Birla Institute of Technology & Science, Pilani
Work Integrated Learning Programmes Division
Third Semester 2020-2021

Assignment 1

Course No.   : S2-20_DSECLZG522
Course Title   : Big Data Systems

Using Hadoop for Historical Sales Data Analysis

Contribution & Team

Group 079

S.No	Name	BITSIDNO
1	Yashodeep Vaidya	2019HC04042
2	Sharad K Laad	2019HC04034
3	Vijay Jaiswal	2019HC04049

Table of Contents

Purpose of study – Assignment Details

Purpose of this report and study is to answer following questions

Business context:

A large multi-national retail chain has sales orders data across regions and different sales channels for a large variety of item types. The business team wants to use this data to analyze various aspects of sales - e.g. top selling items in a region, regions with maximum profit in a certain item type, if there is a significant difference in revenue in two item types across regions etc.

Problem statement:

As the data analytics team, use the sales transaction data set with about 100K records to answer these questions below —

Q 1.   Average unit_price by country for a given item type in a certain year
Q 2.   Total units_sold by year for a given country and a given item type
Q 3.   Find the max and min units_sold in any order for each year by country for a given item type. Use a custom partitioner class instead of default hash based.
Q 4.   What are the top 10 order id for a given year by the total_profit

You have to show the above analysis working on a Hadoop system using map reduce code, preferably in Java or Python. You can do data preparation steps as required before running a MapReduce job to answer these questions above.

Submission materials:
• Working documented code

Data set:
https://drive.google.com/drive/folders/1fJyk6GE13xw0P-OKWwW8fxa-ny1G0TPp?usp=sharing

geosales

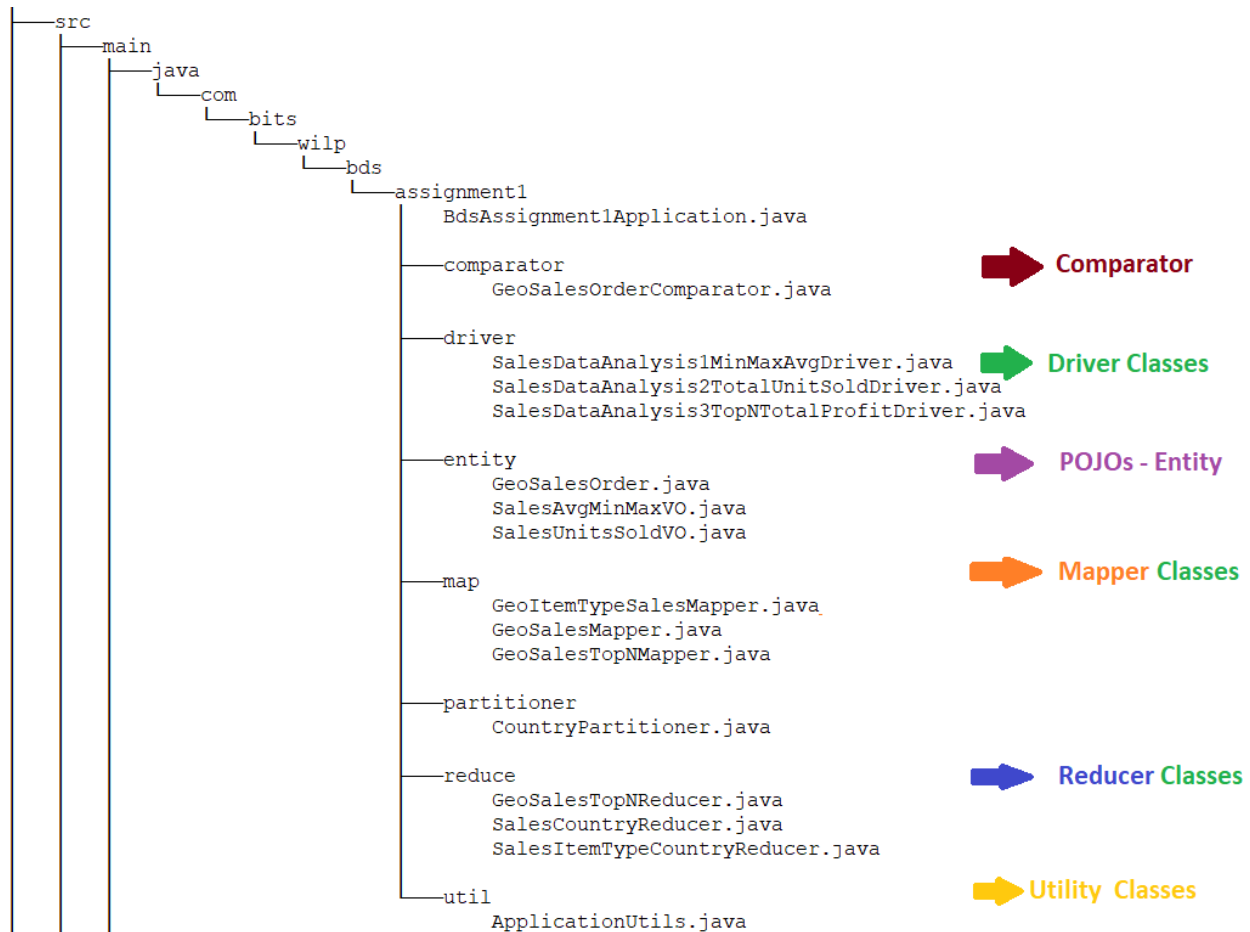| index	region	country	item_type	sales_channel	order_priority	order_date	order_id	ship_date	units_sold	unit_price	unit_cost	total_revenue	total_cost	total_profit
13465	Australia and Oceania	Tonga	Snacks	Online	C	2017-05-04 00:00:00	659991061	2017-05-27 00:00:00	4876	152.58	97.44	743980.08	475117.44	268862.64
84289	Europe	Czech Republic	Baby Food	Online	L	2015-06-01 00:00:00	140080249	2015-06-11 00:00:00	1713	255.28	159.42	437294.64	273086.46	164208.18
6703	Europe	Ukraine	Baby Food	Offline	L	2018-08-17 00:00:00	614736831	2018-08-28 00:00:00	3144	255.28	159.42	802600.32	501216.48	301383.84
81234	Europe	Romania	Personal Care	Online	C	2018-04-10 00:00:00	543800771	2018-05-11 00:00:00	1587	81.73	56.67	129705.51	89935.29	39770.22
75679	Sub-Saharan Africa	Tanzania	Personal Care	Offline	L	2020-03-24 00:00:00	402419173	2020-04-30 00:00:00	3631	81.73	56.67	296761.63	205768.77	90992.86
55273	Europe	France	Beverages	Offline	L	2017-01-18 00:00:00	376248538	2017-02-22 00:00:00	9158	47.45	31.79	434547.1	291132.82	143414.28
67845	Sub-Saharan Africa	Senegal	Clothes	Online	M	2017-02-26 00:00:00	962765681	2017-03-31 00:00:00	2615	109.28	35.84	285767.2	93721.6	192045.6
25431	Sub-Saharan Africa	Sudan	Household	Online	C	2020-07-23 00:00:00	125862395	2020-08-01 00:00:00	4335	668.27	502.54	2896950.45	2178510.9	718439.55

Solution:

Question 1 & 3

Q 1.   Average unit_price by country for a given item type in a certain year

Q 3.   Find the max and min units_sold in any order for each year by country for a given item type. Use a custom partitioner class instead of default hash based.

Answer 1&3:

Since the input and parameter were same for Q1 and Q3 we have used same implementation for both.

Java Class Structure

```
├──src
│   └──main
│       └──java
│           └──com
│               └──bits
│                   └──wilp
│                       └──bds
│                           └──assignment1
│                                   BdsAssignment1Application.java
│                               ├──comparator
│                                   GeoSalesOrderComparator.java          ➡ Comparator
│                               ├──driver
│                                   SalesDataAnalysis1MinMaxAvgDriver.java      ➡ Driver Classes
│                                   SalesDataAnalysis2TotalUnitSoldDriver.java
│                                   SalesDataAnalysis3TopNTotalProfitDriver.java
│                               ├──entity
│                                   GeoSalesOrder.java                    ➡ POJOs - Entity
│                                   SalesAvgMinMaxVO.java
│                                   SalesUnitsSoldVO.java
│                               ├──map
│                                   GeoItemTypeSalesMapper.java           ➡ Mapper Classes
│                                   GeoSalesMapper.java
│                                   GeoSalesTopNMapper.java
│                               ├──partitioner
│                                   CountryPartitioner.java
│                               ├──reduce
│                                   GeoSalesTopNReducer.java              ➡ Reducer Classes
│                                   SalesCountryReducer.java
│                                   SalesItemTypeCountryReducer.java
│                               └──util                                   ➡ Utility Classes
│                                   ApplicationUtils.java
```

Implementation Code
Attached with this document.

Command to Run the Driver
java com.bits.wilp.bds.assignment1.driver.SalesDataAnalysis1MinMaxAvgDriver geosales.csv geosales_output_1

Here :

1. **geosales.csv:** input file having historical sales data
2. **geosales_output_1:** Output folder where the analysis results will be saved

Taking User Input: Item Type

```
Run:     SalesDataAnalysis1MinMaxAvgDriver ×

         D:\installed\java-1.8.0-openjdk-1.8.0.252\bin\java.exe ...
         1. Finding Average unit_price by country for a given item type in a certain year:
         Please enter following:
         1. (a) Item Type (Baby Food, Beverages, Cereal, Clothes, Cosmetics
                         Fruits, Household, Meat, Office Supplies, Personal Care, Snacks, Vegetables): Cereal
```

Running Logs:

```
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Tuvalu~2013 salesAvgMinMaxVO:[averageUnitPrice=205.7,maxUnitSold=6979.0,minUnitSold=697
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Tuvalu~2014 salesAvgMinMaxVO:[averageUnitPrice=205.7,maxUnitSold=9931.0,minUnitSold=588
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Tuvalu~2015 salesAvgMinMaxVO:[averageUnitPrice=205.70000000000002,maxUnitSold=7870.0,mi
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Tuvalu~2016 salesAvgMinMaxVO:[averageUnitPrice=205.7,maxUnitSold=9304.0,minUnitSold=419
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Tuvalu~2017 salesAvgMinMaxVO:[averageUnitPrice=205.7,maxUnitSold=9930.0,minUnitSold=105
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Tuvalu~2018 salesAvgMinMaxVO:[averageUnitPrice=205.7,maxUnitSold=9552.0,minUnitSold=138
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Tuvalu~2019 salesAvgMinMaxVO:[averageUnitPrice=205.70000000000002,maxUnitSold=8395.0,mi
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Tuvalu~2020 salesAvgMinMaxVO:[averageUnitPrice=205.70000000000002,maxUnitSold=9830.0,mi
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Tuvalu~2021 salesAvgMinMaxVO:[averageUnitPrice=205.70000000000002,maxUnitSold=9595.0,mi
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Ukraine~2013 salesAvgMinMaxVO:[averageUnitPrice=205.69999999999996,maxUnitSold=9384.0,m
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Ukraine~2014 salesAvgMinMaxVO:[averageUnitPrice=205.7,maxUnitSold=9254.0,minUnitSold=11
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Ukraine~2015 salesAvgMinMaxVO:[averageUnitPrice=205.7,maxUnitSold=9471.0,minUnitSold=11
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Ukraine~2016 salesAvgMinMaxVO:[averageUnitPrice=205.70000000000002,maxUnitSold=9253.0,m
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Ukraine~2017 salesAvgMinMaxVO:[averageUnitPrice=205.7,maxUnitSold=8858.0,minUnitSold=42
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Ukraine~2018 salesAvgMinMaxVO:[averageUnitPrice=205.70000000000002,maxUnitSold=9620.0,m
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Ukraine~2019 salesAvgMinMaxVO:[averageUnitPrice=205.70000000000002,maxUnitSold=8303.0,m
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:Ukraine~2020 salesAvgMinMaxVO:[averageUnitPrice=205.70000000000002,maxUnitSold=8888.0,m
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:United Arab Emirates~2013 salesAvgMinMaxVO:[averageUnitPrice=205.70000000000002,maxUnit
27-08-2021 20:08:07.481 [pool-5-thread-1] INFO  c.b.w.b.a.reduce.SalesCountryReducer - key:United Arab Emirates~2014 salesAvgMinMaxVO:[averageUnitPrice=205.69999999999996,maxUnit
```

Summary of MapRaduce Job:

```
27-08-2021 20:17:25.471 [main] INFO  org.apache.hadoop.mapreduce.Job -  map 100% reduce 100%
27-08-2021 20:17:25.471 [main] INFO  org.apache.hadoop.mapreduce.Job - Job job_local579692035_0001 completed successfully
27-08-2021 20:17:25.477 [main] INFO  org.apache.hadoop.mapreduce.Job - Counters: 30
    File System Counters
        FILE: Number of bytes read=49482538
        FILE: Number of bytes written=8111825
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
    Map-Reduce Framework
        Map input records=100001
        Map output records=8421
        Map output bytes=1540572
        Map output materialized bytes=1565847
        Input split bytes=177
        Combine input records=0
        Combine output records=0
        Reduce input groups=1587
        Reduce shuffle bytes=1565847
        Reduce input records=8421
        Reduce output records=1587
        Spilled Records=16842
        Shuffled Maps =2
        Failed Shuffles=0
        Merged Map outputs=2
        GC time elapsed (ms)=7
        Total committed heap usage (bytes)=861929472
    Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Input Format Counters
        Bytes Read=14973826
    File Output Format Counters
        Bytes Written=134654
27-08-2021 20:17:25.477 [main] INFO  c.b.w.b.a.d.SalesDataAnalysis1MinMaxAvgDriver - Job was successful

Process finished with exit code 0
```

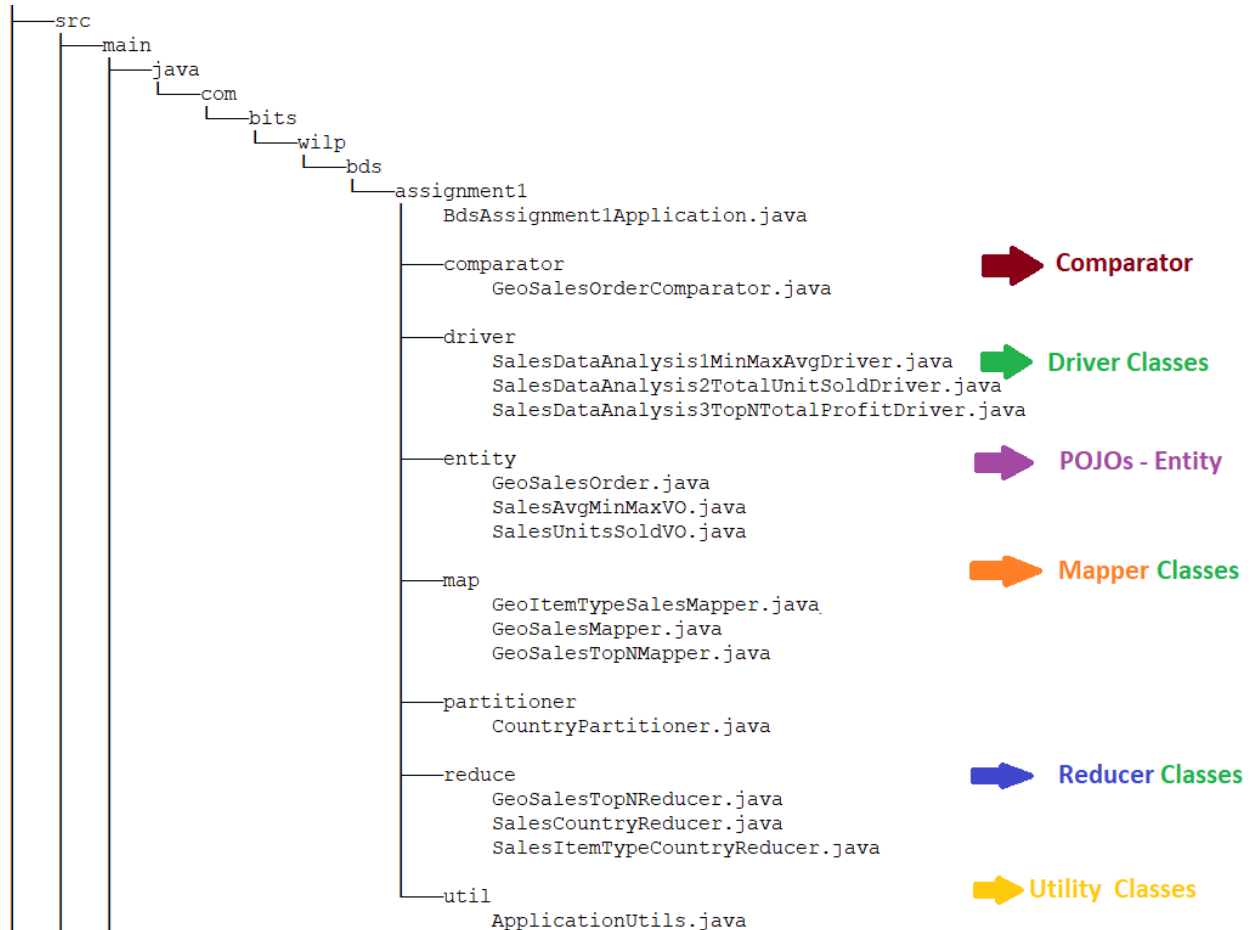Output Folder Created with Analysis:
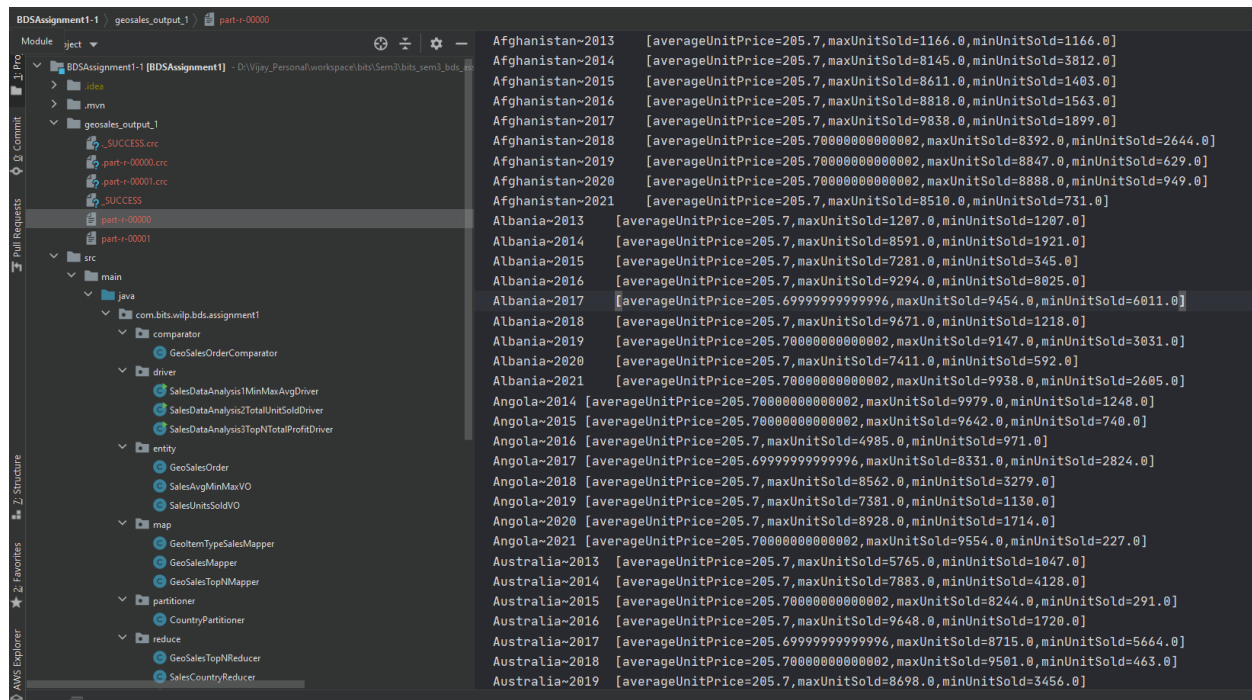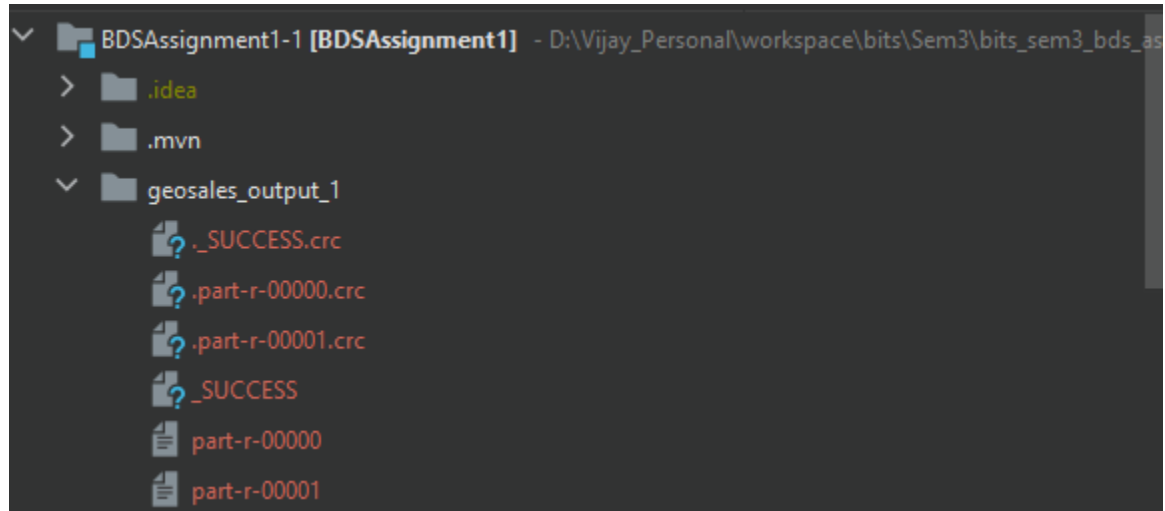
Question 2

Q 2.   Total units_sold by year for a given country and a given item type

Answer 2:

Java Class Structure

```
├──src
│   └──main
│       └──java
│           └──com
│               └──bits
│                   └──wilp
│                       └──bds
│                           └──assignment1
│                                   BdsAssignment1Application.java
│
│                               ├──comparator                    ➤ Comparator
│                                   GeoSalesOrderComparator.java
│
│                               ├──driver                        ➤ Driver Classes
│                                   SalesDataAnalysis1MinMaxAvgDriver.java
│                                   SalesDataAnalysis2TotalUnitSoldDriver.java
│                                   SalesDataAnalysis3TopNTotalProfitDriver.java
│
│                               ├──entity                        ➤ POJOs - Entity
│                                   GeoSalesOrder.java
│                                   SalesAvgMinMaxVO.java
│                                   SalesUnitsSoldVO.java
│
│                               ├──map                           ➤ Mapper Classes
│                                   GeoItemTypeSalesMapper.java
│                                   GeoSalesMapper.java
│                                   GeoSalesTopNMapper.java
│
│                               ├──partitioner
│                                   CountryPartitioner.java
│
│                               ├──reduce                        ➤ Reducer Classes
│                                   GeoSalesTopNReducer.java
│                                   SalesCountryReducer.java
│                                   SalesItemTypeCountryReducer.java
│
│                               └──util                          ➤ Utility Classes
│                                   ApplicationUtils.java
```

Implementation Code

Attached with this document.

Command to Run the Driver

java com.bits.wilp.bds.assignment1.driver.SalesDataAnalysis2TotalUnitSoldDriver geosales.csv geosales_output_2

Here :

1. **geosales.csv:** input file having historical sales data
2. **geosales_output_2:** Output folder where the analysis results will be saved

Taking User Input: Item Type


```
D:\installed\java-1.8.0-openjdk-1.8.0.252\bin\java.exe -javaagent:D:\installed\JetBrains\IntelliJIDEACommunityEdition2020.2.3\lib\idea_rt.jar=60770:D:\installed\JetBrains\Intelli
Q2. total units_sold by year for a given country and a given item type:
Please enter following:
1. (a) Enter Country
   (Afghanistan, Albania, Algeria, Andorra, Angola, Antigua and Barbuda , Armenia, Australia, Austria, Azerbaijan, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize,
Benin, Bhutan, Bosnia and Herzegovina, Botswana, Brunei, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Cape Verde, Central African Republic, Chad, China, Comoros,
Cote d'Ivoire, Croatia, Cuba, Cyprus, Czech Republic, Democratic Republic of the Congo, Denmark, Djibouti, Dominica, Dominican Republic, East Timor, Egypt, El Salvador, Equatoria
Eritrea, Estonia, Ethiopia, Federated States of Micronesia, Fiji, Finland, France, Gabon, Georgia, Germany, Ghana, Greece, Greenland, Grenada, Guatemala, Guinea, Guinea-Bissau, H
Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kiribati, Kosovo, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Le
Libya, Liechtenstein, Lithuania, Luxembourg, Macedonia, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Marshall Islands, Mauritania, Mauritius , Mexico, Moldova , Monaco, M
Morocco, Mozambique, Myanmar, Namibia, Nauru, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, North Korea, Norway, Oman, Pakistan, Palau, Panama, Papua New Guinea, Ph
Portugal, Qatar, Republic of the Congo, Romania, Russia, Rwanda, Saint Kitts and Nevis , Saint Lucia, Saint Vincent and the Grenadines, Samoa , San Marino, Sao Tome and Principe,
Serbia, Seychelles , Sierra Leone, Singapore, Slovakia, Slovenia, Solomon Islands, Somalia, South Africa, South Korea, South Sudan, Spain, Sri Lanka, Sudan, Swaziland, Sweden, Sw
Syria, Taiwan, Tajikistan, Tanzania, Thailand, The Bahamas, The Gambia, Togo, Tonga, Trinidad and Tobago, Tunisia , Turkey, Turkmenistan, Tuvalu, Uganda, Ukraine, United Arab Emi
United Kingdom, United States of America, Uzbekistan, Vanuatu, Vatican City, Vietnam, Yemen, Zambia, Zimbabwe): India
2. (b) Item Type (Baby Food, Beverages, Cereal, Clothes, Cosmetics
                  Fruits, Household, Meat, Office Supplies, Personal Care, Snacks, Vegetables): Cereal
```

Running Logs:


```
27-08-2021 20:42:28.266 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"24862","itemType":"Cereal","ord
27-08-2021 20:42:28.291 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"74870","itemType":"Cereal","ord
27-08-2021 20:42:28.333 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"50804","itemType":"Cereal","ord
27-08-2021 20:42:28.342 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"7357","itemType":"Cereal","orde
27-08-2021 20:42:28.353 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"92249","itemType":"Cereal","ord
27-08-2021 20:42:28.390 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"85075","itemType":"Cereal","ord
27-08-2021 20:42:28.400 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"28901","itemType":"Cereal","ord
27-08-2021 20:42:28.422 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"35836","itemType":"Cereal","ord
27-08-2021 20:42:28.423 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"95885","itemType":"Cereal","ord
27-08-2021 20:42:28.425 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"38809","itemType":"Cereal","ord
27-08-2021 20:42:28.427 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"73857","itemType":"Cereal","ord
27-08-2021 20:42:28.436 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"75602","itemType":"Cereal","ord
27-08-2021 20:42:28.443 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"86107","itemType":"Cereal","ord
27-08-2021 20:42:28.444 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"13457","itemType":"Cereal","ord
27-08-2021 20:42:28.468 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"88110","itemType":"Cereal","ord
27-08-2021 20:42:28.471 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"36786","itemType":"Cereal","ord
27-08-2021 20:42:28.474 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"35561","itemType":"Cereal","ord
27-08-2021 20:42:28.485 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"64181","itemType":"Cereal","ord
27-08-2021 20:42:28.489 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"8924","itemType":"Cereal","orde
27-08-2021 20:42:28.513 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"32570","itemType":"Cereal","ord
27-08-2021 20:42:28.545 [LocalJobRunner Map Task Executor #0] INFO  c.b.w.b.a.map.GeoItemTypeSalesMapper - Processing: {"country":"India","index":"61551","itemType":"Cereal","ord
```

Summary of MapRaduce Job:


```
27-08-2021 20:42:28.876 [pool-5-thread-1] INFO  c.b.w.b.a.r.SalesItemTypeCountryReducer - key:India~Cereal~2020 salesunitssold&vu:[TotalUnitsSold=46275]
27-08-2021 20:42:28.882 [pool-5-thread-1] INFO  o.a.h.m.l.output.FileOutputCommitter - Saved output of task 'attempt_local1698769526_0001_r_000000_0' to file:/D:/Vijay_Personal/w
27-08-2021 20:42:29.000 [main] INFO  org.apache.hadoop.mapreduce.Job - Job job_local1698769526_0001 running in uber mode : false
27-08-2021 20:42:29.001 [main] INFO  org.apache.hadoop.mapreduce.Job -  map 100% reduce 100%
27-08-2021 20:42:29.001 [main] INFO  org.apache.hadoop.mapreduce.Job - Job job_local1698769526_0001 completed successfully
27-08-2021 20:42:29.009 [main] INFO  org.apache.hadoop.mapreduce.Job - Counters: 30
        File System Counters
                FILE: Number of bytes read=29962380
                FILE: Number of bytes written=653125
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
        Map-Reduce Framework
                Map input records=100001
                Map output records=41
                Map output bytes=6985
                Map output materialized bytes=7114
                Input split bytes=177
                Combine input records=0
                Combine output records=0
                Reduce input groups=7
                Reduce shuffle bytes=7114
                Reduce input records=41
                Reduce output records=7
                Spilled Records=82
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=5
                Total committed heap usage (bytes)=611319808
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=14973826
        File Output Format Counters
                Bytes Written=299
27-08-2021 20:42:29.010 [main] INFO  c.b.w.b.a.d.SalesDataAnalysis2TotalUnitSoldDriver - Job was successful

Process finished with exit code 0
```
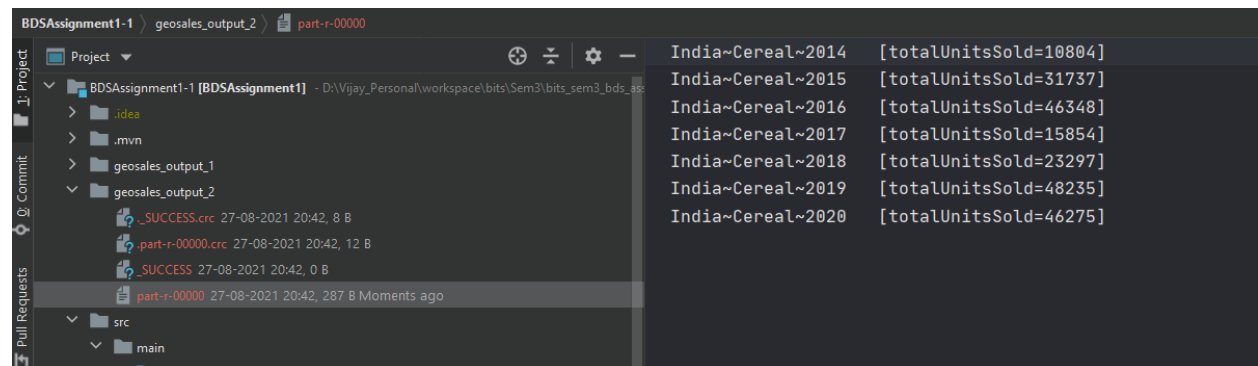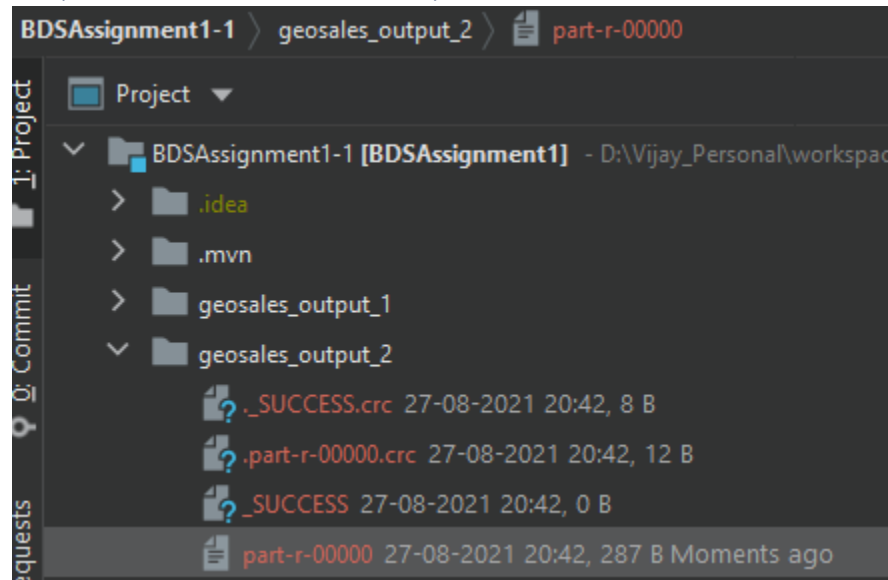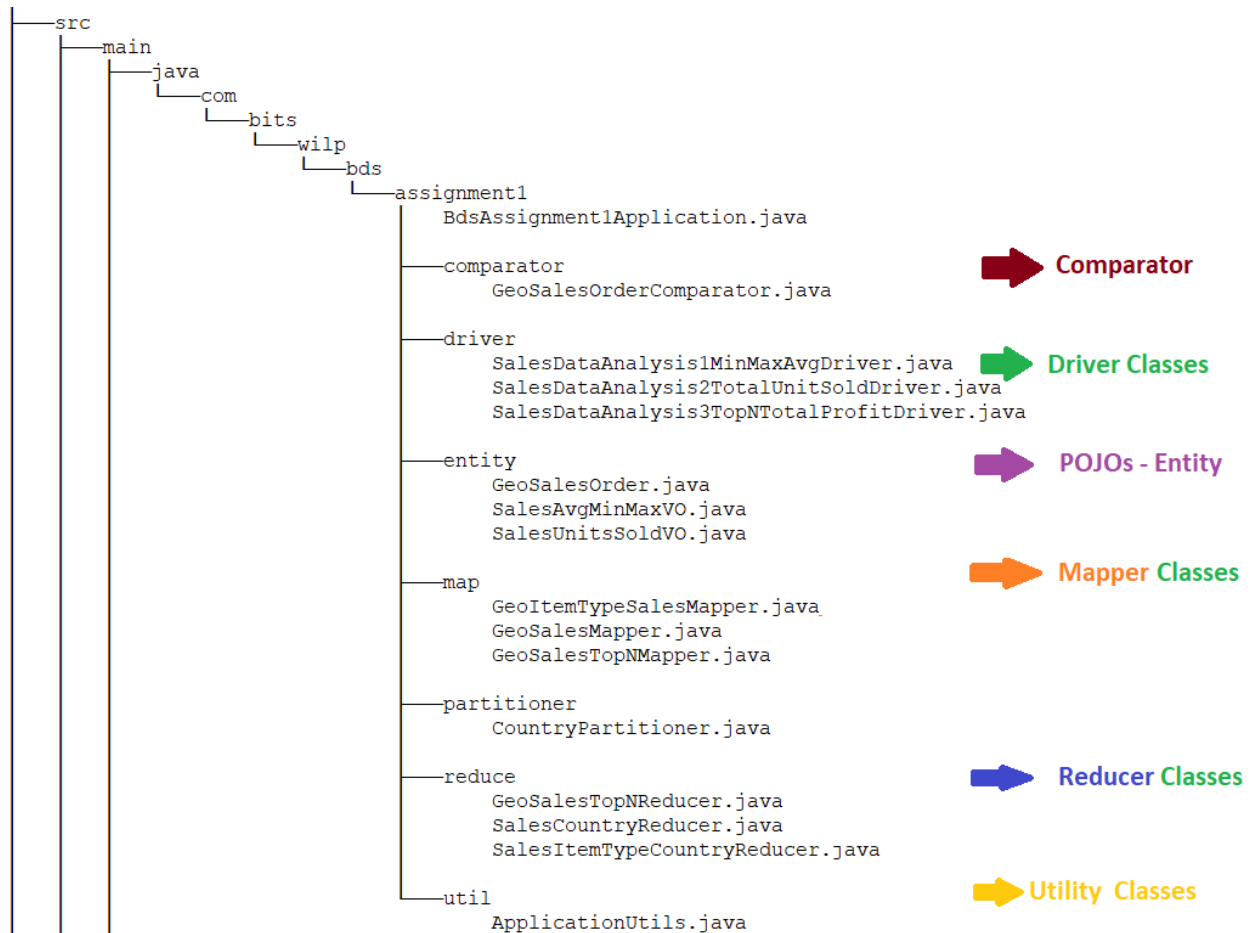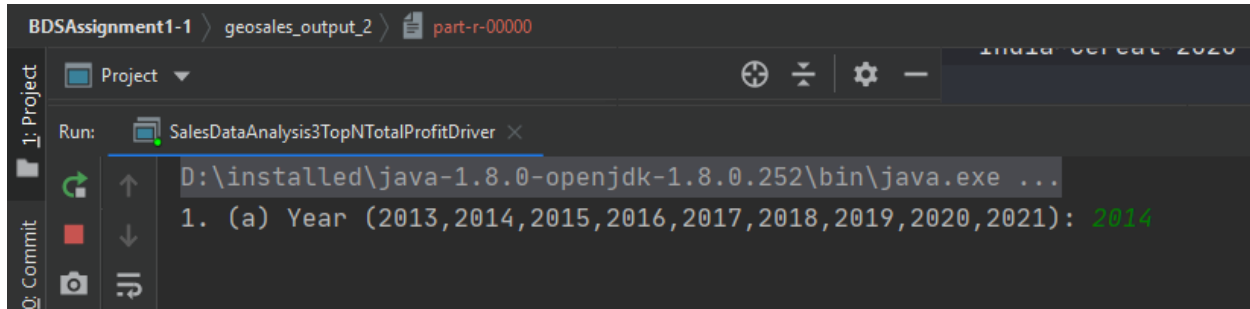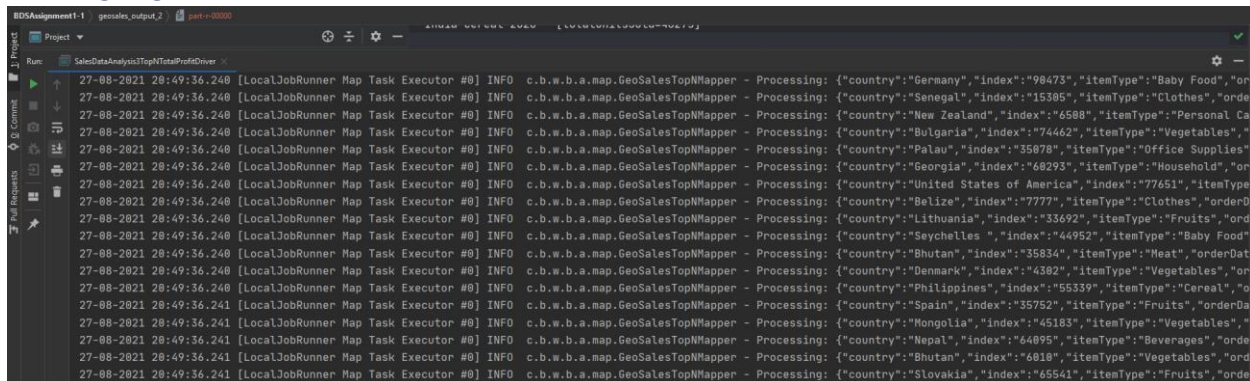
Output Folder Created with Analysis:

India~Cereal~2014   [totalUnitsSold=10804]
India~Cereal~2015   [totalUnitsSold=31737]
India~Cereal~2016   [totalUnitsSold=46348]
India~Cereal~2017   [totalUnitsSold=15854]
India~Cereal~2018   [totalUnitsSold=23297]
India~Cereal~2019   [totalUnitsSold=48235]
India~Cereal~2020   [totalUnitsSold=46275]

Question 4

Q 4.   What are the top 10 order id for a given year by the total_profit

Answer 4:

Java Class Structure

```
├──src
│  ├──main
│  │  ├──java
│  │  │  └──com
│  │  │     └──bits
│  │  │        └──wilp
│  │  │           └──bds
│  │  │              └──assignment1
│  │  │                 │  BdsAssignment1Application.java
│  │  │                 │
│  │  │                 ├──comparator
│  │  │                 │     GeoSalesOrderComparator.java
│  │  │                 │
│  │  │                 ├──driver
│  │  │                 │     SalesDataAnalysis1MinMaxAvgDriver.java
│  │  │                 │     SalesDataAnalysis2TotalUnitSoldDriver.java
│  │  │                 │     SalesDataAnalysis3TopNTotalProfitDriver.java
│  │  │                 │
│  │  │                 ├──entity
│  │  │                 │     GeoSalesOrder.java
│  │  │                 │     SalesAvgMinMaxVO.java
│  │  │                 │     SalesUnitsSoldVO.java
│  │  │                 │
│  │  │                 ├──map
│  │  │                 │     GeoItemTypeSalesMapper.java
│  │  │                 │     GeoSalesMapper.java
│  │  │                 │     GeoSalesTopNMapper.java
│  │  │                 │
│  │  │                 ├──partitioner
│  │  │                 │     CountryPartitioner.java
│  │  │                 │
│  │  │                 ├──reduce
│  │  │                 │     GeoSalesTopNReducer.java
│  │  │                 │     SalesCountryReducer.java
│  │  │                 │     SalesItemTypeCountryReducer.java
│  │  │                 │
│  │  │                 └──util
│  │  │                       ApplicationUtils.java
```

➡ **Comparator**

➡ **Driver Classes**

➡ **POJOs - Entity**

➡ **Mapper Classes**

➡ **Reducer Classes**

➡ **Utility Classes**

Implementation Code

Attached with this document.

Command to Run the Driver

java com.bits.wilp.bds.assignment1.driver.SalesDataAnalysis3TopNTotalProfitDriver geosales.csv geosales_output_3

Here :

3. **geosales.csv:** input file having historical sales data
4. **geosales_output_3:** Output folder where the analysis results will be saved

Taking User Input: Year


```
BDSAssignment1-1 > geosales_output_2 > part-r-00000
Project ▼                                                    ⊕ ÷ | ✿ —      India cereal 2020
Run:      SalesDataAnalysis3TopNTotalProfitDriver ✕
    D:\installed\java-1.8.0-openjdk-1.8.0.252\bin\java.exe ...
    1. (a) Year (2013,2014,2015,2016,2017,2018,2019,2020,2021): 2014
```

Running Logs:

Summary of MapRaduce Job:


```
27-08-2021 20:49:36.487 [pool-5-thread-1] INFO  o.a.h.m.l.output.FileOutputCommitter - Saved output of task 'attempt_local1541764154_0001_r_000000_0' to f
27-08-2021 20:49:37.147 [main] INFO  org.apache.hadoop.mapreduce.Job -  map 100% reduce 100%
27-08-2021 20:49:37.147 [main] INFO  org.apache.hadoop.mapreduce.Job - Job job_local1541764154_0001 completed successfully
27-08-2021 20:49:37.152 [main] INFO  org.apache.hadoop.mapreduce.Job - Counters: 30
    File System Counters
        FILE: Number of bytes read=29951628
        FILE: Number of bytes written=644438
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
    Map-Reduce Framework
        Map input records=100001
        Map output records=10
        Map output bytes=1702
        Map output materialized bytes=1738
        Input split bytes=177
        Combine input records=0
        Combine output records=0
        Reduce input groups=10
        Reduce shuffle bytes=1738
        Reduce input records=10
        Reduce output records=10
        Spilled Records=20
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=10
        Total committed heap usage (bytes)=591396864
    Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Input Format Counters
        Bytes Read=14973826
    File Output Format Counters
        Bytes Written=4556
27-08-2021 20:49:37.152 [main] INFO  c.b.w.b.a.d.SalesDataAnalysis3TopNTotalProfitDriver - Job was successful

Process finished with exit code 0
```

Output Folder Created with Analysis:

1738004.52 {"country":"Kazakhstan","index":"85007","itemType":"Cosmetics","orderDate":"Thu May 29 00:00:00 IST 2014","ord
1736091.95 {"country":"San Marino","index":"35250","itemType":"Cosmetics","orderDate":"Mon Oct 06 00:00:00 IST 2014","orderI
1735570.34 {"country":"Syria","index":"30676","itemType":"Cosmetics","orderDate":"Thu Mar 27 00:00:00 IST 2014","orderId":"1
1733657.77 {"country":"South Sudan","index":"60640","itemType":"Cosmetics","orderDate":"Wed Aug 13 00:00:00 IST 2014","order
1731397.46 {"country":"Mexico","index":"15646","itemType":"Cosmetics","orderDate":"Fri Nov 07 00:00:00 IST 2014","orderId":"
1726702.97 {"country":"Republic of the Congo","index":"57650","itemType":"Cosmetics","orderDate":"Mon May 12 00:00:00 IST 20
1726181.36 {"country":"Croatia","index":"90565","itemType":"Cosmetics","orderDate":"Sat Jan 18 00:00:00 IST 2014","orderId":
1724268.79 {"country":"Cyprus","index":"87177","itemType":"Cosmetics","orderDate":"Fri Mar 28 00:00:00 IST 2014","orderId":"
1723747.18 {"country":"Equatorial Guinea","index":"79328","itemType":"Cosmetics","orderDate":"Wed Jan 08 00:00:00 IST 2014",
1722530.09 {"country":"Lesotho","index":"53478","itemType":"Cosmetics","orderDate":"Fri Oct 31 00:00:00 IST 2014","orderId":