**Week\_05\_assignment\_02\_from\_LSpV2\_Ch3\_part2**

**My work**

**\*\*\*\*\*\*\*\*\*No prof instructions and demo fro this lab…Bcoz prof left the class early..He needs to go somewhere**

**So the class is just for 10 – 11 min..he just explained the instructions given…no demo was done by him\*\*\*\*\*\*\*\*\*\***

**PYSPARK instructions:**

A screenshot of a computer program

Description automatically generated

**SCALA INSTRUCTIONS:**

A screenshot of a computer program

Description automatically generated

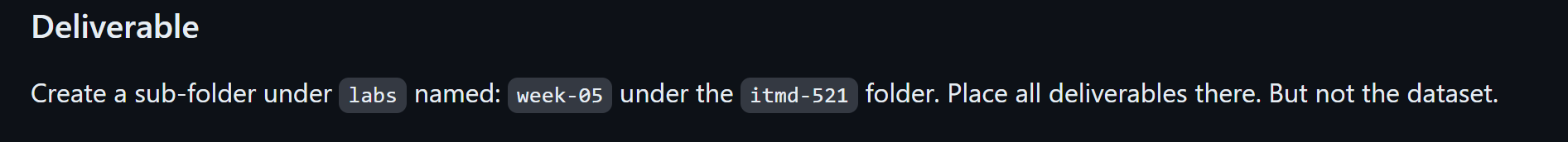
**MYWORK**

**Did git pull from the prev lab-02 work in vagrant box**

A screenshot of a computer program

Description automatically generated

**As per the requirement I created this folder in my vs code**



**I created a folder called week-05 under labs where I will place assignment\_02.py file and scala folder with .scala file following the same hierarchy and placing the build sbt file in the iitial scala folder**

A screenshot of a computer program

Description automatically generated

**The I created assignment\_02.py file in the week-05 folder where I need to develop the code and do the 7 questions.**

A screenshot of a computer

Description automatically generated

**Lets fill the code step by step by doing the git commits…**

**And paste the screenshot here after full completion.**

**But before the full code screenshot….If we get any errors in b/w we will document and take screenshots and explain the solution for future ref**

**I have written the code till create schema(without the 7 questions) and trying to run the code to test with spark submit command in week-05 > py directory**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*testing the code \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

A screen shot of a computer

Description automatically generated

**56 lines of code**

**Before doing the spark submit command lets pull all this code in our vagrant box with git pull command from nannadevara directory**

A screenshot of a computer program

Description automatically generated

**Lets run spark submit command to test our code(without the 7 questions)**

**\*\*\*\*before testing the code I made small change in py file hierarchy:\*\*\*\*\*\*\*\*\*\*\*\*\*\***

A screenshot of a computer program

Description automatically generated

**Lets do git pull again to update this small hierarchy change we did**

A screenshot of a computer program

Description automatically generated

**NoW LETS RUN SPARK SUBMIT TO TEST THE CODE: in week-05 > py directory**

Spark submit code for this assignment\_02 is: (for .py file ~ that’s is pyspark application for fire calls csv):

**spark-submit assignment\_02.py ~/LearningSparkV2/databricks-datasets/learning-spark-v2/sf-fire/sf-fire-calls.csv**

**lets go into week-05 > py directory and run this spark submit command**

**YAYYYYYYYYYYYYYYYYYYYY success…..pyspark code run success**

A screenshot of a computer screen

Description automatically generated

A screen shot of a computer

Description automatically generated

**Now lets answer the 7 questions from Textbook page 68 for pyspark code:**

A screenshot of a computer

Description automatically generated

1. What were all the different types of fire calls in 2018?
2. What months within the year 2018 saw the highest number of fire calls?
3. Which neighborhood in San Francisco generated the most fire calls in 2018?
4. Which neighborhoods had the worst response times to fire calls in 2018?
5. Which week in the year in 2018 had the most fire calls?
6. Is there a correlation between neighborhood, zip code, and number of fire calls?
7. How can we use Parquet files or SQL tables to store this data and read it back?

DONE with answers.. got to github for code and answers. Copy pasted the output in the .py file as multi line comment

**Now scala code for iot devices.json data set**

Created scala directory and followed the hierarchy and wrote/developed the scala code for iot device.json in assignment\_02.scala file:

A screen shot of a computer program

Description automatically generated

I also added the sbt file and filled the content.

A screenshot of a computer

Description automatically generated

Then we need to answer these questions from the textbook page 74

1. Detect failing devices with battery levels below a threshold.

2. Identify offending countries with high levels of CO2 emissions.

3. Compute the min and max values for temperature, battery level, CO2, and

humidity.

4. Sort and group by average temperature, CO2, humidity, and country.

A screenshot of a computer program

Description automatically generated

Please note that we can run this file and get these outputs after we run sbt clean package

Which is gonna generate a jar file which will compile our code

We take that jar file and do the spark submit command….

\*\*\*ALSO please note that do not forget to do git pull in vm whenever you are developing code and trying to run the code to see the results\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

THEN OUR SCALA CODE WILL RUN …. AND RUNS PROPERLY IF OUR CODE AND ANSWERS ARE CORRECT

THEN WE WILL GET THESE OTPUTS  
  
THEN WE NEED TOCOPY THOSE OUTPUTS AND PASTE IN OUR VS CODE .scala file UNDER EACH QUESTION IN MULTILINE COMMENT AS HOWN IN SC

A screenshot of a computer program

Description automatically generated

The spark submit command for this iot devices is:

spark-submit --class main.scala.assignment\_02.assignment\_02 /home/vagrant/itmd-521/nannadevara/itmd-521/labs/week-05/scala/target/scala-2.12/main-scala-assignment\_02\_2.12-1.0.jar /home/vagrant/LearningSparkV2/databricks-datasets/learning-spark-v2/iot-devices/iot\_devices.json

A screenshot of a computer program

Description automatically generated

The scala file successfully ran and showed the schema and answerd to the questions.

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Then copy these outputs in to our scala file in a multi line comment…

But nenu munde outputs copy chesesa…motham scala code..answers comments anni ipoyaka malli sbt clean package run chesi motham code ni compile chesi spark submit kotta final ga test ki….

Motham run indi…

A screen shot of a computer program

Description automatically generated