Danalitic
Crafting Intelligent AI Platform

PROGRAMING TEST

Danalitic India Pvt Ltd,
World Trade Center, Level 9, Tower – 2,
Kharadi, Pune (M.H.)

Problem Statement

You work for a manufacturing company which produces three industrial machinery parts code named Rocks, Papers and Scissors. The manufacturing process is divided among different production units spread across the country. Each production unit has a quality control process in place which discards items that are not upto the company standards from each batch produced. The number of items produced or discarded of each type are recorded by the staff at all production units.

Recently, as the company sales went up, production was expanded to include more manufacturing units to keep up with the demand. The customer care department started receiving complaints from customers regarding low quality products. As a short term measure, the support staff was instructed to record all the complaints in a database in the form of the invoice Id and the name of the defected item.

Since the reputation of the company is at stake, the management wants to identify the low quality production units so that their manufacturing practices can be thoroughly examined. You have been given access to the sales data along with the complaint records and production logs of all production units. The critical reports you have to produce are defined below.

The customers are grouped according to the organisation they belong to. A customer group has customer Ids beginning with the same letter followed by a different 3 digit number for each individual buyer in the organisation.

Assume that a complaint about a particular sale means all items of that type sold were defective. Eg, if a customer complains about rocks in invoice id 873uwg then that implies all rocks sold in that transaction were defective.

For simplicity’s sake, we will also assume each sale included items from one batch only.

TASK

Write a program using PySpark, Java SDK for Spark, Python(use suitable data manipulation libraries), R or other suitable options to create the reports. **Make sure you submit the three output reports in csv or tsv or xls/xlsv format.**

INPUT

As part of the problem statement you should receive three files corresponding to the sales data, complaint records and production logs : sales.tsv, complaints.tsv and production_logs.tsv. The format of each dataset is as follows,

Sales.tsv

Invoice Id	**Customer Id**	**Items Summary**	**Batch Id**
848777	n488	{“rock”:5, “paper”:7}	Gf563
872013	w900	{“scissor”:67, “rock”:2}	Gf563
567290	w562	{“paper”:52, “rock”:8, “scissor”:43}	Uy638

Complaints.tsv

Invoice Id	**Defective Item**
848777	paper
872013	rock

Production_logs.tsv

Production Unit Id	**Batch Id**	**Items produced**	**Items discarded**
P018	Gf563	{“rock”:650, “paper”:900, “scissor”:400}	{“rock”:67, “paper”:6}
P902	Uy638	{“rock”:300, “paper”:1200, “scissor”:350}	{“rock”:3, “paper”:63, “scissor”:5}

OUTPUT

The following reports have to be generated and submitted.

Report 1

The percentage of each defective item produced by each production unit. The total defective items produced by a production unit is the the number of defective items detected and removed at the unit plus the defective items reported by customers.

Production Unit Id	Item	Defect %
P018	rock	10.6
P018	paper	1.4

Report 2

The number of complaints by each customer group contrasted with the number of items they have bought.

Customer Group	Number of complaints	Rocks bought	Paper Bought	Scissors bought
n	1	5	7	0
w	1	10	52	110

Report 3

The percentage of total defective items that were detected by Quality Control on the factory floor. This is the number detected by QC out of all defective items count.

Production Unit	% detected defects by QA
P018	89.02
P902	100

NOTES

- The Invoice Id is a 6 digit number.
- The Customer Id, Batch Id and Production unit Id are alphanumeric strings.
- A customer group has customer Ids beginning with the same letter followed by a different 3 digit number for each individual buyer in the organisation.
- The details of items sold, produced and discarded are stored as JSON strings in string format.
- One Production Unit can correspond to multiple batches.
- There can be multiple complaints about products that come from the same batch but were sold in different transactions.
- **The sample input and output given can be used for validating the logic of your solution**