Code Logic - Retail Data Analysis

PySpark Code Logic

Step 1: Importing the required modules - *System dependencies for CDH and declaring PySpark environment variables*

 # importing the modules - System dependencies for CDH

```
from pyspark.sql import SparkSession from
pyspark.sql.functions import * from
pyspark.sql.types import * from
pyspark.sql.functions import from_json from
pyspark.sql.window import Window
```

Step 2: Creating Initial Spark Session

 # Initializing Spark Session

```
spark = SparkSession \
    .builder \
    .appName("RetailDataAnalysis") \
    .getOrCreate() spark.sparkContext.setLogLevel('ERROR')
```

Step 3: Reading the Input from Kafka

 # Read Input from kafka

```
rawOrder = spark \
    .readStream \
    .format("kafka") \
    .option("kafka.bootstrap.servers","18.211.252.152:9092") \
    .option("startingOffsets", "latest") \
    .option("subscribe","real-time-project") \
    .load()
```

Step 4: _Defining the schema for incoming data JSON file data_

 # Defining the Schema

```
jsonSchema = StructType() \
        .add("invoice_no", LongType()) \
        .add("country", StringType()) \
        .add("timestamp", TimestampType()) \
        .add("type", StringType()) \
        .add("items", ArrayType(StructType([
        StructField("SKU", StringType()),  StructField("title",
        StringType()),
        StructField("unit_price", DoubleType()),
        StructField("quantity", IntegerType())
])))
```

Step 5: _Creating dataframe from incoming data_

 # Creating dataframe from input data after applying the schema

```
orderStream = orderRaw.select(from_json(col("value").cast("string"),
jsonSchema).alias("data")).select("data.*")
```

Step 6: _Defining UDF functions (total_tems, total_cost, is_order, is_return), conversion to UDF types and calculating columns_

_# UDF for calculating total_items_

```
def items_TotalCount(items):
        total_count = 0 for item in items: total_count =
        total_count + item['quantity']
        return total_count
```

 # UDF for calculating order type

```python
def is_order(type):
        if type=="ORDER":
                return 1
        else:
                return 0
```

UDF for calculating return type

```python
def is_return(type):
        if type=="RETURN":
                return 1
        else:
                return 0
```

UDF for calculating total_cost

```python
def TotalCostSum (items,type):\ total_sum
        = 0
        for item in items:
     total_sum = total_sum + item['unit_price'] * item['quantity']    if
type=="RETURN":
                return total_sum * (-1)
else:
                return total_sum
```

Converting to UDF's with the utility functions

```python
isorder = udf(is_order, IntegerType()) isreturn =
udf(is_return, IntegerType()) totalcount =
udf(items_TotalCount, IntegerType()) totalcost =
udf(TotalCostSum, DoubleType())
```

Calculating columns(total_cost, total_items, is_order, is_return)

```
order_stream = orderStream \
        .withColumn("total_cost", totalcost(orderStream.items, orderStream.type)) \
        .withColumn("total_items", totalcount(orderStream.items)) \
        .withColumn("is_order", isorder(orderStream.type)) \
        .withColumn("is_return", isreturn(orderStream.type))
```

Step 7: Writing intermediate dataset to console with 1 Minute Interval

```
orderStreamOutput = order_stream \
        .select("invoice_no", "country",
        "timestamp","total_items","total_cost","is_order","is_return") \
        .writeStream \
        .outputMode("append") \
        .format("console") \
        .option("truncate", "false") \
        .trigger(processingTime="1 minute") \
        .start()
```

Step 8:*Calculating time-based KPIs using with Watermark and groupBy*

 # Calculating time based KPIs

```
timeBasedKPIs = order_stream \
        .withWatermark("timestamp", "1 minute") \
        .groupby(window("timestamp", "1 minute", "1 minute")) \
        .agg(count("invoice_no").alias("OPM"), sum("total_cost").alias("total_sales_volume"),
                avg("total_cost").alias("average_transaction_size"),
avg("is_return").alias("rate_of_return")) \
        .select("window", "OPM", "total_sales_volume", "average_transaction_size",
        "rate_of_return")
```

 # write stream for time based KPIs

```python
timeBasedKPIsOutput = timeBasedKPIs \ .writeStream
        \
        .outputMode("append") \
        .format("json") \
        .option("format","append") \
        .option("truncate", "false") \
        .option("path", "time-wise-kpi") \
        .option("checkpointLocation", "time-kpi") \
        .trigger(processingTime="1 minute") \
        .start()
```

Step 9: *Calculating time-based and country-based KPIs using withWatermark and groupBy*

```python
# Calculating time-based and country-based KPIs

timeAndCountryBasedKPIs = order_stream \
        .withWatermark("timestamp", "1 minute") \
        .groupby(window("timestamp", "1 minute", "1 minute"), "country") \
        .agg(count("invoice_no").alias("OPM"),
sum("total_cost").alias("total_sales_volume"),
avg("is_return").alias("rate_of_return")) \
        .select("window", "country", "OPM", "total_sales_volume", "rate_of_return")
```

```python
# write stream for time and country based KPIs

timeAndCountryBasedKPIsOutput = timeAndCountryBasedKPIs \
        .writeStream \
        .outputMode("Append") \
        .format("json") \
        .option("format","append") \
        .option("truncate", "false") \
        .option("path", "time-country-wise-kpi") \
        .option("checkpointLocation","time-country-kpi") \
.trigger(processingTime="1 minute") \
        .start()
```

Step 10: Waiting for the termination of stream infinitely

 # Waiting infinitely to read the data

timeAndCountryBasedKPIsOutput.awaitTermination()

Console Commands and Analysis

Spark Submit Command and too Console-output file

spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 spark-streaming.py
18.211.252.152 9092 real-time-project > Console-output

Console Output

```
Krishna@Krishnas-MacBook-Pro retail_data_analysis % cat Console-output
---------------------------------------------
Batch: 0
---------------------------------------------
+----------+-------+---------+----------+-----------+--------+---------+
|invoice_no|country|timestamp|total_cost|total_items|is_order|is_return|
+----------+-------+---------+----------+-----------+--------+---------+
+----------+-------+---------+----------+-----------+--------+---------+

---------------------------------------------
Batch: 1
---------------------------------------------
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+
|invoice_no    |country       |timestamp          |total_cost         |total_items|is_order|is_return|
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+
|154132550934204|United Kingdom|2022-08-27 09:02:04|43.870000000000005|7          |1       |0        |
|154132550934205|EIRE          |2022-08-27 09:02:10|34.8              |24         |1       |0        |
|154132550934206|United Kingdom|2022-08-27 09:02:10|32.11             |9          |1       |0        |
|154132550934207|United Kingdom|2022-08-27 09:02:11|1.25              |1          |1       |0        |
|154132550934208|United Kingdom|2022-08-27 09:02:21|17.439999999999998|4          |1       |0        |
|154132550934209|United Kingdom|2022-08-27 09:02:22|67.95             |35         |1       |0        |
|154132550934210|United Kingdom|2022-08-27 09:02:31|51.81             |31         |1       |0        |
|154132550934211|United Kingdom|2022-08-27 09:02:38|1.45              |1          |1       |0        |
|154132550934212|United Kingdom|2022-08-27 09:02:50|41.169999999999995|21         |1       |0        |
|154132550934213|United Kingdom|2022-08-27 09:02:52|112.38            |242        |1       |0        |
|154132550934214|United Kingdom|2022-08-27 09:02:53|49.67999999999999 |28         |1       |0        |
|154132550934215|United Kingdom|2022-08-27 09:02:59|220.54999999999995|92         |1       |0        |
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+

---------------------------------------------
Batch: 2
---------------------------------------------
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+
|invoice_no    |country       |timestamp          |total_cost         |total_items|is_order|is_return|
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+
|154132550934216|United Kingdom|2022-08-27 09:03:02|30.180000000000003|12         |1       |0        |
|154132550934217|United Kingdom|2022-08-27 09:03:03|163.93999999999997|37         |1       |0        |
|154132550934218|United Kingdom|2022-08-27 09:03:12|239.37000000000003|21         |1       |0        |
|154132550934219|Germany       |2022-08-27 09:03:31|24.0              |10         |1       |0        |
|154132550934220|United Kingdom|2022-08-27 09:03:48|200.56999999999996|20         |1       |0        |
|154132550934221|United Kingdom|2022-08-27 09:03:52|-15.6             |8          |0       |1        |
|154132550934222|United Kingdom|2022-08-27 09:03:55|157.51000000000002|41         |1       |0        |
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+

---------------------------------------------
Batch: 3
---------------------------------------------
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+
|invoice_no    |country       |timestamp          |total_cost         |total_items|is_order|is_return|
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+
|154132550934223|United Kingdom|2022-08-27 09:04:05|33.44999999999996 |51         |1       |0        |
|154132550934224|United Kingdom|2022-08-27 09:04:07|66.61             |57         |1       |0        |
|154132550934225|EIRE          |2022-08-27 09:04:07|32.92             |29         |1       |0        |
|154132550934226|United Kingdom|2022-08-27 09:04:22|89.88999999999999 |39         |1       |0        |
|154132550934227|United Kingdom|2022-08-27 09:04:24|19.79             |25         |1       |0        |
|154132550934228|United Kingdom|2022-08-27 09:04:25|4.949999999999999 |3          |1       |0        |
|154132550934229|United Kingdom|2022-08-27 09:04:30|27.330000000000002|9          |1       |0        |
|154132550934230|United Kingdom|2022-08-27 09:04:32|13.16             |4          |1       |0        |
|154132550934231|United Kingdom|2022-08-27 09:04:35|50.51999999999996 |12         |1       |0        |
|154132550934232|United Kingdom|2022-08-27 09:04:41|11.649999999999999|7          |1       |0        |
|154132550934233|United Kingdom|2022-08-27 09:04:50|79.67999999999999 |96         |1       |0        |
```

```
Batch: 3
---------------------------------------------
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+
|invoice_no    |country       |timestamp          |total_cost         |total_items|is_order|is_return|
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+
|154132550934223|United Kingdom|2022-08-27 09:04:05|33.44999999999996 |51         |1       |0        |
|154132550934224|United Kingdom|2022-08-27 09:04:07|66.61             |57         |1       |0        |
|154132550934225|EIRE          |2022-08-27 09:04:07|32.92             |29         |1       |0        |
|154132550934226|United Kingdom|2022-08-27 09:04:22|89.88999999999999 |39         |1       |0        |
|154132550934227|United Kingdom|2022-08-27 09:04:24|19.79             |25         |1       |0        |
|154132550934228|United Kingdom|2022-08-27 09:04:25|4.949999999999999 |3          |1       |0        |
|154132550934229|United Kingdom|2022-08-27 09:04:30|27.330000000000002|9          |1       |0        |
|154132550934230|United Kingdom|2022-08-27 09:04:32|13.16             |4          |1       |0        |
|154132550934231|United Kingdom|2022-08-27 09:04:35|50.51999999999996 |12         |1       |0        |
|154132550934232|United Kingdom|2022-08-27 09:04:41|11.649999999999999|7          |1       |0        |
|154132550934233|United Kingdom|2022-08-27 09:04:50|79.67999999999999 |96         |1       |0        |
|154132550934234|United Kingdom|2022-08-27 09:04:51|0.85              |1          |1       |0        |
|154132550934235|United Kingdom|2022-08-27 09:04:56|58.55999999999999 |57         |1       |0        |
|154132550934236|United Kingdom|2022-08-27 09:04:57|17.4              |12         |1       |0        |
|154132550934237|EIRE          |2022-08-27 09:04:57|23.79             |9          |1       |0        |
|154132550934238|United Kingdom|2022-08-27 09:04:58|22.66             |12         |1       |0        |
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+

---------------------------------------------
Batch: 4
---------------------------------------------
+--------------+--------------+-------------------+----------+-----------+--------+---------+
|invoice_no    |country       |timestamp          |total_cost|total_items|is_order|is_return|
+--------------+--------------+-------------------+----------+-----------+--------+---------+
|154132550934239|United Kingdom|2022-08-27 09:05:14|738.95    |187        |1       |0        |
|154132550934240|United Kingdom|2022-08-27 09:05:20|79.01     |87         |1       |0        |
|154132550934241|United Kingdom|2022-08-27 09:05:29|87.76     |28         |1       |0        |
|154132550934242|United Kingdom|2022-08-27 09:05:32|-91.75    |77         |0       |1        |
|154132550934243|United Kingdom|2022-08-27 09:05:35|19.8      |4          |1       |0        |
|154132550934244|United Kingdom|2022-08-27 09:05:35|41.5      |42         |1       |0        |
|154132550934245|United Kingdom|2022-08-27 09:05:45|73.28     |37         |1       |0        |
+--------------+--------------+-------------------+----------+-----------+--------+---------+

---------------------------------------------
Batch: 5
---------------------------------------------
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+
|invoice_no    |country       |timestamp          |total_cost         |total_items|is_order|is_return|
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+
|154132550934246|United Kingdom|2022-08-27 09:06:08|20.279999999999998 |52         |1       |0        |
|154132550934247|United Kingdom|2022-08-27 09:06:09|25.86              |6          |1       |0        |
|154132550934248|United Kingdom|2022-08-27 09:06:17|49.6               |10         |1       |0        |
|154132550934249|United Kingdom|2022-08-27 09:06:28|26.69              |9          |1       |0        |
|154132550934250|United Kingdom|2022-08-27 09:06:29|5.04               |12         |1       |0        |
|154132550934251|Netherlands   |2022-08-27 09:06:33|45.55              |15         |1       |0        |
|154132550934252|United Kingdom|2022-08-27 09:06:33|103.73000000000002 |59         |1       |0        |
|154132550934253|United Kingdom|2022-08-27 09:06:34|19.55              |21         |1       |0        |
|154132550934254|United Kingdom|2022-08-27 09:06:35|118.31             |33         |1       |0        |
|154132550934255|United Kingdom|2022-08-27 09:06:41|30.990000000000002 |9          |1       |0        |
|154132550934256|United Kingdom|2022-08-27 09:06:43|1.25               |1          |1       |0        |
|154132550934257|Cyprus        |2022-08-27 09:06:44|2.95               |1          |1       |0        |
|154132550934258|United Kingdom|2022-08-27 09:06:55|41.870000000000005 |13         |1       |0        |
+--------------+--------------+-------------------+-------------------+-----------+--------+---------+

---------------------------------------------
Batch: 6
```

```
------------------------------------------
Batch: 6
------------------------------------------

+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|invoice_no    |country       |timestamp           |total_cost       |total_items|is_order|is_return|
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|1541325509342591|Austria     |2022-08-27 09:07:04|19.08            |2          |1       |0        |
|1541325509342601|United Kingdom|2022-08-27 09:07:07|51.45            |35         |1       |0        |
|1541325509342611|United Kingdom|2022-08-27 09:07:09|13.2000000000001 |24         |1       |0        |
|1541325509342621|Hong Kong     |2022-08-27 09:07:20|49.8999999999999 |38         |1       |0        |
|1541325509342631|Hong Kong     |2022-08-27 09:07:20|63.65            |49         |1       |0        |
|1541325509342641|United Kingdom|2022-08-27 09:07:20|27.3699999999997 |13         |1       |0        |
|1541325509342651|United Kingdom|2022-08-27 09:07:22|19.55            |3          |1       |0        |
|1541325509342661|United Kingdom|2022-08-27 09:07:29|-99.56           |62         |0       |1        |
|1541325509342671|United Kingdom|2022-08-27 09:07:32|-27.080000000002 |16         |0       |1        |
|1541325509342681|United Kingdom|2022-08-27 09:07:41|15.04            |12         |1       |0        |
|1541325509342691|United Kingdom|2022-08-27 09:07:43|10.02            |8          |1       |0        |
|1541325509342701|United Kingdom|2022-08-27 09:07:50|288.45000000000005|149       |1       |0        |
|1541325509342711|United Kingdom|2022-08-27 09:07:50|14.85000000000001|3          |1       |0        |
|1541325509342721|United Kingdom|2022-08-27 09:07:51|12.48            |6          |1       |0        |
|1541325509342731|United Kingdom|2022-08-27 09:07:54|66.64            |20         |1       |0        |
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+

------------------------------------------
Batch: 7
------------------------------------------

+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|invoice_no    |country       |timestamp           |total_cost       |total_items|is_order|is_return|
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|1541325509342741|United Kingdom|2022-08-27 09:08:03|1.25             |1          |1       |0        |
|1541325509342751|United Kingdom|2022-08-27 09:08:06|58.5             |38         |1       |0        |
|1541325509342761|United Kingdom|2022-08-27 09:08:07|0.85             |1          |1       |0        |
|1541325509342771|United Kingdom|2022-08-27 09:08:09|28.4599999999999 |8          |1       |0        |
|1541325509342781|United Kingdom|2022-08-27 09:08:10|20.65            |17         |1       |0        |
|1541325509342791|United Kingdom|2022-08-27 09:08:11|15.0             |4          |1       |0        |
|1541325509342801|United Kingdom|2022-08-27 09:08:11|512.400000000001 |146        |1       |0        |
|1541325509342811|United Kingdom|2022-08-27 09:08:13|62.5199999999996 |14         |1       |0        |
|1541325509342821|United Kingdom|2022-08-27 09:08:16|42.85            |18         |1       |0        |
|1541325509342831|United Kingdom|2022-08-27 09:08:26|30.3100000000002 |17         |1       |0        |
|1541325509342841|France        |2022-08-27 09:08:35|12.5             |2          |1       |0        |
|1541325509342851|United Kingdom|2022-08-27 09:08:44|21.5             |10         |1       |0        |
|1541325509342861|United Kingdom|2022-08-27 09:08:54|429.6            |109        |1       |0        |
|1541325509342871|United Kingdom|2022-08-27 09:08:55|55.61            |45         |1       |0        |
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+

------------------------------------------
Batch: 8
------------------------------------------

+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|invoice_no    |country       |timestamp           |total_cost       |total_items|is_order|is_return|
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|1541325509342881|United Kingdom|2022-08-27 09:08:56|23.53            |8          |1       |0        |
|1541325509342891|United Kingdom|2022-08-27 09:09:08|18.75            |5          |1       |0        |
|1541325509342901|United Kingdom|2022-08-27 09:09:14|58.61            |39         |1       |0        |
|1541325509342911|United Kingdom|2022-08-27 09:09:15|165.66           |71         |1       |0        |
|1541325509342921|United Kingdom|2022-08-27 09:09:18|46.010000000000005|37        |1       |0        |
|1541325509342931|United Kingdom|2022-08-27 09:09:20|244.9500000000002|42         |1       |0        |
|1541325509342941|United Kingdom|2022-08-27 09:09:21|33.7             |16         |1       |0        |
|1541325509342951|United Kingdom|2022-08-27 09:09:22|2.46             |1          |1       |0        |
|1541325509342961|United Kingdom|2022-08-27 09:09:28|170.35           |195        |1       |0        |
|1541325509342971|United Kingdom|2022-08-27 09:09:29|-58.14           |12         |0       |1        |
```

```
|1541325509342861|United Kingdom|2022-08-27 09:08:54|429.6            |109        |1       |0        |
|1541325509342871|United Kingdom|2022-08-27 09:08:55|55.61            |45         |1       |0        |
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+

------------------------------------------
Batch: 8
------------------------------------------

+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|invoice_no    |country       |timestamp           |total_cost       |total_items|is_order|is_return|
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|1541325509342881|United Kingdom|2022-08-27 09:08:56|23.53            |8          |1       |0        |
|1541325509342891|United Kingdom|2022-08-27 09:09:08|18.75            |5          |1       |0        |
|1541325509342901|United Kingdom|2022-08-27 09:09:14|58.61            |39         |1       |0        |
|1541325509342911|United Kingdom|2022-08-27 09:09:15|165.66           |71         |1       |0        |
|1541325509342921|United Kingdom|2022-08-27 09:09:18|46.010000000000005|37        |1       |0        |
|1541325509342931|United Kingdom|2022-08-27 09:09:20|244.9500000000002|42         |1       |0        |
|1541325509342941|United Kingdom|2022-08-27 09:09:21|33.7             |16         |1       |0        |
|1541325509342951|United Kingdom|2022-08-27 09:09:22|2.46             |1          |1       |0        |
|1541325509342961|United Kingdom|2022-08-27 09:09:28|170.35           |195        |1       |0        |
|1541325509342971|United Kingdom|2022-08-27 09:09:29|-58.14           |12         |0       |1        |
|1541325509342981|United Kingdom|2022-08-27 09:09:44|17.25            |3          |1       |0        |
|1541325509342991|United Kingdom|2022-08-27 09:09:46|230.4299999999995|44         |1       |0        |
|1541325509343001|United Kingdom|2022-08-27 09:09:54|29.0             |12         |1       |0        |
|1541325509343011|Netherland    |2022-08-27 09:09:54|30.3             |6          |1       |0        |
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+

------------------------------------------
Batch: 9
------------------------------------------

+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|invoice_no    |country       |timestamp           |total_cost       |total_items|is_order|is_return|
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|1541325509343021|United Kingdom|2022-08-27 09:10:03|63.43000000000001|21         |1       |0        |
|1541325509343031|United Kingdom|2022-08-27 09:10:18|138.91           |40         |1       |0        |
|1541325509343041|United Kingdom|2022-08-27 09:10:19|35.53            |5          |1       |0        |
|1541325509343051|United Kingdom|2022-08-27 09:10:43|34.75            |11         |1       |0        |
|1541325509343061|United Kingdom|2022-08-27 09:10:51|-65.38           |26         |0       |1        |
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+

------------------------------------------
Batch: 10
------------------------------------------

+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|invoice_no    |country       |timestamp           |total_cost       |total_items|is_order|is_return|
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+
|1541325509343071|United Kingdom|2022-08-27 09:11:02|34.28            |14         |1       |0        |
|1541325509343081|United Kingdom|2022-08-27 09:11:05|51.21            |29         |1       |0        |
|1541325509343091|United Kingdom|2022-08-27 09:11:13|345.1            |138        |1       |0        |
|1541325509343101|United Kingdom|2022-08-27 09:11:25|3.75             |1          |1       |0        |
|1541325509343111|United Kingdom|2022-08-27 09:11:28|22.08            |14         |1       |0        |
|1541325509343121|United Kingdom|2022-08-27 09:11:30|62.9             |34         |1       |0        |
|1541325509343131|EIRE          |2022-08-27 09:11:33|19.3             |6          |1       |0        |
|1541325509343141|United Kingdom|2022-08-27 09:11:35|9.72             |7          |1       |0        |
|1541325509343151|United Kingdom|2022-08-27 09:11:41|60.7499999999999 |5          |1       |0        |
|1541325509343161|United Kingdom|2022-08-27 09:11:43|220.86           |19         |1       |0        |
|1541325509343171|Switzerland   |2022-08-27 09:11:46|36.98000000000049|9          |1       |0        |
|1541325509343181|Switzerland   |2022-08-27 09:11:53|60.64            |12         |1       |0        |
|1541325509343191|Italy         |2022-08-27 09:11:54|15.0             |12         |1       |0        |
+--------------+--------------+--------------------+-----------------+-----------+--------+---------+

Hkrishna@Krishnas-MacBook-Pro retail_data_analysis %
```

Hadoop part files

```
hkrishna@Krishnas-MacBook-Pro Output % tree
 — Country-and-timebased-KPI
   └── time-country-wise-kpi
       └── time-country-wise-kpi
           ├── _spark_metadata
           │   ├── 0
           │   ├── 1
           │   ├── 10
           │   ├── 11
           │   ├── 12
           │   ├── 13
           │   ├── 14
           │   ├── 2
           │   ├── 3
           │   ├── 4
           │   ├── 5
           │   ├── 6
           │   ├── 7
           │   ├── 8
           │   └── 9.compact
           ├── part-00000-0196de06-0879-4821-be0e-081ec5c256f8-c000.json
           ├── part-00000-25284aab-6393-40a2-bf36-a189be1bf1cd-c000.json
           ├── part-00000-32296765-2dd2-42bb-91ee-a13bdf1e8327-c000.json
           ├── part-00000-3519f0a1-5b0d-4cd2-9a10-119a5a0e23ed-c000.json
           ├── part-00000-45c70e01-c04e-46ed-872d-c713e0e7f61d-c000.json
           ├── part-00000-7c4ae1c2-8500-4673-bf5f-1c906ba06135-c000.json
           ├── part-00000-831b5819-7a4e-4212-ba35-2e784c0d0421-c000.json
           ├── part-00000-83ca12d0-59d2-4348-8885-031942b8874b-c000.json
           ├── part-00000-8b88acea-afbe-4441-8b2e-0ef5fcff16af-c000.json
           ├── part-00000-a237f2cb-7521-4a8c-bca3-615a64b3cb3a-c000.json
           ├── part-00000-a724ec59-bc11-4b95-90f5-33f17d86bff2-c000.json
           ├── part-00000-de17eb37-5f53-4098-9379-b39f45c39127-c000.json
           ├── part-00000-df236787-21b6-4aa8-9647-022f669e29ed-c000.json
           ├── part-00000-e6353d2d-5c90-483d-ab9e-8c01319f12d7-c000.json
           ├── part-00000-f57ea320-21d3-47e9-adbb-e76562979eea-c000.json
           ├── part-00002-3f65b776-731d-4f39-b3bd-5fd1c2841a61-c000.json
           ├── part-00012-c34a4fba-31b1-414a-9b7b-99173af7b5d1-c000.json
           ├── part-00019-befc29a8-8460-4b5c-87b9-c7db233ada10-c000.json
           ├── part-00022-4c57b3e6-144b-42ff-aeb6-68485b6835f0-c000.json
           ├── part-00034-a759390a-8103-4b5b-a305-9d7497485d32-c000.json
           ├── part-00045-52a168c7-5564-4842-81f1-d9c16a76d765-c000.json
           ├── part-00049-964051e4-e8a7-47a0-9233-8645a42c8088-c000.json
           ├── part-00065-6191ce46-d0e2-4b1e-ba50-aab4579a73fb-c000.json
           ├── part-00073-2631e7e5-c252-42db-8912-6ff46820f705-c000.json
           ├── part-00081-41c88ae7-58a1-4d76-b583-b9f378ad239a-c000.json
           ├── part-00082-59a9a162-f76f-419b-9755-664bb9003638-c000.json
           ├── part-00086-deb0d215-50b8-4b76-9477-2c8f26f00021-c000.json
           ├── part-00095-ded0b8be-ab41-4a41-9e60-9542f94578fb-c000.json
           ├── part-00108-83c35027-a744-41ba-9f88-e46da4787b58-c000.json
           ├── part-00132-c347d6b5-1232-4a36-b5ad-148d2251e3ab-c000.json
           ├── part-00146-ed982e74-64a2-439a-b69c-ac0371366934-c000.json
           ├── part-00153-14d8dcd9-b8ca-433d-ab49-be93600d9fc6-c000.json
           ├── part-00171-619e513b-bd5a-47c0-b396-4734228160ee-c000.json
           ├── part-00175-44606e86-3c64-4fcb-80f1-6eb766f281db-c000.json
           ├── part-00180-328b9507-76e6-4f05-b506-3c16e01d4285-c000.json
           └── part-00185-ffa2a99d-0618-41c2-abab-cb73865996b7-c000.json
 — Timebased-KPI
   └── time-wise-kpi
       └── time-wise-kpi
```

```
           ├── part-00081-5d66a435-714e-4bf9-bc42-c8e3e5ccd19f-c000.json
           ├── part-00082-59a9a162-f76f-419b-9755-664bb9003638-c000.json
           ├── part-00086-deb0d215-50b8-4b76-9477-2c8f26f00021-c000.json
           ├── part-00095-ded0b8be-ab41-4a41-9e60-9542f94578fb-c000.json
           ├── part-00108-83c35027-a744-41ba-9f88-e46da4787b58-c000.json
           ├── part-00132-c347d6b5-1232-4a36-b5ad-148d2251e3ab-c000.json
           ├── part-00146-ed982e74-64a2-439a-b69c-ac0371366934-c000.json
           ├── part-00153-14d8dcd9-b8ca-433d-ab49-be93600d9fc6-c000.json
           ├── part-00171-619e513b-bd5a-47c0-b396-4734228160ee-c000.json
           ├── part-00175-44606e86-3c64-4fcb-80f1-6eb766f281db-c000.json
           ├── part-00180-328b9507-76e6-4f05-b506-3c16e01d4285-c000.json
           └── part-00185-ffa2a99d-0618-41c2-abab-cb73865996b7-c000.json
 — Timebased-KPI
   └── time-wise-kpi
       └── time-wise-kpi
           ├── _spark_metadata
           │   ├── 0
           │   ├── 1
           │   ├── 10
           │   ├── 11
           │   ├── 12
           │   ├── 13
           │   ├── 14
           │   ├── 2
           │   ├── 3
           │   ├── 4
           │   ├── 5
           │   ├── 6
           │   ├── 7
           │   ├── 8
           │   └── 9.compact
           ├── part-00000-021e69bd-3cea-4d04-b150-3e0555fceb7b-c000.json
           ├── part-00000-041ea273-20ce-458d-b1d1-8cdaa8de5684-c000.json
           ├── part-00000-044b7eed-e577-40a7-8ce0-049ad71f3231-c000.json
           ├── part-00000-18a9fa97-1fcb-45dc-9758-285d16f62bc0-c000.json
           ├── part-00000-1ff10c56-1736-438e-bc2a-dc033dd55c4d-c000.json
           ├── part-00000-461e169e-a761-410c-aa8b-7c39ba3219f9-c000.json
           ├── part-00000-4c57f2bd-1d14-4682-bb54-5282456c30e7-c000.json
           ├── part-00000-707c36c6-064d-4b14-ba70-2c5097151913-c000.json
           ├── part-00000-76e4dbcc-5dc6-4592-8795-2215d69b017c-c000.json
           ├── part-00000-7e61af6e-a000-4e79-a4a6-352965e17f47-c000.json
           ├── part-00000-84c81f63-4f94-4980-84fd-87eae935862e-c000.json
           ├── part-00000-9af1306e-cfd8-4fae-a807-3a4cbb818185-c000.json
           ├── part-00000-a5e7fe15-1c93-402f-961f-87adf97041da-c000.json
           ├── part-00000-bc7b91d7-20af-499e-aabf-2eb8cb39d570-c000.json
           ├── part-00000-d3334c4d-4c11-4bef-a0dd-b000foc8d6a3-c000.json
           ├── part-00004-dafd43a5-acea-4340-9d0d-8698c3a40771-c000.json
           ├── part-00004-0341bcd9-6bf2-4d98-840f-f277aeda81cf-c000.json
           ├── part-00012-0f16fdf4-2b77-435c-a9e6-e051316d707c-c000.json
           ├── part-00054-48e3dd59-f404-463c-b48d-bb9b3cf0cd1c-c000.json
           ├── part-00067-7d91238-3e40-4d11-843f-3fd8f256e946-c000.json
           ├── part-00087-80fe8e25-b743-408b-811e-0262051dec0f-c000.json
           ├── part-00088-b73849a3-c467-42f3-bb72-75f0f0b92d43-c000.json
           ├── part-00099-0b1a7e0d-9777-4b29-b7d5-e76266e967b1-c000.json
           ├── part-00106-5b2d8b71-4acd-44da-8c6b-fe4892063e23-c000.json
           ├── part-00133-6a735a5d-2429-4e1a-858c-e433640e3d76-c000.json
           ├── part-00154-203a6b62-7fee-4c10-8058-8d5b1be82a84-c000.json
           └── part-00198-21224b93-1ac7-4748-806b-d39e02d0b82a-c000.json

8 directories, 94 files
hkrishna@Krishnas-MacBook-Pro Output %
```