Question:

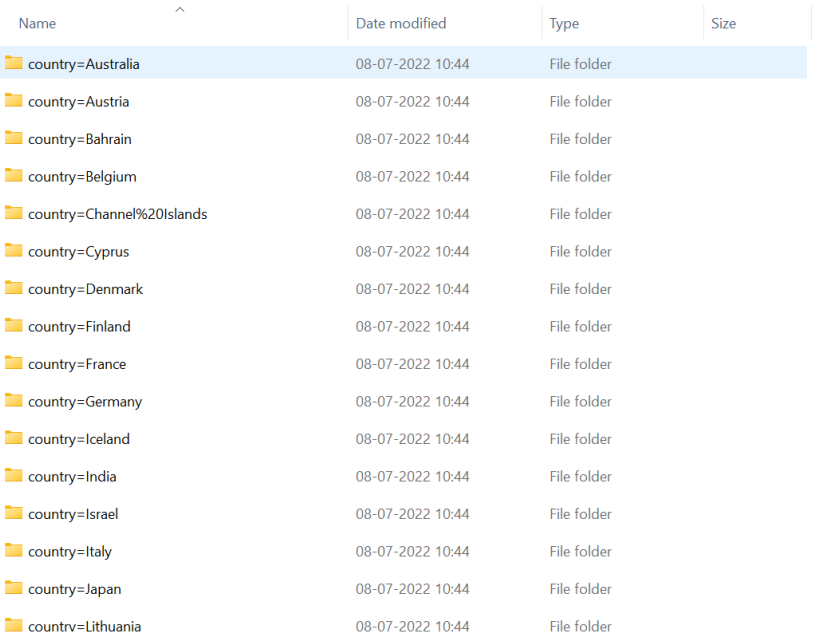
We have a file windowdata.csv and the field names are country, weeknum, numinvoices, totalquantity, invoicevalue

Step 1: create spark session
Step 2: set the logging level to error
Step 3: Using the standard dataframe reader API load the file and create a dataframe.
Step 4: Use the standard dataframe writer api to save it in parquet format. While saving make sure data is stored where we should have a folder for each country, weeknum (combination)
Step 5: Also use the dataframe write api to save the data in Avro format. While saving make sure data is stored where we should have a folder for each country.
Step 6: Apply header
Step 7: Convert dataframe to dataset(Specific type)

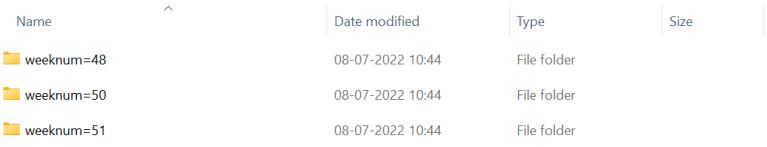
Code:

```
import org.apache.spark.sql.{SaveMode, SparkSession}

object WindowDataWrite extends App{

  case class
Customer(country:String,weeknum:Int,numinvoices:Int,totalquantity:Int,invoi
cevalue:Double)
   val session = SparkSession
  .builder.appName("Window Data")
  .master("local[*]")
  .getOrCreate()

  val dataframe = session
    .read
    .option("inferSchema","true")
    .csv("E://Data//windowdata.csv")

  // adding column names to the data
  val withColumns = dataframe.toDF("country", "weeknum", "numinvoices",
"totalquantity", "invoicevalue");

  // Now displaying the data
  withColumns.show(10,false);

  // Reading the data with headers
  val dataframeHeader = session
    .read
    .option("header","true")
    .option("inferSchema","true")
    .csv("E://Data//windowdata_columns.csv")

  import session.implicits._
  val datasetHeader = dataframeHeader.as[Customer]

  datasetHeader.show(10,false);
```

```
  // Each folder country, week num combination
  withColumns
    .write
    .mode("overwrite")
    .format("parquet")
    .partitionBy("country","weeknum")
    .option("path","E://Data//WriteData//Sample1")
    .save()


  session.close();
}
```



Name	Date modified	Type	Size
country=Australia	08-07-2022 10:44	File folder	
country=Austria	08-07-2022 10:44	File folder	
country=Bahrain	08-07-2022 10:44	File folder	
country=Belgium	08-07-2022 10:44	File folder	
country=Channel%20Islands	08-07-2022 10:44	File folder	
country=Cyprus	08-07-2022 10:44	File folder	
country=Denmark	08-07-2022 10:44	File folder	
country=Finland	08-07-2022 10:44	File folder	
country=France	08-07-2022 10:44	File folder	
country=Germany	08-07-2022 10:44	File folder	
country=Iceland	08-07-2022 10:44	File folder	
country=India	08-07-2022 10:44	File folder	
country=Israel	08-07-2022 10:44	File folder	
country=Italy	08-07-2022 10:44	File folder	
country=Japan	08-07-2022 10:44	File folder	
countrv=Lithuania	08-07-2022 10:44	File folder	



Name	Date modified	Type	Size
weeknum=48	08-07-2022 10:44	File folder	
weeknum=50	08-07-2022 10:44	File folder	
weeknum=51	08-07-2022 10:44	File folder	