- ***Introduction to Data Science***

In this part, we introduce the essential concepts such as the data science cycle, regression/classification models, evaluation criteria, and overfitting/underfitting avoidance strategies.

- ***Extract, Transform, Load (ETL)***

There is no data science without data, and we need to learn how to get the data and prepare it. We review various data sources and formats that we might face in a problem and how to extract data from them. We discuss how to read these structured and unstructured data and convert them to a form that is suitable for a modeling procedure. This includes data shaping, cleaning, and manipulation using python and R tools.

- ***Exploratory Data Analysis (EDA)***

Next, we explore the data to better understand its structure, features, and how these features are related to each other. In this part, we review some of the important EDA techniques, from statistical analysis to visualization techniques that help us find patterns in the data.

- ***Feature Engineering***

Predictive models rely on features, and better features can generate better models. In this part, we review some of the main data transformations for generating and modifying features that increase the predictive power of the modeling algorithm. Also, we discuss feature selection methods to eliminate redundant features.

- ***Linear models, Generalized Linear Models (GLM)***

Linear models are simple but intuitive. We review the linear model, and some of its extensions such as the Generalized Linear Model (GLM), Generalized Additive Model (GAM), and Generalized Linear Mixed Model (GLMM). We discuss the techniques that help us to interpret linear models such as the concept of elasticity and various visualization types. We also discuss regularized linear models, Lasso and Ridge regression models.

- ***Decision Trees, Random Forests, Gradient Boosting***

Tree-based models are another powerful method in real applications. We review the decision trees models, their learning, and interpretation. Then, we discuss ensemble tree methods, random forest and gradient boosting, their implementation, and hyper-parameter tuning.

- ***Artificial Neural Networks, Deep Learning***

Neural Networks are popular nonlinear methods that used many data science problems. While this area is very vast, we discuss the main concepts such as how to train a network, tune the hyper-parameter, and interpret the model.

- ***Unsupervised Learning Methods***

Unlike classification and regression problems, many real-world data do not have specific labels. Clustering is the main technique to find patterns in such data. We review important clustering approaches and how to do the model selection on them.

- ***Introduction to Clouds, Model Deployments***

We want to make our models alive, and we need to know how to implement and run them in a cloud environment. We review how to start working in the IBM Cloud and how to use Watson Studio to train and deploy your AI model.