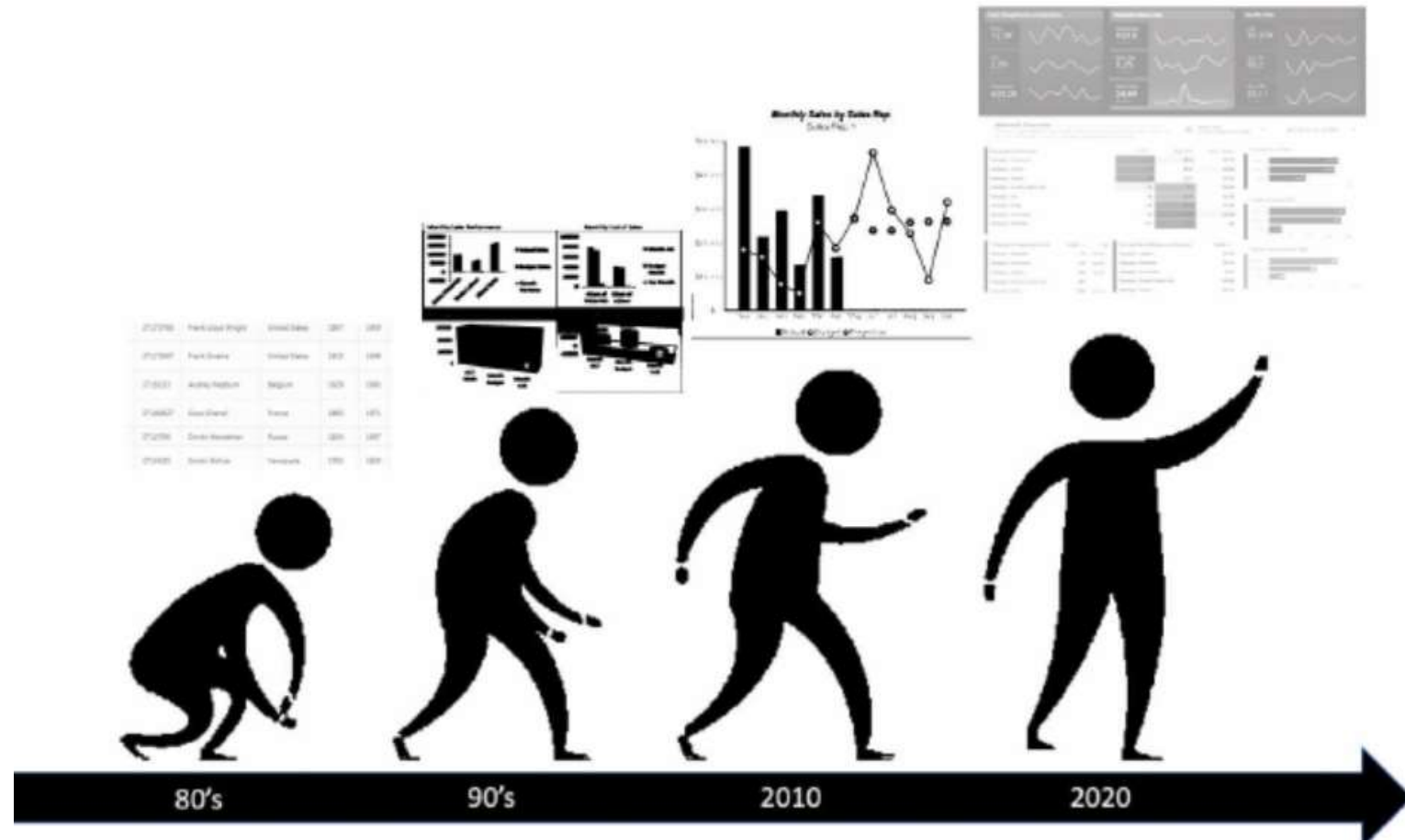# INTRODUCTION

Introduction to Data Analysis

Data Analysis Evolution

- Data analysis has evolved from basic dummy report to interactive and quipped beyond display data in graphical format.
- It has become one of the critical tools:
  - required to assist management in making best decision and the best time,
  - predicting the future outcome for every decision make and
  - propose action based on the input and the need.

Introduction to Data Analysis

- In the early 80s, data analysis is just a process of getting data, arrange and publish in tabular (also known as grid) format. Most widely used tool are Spreadsheet software like Microsoft Excel, and tool that require programming such as Crystal Report and Table using HTML tag.

- In the late 90s, enterprise start to look on detail analysis. This is the era of visualization started. Flash and various visualization library to convert the traditional method of viewing data into graphical format was used. Managers started to use bar chart and line chart to view the performance and analysis.
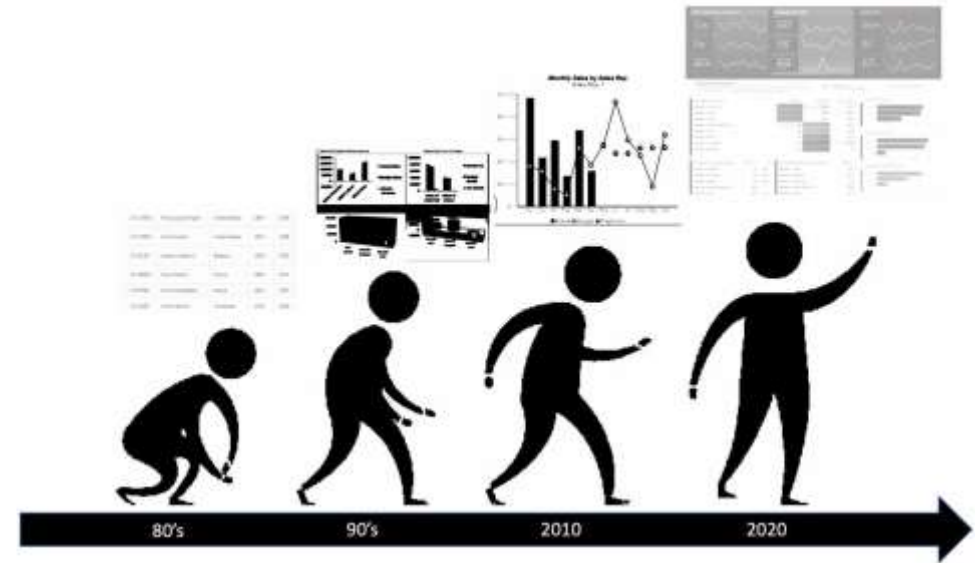
Introduction to Data Analysis

- In the new era, visualization has becoming norm and useful tool for managers in making decision.

- Yet, it is insufficient until the predictive analysis came into the picture.

- Tableau, Fusionex (Malaysian BI tool), Qlik Sense, Sisense, Microsoft Power BI, BusinessObject, Microstrategy, etc. are the common tool used for developing and presenting reports in an interactive dashboard interface. This tool too evolved and support predictive analysis which allows managers to perform what-if scenario to find the future outcome when some changeable parameters were changed.

- In the late 2010, with the processing capability increase, implementation of cloud-based infrastructure, faster processing power, and breakthrough of data analysis (big data analytics), data analysis evolved into real-time analytics.

FIGURE 2: VISUALIZATION (IMAGE SOURCE: HTTPS://WWW.SPRINGBOARD.COM/BLOG/DATA-ANALYTICSTOOLS/)

Data Analysis Definition

Data analysis has evolved from simple reports to predictive, to prescriptive and later with actionable insight, but the definition of data analysis remains at it is. Below are the few definitions found on the net.

1. **Wikipedia**

   Data analysis is a process of inspecting, cleansing, transforming and modeling data with the goal of discovering useful information, informing conclusion and supporting decision-making.
   Source (https://en.wikipedia.org/wiki/Data_analysis)

2. **Guru99**

   Data analysis is defined as a process of cleaning, transforming, and modelling data to discover useful information for business decision-making.
   Source (https://www.guru99.com/what-is-data-analysis.html)

Data Analysis Definition

3. **ORI**

   Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.
   Source (https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/datopic.html)

4. **Business Dictionary**

   Data analysis is the process of evaluating data using analytical data and logical reasoning to examine each component of the data provided.
   Source (http://www.businessdictionary.com/definition/data-analysis.html)

Data Analysis Definition

Based on many definition, it can be summarize that Data Analysis is defined as a process of **cleaning**, **transforming**, and **modelling** data to discover **useful information** for business decision-making. It is the primary component that transform data into **Actionable Insight**.

Data analysis involves asking questions about what happened, what is happening, and what will happen next.

The cycle of data analysis continues as the business needs to ensure the business persistent and resistance towards the change in the volatile economy.
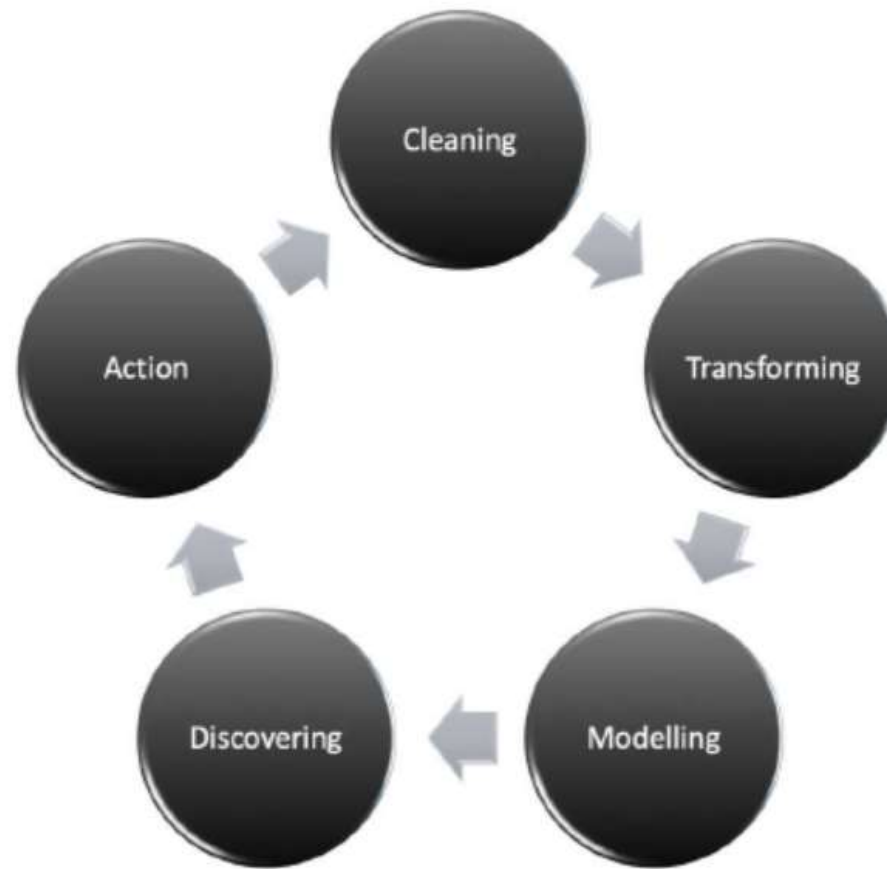

FIGURE 3: THE DEFINITION OF DATA ANALYSIS

Types of Data Analysis

There are many types of analysis; of which it can be grouped into four (4) main types of analysis. The types of analysis are based on the value of information it produces and the complexity in producing the information.
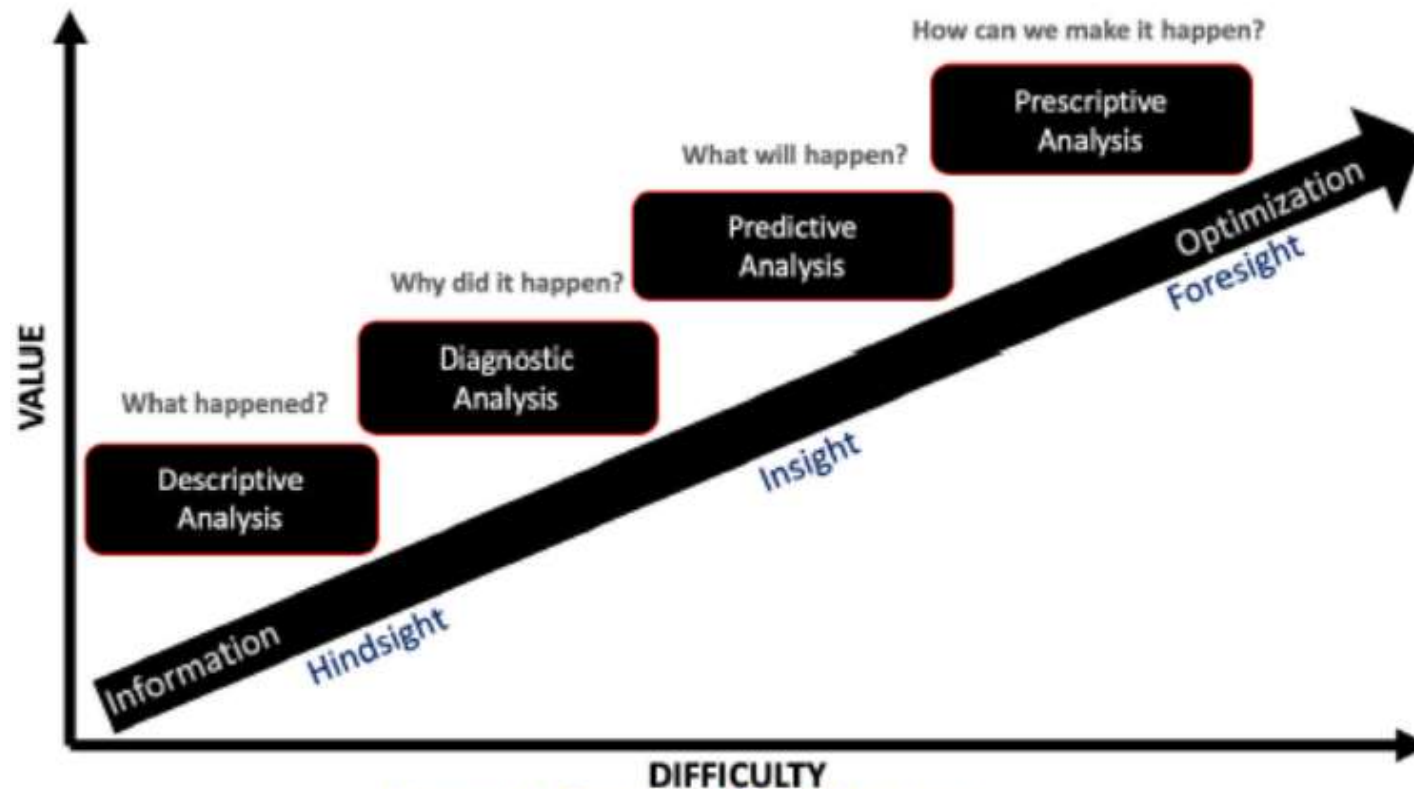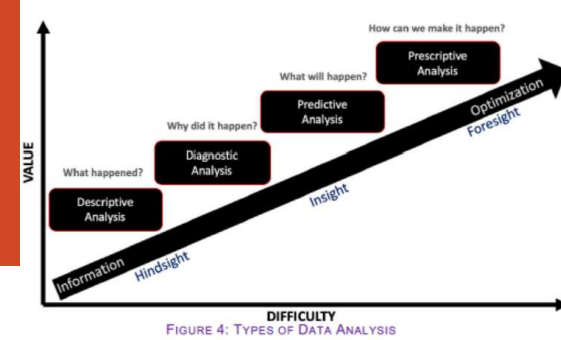

FIGURE 4: TYPES OF DATA ANALYSIS

Types of Data Analysis


FIGURE 4: TYPES OF DATA ANALYSIS

1. Descriptive Analysis

- As the word descriptive, Descriptive analysis explain briefly the data in visualization format.
- Descriptive analysis answers the "**what happened**" by summarizing past data usually in the form of dashboards.
- Descriptive Statistics are used to present quantitative descriptions in a manageable form.
- Descriptive statistics help us to simplify large amounts of data in a sensible way.
- Each descriptive statistic reduces lots of data into a simpler summary.

FIGURE 5: EXAMPLE OF DESCRIPTIVE ANALYSIS (COURTESY OF HTTPS://BLOG.HUBSPOT.COM/MARKETING/KPI-DASHBOARD)

Types of Data Analysis

FIGURE 4: TYPES OF DATA ANALYSIS
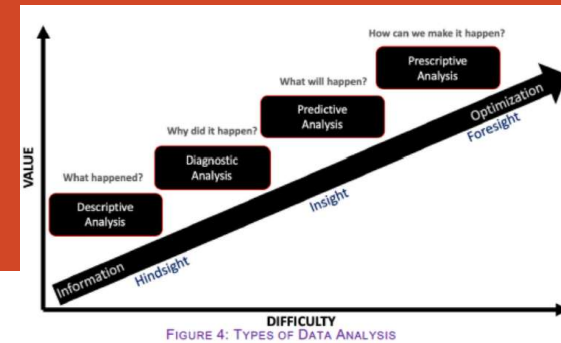
2. Diagnostic Analysis

Also known as explanatory analysis, whereby it allow user to interactively drill-down the analysis to find the root cause of it (or at least to be able to provide a gist of event causing it). It is a continuation of descriptive analysis, whereby it answer the question of '**Why did it happen**". Diagnostic analysis takes the insight found from descriptive analytics and drills down to find the cause of that outcome. Organizations make use of this type of analytics as it creates more connections between data and identifies patterns of behaviour.

A critical aspect of diagnostic analysis is creating detailed information. When new problems arise, it is possible you have already collected certain data pertaining to the issue. By already having the data at your disposal, it ends having to repeat work and makes all problems interconnected.
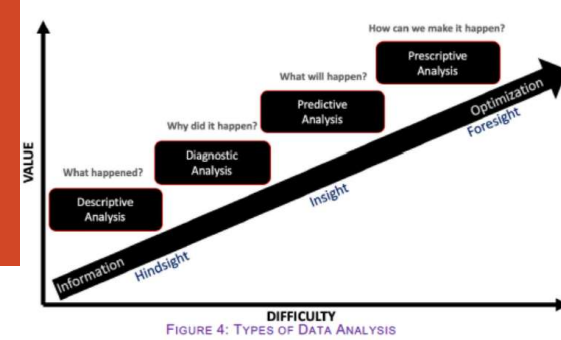
Types of Data Analysis

FIGURE 4: TYPES OF DATA ANALYSIS

- The figure is an example of diagnostic analysis dashboard.
- The dashboard will describe the current situation and point out the main issue through an indicator.
- User shall then be able to diagnose to find the root cause by drilling the pointed issue as deep as it goes depending on the design and availability of the data.
- The more experience data analyst is, the more comprehensive diagnostic analysis dashboard can be developed.
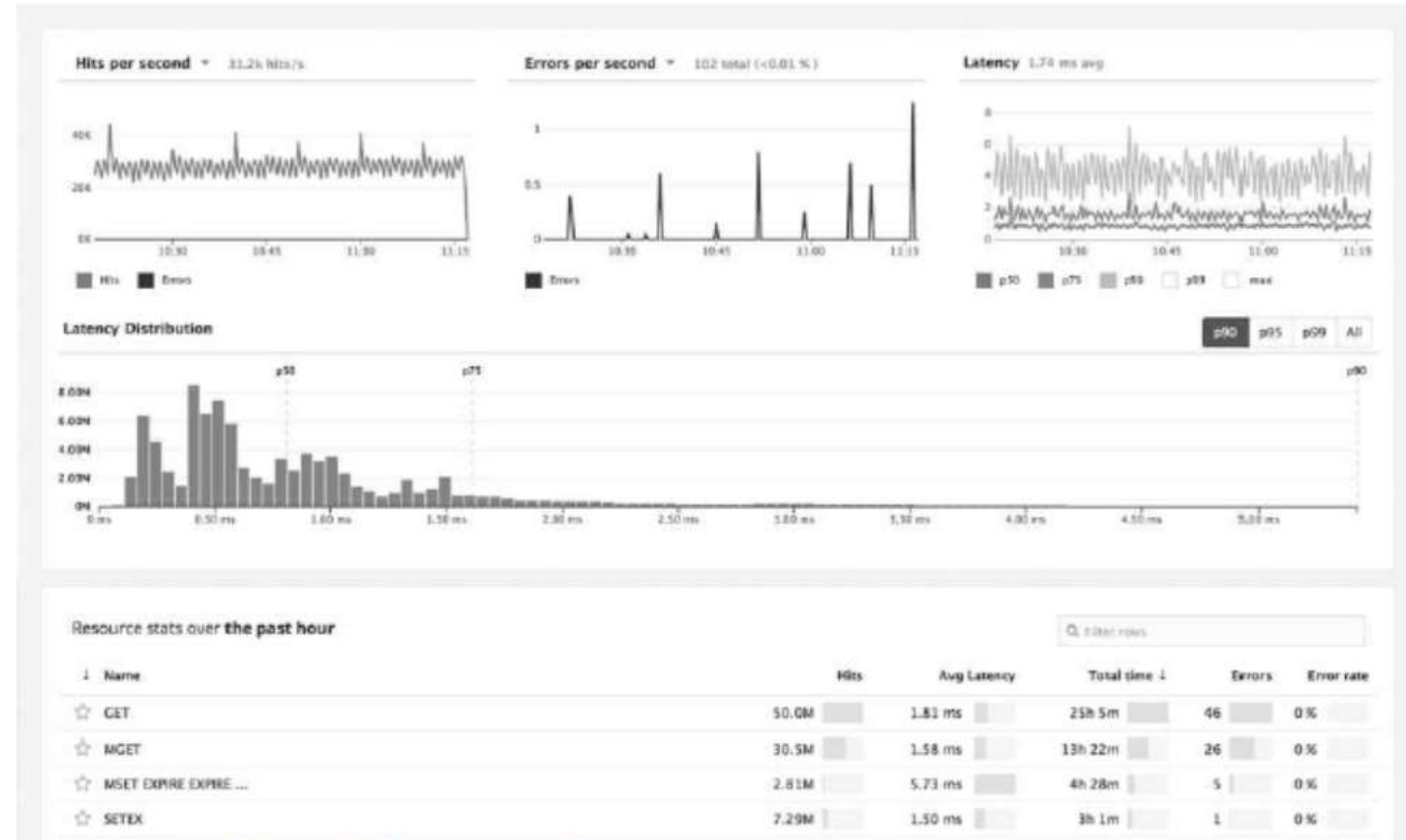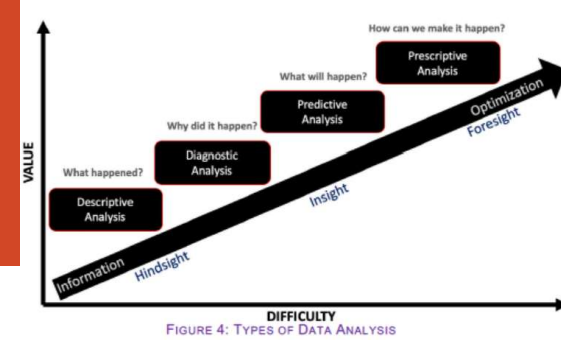

FIGURE 6: EXAMPLE OF DIAGNOSTIC ANALYSIS (COURTESY OF
HTTPS://RAYGUN.COM/BLOG/PERFORMANCE-MONITORING-TOOLS/)

Types of Data Analysis


FIGURE 4: TYPES OF DATA ANALYSIS

3. Predictive Analysis

Predictive analysis attempts to answer the question "**what is likely to happen**". Predictive analysis is also known as **Causal Analysis**. This type of analytics utilizes previous data to make predictions about future outcomes. It is extended from the descriptive and diagnostic analyses. Predictive analysis uses the summarize data to make logical predictions of the outcomes of events.

Predictive analysis relies on statistical modelling, which requires added technology and manpower to forecast. It is also important to understand that forecasting is only an estimate; the accuracy of predictions relies on quality and detailed data, the statistical models used, and the knowledge and experience of data scientist. Practical use of predictive is risk assessment, sale forecast, churn analysis, sentiment analysis, etc.

Types of Data Analysis
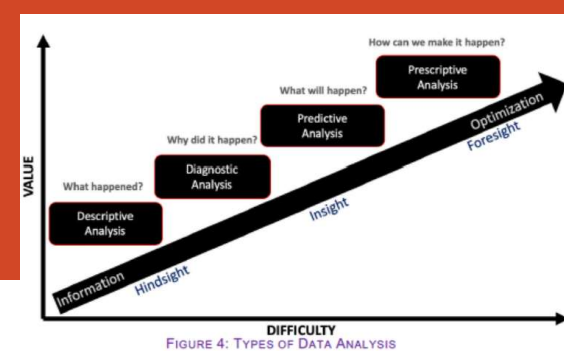
As shown in the following figure, the system predict the behaviour of an asset and potential failure. To ensure that the prediction is less error, the system required to be improve from time to time by monitoring and adjusting the model especially when an outlier is detected.
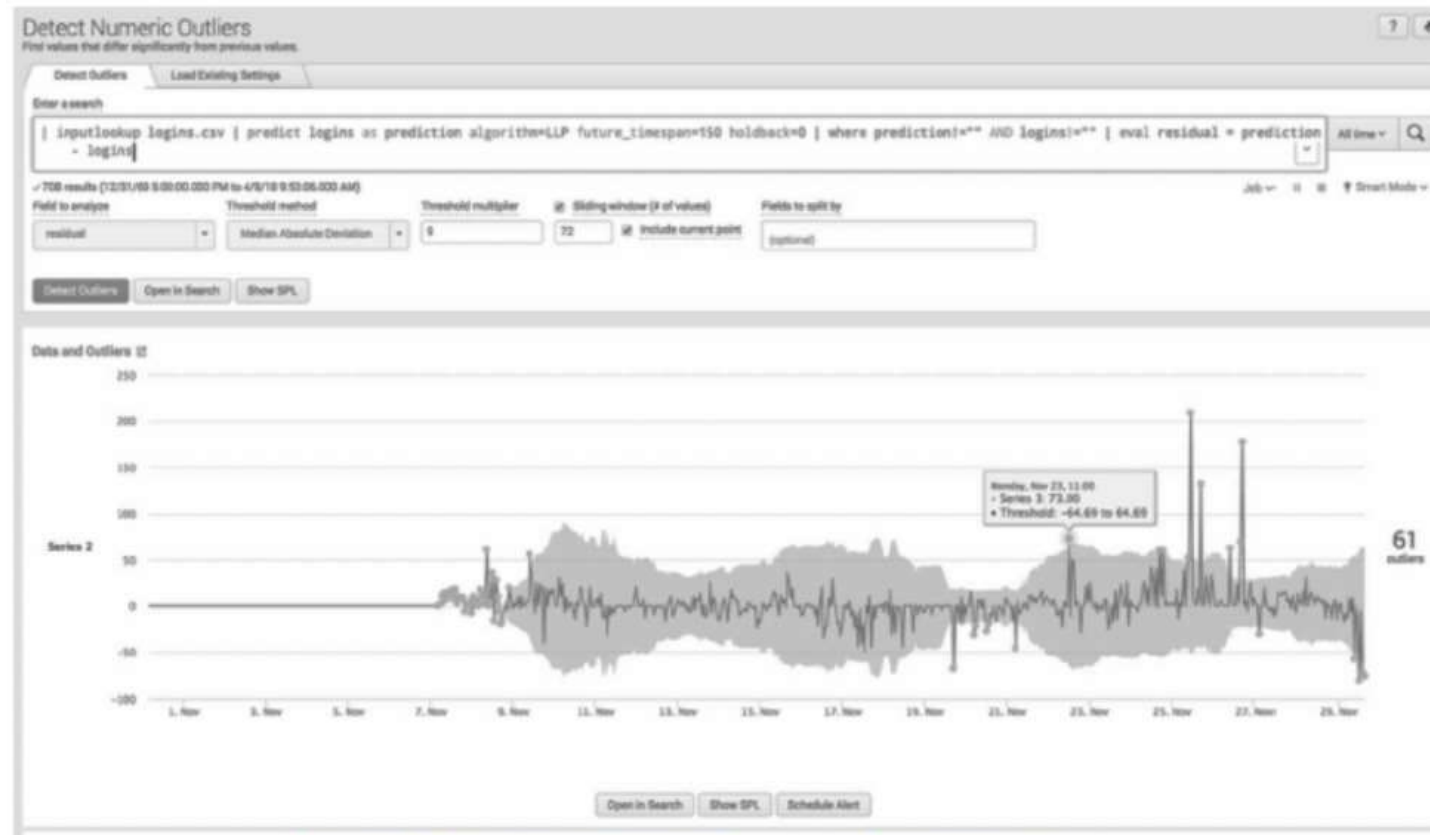

FIGURE 7: EXAMPLE OF PREDICTIVE ANALYSIS (COURTESY OF HTTPS://WWW.SPLUNK.COM/EN_US/IOT/PREVENTATIVE-MAINTENANCE-AND-PREDICTIVE-ANALYTICS.HTML)
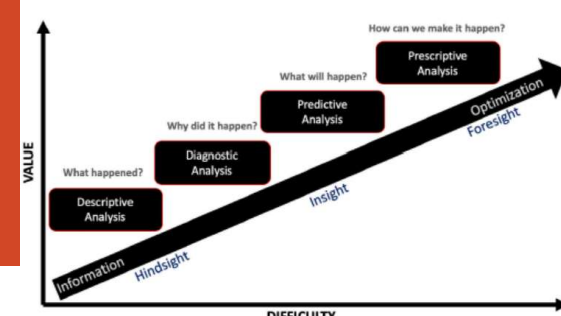
Types of Data Analysis

FIGURE 4: TYPES OF DATA ANALYSIS

4. Prescriptive Analysis

It is the frontier of data analysis, combining the insight from all previous analyses to determine the course of action to take in a current problem or decision. Recent technology allow prescriptive analysis to be implemented faster and at affordable investment. Practical use of prescriptive analysis is through Artificial Intelligent (AI), Machine Learning (ML) and/or Deep Learning (DL).

Prescriptive Analysis systems consume a large amount of data to continuously learn and use this information to make informed decisions. Well-designed Prescriptive Analysis systems are capable of communicating these decisions and even putting those decisions into action. Business processes can be performed and optimized daily without a human doing anything with artificial intelligence.

Types of Data Analysis

Figure 4: Types of Data Analysis
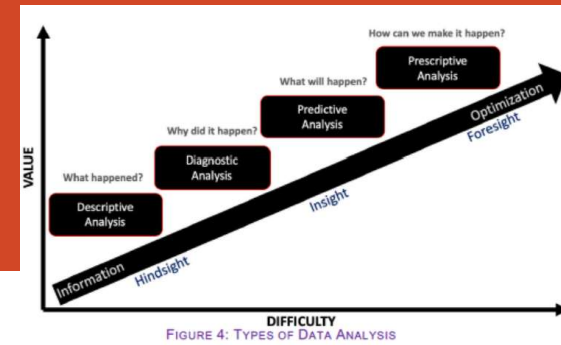
- Currently, most of the big data-driven companies (Apple, Facebook, Netflix, etc.) are utilizing prescriptive analytics to improve decision making and promoting new offering to their customers.

- It is not easy and straight-forward in implementing the prescriptive analysis.

- Any company that wishes to implement prescriptive analysis shall have a group of dedicated personnel comprises of:
    - Data Scientist,
    - Subject Matter Expert,
    - Data Engineer,
    - DBA (Doing Business As), and
    - Infra Engineer.

- A single person or small team will not able to achieve the full potential of prescriptive analysis.

Types of Data Analysis

FIGURE 4: TYPES OF DATA ANALYSIS

Following figure is an example of prescriptive analysis on LAN/WAN connection. The complexity of the analysis requires experience person to analyse and provide the outcome to management.
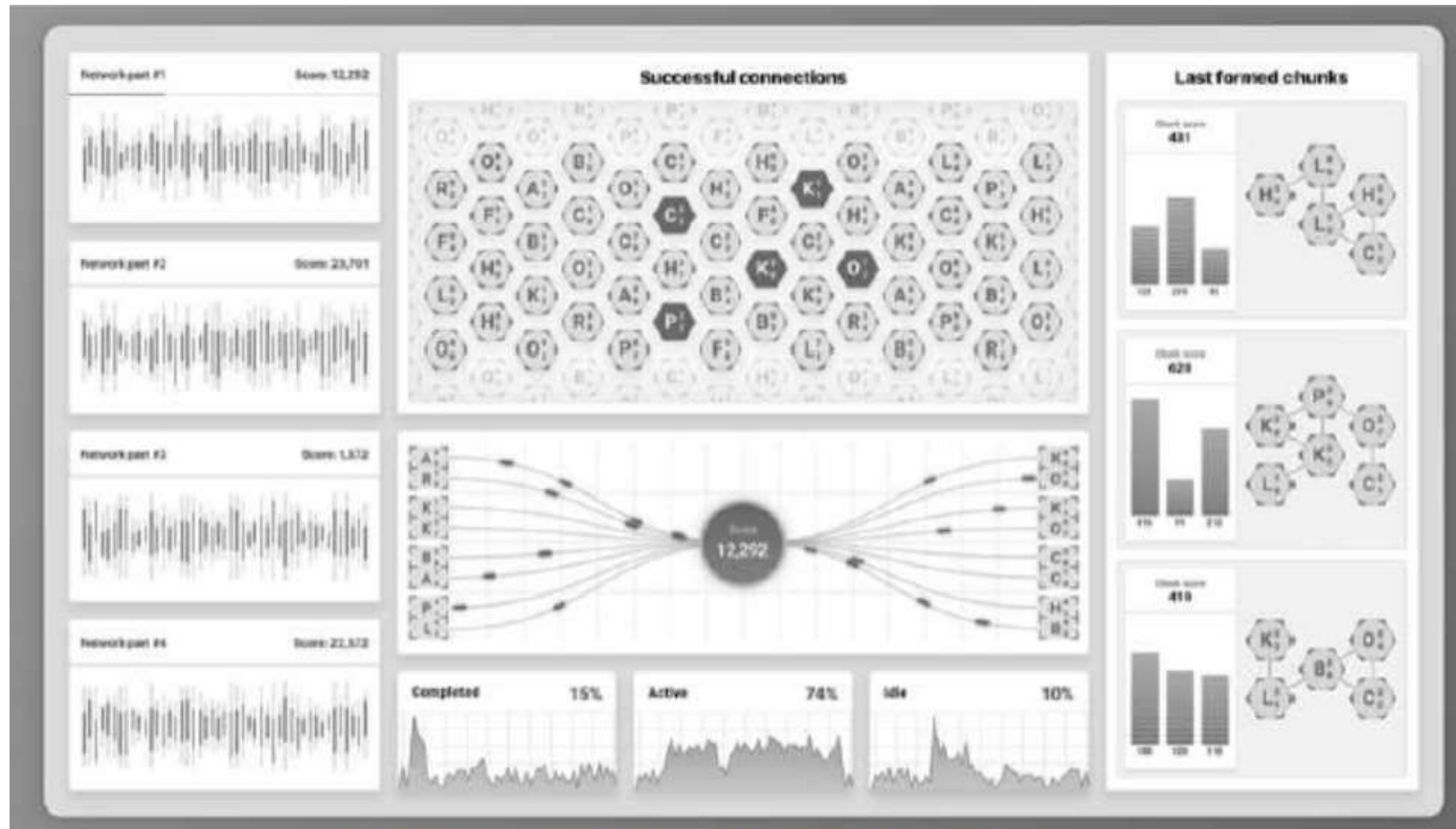
FIGURE 8: EXAMPLE OF PRESCRIPTIVE ANALYSIS

Data Visualization versus Data Analysis

- Data Visualization and Data Analysis are two data terminology that sometimes confused the data analysis people.

- These two terms seem related, interconnected but serve different purpose and for different perspective.

- From definition itself, despite data visualization and data analysis are focusing on visualizing data, the latter deal with data at an immense quantity, a much deeper level and required tools, models and algorithms for an in-depth analysis.

Data Visualization versus Data Analysis

Based on Eva Murray (published online at Forbes), data analysis is an exploratory process that often starts with specific questions. It requires curiosity, the desire to find answers and a good level of tenacity, because those answers aren't always easy to come by. It is the method of examining data sets (structured or unstructured) in order to get useful insights to draw conclusions about the datasets.

Data visualization involves the visual representation of data, ranging from single charts to comprehensive dashboards. This visual form can be a chart, graphs, lists or a map etc. This representation helps people to understand the magnitude of the data. An effective visualizations significantly reduce the amount of time it takes for your audience to process information and access valuable insights.

Data Visualization versus Data Analysis

Based on Eva Murray (published online at Forbes), data analysis is an exploratory process that often starts with specific questions. It requires curiosity, the desire to find answers and a good level of tenacity, because those answers aren't always easy to come by. It is the method of examining data sets (structured or unstructured) in order to get useful insights to draw conclusions about the datasets.

Data visualization involves the visual representation of data, ranging from single charts to comprehensive dashboards. This visual form can be a chart, graphs, lists or a map etc. This representation helps people to understand the magnitude of the data. An effective visualizations significantly reduce the amount of time it takes for your audience to process information and access valuable insights.

Data Visualization versus Data Analysis

Data Visualization versus Data Analysis	
Data Visualization	**Data Analysis**
Used For	
To communicate information clearly and efficiently to users by presenting them in visual	To assist the users to make-informed business decision by analysing the data
Relation	
Help data analytics to get better insights	The source of data visualization and make use of Data Visualization to draw conclusions about the datasets.

Data Visualization versus Data Analysis

Data Visualization versus Data Analysis	
Data Visualization	**Data Analysis**
Tools, Technique & Method	
Static / Interactive chart (bar chart, line chart, pie chart, heat map, etc) Plotly, Tableau, QlickView, etc.	Descriptive, Diagnosis, Predictive, Prescriptive (use similar chart presented in dashboard or storyboard) Hive, Ploybase, Tableau, R, Python, etc.
Industry	
Finance, Banking, Retailer, Healthcare, etc., where past history is more relevant concern for descriptive and diagnosis view.	Commercial, Crime detection, Services, etc., where future is more significant to predict and prescribe process.

Data Visualization versus Data Analysis

Data Visualization versus Data Analysis	
Data Visualization	**Data Analysis**
Who Perform	
Data Engineers	Data Analyst, Data Scientist
Platform	
Big data processing, Service Management Dashboard	Big data processing, data mining, AI/ML/DL
Benefits	
1. Identify areas need for improvement/attention 2. Clarity on situation/issue	1. Identify the model or pattern 2. Predict the needs

Data Analysis Processes

There are many processes in data analysis depending on the school of thought or preferences. However, basic and most relevant data analysis processes consist of 5 processes or activities as shown in figure.
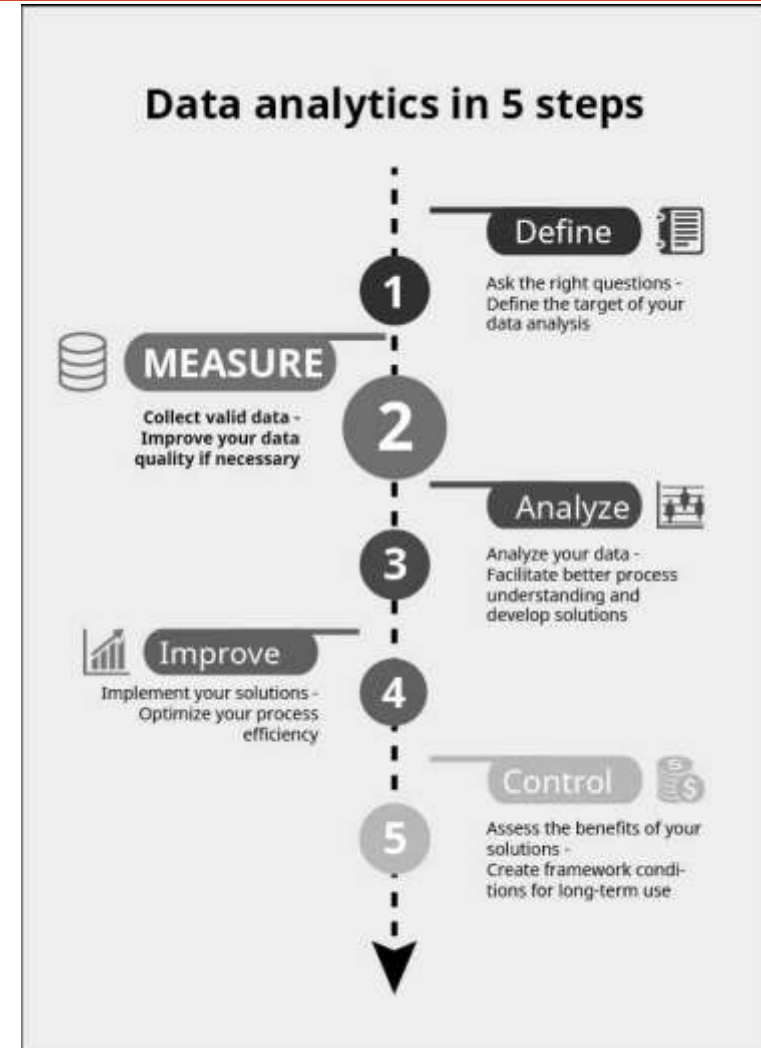
Data Analysis Processes

1. **Define**

    The processes start with defining the business objectives, output expectation, the need, data sources required and metrics or measurement to be used. This is the most critical activities as it will drive the data analysis processes and ensure the successful implementation of data analysis.

    In define stage, critical question shall be raised and answered such as

    a. What are the business objectives or goal?

    b. What are the data available, the sources and the ownership of data?

    c. What is the level of data integrity?

    d. What does the measurement need to measure the data?

    e. What is the expected outcome of the data analysis?

2. **Measure/Collect**

    It is the process of gathering, collecting, cleansing and organizing the data from sources defined earlier in the first step. This process will determine the quality of the result. Subject Matter Expert (SME), data engineer and database administrator play the important role for the success of data collection.

Data analytics in 5 steps

Define — 1
Ask the right questions - Define the target of your data analysis

MEASURE — 2
Collect valid data - Improve your data quality if necessary

Analyze — 3
Analyze your data - Facilitate better process understanding and develop solutions

Improve — 4
Implement your solutions - Optimize your process efficiency

Control — 5
Assess the benefits of your solutions - Create framework conditions for long-term use

Data Analysis Processes

3. Analyze

Once the data is properly organize and gathered, the data analyst, data scientist and/or data statistician shall start to analyze and design the right visual that shall help to provide the answer to the questions raised before and/or giving the viewer the ability to see things happen in the past, the cause and what can happen in the future.

4. Improve/Visualization

Once the data analyst defines the visual, data engineer will develop the visualization and improve it together with data scientist and SME, to ensure the visual used is able to provide the viewer the answers needed based on question derived in the first process.

5. Control

Finally, the processes are completed and frequent review on the data and visual is required to control the integrity of the data. It is also the process whereby the user interprets the visual or result of the data analysis and make necessary decision based on the result or recommendation.

Data analytics in 5 steps

1 Define
Ask the right questions - Define the target of your data analysis

MEASURE

2 Collect valid data - Improve your data quality if necessary

3 Analyze
Analyze your data - Facilitate better process understanding and develop solutions

4 Improve
Implement your solutions - Optimize your process efficiency

5 Control
Assess the benefits of your solutions - Create framework conditions for long-term use

Data Analysis Tool

- Depending on needs, data analysis tool may vary for each individual and companies' requirements.
- It also depends on the willingness to spend and the available key personnel in doing the analysis.
- Some may use 1 tool while others may implement multiple tools. Below are some of the tools which can be used individually or together.
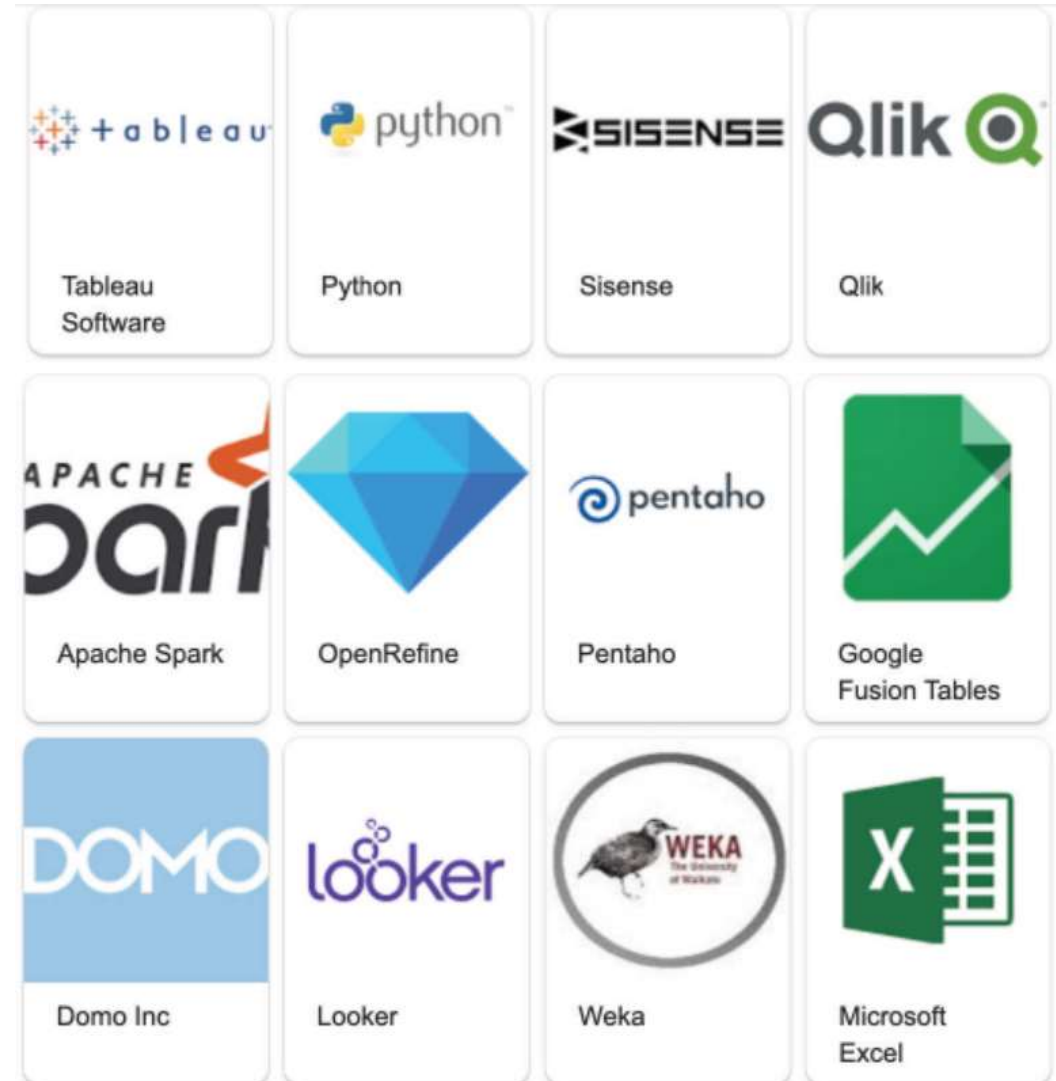

FIGURE 10: EXAMPLE OF DATA ANALYSIS TOOL (COURTESY OF GOOGLE)

Data Visualization Tool

- Data visualization tool on the other end is actually the object to represent/visualize data.
- Example of the tool is grid, bar chart, pie chart, etc.
- Used of data visualization tool is purely dependent on creativity of the data engineer and capability of data analyst to convert the raw and unreadable data into presentable visual format.
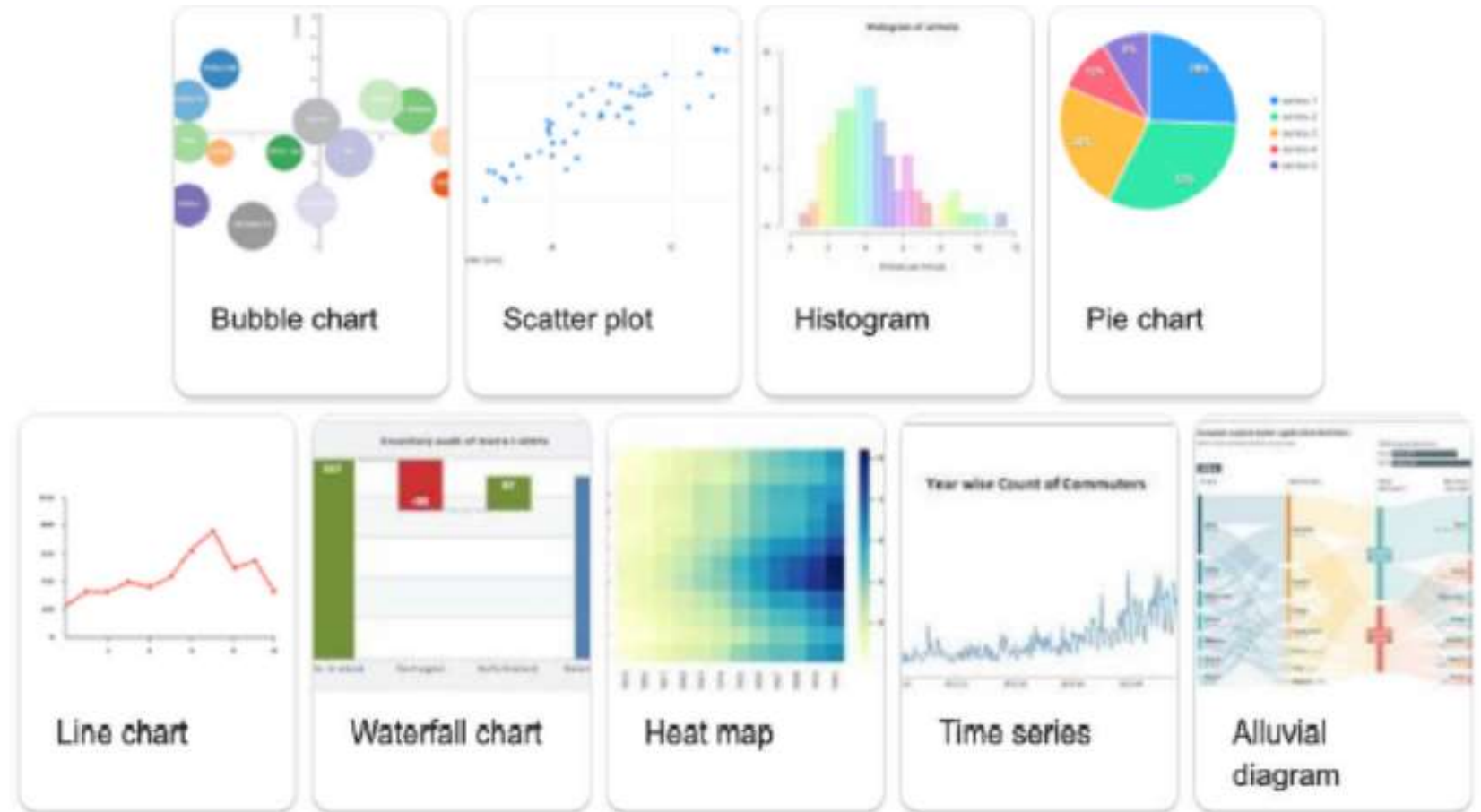- However, there are some guidelines on which visual is suitable to represent an information.

FIGURE 11: EXAMPLE OF DATA VISUALIZATION TOOL

Data Visualization Tool

- Jānis Gulbis provide a very comprehensive explanation on the usability of each type of data visualization and accessible at https://eazybi.com/blog/data_visualization_and_chart_types/.
- However, Tableau has published a simpler data visualization guidelines and is accessible freely at https://www.tableau.com/learn/whitepapers/which-chart-or-graph-is-right-for-you.

1. Bar Chart

Bar charts is use to compare data across categories, highlight differences, show trends and outliers, and reveal historical highs and lows at a glance. Bar charts are effective when you have data that can be split into multiple categories.

2. Line Chart

The line chart, or line graph, connects several distinct data points, presenting them as one continuous evolution. Use line charts to view trends in data, usually over time and financial related like stock price, number of ridership, etc.

3. Pie Chart

Pie charts are powerful for adding detail to other visualizations. Alone, a pie chart doesn't give the viewer a way to quickly and accurately compare information but will provide a gist of important information. Pie chart is best to use when number of categories or element of comparison is less than five (5).

4. Maps

Maps is suitable when you have the location information and wish to show the location correlation with your data. However, the more scatter your location are, the more difficult to use maps. You need to categories the location into country, state, city, and only zoom into the exact address when it is more visualizable.

5. Density Maps

Density maps reveal patterns or relative concentrations that might otherwise be hidden due to an overlapping mark on a map—helping you identify locations with greater or fewer numbers of data points. Density maps are most effective when working with a data set containing many data points in a small geographic area.

Data Visualization Tool

6. Scatter Plot

Scatter plots are an effective way to investigate the relationship between different variables, showing if one variable is a good predictor of another, or if they tend to change independently. A scatter plot presents lots of distinct data points on a single chart. The chart can then be enhanced with analytics like cluster analysis or trend lines.

7. Gantt Chart

Gantt charts display a project schedule or show changes in activity over time. A Gantt chart shows steps that need to be completed before others can begin, along with resource allocation.

8. Bubble Chart

Although bubbles aren't technically their own type of visualization, using them as a technique adds detail to scatter plots or maps to show the relationship between three or more measures. Varying the size and colour of circles creates visually compelling charts that present large volumes of data at once.

9. Treemap

Treemaps relate different segments of your data to the whole. As the name of the chart suggests, each rectangle in a treemap is subdivided into smaller rectangles, or sub-branches, based on its proportion to the whole. They make efficient use of space to show percent total for each category.

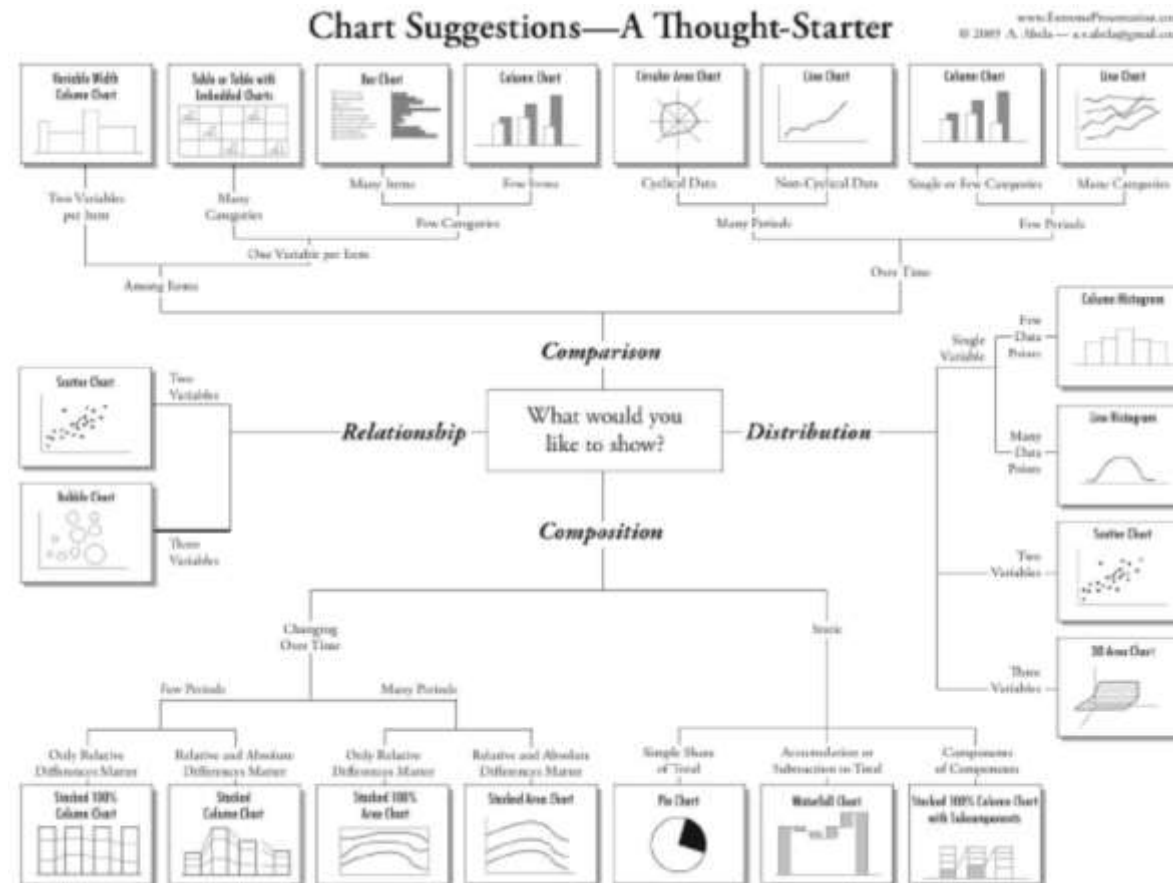# Data Visualization Tool

How to Choose the Right Chart for Your Data (infogram.com)

FIGURE 12: A PRACTICAL USE CASE FOR EACH CHART/VISUAL