# Learn Something Every Day

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Implementation of DS

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Data Science in Search Engine

# Data Science in Finance

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

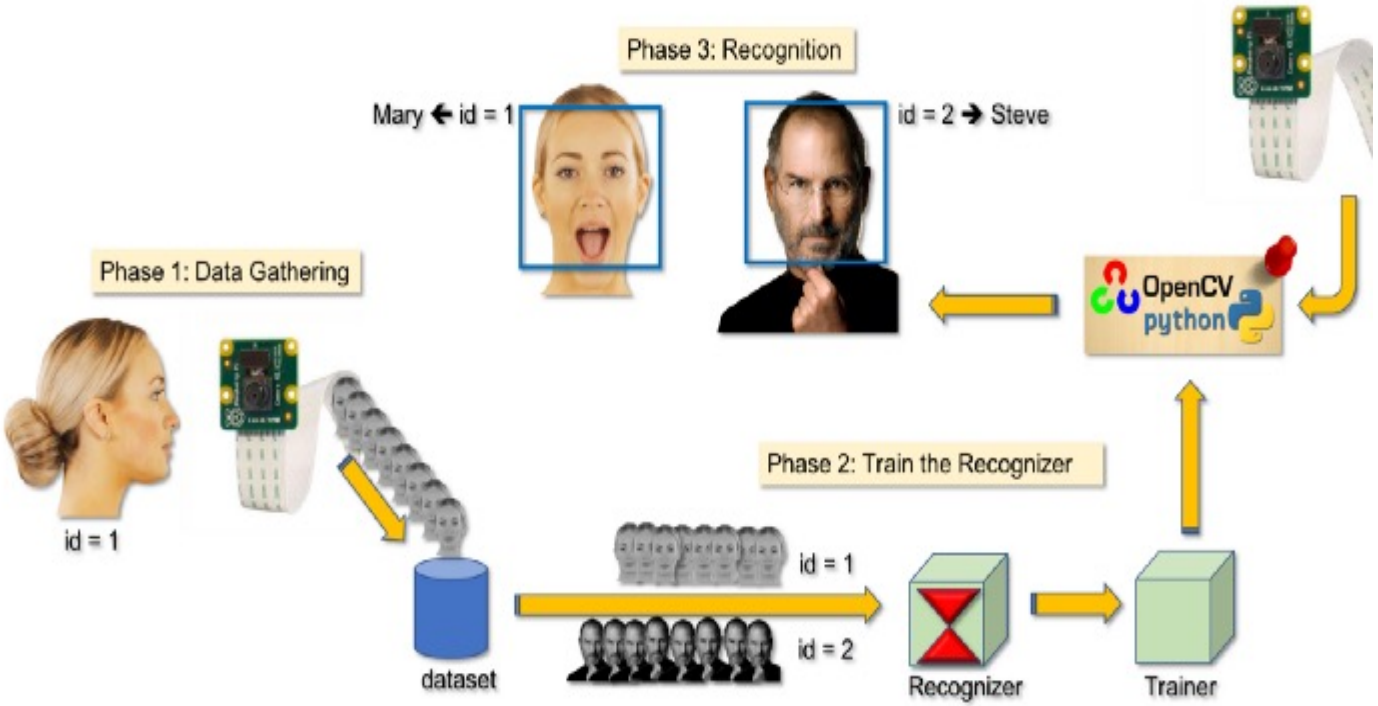
# Data Science in Transport

# Data Science in E-Commerce

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

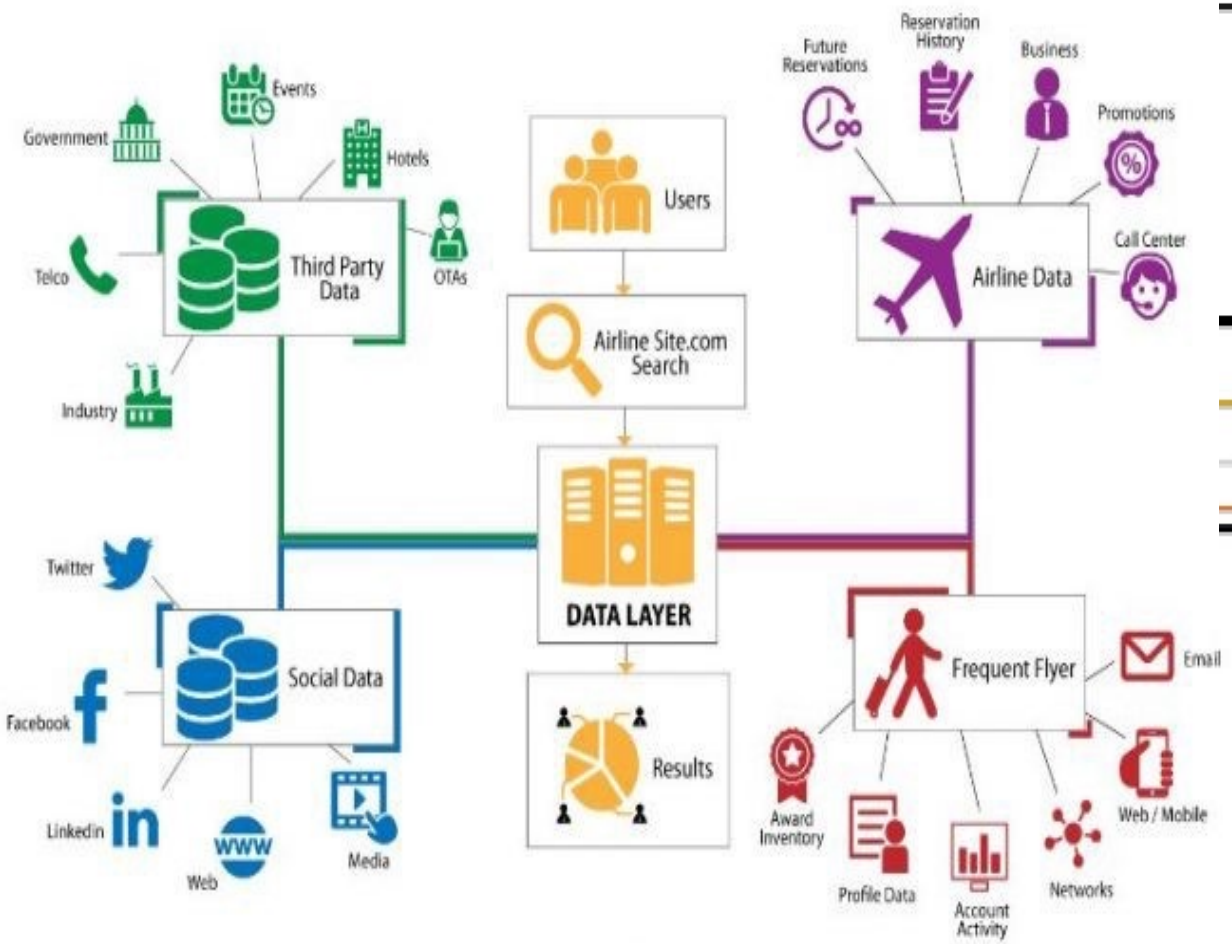
# Data Science in Health Care

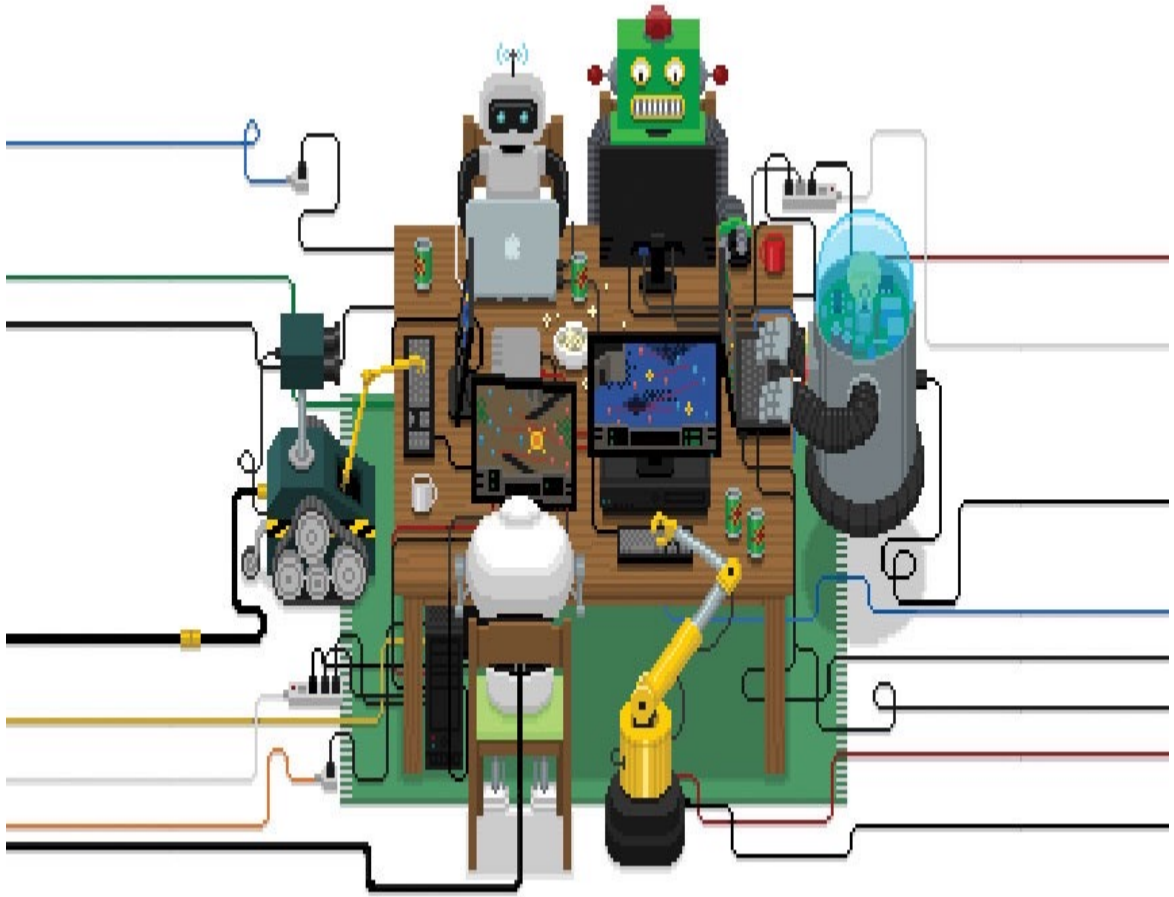
# Data Science in Image recognition




CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Data Science in Airline Routing Planning



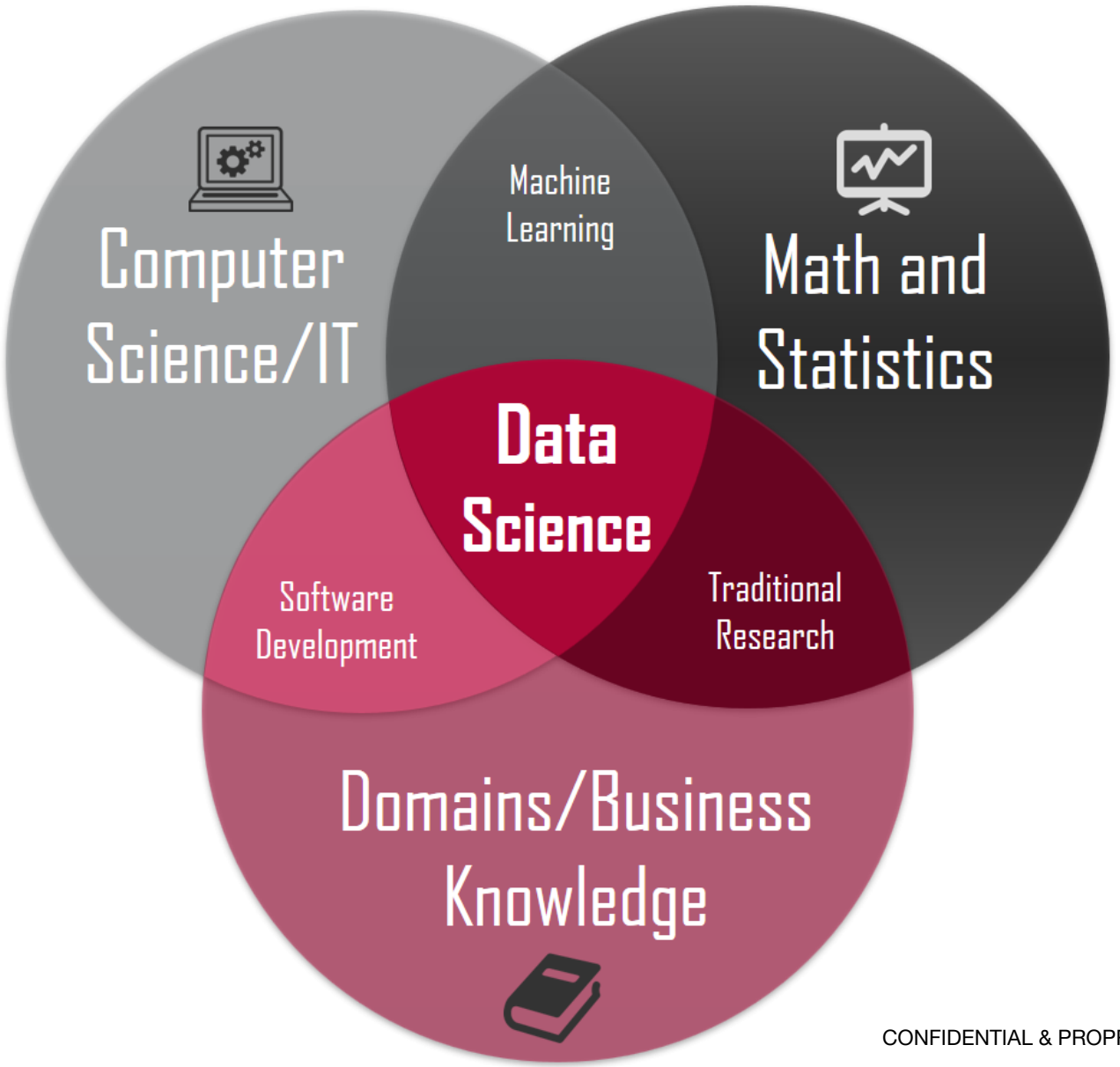



# Data Science in Gaming

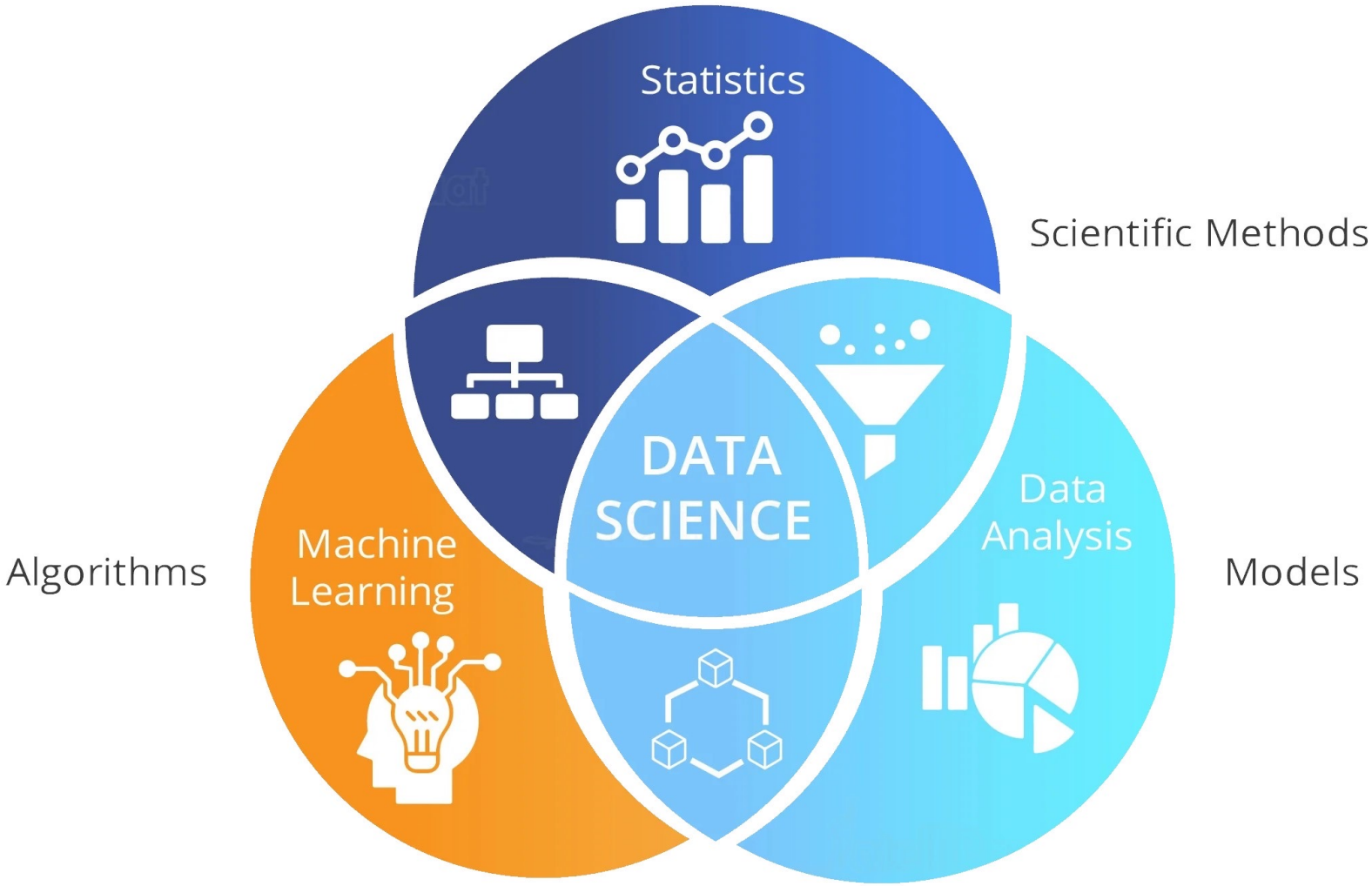
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Introduction to DS

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies




CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies



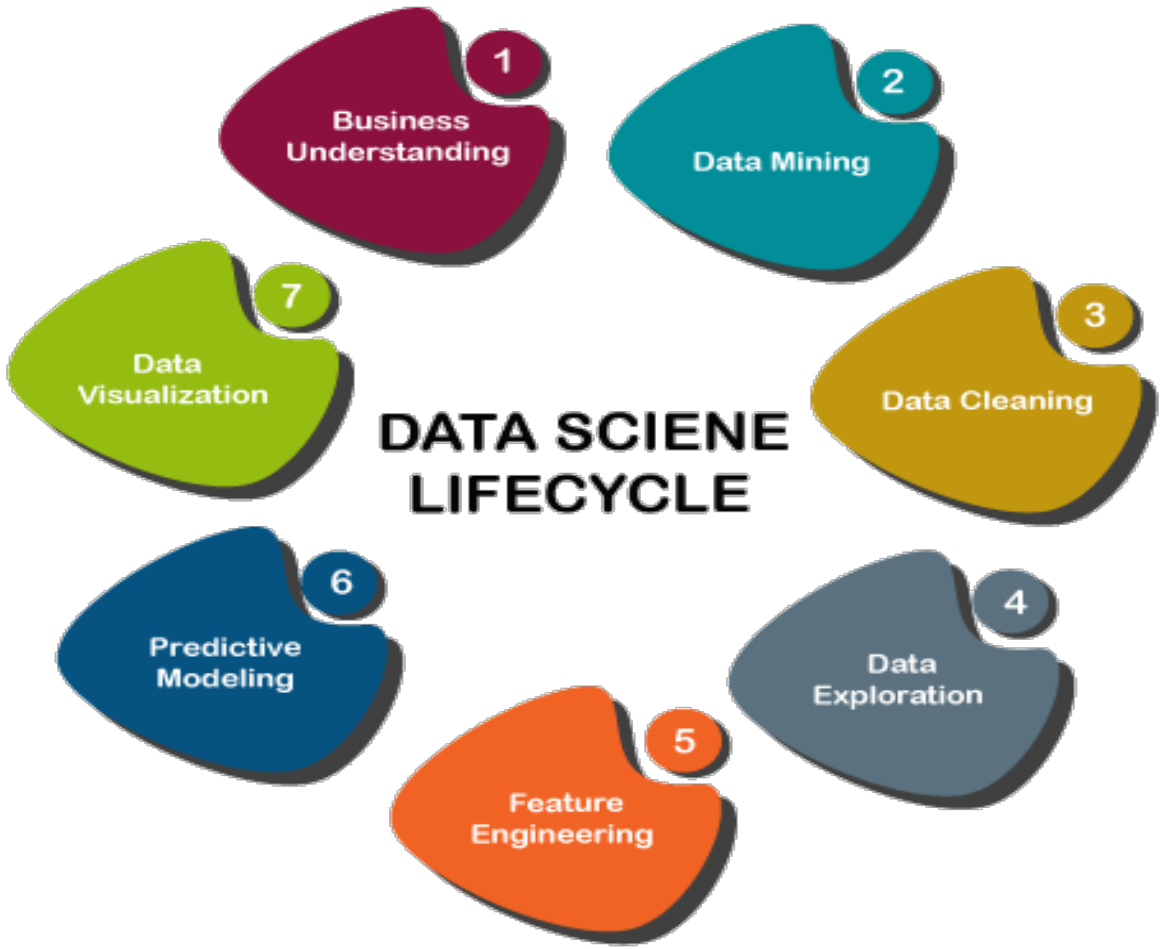

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Data Science Basic

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

1. Understand the project objectives and requirements from a business perspective, and then convert this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

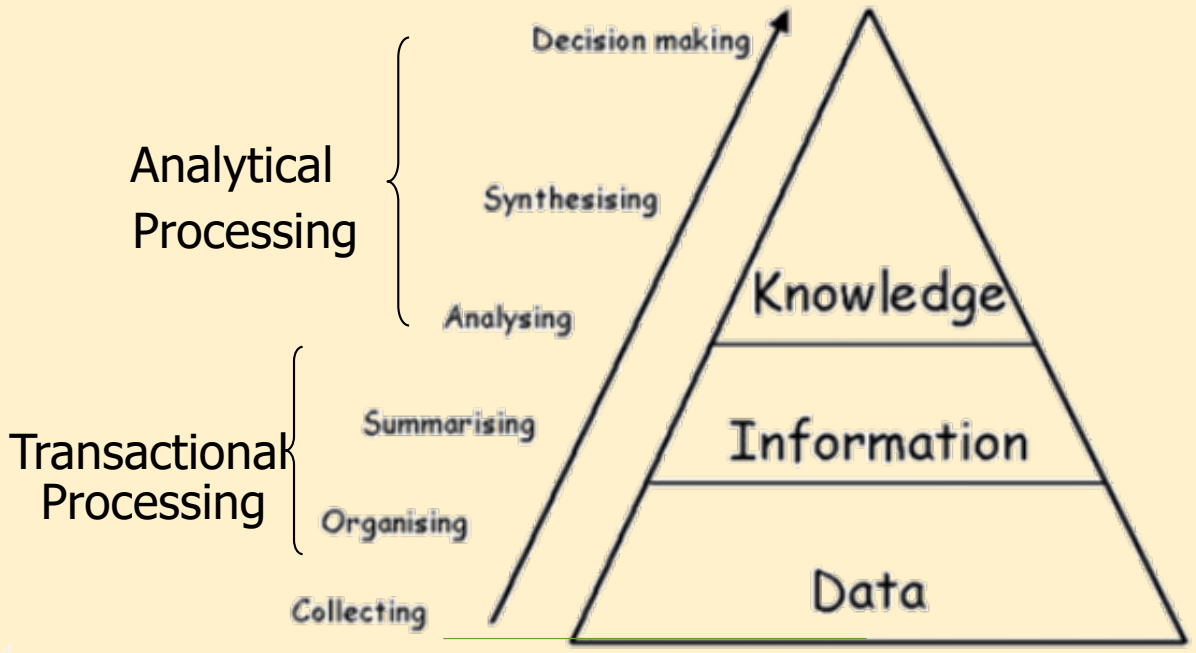



CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

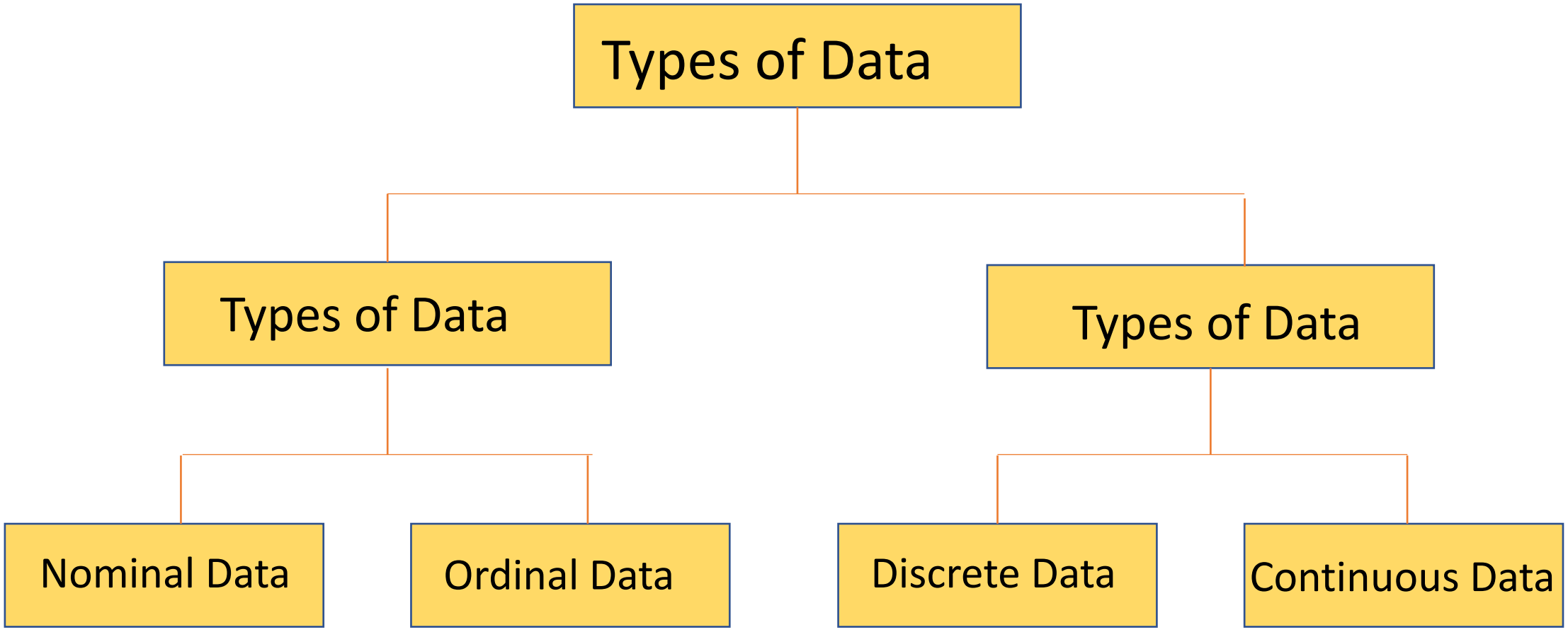
2. Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

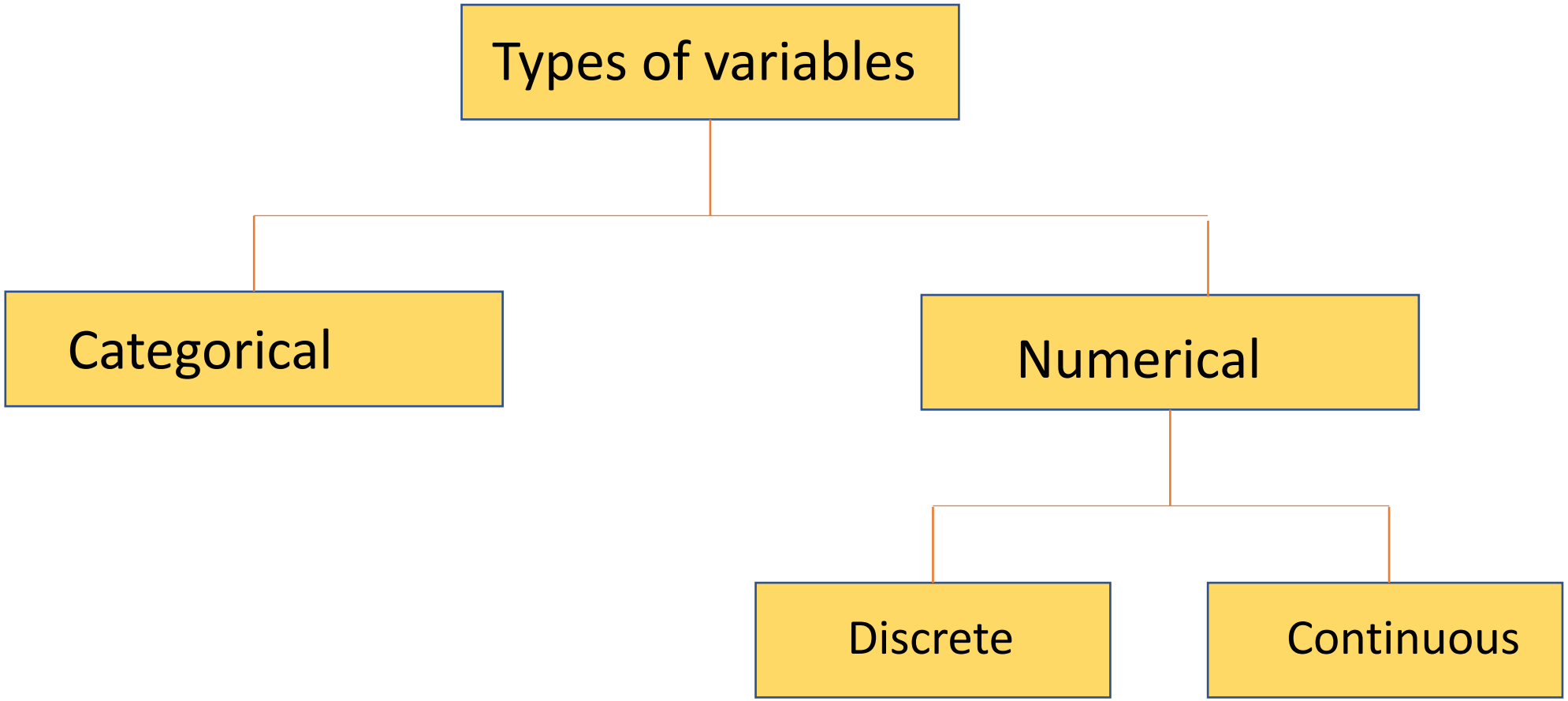
# DIKW




CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies




CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

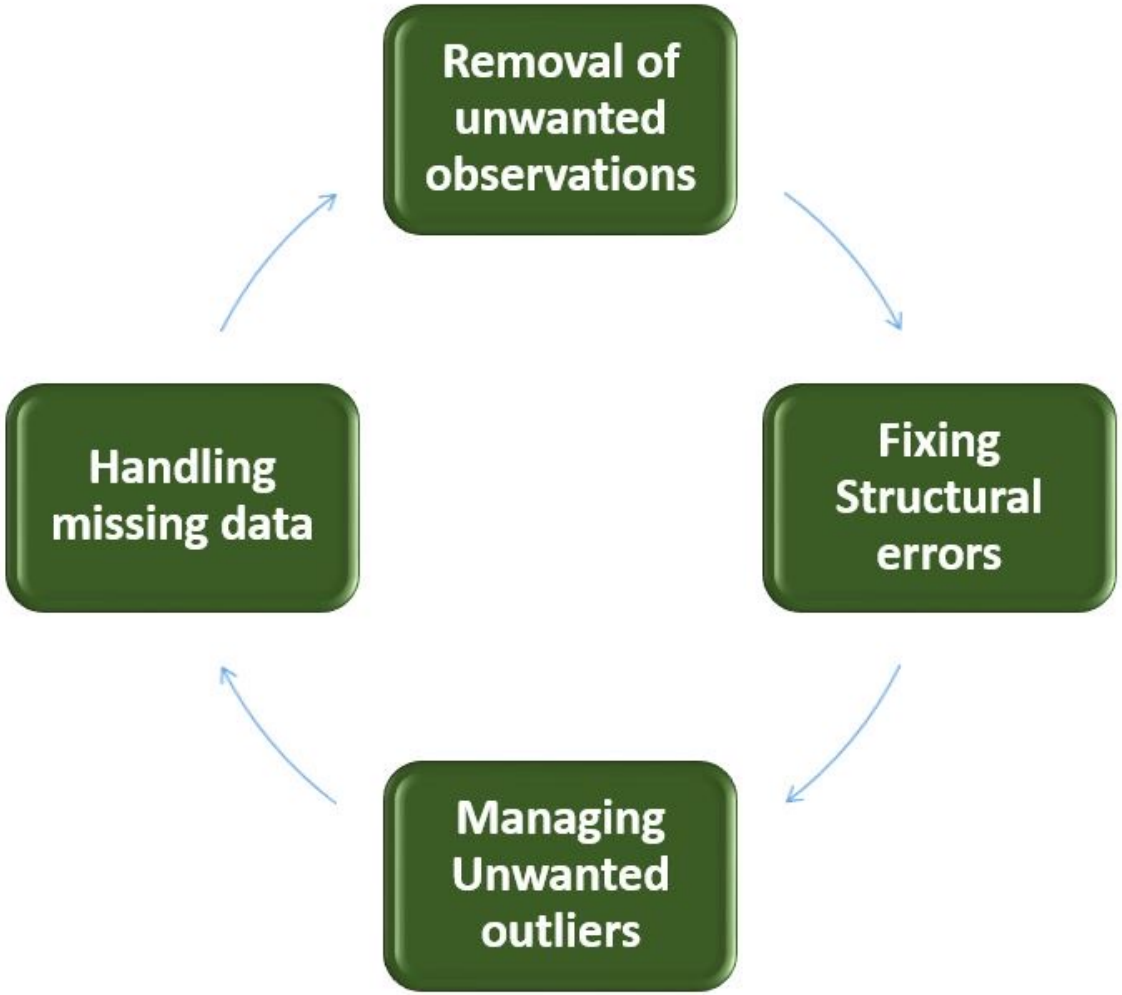



CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

3. Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Data Cleaning

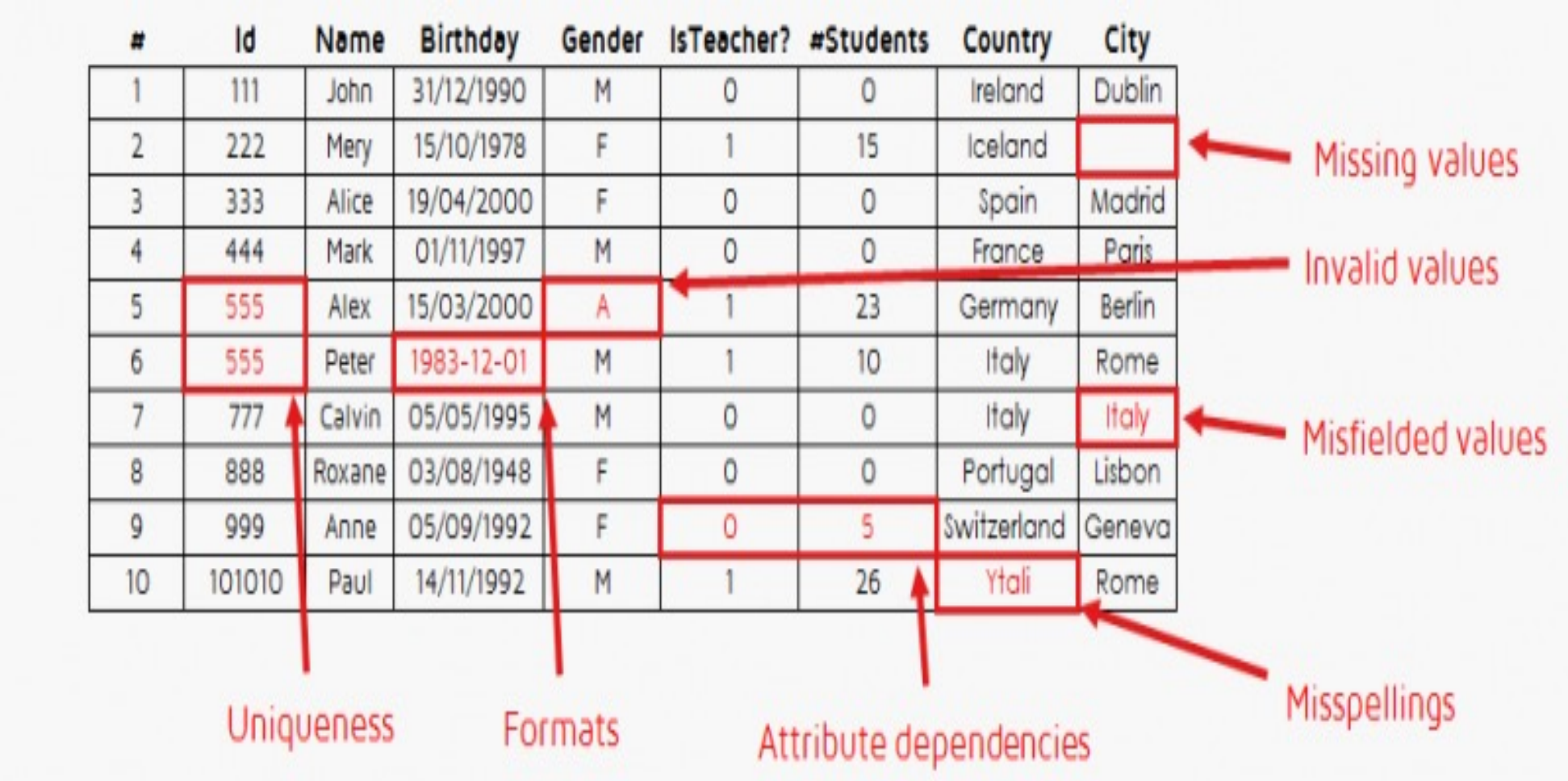



CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# DATA CLEANING STEPS

| | |
|---|---|
| **Removing unwanted observations** | • Duplicate/ redundant or irrelevant values deletion . |
| **Missing Data handling** | • Fixing issue of unknown missing values |
| **Structural error solving** | • Fixing problems with mislabeled classes, typesin names of features, same attribute with different name etc. |
| **Outliers Management** | • Unwanted values which are not fiting in datasets. |

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies




CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

4. Data exploration refers to the initial step in data analysis in which data analysts use data visualization and statistical techniques to describe dataset characterizations, such as size, quantity, and accuracy, in order to better understand the nature of the data.

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Steps in the Data Analysis Process

Step 1: Decide on the objectives or Pose a Question

Step 2: What to Measure and How to Measures

Step 3: Data Collection

Step 4: Summarizing and Visualizing Data

Step 5: Data Modelling

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

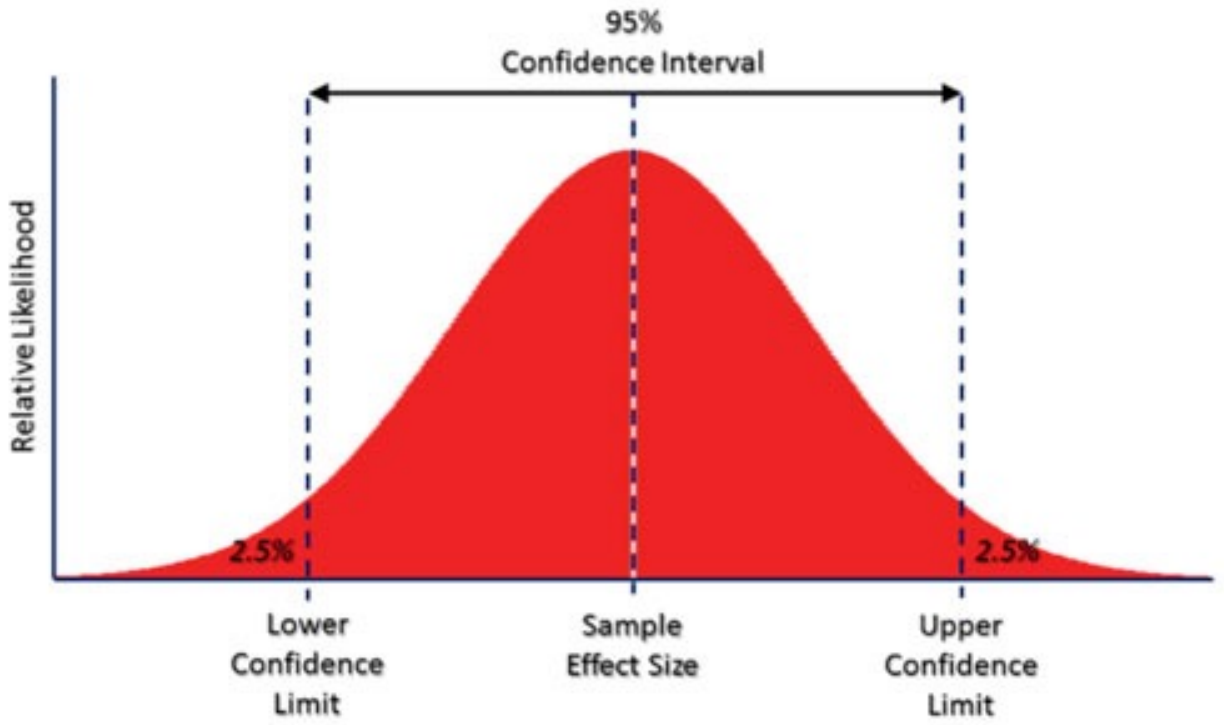
5. Feature engineering is a machine learning technique that leverages data to create new variables that aren't in the training set.

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

6. Predictive modeling is a mathematical process used to predict future events or outcomes by analyzing patterns in a given set of input data.

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

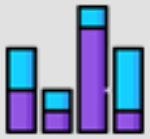
A Confidence Score is a number between 0 and 1 that represents the likelihood that the output of a Machine Learning model is correct and will satisfy a user's request.




CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

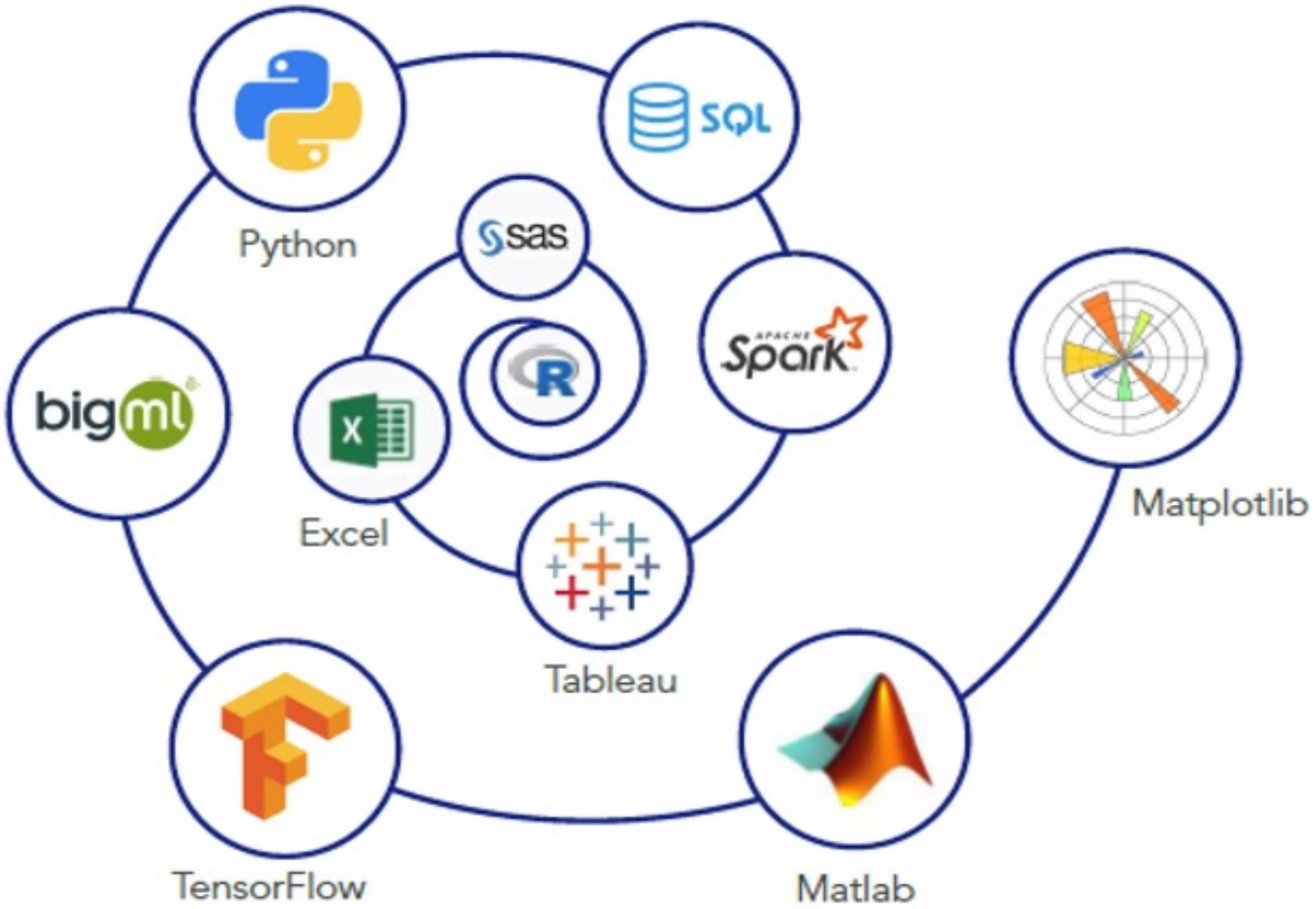
7. Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights.

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

| Chart | Visual | X axis | Y axis | Analysis | Example |
|---|---|---|---|---|---|
| Scatter plot/Line Plot | | Continuous | Continuous | - Understanding linear, non-linear relationship between two variables<br>- Trend analysis, change in KPI over time | - How does heart rate change with age?<br>- How sales of a company varied over a period of time? |
| Bar Graph | | Categorical /Discrete Continuous | Continuous | - How Y (can be any performance indicator) varies across different categories? | - How sales in 2019 varied for different mobile phone brands? i.e. mobile phone brand is the category and sales is the KPI |
| Stack Bar Graph |  | Categorical | Continuous | - Relative comparison of multiple categories within a category | - Comparison of revenue generated by Apple, Samsung & Xiaomi across different products like mobile phone, laptops, television, and headsets |
| Box Plot | | Continuous | | - Outlier detection<br>- Analysing data distribution across Median and Inter Quartile Range | - How different sales figures across a year is distributed? |
| Pie Chart | | Categorical& Continuous | | - Relative comparison of different categories for one single entity in terms of proportion/percentages | - What percentage of Sales in 2019 is constituted by different products under Apple? |
| Histogram Plot | | Continuous | - | - How distribution of values of x varies across different range buckets? | - Distribution of income across income buckets for developing countries |

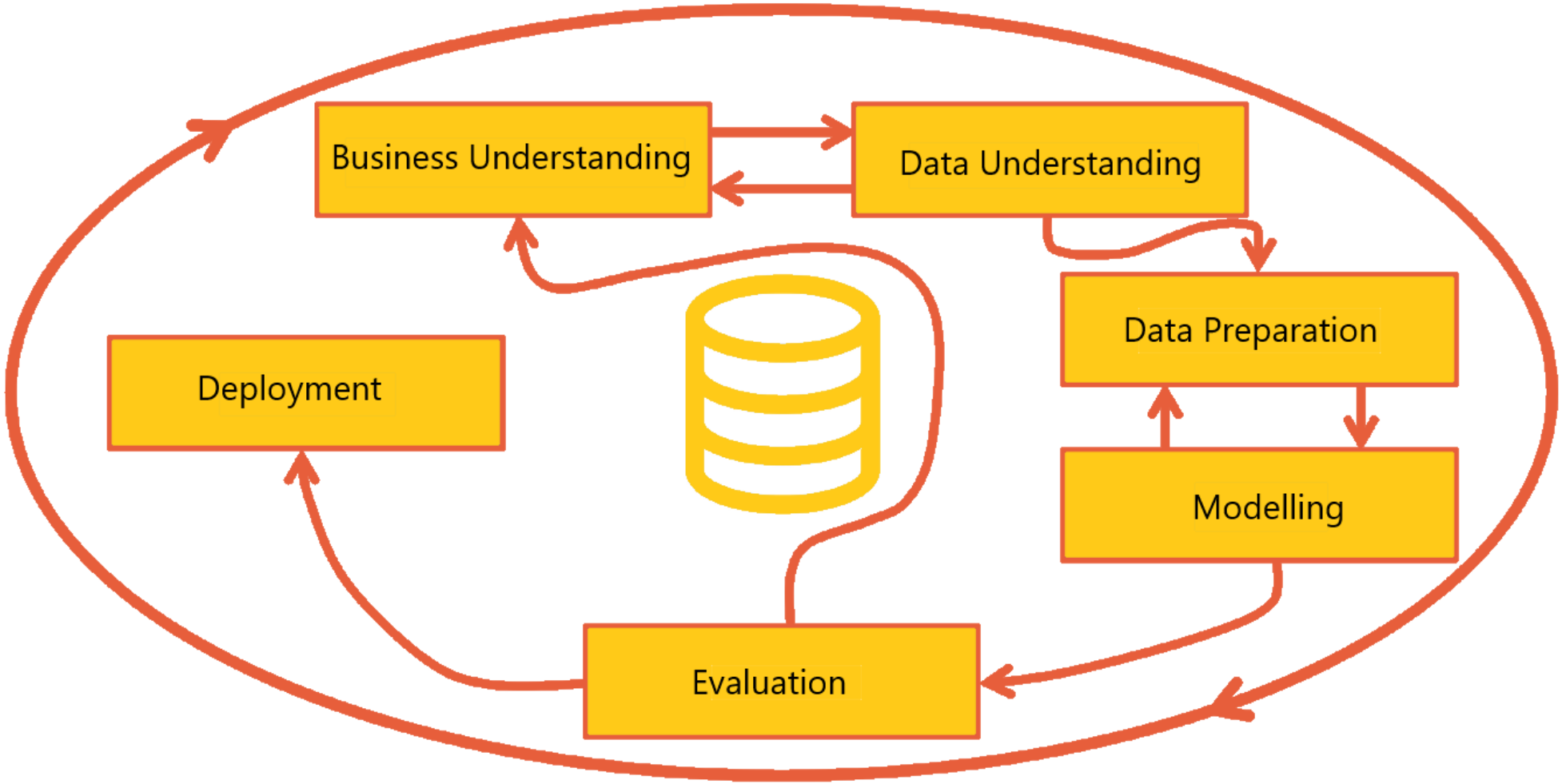
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Data Science Tools



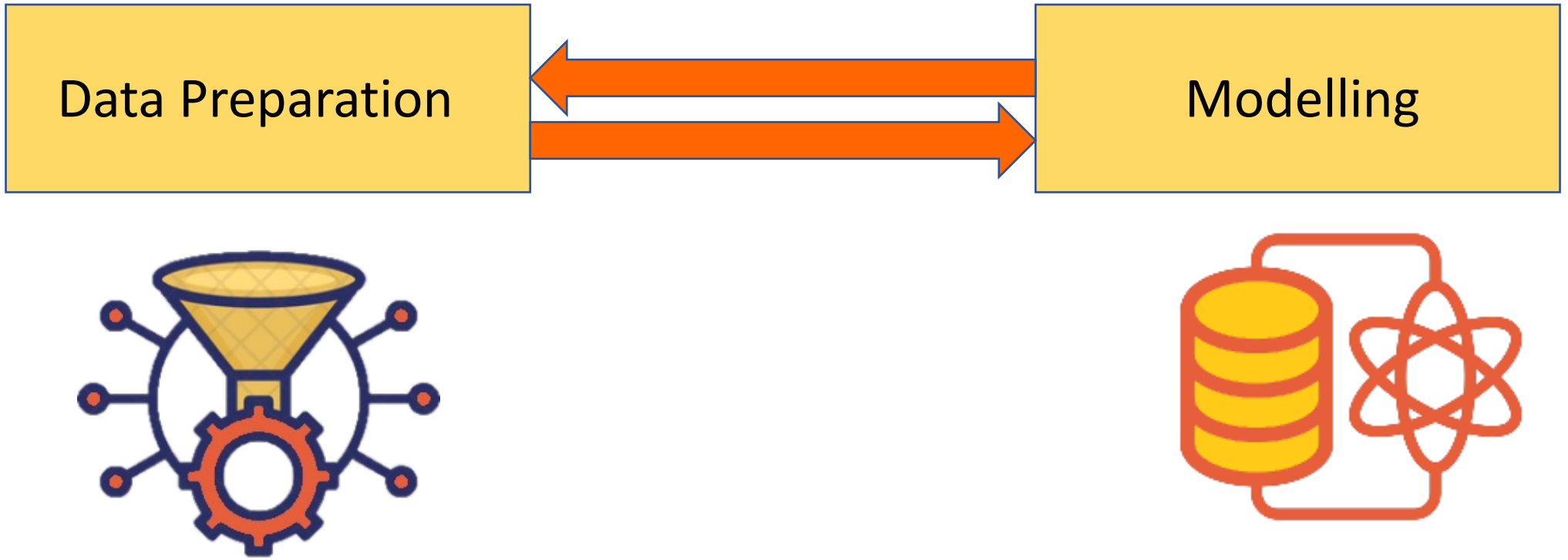

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Data Science Basics




CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Modelling in Data Science




CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

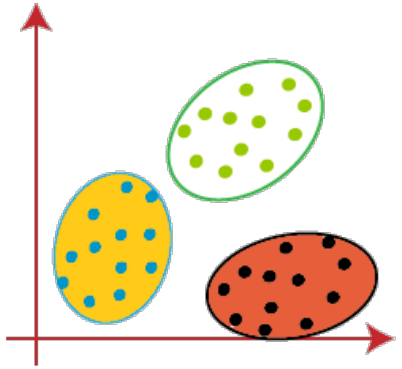
# Data Science Task

ADVERK TECHNOLOGIES

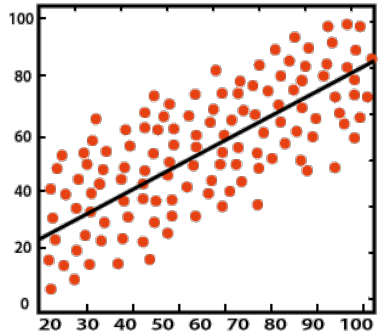
## ML Algorithms

- Classification

- Clustering



- Regression

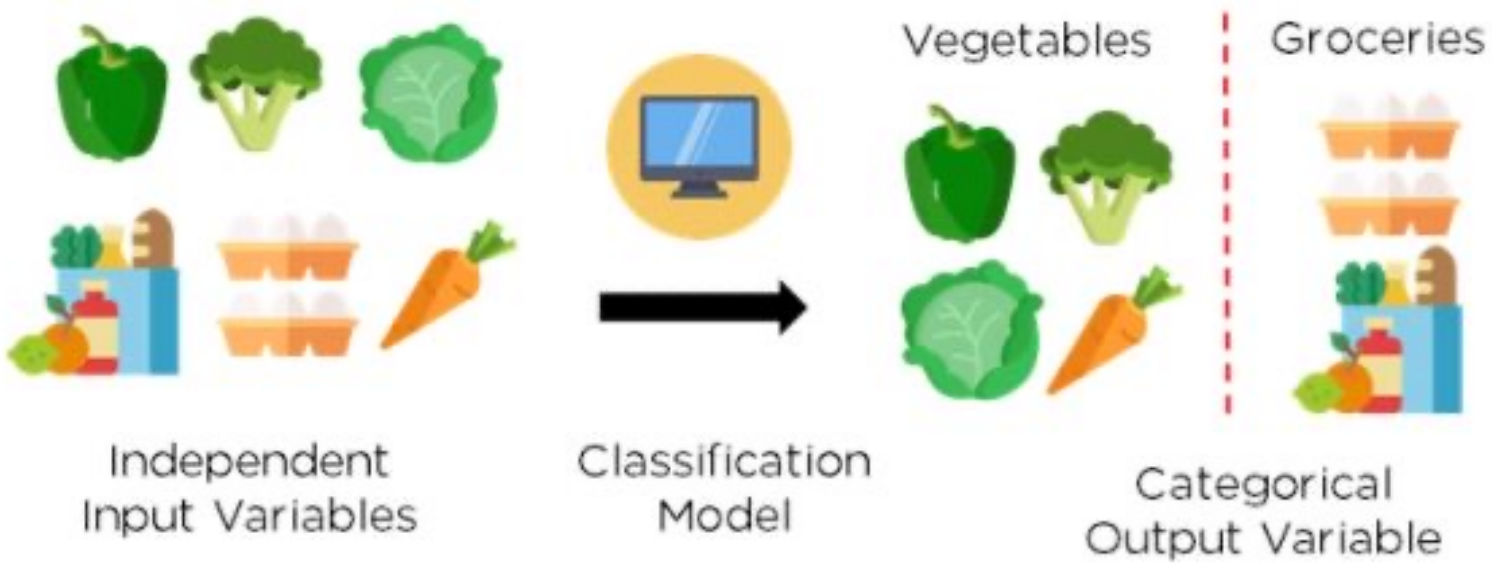



- Association

(A, B) ⟶ C
(D, E) ⟶ F
(A, E) ⟶ G

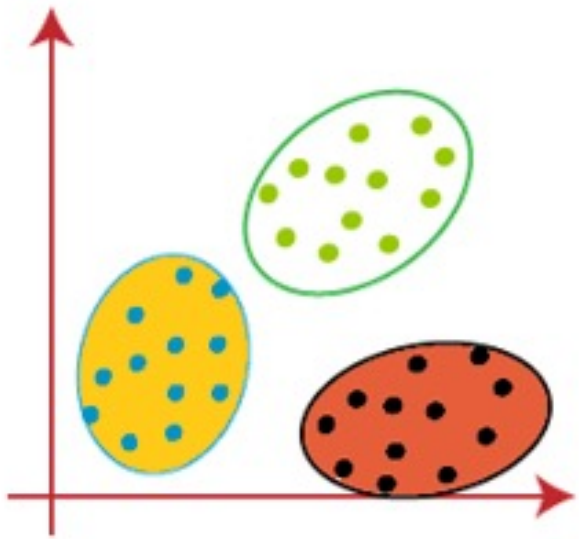
CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

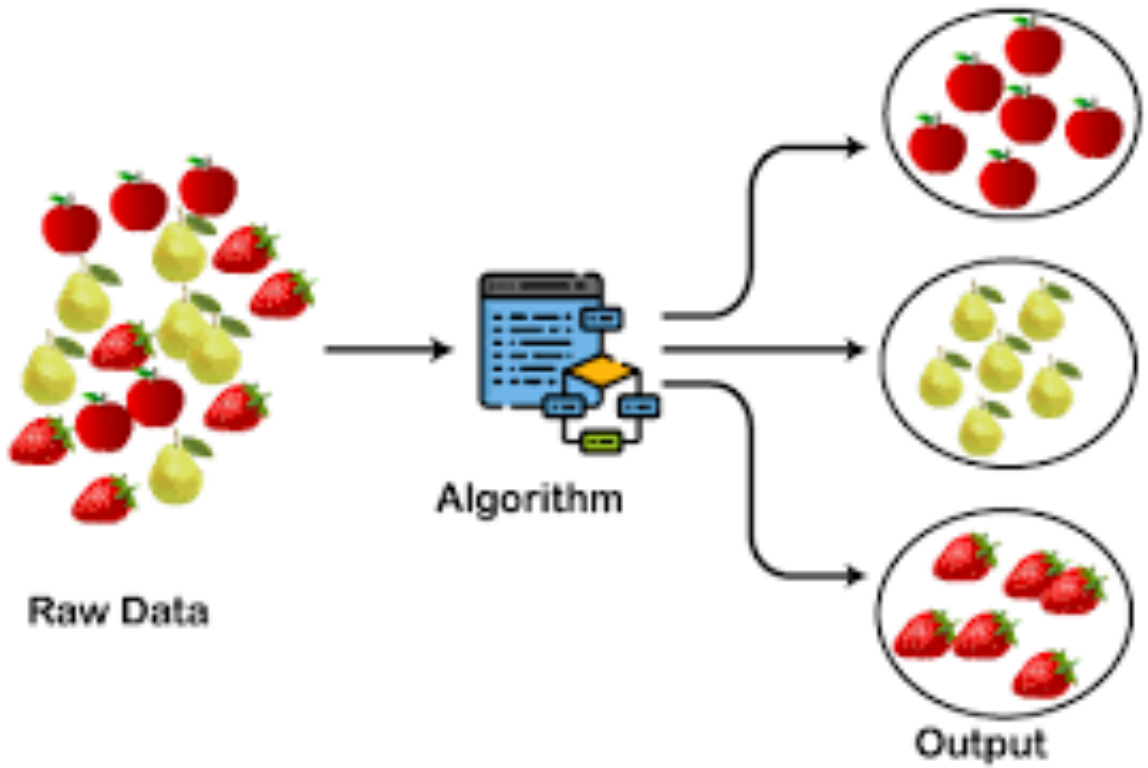
# Classification




CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

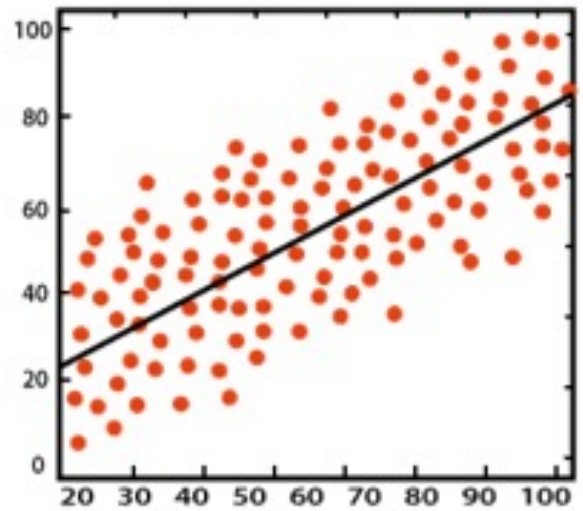
# Clustering

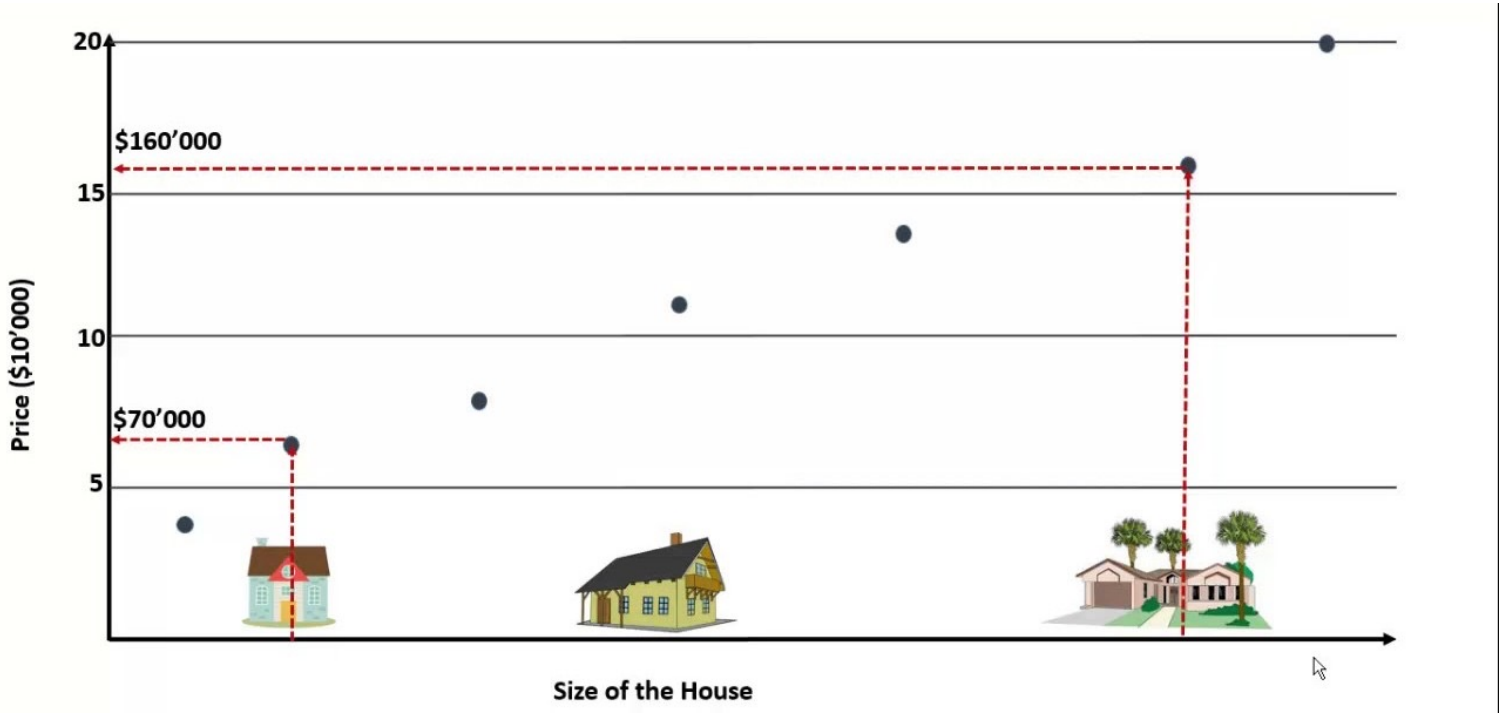





CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Regression




CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# Association

(A, B) ⟶ C

(D, E) ⟶ F

(A, E) ⟶ G

Customer 1

Customer 2

Customer 3

Customer n

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies

# THANK YOU!

CONFIDENTIAL & PROPRIETARY 2022 – Adverk Technologies