Great Expectation Tutorial!

**Step 1: Setup**

**Create a Virtual Environment**

Python3 -m venv venv

source venv/bin/activate

**Install dependencies**

Pip3 install great\_expectations

If there a package in red with version error please install the specified version of that package and reinstall great expectation

Pip3 install mistune=0.8.4

**Check the Version**

great\_expectations –version

confirm it with pip3 list

### Create a Data Context

The folder structure that contains the entirety of your Great Expectations project.

great\_expectations init

This will create a folder in your directory which will store everything about your assert statements and expectations

Congratulations, that's all there is to Step 1: Setup with Great Expectations. You've finished the first step!

### Step 2: Connect to Data

### Create a Data source with the CLI

A connection to the data directory we've provided in the repo. This could also be a database connection or AWS or GCP

In this we will have a csv file so we will use a simple data storage

great\_expectations datasource new

### Select appropriate options shown on the screen

### For this tutorial we will using” Files on a filesystem” and “Pandas”

This will now **open a new Jupyter Notebook** to complete the Data source configuration. Your console will display a series of messages as the Jupyter Notebook is loaded, but you can disregard them.

In the future, you can modify or delete your configuration by editing your great\_expectations.yml and config\_variables.yml files directly.

### Step 3: Create Expectations

[Expectations](https://docs.greatexpectations.io/docs/terms/expectation) are the workhorse abstraction in Great Expectations.

* [**Expectation Suite**](https://docs.greatexpectations.io/docs/terms/expectation_suite): A collection of Expectations.
* [**Profiler**](https://docs.greatexpectations.io/docs/terms/profiler): A tool that automatically generates Expectations from a [Batch](https://docs.greatexpectations.io/docs/terms/batch) of data.

great\_expectations suite new

After doing this you will be asked a couple of questions.

For this tutorial we will be using the profiles which is made by GE in order to quickly setup some expectations

When you answer with Y (or just press enter) Great Expectations will **open a Jupyter Notebook** that helps you populate the new suite.

If you want to edit your suite

great\_expectations suite edit Lab3Suite

This will open a new jupyter notebook and let you add or delete stuff

### Step 4: Validate Data

You created a Checkpoint which you used to validate new data. You then viewed the Validation Results in Data Docs.

A Checkpoint runs an [Expectation Suite](https://docs.greatexpectations.io/docs/terms/expectation_suite) against a [Batch](https://docs.greatexpectations.io/docs/terms/batch) (or [Batch Request](https://docs.greatexpectations.io/docs/terms/batch_request)).

great\_expectations checkpoint new getting\_started\_checkpoint

This will **open a Jupyter Notebook** that will allow you to complete the configuration of your Checkpoint.

Edit the data\_asset\_name to reference the data we want to validate.

In order to build [Data Docs](https://docs.greatexpectations.io/docs/terms/data_docs) and get your results in a nice, human-readable format, you can simply uncomment and run the last cell in the notebook.

**Congratulations! You have now completed the “Getting started with Great Expectations” tutorial.**