ASSIGNMENT – 3 SPARK SQL using json file

Note : provide the commands along with output.

Consider the following json file as datasource.
  {"id" : "101", "ename" : "Ram", "age" : "29"}
  {"id" : "102", "ename" : "Sai", "age" : "30"}
  {"id" : "103", "ename" : "joe", "age" : "32"}
  {"id" : "104", "ename" : "sirish", "age" : "30"}
  {"id" : "105", "ename" : "shekhar", "age" : "32"}

1.   From the above json file answer the following.
a)   Display file data using data frame.
scala> file.show()

```
scala> val file = spark.read.json("input.json")
file: org.apache.spark.sql.DataFrame = [age: string, ename: string ... 1 more field]

scala> file.printSchema
root
 |-- age: string (nullable = true)
 |-- ename: string (nullable = true)
 |-- id: string (nullable = true)


scala> file.show()
+---+-------+---+
|age|  ename| id|
+---+-------+---+
| 29|    Ram|101|
| 30|    Sai|102|
| 32|    joe|103|
| 30| sirish|104|
| 32|shekhar|105|
+---+-------+---+
```

b)   Display the structure of data frame .
scala> file.printSchema

```
scala> file.printSchema
root
 |-- age: string (nullable = true)
 |-- ename: string (nullable = true)
 |-- id: string (nullable = true)
```

c)   Display only enames from the given file using select.
scala> file.select("ename").show

```
scala> file.select("ename").show
+-------+
|  ename|
+-------+
|    Ram|
|    Sai|
|    joe|
| sirish|
|shekhar|
+-------+
```

2.   From the json file given, answer the following.
a)   Provide the count of each age.
scala> file.select("age").count

```
scala> file.select("age").count
res6: Long = 5
```

b)   Provide all records in sorted order.
(Comment: Since no sort column is given i sorted the data on first column i.e "age")
scala> file.orderBy("age").show()
OR
scala> file.sort("age").show()

```
scala> file.orderBy("age").show()
+---+-------+---+
|age|  ename| id|
+---+-------+---+
| 29|    Ram|101|
| 30| sirish|104|
| 30|    Sai|102|
| 32|    joe|103|
| 32|shekhar|105|
+---+-------+---+
```

```
scala> file.sort("age").show()
+---+-------+---+
|age|  ename| id|
+---+-------+---+
| 29|    Ram|101|
| 30| sirish|104|
| 30|    Sai|102|
| 32|    joe|103|
| 32|shekhar|105|
+---+-------+---+
```

c)   Display records whose age <=30.
scala> file.filter("age <= 30").show

```
scala> file.filter("age <= 30").show
+---+-----+---+
|age|ename| id|
+---+-----+---+
| 29|  Ram|101|
| 30|  Sai|102|
| 30|sirish|104|
+---+-----+---+
```