**ROAD MAP**

**PROBLEM UNDERSTANDING:**

* Data science problem discussion and setting project objectives
  + Creating procedures, project road map
  + Performance goal definitions
  + Time & resource planning and management

**DATA UNDERSTANDING:**

* Creating data dictionary and updating when necessary
* Data analysis, descriptive statistics
  + Missing value checking and value replacement if necessary
  + Variable distributions
  + Frequency analysis
  + Descriptive statistics

**DATA PROCESSING:**

* Feature engineering and creating new variables
  + Creating new variables by combining existing ones
  + Defining independent variables (explanatory variables) and dependent variable (target variable) for supervised learning
  + Understanding the effects of individual variables on target variable for supervised learning
  + Checking co-linearity among variables for elimination and obtaining most important variables by using:
    - Correlation analysis
    - Decision trees
    - PCA, low/high variance methods
  + After this step, ML model decision will be made
  + Before applying model, data will be divided into 2 parts, namely: Training and Testing Sets

**STATISTICAL MODELING OF DATA:**

* Machine learning/statistical modeling of data

**EVALUATION PHASE:**

* Evaluating results
* Testing and validation
* Creating use cases and result presentations

**VISUALISATION OF RESULTS AND FINDINGS:**

**DEOPLOYMENT/PRODUCTION ENVIRONMENT:**

* Finalizing the project with all steps
* Models are put into operation for the intended usage