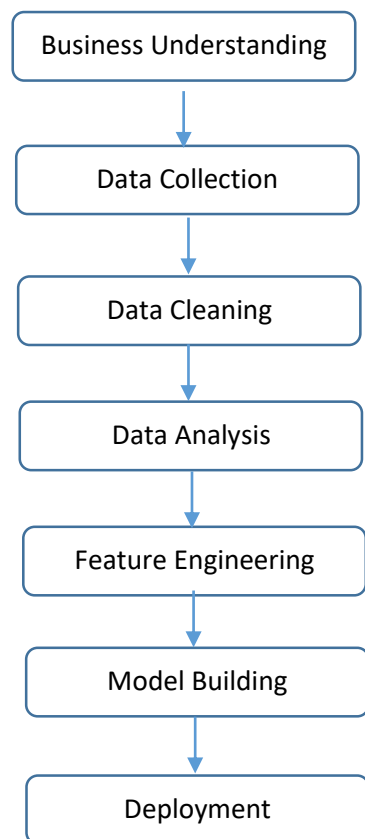


## **Case Study 5: Trainer Statistics**

**Description:** Trainers are available in ARICH who have different skill set & background. While many of them are in-house employees few others are consultants. Trainers are allocated to in-house as well as corporate training requirements based on demands & duration. Collect data relevant to this business case and do a projection on the demand for trainer on specific technologies, matching skill sets of trainers & requirements fulfilled by in-house & consultant trainers and in what time they are fulfilled and what up-skilling needed for trainer based on requirement that could not be fulfilled.

**Pre-requisite:** Clear understanding of the lifecycle of Data Science & Analysis, as follows.



### **General Instructions:**



- ✓ For the given Case Study, you are expected to follow all the instructions & implementation given in this document
  - ✓ Every individual is expected to complete the learning before working on this project
  - ✓ Also ensure you learn the Lifecycle & follow it
  - ✓ Identify what all planned to be used – plug ins / IDE / Tools
- 
- ✓ Feel free to include additional flows or tweak any existing flows to the given problem statement to add flavours to the functional scope of work

- ✓ Identify any suggestions / recommendations to process, new or best practices to make it more interesting

### **Data to be collected:**

Trainer:	Company:	Trainer Allocation:
Trainer Id, Trainer Name, Education, Contact Number, Work Experience, Technologies, Total Batches, Employee Type, etc	Company Id, Company Name, Company Type, Technology, Trainers Count, Expected Start Date, Duration, Status, etc	Trainer Id, Company Id, Start Date, Technology, Status, etc

### **Prediction Analysis:**

- Specific background of trainers with experience
- Preferences of technologies & trainers demanded by company
- On-time fulfilment of demand & gap analysis

### **Process:**

- Step 1: Collect set of data as given above & store in excel
- Step 2: Read the data; Check for validity & integrity of the data
- Step 3: Apply mathematical formula as needed for any calculation
- Step 4: Create a Python Program by including necessary python functions
- Step 5: Fetch the data from Data Source (eg: Database, Excel Sheet, etc)
- Step 6: Include packages as needed for any Data Application process.
- Step 7: Build the program to assess the data & predict values
- Step 8: Finally print the data as measured & validate for accuracy.

### **Expected Components in the Project:**

- Datasets & Resources
- Exploratory Data Analysis
- Semantic Analysis
- Decision Tree
- Linear Regression
- Cross Validation
- Intuition behind Logistic Regression
- Accuracy of Model
- Pandas for Data Manipulation
- Numpy for Mean & Mediation
- Matplotlib, Seaborn, Plotty etc for Data Visualization
- ChatBot