**Module 1**

**From Problem to Approach**

**Business Understanding**

Business Understanding

What is the problem that you are trying to solve?

Analytical Approach

How can you use data to asnwer the question?

What is the goal? Profit? Reduce cost?

**Analytic Approach**

How can you use data to answer the question.

Identifyinf what type of patters will be needed

Analytic approach:  
Predictive model – probabilities of an action

Descriptive model – Show relationship

Classification model – classification model

Case study – applying concept  
decision tree classification – likelihood of classified outcome

Lesson Summary

In this lesson, you have learned:  
The need to understand and prioritize the business goal.

The way stakeholder support influences a project.

The importance of selecting the right model.

When to use a predictive, descriptive, or classification model.

**Data Requirements**

Data requirements: What are data requirements?

Prior to undertaking the data collection and data preparation stages of the methodology, it's vital to define the data requirements for decision-tree classification. This includes identifying the necessary data content, formats and sources for initial data collection.

**Data collection**

What occurs during data collection

When collecting data, it is alright to defer decisions about unavailable data, and attempt to acquire it at a later stage.

Lesson Summary

In this lesson, you have learned:  
The significance of defining the data requirements for your model.

Why the content, format, and representation of your data matter.

The importance of identifying the correct sources of data for your project.

How to handle unavailable and redundant data.

To anticipate the needs of future stages in the process.

**Module 2**

**From Understanding to Preparation**

**Data Understanding**

Descriptive statistics-

Univariate statistics: statistics included Hearst, univariates, and statistics on each variable, such as mean, median, minimum, maximum, and standard deviation.

Pairwise correlations: How closely certain variables were related, and which ones, if any, were very highly correlated, meaning that they would be essentially redundant, thus making only one relevant for modeling.

Histograms: understand how values or a variable are distributed, and which sorts of data preparation may be needed to make the variable more useful in a model.

Data quality: Missing values, invalid or misleading values

**Data Preparation - Concepts**

Most time consuming

What are the ways in which the data is prepared

**Data Preparation – Case Study**

Lesson Summary

In this lesson, you have learned:  
The importance of descriptive statistics.

How to manage missing, invalid, or misleading data.

The need to clean data and sometimes transform it.

The consequences of bad data for the model.

Data understanding is iterative; you learn more about your data the more you study it.

**From Modelling to Evaluation**

**Modelling-Concepts**

Modelling – In what way can be visualized to get to the answer that is required

Evaluation – Does the model used really answer the initial question or does it need to be adjusted

Data Modelling focuses on developing models that are either descriptive or predictive.

A descriptive model might examine things like: if a person did this, then they're likely to prefer that.

A predictive model tries to yield yes/no, or stop/go type outcomes. These models are based on the analytic approach that was taken, either statistically driven or machine learning driven.

Predictive modelling – use training set to calibrate model

**Modelling-Case Study**

Type 1 error- false positve

Type 2 error – false negative

**Evaluation**

When and how to adjust the model?

optimal model is the one giving the maximum separation between the blue ROC curve relative to the red base line.

Receiver operating characteristics curve

In this lesson, you have learned:  
The difference between descriptive and predictive models.

The role of training sets and test sets.

The importance of asking if the question has been answered.

Why diagnostic measures tools are needed.

The purpose of statistical significance tests.

That modeling and evaluation are iterative processes.

**Module 3**

**From Deployment to Feedback**

**Deployment**

involves getting the stakeholders familiar with the tool produced

**Feedback**

**Summary**

These steps have included: forming a concrete business or research problem, collecting and analyzing data, building a model, and understanding the feedback after model deployment.

learned methodical ways of working with the data, specifically, determining the data requirements, collecting the appropriate data, understanding the data, and then preparing the data for modeling

learned how to model the data by using the appropriate analytic approach, based on the data requirements and the problem that you were trying to solve

Once the approach was selected, you learned the steps involved in evaluating and deploying the model, getting feedback on it, and using that feedback constructively so as to improve the model.

this methodology are iterative! This means that the model can always be improved for as long as the solution is needed, regardless of whether the improvements come from constructive feedback, or from examining newly available data sources

Using a real case study, you learned how data science methodology can be applied in context, toward successfully achieving the goals that were set out in the business requirements stage.

methodology contributed additional value to business units by incorporating data science practices into their daily analysis and reporting functions

That its purpose is to explain how to look at a problem, work with data in support of solving the problem, and come up with an answer that addresses the root problem.

a methodology can help you solve not only your data science problems, but also any other problem. Your success within the data science field depends on your ability to apply the right tools, at the right time, in the right order, to the address the right problem.

In this lesson, you have learned:  
The importance of stakeholder input.

To consider the scale of deployment.

The importance of incorporating feedback to refine the model.

The refined model must be redeployed.

This process should be repeated as often as necessary.