

Lab 1: Environment Setup & Data Handling

Part 2 of this assignment is due to Gradescope at the end of lab (5p ET). You may work with others in your breakout room, but you must submit your own solution.

Introduction

The purpose of this lab is to get you started running Jupyter notebooks with the Python libraries we will use throughout the semester. You will also get some practice loading and analyzing data – a prerequisite for training machine learning models.

Provided Files

- [Lab1.pdf](#): This file
- [Lab1.ipynb](#): Jupyter notebook with Part 2 instructions
- [cta-ridership.csv](#): Data for exploration

Part 1. Environment Setup

The first part of this lab involves preparing your programming environment. The following sections provide links to online installation instructions. You may already have some of these libraries installed, in which case you do not need to reinstall, but you may want to update to the most recent version.

Python 3

We will be using Python 3 for all class and lab programming. You can download and install Python 3 here: <https://www.python.org/downloads/>.

JupyterLab

JupyterLab is a browser-based interface for interactive Python notebooks. Jupyter notebooks are cell-based, meaning that your code (or text markup!) can be subdivided into cells that can be run in any order. The notebooks are backed by a Python kernel that maintains state between cell executions. Jupyter Notebooks are a very common programming environment for machine learning and data science. The cell-based execution and in-place plotting makes data exploration particularly convenient. We will be using Jupyter Notebooks for some labs and most in-class exercises.

You can download and install JupyterLab with `pip3 install jupyterlab`, instructions here: <https://jupyter.org/install>. I recommend using this method instead of installing with `conda` (Anaconda), but you can choose whichever you prefer.

Once you have JupyterLab installed, try opening it with the `jupyter-lab` command and playing around with the interface.

Additional Python Libraries

Install the following Python libraries:

- **NumPy.** `pip3 install numpy`. Documentation: <https://numpy.org/>
- **SciPy.** `pip3 install scipy`. Documentation: <https://docs.scipy.org/doc/scipy-1.6.0/reference/>
- **Pandas.** `pip3 install pandas`. Documentation: <https://pandas.pydata.org/>
- **Excel Read.** `pip3 install xlrd`. Documentation: <https://pypi.org/project/xlrd/>
- **Matplotlib.** `pip3 install matplotlib`. Documentation: <https://matplotlib.org/>
- **Seaborn.** `pip3 install seaborn`. Documentation: <https://seaborn.pydata.org/>
- **TensorFlow.** `pip3 install tensorflow`. Documentation: <https://www.tensorflow.org/>
- **Keras.** `pip3 install keras`. Documentation: <https://keras.io/>

Part 2. Data Handling

The second part of this lab involves practice importing, manipulating, and plotting data using the provided `Lab1.ipynb` file. You should open and work through this file in JupyterLab.

Deliverables

Upload the following files to Gradescope:

- Completed `Lab1.ipynb`

Extra Credit Opportunity

If you find a bug anywhere in this lab, please inform Prof. Aphorpe. The first student (or breakout room) to find any particular bug will be given a small amount of extra credit. This will help make the course better for students in future years.