// Running the Code on Interactive Shell //

1. Once the lab is spun up, open a terminal window.

2. To execute the code on an interactive pyspark shell, we need to login to Spark docker container. Execute following command on terminal to do the same.

 # docker exec -i -t hdp_spark-master bash

3. Once we are inside the Spark container, execute following command to launch a pyspark shell.

 # ./spark/bin/pyspark

4. Once the shell is launced, navigate into Project folder on Desktop, for getting the commands to run in the interactive shell.

// Running the Code on JupyterLab //

1. Once the lab is spun up, open a terminal window.

2. To run the code in jupyterlab we need to copy the data and code files to the jupyterlab's docker container. Execute following command on the terminal to do the same.

   # docker cp /home/labuser/Project/Code hdp_jupyterlab:/opt/workspace/

3. Open Chrome browser to launch an instance of jupyterlab. Type following in the address bar and press enter.

   localhost:4888/

4. Click on the Data folder on the left panel, and navigate to the folder where all the data and code files are present.

5. Execute the code as instructed in the project videos.