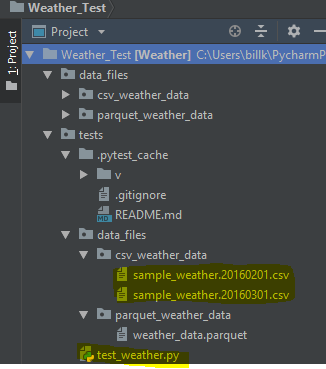
**DATA ENGINEER PYTHON TEST**

My approach to the Data Engineer Python Test was to read the questions and workout the subset of data that was required to answer all the questions. I decided to extract ObservationDate, ScreenTemperature and Region columns from the csv files.

Looking at the data size and quality I decided using Pandas would be the best option, as Pandas are an efficient way of working with small datasets and a great tool for data analysing and manipulating data.

After analysing the data, I found ‘-99’ values in the ScreenTemperature column, these values seemed out of line with the rest of the data. Looking at the data context I decided to set these values to 0.

I have created unit tests for the functions and created test resource files by sampling the csv data files provided.



I returned the answers to the questions in the form of a Pandas dataframe, please see below -

