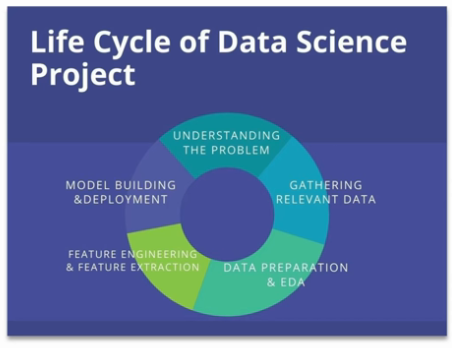
**Introduction to the data science workflow**

****

Understanding the problem:

* Define the objectives clearly.
* Determine how to measure the project’s success.
* Identify the problem type (e.g. classification, regression).

Gathering relevant data:

* Collect data from available sources.
* Ensure data is relevant to the problem.
* Consider the volume, variety, velocity, and veracity of the data.

Data Preparation and EDA(Exploratory Data Analysis):

* Clean the data (handle missing values, remove duplicates).
* Perform exploratory analysis to understand the data.
* Normalize or scale the data if necessary.

Feature Engineering and Feature Extraction:

* Create new features that can help improve model performance.
* Reduce dimensionality if the feature space is too large.
* Select the most important features to be used for modelling.

Model Building and Deployment:

* Choose appropriate algorithms and train models.
* Validate model performance using cross-validation.
* Deploy the model for real-time use or batch processing.