

## Practical. No. 1.

Title: Comparative Study of Open-Source Data analysis tools.

Theory:

Data analysis:

Data analysis is now a priority for businesses, and choosing the right data analysis tool is key to turning troves of data into usable information.

Data analytics is the science of analyzing raw to make conclusions about their information. Many of the techniques & processes of data analytics have been automated into mechanical processes & algorithms that work over raw data for human consumption.

• What are Data Analysis Tools?

Data analysis tools are software and programs that collect & analyze data about a business, its customers & its competition, in order to improve process & help uncover insights to make data-driven decisions.

Data Analysis tool enables businesses to analyse vast stores of data for great competitive advantage.

1] Grafana: Grafana is an open-source platform primarily used for monitoring & real-time visualization of time-series data. It is

widely adopted in IT operations, DevOps & system monitoring to track performance metrics, infrastructure health, and application logs. Unlike Power BI and Tableau, which focus on business intelligence, Grafana specializes in live data streaming & integrates seamlessly with time-series databases such as Prometheus, influxDB, Elasticsearch, Graphite & Loki.

Grafana provides high customization options, allowing users to create dynamic dashboards, with various visualization plugins. It supports alerts & notifications, which are critical for system administrators to monitor anomalies. However it requires some technical expertise for setup & configuration, making it less user-friendly for non-technical users.

## 2. Power BI.

Power BI developed by Microsoft, is a powerful business intelligence & data analytics tool designed for enterprises & business users. It enables organizations to collect, analyze & visualize structured and unstructured data from various sources such as SQL databases, Excel, Azure, Google Analytics & cloud-based services. It integrates well with Microsoft's ecosystem, making it an excellent choice.

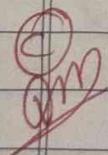
for business already using office 365, sharepoint & Azure service.

### 3. Tableau:

Tableau: is a leading data visualization & business intelligence tool known for its advanced analytics capabilities. It is widely used by data analysts, researchers & enterprises for in-depth data exploration, storytelling, & interactive dashboards. Tableau supports a broad range of data sources, including SQL databases, spreadsheets, cloud platforms, & big data sources like Hadoop & spark.

~~Tableau's key strength is its high-quality visualization, which allows users to create detailed, interactive reports, with drag-and-drop functionality. Unlike Power BI, which is more focused on Microsoft integrations, Tableau provides cross-platform flexibility, making it suitable for organizations with diverse tech stacks.~~

Conclusion: Here we learned about Data Analysis tools using comparative study.



## Practical No. 2.

Title: Identify key performance indicators for any real time case study & present analysis of the same.

Aim: Identify

objectives: To study key performance indicators.

- Present analysis for real time case study.

Software: Microsoft power BI.

Theory:

- Brief about KPI.
- A key performance indicator is quantifiable measure of performance over time for a specific strategic objective. Business leaders & senior executives use KPI to judge the effectiveness of their efforts & make better informed decisions.

KPI stands for key performance indicator. KPIs provide targets for teams to shoot for, milestones to gauge progress, and insights that help people across the organization make better decisions.

• Steps to identify KPI.

1. ~~Who, what, how~~: Be clear about who the audience is, what they want, & how they're going to use the KPIs. This means working with your stakeholders to identify the core KPIs that map directly to their goals & strategy.

2. Be SMART: The popular acronym stands for Specific, Measurable, Attainable, Realistic

# DPU

Time-bound. Iterate & revote. Over time, see how you or your audience are using the set of KPIs & if you find that certain ones aren't relevant, remove or replace them.

3. Iterate & evolve: over time, see how you or your audience are using the set of KPIs & if you find that certain ones aren't relevant, remove or replace them.

KPI requirements:

A designer bases a KPI visual on a specific measure. The intention of the KPI is to help you evaluate the current value & status of a metric against a defined target. A KPI visual requires a base measure that evaluates to a value, a target measure or value, & a threshold or goal.

A KPI dataset needs to contain goal values for a KPI. If your dataset doesn't contain goal values, you can create them by adding an Excel sheet with goals to your data model or PBIX file.

• For Power BI desktop:

To follow along, use the Retail Analysis sample PBIX file.

- 1] From the upper left section of the menubar, select File > open report.
- 2] Find your copy of the Retail Analysis Sample PBIX file.
- 3] Open the Retail Analysis sample PBIX file in report view.

4. Select + to add a new page.

• How to create a KPI.

• In this example, you'll create a KPI that measures the progress you have made toward Covid Trends in female Age wise.

To Get KPI:

1. Open Power BI.

2. Click on Get data icon & insert the selected data set.

3. Once the Dataset is inserted successfully, To get the visual in KPI select the KPI icon from the visualization pane.

4. To add a goal, drag female Age group to the Target Field. Grand Total to value & month we have to compare in Trend Axis.

5. Optionally, format the KPI by selecting the paint brush icon to open the format visual pane.

KPI:

A Key Performance Indicator is measurable value that indicates how effectively an individual, team or organization is achieving specific objectives. KPIs help track the progress toward strategic goals & can be categorized into various types, such as financial, operational, customer-focused or employee-related metrics.

Types of KPI:

1] Leading KPI: Predict future performance.

- 2) lagging KPIs → reflect past performance.
- 3) Quantitative KPIs → measurable with numbers.
- 4) Qualitative KPIs → Subjective assessments.

Conclusion: We studied about what is KPI & How to create it.

## Practical No. 3.

**Title:** Create, model & analyze Petri nets with a Standards-compliant Petri net tool for Producer / consumer OR Dining philosophers problem.

**Theory:** Modelling the petri net:

- Next, we will model the petri net using a standards-compliant petri net tool.
- There are several options available, such as CPN Tools, Renew
- Let's use PIPER for this example.
- Open PIPER & create a new petri net.
- Add two places & name them 'buffer' & 'resources'.
- Add two transitions & name them 'produce' & 'consume'.
- Add input arcs from 'buffer' to 'produce' & from 'resource' to 'consume'.
- Add output arcs from 'produce' to 'resource' & from 'consume' to 'buffer'.
- The petri net should look like this:
- Analyze the petri net:
  - Once we have modelled the petrinet, we can analyze it to determine its behaviour.
  - One useful analysis is to check for liveness, which means that all transitions can eventually be fired.
  - To check for liveness in PIPER, go to the 'Analysis' menu & select 'liveness'.
  - The result should show that the petri net is live

# DPU

Which means that both the 'produce' & 'consume' transition can fire indefinitely.

A petri net is a mathematical modeling tool used to describe & analyze system with concurrent, parallel, or distributed processes. It is widely applied in workflow modeling, distributed computing, communication protocol & manufacturing system.

Components of Petri net:

- 1] places (circle,  $(o)$ ) - Represent conditions or system states.
- 2] Transitions: (rectangle or 'Bar',  $\square$ ). Represent events that cause state changes.
- 3] Token (Dot,  $\bullet$ ): Represent dynamic resources or conditions moving between places.
- 4] Arcs (Directed Arrows " $\rightarrow$ ") - Define the flow between place & transitions.

Working of a Petri Net:

- A transition is enabled when all its input places contains the required number of tokens.
- When the transition fires, it consumes tokens from input places & produces tokens in output places.
- This models how a system evolves over time with discrete state changes.
- Types of Petri nets:
  - 1] Ordinary Petri Nets: Basic form with simple token movement.

2. Colored Petri Nets (CPN) - Tokens carry data values, enhancing expressiveness.
3. Timed Petri Nets: Include timing constraints for performance modeling.
4. Stochastic Petri Nets: Use probabilities for event occurrence.

Applications:

- Computer Science: Process synchronization, deadlock detection.
- Manufacturing: Work flow & production line analysis.
- Networking: Communication protocol, packet switching.
- Biology: Modeling biochemical reactions & metabolic pathways.
- Formal Definition of Petri net:

A Petri-Net is formally defined as 5-tuple:

$$PN = (P, T, F, w, m_0)$$

Where,

$P$ : Finite set of places

$T$ : Finite set of transitions

$F \subseteq (P \times T) \cup (T \times P)$  = Flow relation

$w$ :  $f \rightarrow N$  = Weight function.

$m_0$ :  $P \rightarrow N$  = Initial marking.

Conditions:

- $P \cap T = \emptyset$  (place & transition are distinct)
- $P \neq \emptyset, T \neq \emptyset$

Conclusion: We studied about Petri nets & how they work.

# DPU

AS Practical No. 4.

Aim: Perform a what-if analysis on Book store Scenario using Excel.

Objective: To study about what-if analysis in Excel.

Software: In Microsoft, Excel.

Theory: In Excel, what-if analysis is process of changing cells values to see how those changes will affect the worksheet's outcome. You can use several set of values to explore all the result in one or more formulas.

What-if Excel is used by almost every data analyst & especially middle to higher management professional to make better, faster & more accurate decisions based on the data. What-if analysis is useful in many situations, such as:

- You can propose different budgets based on revenue.
- You can predict the future values based on the given historical values.
- If you expect a certain value due to formula, you can find different sets of input values that produce the desired result.

Scenario: A book store want to analyze the impact of changing the price & the quantity sold of a popular book on their profit.

What-if Analysis ~~&~~ Scenario-Manager:

Scenario manager is one of the what-if analysis

# DPU

tool is in Excel. Scenario manager is useful in case where you have more than two variable in the sensitivity analysis. Scenario manager creates scenario for each set of the input values for the variables under considerations.

Step 2: Set up the data. Create a table with the following:

- 1] Price
- 2] Quantity sold
- 3] Revenue
- 4] Total cost
- 5] Profit.

Enter the current data for the book.

Price:

Quantity sold:

100 Total cost:

\$400

use the following formulas to calculate the revenue & profit.

To create an analysis report with Scenario Manager, follow the following steps such as.

Step 1: Click the Data tab

Step 2: Go to the What-if Analysis button & click on the Scenario Manager from the dropdown list.

Step 3: Now a scenario manager dialog box appears, click on the Add button to create a scenario.

Step 4: Create the scenario, name the scenario, enter the value for each changing input cell for

that scenario & then click the ok button.

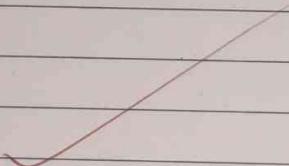
steps: click on OK. Then click on Summary.

Steps: Click on Show button. Set the result cell number to B6:B16.

Steps: Click on OK. And scenario 4 to Summary will be displayed.

NOTE: Elaborate in more details of if-else analysis.

Conclusion: Thus we studied about What-if analysis.



## Practical No. 5.

Aim: Create a decision tree for predicting the loan eligibility process using python.

Objective: Predict loan eligibility using a decision tree.

Software used: Jupyter notebook:

Theory: ~~NOTE: Write basics of python language.~~

~~Identify the problem & gather data:~~

- The loan eligibility process involves determining whether a person is eligible for a loan based on certain criteria, such as income, credit score, & employment status.
- To create a decision tree, we need together dataset includes these criteria as well as the loan eligibility status.
- The predictor variable are the criteria used to determine loan eligibility, such as income, credit score, and employment status.
- To create the decision tree, we will use an algorithm that recursively splits the data into subsets based on the predictor variable.

Decision Tree :

A Decision tree: is a supervised machine learning algorithm used for classification & regression tasks. It is a tree-like

Structure where each internal node represents a decision each branch represents an outcome of that decision. & each leaf node represents a class label or predicted value.

Key components:

- 1) Root Node: The starting point of the tree containing the entire dataset.
- 2) Internal Nodes - Represents decision points based on feature conditions.
- 3) Branches - Show the possible outcomes for a decision.
- 4) Leaf Nodes: Represent the final output

Working of Decision Tree:

- 1) Feature Selection: The algorithm selects the best feature using criteria like:
  - Gini Index
  - Entropy: for classification
  - Variance Reduction (for regression)
- 2) splitting: The dataset is recursively split based on the chosen feature.
- 3) Stopping Criteria: The process stops when a predefined condition is met.
- 4) Prediction: For a new input, it follows the decision path from the root to a leaf node.

Advantages:

- Easy to interpret & visualize.
- Handles both numerical & categorical data.
- Requires minimal data processing.

Disadvantages:

- Prone to overfitting
- Sensitive to noisy data.
- Common variants:
  - ID3
  - C4.5 & C5.0
- CART - Classification & Regression Tree.

Conclusion: Thus we studied about regression tree.

QM

## Practical No. 6.

Title: Create visualization using Excel.

1. Combo charts
2. Band chart
3. Thermometer chart.
4. Gantt chart
5. Waterfall chart
6. Sparklines
7. Pivot chart.

Objective: To learn visualization tool in excel.

Software used: Microsoft Excel.

Theory: Data visualization can be done using tools like Tableau, Google charts, Data Wrapper & many more. Excel is a spreadsheet that is used for data organization & data visualization as well. In this article let's understand Data visualization in Excel.

Excel provides various type of charts like, ~~combo chart~~, Band chart, Gantt chart, Line chart & much more steps for visualizing data in Excel.

Open the Excel Spreadsheet & enter the data or select the data you want to visualize. Click on the insert tab & select the chart from the list of charts available. Here we are going to learn about few

charts that are:

1. Combo charts.

Combo charts combine two or more chart types to make the data easy to understand. Shown with a secondary axis, this chart is even easier to read. You can use combo charts when the numbers in your data vary widely from data series, or you have mixed type of data.

Band charts:

A Band chart is a line chart with added shaded areas are the Bands to display the upper & lower boundaries of the defined data ranges. The shaded areas are the Bands.

Thermometer chart:

A Thermometer chart keeps track of a single task, for example, completion of work representing the current status as compared to the target. It displays the percentage of the task completed, taking target as 100%.

Gantt chart

A Gantt chart is a chart in which a series of horizontal lines show the amount of work done in certain periods of time with relation to the amount of work planned.

for those period. The horizontal lines depict tasks, task duration & task hierarchy.

### Waterfall chart:

A waterfall chart is a form of data visualization that helps in understanding the cumulative effect of sequentially introduced positive or negative values. A typical waterfall chart is used to show how an initial value is increased/decreased by a series of intermediate values, leading to a final value.

In a waterfall chart, the columns are colour coded so that you can quickly tell positive from negative numbers. The initial & the final value column start on the horizontal axis, while the intermediate value are floating columns.

### Sparklines charts:

Sparklines are tiny charts placed in single cells each representing a row of data in your selection. They provide a quick way to see trends.

Sparklines have the following types:

- Line Sparkline
- Column Sparkline.
- Gain/Loss Sparkline.

## Pivot chart:

A pivot chart Table in Excel is a powerful tool used to summarize, analyse & present large amounts of data. It allows you to transform & analyse data sets into useful information, by summarizing & grouping data according to specific categories & variables.

Conclusion: The we studied the different types of charts.

Q  
Ans

## Practical No. 7

Aim: Create interactive visualization using any open-source tool.

Objectives: To study the open-source visualization tool KNIME for interactive visualization.

Software: KNIME

Theory: KNIME has brought the development of machine learning models in the purview of common man.

KNIME provides a graphical interface for the entire development. In KNIME, you simply have to define the workflow between the various predefined nodes provided in its repository. KNIME provides several predefined components called nodes for various tasks such as reading data, applying various machine algorithms & visualizing data in various formats.

Thus for working with KNIME, no programming knowledge is required.

KNIME Analytics platform is available for windows, Linux & Mac OS.

The KNIME views which are of immediate use to us are marked in the screenshot & listed below.

- workspace
- outline

- Nodes repository.
- KNIME Explorer
- Console
- Description

We will build your own machine learning model to categorize the plants based on few observed features. We will known into dataset from UCI machine learning Repository for this purpose.

The dataset contains three different classes of plants. We will train our model to classify an unknown plant into one of these three classes. We will start with creating a new workflow in KNIME for creating our machine learning models.

To create a new workflow, select the following menu option in the KNIME workbench.

Select the New KNIME Workflow option & click on the Next button. On the next screen, you will be asked for the desired name for the workflow & the destination folder for saving it. Enter this information as desired & click Finish to create a new workspace.

A new workspace with the given name would be added to the workspace view.

Preparing Dataset.

Download the iris dataset from the UCI

machine learning repository site Download Iris Dataset. The download iris.csv file is in CSV format. we will make some changes in it to add the column names.

Open the downloaded file in your favorite text editor & add the following line at the beginning sepall length, petal length, sepall width, petal width, class.

When our file Reader node reads this file, it will automatically take the above fields as column names. Now you will start adding various nodes. Adding File Reader Go to the Nodes Repository view, type 'File' in the search box to locate the File Reader node.

Select & double click the File Reader to add the node into the workspace.

Alternatively, you may use drag-n-drop feature to add the node into the workspace. After the node is added, you will have to configure it. Right click on the node & select the Configure menu option. You have done this in the earlier lesson.

Conclusion: Thus we studied about KNIME software tool.

Om

Practical No. 8

Aim: Create a dashboard / report using Google Data studio on YouTube channel data / Google Ad Data Search console data.

Objective: To study the visualization tool Google Data Studio & create a dashboard for google Ad Data

Tool: Google Data studio.

Theory: Google data studio is a web-based data visualization tool that helps users build customized dashboard & easy-to-understand reports. It helps in tracking key KPI's for customers visualizing trends & comparing performance over time.

Data studio, offered by google makes every bit of data useful so that accessing & sharing of reports is much simpler.

Here is how you can create a dashboard on google data studio. we will create a report for google Ad Data click on the click & open the google data studio.

You will see the templates on dashboard, click on Google Ad Overview template to

open that dataset.

You will get the view of data dashboard once you open it click on edit & view option on upper right corner.

After clicking you will get the editable version.

Add new page to create dashboard by clicking on add page button on the ribbon.

Once you add a page a blank page will appear then you can add the data you are interested in by clicking on datapane or add data option on the ribbon as shown in the picture.

or use the pre uploaded data for visualizing.

Once you add data click on the add chart option on ribbon. And add the table from drop down list choose on the screen where you want to place the table.

On right hand side of the screen you will see a pane from where you can

## DYP DPU

add & sort the data for visualizing. By clicking on dimensions we will add more feature & detail in the table. Here we will select

- Adgroup ID
- Ad Status
- Ad type.

Data along with the campaign column.

Once you add the table elements of your interest click on do add chart option or ribbon. And select the charts you want for visualizing the data.

~~Here we have taken column & pie chart for visualizing the google Ad overview.~~

Conclusion: Thus we studied about how to search the data using Google data studio.

Q  
Ans