# This is your Editor pane. Write the Dockerfile here and

# use the command line to execute commands

$ cd goodwill\_python\_data\_science\_container

$ vim Dockerfile

FROM ubuntu:16.04

RUN apt-get update \

&& apt-get install -y python3-pip build-essential \

&& apt-get -y autoremove \

&& apt-get -y clean \

&& rm -rf /var/lib/apt/lists/\*

RUN pip3 install --upgrade pip

RUN pip3 install tensorflow

RUN pip3 install numpy pandas sklearn matplotlib seaborn jupyter pyyaml h5py

RUN pip3 install keras --no-deps

RUN ["mkdir", "notebooks"]

COPY jupyter\_notebook\_config.py /root/.jupyter/

COPY run\_jupyter.sh /

# Jupyter and Tensorboard ports

EXPOSE 9000 6006

# Store notebooks in this mounted directory

VOLUME /notebooks

CMD ["/run\_jupyter.sh"]

c = get\_config() # get the config object

c.IPKernelApp.pylab = 'inline' # in-line figure when using Matplotlib

c.NotebookApp.ip = '\*'

c.NotebookApp.open\_browser = False # do not open a browser window by default when using notebooks

c.NotebookApp.token = '' # No token. Always use jupyter over ssh tunnel

c.NotebookApp.notebook\_dir = '/notebooks'

c.NotebookApp.allow\_root = True # Allow to run Jupyter from root user inside Docker container

#!/usr/bin/env bash

jupyter notebook "$@"

$ chmod +x run\_jupyter.sh

$ docker build -f Dockerfile -t python\_data\_science\_container

$ docker run -it -p 9000:9000 -p 6006:6006 -d -v $(pwd)/notebooks:/notebooks goodwill\_python\_data\_science\_container

#to add spark, you need to do the following

$ cd ~

$ pwd

/Users/maxmelnick/apps

$ mkdir spark-docker && cd $\_

$ pwd

/Users/goodwill\_python/apps/spark-docker

#execute the following

$ docker run -d -p 9000:9000 -v $PWD:/home/goodwillmumvenge/work --name spark jupyter/pyspark-notebook

#now run the following code to test spark

import pyspark

sc = pyspark.SparkContext('local[\*]')

# do something to prove it works

rdd = sc.parallelize(range(1000))

rdd.takeSample(False, 5)

#to stop spark, enter the following

$ docker stop spark