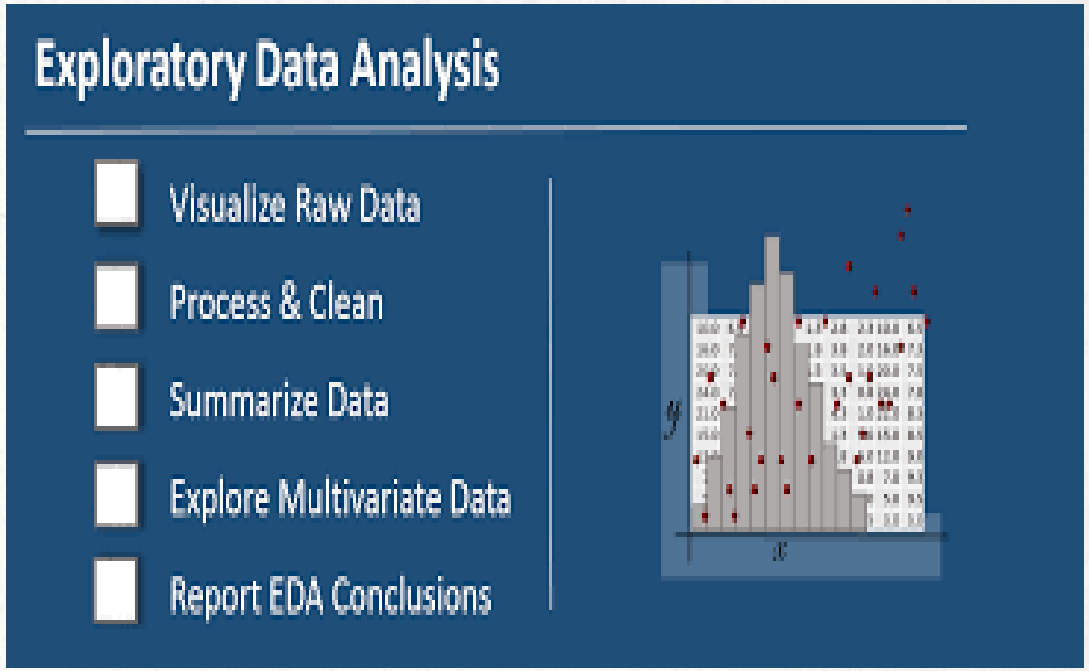
PRESENTATION

BY KRISHNENDU R

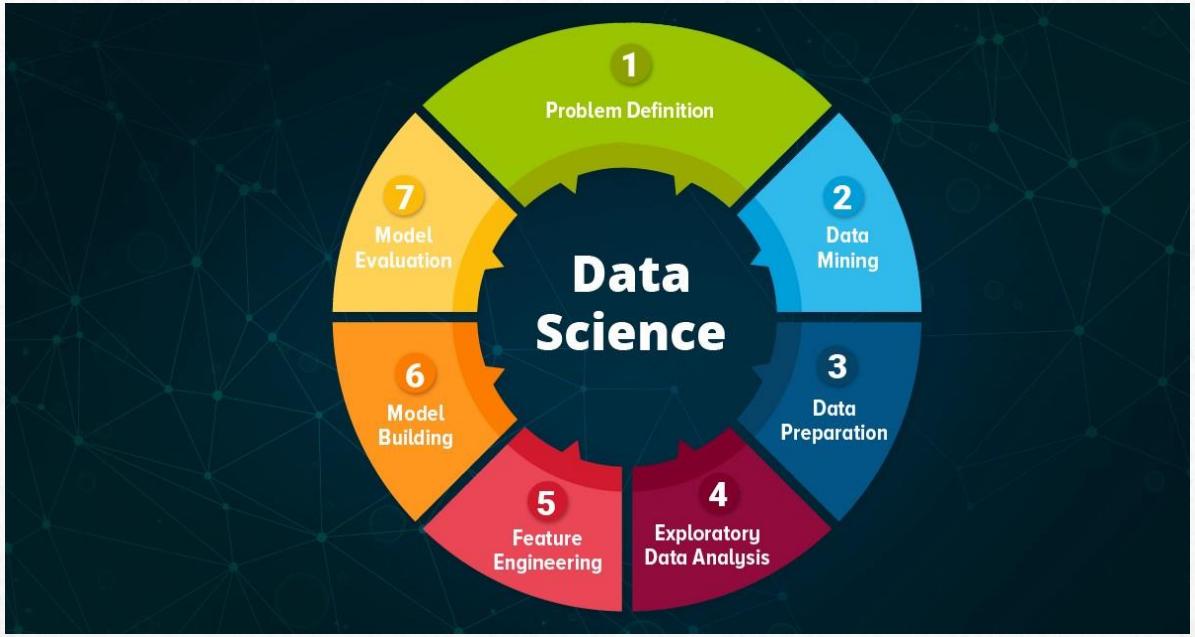
EXPLORATORY DATA ANALYTICS AND EXPLANATORY DATA ANALYTICS

1. EXPLORATORY DATA ANALYTICS

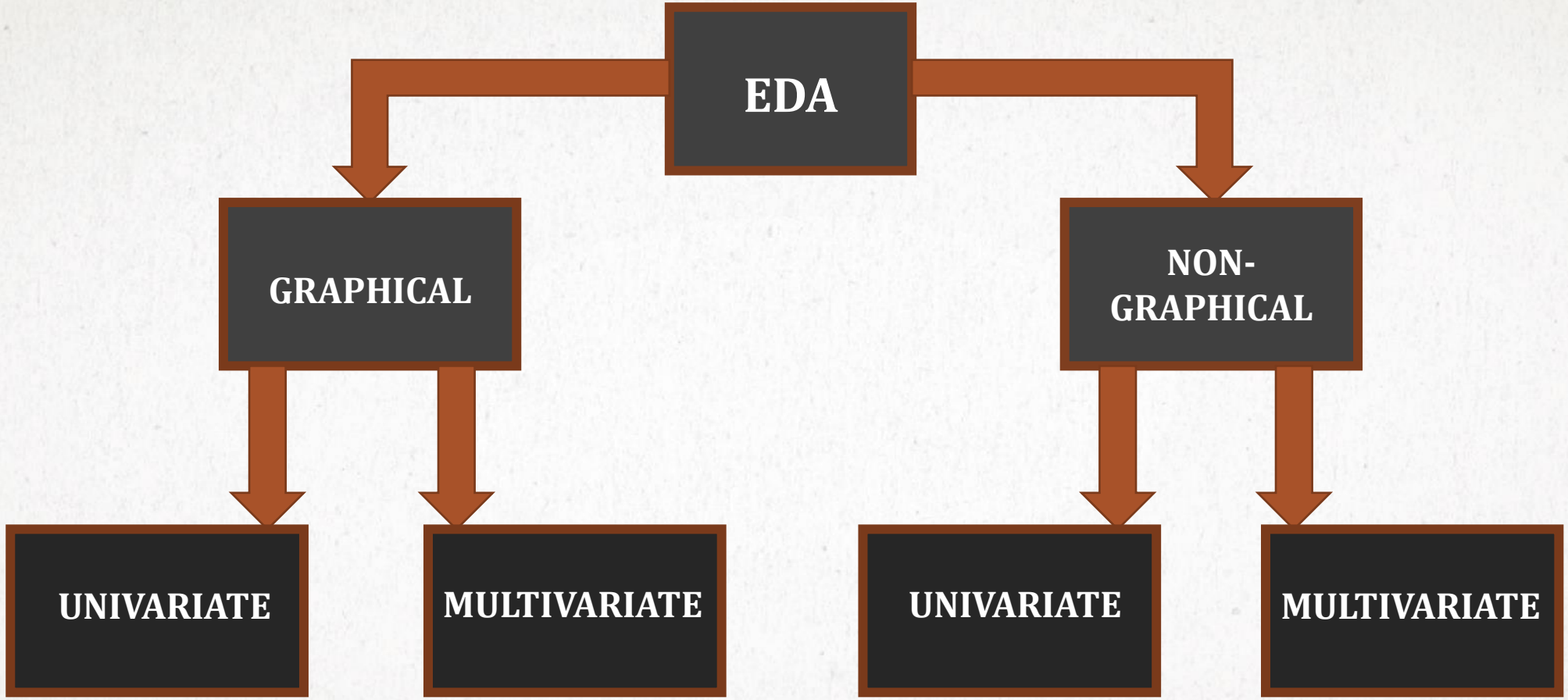
- Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.
- It helps determine how best to manipulate data sources to get the answers you need, making it easier for data scientists to discover patterns, spot anomalies, test a hypothesis, or check assumptions.



- EDA is primarily used to see what data can reveal beyond the formal modeling or hypothesis testing task and provides a provides a better understanding of data set variables and the relationships between them.
- It can also help determine if the statistical techniques you are considering for data analysis are appropriate.
- Originally developed by American mathematician John Tukey in the 1970s, EDA techniques continue to be a widely used method in the data discovery process today.



TYPES OF EDA


EDA
GRAPHICAL
NON-
GRAPHICAL
UNIVARIATE
MULTIVARIATE
UNIVARIATE
MULTIVARIATE

CONT...

There are four primary types of EDA:

- **Univariate non-graphical.** This is simplest form of data analysis, where the data being analyzed consists of just one variable. Since it's a single variable, it doesn't deal with causes or relationships. The main purpose of univariate analysis is to describe the data and find patterns that exist within it.
- **Univariate graphical.** Non-graphical methods don't provide a full picture of the data. Graphical methods are therefore required. Common types of univariate graphics include:
  - Stem-and-leaf plots, which show all data values and the shape of the distribution.
  - Histograms, a bar plot in which each bar represents the frequency (count) or proportion (count/total count) of cases for a range of values.
  - Box plots, which graphically depict the five-number summary of minimum, first quartile, median, third quartile, and maximum.

CONT...

- **Multivariate nongraphical:** Multivariate data arises from more than one variable. Multivariate non-graphical EDA techniques generally show the relationship between two or more variables of the data through cross-tabulation or statistics.
- **Multivariate graphical:** Multivariate data uses graphics to display relationships between two or more sets of data. The most used graphic is a grouped bar plot or bar chart with each group representing one level of one of the variables and each bar within a group representing the levels of the other variable.

EXPLORATORY DATA ANALYSIS TOOLS

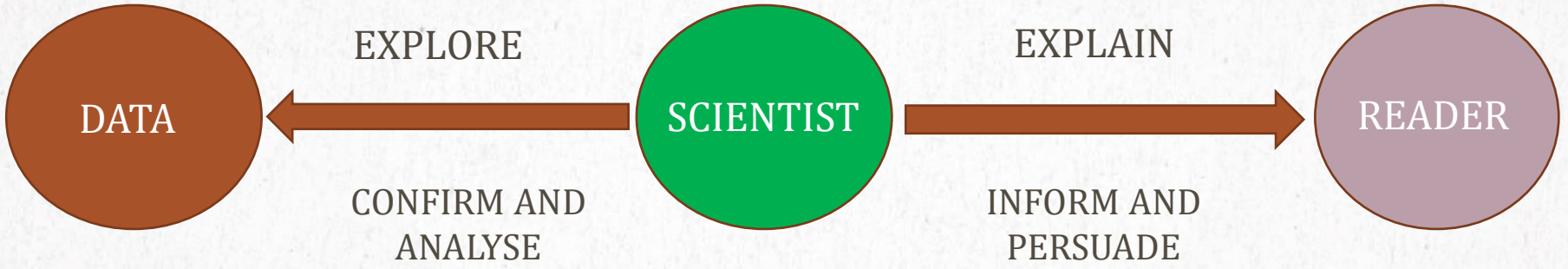
Some of the most common data science tools used to create an EDA include:

- **Python:** An interpreted, object-oriented programming language with dynamic semantics Python and EDA can be used together to identify missing values in a data set, which is important so you can decide how to handle missing values for machine learning.
- **R:** An open-source programming language and free software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians in data science in developing statistical observations and data analysis.

2. EXPLANATORY DATA ANALYTICS

- Explanatory analysis is the step beyond exploratory.
- Instead of explaining what happened, you're more focused on how and why it happened and what should happen next and, in most cases, communicating that to the necessary decision-makers and stakeholders.
- Exploratory Data Analysis is an important step before starting to analyze or modeling of the data.
- It provides the context needed to develop an appropriate model and interpret the results correctly

CONT...

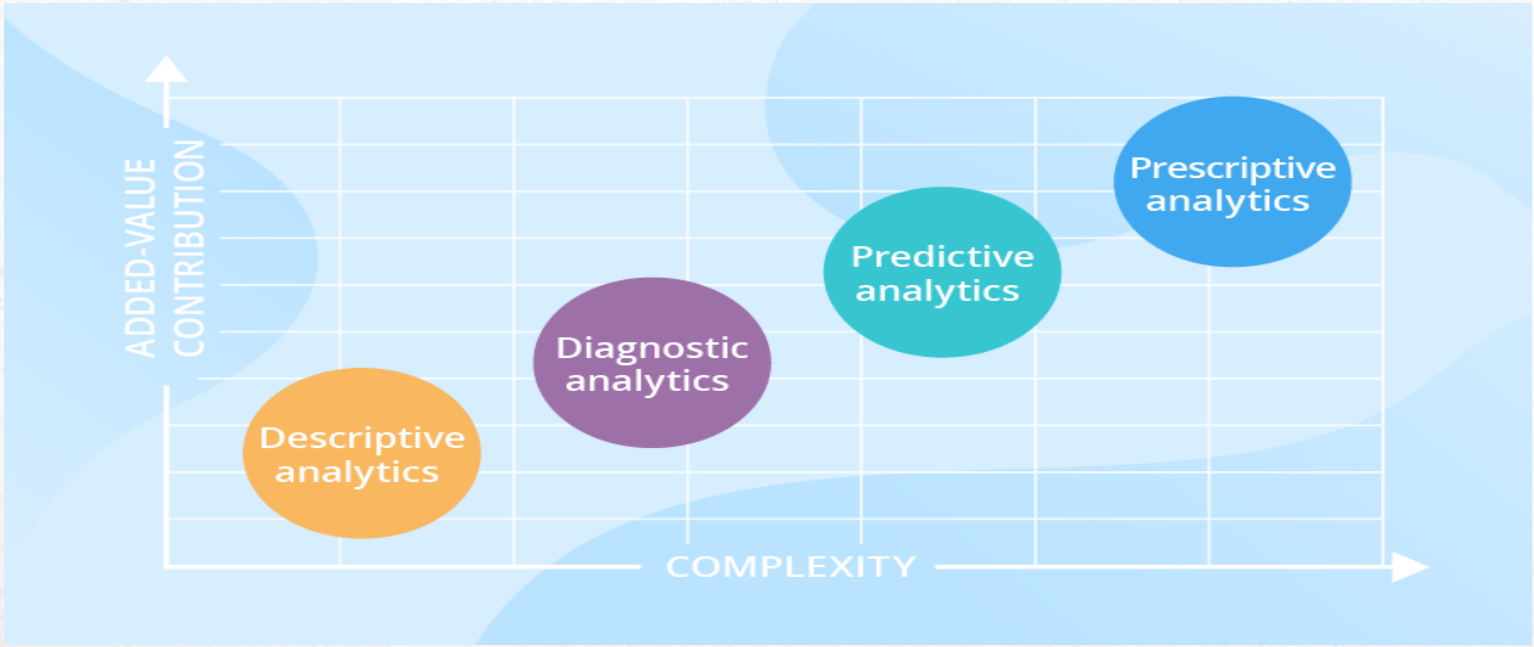

DATA
EXPLORE
CONFIRM AND ANALYSE
SCIENTIST
EXPLAIN
INFORM AND PERSUADE
READER

DATA ANALYSIS

- **Data analysis** is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.
- In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

TYPES OF DATA ANALYSIS

Data can be used to answer questions and support decisions in several different ways. It can help to group these types of analysis into four categories commonly used in the field.



1. Descriptive analysis

- Descriptive analysis tells us what happened. This type of analysis helps describe or summarize quantitative data by presenting statistics.
- For example, statistical analysis could show the distribution of sales across a group of employees and the average sales figure per employee.
- Descriptive analysis answers the question, “what happened?”

CONT…

2. Diagnostic analysis

- If the descriptive analysis determines the “what,” diagnostic analysis determines the “why.”
- Let’s say a descriptive analysis shows an unusual influx of patients in a hospital.
- Drilling into the data further might reveal that many of these patients shared symptoms of a particular virus.
- This diagnostic analysis can help you determine that an infectious agent — the “why” — that led to the influx of patients.
- Diagnostic analysis answers the question, “why did it happen?”

CONT…

3. Predictive analysis

- So far, we've looked at types of analysis that examine and draw conclusions about the past.
- Predictive analytics uses data to form projections about the future.
- Using predictive analysis, you might notice that a given product has had its best sales during the months of September and October each year, leading you to predict a similar high point during the upcoming year.
- Predictive analysis answers the question, "what might happen in the future?"

CONT…

4. Prescriptive analysis

- Prescriptive analysis takes all the insights gathered from the first three types of analysis and uses them to form recommendations for how a company should act.
- Using our previous example, this type of analysis might suggest a market plan to build on the success of the high sales months and harness new growth opportunities in the slower months.
- Prescriptive analysis answers the question, “what should we do about it?”
- This last type is where the concept of data-driven decision-making comes into play

THANK YOU