Getting started with Python, Jupyter Notebook & Turi Create

It's important to emphasize that this specialization is **not** about providing training for a specific software package. The goal of the specialization is for your effort to be spent on learning the fundamental concepts and algorithms behind machine learning in a hands-on fashion. These concepts transcend any single package. What you learn here you can use whether you write code from scratch, use any existing ML packages out there, or any that may be developed in the future.

The learning approach in this specialization is to start from use cases and then dig into algorithms and methods, what we call a *case-studies approach.* We are very excited about this approach, since it has worked well in several other courses. The first course is focused on understanding how ML can be used in various cases studies, and the follow on courses will dig into the details of algorithms and methods for each of the main ML areas. In the first course, you will not be implementing algorithms from scratch, but rather building intelligent applications that use ML. In the subsequent course, we will be implementing and comparing a wide range of algorithms. To make it easy to implement the use cases we will be covering, we are recommending a particular set of software tools, but you can successfully complete the course with other tools out there.

Why Python

In this course, we are going to use the Python programming language to build several intelligent applications that use machine learning. Python is a simple scripting language that makes it easy to interact with data. Furthermore, Python has a wide range of packages that make it easy to get started and build applications, from the simplest ones to the most complex. Python is widely used in industry, and is becoming the *de facto* language for data science in industry. (R is another alternative language. However, R tends to be significantly less scalable and has very few deployment tools, thus it is seldom used for production code in industry. It is possible, but highly discouraged to use R in this specialization.)

We will also use the Jupyter Notebook in our videos. The Jupyter Notebook is a simple interactive environment for programming with Python, which makes it really easy to share your results. Think about it as a combination of a Python terminal and a wiki page. Thus, you can combine code, plots and text to explain what you did. (You are not required to use Jupyter Notebook in the assignments, and should have no problem using straight up Python if you prefer.)

Why SFrame & Turi Create

There are many excellent machine learning libraries in Python. One of the most popular one today is [scikit-learn](http://scikit-learn.org/stable/" \t "_blank). Similarly, there are many tools for data manipulations in Python; a popular example is [Pandas](http://pandas.pydata.org/). However, most of these tools do not scale to large datasets, including some we will tackle in this specialization. In addition, in this specialization, we will cover a wide range of ML models, feature engineering transformation, and evaluation metrics. With most existing packages, you will have to install a combination of packages to get the tools that we need to tackle the use cases in this course. This is possible, but requires advanced knowledge of Python, which we feel will slow down most people's learning of the core concepts.

The main goal of this course is to learn core ML concepts, not how to use a specific software package. Thus, in this course, we recommend you use Turi Create, a package we have been working on for many years now, and has seen an exciting adoption curve, especially in industry with folks building real applications. Turi Create is a highly scalable machine learning library for Python, which also includes the SFrame, a highly-scalable library for data manipulation. A huge advantage of SFrame over Pandas is that with SFrame, you are not limited to datasets that fit in memory, which allows you to deal with large datasets, even on a laptop. (The SFrame API is very similar to Pandas' API. [Here is a doc showing the relationship between the two of them.](https://turi.com/learn/translator/))

The reason we suggest you use Turi Create is because we very strongly believe using this software will make it much easier for us to follow the "case-study approach" we are taking in this specialization. In particular, it will let you focus on exploring each case study in this first course, without having to implement your own algorithms from scratch, and benefiting from the performance advantages that Turi Create provides. ***In subsequent courses in the specialization, you will be implementing many of these algorithms from scratch, having had the foundation of seeing them perform in practice on real applications.***

Licenses for SFrame & Turi Create

Turi Create and the SFrame package are available in [open-source under a permissive BSD license](https://github.com/turi-code/SFrame). So, you will always be able to use Turi Create and SFrames for free.

We are happy, however, for you to use any tool(s) of your liking, by following the steps below. As you will notice, we are only grading the output of your programs, so the specific software tool is not the focus of the course.

It's important to emphasize that this specialization is **not** about providing training for a specific software package. The goal of the specialization is for your effort to be spent on learning the fundamental concepts and algorithms behind machine learning in a hands-on fashion. These concepts transcend any single package. What you learn here you can use whether you write code from scratch, use any existing ML packages out there, or any that may be developed in the future. We are happy to hear that so many of you are enjoying this approach so far!

Using other ML packages

We strongly encourage you to use SFrame for this course.

You are welcome to use other ML packages, like [scikit-learn](http://scikit-learn.org/stable/" \t "_blank), instead of Turi Create. However, we believe this will significantly slow down the your implementation tasks, especially for this first course.

The first course is focused on exploring the use cases we'll tackle throughout the specialization. A huge goal here is to familiarize ourselves with the core ML concepts that we will use the 5 follow-on courses. In those course, there will be much more implementation of ML algorithms, so the specific ML package becomes less important. But, in this first course, we want to move quickly through all the use cases, and Turi Create will help us do just that.

If you choose to use a different package, we will provide the data sets and the assignment questions will not depend specifically on Turi Create.

Learning outcomes

This reading will walk you through the steps you will need to follow to install and get started with Python, Jupyter Notebook, and Turi Create.

* Installing Python, Jupyter Notebook, and Turi Create
* Starting Jupyter Notebook
* Writing variables, functions and loops in Python
* Doing basic data manipulations in Python with SFrames

**Getting started using these resources**

Downloading and installing Python, Jupyter Notebook and Turi Create on your own machine

1. If you do not already have Python installed, download and install Python 3.7: <https://www.python.org/downloads/>.
2. Download and install Jupyter Notebook: <http://jupyter.org/install>. Follow the instructions for "Installing Jupyter with pip", use the commands under the section for Python 3
3. Download and install Turi Create: <https://github.com/apple/turicreate#installation>. **Note: it is not required that you use virtualenv, but it might be helpful, especially if you run into installation issues due to conflicting versions of software.**

Other resources

* There are many Python resources available online. [Here is a good place for documentation](https://docs.python.org/3/).
* For Turi Create, there is also a lot of information available online. Here are some starting points.

|  |  |
| --- | --- |
| The User Guide | <https://apple.github.io/turicreate/docs/userguide/> |
| More Detailed API Docs | <https://apple.github.io/turicreate/docs/api/> |

Watch the videos on getting started with Jupyter Notebook and SFrames

If you haven’t done so yet, before you start, we recommend you watch the videos where we go over Python, Jupyter notebook and SFrames using Turi Create.

Download the data and sample code and familiarize yourself with the notebooks

Before doing the assignments in this course, familiarize yourself with the two notebooks we covered in the videos:

* Download the notebook that covers getting started with Python:

[Getting started with Jupyter Notebook.ipynb.zip](https://d18ky98rnyall9.cloudfront.net/cyoFD-efEeiaxBKyA9PBAg_737ba720e79f11e8b282118f44bfde12_Getting-started-with-Jupyter-Notebook.ipynb.zip?Expires=1597968000&Signature=JbdbKqATzBZS00mTV3Y5-l5p2KzTZ2rSlgCXwcI2rXtjNLAtYsyC1g59J4SpcdHLF2Qfwu~W-5BtPqhbbQEO7v9XWybFlDyeoyY4ngYTM0H9RkPKPgyGPJuVDZn9PIktIiYayzW6aQWvdR544aDVE8MRxHPmVrBGVNjSrE~-hfI_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

* Download the notebook that covers getting started with SFrames:

[Turi Getting Started with SFrames.ipynb.zip](https://d18ky98rnyall9.cloudfront.net/cyqhUOefEeiaxBKyA9PBAg_737fc5d0e79f11e8bc932f9e6898fd60_Turi-Getting-Started-with-SFrames.ipynb.zip?Expires=1597968000&Signature=Xp2pDQr65LX9zpk5x6U70sOt24HSUK82hpU6htiikTl9i-8sxwlF4LopFNUnu4i3XoKQKh3SjFT3hZ~aiqNj3mK4OwM9zplpmYZWBVGUp7dzCtm3d~4Szk8nvsSzdLhzMsCvQh-q7L7G-LM2KyB88lOrf84ILYAXfsEM3~Y1fVw_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

* Download the simple people dataset: [people-example.csv](https://d396qusza40orc.cloudfront.net/phoenixassets/course1-for-students/people-example.csv)
* Save all these files in the same directory (where you are calling Jupyter notebook from). **Not sure where to save the files? See**[**this guide**](https://www.coursera.org/learn/ml-foundations/supplement/kRD6B/where-should-my-files-go)**.**

Familiarize yourself with the notebooks

Make sure:

* You’ve downloaded and installed Python, Jupyter Notebook, and Turi Create.
* Started up Jupyter Notebook from the directory you downloaded the files above, by typing

jupyter notebook

in your command line (for Mac/Linux) or Command Prompt (for Windows).

Now you are ready to get started! Familiarize yourself with Python and SFrames, as well as writing code with the Jupyter Notebook. From here, you will be ready to do all the assignments in the course, and build awesome intelligent applications that use machine learning

Where should my files go?

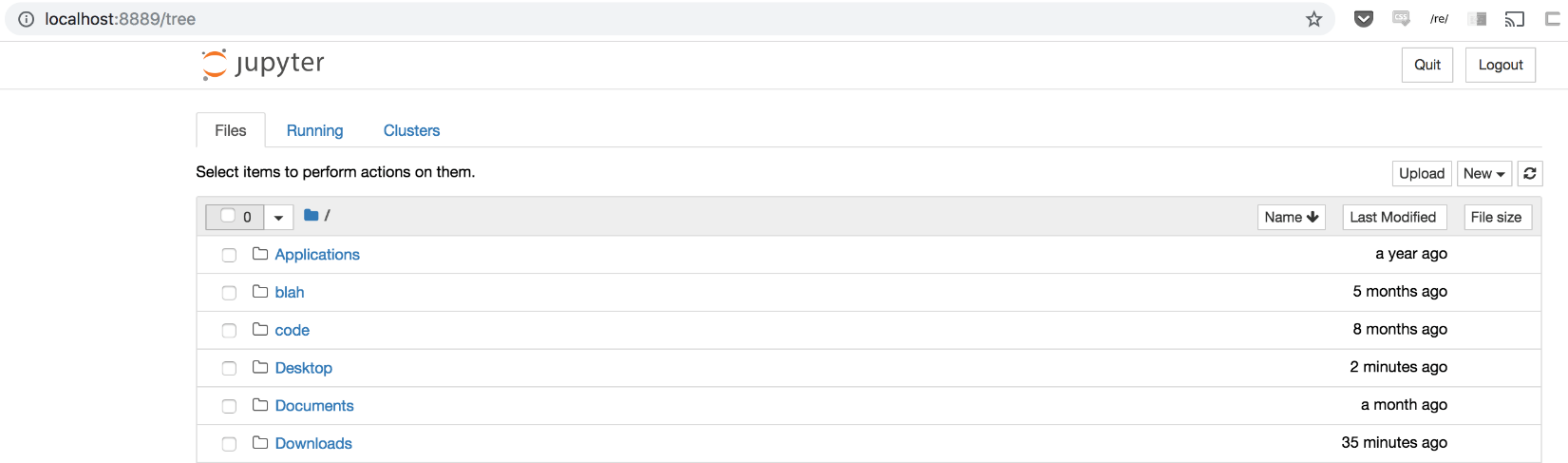
In demos and programming assignments, you will be asked to download notebooks and datasets. **This guide will show you where to place those files, so that IPython notebook can find them.**

Option 1: Place your folders and files where Jupyter can find them by default

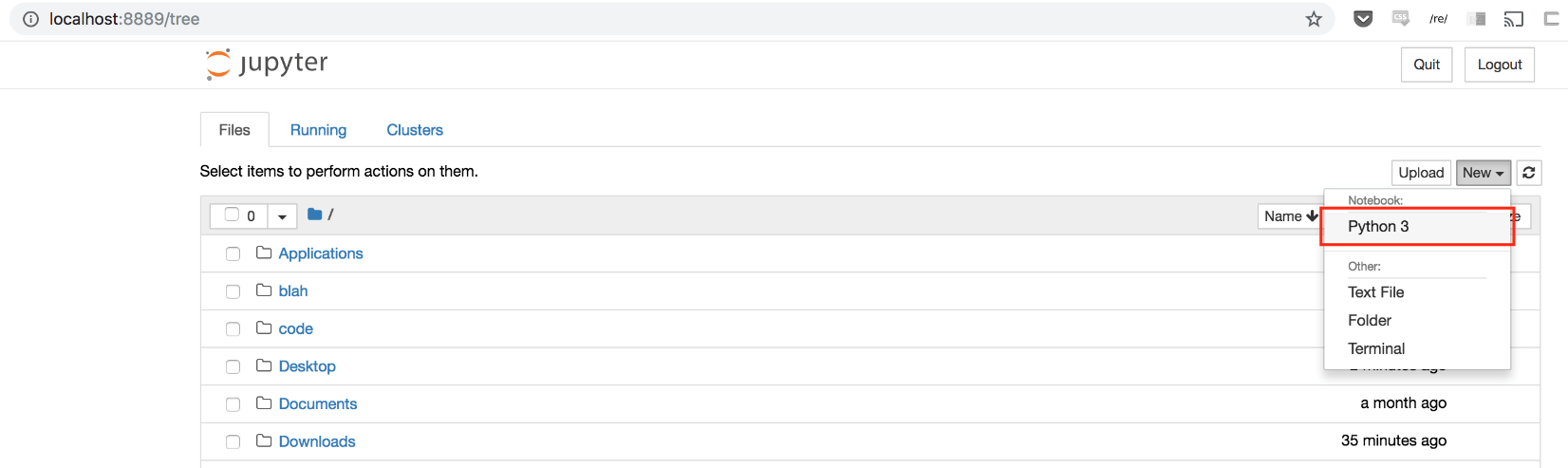
1. Launch the Jupyter Notebook by typing

jupyter notebook

in the command line (Mac & Linux) or Command Prompt (Windows). You should be greeted in the web browser with the main page:



2. From the top right, find the button labeled "New▾". Click the button to get a drop-down menu, and select "Python 3" under the sub-heading "Notebook" This should create a new notebook inside the home directory of Jupyter notebook.



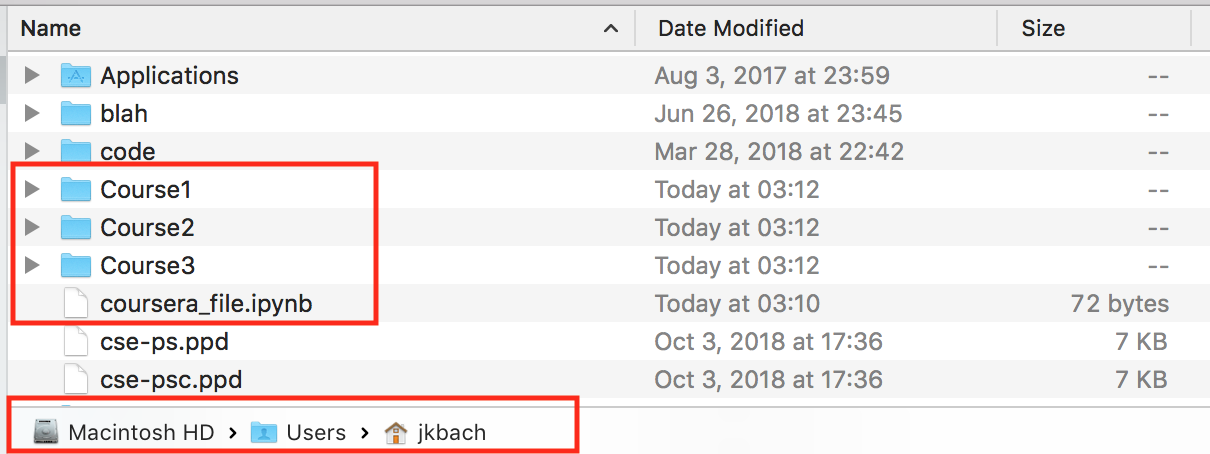
3. In the new notebook, run

import os

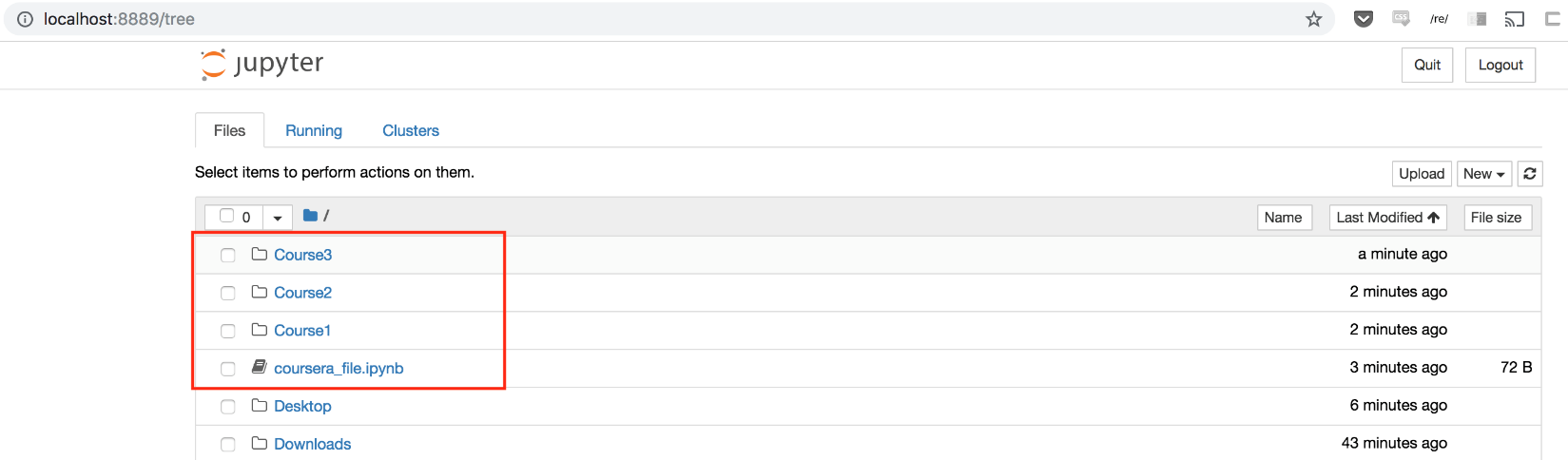
print os.getcwd()

to obtain the full path of the **home directory of Jupyter notebook**. **This path is where your files should go.** Highlight the path and copy it.

4. Place any files (notebooks and datasets) under this home directory using your file browser. You may organize your files using sub-folders.



5. All files and folders placed inside the home folder will appear in the main page:



Option 2: run Jupyter Notebook from where the files are

If you feel comfortable moving around the file system via command line, you may also choose to store your files wherever you like, then simply navigate to that directory and run the notebook there. For example, on command line in Mac/Linux, this looks like:

cd <directory\_where\_my\_files\_are\_stored>

jupyter notebook

This option has the advantage of reducing clutter in your home directory.