

The Data Science Process

St. Joseph's University, Bengaluru

Basics of Data Science

The Data Science Process

Setting the Research Goal → Retrieving Data → Data Preparation → Data Exploration → Data Modeling → Presentation

Setting the Research Goal

Prepare a Project Charter

1. What will you research?
2. What Data?
3. What resources?
4. Timeline
5. Deliverable



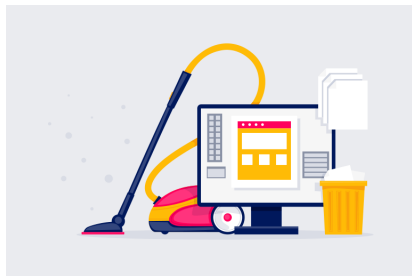
Retrieving Data



1. Government websites/databases
2. Built in data sets
3. <https://www.kaggle.com>
4. <https://archive.ics.uci.edu/ml/index.php>

Data Preparation

1. Data Cleansing
2. Data Integration
3. Data Transformation



Data Exploration

Primarily, data visualization techniques are used in this step.
This step is also called **Exploratory Data Analysis (EDA)**.



Data Modeling and Model Building

This step has three phases:

1. Selecting Variables 2. Executing the model 3. Model Diagnosis

