

# Assignment: 1

---

TEAM7: SNIGDHA JOSHI & VIPRA SHAH



This assignment is divided into two parts:

- 1) Data Ingestion
- 2) Data Wrangling

For both docker images perform following steps:

- 1) Pull the image from the dockerhub
- 2) Copy the config.json and configwrangle.json file to your local machine
- 3) Add your AWS access and secret keys, create a bucket called "team7pa\_assignment1" then run the following codes.
- 4) Commit changes
- 5) Execute the code

For detailed description of above steps please refer below:

## Data Ingestion

1. Pull image from docker hub

```
Snigdha@DESKTOP-T23DDG5 MINGW64 ~  
$ docker pull joshisn/assignment1:final  
final: Pulling from joshisn/assignment1  
9f0706ba7422: Already exists  
d3942a742d22: Already exists  
62b1123c88f6: Already exists  
2dac6294ef18: Already exists  
a7bb658fb099: Already exists  
a811de274338: Already exists  
771f11f32dc9: Already exists  
a16d4d6b543c: Already exists  
f0b7951cc55d: Pull complete  
ce2bc6ff5564: Pull complete  
2784ab0b4421: Pull complete  
0c2ba9465c05: Pull complete  
eb42dd3fb11d: Pull complete  
3c4082260b5e: Pull complete  
cc2054b4362f: Pull complete  
b07c492e611d: Pull complete  
14130aa31fd8: Pull complete  
108a648cac70: Pull complete  
fb75714b7629: Pull complete  
326b4a4ed0ff: Pull complete  
18301873a3a8: Pull complete  
cf840d1e52af: Pull complete  
d6f0d1190f67: Pull complete  
Digest: sha256:1efe4239fcf86f4da963abf84922835c68a33e906f5e98d0da37bfb431dfcf1a  
Status: Downloaded newer image for joshisn/assignment1:final
```

2. Create the container

```
Snigdha@DESKTOP-T23DDG5 MINGW64 ~/Documents/ADS/Assignment1/Raw_Data (master)
$ docker create --name="rawdata13" joshisn/assignment1:final
1a0eeb55800aae359b25741e286f4c2f39f75727289fe941c86dff88dfc0296f
```

3. Copy config file to your local machine and edit as mentioned below:

```
Snigdha@DESKTOP-T23DDG5 MINGW64 ~/Documents/ADS/Assignment1/Raw_Data (master)
$ docker cp config.json rawdata13:/Assignment1/
```

4. Start the container

```
Snigdha@DESKTOP-T23DDG5 MINGW64 ~/Documents/ADS/Assignment1/Raw_Data (master)
$ docker start -i rawdata13
/Assignment1
1
2017-06-25
```

5. Commit the container to save the changes

```
Snigdha@DESKTOP-T23DDG5 MINGW64 ~/Documents/ADS/Assignment1/Raw_Data (master)
$ docker commit rawdata13 joshisn/assignment1newimage
sha256:07493232db067285be1d8614ba6c626d1481d0649e05070ebe26baea9abc450e
```

6. Run jupyter notebook


```
Snigdha@DESKTOP-T23DDG5 MINGW64 ~/Documents/ADS/Assignment1/Raw_Data (master)
$ docker run -it -d --name "rawdata_2" -p 8888:8888 joshisn/assignment1newimage /bin/bash -c 'jupyter notebook --no-browser --allow-root --ip=* --NotebookApp.password="$PASSWD" "$@"'
272d75e6db86843baa4373c1a8874b33585d3e78462240531d0b24b0653b8561
```

7. Connect jupyter notebook

<http://<yourdocker ip address>:8888>

← → ↻ ① Not secure | 192.168.99.100:8888/login?

---

 jupyter

---

Password or token:

**Token authentication is enabled**

If no password has been configured, you need to open the notebook server with its login token in the URL, or paste it above. This requirement will be lifted if you [enable a password](#).

The command:

```
jupyter notebook list
```

will show you the URLs of running servers with their tokens, which you can copy and paste into your browser. For example:

Currently running servers:  
`http://localhost:8888/?token=c8de56fa... :: /Users/you/notebooks`

or you can paste just the token value into the password field on this page.

See [the documentation on how to enable a password](#) in place of token authentication, if you would like to avoid dealing with random tokens.

Cookies are required for authenticated access to notebooks.

## Data Wrangling

1. Run following commands to pull the image and create container and run wrangle.py file

Note: before running Create command, copy and edit the configwrangle.json file as mentioned in Data Ingestion part.

```
Snigdha@DESKTOP-T23DDG5 MINGW64 ~/Documents/ADS/Assignment1/Clean_Data (master)
$ docker pull joshisn/assignment1:final
final: Pulling from joshisn/assignment1
Digest: sha256:71464840f073694af87742782c47e0c3087332252a9c8b9df9bfff91d9b7b48ea
Status: Image is up to date for joshisn/assignment1:final
```

```
Snigdha@DESKTOP-T23DDG5 MINGW64 ~/Documents/ADS/Assignment1/Clean_Data (master)
$ docker create --name="cleandatacontainer_new" joshisn/assignment1:final
2ef4743b59e70f425a1e758f2e51a55c5c30c6df6883455671dd809a6aec5320

Snigdha@DESKTOP-T23DDG5 MINGW64 ~/Documents/ADS/Assignment1/Clean_Data (master)
$ docker start -i cleandatacontainer_new
```

After performing above command clean.csv file will be uploaded on S3 bucket.

🔍 Type a prefix and press Enter to search. Press ESC to clear.		
<div><div>📁 Upload</div><div>+ Create folder</div><div>More ▾</div></div>		
<input type="checkbox"/>	Name ↑	Last modified ↑
<input type="checkbox"/>	📄 PA1_250617_WBAN_14737.csv	Jun 25, 2017 2:03:15 AM
<input type="checkbox"/>	📄 PA_240617_WBAN_14737_clean.csv	Jun 24, 2017 11:10:14 PM
<input type="checkbox"/>	📄 PA_250617_WBAN_14737_clean.csv	Jun 25, 2017 3:42:01 AM

To open jupyter notebook follow command mentioned in Data Ingestion part.