**Problem Statement**

**Analyze the Retail Purchase data set and implement the KPI’s as defined below:**

1. Calculate sales breakdown by product category across all of the stores.

2. Calculate sales breakdown by store across all of the stores (handle bad records)

3. Find the total sales values across all the stores and the total number of sales. Assume there is only one reducer.

**Sample Input Data:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Time** | **City** | **Product Category** | **Total** | **Payment\_Mode** |
| 2012-01-01 | 09:00 | San Jose | Men's Clothing | 214.05 | Amex |
| 2012-01-01 | 09:00 | Fort Worth | Women's Clothing | 153.57 | Visa |
| 2012-01-01 | 09:00 | San Diego | Music | 66.08 | Cash |
| 2012-01-01 | 09:00 | Pittsburgh | Pet Supplies | 493.51 | Discover |
| 2012-01-01 | 09:00 | Omaha | Children's Clothing | 235.63 | MasterCard |
| 2012-01-01 | 09:00 | Stockton | Men's Clothing | 247.18 | MasterCard |
| 2012-01-01 | 09:00 | Austin | Cameras | 379.6 | Visa |
| 2012-01-01 | 09:00 | New York | Consumer Electronics | 296.8 | Cash |
| 2012-01-01 | 09:00 | Corpus Christi | Toys | 25.38 | Discover |

**How to run the project**

**Step1: Copy the retail data to specific dir**

Hdfs dfs –put Retail\_Sample\_Data\_Set /hadoop\_dir

**Solution 1: Calculate sales breakdown by product category across all of the stores.**

**Command:**

Yarn jar Retail\_Project.jar ProductCatagory.ProductCategorySalesDriver /hadoop\_dir/ Retail\_Sample\_Data\_Set /Output\_dir

**How to check output:**

**Command:** Hdfs dfs –cat /Output\_dir/part\*

**Solution2: Calculate sales breakdown by store across all of the stores (handle bad records)**

**Command:**

Yarn jar Retail\_Project.jar StoreAnalysis.StoreSalesDriver /hadoop\_dir/ Retail\_Sample\_Data\_Set /Output\_dir

**How to check output:**

**Command:** Hdfs dfs –cat /Output\_dir/part\*

**Solution3: Find the total sales values across all the stores and the total number of sales. Assume there is only one reducer**

Yarn jar Retail\_Project.jar TotalSaleAnalysis.SalesDriver /hadoop\_dir/ Retail\_Sample\_Data\_Set /Output\_dir

**How to check output:**

**Command:** Hdfs dfs –cat /Output\_dir/part\*

**For More Detail Contact**

**Er.NITIN VERMA**

**9988774296**