# MLflow and Hydra Notes

## Introduction

This document is primarily concerned with giving me a template to run these commands. One that I can apply to new situations. We are going to use the NYC\_Predict\_Rental\_Prices as the example.

### Steps

We are going to run this in steps to start with.

#### download

There is no directory for this. This is a preliminary step. The code was already set up for me. To run this:  
mlflow run . -P steps=download

The code in main.py looks like this.

##### if "download" in active\_steps:

##### # Download file and load in W&B

##### \_ = mlflow.run(

##### f"{config['main']['components\_repository']}/get\_data",

##### "main",

##### version='main',

##### parameters={

##### "sample": config["etl"]["sample"],

##### "artifact\_name": "sample.csv",

##### "artifact\_type": "raw\_data",

##### "artifact\_description": "Raw file as downloaded"

##### },

##### )

This goes to config.yaml and puts together the path to get sample1.csv out of github. This command uploads sample1 as sample.csv:latest to wandb

Now that sample.csv is on wandb, you next download it and look at it in a Jupyter Notebook using this command.

##### mlflow run src/eda

You need to be in the parent directory to do this. It needs to look at the MLproject file at the top and then go down 2 levels to eda and look at that one. I found this would start a very poor Jupyter environment. I did not like working in it and it seemed to have problems saving the code to wandb. I just gave up and saved the code as an artifact.

#### EDA

We are doing a preliminary look at the data with this step. This is in the directory eda. There are no further Hydra or MLflow commands here.

#### basic\_cleaning

You need to be in the root directory of this project. Then you can run cookiecutter.

##### cookiecutter cookie-mlflow-step -o src

##### step\_name [step\_name]: basic\_cleaning

##### script\_name [run.py]: run.py

##### job\_type [my\_step]: basic\_cleaning

##### short\_description [My step]: A very basic data cleaning

##### long\_description [An example of a step using MLflow and Weights & Biases]: Download from W&B the raw dataset and apply some basic data cleaning, exporting the result to a new artifact

##### parameters [parameter1,parameter2]: input\_artifact,output\_artifact,output\_type,output\_description,min\_price,max\_price

The MLflow command for this RAN from the root directory is:

##### mlflow run . -P steps="basic\_cleaning"

There was some work to get all of the run.py, main.py, conda.yaml (2 files basic\_cleaning and root directory), MLflow file (local to basic\_cleaning) in shape. Just follow those as the examples.