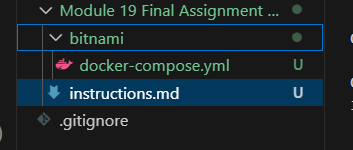
**Submission Instructions:**

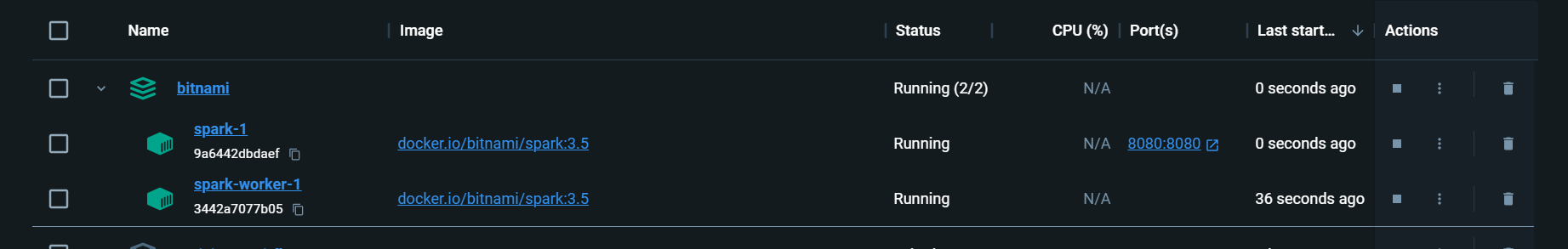
Your submission for this assignment should be a Word document that includes the following screenshots, each labeled for the step that the screenshot represents:

**Part 1: PySpark**

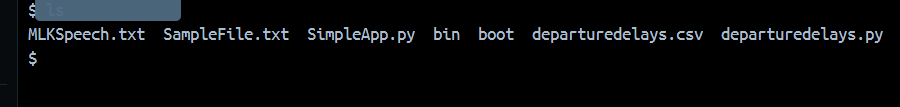
1. Provide a screenshot to show that you correctly pulled the *image* and that the docker-compose.yaml file is present.



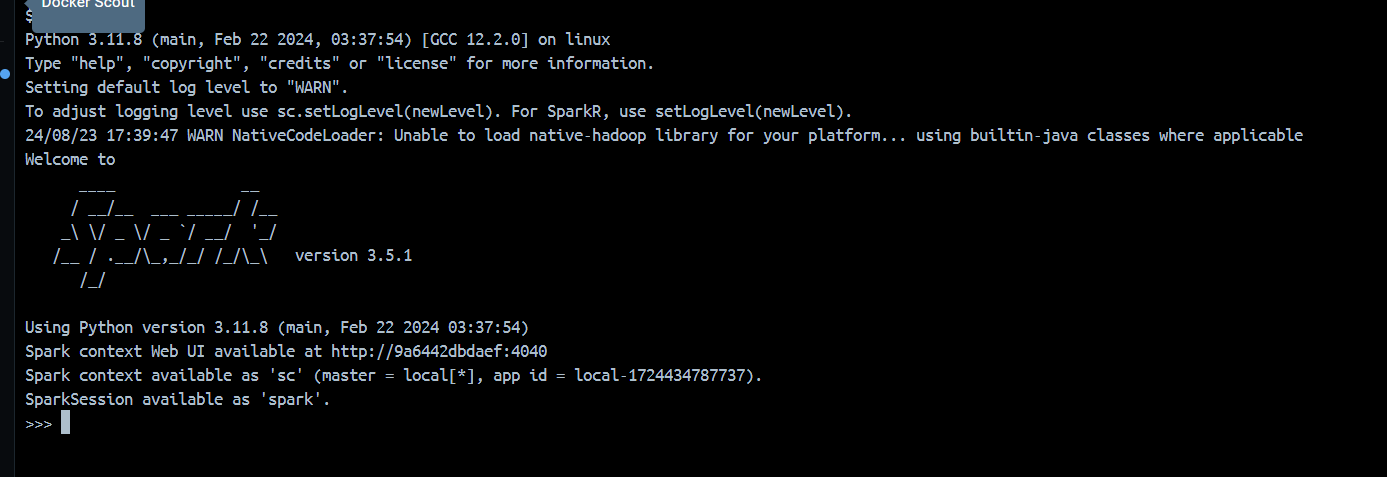
1. Provide a screenshot of your Docker Desktop to show that you correctly pulled the *containers.*



1. Provide a screenshot to show that you successfully copied the departuredelays.csv file to the bitnami\_spark\_1 *container*.



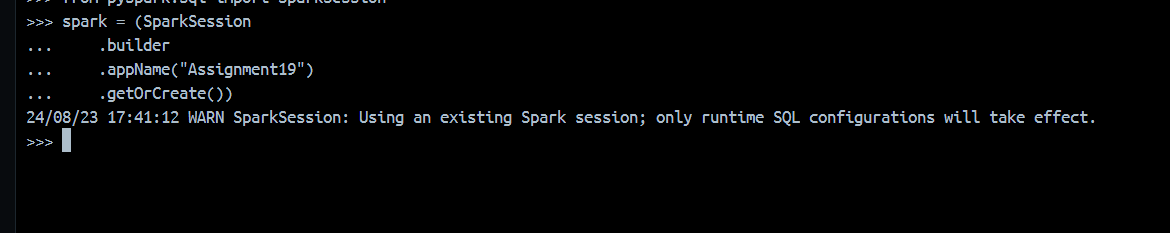
1. Provide a screenshot to show that you successfully opened PySpark.



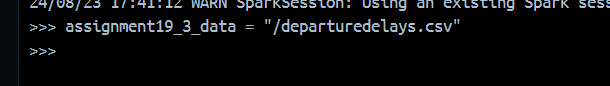
1. Provide a screenshot to show that you successfully started a PySpark session.



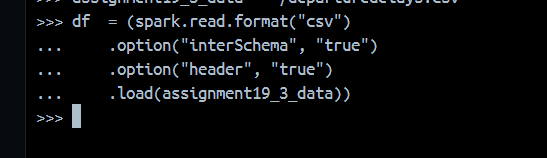
1. Provide a screenshot to show that you successfully defined the spark PySpark session.



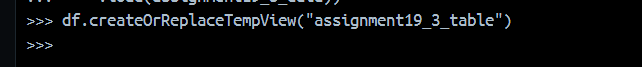
1. Provide a screenshot to show that you successfully defined the assignment19\_3\_data variable.



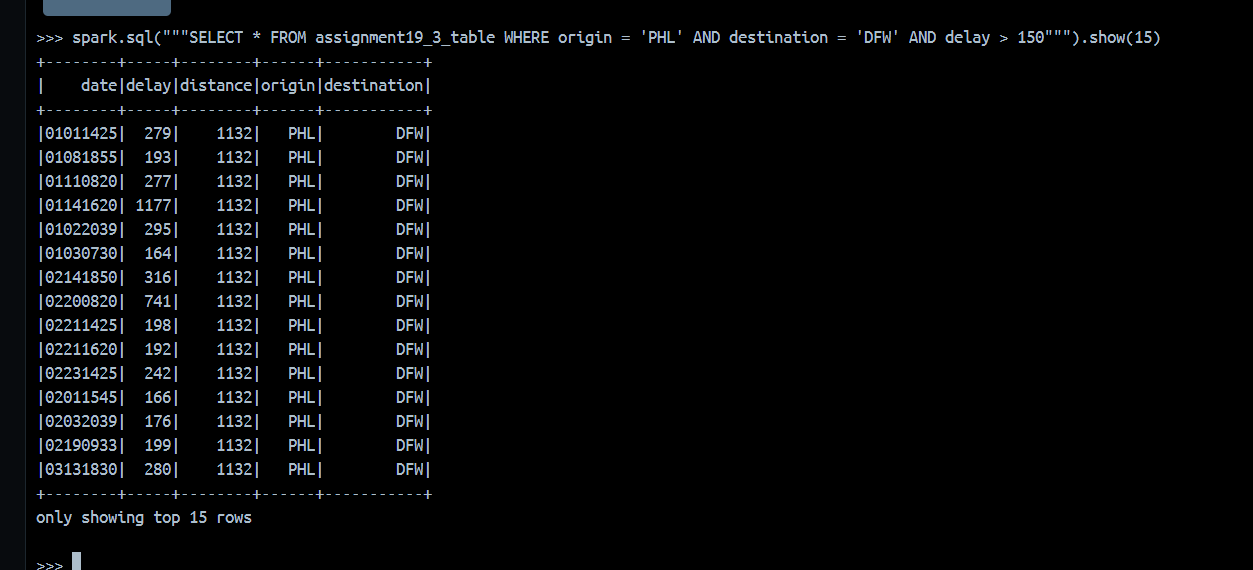
1. Provide a screenshot to show that you successfully defined the df *dataframe* that contains all of the entries in the departuredelays.csv file.



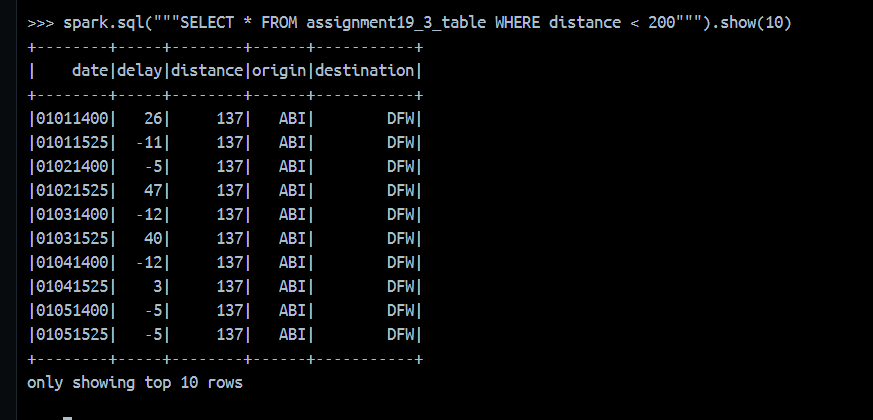
1. Provide a screenshot to show that you successfully created a view of the assignment19\_3\_table *dataframe*.



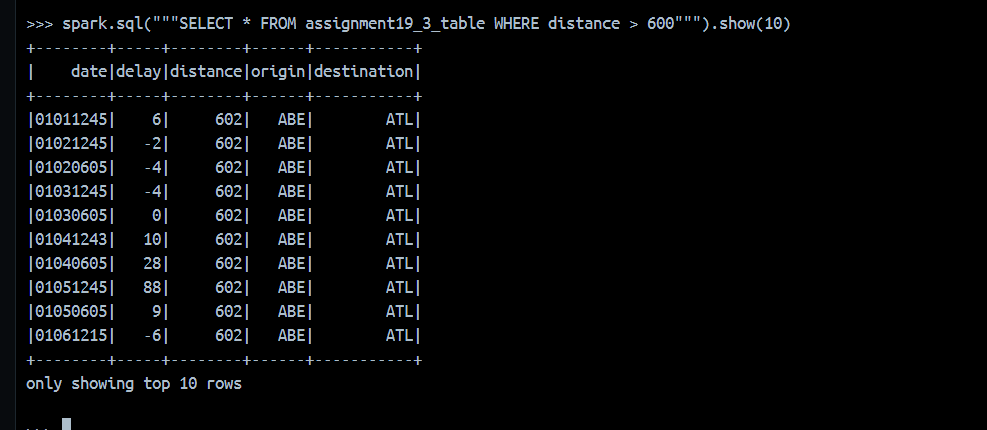
1. Provide a screenshot to show that you selected the correct entries from your data. Your data should display the first 15 flights from PHL to DFW that had a delay of greater than 150 minutes.



1. Provide a screenshot to show that you selected the correct entries from your data. Your data should display the first 10 flights that have a distance of less than 200 miles and the resulting table should contain all of the columns in the original dataset.

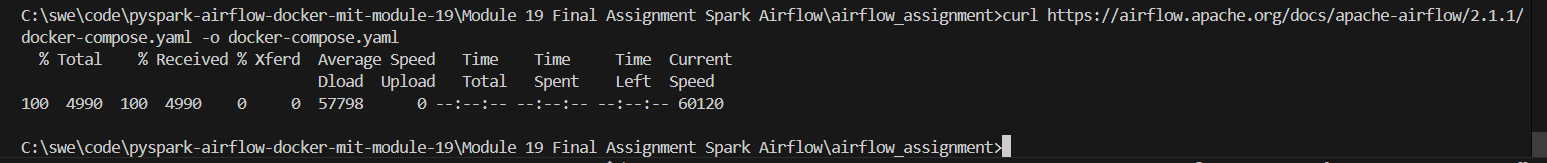


1. Provide a screenshot to show that you selected the correct entries from your data. Your data should display the first 10 flights that have a distance greater than 600 miles and the resulting table should contain all of the columns in the original dataset.

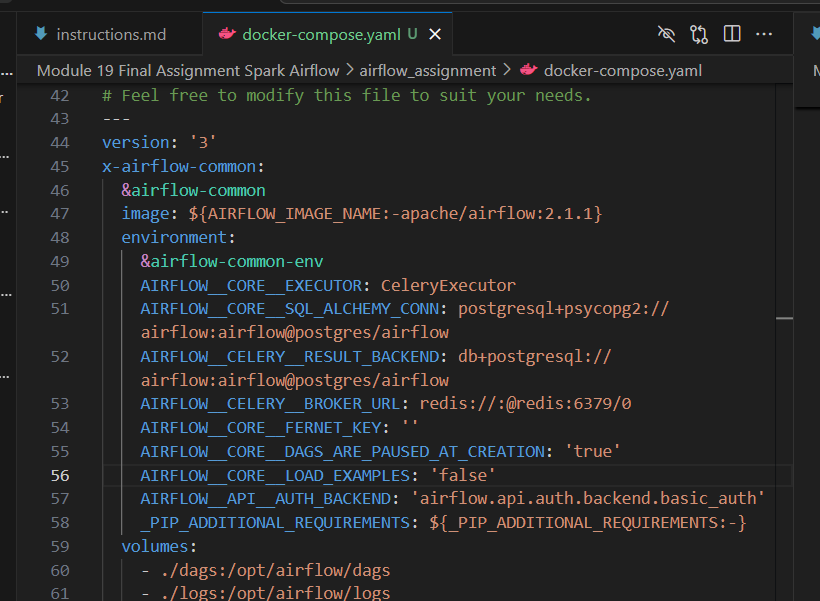


**Part 2: Airflow**

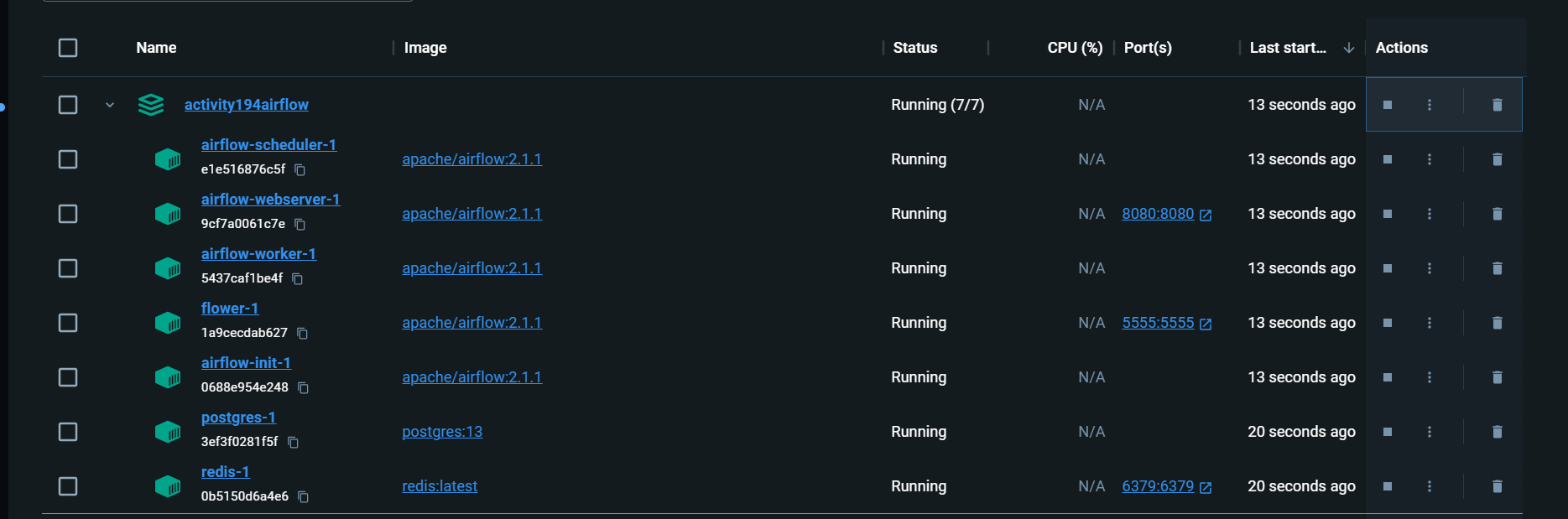
1. Provide a screenshot of your Terminal window response to show that you correctly pulled the Airflow file.



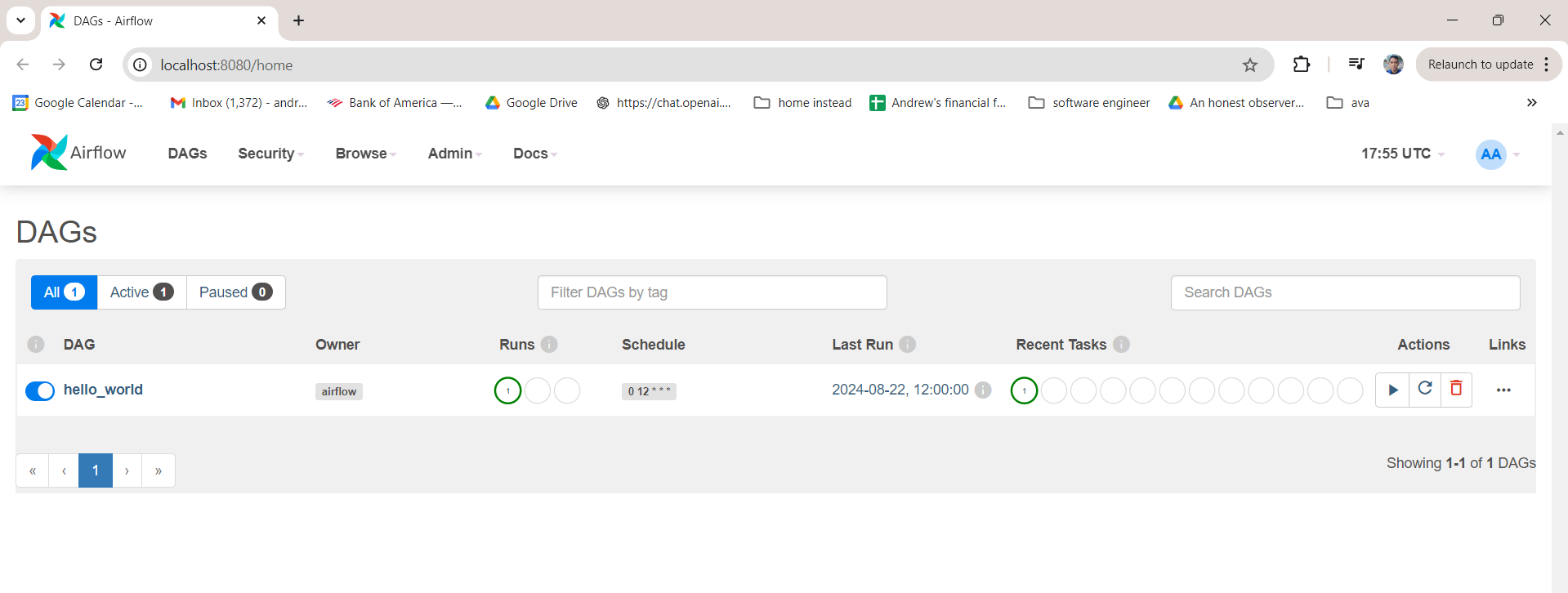
1. Provide a screenshot to show the changed example value ( AIRFLOW\_\_CORE\_\_LOAD\_EXAMPLES set to false).



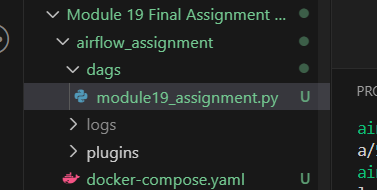
1. Provide a screenshot to show that the Airflow Docker *containers* are running.



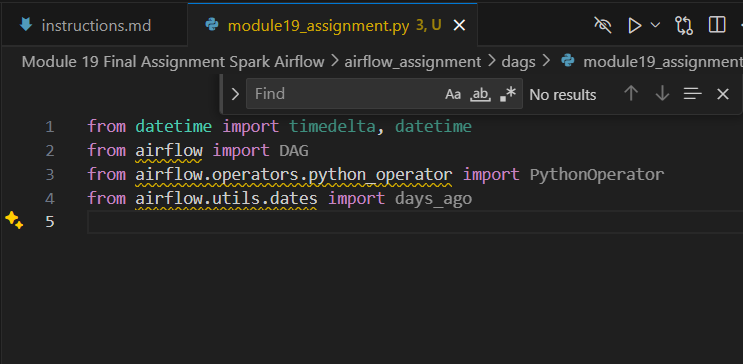
1. Provide a screenshot of your browser window to show that you have successfully logged in to Airflow.



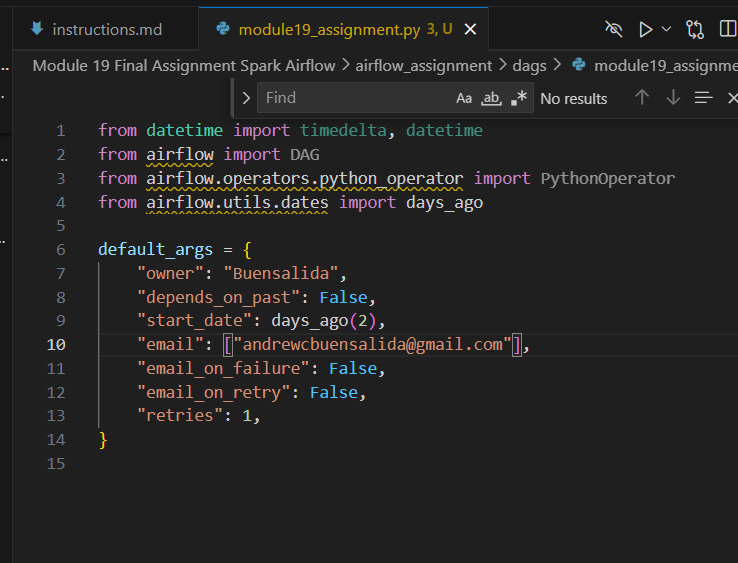
1. Provide a screenshot to show that you created the module19\_assignment.py file.



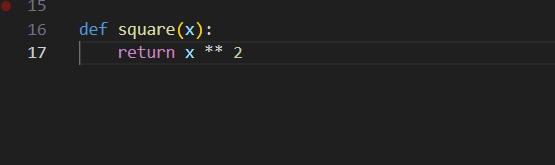
1. Provide a screenshot to show that you correctly imported the required *libraries*.



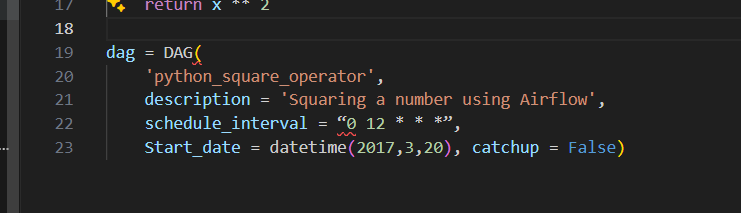
1. Provide a screenshot to show that you set up your DAG correctly, including your last name and email address.



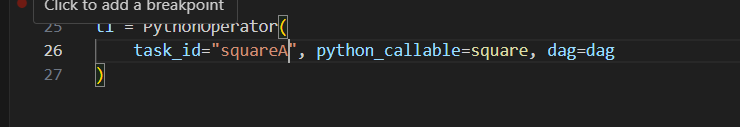
1. Provide a screenshot to show that you defined the square() *function* correctly.



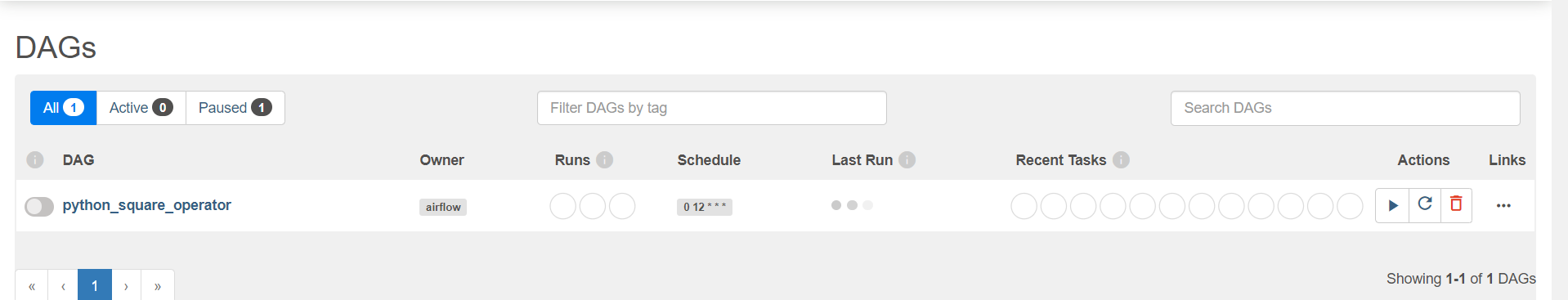
1. Provide a screenshot to show that you correctly defined the DAG object.



1. Provide a screenshot to show that you defined the DAG *Task* correctly.



1. Provide a screenshot of the Airflow UI to show that your DAG is configured correctly.



1. Provide a screenshot of the log to show that the DAG ran successfully.

