysis Expressions) Implementation**: DAX functions are used to create custom measures and calculated columns for in-depth data analysis.

7. **Interactivity and Filtering**: Features like drill-through, filters, and slicers allow dynamic data exploration.

8. **Dashboard Optimization**: Performance tuning techniques such as aggregations, indexing, and query reduction ensure efficient data loading and visualization rendering.

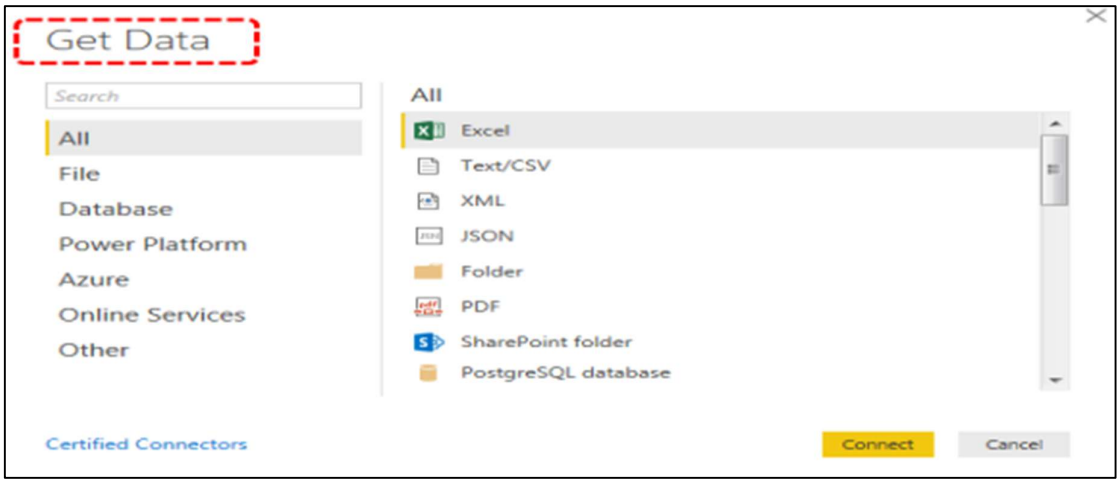
9. **Report Publishing and Sharing**: The finalized Power BI dashboard is published and shared with stakeholders via Power BI Service or local export options.

This approach ensures a structured, insightful, and well-optimized dashboard that enhances decision-making based on clean and processed data.

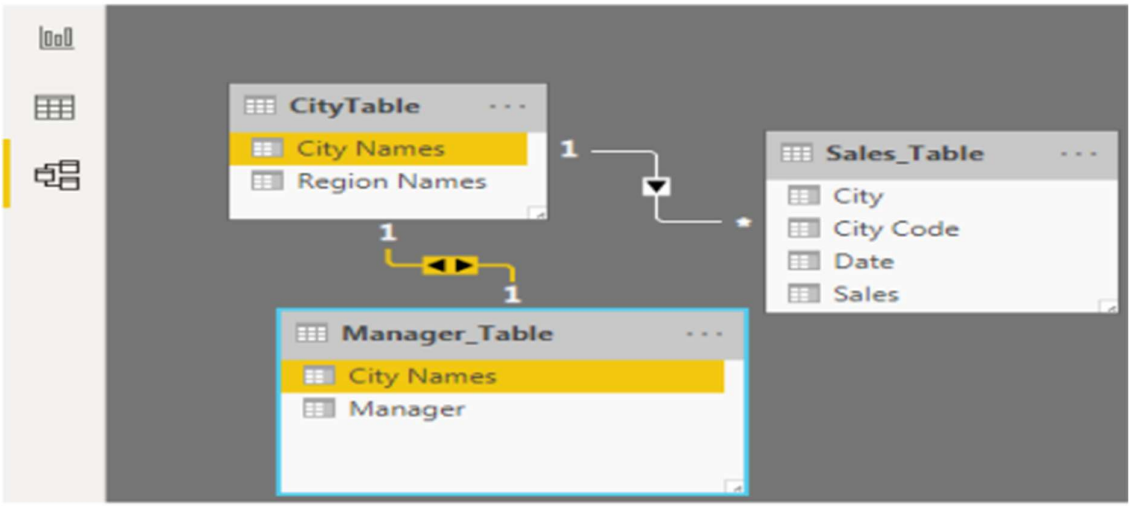
4. Procedure:

Features of Microsoft Power BI

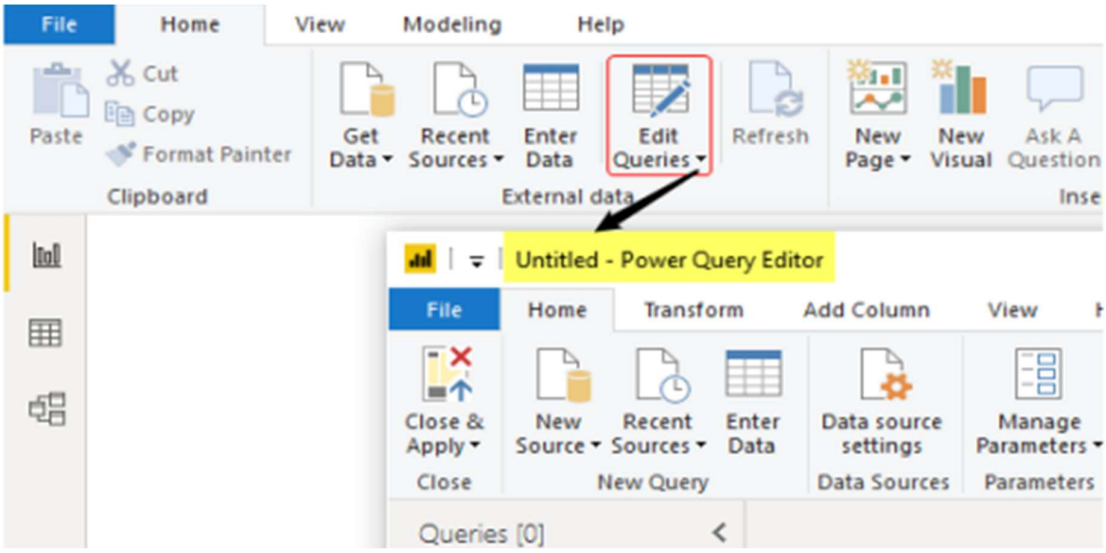
1. Data Connection : As Excel users, we usually get the data in an “Excel Workbook” or a Text file or CSV file. But this is not the case with Power BI because it can get data almost everywhere. Microsoft has developed data source connections with a huge list of sources like SQL, Azure, Excel, Text, CSV, PDF, Cloud, and on premises data. It does not matter where the data is and what format it is; we will get a detailed view of it.



2. Data Relationships: In Power BI, if you have data in multiple columns, then we can define the relationship between those tables based on at least one matching column from either table. Below is the screenshot of creating relationships between tables. We call this process "Data Modelling" in Power BI.



3. Power Query, Power Pivot Access: We can edit the data in Power BI using "Power Query in Excel" and "Power Pivot." These are all different components of Power BI, which helps the user alter the data to fit their needs. Power BI Query is used for data transformation and manipulation tools, and Power Pivot is a memory tool to model the data.

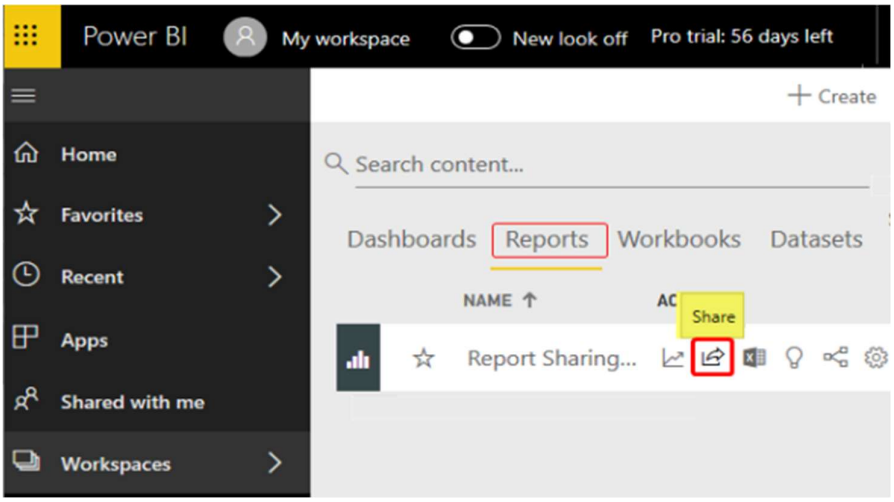


4. Cost of Power BI: We would consider Power BI cost a key feature because when you compare Power BI with other data visualization tools like Tableau, QlikView, and other products cost much less than that. And one more thing about Power BI is its free version of Power BI Desktop comes free of cost with enough features for the starter to get started with data visualization.

5. Custom Visualizations: Power BI has many built-in visuals to build dashboards and reports. Apart from these built-in visuals, users can download the custom visuals from marketplaces per their requirements.

6. Power BI Question and Answers: Power BI can answer your queries regarding the uploaded data. For example, after the data is uploaded to Power BI, it would help you with the information required, like “What are the total sales for the year 2018?” you have to type the question. Then, it would show the result with a beautiful visual or chart

7. Report Sharing: When you share a report in Excel, you share it with the data sets, but in Power BI, using Power BI Pro services, we can share the report with team members without worrying about data security.



8. Mobile App: The Power BI app is available to download. It makes the report reaching the end-users much easier and simpler. Once the report is shared with the end-user, they must sit in front of the laptop or system. Rather, they can open the Power BI app on their mobile and start reading the report instantly.

5. Learning Outcomes :

- Understand data preprocessing techniques for effective analysis.
- Develop interactive dashboards using Power BI features.
- Apply DAX functions for custom calculations and insights.
- Enhance decision-making through data visualization and reporting.

Experiment – 2.2

Student Name: Prabhjot Kaur **UID: 24MCS10022**

Branch: ME-CSE **Section/Group: 24MCS-1/A**

Semester: 2nd **Date of Performance: /02/2025**

Subject: Business Analytics **Subject Code: 24CSH-661**

1. Aim : Produce the graph and visualize the selected parameters on the basis of their high relevance while extracting useful intelligence.

2. Task to be done :

- Preprocess the dataset by handling missing values, removing duplicates, and applying transformations.
- Create parameters in Power BI to allow dynamic selection of relevant data fields.
- Generate various visualizations such as heatmaps, histograms, bar charts, and scatter plots to analyze key patterns.
- Implement interactive elements like filters, slicers, and drill-through analysis for deeper data exploration.

3. Description :

Data visualization plays a crucial role in extracting meaningful insights from raw datasets. In this practical, we aim to generate graphs and visualize key parameters based on their high relevance to derive useful intelligence. Using Power BI Desktop, we will preprocess the dataset, clean missing values, and apply transformations to ensure data quality.

Terminologies:

1. **Graph** – A visual representation of data that helps in identifying trends, patterns, and relationships between variables.
2. **Nodes (Data Points)** – Individual elements or data entities represented in a graph, such as customers, products, or transactions.

3. **Edges (Connections)** – The links between nodes, showing relationships or dependencies between different data points.
4. **Drill-Through Analysis** – A technique that allows users to explore detailed insights by clicking on summarized data points.
5. **Heatmap** – A graphical representation of data where values are depicted using color intensity, helping in identifying patterns and correlations.
6. **Scatter Plot** – A type of visualization used to display relationships between two numerical variables and detect correlations.
7. **Hierarchical Relationships** – A structured way of representing data in a parent-child format, commonly used in organizational charts and product categories.
8. **Data Aggregation** – The process of summarizing large datasets by computing values such as sum, average, count, or percentage.
9. **Trend Analysis** – A technique used to identify patterns and movements over time, often visualized using line charts.
10. **Outlier Detection** – Identifying data points that deviate significantly from other observations, helping in fraud detection and anomaly identification.

4. Procedure:

Create the parameters:

These steps use the Superstore sample to create new parameters.

1. In the Data pane, click the drop-down arrow in the upper right corner and select Create Parameter.
2. In the Create Parameter dialog box, complete the following steps:

    a. Name the parameter so that viewers can tell what changing it will do. This example uses Select Column 1 Heading.

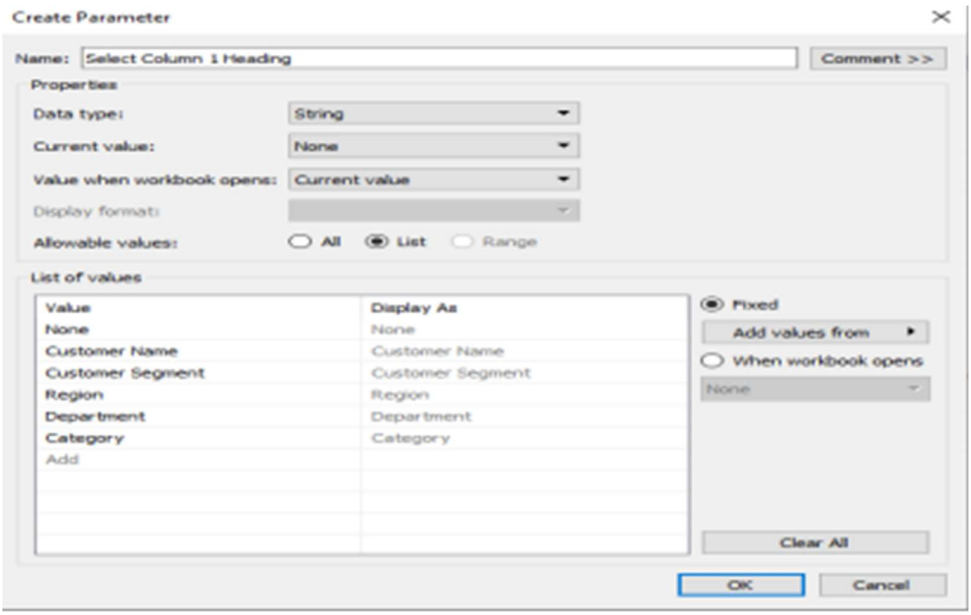
    b. For Data type, select String.

    c. For Allowable Values, select List, type None as the first value in the list, and then press Enter.

    d. Complete the list by typing the names of the additional dimension fields that you want to expose

    through the parameter.

The **Display As** aliases default to the field name, and for this exercise you can leave them as they are.



e. Click OK to return to the Calculated Field dialog box.

3. Repeat the previous step to create the following additional parameters:

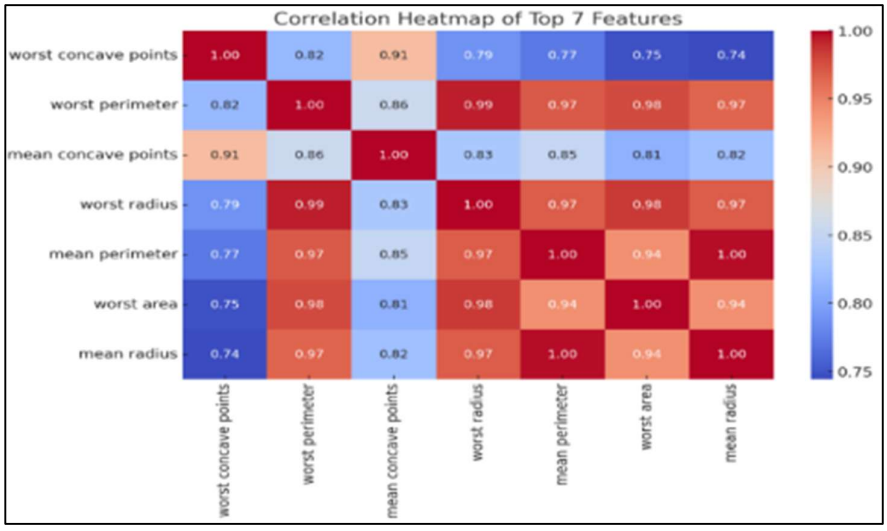
a. Select Column 2 Heading

b. Select Row 1 Heading

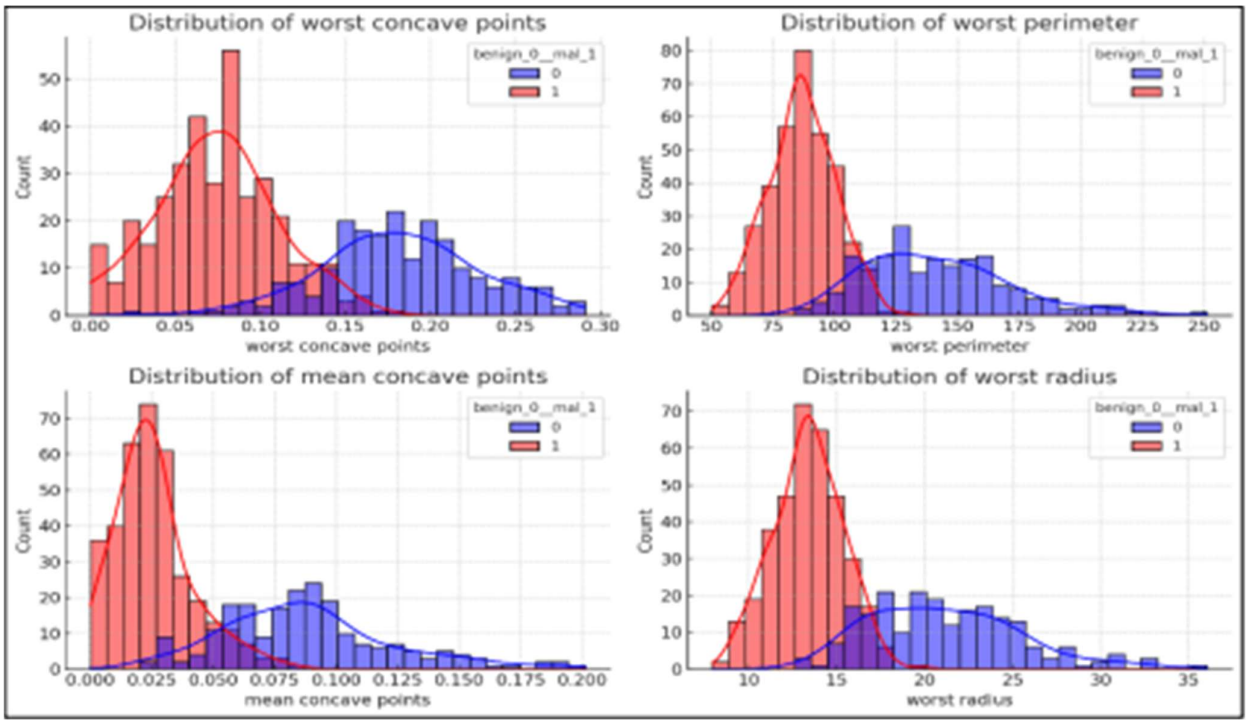
c. Select Row 2 Heading

Example: Cancer Classification Dataset

The dataset contains 569 entries with 31 columns, including features related to cancer classification. There are no missing values. The last column, benign_0__mal_1, is the target variable indicating whether the tumor is benign (0) or malignant (1).



The heatmap shows the correlation among the top 7 most relevant features with the target variable. These features have strong relationships with cancer classification, which can be useful for predictive modeling. Next, will visualize the distribution of these features based on the benign and malignant classes.



The histograms show the distribution of the top four most relevant features for benign (blue) and malignant (red) cases. These visualizations help in understanding how each feature contributes to cancer classification.

5. Learning Outcomes :

- Understand data preprocessing techniques for cleaning and structuring datasets.
- Gain expertise in creating dynamic parameters and relationships in Power BI.
- Develop skills in visualizing correlations and trends to extract meaningful insights.
- Enhance analytical capabilities using interactive dashboards for data-driven decision-making.

Experiment – 2.3

Student Name: Prabhjot Kaur **UID: 24MCS10022**

Branch: ME-CSE **Section/Group: 24MCS-1/A**

Semester: 2nd **Date of Performance: /0 /2025**

Subject: Business Analytics **Subject Code: 24CSH-661**

1. Aim : Produce the graph and visualize the selected parameters on the basis of their high relevance while extracting useful intelligence.

2. Task to be done:

- Install a LaTeX editor like TeXMaker or TeXStudio.
- Create a basic LaTeX document with title, sections, and paragraphs.
- Learn to write mathematical equations and insert images/tables.
- Generate a PDF output from the .tex file.

3. Description :

LaTeX is a high-quality typesetting system widely used for producing scientific and technical documents. Unlike WYSIWYG editors, LaTeX uses markup commands to format documents, allowing for precise control over structure and layout. In this activity, students will install a LaTeX editor such as TeXMaker or TeXStudio, which provides an interface to write, preview, and compile LaTeX code. They will learn how to organize reports using sections, format text, create lists, insert figures, and write complex equations—an essential skill for research, academic writing, and engineering documentation.

Key Features of LaTeX

- Professional typesetting for journal articles, books, reports, and slides.
- Handles complex mathematical equations with precision.
- Excellent for large structured documents with cross-references, sections, tables, and figures.
- Built-in support for bibliography generation, multilingual content, and indexing.
- Compatible with PostScript, PDF, and vector graphics inclusion.

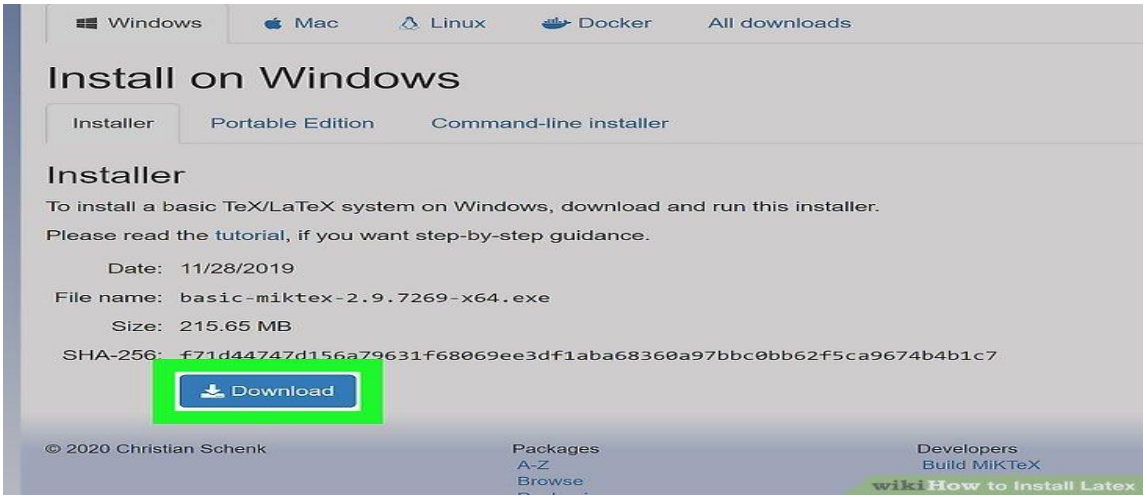
4. Procedure :

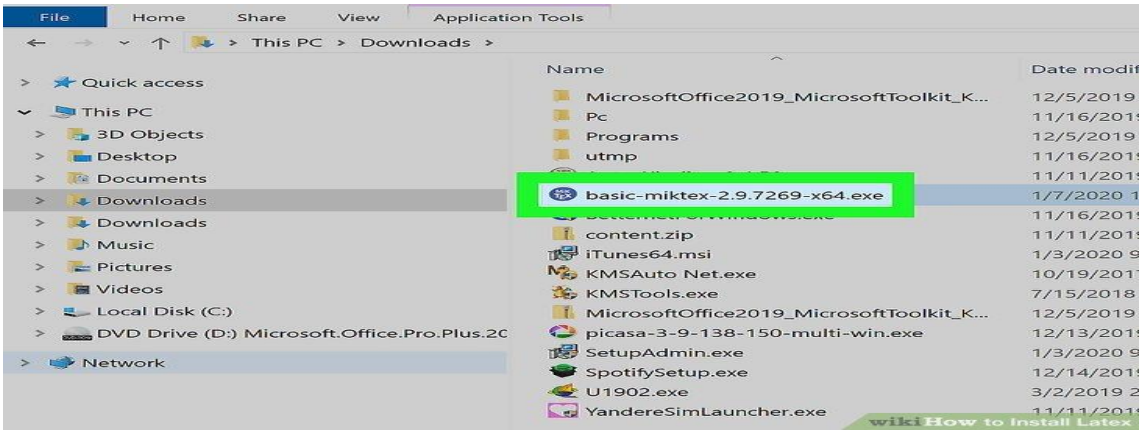
Step-by-Step Installation Process (Windows using MiKTeX)

1. **Open the LaTeX Website:**
   Visit https://www.latex-project.org in your web browser.

2. **Navigate to the Download Page:**
   Click on the Get tab at the top and select the MiKTeX link under the Windows section. This redirects to the official MiKTeX page.

3. **Download the Installer:**
   Click on the **Downloads** tab and then hit the **blue Download button** to begin downloading the basic-miktex.exe installer.

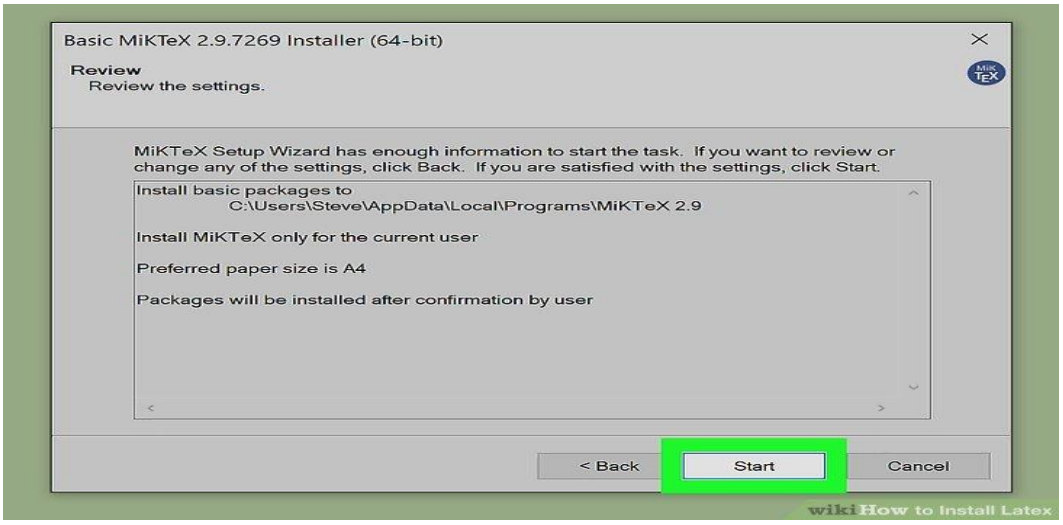




4. **Run the Installer:**

   Open the downloaded file. In the installation wizard:

   - Accept the license agreement.
   - Choose installation scope (Current User or All Users).
   - Select the installation path if needed.
   - Set the preferred paper type (A4/Letter).
   - Click **Start** to begin installation.



5. Learning Outcomes :

- Understand the LaTeX environment and how it differs from word processors.
- Structure academic/technical documents professionally.
- Write and format mathematical content using LaTeX syntax.
- Compile LaTeX code into a polished PDF report.

Experiment – 2.4

Student Name: Prabhjot Kaur **UID: 24MCS10022**

Branch: ME-CSE **Section/Group: 24MCS-1/A**

Semester: 2nd **Date of Performance: /0 /2025**

Subject: Business Analytics **Subject Code: 24CSH-661**

1. Aim : Install and learn Weka Tool for data mining while exploring its features.

2. Task to be done:

- Install the Weka tool on a Windows/Linux/Mac system.
- Open Weka GUI and explore different panels (Explorer, Experimenter, etc.).
- Load a sample dataset and apply classification and clustering.
- Analyze and interpret the output and visual results.

3. Description :

Weka (Waikato Environment for Knowledge Analysis) is a free and open-source software tool developed by the University of Waikato for performing data mining and machine learning tasks. It provides a simple graphical interface that allows users to analyze data, build machine learning models, and visualize results without writing code.

Weka supports a wide variety of data mining techniques, including:

- Classification (e.g., J48 decision trees, Naive Bayes)
- Clustering (e.g., K-means)
- Association rule mining (e.g., Apriori)
- Data preprocessing and filtering
- Model evaluation and visualization

Key features of Weka:

- Easy-to-use GUI for beginners
- Accepts .arff, .csv, and .xrff formats
- Offers multiple evaluation methods like cross-validation and confusion matrices
- Built-in datasets for quick experimentation

By exploring Weka's core functionalities, students will develop a strong foundation in data analysis and A model interpretation, making it easier to transition into advanced tools or real-world applications.

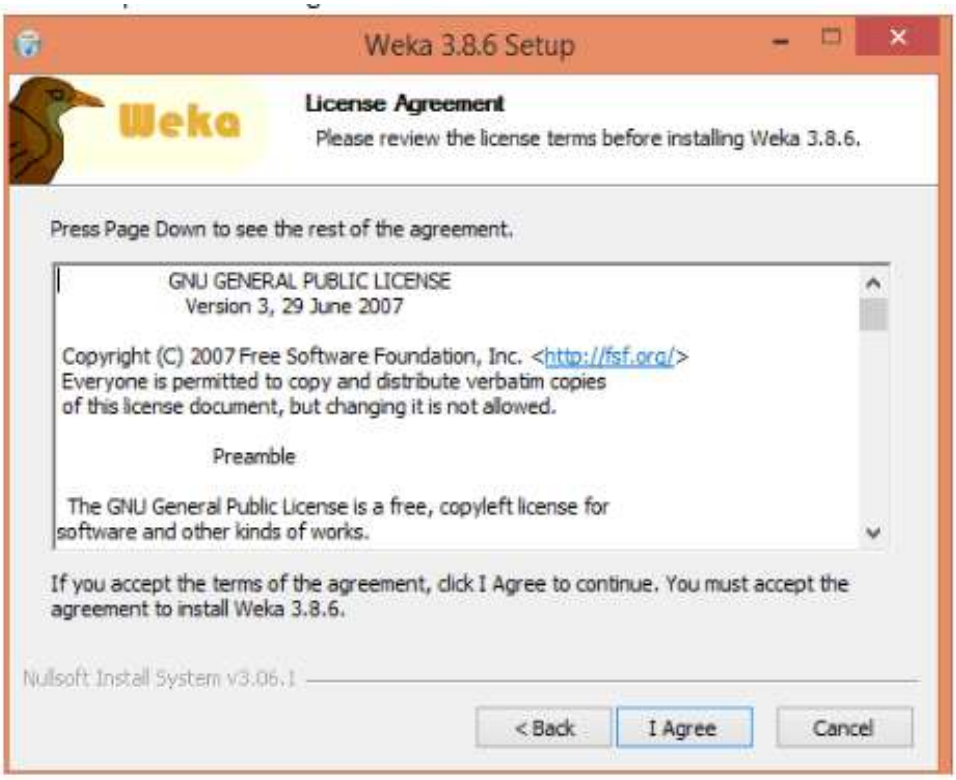
4. Procedure :

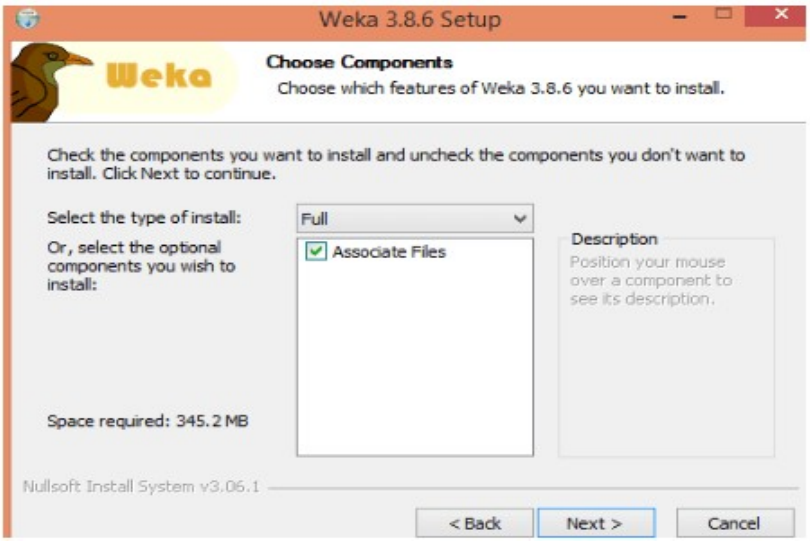
Download Weka:

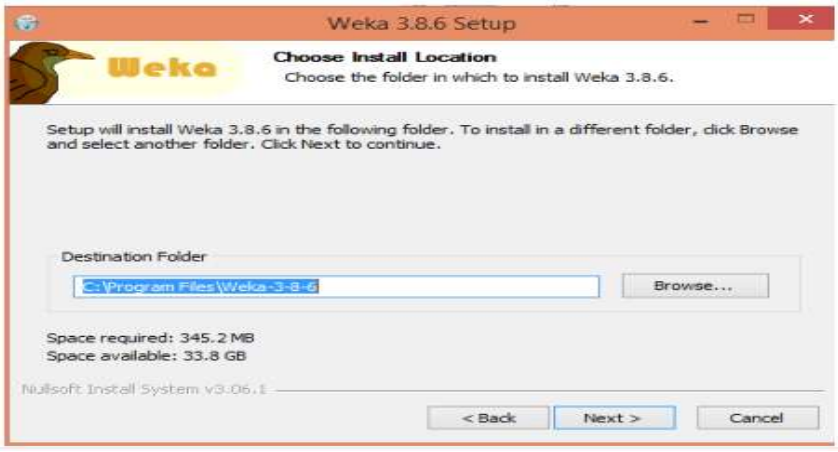
- Visit the official website: https://www.cs.waikato.ac.nz/ml/weka/
- Click on *Download* and choose the version suitable for your OS.

Install Weka:

- Run the downloaded installer.
- Follow the installation wizard and complete the setup.




Weka 3.8.6 Setup
Weka
Choose Components
Choose which features of Weka 3.8.6 you want to install.
Check the components you want to install and uncheck the components you don't want to install. Click Next to continue.
Select the type of install:
Full
Or, select the optional components you wish to install:
Associate Files
Description
Position your mouse over a component to see its description.
Space required: 345.2 MB
Nullsoft Install System v3.06.1
< Back
Next >
Cancel


Weka 3.8.6 Setup
Weka
Choose Install Location
Choose the folder in which to install Weka 3.8.6.
Setup will install Weka 3.8.6 in the following folder. To install in a different folder, click Browse and select another folder. Click Next to continue.
Destination Folder
C:\Program Files\Weka-3-8-6
Browse...
Space required: 345.2 MB
Space available: 33.8 GB
Nullsoft Install System v3.06.1
< Back
Next >
Cancel


Weka 3.8.6 Setup
Completing Weka 3.8.6 Setup
Weka 3.8.6 has been installed on your computer.
Click Finish to close Setup.
< Back
Finish
Cancel

- **Launch Weka:**
  - Open the Weka application.

  - Choose the **Explorer** panel to begin working with data.
- **Load Dataset:**
  - Use built-in datasets like iris.arff or import your own .csv or .arff file.
- **Apply Algorithms:**
  - Navigate to the **Classify** tab to apply classification (e.g., J48 Decision Tree).
  - Try clustering using the **Cluster** tab (e.g., K-means algorithm).
- **Analyze Results:**
  - Observe the textual output, accuracy metrics, and tree/graph visualizations.
  - Reflect on what the model has learned and how it performed.

5. Learning Outcomes :

- Gain hands-on experience with the Weka data mining tool.
- Understand how to load and preprocess datasets.
- Apply basic machine learning algorithms like J48 and K-means.
- Evaluate model performance through Weka's output metrics.

Experiment – 3.1

Student Name: Prabhjot Kaur **UID: 24MCS10022**

Branch: ME-CSE **Section/Group: 24MCS-1/A**

Semester: 2nd **Date of Performance: /0 /2025**

Subject: Business Analytics **Subject Code: 24CSH-661**

1. Aim : Perform the preprocessing in the selected dataset using Weka.

2. Task to be done:

- Load a dataset (e.g., .arff or .csv) into Weka.
- Identify and handle missing or inconsistent values.
- Apply filters for normalization or attribute selection.
- Save the preprocessed dataset for further analysis.

3. Description :

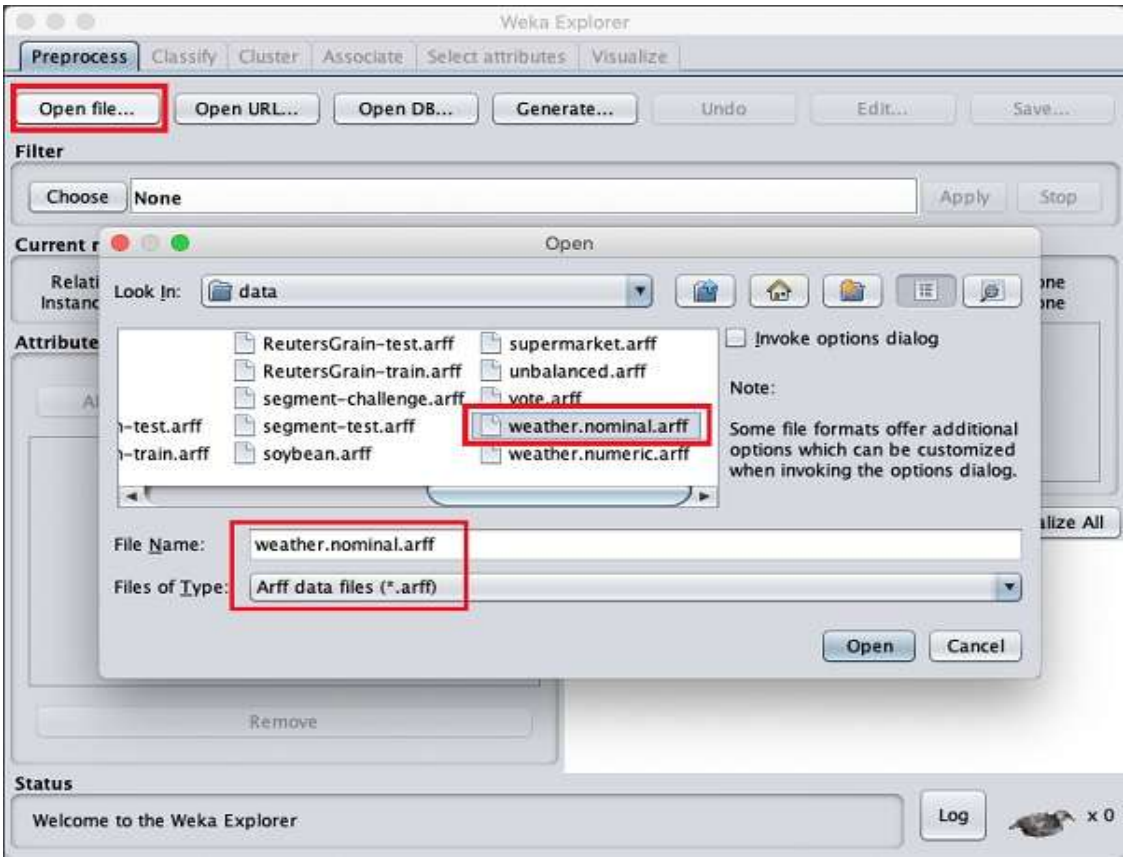
Data preprocessing is a crucial step in the data mining process, as it prepares raw data for analysis by transforming it into a suitable format. Weka provides a powerful set of preprocessing tools that allow users to clean, normalize, transform, and filter data before applying machine learning algorithms. Through Weka's *Preprocess* panel, users can handle missing values, convert attribute types, apply filters, and select relevant features. In this activity, students will use Weka to load a dataset and perform various preprocessing operations to enhance the quality and consistency of the data. This step improves model accuracy and ensures more meaningful outcomes during the data mining process.

4. Procedure :

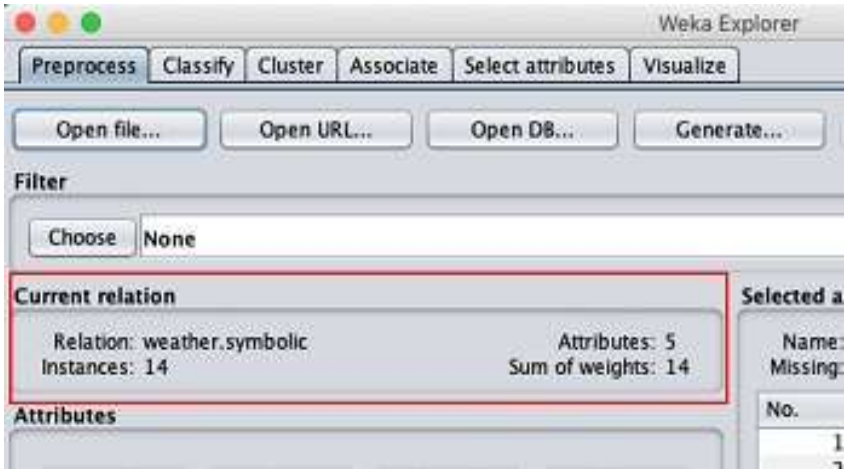
The data that is collected from the field contains many unwanted things that leads to wrong analysis. For example, the data may contain null fields, it may contain columns that are irrelevant to the current analysis, and so on. Thus, the data must be preprocessed to meet the requirements of the type of analysis you are seeking. This is the done in the preprocessing module.

To demonstrate the available features in preprocessing, we will use the **Weather** database that is provided in the installation.

Using the **Open file** ... option under the **Preprocess** tag select the **weather-nominal.arff** file.



When you open the file, your screen looks like as shown here −



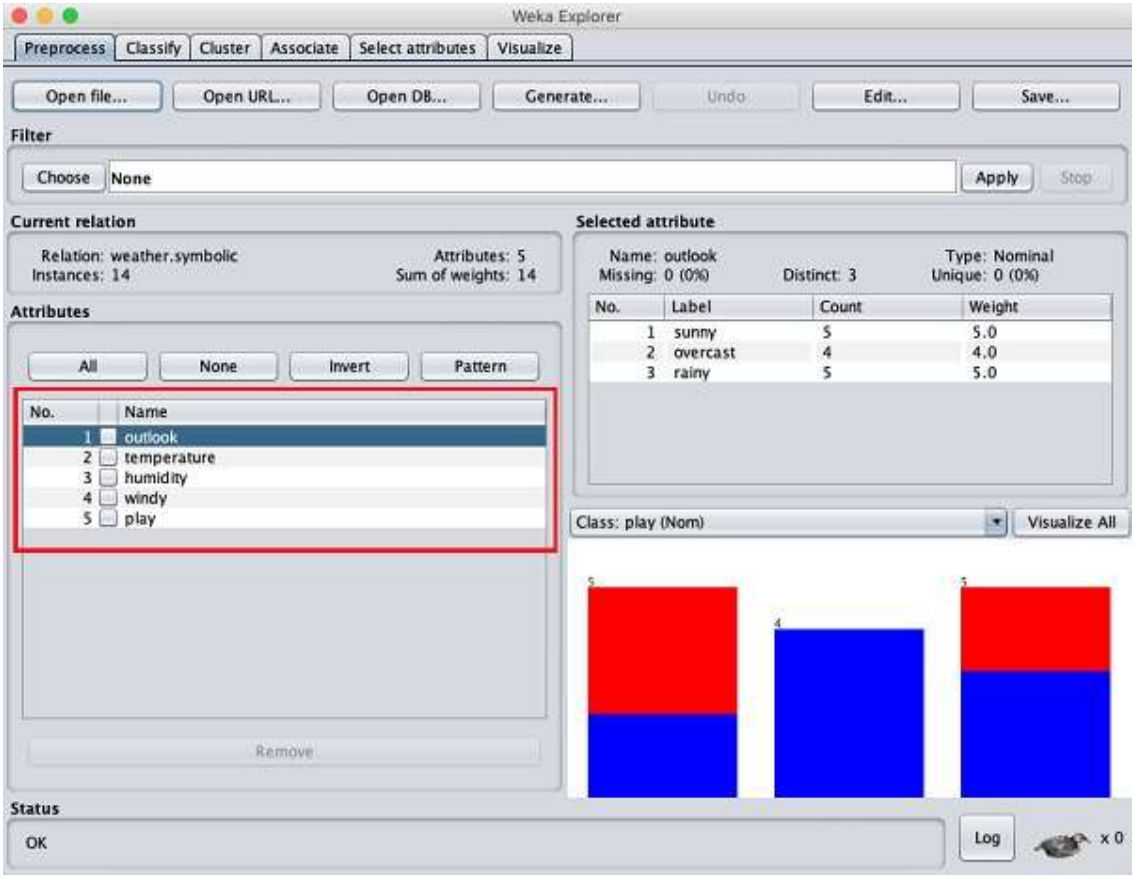
This screen tells us several things about the loaded data, which are discussed further in this chapter.

Understanding Data

Let us first look at the highlighted Current relation sub window. It shows the name of the database that is currently loaded. You can infer two points from this sub window −

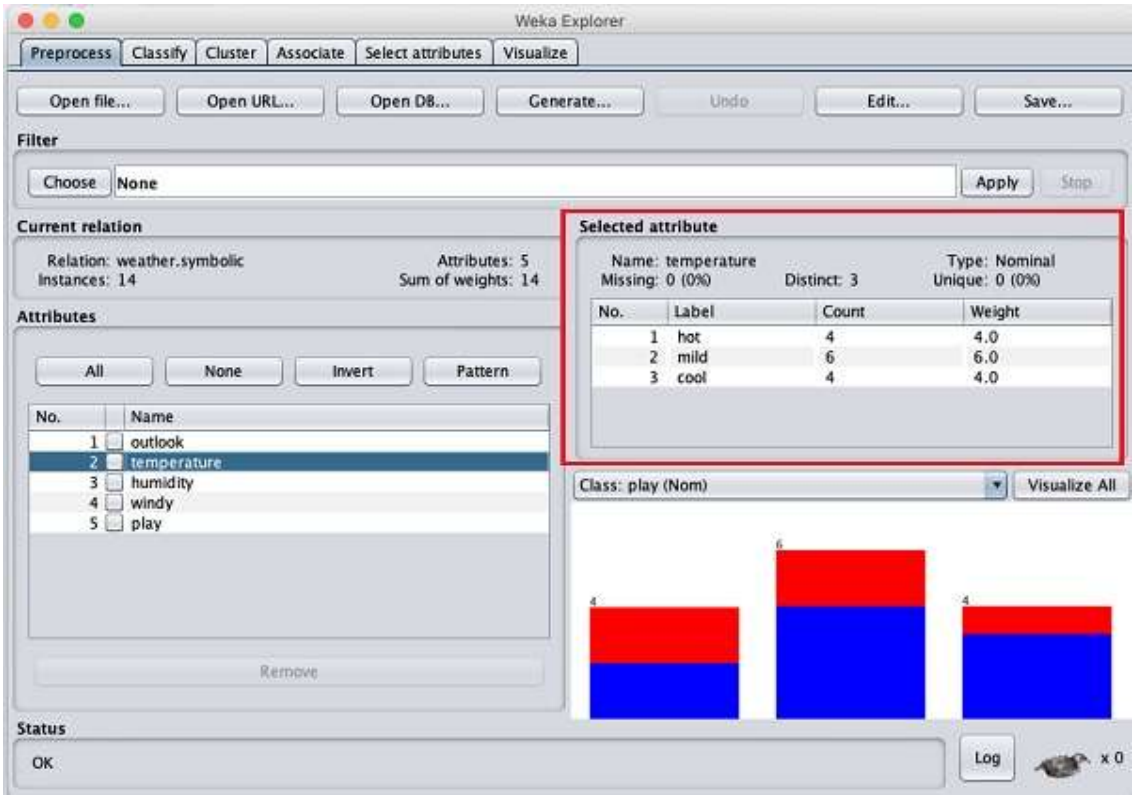
- There are 14 instances - the number of rows in the table.
- The table contains 5 attributes - the fields, which are discussed in the upcoming sections.

On the left side, notice the Attributes sub window that displays the various fields in the database.



The **weather** database contains five fields - outlook, temperature, humidity, windy and play. When you select an attribute from this list by clicking on it, further details on the attribute itself are displayed on the right hand side.

Let us select the temperature attribute first. When you click on it, you would see the following screen −

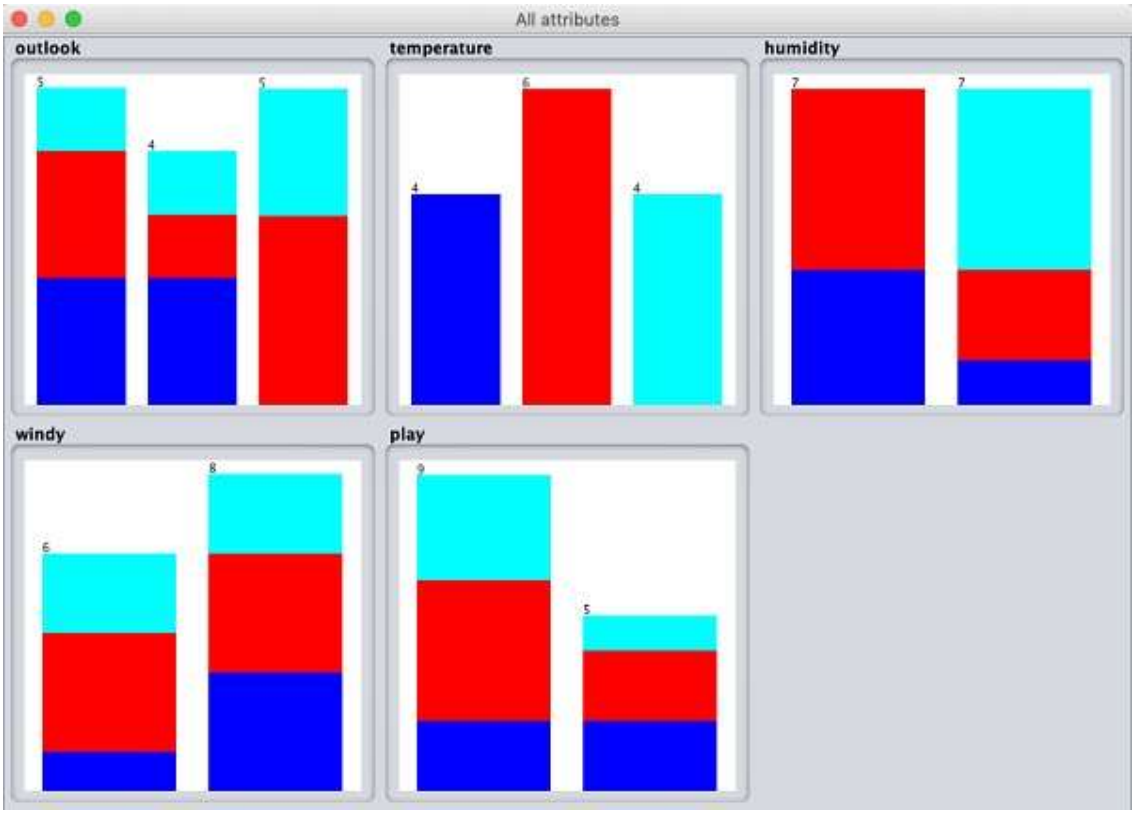


In the Selected Attribute subwindow, you can observe the following −

- The name and the type of the attribute are displayed.
- The type for the temperature attribute is Nominal.
- The number of Missing values is zero.
- There are three distinct values with no unique value.
- The table underneath this information shows the nominal values for this field as hot, mild and cold.
- It also shows the count and weight in terms of a percentage for each nominal value.

At the bottom of the window, you see the visual representation of the class values.

If you click on the Visualize All button, you will be able to see all features in one single window as shown here −



Removing Attributes

Many a time, the data that you want to use for model building comes with many irrelevant fields. For example, the customer database may contain his mobile number which is relevant in analysing his credit rating.

To remove Attribute/s select them and click on the **Remove** button at the bottom.

The selected attributes would be removed from the database. After you fully preprocess the data, you can save it for model building.

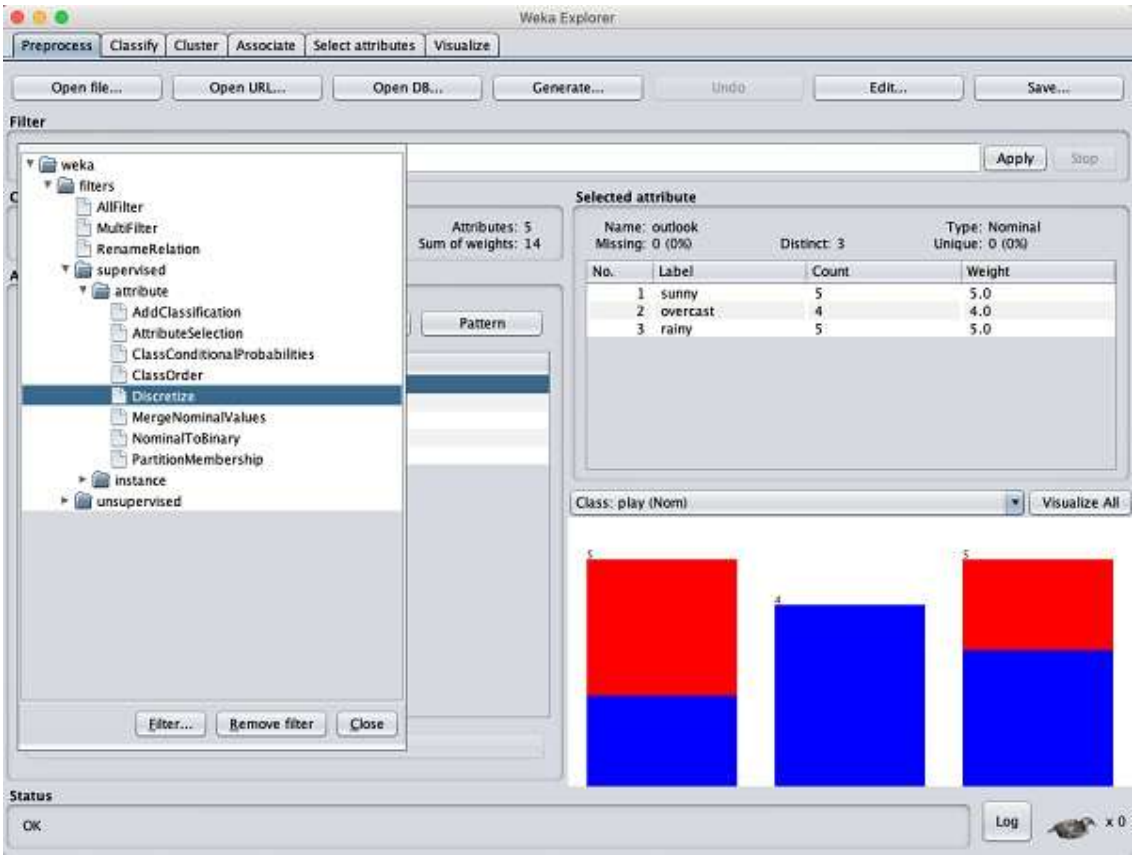
Next, you will learn to preprocess the data by applying filters on this data.

Applying Filters

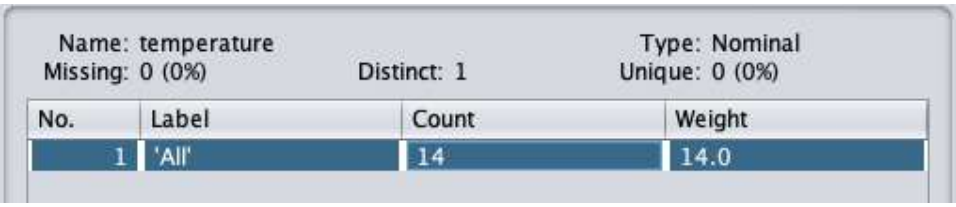
Some of the machine learning techniques such as association rule mining requires categorical data. To illustrate the use of filters, we will use **weather-numeric.arff** database that contains two **numeric** attributes - **temperature** and **humidity**.

We will convert these to **nominal** by applying a filter on our raw data. Click on the **Choose** button in the **Filter** subwindow and select the following filter −

weka→filters→supervised→attribute→Discretize



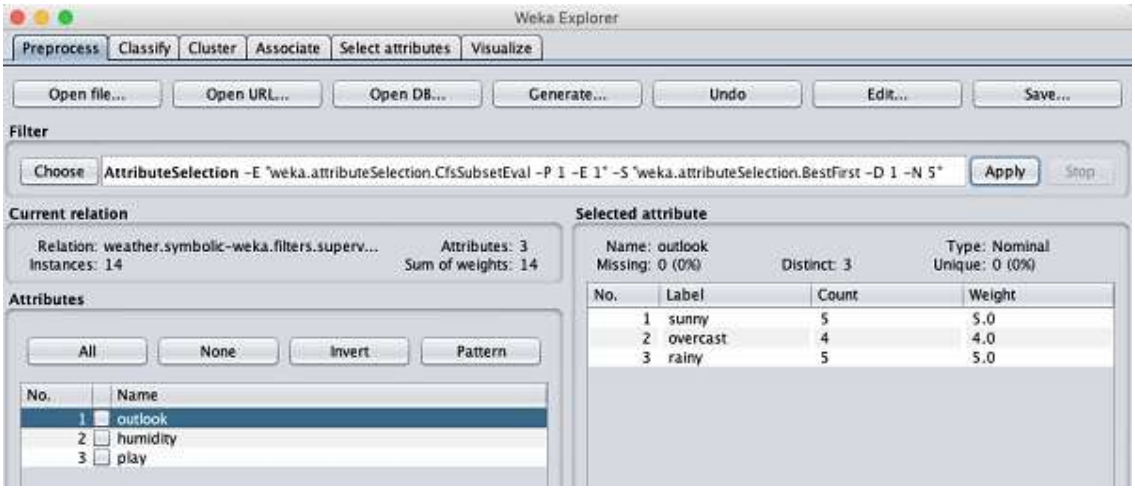
Click on the **Apply** button and examine the **temperature** and/or **humidity** attribute. You will notice that these have changed from numeric to nominal types.



Let us look into another filter now. Suppose you want to select the best attributes for deciding the **play**. Select and apply the following filter −

weka→filters→supervised→attribute→AttributeSelection

You will notice that it removes the temperature and humidity attributes from the database.



After you are satisfied with the preprocessing of your data, save the data by clicking the **Save** button. You will use this saved file for model building

5. Learning Outcomes :

- Understand the importance of data preprocessing.
- Learn to identify and handle missing or noisy data.
- Apply Weka filters for transforming and selecting attributes.
- Gain skills in preparing clean data for model building.

Experiment – 3.2

Student Name: Prabhjot Kaur **UID: 24MCS10022**

Branch: ME-CSE **Section/Group: 24MCS-1/A**

Semester: 2nd **Date of Performance: /0 /2025**

Subject: Business Analytics **Subject Code: 24CSH-661**

1. Aim : Perform classification on the selected dataset using Weka.

2. Task to be done:

- Load the selected dataset in Weka Explorer.
- Choose and apply a classification algorithm (e.g., J48).
- Perform model evaluation using cross-validation.
- Analyze the output results and accuracy metrics.

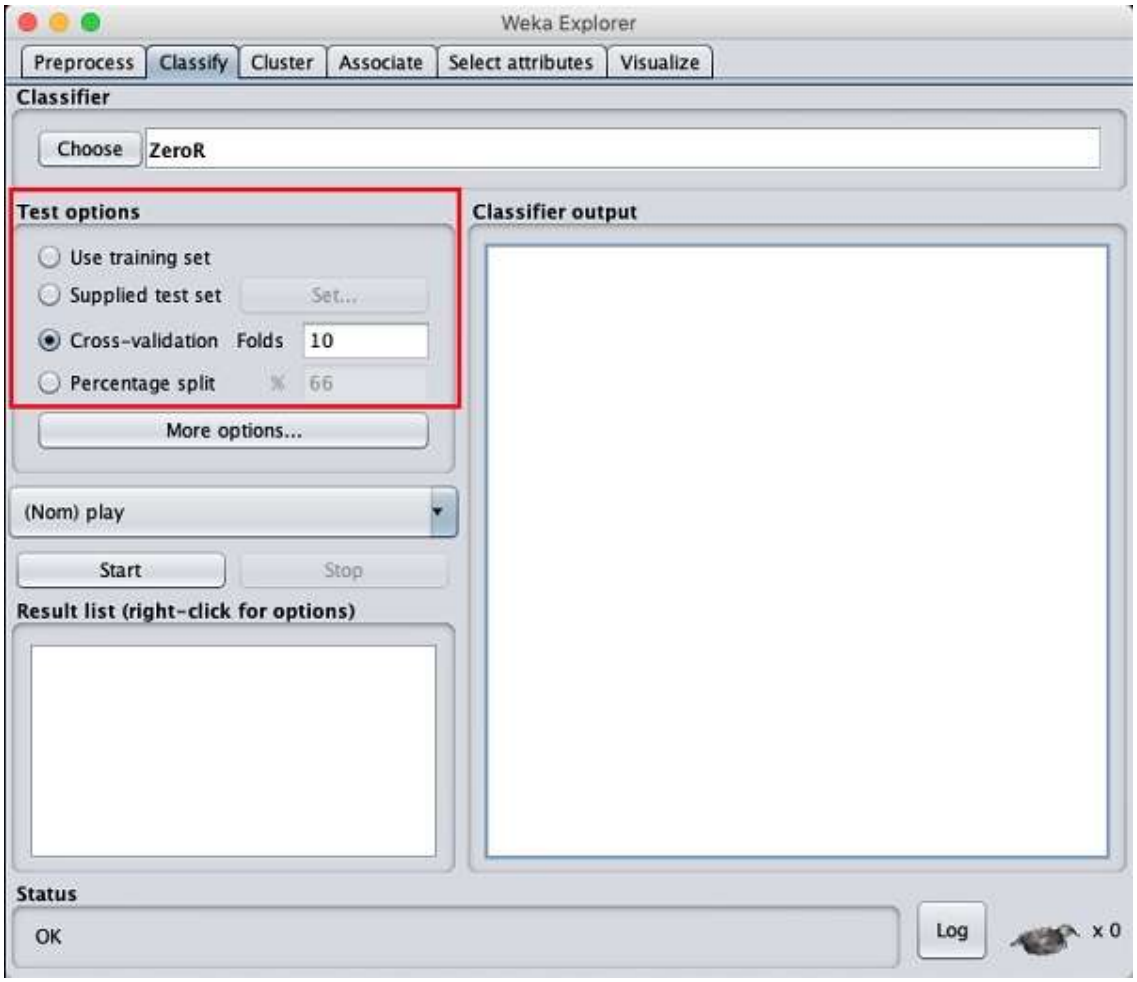
3. Description :

Classification is a supervised machine learning technique used to assign labels or categories to data instances based on input features. Weka provides a variety of built-in classification algorithms such as J48 (Decision Tree), Naive Bayes, Random Forest, and more, which can be easily applied through its graphical interface.

Learn to load a dataset into Weka and apply classification algorithms to build predictive models. The performance of these models will be evaluated using tools like confusion matrices, accuracy scores, and cross-validation, helping users understand model behavior and effectiveness in real-world scenarios.

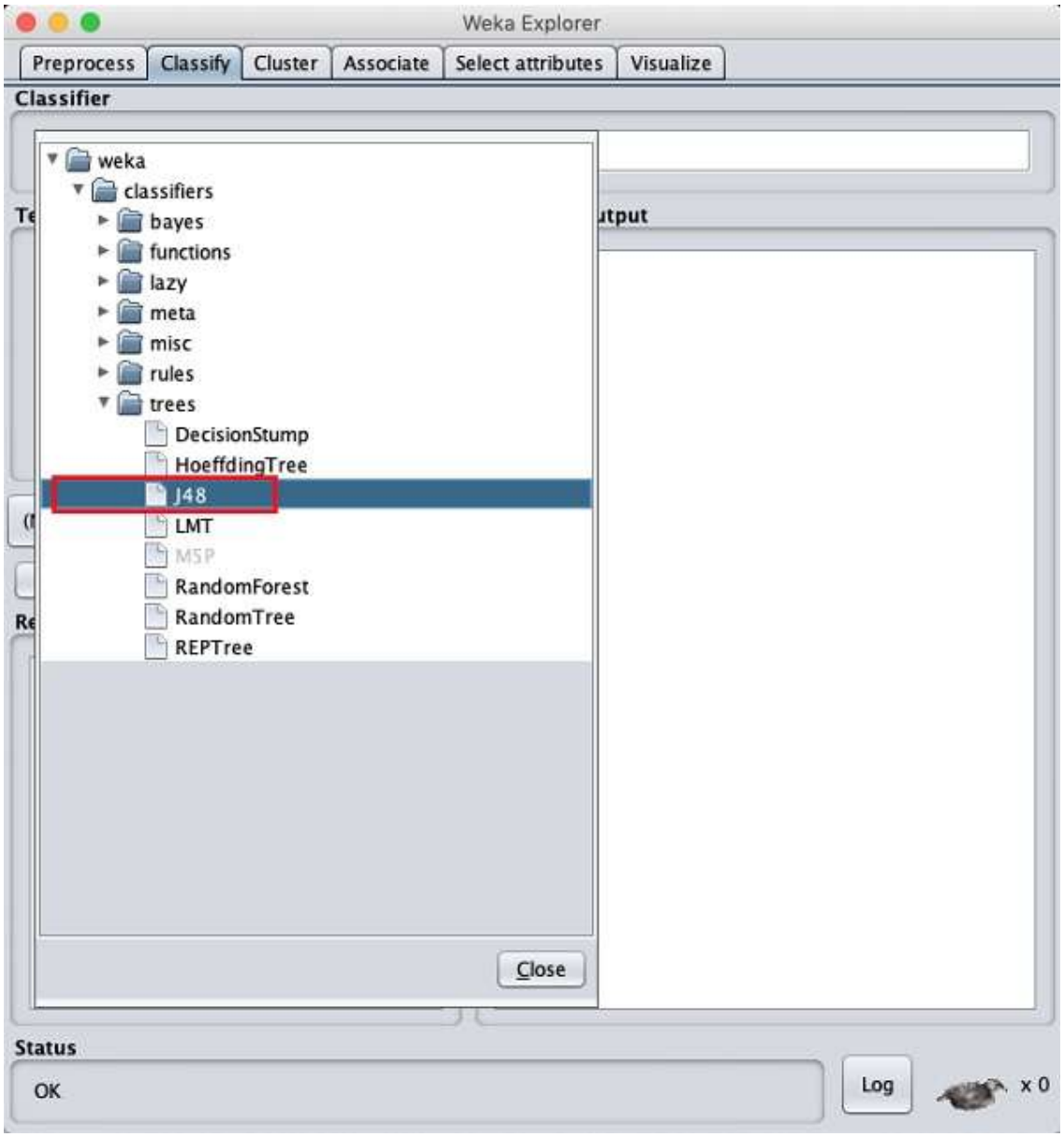
4. Procedure :

1. **Load Preprocessed Dataset**
    - Open Weka and go to the Preprocess tab.
    - Click Open file... and load the previously preprocessed dataset (e.g., *weather.arff*).
    - Ensure that the class attribute (e.g., play) is selected as the output class.



2. **Select the Classification Tab**
    - Click on the Classify tab.
    - You will see options to set up the testing method and to choose a classifier.
3. **Choose a Testing Method**
    - Select one of the following options for model evaluation:
        - Training set (uses full dataset for training — not ideal for real testing)
        - Supplied test set (if you have an external test dataset)
        - Cross-validation (recommended; default is 10-fold)
        - Percentage split (e.g., 66% training, 34% testing)
    - For most cases, use Cross-validation (10 folds).
4. **Select a Classifier**
    - Click on the Choose button.

- Navigate to: trees → J48 (this is Weka's version of the C4.5 Decision Tree algorithm).

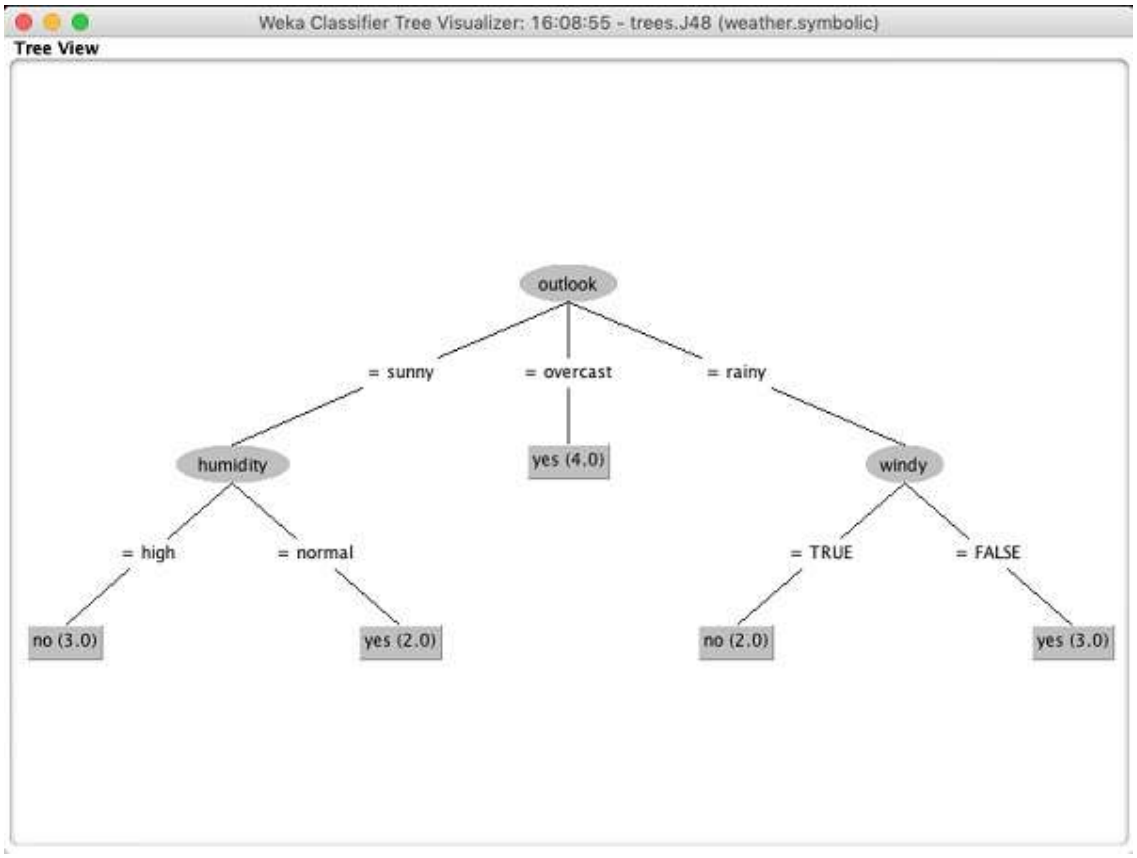


- You can optionally configure J48 parameters before proceeding.

5. **Start Classification**
   - Click the Start button.
   - Weka will run the selected classifier and display the results on the right panel.
6. **Interpret the Results**

- Check the Summary, including:
  - Number of correctly and incorrectly classified instances
  - Relative absolute error
  - Confusion matrix
- Use these metrics to evaluate model performance.

7. **Visualize the Tree and Errors**
   - In the Result list, right-click on the result entry.
   - Select Visualize tree to see the decision tree structure.
   - Select Visualize classifier errors to view a plot showing correct vs. incorrect classifications.
     - Crosses (X) indicate correct predictions
     - Squares (■) indicate misclassifications
   - Use the jitter slider or drop-down lists to adjust plot views and axes.



8. **Refine the Model (Optional)**
    - If performance is poor, consider:
        - Using more data
        - Feature selection or transformation
        - Trying different classifiers (e.g., Naive Bayes, Random Forest)

5. Learning Outcomes :

- Understand the concept of classification in machine learning.
- Learn how to apply classifiers using Weka.
- Interpret model evaluation metrics like accuracy and confusion matrix.
- Compare performance of different classification algorithms.

Experiment – 3.3

Student Name: Prabhjot Kaur **UID: 24MCS10022**

Branch: ME-CSE **Section/Group: 24MCS-1/A**

Semester: 2nd **Date of Performance: /0 /2025**

Subject: Business Analytics **Subject Code: 24CSH-661**

1. Aim :
Perform Clustering and KNN on the selected dataset using Weka.

2. Task to be done:

- Load the iris.arff dataset in Weka's Preprocess tab.
- Go to the Cluster tab and apply Simple K-Means clustering.
- Switch to the Classify tab and apply the IBk classifier (KNN).
- Analyze and visualize the output clusters and KNN classification results.

3. Description :

Clustering and K-Nearest Neighbors (KNN) are two fundamental machine learning techniques used for analyzing and grouping data. We use Weka to perform K-Means Clustering, an unsupervised algorithm that groups similar data points into clusters based on their features. We also apply KNN, a supervised learning algorithm that classifies a new instance based on the majority class among its nearest neighbors in the dataset.

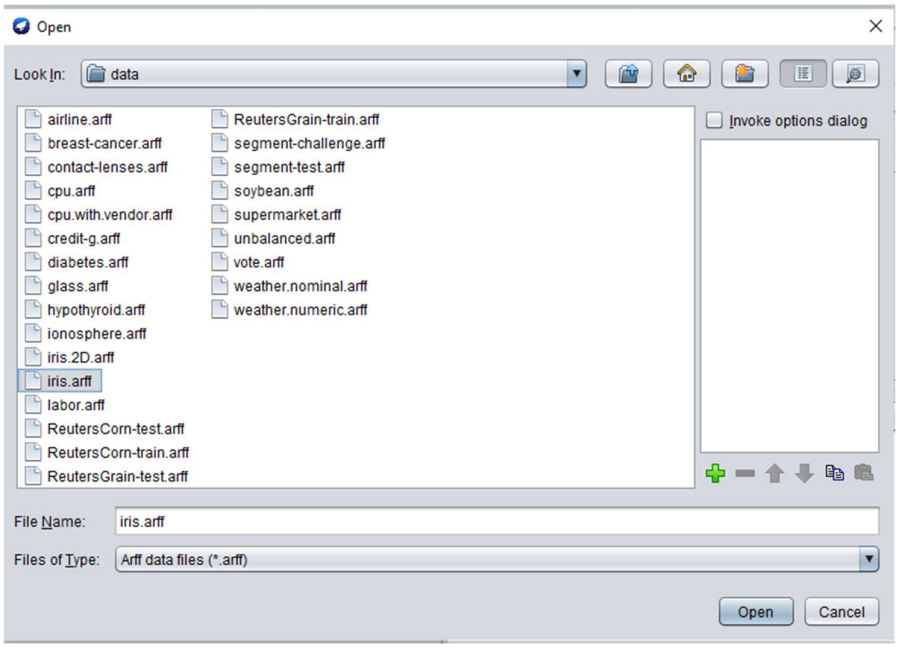
The dataset used (e.g., *iris.arff*) is first preprocessed, and then both techniques are applied using Weka's graphical interface. This activity demonstrates how to use Weka to discover natural data groupings and classify instances based on proximity, helping learners better understand pattern discovery and classification.

4. Procedure :

Step 1: Load Dataset

- Open Weka Explorer.

- Click on the Preprocess tab.
- Load the iris.arff dataset (or any other preprocessed dataset).

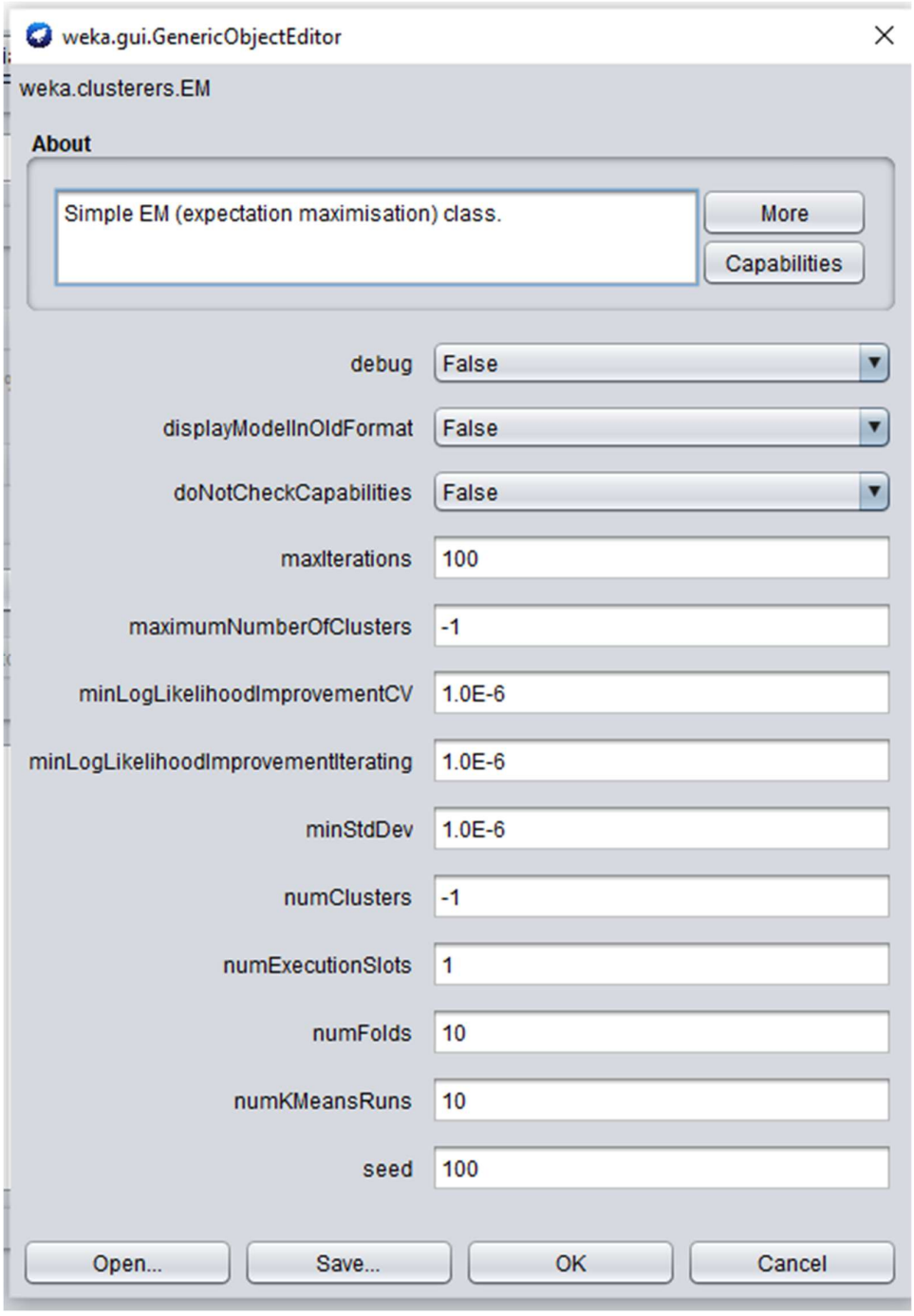


Step 2: Select Clustering Algorithm

- Click the Cluster tab.
- Press the Choose button.
- From the list, select:
  weka → clusterers → SimpleKMeans

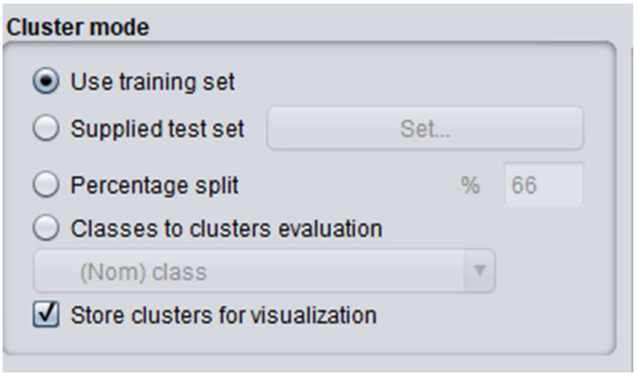
Step 3: Set Number of Clusters

- Next to the Choose button, click on the text box icon to open algorithm settings.
- Set the number of clusters to 3.
- Leave the seed value unchanged (default is fine).
- Click OK to save settings.



Step 4: Run the Clustering

- Ensure Cluster Mode is set to Use training set.
- Click the Start button to execute clustering.



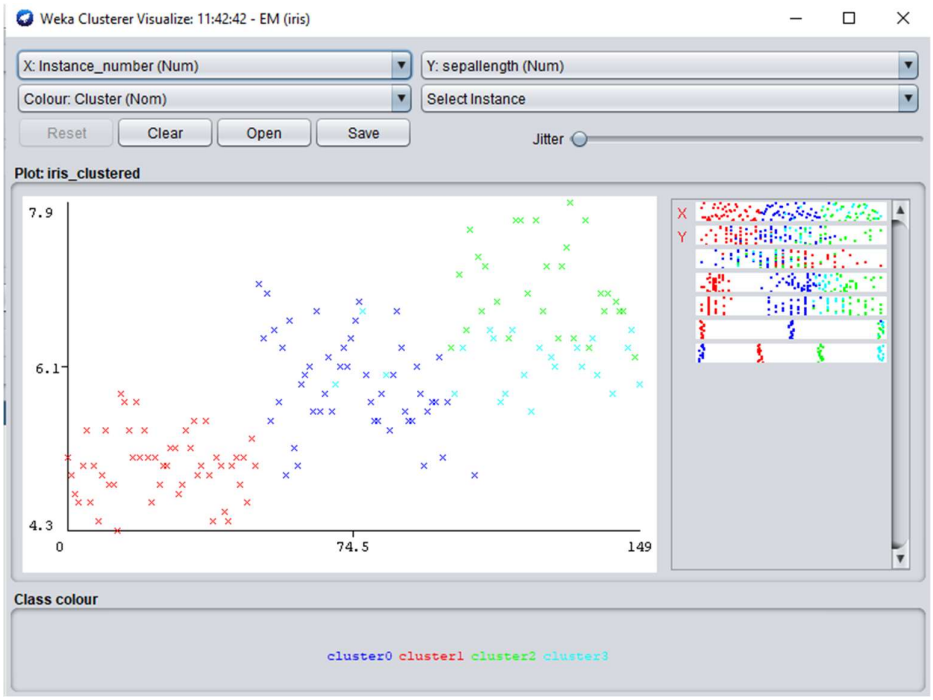
Step 5: View Results

- After processing, Weka will display:
    - Cluster centroids (mean vectors)
    - Count and percentage of instances in each cluster

```
Number of clusters selected by cross validation: 4
Number of iterations performed: 16


                    Cluster
Attribute                0       1       2       3
                    (0.32)  (0.33)   (0.2)  (0.14)
=====================================================
sepallength
  mean                5.897   5.006  6.9426  6.1304
  std. dev.          0.5279  0.3489   0.498  0.2943

sepalwidth
  mean                2.7519   3.418  3.1103  2.8088
  std. dev.          0.3103  0.3772  0.2952  0.2361

petallength
  mean                4.2267   1.464  5.8559  5.0993
  std. dev.           0.445  0.1718  0.4626  0.2462

petalwidth
  mean                1.3134   0.244  2.1495  1.8254
  std. dev.          0.1864  0.1061   0.232  0.2152

class
  Iris-setosa              1      51       1       1
  Iris-versicolor    48.1125       1  1.0182  3.8693
  Iris-virginica      2.0983       1 31.0375 19.8641
  [total]            51.2108      53 33.0557 24.7335
```

Step 6: Visualize Clusters

- Right-click on the result entry in the Result list.
- Select Visualize cluster assignments to see the data points grouped visually.
- You can change axes or add jitter for better clarity.



5. Learning Outcomes :

- Understand how K-Means groups data into clusters based on similarity.
- Learn to classify instances using the K-Nearest Neighbors algorithm.
- Gain hands-on experience with Weka's Cluster and Classify interfaces.
- Interpret cluster centroids, classification accuracy, and visual outputs.