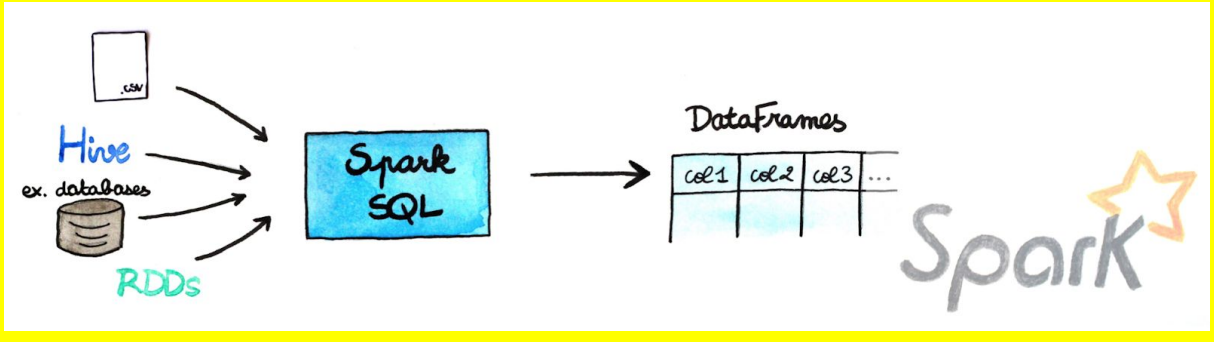
Spark SQL-Java Application: Read CSV file into Data Frame and Execute some Queries With Spark SQL And Java

I. Introduction:

In this documentation, we are focused to parse data from a CSV file, perform some queries and output the result in the output using the Spark Core and Spark SQL APIs, and also Java.



II. Technologies:

- Java 8
- Spark Core 2.4.7
- Spark SQL 2.4.7
- Maven
- Intellij IDEA

III. Implement some queries using java and spark:

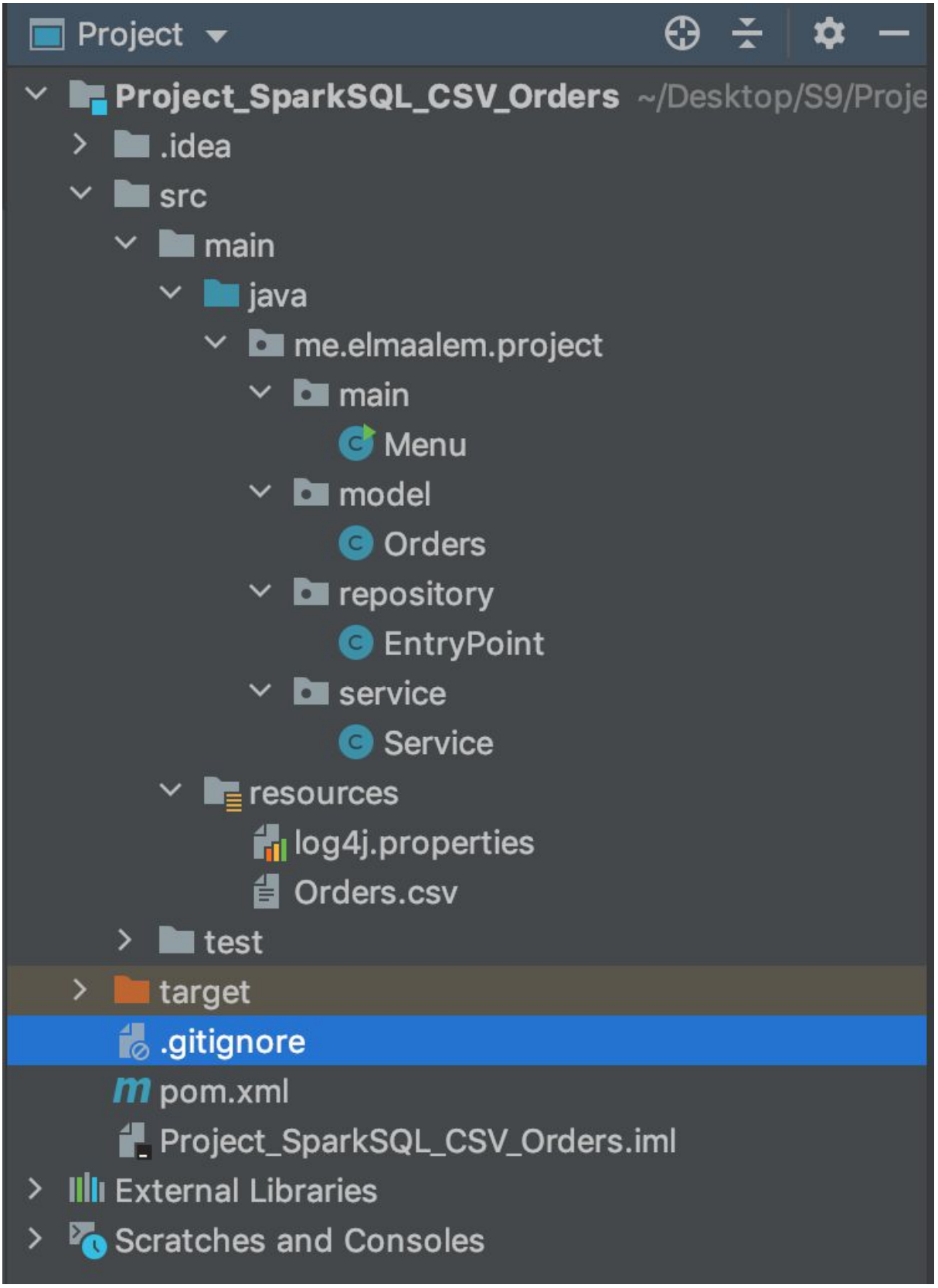
-Let's have a look at the **orders** dataset which we will use for this queries:

orderId	date	quantity	sales	mode	profit	unitPrice	customerName	customerSegment	productCategory
3	10-13-2010	6	261.54	Regular Air	-213.25	38.94	Muhammed MacIntyre	Small Business	Office Supplies
6	02-20-2012	2	6.93	Regular Air	-4.64	2.08	Ruben Dartt	Corporate	Office Supplies
32	07-15-2011	26	2808.08	Regular Air	1054.82	107.53	Liz Pelletier	Corporate	Furniture
32	07-15-2011	24	1761.4	Delivery Truck	-1748.56	70.89	Liz Pelletier	Corporate	Furniture
32	07-15-2011	23	160.2335	Regular Air	-85.129	7.99	Liz Pelletier	Corporate	Technology
32	07-15-2011	15	140.56	Regular Air	-128.38	8.46	Liz Pelletier	Corporate	Technology
35	10-22-2011	30	288.56	Regular Air	60.72	9.11	Julie Creighton	Corporate	Office Supplies
35	10-22-2011	14	1892.848	Regular Air	48.987	155.99	Julie Creighton	Corporate	Technology
36	11-02-2011	46	2484.7455	Regular Air	657.477	65.99	Sample Company A	Home Office	Technology
65	03-17-2011	32	3812.73	Regular Air	1470.3	115.79	Tamara Dahlen	Corporate	Technology
32	07-15-2008	26	<null>	Regular Air	<null>	107.53	Liz Pelletier	Corporate	Furniture
32	07-15-2008	24	<null>	Delivery Truck	<null>	70.89	Liz Pelletier	Corporate	Furniture
32	07-15-2008	23	<null>	Regular Air	<null>	7.99	Liz Pelletier	Corporate	Technology
32	07-15-2008	15	<null>	Regular Air	<null>	8.46	Liz Pelletier	Corporate	Technology
35	10-22-2008	30	<null>	Regular Air	<null>	9.11	Julie Creighton	Corporate	Office Supplies
35	10-22-2008	14	<null>	Regular Air	<null>	155.99	Julie Creighton	Corporate	Technology
36	10-22-2008	46	<null>	Regular Air	<null>	65.99	Sample Company A	Home Office	Technology
65	10-22-2008	32	<null>	Regular Air	<null>	115.79	Tamara Dahlen	Corporate	Technology
66	01-19-2009	41	108.15	Regular Air	7.57	2.88	Arthur Gainer	Consumer	Office Supplies
69	06-03-2009	42	1186.06	Regular Air	511.69	30.93	Jonathan Doherty	Corporate	Furniture
69	06-03-2009	28	51.53	Express Air	0.35	1.68	Jonathan Doherty	Corporate	Office Supplies
70	12-17-2010	48	90.05	Regular Air	-107	1.86	Helen Wasserman	Home Office	Office Supplies
70	12-17-2010	46	7804.53	Regular Air	2057.166	205.99	Helen Wasserman	Home Office	Technology
96	04-16-2009	37	4158.1235	Regular Air	1228.887	125.99	Keith Dawkins	Home Office	Technology
97	01-28-2010	26	75.57	Regular Air	28.24	2.89	Craig Yedwab	Consumer	Office Supplies
129	11-18-2012	4	32.72	Regular Air	-22.59	6.48	Pauline Chand	Corporate	Office Supplies

-These are **queries** to be exported:

- Get all Orders from csv file
- Get Orders By Customer Name
- Get Orders(Name,Sales,Profit,Category) By Customer Name and Order Date
- Get Orders By Product Category When Profit must be greater than 0 And Sorted by Customer Name
- Get Orders By Product Category In period and sorted by Sales

IV. Project Structure :



V. Setup Dependencies on pom.xml:

After Adding the below dependencies on pom.xml, It will download all the required packages.

```
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/maven-4.0.0.xsd">
    <modelVersion>4.0.0</modelVersion>

    <groupId>me.elmmalem</groupId>
    <artifactId>Prject_SparkSQL_With_Java</artifactId>
    <version>1.0-SNAPSHOT</version>
    <build>
        <plugins>
            <plugin>
                <groupId>org.apache.maven.plugins</groupId>
                <artifactId>maven-compiler-plugin</artifactId>
                <configuration>
                    <source>8</source>
                    <target>8</target>
                </configuration>
            </plugin>
        </plugins>
    </build>
    <dependencies>
        <!-- Dependency of Apache Spark Core-->
        <dependency>
            <groupId>org.apache.spark</groupId>
            <artifactId>spark-core_2.11</artifactId>
            <version>2.4.7</version>
        </dependency>
        <!-- Dependency of Apache Spark SQL-->
        <dependency>
            <groupId>org.apache.spark</groupId>
            <artifactId>spark-sql_2.11</artifactId>
            <version>2.4.7</version>
        </dependency>
    </dependencies>
</project>
```

VI. Configure Log4j file on spark console :

I'd like to stop various **INFO messages** that are coming on the spark console to get just the result on the console without logging messages.

```
21/01/24 00:05:57 INFO TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 430 ms on localhost (executor driver) (1/1)
21/01/24 00:05:57 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
21/01/24 00:05:57 INFO DAGScheduler: ResultStage 0 (show at Service.java:14) finished in 0.669 s
21/01/24 00:05:57 INFO DAGScheduler: Job 0 finished: show at Service.java:14, took 0.729501 s
+-------+----------+--------+---------+-------------+------------------+---------+----------------+---------------+---------------+
|orderId|      date|quantity|    sales|         mode|            profit|unitPrice|    customerName|customerSegment|productCategory|
+-------+----------+--------+---------+-------------+------------------+---------+----------------+---------------+---------------+
```

I edit the **log4j.properties** file in order to stop these messages. Here are the contents of **log4j.properties:**

```
#Stop INFO messages displaying on Spark console to get just the result expected
log4j.rootCategory=ERROR, console
log4j.appender.console=org.apache.log4j.ConsoleAppender
log4j.appender.console.target=System.err
log4j.appender.console.layout=org.apache.log4j.PatternLayout
log4j.appender.console.layout.ConversionPattern=%d{yy/MM/dd HH:mm:ss} %p %c{1}: %m%n
```

VII. Define Data Model:

In the **model** package, we define **Orders** class.
model/Orders.class:

```
public class Orders implements Serializable {

    private int orderId;
    private LocalDate date;
    private int quantity;
    private Optional<Double> sales;
    private String mode;
    private Optional<Double> profit;
    private double unitPrice;
    private String customerName;
    private String customerSegment;
    private String productCategory;
    private static final DateTimeFormatter FORMATTER = DateTimeFormatter.ofPattern("MM-dd-yyyy");

    public Orders(){}
    public Orders(int orderId, String date, int quantity, double sales, String mode, double profit, double unitPrice,
                  String customerName, String customerSegment, String productCategory) {...}

    public String getProductCategory() { return productCategory; }

    public void setProductCategory(String productCategory) { this.productCategory = productCategory; }

    public int getOrderId() { return orderId; }

    public void setOrderId(int orderId) { this.orderId = orderId; }

    public LocalDate getDate() { return date; }

    public void setDate(String date) { this.date = LocalDate.parse(date,FORMATTER); }

    public int getQuantity() { return quantity; }

    public void setQuantity(int quantity) { this.quantity = quantity; }

    public Optional<Double> getSales() { return sales; }
```

VIII. Create a Repository to working with Dataframe(Orders.csv):

Let's create a repository to interact with **Orders** from the csv file.
In the **repository** package, create a class **EntryPoint** which is responsible for reading **CSV file** and loading the data into a **spark dataframe** with a custom schema.

```java
public class EntryPoint {

    public EntryPoint() { }

    private static SparkSession sparkSession(){
        return SparkSession
                .builder()
                .appName(" Application with Spark SQL and Java")
                .master("local[*]")
                .getOrCreate();
    }

    private static StructType customSchema(){
        return new StructType(new StructField[] {
                new StructField("orderId", DataTypes.IntegerType, true,
Metadata.empty()),
                new StructField("date", DataTypes.DateType, true,
Metadata.empty()),
                new StructField("quantity", DataTypes.IntegerType, true,
Metadata.empty()),
                new StructField("sales", DataTypes.DoubleType, true,
Metadata.empty()),
                new StructField("mode", DataTypes.StringType, true,
Metadata.empty()),
                new StructField("profit", DataTypes.DoubleType, true,
Metadata.empty()),
                new StructField("unitPrice", DataTypes.DoubleType, true,
Metadata.empty()),
                new StructField("customerName", DataTypes.StringType, true,
Metadata.empty()),
                new StructField("customerSegment", DataTypes.StringType,
true, Metadata.empty()),
                new StructField("productCategory", DataTypes.StringType,
true, Metadata.empty())
        });
    }

    public static Dataset<Orders> getDataset(){
        Encoder<Orders> orderEncoder = Encoders.bean(Orders.class);

        return sparkSession().read()
                .option("header","true")
                .option("treatEmptyValuesAsNulls", "true")
                .schema(customSchema())
                .option("mode","DROPMALFORMED")
                .option("dateFormat", "MM-dd-yyyy")
                .option("delimiter",",")
                .csv("src/main/resources/Orders.csv")
                .as(orderEncoder);
    }
}
```

IX. Create a Spark Service:

SparkService class uses **Repository/EntryPoint** class for 5 functions:

→ **listOrders(int numberRows)**: Get all Orders from csv file
→ **listOrdersMatchCustomerName(String customerName)**: Get Orders By Customer Name
→ **listOrdersMatchCustomerNameAndOrderDate(String customerName, String orderDate)**: Get Orders(Name,Sales,Profit,Category) By Customer Name and Order Date
→ **listOrdersMatchCategoryAndProfitPositiveAndSortByCustomerName(String productCategory):** Get Orders By Product Category When Profit must be greater than 0 And Sorted by Customer Name
→ **listOrdersMatchCategoryAndPeriodDateAndSortBySales(String productCategory, String startDate, String endDate):** Get Orders By Product Category In period and sorted by Sales

Here is the code of **service/SparkService.java:**

```
public class SparkService {

   public void listOrders(int numberRows){
       getDataset().show(numberRows);
   }

   public void listOrdersMatchCustomerName(String customerName) {
       Dataset<Orders> orders =
getDataset().filter((FilterFunction<Orders>) order ->
order.getCustomerName().equals(customerName));
       orders.show((int) getDataset().count());
   }

   public void listOrdersMatchCustomerNameAndOrderDate(String
customerName, String orderDate) {
       Dataset<Row> orders = getDataset()
               .filter((FilterFunction<Orders>) order ->
order.getCustomerName().equals(customerName))

.select("customerName","date","sales","profit","productCategory")
               .where("date == \"" + orderDate + "\"");
       orders.show((int) getDataset().count());
   }

   public void
listOrdersMatchCategoryAndProfitPositiveAndSortByCustomerName(String
productCategory) {
```

```
        Dataset<Orders> orders = getDataset()
                .filter((FilterFunction<Orders>) order ->
order.getProductCategory().equals(productCategory))
                .filter("profit > 0.0")
                .sort("customerName");
        orders.show((int) getDataset().count());
    }

    public void listOrdersMatchCategoryAndPeriodDateAndSortBySales(String
productCategory, String startDate, String endDate) {
        Dataset<Row> orders = getDataset()
                .filter((FilterFunction<Orders>) order ->
order.getProductCategory().equals("Office Supplies"))
                .sort("sales")
                .where("date < \""+endDate+"\" and date >
\""+startDate+"\"")

.select("customerName","date","sales","quantity","profit","unitPrice","cu
stomerSegment");
        orders.show((int) getDataset().count());
    }
}
```

X. Creating a Menu Driven Program :

Let's create a **Menu** class under package **Main** to obtain input from a user by displaying a list of options.

main/Menu.java:

```java
public class Menu {

    public static Scanner scanner = new Scanner(System.in);
    public static SparkService sparkService = new SparkService();

    public static void main(String[] args) {

        try {
            int menuOption = 0;
            String customerName;
            String productCategory;
            String orderDate;
            String startDate;
            String endDate;

            do {
                // Setting menuOption equal to return value from showMenu();
                menuOption = showMenu();

                switch (menuOption) {...}

            } while (menuOption != 6);

            // Exiting message when user decides to quit Program
            System.out.println("Thanks for using this Program...");

        } catch (Exception ex) {
            System.out.println("Sorry problem occured within Program");
            scanner.next();
        } finally {
            scanner.close();
        }
    }
    public static int showMenu() {...}
}
```

XI. Output:

While executing each query , you will be able to see below its content in the console.

1. Menu:

```
Menu:
1. Get All Orders form CSV file
2. Get Orders By Customer Name
3. Get Orders(Name,Sales,Profit,Category)  By Customer Name and Order Date
4. Get Orders By Product Category When Profit must be greater than 0 And Sorted by Customer Name
5. Get Orders By Product Category In period and sorted by Sales
6. Quit Program
Enter the number of Query from above...
```

2. First Query:

```
Enter the number of Query from above...
1
Enter the rows number of Orders that you want show :
10
+-------+----------+--------+---------+--------------+--------+---------+-----------------+--------------+---------------+
|orderId|      date|quantity|    sales|          mode|  profit|unitPrice|     customerName|customerSegment|productCategory|
+-------+----------+--------+---------+--------------+--------+---------+-----------------+--------------+---------------+
|      3|2010-10-13|       6|   261.54|   Regular Air| -213.25|    38.94|Muhammed MacIntyre| Small Business|Office Supplies|
|      6|2012-02-20|       2|     6.93|   Regular Air|   -4.64|     2.08|       Ruben Dartt|      Corporate|Office Supplies|
|     32|2011-07-15|      26|  2808.08|   Regular Air| 1054.82|   107.53|     Liz Pelletier|      Corporate|      Furniture|
|     32|2011-07-15|      24|   1761.4|Delivery Truck|-1748.56|    70.89|     Liz Pelletier|      Corporate|      Furniture|
|     32|2011-07-15|      23| 160.2335|   Regular Air| -85.129|     7.99|     Liz Pelletier|      Corporate|     Technology|
|     32|2011-07-15|      15|   140.56|   Regular Air| -128.38|     8.46|     Liz Pelletier|      Corporate|     Technology|
|     35|2011-10-22|      30|   288.56|   Regular Air|   60.72|     9.11|   Julie Creighton|      Corporate|Office Supplies|
|     35|2011-10-22|      14| 1892.848|   Regular Air|  48.987|   155.99|   Julie Creighton|      Corporate|     Technology|
|     36|2011-11-02|      46|2484.7455|   Regular Air| 657.477|    65.99|  Sample Company A|    Home Office|     Technology|
|     65|2011-03-17|      32|  3812.73|   Regular Air|  1470.3|   115.79|     Tamara Dahlen|      Corporate|     Technology|
+-------+----------+--------+---------+--------------+--------+---------+-----------------+--------------+---------------+
only showing top 10 rows

Menu:
1. Get All Orders form CSV file
2. Get Orders By Customer Name
3. Get Orders(Name,Sales,Profit,Category)  By Customer Name and Order Date
4. Get Orders By Product Category When Profit must be greater than 0 And Sorted by Customer Name
5. Get Orders By Product Category In period and sorted by Sales
6. Quit Program
Enter the number of Query from above...
```

3. Second Query:

```
Enter the number of Query from above...
2
Enter the customer name :  Ex= Liz Pelletier
Liz Pelletier
+-------+----------+--------+--------+--------------+--------+---------+-------------+--------------+---------------+
|orderId|      date|quantity|   sales|          mode|  profit|unitPrice| customerName|customerSegment|productCategory|
+-------+----------+--------+--------+--------------+--------+---------+-------------+--------------+---------------+
|     32|2011-07-15|      26| 2808.08|   Regular Air| 1054.82|   107.53|Liz Pelletier|      Corporate|      Furniture|
|     32|2011-07-15|      24|  1761.4|Delivery Truck|-1748.56|    70.89|Liz Pelletier|      Corporate|      Furniture|
|     32|2011-07-15|      23|160.2335|   Regular Air| -85.129|     7.99|Liz Pelletier|      Corporate|     Technology|
|     32|2011-07-15|      15|  140.56|   Regular Air| -128.38|     8.46|Liz Pelletier|      Corporate|     Technology|
|     32|2008-07-15|      26|    null|   Regular Air|    null|   107.53|Liz Pelletier|      Corporate|      Furniture|
|     32|2008-07-15|      24|    null|Delivery Truck|    null|    70.89|Liz Pelletier|      Corporate|      Furniture|
|     32|2008-07-15|      23|    null|   Regular Air|    null|     7.99|Liz Pelletier|      Corporate|     Technology|
|     32|2008-07-15|      15|    null|   Regular Air|    null|     8.46|Liz Pelletier|      Corporate|     Technology|
+-------+----------+--------+--------+--------------+--------+---------+-------------+--------------+---------------+

Menu:
```

4. Third Query:

```
6. Quit Program
Enter the number of Query from above...
3
Enter the customer name : Ex= Liz Pelletier
Liz Pelletier
Enter the order date with this format (yyyy-MM-dd) : Ex= 2008-07-15
2008-07-15
+-------------+----------+-----+------+---------------+
| customerName|      date|sales|profit|productCategory|
+-------------+----------+-----+------+---------------+
|Liz Pelletier|2008-07-15| null|  null|      Furniture|
|Liz Pelletier|2008-07-15| null|  null|      Furniture|
|Liz Pelletier|2008-07-15| null|  null|     Technology|
|Liz Pelletier|2008-07-15| null|  null|     Technology|
+-------------+----------+-----+------+---------------+

Menu:
1. Get All Orders form CSV file
2. Get Orders By Customer Name
3. Get Orders(Name,Sales,Profit,Category)  By Customer Name and Order Date
4. Get Orders By Product Category When Profit must be greater than 0 And Sorted by Customer Name
5. Get Orders By Product Category In period and sorted by Sales
6. Quit Program
Enter the number of Query from above...
```

5. Fourth Query :

```
6. Quit Program
Enter the number of Query from above...
4
Enter the product category : Ex: Office Supplies
Office Supplies
+-------+----------+--------+--------+-------------+----------------+---------+-------------------+---------------+---------------+
|orderId|      date|quantity|   sales|         mode|          profit|unitPrice|       customerName|customerSegment|productCategory|
+-------+----------+--------+--------+-------------+----------------+---------+-------------------+---------------+---------------+
|   4132|2011-05-28|       5|   14.76|  Regular Air|            1.32|     2.88|      Aaron Bergman|      Corporate|Office Supplies|
|   2436|2011-09-10|      14|  330.21|  Express Air|           83.24|    22.84|         Adam Hart|      Corporate|Office Supplies|
|   2532|2011-10-10|      39|  282.07|  Regular Air|          140.01|     7.31|        Alan Barnes|      Corporate|Office Supplies|
|   2978|2011-05-02|      36| 5410.95|  Regular Air|         2077.91|   162.93|  Alan Schoenberger|      Corporate|Office Supplies|
|   6465|2011-08-12|      36|  240.14|  Express Air|           57.78|     6.88|       Alan Shonely| Small Business|Office Supplies|
|   6912|2011-12-10|      14|  294.52|  Regular Air|           15.34|    21.78|       Alan Shonely| Small Business|Office Supplies|
|   1637|2011-02-07|      47| 1348.57|  Regular Air|           19.57|    30.98|Aleksandra Gannaway|      Corporate|Office Supplies|
|   2659|2010-12-17|      50|  881.65|  Regular Air|           41.27|    18.97|     Alice McCarthy|      Corporate|Office Supplies|
|   4128|2011-10-06|      38|  391.42|  Regular Air|           25.03|     9.99|       Allen Golden|       Consumer|Office Supplies|
|    998|2009-11-25|      16|  248.26|  Regular Air|            93.8|    15.74|   Allen Rosenblatt| Small Business|Office Supplies|
|   2438|2010-07-14|      20|  278.68|  Regular Air|          8.8995|    14.48|       Andrew Allen|      Corporate|Office Supplies|
|   2084|2010-03-16|      16|  692.73|  Regular Air|           94.97|    40.99|        Andy Reiter|      Corporate|Office Supplies|
|   3585|2009-11-22|      27|  259.72|  Regular Air|           78.05|    10.06|          Ann Blume|      Corporate|Office Supplies|
|   5894|2009-08-12|       7|  384.33|  Regular Air|         87.6775|     58.1|          Ann Chong|      Corporate|Office Supplies|
|   1702|2011-05-06|      23|   67.24|  Regular Air|             4.9|     2.84|       Annie Cyprus|    Home Office|Office Supplies|
|   4676|2011-08-31|      50|  187.83|  Regular Air|           85.96|     3.75|       Annie Cyprus|    Home Office|Office Supplies|
|   5860|2012-02-19|      12|   56.73|  Regular Air|            8.33|     4.84|      Annie Thurman|       Consumer|Office Supplies|
|    928|2011-03-01|      21| 1222.68|  Express Air|          300.97|    59.98|  Anthony O'Donnell|       Consumer|Office Supplies|
|   6020|2009-09-21|      23|   66.09|  Regular Air|           13.71|     2.88|     Arianne Irving|      Corporate|Office Supplies|
|   1699|2010-12-29|      40|  430.88|  Regular Air|            39.0|    11.09|       Art Ferguson|      Corporate|Office Supplies|
|   6309|2010-06-09|      10|   29.79|  Regular Air|            3.76|     2.88|         Art Miller|      Corporate|Office Supplies|
|     66|2009-01-19|      41|  108.15|  Regular Air|            7.57|     2.88|      Arthur Gainer|       Consumer|Office Supplies|
```

6. Fifth Query:

```
6. Quit Program
Enter the number of Query from above...
5
Enter the product category : Ex: Office Supplies
Office Supplies
Enter the start date : Ex: 2010-03-16
2010-03-16
Enter the end date : Ex: 2011-08-31
2011-08-31
+------------------+----------+--------+--------+-------------------+---------+---------------+
|      customerName|      date|   sales|quantity|             profit|unitPrice|customerSegment|
+------------------+----------+--------+--------+-------------------+---------+---------------+
|   Ricardo Emerson|2011-08-07|    2.24|       1|              -1.97|     1.48|      Corporate|
|       Bobby Elias|2011-08-04|   11.15|       6|              -2.62|     1.81| Small Business|
|        Sean Wendt|2010-08-13|    12.8|       4|              -2.21|     2.88|      Corporate|
|     Aaron Bergman|2011-05-28|   14.76|       5|               1.32|     2.88|      Corporate|
|      Tracy Poddar|2011-01-24|   14.96|       4|-0.0300000000000002|     3.69|      Corporate|
|      Tracy Poddar|2011-01-24|   15.69|       4|              -1.59|     3.85|      Corporate|
|  Maureen Fritzler|2010-12-11|   16.08|       4|             -3.749|      3.8|      Corporate|
|     Edward Becker|2011-03-16|   17.12|       4|             -2.553|      3.8|      Corporate|
|       Rick Wilson|2010-11-24|   17.89|       4|              -1.15|     3.98|       Consumer|
|    Karen Ferguson|2011-05-24|    23.7|       1|              -8.97|    19.84|    Home Office|
|        Evan Henry|2010-11-25|   24.72|       8|             -21.47|     2.74|      Corporate|
|     Deirdre Greer|2010-05-25|   26.53|       3|              -15.0|     6.68|    Home Office|
|     Thomas Boland|2011-07-15|   28.01|       6|               3.46|     4.98|      Corporate|
| Michael Dominguez|2010-07-25|   29.42|       4|              14.43|     3.69|    Home Office|
|        Art Miller|2010-06-09|   29.79|      10|               3.76|     2.88|      Corporate|
|     Brenda Bowman|2010-11-18|   31.77|      12|               8.34|     2.61| Small Business|
|       Peter McVee|2011-03-17|   31.87|       8|              -1.25|     4.26| Small Business|
|     Eileen Kiefer|2010-10-21|   34.59|      14|             -60.13|     2.18|       Consumer|
|      Liz Carlisle|2010-12-20|   35.58|       7|             -19.95|     4.42|      Corporate|
|    Aaron Smayling|2010-12-04|   38.26|      22|              -2.34|     1.76| Small Business|
```

7. Quit Program:

```
Menu:
1. Get All Orders form CSV file
2. Get Orders By Customer Name
3. Get Orders(Name,Sales,Profit,Category)  By Customer Name and Order Date
4. Get Orders By Product Category When Profit must be greater than 0 And Sorted by Customer Name
5. Get Orders By Product Category In period and sorted by Sales
6. Quit Program
Enter the number of Query from above...
6
Quitting Program...
Thanks for using this Program...

Process finished with exit code 0
```

XII. Wrapping Up:

In this project, we have created a spark application using **Spark Core** and **Spark SQL** with **Java**. Here, we have loaded the CSV file into **Data Frame** without using any external package. Also, The CSV format is the common file format which gets used as a source file in most cases.
If you want to test the examples above, you will find my Github code link:
Read CSV file into Dataframe And Perform some Queries