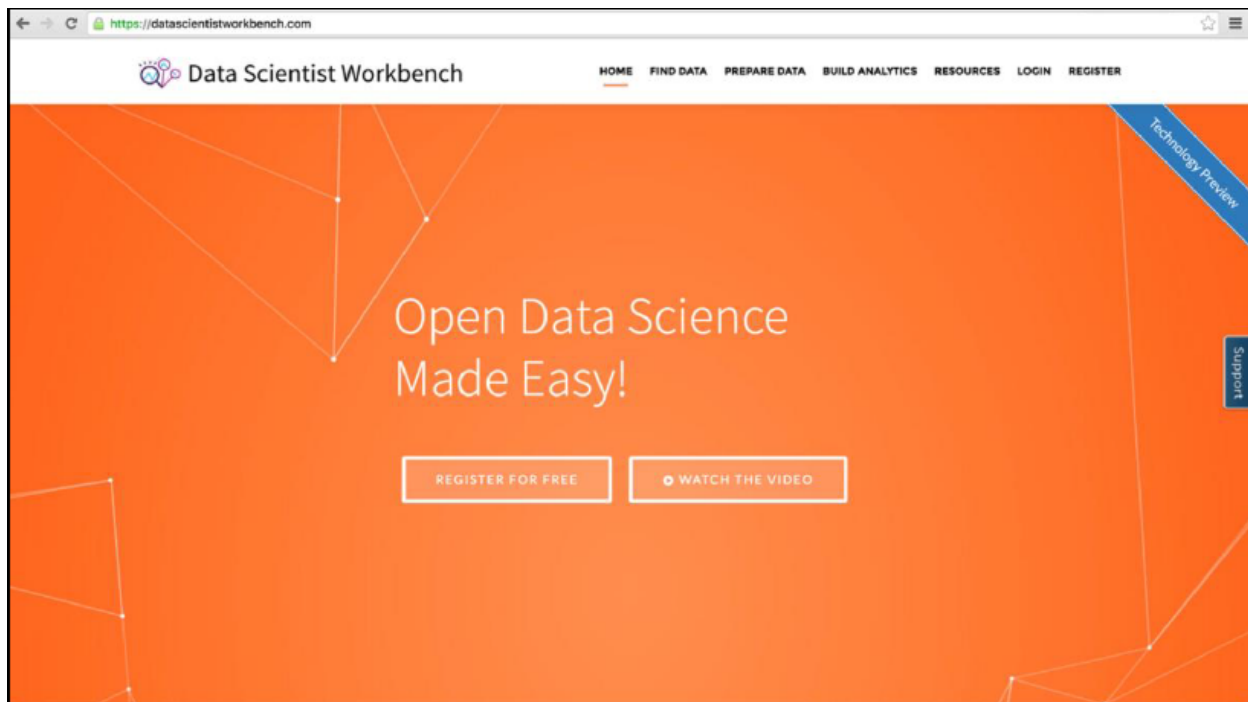




## Lab: From Modeling to Evaluation

### *Hands-on exercises using data on food ingredients*

Normally this hands-on lab would be done using the Data Scientist Workbench (DSWB), which provides a great way of organizing your models, displaying the output, and showing visualizations all in the same interface. However, since this is a fundamentals course without prerequisites, you might not yet be familiar with this tool. Therefore, this lab is mainly theoretical, describing the process you would following if you were utilizing the workbench.



If in future, you are so inclined to learn more about Data Science and the Data Scientist Workbench, you can start by taking the **Data Science Hands-on with Open-Source Tools** course and then come back to this course to complete the two optional hands-on labs that use the DSWB.



## Contents

<b>LAB 1</b>	<b>FROM MODELING TO EVALUATION .....</b>	<b>3</b>
	1.1 FROM MODELING TO EVALUATION.....	4
	1.2 SUMMARY .....	4
<b>LAB 2</b>	<b>OPTIONAL: DECISION TREES AND CLUSTERING USING DATA SCIENCE WORKBENCH.....</b>	<b>5</b>
	1.3 DATA UNDERSTANDING AND DATA PREPARATION .....	6
	1.4 SUMMARY .....	6



## Lab 1 From modeling to evaluation

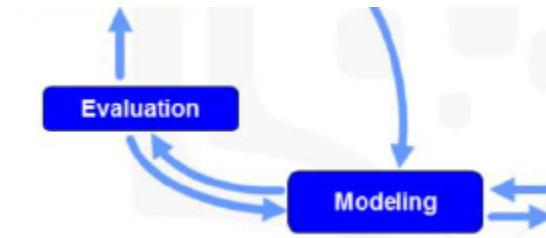
After completing this hands-on lab, you will know:

- What the purpose of data modeling is.
- Some characteristics of the modeling process.
- How to apply modeling to any data science problem (cuisine, in this lab).
- What it means to “evaluate” a model.
- Ways in which a model is evaluated.
- How to apply evaluation to any data science problem (cuisine, in this lab).

Allow 30 minutes to complete this section of the lab.



## 1.1 From modeling to evaluation



### \_\_1. **Modeling** focuses on developing models:

- **Predictive** or **descriptive** models, according to the previously-defined analytic approach
- **Training set** for predictive modeling

The modeling stage focuses on developing predictive or descriptive models according to the previously-defined analytic approach.

With predictive models, data scientists use a *training* set to build the model. A training set consists of historical data in which the outcome is known, so that we can model it.

The modeling process is typically highly iterative as organizations gain intermediate insights, leading to refinements in data preparation and model specification.

For a given technique, data scientists may try multiple algorithms with their respective parameters to find the best model for the available variables.

### \_\_2. **Model evaluation**

- Model evaluation is performed during model development and before model deployment in order to:
  - Understand the model's quality
  - Ensure that it properly addresses the business problem
- Diagnostic measures to be considered:
  - Suitable to the modeling technique used
  - Testing set
  - Refine model as needed

## 1.2 Summary

Congratulations! ...



## Lab 2      OPTIONAL: Decision Trees and Clustering using Data Science Workbench

After completing this hands-on lab, you will know:

- How to model data using decision trees using the Data Scientist Workbench.
- How to test the model(s) using the Data Scientist Workbench.
- How to predict the cuisine of your own recipe, “my\_recipe” using the Data Scientist Workbench.

Allow 30 minutes to complete this section of the lab.



## 1.3 Data understanding and data preparation

- A. Login to your Data Scientist workbench account at [my.datascientistworkbench.com](https://my.datascientistworkbench.com). If you have not signed up yet, go to [datascientistworkbench.com](https://datascientistworkbench.com) and watch the video; then click the SIGN UP FOR FREE button to create your account.
- B. Upload the Jupyter Notebook “*DS0103ENData Science Methodology 101 Module 4 Hands-on Lab using R in DSWB*” to your data folder to your Data Scientist Workbench.
- C. Click on **Jupyter Notebook** and follow the instructions in the uploaded notebook (from step B): execute each of the code cells in the Notebook and review the results.
- D. These are the steps within the Jupyter Notebook:
  1. Import recipes.csv into R.
  2. Model data using decision trees
    - i. Import the required libraries
    - ii. Display the decision tree
    - iii. Summarize the decision tree
    - iv. Create refined decision trees
  3. Testing our model(s)
  4. Predict the cuisine of your own recipe, “my\_recipe”

## 1.4 Summary

Congratulations! ...