**Docker:**

Docker is a technology built on containerization. Containerization is a virtual environment without guest OS and hassles of virtualization. Containerization provides isolated environments called containers. Containers are available as images in docker repositories. Downloaded image becomes container when it’s instantiated. Multiple containers can be run in parallel without conflicting with each other. Clients communicate with containers using Rest API.

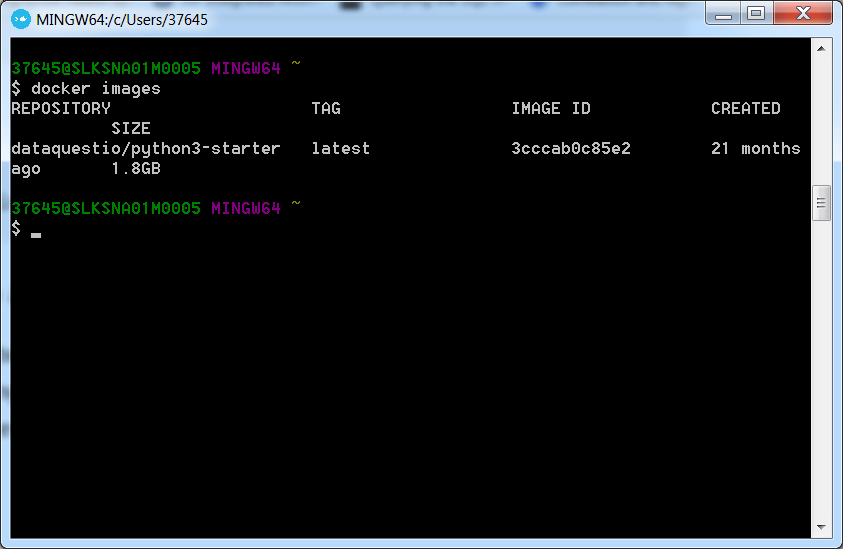
**Docker for data science:**

There are different readymade packages available to kick start data science work without muscling through the setup/installation process. One among them used for assignment 1 is dataquestio/python3-starter, whose image comes with built-in python3, jupyter, and other necessary packages like numpy, sklearn, pandas, etc. If needed, another image for python2 can be downloaded and instantiated by downloading dataquestio/python2-starter. There are also images for MongoDB, MySQL etc, in docker repositories.

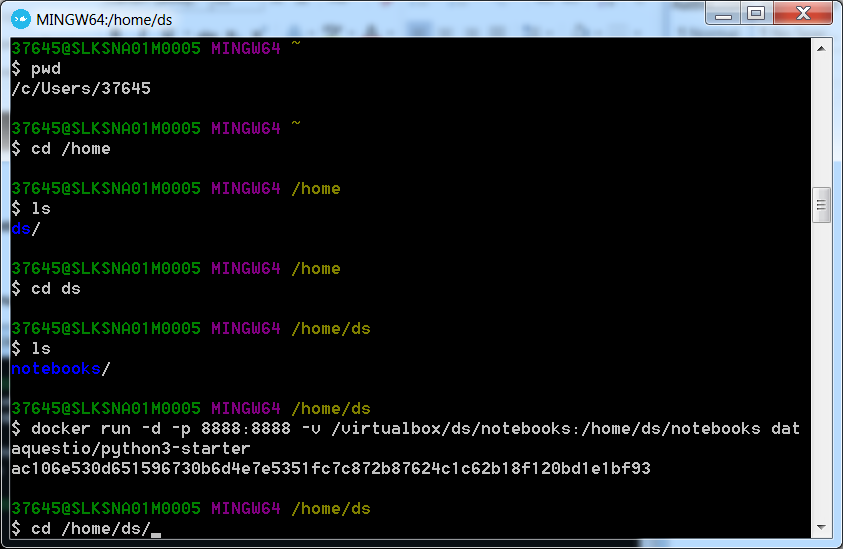
**Docker disadvantages:**

As with any other technologies, there is a learning curve involved with Docker and it can be deeper for people not familiar with UNIX background. Since the technology is built on Linux, it requires a Linux virtual environment on windows, which partly defeats the advantage of containerization.

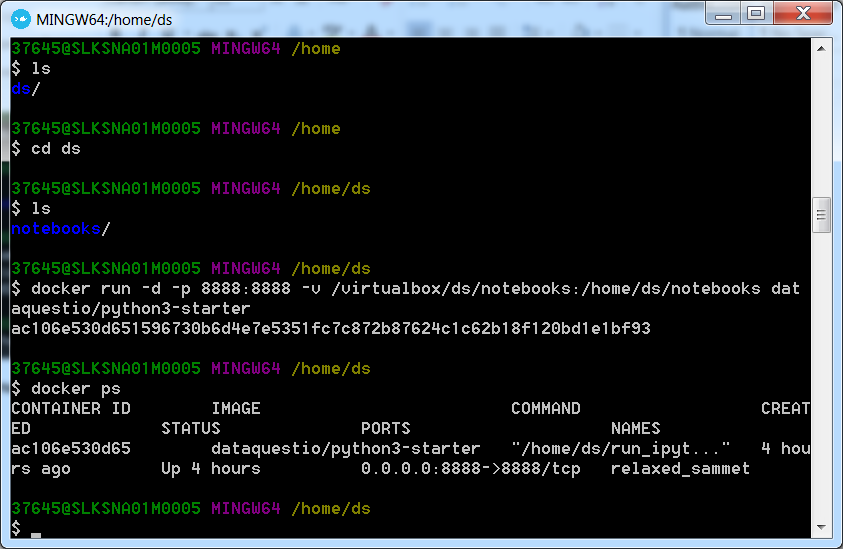
**Docker downloaded image screenshot:**



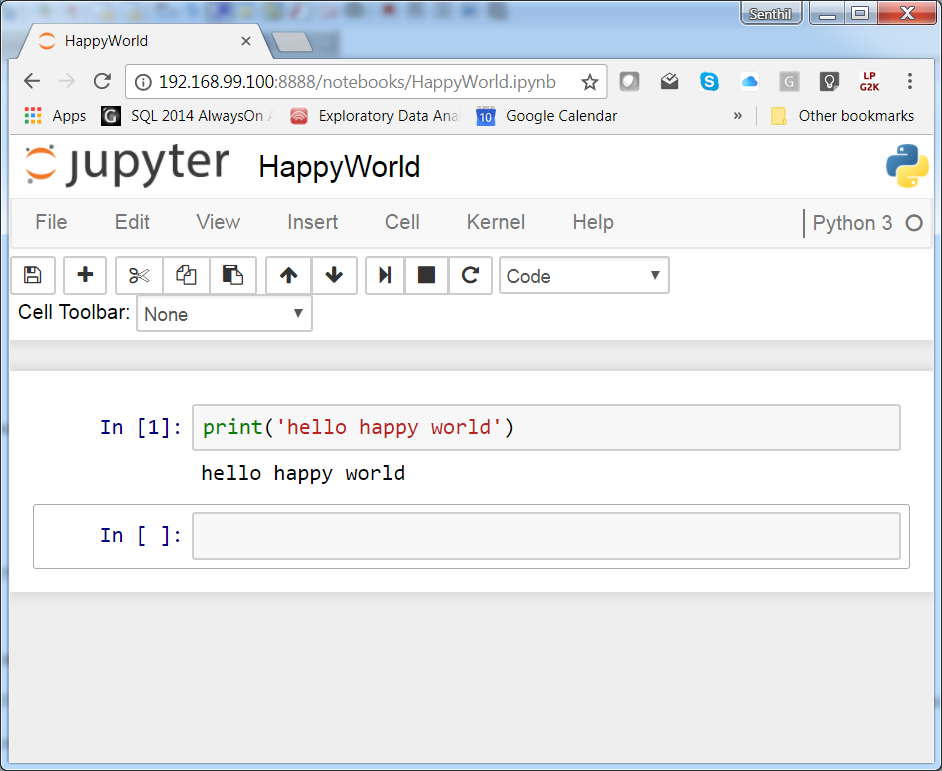
**Docker start command screenshot:**



**Docker running screenshot:**



**Hello World screenshot: Jupyter client communicating with container process**



**Screenshot of blog example:**

