PIM Training Program

Data Analysis - Getting Started

Data Analysis/Analytics Goal & Approach

This session is on **Getting started with Data Analysis**.

Data Analysis – Getting Started

1. Define Analysis Goal
2. Define Analysis Approach

Session Details

Session Facilitator: Shriram Aravamudhan (L5 Operations Manager)

Session Duration: 4 hours

Objectives of Training

Data Analysis helps the business to understand problems facing an organization, and explore data in meaningful ways. As a PIM you will collect data, organize, interpret, structure, and present the data into useful information to provide business context. This session will introduce you to the concept of Data Analysis. At the end of this session, you should be able to:

- Explain the need for Data Analysis
- List down the types of Data Analysis
- Explain the application of Data Science in Business Context
- Create goals for Data Analysis

Need for Data Analysis

Data Analysis is required in business for the following reasons.

- Understand business direction and objectives
- Explore the meaning behind the numbers and figures in data
- Analyze the causes of certain events based on data findings
- Contribute to business decision-making by offering educated opinions

Why Data Science?

Data Science: Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data.

Three significant events triggered the current meteoric growth in the use of data analytics in decision making and ***Statistics is Heart of Analytics***

Event 1

- *Large Amount of Data resulting from Technological developments, Revolution of Internet, social networks, and from mobile phones requiring Analytics to derive insights*
- *The discovery of pattern and trends from these data can be leveraged by organizations will pave the way for improving profitability, and performance.*

Event 2

- *Advances in enormous computing power to effectively process and analyze massive amounts of data*
- *Sophisticated and faster algorithms for solving problems*
- *Data Visualization for Business Intelligence*

Event 3

- *Large data storage capability*
- *Parallel computing, and cloud computing coupled with better computer hardware have enabled businesses to solve large scale problems faster than ever before*

An Example

I think customer experience in my category are mostly negative off late and I need to fix it. Can we come up with a solution?

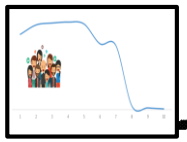
Ok! Let me,

1. Define the problem
2. Get relevant data
3. Draw Hypothesis
4. Size the problem
5. Get it vetted!
6. Do RCA
7. Get it vetted!
8. Design Solution as required

An Example

Alright! Here are my findings from the initial Data Analysis, Of all the customers who have written a review in the last 1 year, 78% of the customer have positive sentiment on the products from this category, and 22% have had negative experience,

And yes, you were right, 14% of 22% negative experience have come in the last 3 months, so seems to be a problem creeping up recently!! Here is a graph…

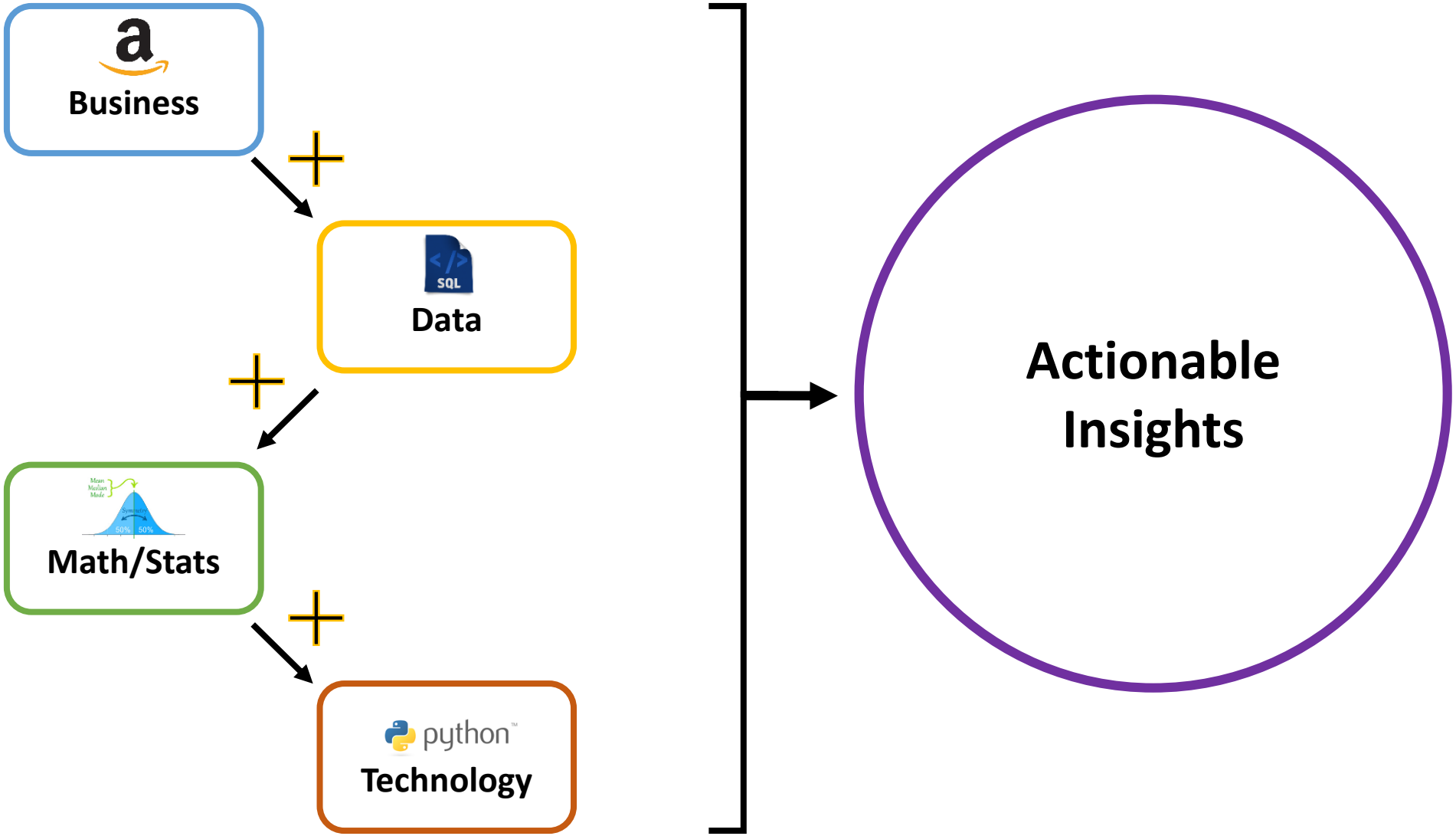


Cool! how did you go about doing this?

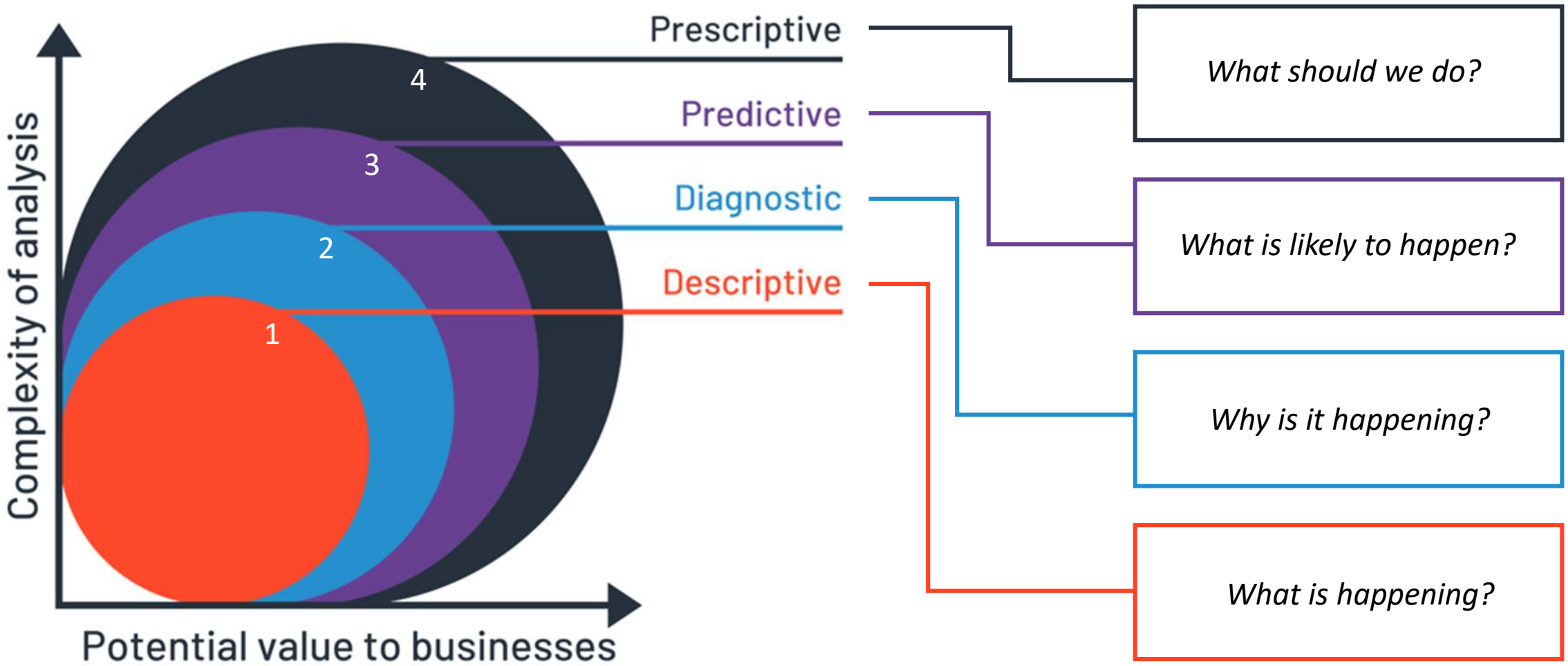
That's what we will learn in this session

Data Analysis

Data Analysis Structure



Types of Data Analysis



Descriptive Analysis

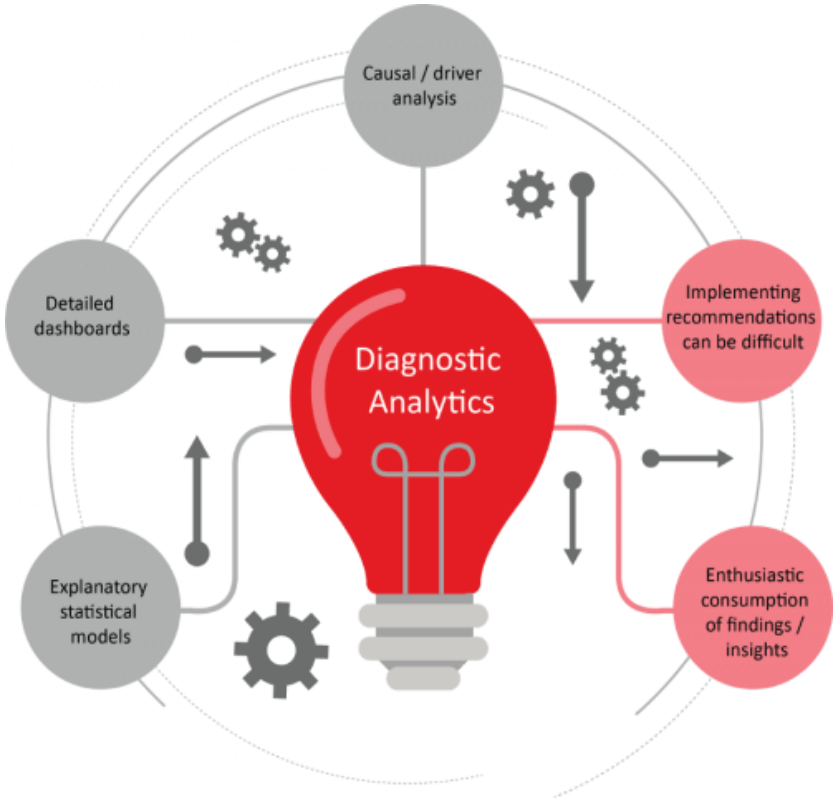
Descriptive analysis gives an idea of the distribution of your data, helps you detect outliers and typos, and enable you identify associations among variables, thus preparing you for conducting further analysis.

The analysis helps you to witness **'What's happening in my business?'** and enhance effective visualization.

Diagnostic Analysis

Diagnostic analysis is a form of advanced analytics which examines data or content to understand the reason behind an incident or problem. It is characterized by techniques such as drill-down, data discovery, data mining and correlations.

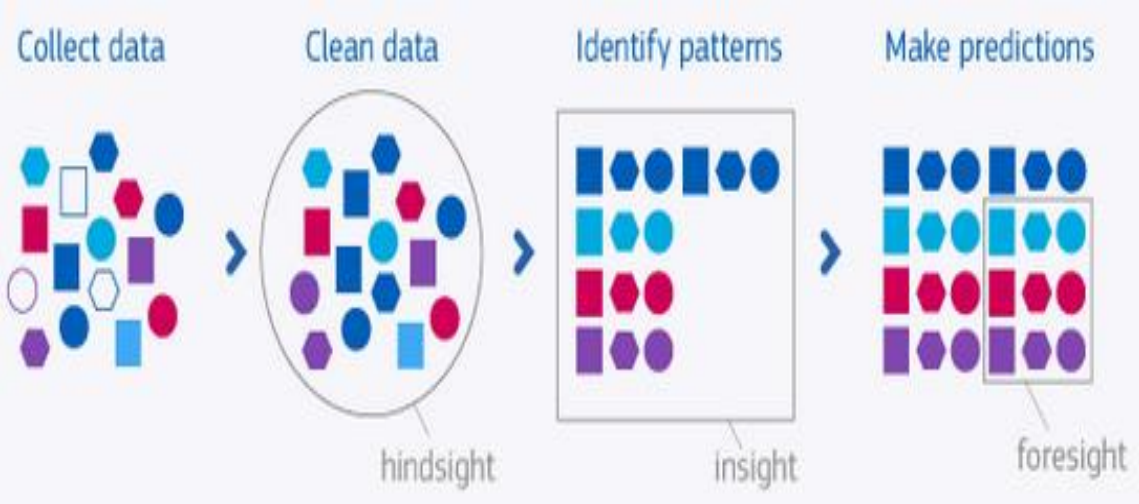
The analysis helps you to witness **'Why did it happen in my business?'**.



Predictive Analysis

Predictive analysis is the branch of the advanced analytics which is used to make predictions about unknown future events. Predictive analytics uses techniques from data mining, statistics, modeling, machine learning (ML), and artificial intelligence to analyze current data to make predictions about future.

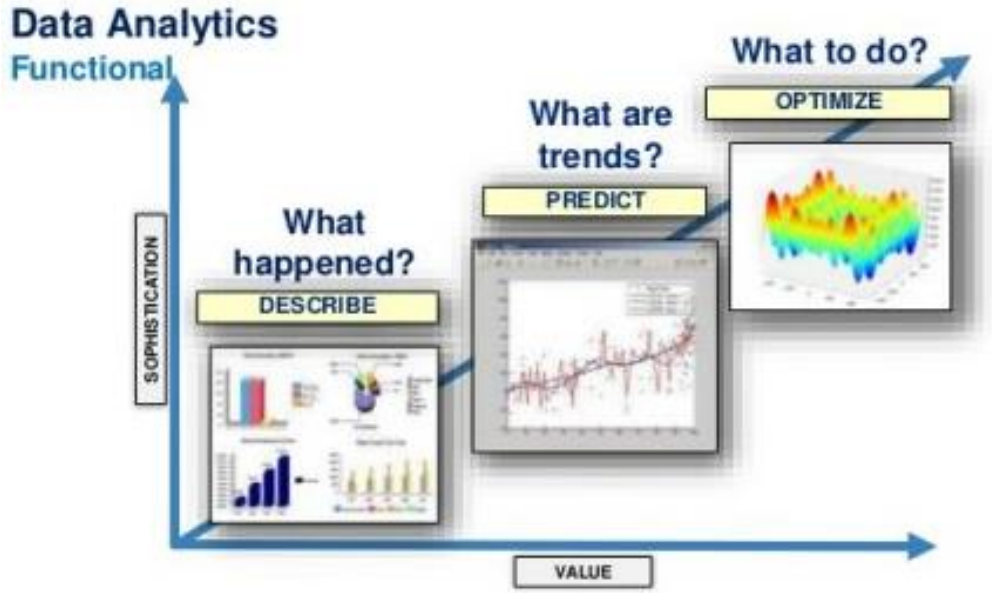
The analysis helps you to witness **'What is likely to happen in the future in my business?'**.



Prescriptive Analysis

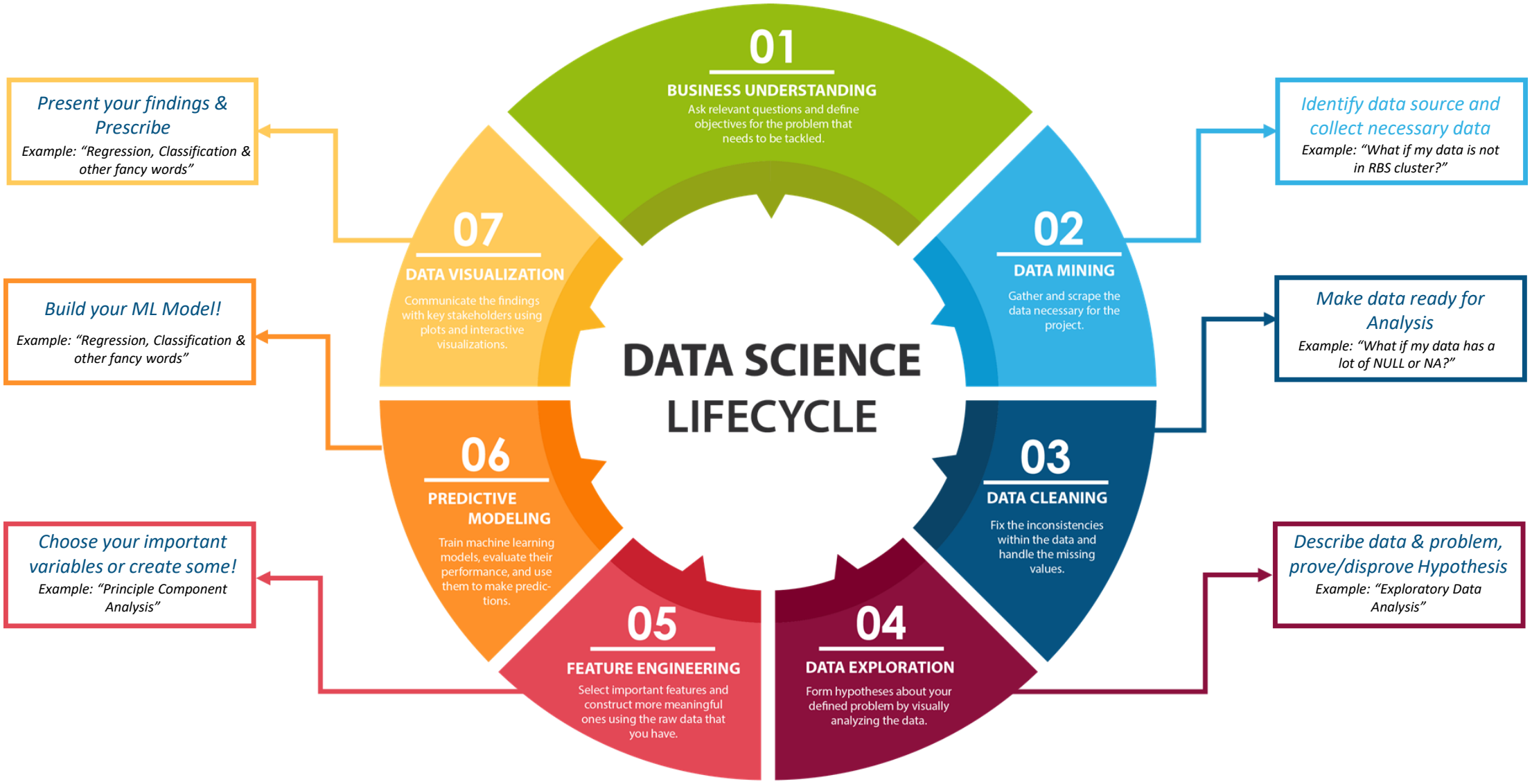
Prescriptive analysis is a type of data analytics which uses technology to help businesses make better decisions through the analysis of raw data. It factors information about possible situations or scenario's, available resources, past & current performance, and suggests a course of action or strategy.

The analysis helps you to witness **'What should my business do?'**.



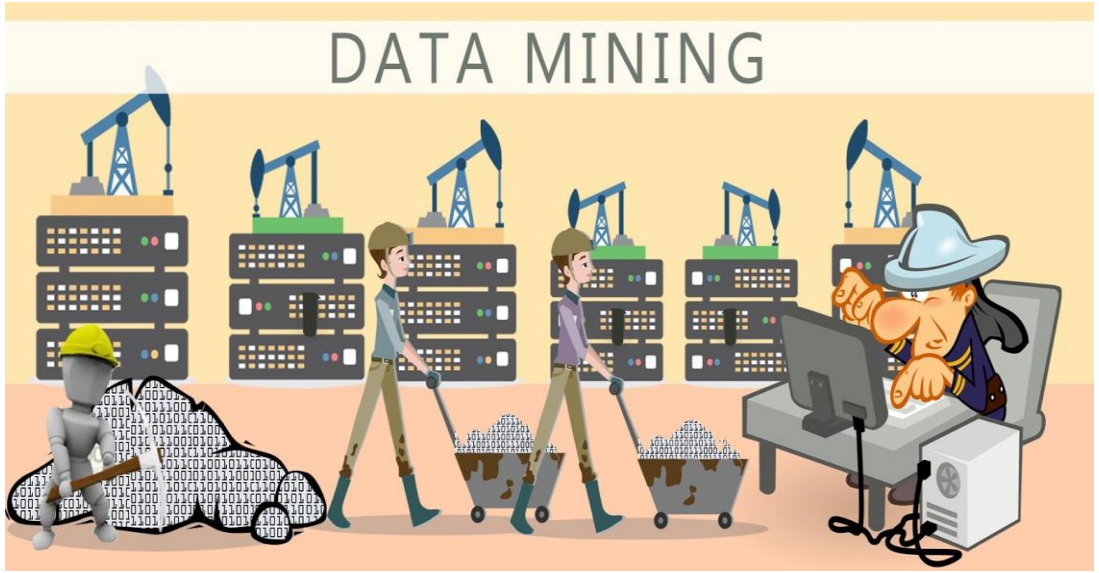
Data Science

Data Science Lifecycle



Data Mining

Data Mining is the practice of examining large pre-existing databases in order to generate new information.



Data Cleaning

Data cleaning is the process of detecting and correcting corrupt or inaccurate records from a record set, table, or database and it refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

Data Exploration

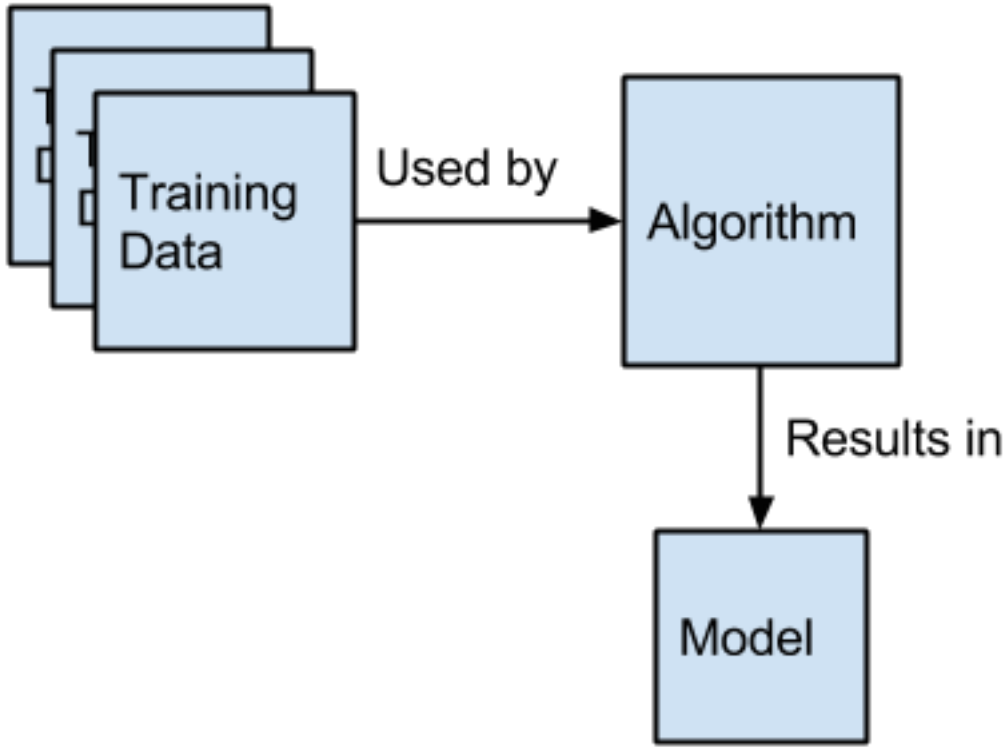
Data Exploration involves summarizing the main characteristics of a data set, uncovering the relationships between the different variables, and simply view most of the relevant features of a data set.

Feature Engineering

Feature engineering is the process of using domain knowledge of the data to create features that make Machine Learning (ML) algorithms work. Feature engineering is fundamental to the application of machine learning.

Predictive Modeling

Predictive Modeling is a process that uses data mining and probability to forecast outcomes. Once data has been collected for relevant predictors, a statistical model is formulated.

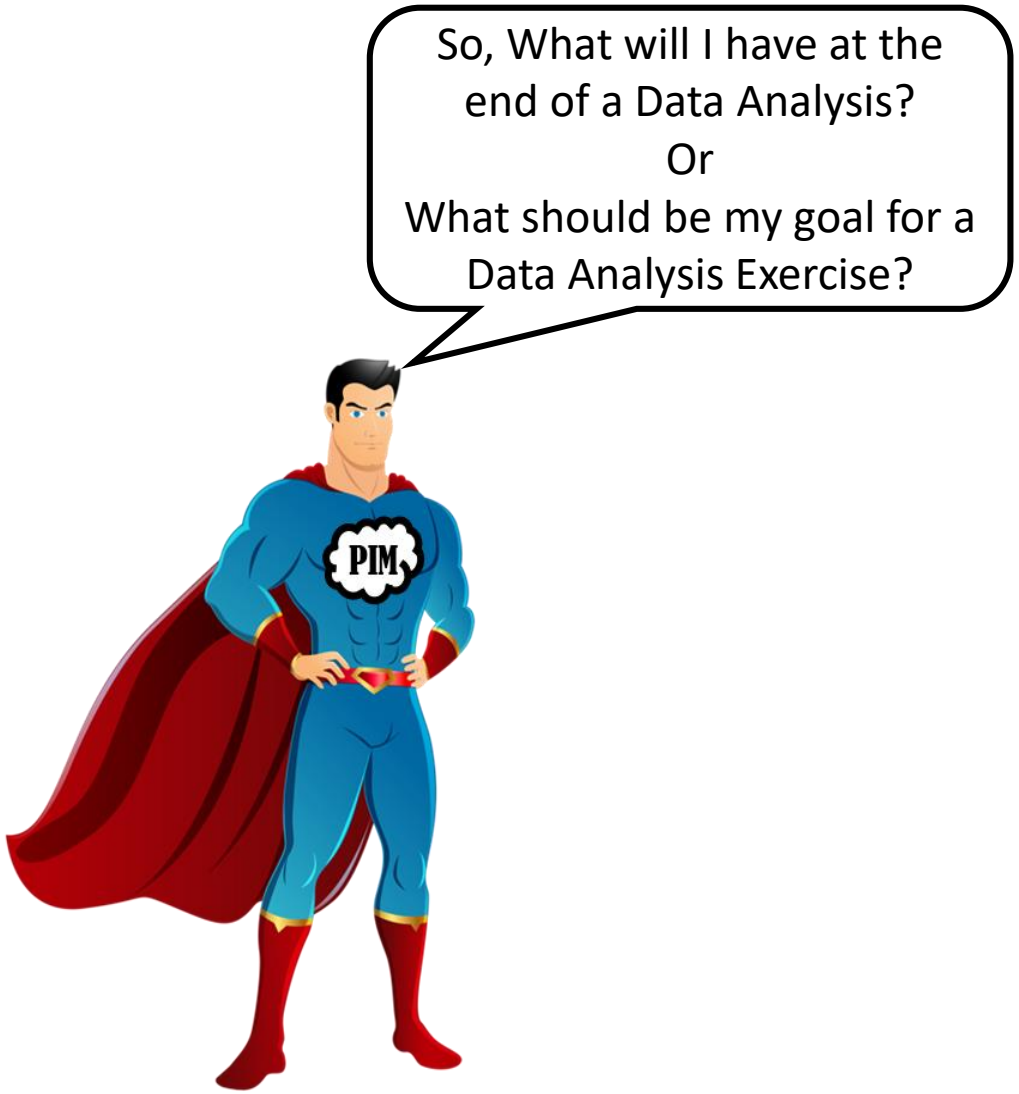


Data Visualization

Data Visualization is the graphical representation of information and data. It involves producing images that communicate relationships among the represented data to viewers of the images.

Data Analysis - Goals

Goal for Data Analysis



- Clean and (if needed) Feature Engineered Data
- A Proven or disproven hypothesis with supporting analysis
- Approach for solution building (model → classification, regression, etc.)
- A Business inference with potential impacts

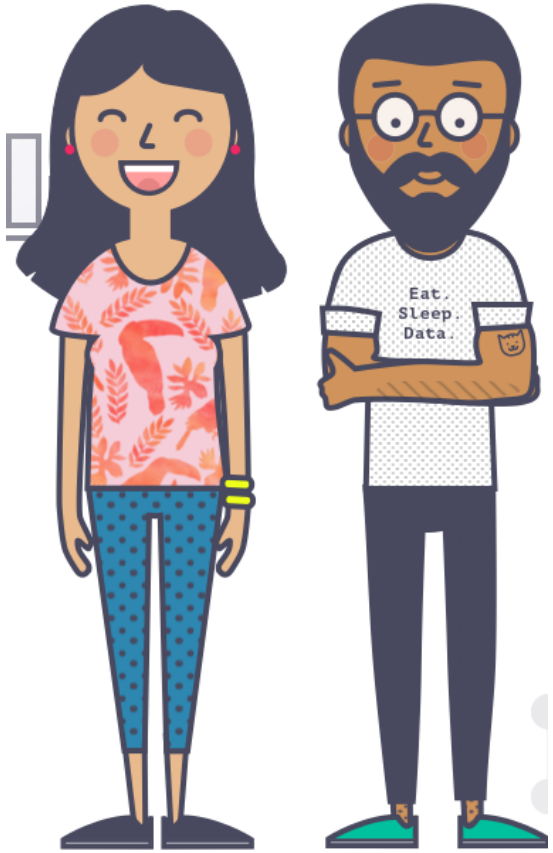
Role Play

Learning through Role Play

STL

PIM

DATA ENGG.

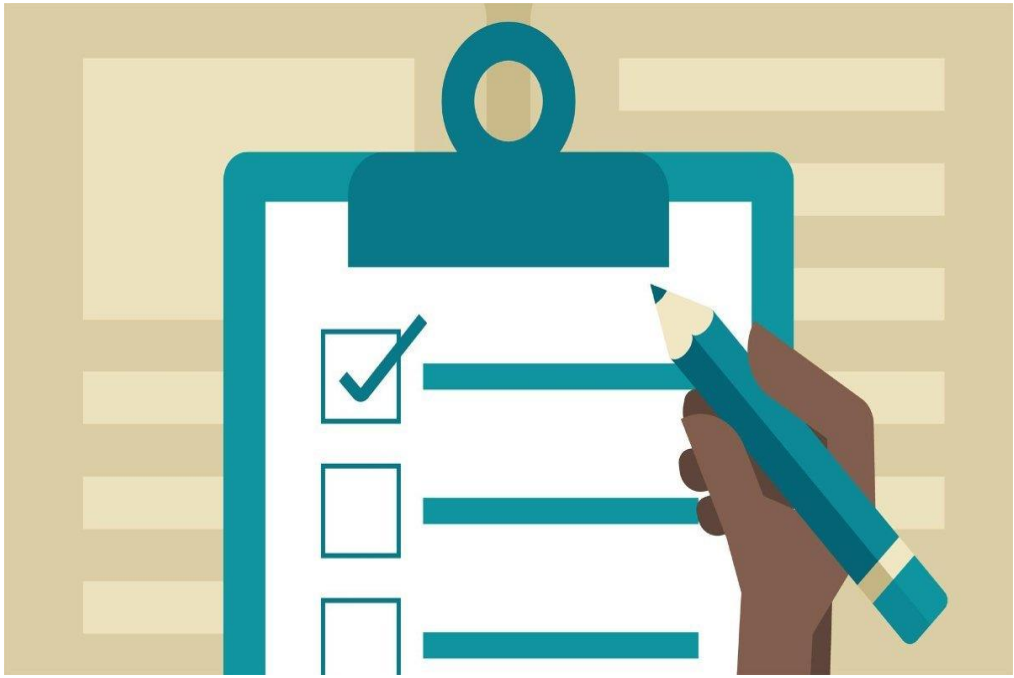


OBSERVER

Key Learnings

What did you Learn?

- Need and importance of Data Analysis in today's world
- What is Data Analysis?
- Types of Data Analysis?
  - Descriptive
  - Diagnostic
  - Predictive
  - Prescriptive
- Data Science Lifecycle
  - Business Understanding
  - Data Mining
  - Data Cleaning
  - Data Exploration
  - Feature Engineering
  - Predictive Modelling
  - Data Visualization
- Setting goals for Data Analysis



Thank You!