

Tutorial and Intro to Data Analytics with Python

In this brief tutorial, we will look at the use of the pandas library in Python for data analytics. We'll first use the `mysql.connector` library to query and retrieve data from an SQL server, then load that data into a pandas dataframe, and ultimately plot the data. *Note: This tutorial is written in and for Python 2.7. It has not been tested with any other version of Python.*

This project is based on an energy-analytics project, which means that we will query energy related data. More specifically, we querying the energy consumption of solar hybrid mini split air conditioner used for produce cooling. You will have access to both temperature and energy data at a temporal resolutions of 15 minutes.

Have fun exploring.

Software preparation

To do this tutorial, you will need to have some software and libraries installed on your computer. If you'd rather use a Unix box on Cloud9, you can skip this step read the [Using Cloud9](#) instructions instead. The following instructions are for Unix-based systems (including Mac). Please refer to the online documentation for doing this in Windows (Sorry, Python is the reason why I traded Windows one night for Arch Linux a few years ago before moving over to Mac.)

You need to make sure that you have the following software/libraries (and their dependencies) installed on your machine:

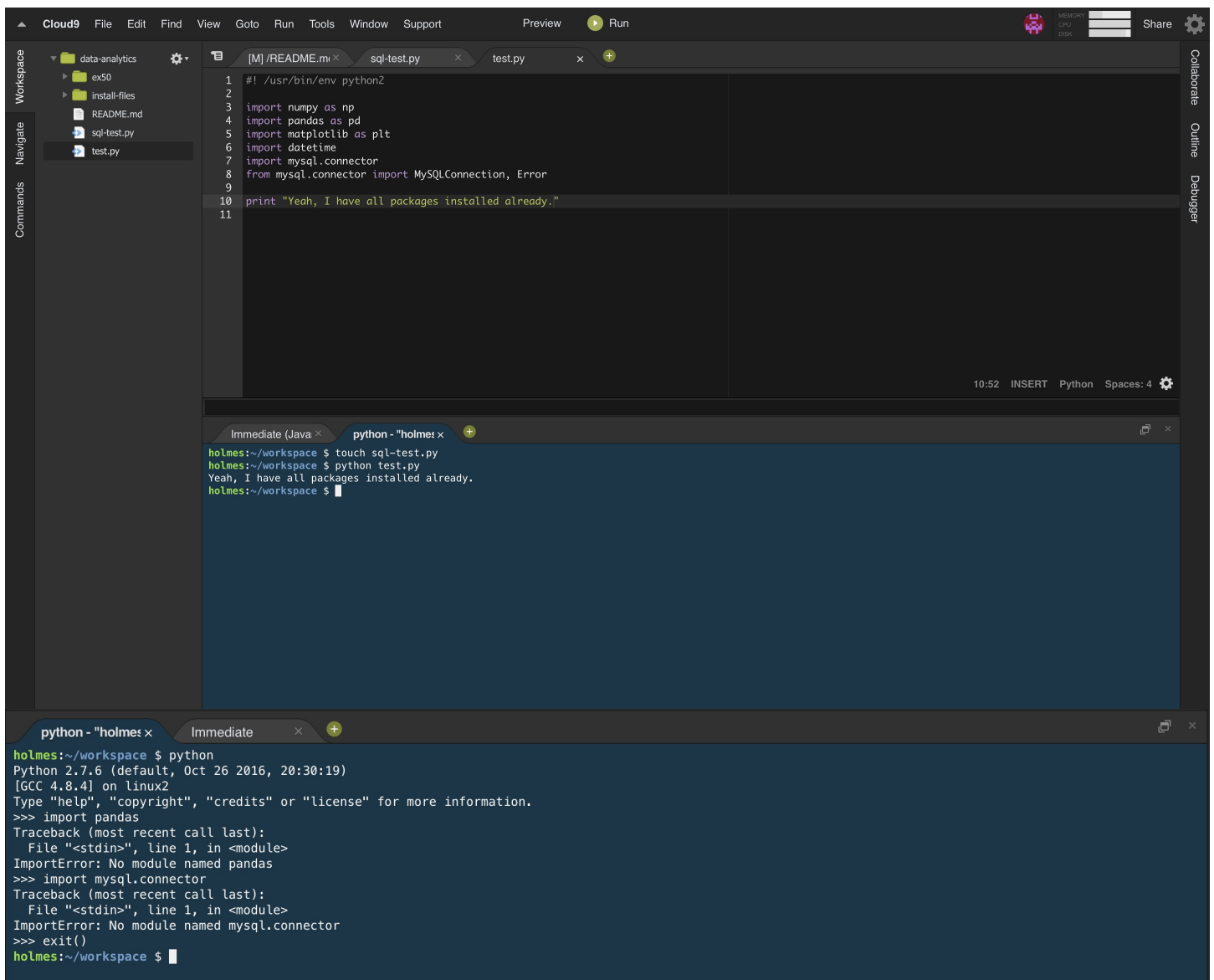
- [Python 2.7](#)
- `pip`
- `numpy` --- `pip install numpy`
- `pandas` --- `pip install pandas` or `conda install pandas` (if you previously install [conda](#))
- `matplotlib` --- `pip install matplotlib`
- `mysql.connector` --- `pip install mysql-connector==2.1.6`

If you're on a Mac, my recommendation is to install Homebrew, then `brew install python2`, and then `pip2 install` instead of `pip install`. Again, there are so many instructions online that are fantastic.

Importing libraries in Python

To test if you have everything, open your Python shell (it should say `Python 2.7.x`), and run the following commands (this is how you import libraris). Hopefully you will not see any error messages.

```
import numpy
import pandas
import matplotlib
```








Using Cloud9

In order for you to use the preconfigured cloud9 environment, email hsmidt@hawaii.edu with the email account that you want to use. You will then receive an invitation of join **Team stemd2** on cloud9. Follow the link and sign up for an account if you don't have one yet.

Once you sign up, you want to click **join team**. You should be redirected to something that looks like the first image below. You then want to click on the stemd2 link on the navigation pane on the left, which should take you to the page shown in the second image below.

9






Holm

Workspaces


Shared With Me

Repositories

YOUR TEAM SUBSCRIPTIONS


STEMD2

YOUR INDIVIDUAL SUBSCRIPTIONS


Free

Welcome Holm!

Start coding now by creating a workspace below. Please don't forget to **set your password** from the email we sent to **holm.smidt@gmail.com**, to prevent losing access to your account!





Filter...


Workspaces



Create a new workspace

9






PRO TEAM **STEMD2**

Workspaces

Welcome Holm!

Start coding now by creating a workspace below. Please don't forget to **set your password** from the email we sent to **holm.smidt@gmail.com**, to prevent losing access to your account!



holmes

stemd2/data-analytics

python workspace

Python box for data analytics

Updated a few seconds ago.



1 CPU

512MB RAM

2GB HDD

You then want to clone the only workspace that you can see.

Create a new workspace

Workspace name

data-analytics-cloned

Description

Python box for data analytics

Team

STEMD2

Hosted workspace

[Clone workspace](#)

Remote SSH workspace

Salesforce

Clone from an existing workspace

Cloud9 will create an exact copy of your workspace. This includes tab-states, project settings and visibility settings.

Clone workspace:

holmes/data-analytics

You are currently cloning workspaces from your team STEM2. To clone your own workspaces, unset the value above.

Create workspace

9

Cloud9

Restoring data from hibernation.

Waking up your workspace from hibernation

What's going on here?

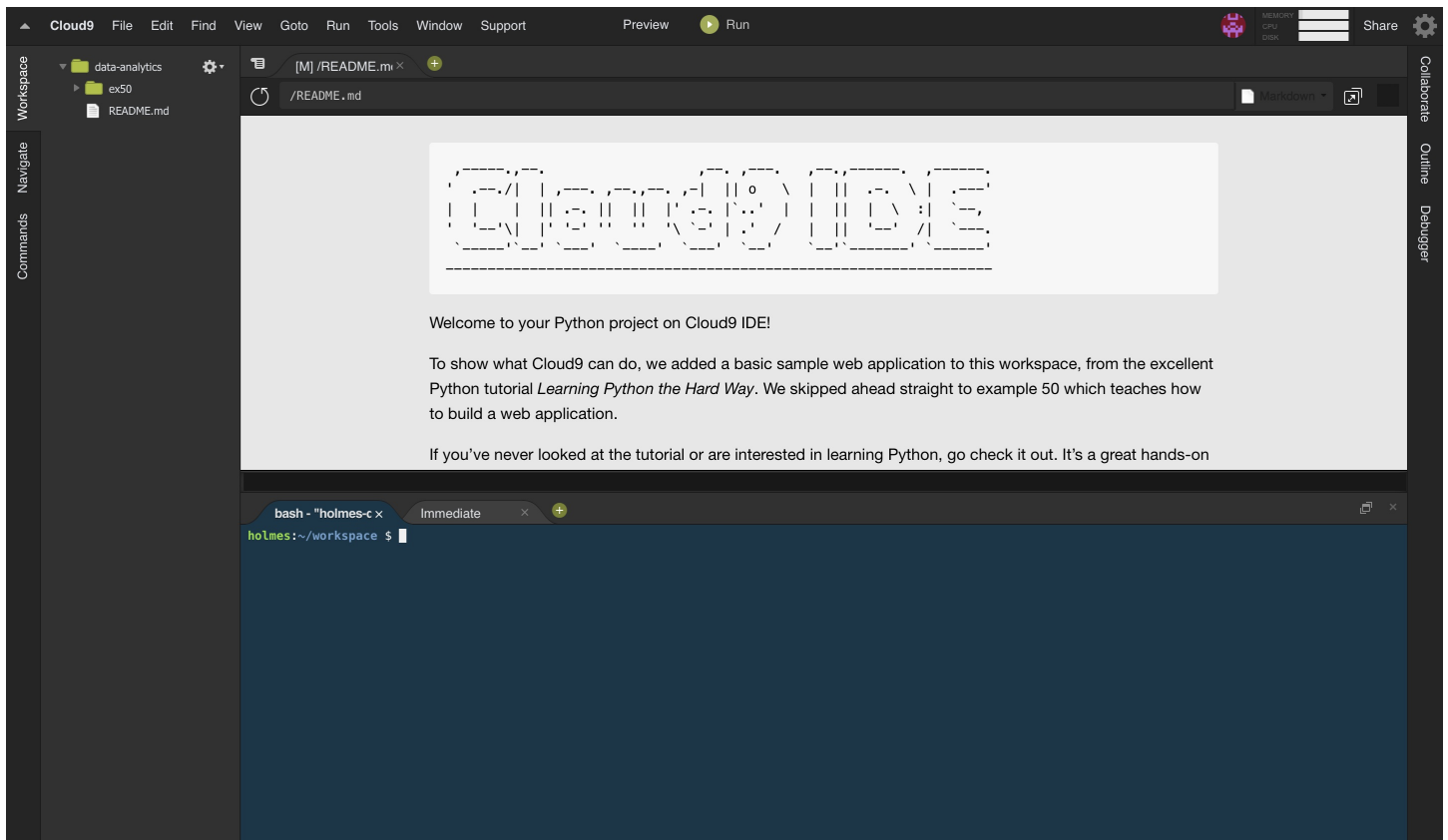
We're migrating your premium workspace to a new server to ensure optimal performance.

Please wait a moment while we move your workspace. It will be just as you left it.

Once you're being redirected to the cloud9 workspace, you should see something like shown below. You can create files by right-clicking on the left and selecting **New File** and then open it in the editor page. Once you save a file, you can execute it by typing `python test.py` in the terminal.

Page 4/5

Friday, Dec 15, 2017, 11:39 AM by Holm Smidt



You can start a python shell by typing `python` into your terminal. To run a python scrip, type `python test.py` into your terminal.

Starter Documents

To work through the tutorial, you should download the provided archive from Google Classroom or [Github](#) and use the provided files in the cloud9 directory. The file structure should be as follows:

```
- output/  
- config.ini  
- energyAnalyzer.py    <-- this is the tutorial file  
- sqlUtils.py
```

Tutorial

The Python file steps you through the analysis process. It is separated in to three section:

- data retrieval using SQL
- data processing using pandas data frames
- data visualization using the matplotlib and pandas Library

Throughout the Python script, you are asked to make changes to existing code or some code.