**VerticaPy Lesson Series**

Teaching data science concepts using VerticaPy

# Introduction

This document explains how to use the Jupyter notebook templates to create brief lessons on data science concepts. The idea is to have a unified layout so that it looks professional, and it is easy to modify before making it online. The goal of these lessons is to enable and attract users to VerticaPy. There are three core type of pages:

1. Course Page

This page contains a video to excite the users, as well as some basic details like the prerequisites and the goals of the lesson.

1. Module Outline Page

This page contains the list of the lessons including exercises, if any.

1. Lesson Pages

There could be multiple lesson pages to break down a complicated topic. One way to break down could be to separate theory from application. Additionally, it will be ideal to have one Exercise page where the audience is challenged to try out some sample problems.

**Prerequisites:** ipywidgets (7.6.5 preferred), Ipython, and voila (0.3.6 preferred).

Note: The current version of Vertica Demo does not have the above prerequisites so it may be better to create your own environment. This means making the lesson on your own environment and running the actual VerticaPy codes on VerticaPy demo, and lastly cutting snippets of code results and pasting into the notebook.

For all the pages, to edit cells, just double click them. To add more cells, copy the type of cell and paste them wherever needed. This will ensure consistency in font and heading styles.

**How to view:** To view the webpage layout, open the notebook, click the *voila* button on the top ribbon. For the links to work, their references need to be updated.

One set of example pages is already placed in the folder: Data Science Essentials/Linear Regression.

# Details

Now let us list down some further details of the each of those pages.

## Course Page

**Video:** Next, an attractive video should be created to lure in the users. The instructions of how to create the video will be shared separately. This video can be stored in the Figures folder to be accessed. The name of the video should be “Video\_1.mp4”.

**Briefly:** Add a brief description of the entire course focusing on its application.

**Highlights:** The most relevant and simple words should be put here.

**Difficulty and Time:** You can add time in minutes or hours. Difficulty levels are: Easy, Intermediate, Hard.

**Prerequisites:** Add most suitable prerequisites. Use bullets if multiple.

**Goals:** Be concise in writing the goals.

**Modules:** List down all the lessons. Add hyperlink to link between different lessons. To get hyperlinks of respective pages, they will first have to be run using voila button. Then copy the web address e.g., “ <http://localhost:8888/voila/render/Documents/Template/Enablement%20Template/Module_v1_2.ipynb>” and replace the existing address. Note that this web address contains “voila”.

## Module-page

Update the lesson title and difficulty level.

Add a video highlighting important aspects of all the lessons in the module. The instructions of how to create the video will be shared separately.

The hyperlinks for the lessons and exercises also need to be updated. That is all for this page.

## Lesson page

There could be multiple such pages depending on the lesson design.

First the **Lesson Name** and **Sub Module Name** need to be updated. The hyperlink for the “Go Back to Main Page” button also needs to be updated so that it directs to the Main Page.

Update the estimated time required to go through the lesson.

Next add a brief intro and explain why this is important. Double click the cell to edit.

**Table of Contents:** Update it and add references to the headers below. To do that you will need to click on header cells and give them unique names in the following format:

<a id='CELL\_NAME'></a>

And reference them using:

text (#CELL\_NAME)

Next, explain what will be learnt in the lesson in bullets. Double click to edit.

**General Writing:** Copy, edit and delete the cells as per requirement. You can use Markdown to write equations. Copy and paste images in separate cells so that they can be easily references.

There are sample Knowledge Check which can be used to enable the reader to interact. To create a multiple-choice question, a function (create\_multipleChoice\_widget) has been created which can be easily called.

To add code snippets in Markdown, use the following format:

‘’’Python

\*INSERT CODE HERE\*

‘’’

Video can also be added if needed. Just need to update the reference of video path.

At the end of the lesson, mention author’s name and contact information.

The last part of the lesson are citations.

**Exercise Pages:** A special type of lesson are exercise pages which the reader can use to challenge themselves. For this part, the readers could be provided a data file which they can ingest and based on that answer some specific questions. Currently there are two types of questions available: (1) multiple choice, and (2) numeric. Functions have been created which can be easily called to create questions and identify the right answers.

## Nomenclature

The naming of the different pages should follow the given nomenclature:

CourseName\_ModuleName\_LessonName

e.g.,

Essentials\_LinearRegression\_Theory

For each module, separate folder should be made inside the Course folder.

## Folders

All the Jupyter notebook pages should be placed inside their respective Module folder, which will be inside the Course folder. The hierarchy is as follows:

Course folder > Module Folder > All the pages

As an example, three pages of the module Linear Regression have already been placed inside the Data Science Essentials folder.

The additional folders “Figures” and “Data” also need to be placed inside the respective module folder. Some figures such as the icons inside the template are essential and so these images should not be deleted from the folders.

# Curriculum

Data Science is a vast field with lots of rabbit holes, so we want to stick to the most common ones initially. Anyone is welcome to suggest a course curriculum. To start off, below is the general layout of the first Course: Data Science Essentials.

(Note: Each bullet is a module and sub-bullet is a lesson.)

**Data Science Essentials**

* *Overview of Data Science*
  + Basic terminologies
  + Datasets
  + Vdataframe

* *Basic data preparation*
  + Formats of data (csv, image, text etc)
  + Basic operations
    - Impute
    - Null or missing values
    - Normali ze
    - Concatenate and Transform
  + Advanced operations
    - Outlier detection
  + Test/Train split
* *Basic data exploration*
  + Visualizations
    - Different types of plots
  + Dimension reduction (TSNE, PCA)
* *Advanced statistics*
  + Hypothesis Testing
  + Bootstrapping
  + Bayes Rule

* *Linear Regression*
  + Theory
  + Example
  + Exercise
* *Classification*
  + Theory
  + Example
  + Exercise
* *Project*