SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)
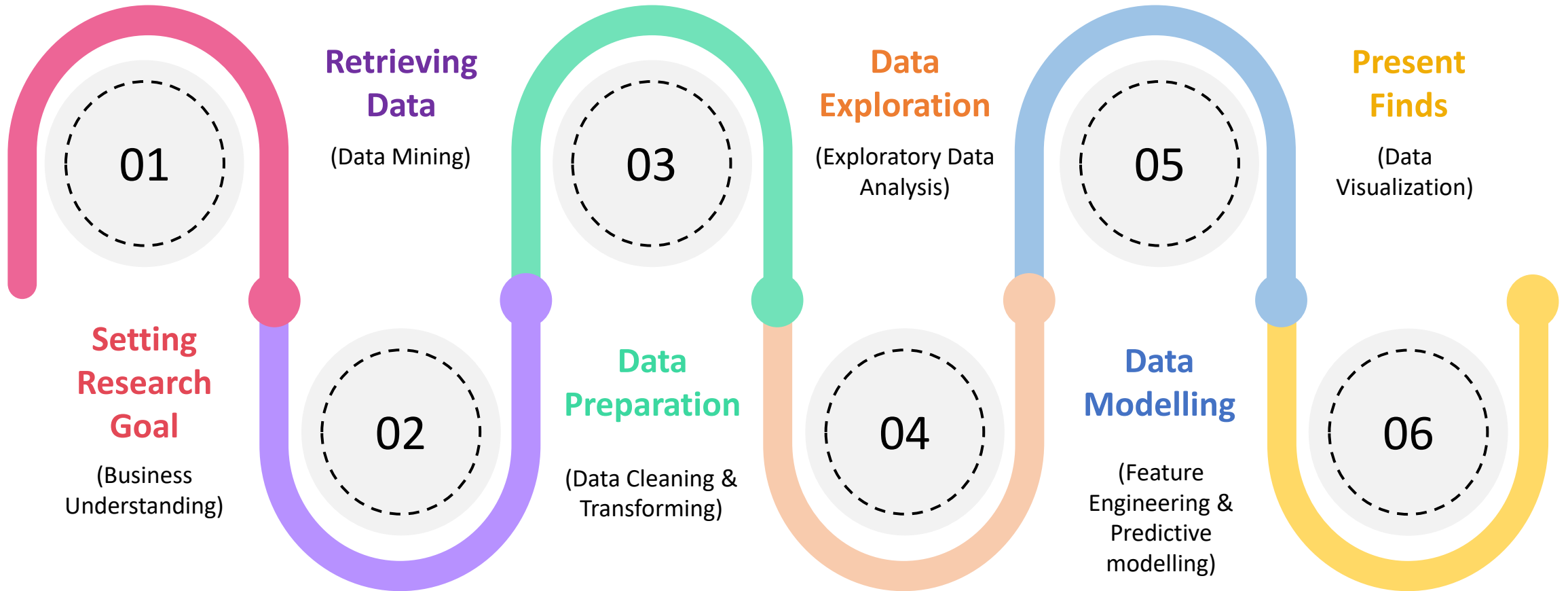
Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai
Accredited by NAAC-UGC with 'A++' Grade (Cycle III) &
Accredited by NBA (B.E - CSE, EEE, ECE, Mech & B.Tech.IT)
COIMBATORE-641 035, TAMIL NADU

DEPARTMENT OF COMPUTER APPLICATIONS

23CAE704 – DATA SCIENCE

UNIT – II: DATA SCIENCE PROCESS

TOPIC: OVERVIEW – SETTING RESEARCH GOALS

01 Setting Research Goal (Business Understanding)

Retrieving Data (Data Mining)

02

03 Data Preparation (Data Cleaning & Transforming)

Data Exploration (Exploratory Data Analysis)

04

05 Data Modelling (Feature Engineering & Predictive modelling)

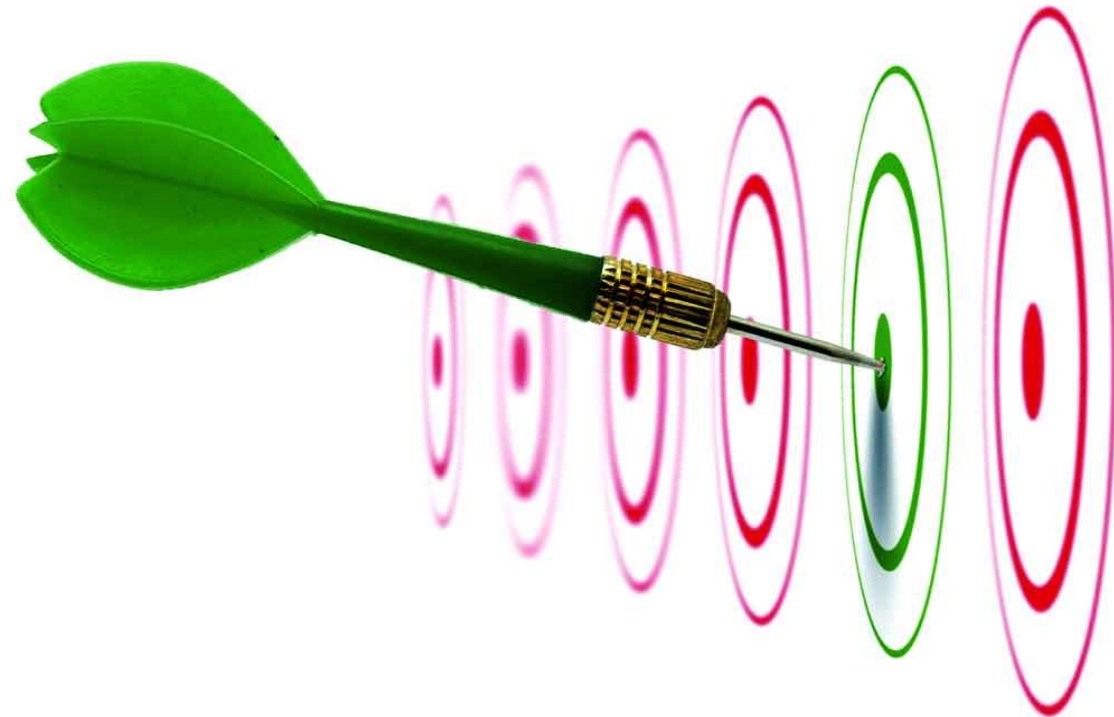Present Finds (Data Visualization)

06

Setting Research Goals

✓ Every data science project should begin with a clear understanding of the problem at hand.

✓ Whether it's optimizing a business process, predicting customer behavior, or identifying patterns in a dataset, articulating the problem statement is essential.

✓ The problem statement sets the stage for the entire research endeavour, providing a focused direction for subsequent steps.

✓ Once the problem is defined, establish the specific objectives you aim to achieve.

✓ These objectives should be measurable and tied to key performance indicators (KPIs).

✓ For instance, if the goal is to enhance customer engagement, metrics like click-through rates, conversion rates, or customer satisfaction scores can be employed to quantify success.
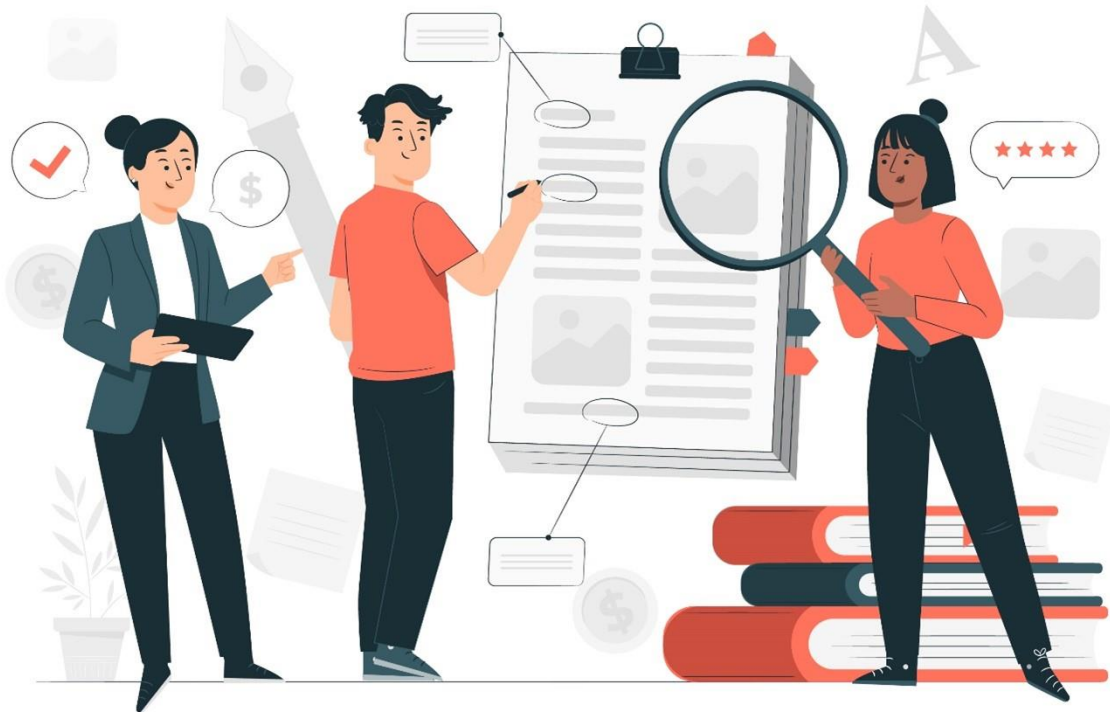
✓ Collaboration with stakeholders is integral to the success of any data science project.

✓ Engage with domain experts, business leaders, and end-users to comprehend their needs and expectations.

✓ This collaborative approach ensures that the research goals align with organizational objectives and address real-world challenges.

Conduct a Literature Review

- Before diving into data analysis, it is beneficial to review existing literature in the field.

- A thorough literature review helps in understanding previous research findings, methodologies, and potential gaps in knowledge.

- This step aids in refining research questions, avoiding redundancy, and incorporating best practices.

Formulate Hypotheses

✓ Based on the problem statement, objectives, and literature review, formulate hypotheses that can be tested using data.

✓ Hypotheses guide the data collection and analysis process, providing a structured approach to uncovering patterns, relationships, or trends in the data.

Data Collection and Preparation

- ✓ Selecting and preparing the right data are critical components of any data science project.

- ✓ Define the data sources, collect relevant datasets, and clean the data to ensure its quality.

- ✓ The quality of insights derived is directly proportional to the quality of the data analyzed.

Choose Appropriate Methods & Models

- Selecting the right analytical methods and models is crucial for achieving accurate and meaningful results.

- Depending on the nature of the problem, choose appropriate statistical techniques, machine learning algorithms, or other analytical methods.

- Consider factors such as interpretability, scalability, and computational efficiency when making these choices.

✓ Establish criteria to evaluate the performance of models and methods.

✓ This step is essential for measuring the success of the research and determining whether the objectives have been met.

✓ Common evaluation metrics include accuracy, precision, recall, and F1 score for classification problems, or Mean Squared Error for regression tasks.

✓ Data science is often an iterative process.

✓ Conduct initial analyses, evaluate results, and refine the approach based on insights gained.

✓ Iterative refinement allows for the improvement of models and methods, ensuring that the research goals are achieved effectively.

✓ The final step in any data science research is to communicate findings effectively.

✓ Whether through reports, visualizations, or presentations, convey the results in a manner that is accessible to both technical and non-technical audiences.

✓ Clearly articulate the implications of the research and any recommendations for future actions.