

Code Logic – Retail Data Analysis

Command used to run the spark job :

```
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.2.1 spark-streaming.py  
18.211.252.152 9092 real-time-project
```

Utility Functions :

get_total_items : to sum up the quantity of items ordered for each invoice.

get_total_cost : quantity * unit_price for each invoice.

is_order : returns 1 if the type is ORDER else return 0

is_return : returns 1 if the type is RETURN else return 0

Streams :

kafkaDF1 : Input Stream

retail_console_result : Stream with extended columns added to raw data.

agg_time : Stream with time-based KPIs with a tumbling window of one minute on orders across the globe.

agg_time_country : Stream with time and country-based KPIs with a tumbling window of one minute on orders across the globe.

Output HDFS directory for agg_time : time_based_data

Output HDFS directory for agg_time_country : time_country_based_data

Console Output :

Batch: 2

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132557951104	United Kingdom	2024-01-08 08:56:31	77.0	52	1	0
154132557951105	United Kingdom	2024-01-08 08:56:40	3.5799999237060547	2	1	0
154132557951106	United Kingdom	2024-01-08 08:56:42	53.87999987602234	25	1	0
154132557951107	France	2024-01-08 08:56:49	22.65000033378601	11	1	0
154132557951108	United Kingdom	2024-01-08 08:56:51	16.380000233650208	6	1	0
154132557951109	United Kingdom	2024-01-08 08:57:05	84.9399995803833	34	1	0
154132557951110	United Kingdom	2024-01-08 08:57:10	-16.450000047683716	13	0	1
154132557951111	United Kingdom	2024-01-08 08:57:12	23.249999165534973	23	1	0

Batch: 3

invoice_no	country	timestamp	total_cost	total_items	is_order	is_return
154132557951112	Portugal	2024-01-08 08:57:16	26.27999973297119	16	1	0
154132557951113	United Kingdom	2024-01-08 08:57:32	126.06999880075455	56	1	0
154132557951114	United Kingdom	2024-01-08 08:57:37	42.170000076293945	11	1	0
154132557951115	United Kingdom	2024-01-08 08:57:39	99.32000041007996	23	1	0
154132557951116	United Kingdom	2024-01-08 08:57:45	356.59000408649445	111	1	0
154132557951117	United Kingdom	2024-01-08 08:57:49	38.47999882698059	74	1	0
154132557951118	United Kingdom	2024-01-08 08:57:55	33.249999046325684	13	1	0
154132557951119	United Kingdom	2024-01-08 08:57:56	-135.8199986219406	60	0	1
154132557951120	United Kingdom	2024-01-08 08:58:02	-39.74999928474426	23	0	1
154132557951121	Portugal	2024-01-08 08:58:04	8.319999694824219	8	1	0
154132557951122	United Kingdom	2024-01-08 08:58:04	86.11999881267548	55	1	0
154132557951123	United Kingdom	2024-01-08 08:58:07	-31.45000010728836	11	0	1
154132557951124	United Kingdom	2024-01-08 08:58:13	18.15999972820282	30	1	0