Task-2: Write a Spark job **without using SQL** to determine the average pressure at 9am (Pressure9am) for Launceston (location).

Step 1: Login to HADOOP CLUSTER and Navigate to Pyspark using pyspark command

```
(base) Achyuthas-MacBook-Air:~ achyuthanagaveti$ ssh anagave@hadoop-nn001.cs.okstate.edu
anagave@hadoop-nn001.cs.okstate.edu's password:
Welcome to Ubuntu 20.04.1 LTS (GNU/Linux 5.4.0-62-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

Last login: Fri Apr  2 02:02:28 2021 from 10.200.206.99
```

```
anagave@hadoop-nn001:~$ pyspark
```

```
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 3.0.1
      /_/

Using Python version 3.8.5 (default, Jan 27 2021 15:41:15)
SparkSession available as 'spark'.
```

Step 2: CSV file is located in /user/common_data/ Spark_Assignment_Dataset.csv path.

Step 3: Import required packages such as SparkSession, pys-ark.sql.functions, pys-ark.sql.types, date time.

Step 4: Create spark session variable to assign sparksession and dataframe variable to create a data frame from csv

---

Step 5: Apply filter on the data frame to show the Average pressure, and print the same. Results are shown below(Highlighted in the red box)