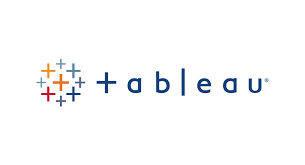
Data Wrangling, Data

Analysis & Visualization

-*Under Guidance of Prof. Sri Krishnamurthy*

** ** **

*Compiled By:*

Mohit Mittal,

Sneha Ravikumar,

Taj Poovaiah

# OBJECTIVE

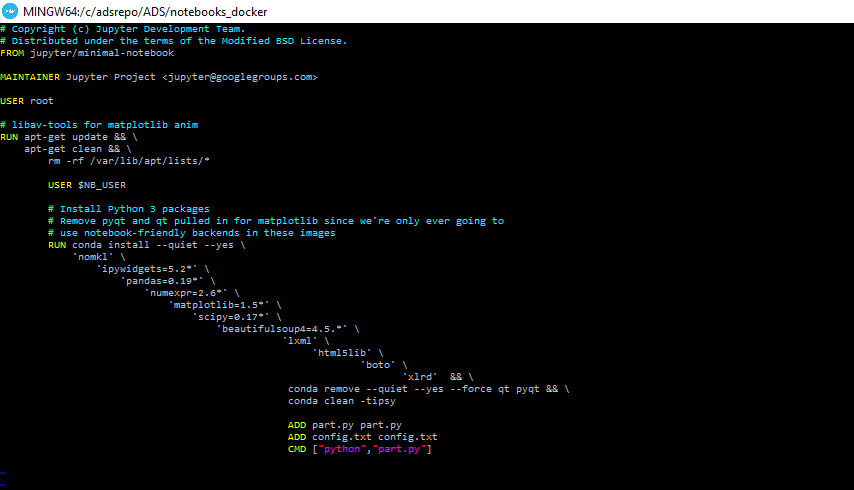
The objective of our assignment is to gather, retrieve, clean, and analyze the data using python code, dockerize the pipeline so as to upload our end results onto Amazon S3 buckets. The use of configuration file and Docker image makes it easy for flexible usage.

# INTRODUCTION TO DOCKER

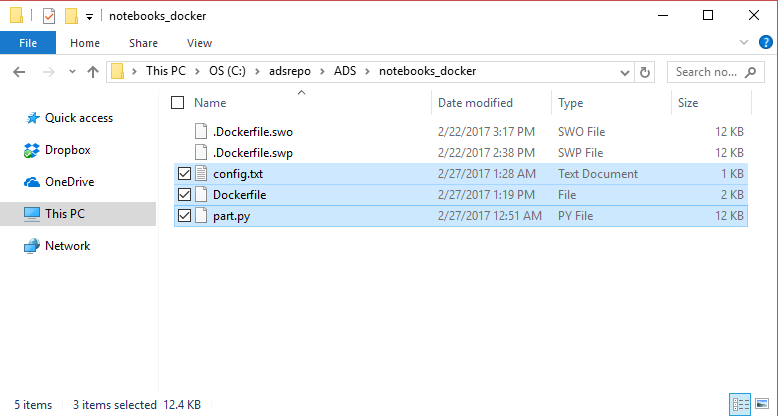
Docker is an open-source project that automates the deployment of applications inside software containers. Quote of features from Docker web pages: “Docker containers wrap up a piece of software in a complete filesystem that contains everything it needs to run: code, runtime, system tools, and system libraries – anything you can install on a server. This guarantees that it will always run the same, regardless of the environment it is running in.”

# DOCKER DESIGN AND IMPLEMENTATION

Docker file is used to specify the configurations and settings required for the application to run:



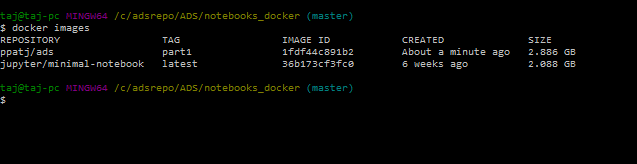
The Docker Directory contains the files required to run the application



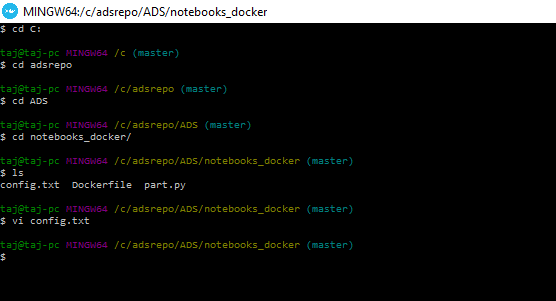
## PULLING A DOCKER IMAGE AND RUNNING IT ON YOUR MACHINE

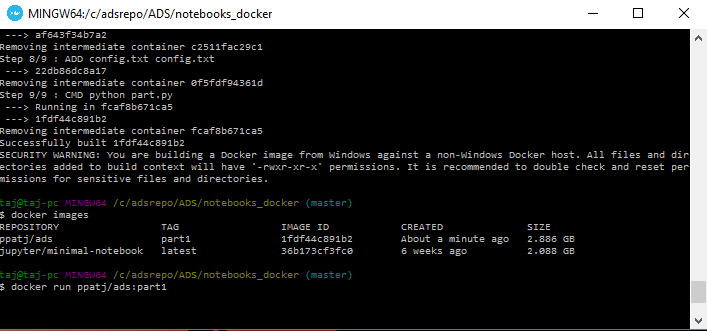
Open the Docker Quickstart terminal on your machine to get started by running the commands on Docker

1. Run **$docker pull “repo name”** to pull the image from any repo on­­to your system
2. Run **$docker images** to check the list of docker images in the system



Update the configuration file with your values



1. Run **$docker run “repo name:tag name”** to run the application on Docker
2. 

The status of the program being executed will be updated on the terminal

