Data Analyst - Assignment

Overview

Please understand the below mentioned real-life scenario and try to solve the assignment. The sample data is attached in the link provided below for your reference.

Business Scenario

You are a data analyst and your client has a large ecommerce company in India (let's call it X). X gets a thousand orders via their website on a daily basis and they have to deliver them as fast as they can. For delivering the goods ordered by the customers, X has tied up with multiple courier companies in India as delivery partners who charge them some amount per delivery.

The charges are dependent upon two factors:
- Weight of the product
- Distance between the warehouse (pickup location) and customer's delivery address (destination location)

On an average, the delivery charges are Rs. 100 per shipment. So if X ships 1,00,000 orders per month, they have to pay approximately Rs. 1 crore to the courier companies on a monthly basis as charges.

As the amount that X has to pay to the courier companies is very high, they want to verify if the charges levied by their Delivery partners **per Order** are correct.

Input Data

Left Hand Side (LHS) Data (X's internal data spread across three reports)
- Website order report- which will list Order IDs and various products (SKUs) part of each order. Order ID is common identifier between X's order report and courier company invoice
- Warehouse pincode to All India pincode mapping -(this should be used to figure out delivery zone (a/b/c/d/e) and during analysis compare against one reported by courier company in their CSV invoice per Order ID
- SKU master with gross weight of each product. This should be used to calculate total weight of each order and during analysis compare against one reported by courier company in their CSV invoice per Order ID. The courier company calculates weight in slabs that is applicable for that delivery zone, so first you have to figure out the total weight of the shipment and then figure out applicable weight (based on zone's weight slab).

For example:

Total Weight	Applicable Zone	Applicable Weight Slab (based on zone)	Applicable Weight
400gm	a	0.25	0.5
400gm	b	0.5	0.5
400gm	c	0.75	0.75
2.2KG	a	0.25	2.25
2.2KG	b	0.5	2.5
2.2KG	c	0.75	2.25

RHS Data (courier company invoice in CSV file)
- Invoice in CSV file mentioning AWB Number (courier company's own internal ID), Order ID (company X's order ID), weight of shipment, warehouse pickup pincode, customer delivery pincode, zone of delivery, charges per shipment, type of shipment
- Courier charges rate card at weight slab and pincode level. If the invoice mentions "Forward charges" then only forward charges ("fwd") should be applicable as per zone and fixed & additional weights based on weight slabs. If the invoice mentions "Forward and rto charges" then forward charges ("fwd") and RTO charges ("rto") should be applicable as per zone and fixed & additional weights based on weight slabs.
- For the first slab of that zone, "fixed" rate as per the slab is applicable. For each additional slab, "additional" weight in the same proportion is applicable. Total charges will be "fixed" + "total additional" if any. For example: weight 2.2KG, Zone C. So for Zone C the slab length = 0.75KG. So the total applicable weight = 2.25KG. For the first 0.75 KG the charge is "fwd", and for each 0.75 after the first, charges will be additional charges.

Output Data 1
Create a resultant CSV/Excel file with the following columns:
- Order ID
- AWB Number
- Total weight as per X (KG)
- Weight slab as per X (KG)
- Total weight as per Courier Company (KG)
- Weight slab charged by Courier Company (KG)
- Delivery Zone as per X

- Delivery Zone charged by Courier Company
- Expected Charge as per X (Rs.)
- Charges Billed by Courier Company (Rs.)
- Difference Between Expected Charges and Billed Charges (Rs.)

Output Data 2

Create a summary table

	Count	Amount (Rs.)
Total orders where X has been correctly charged	<count>	<total invoice amount>
Total Orders where X has been overcharged	<count>	<total overcharging amount>
Total Orders where X has been undercharged	<count>	<total undercharging amount>

Assignment data available in 6 CSV files as attached.

Submission Please submit the result in an Excel with two workbooks (summary table in one and order level calculation in another) and your code in any programming language such as Python.

Submit to Email id :- tensorflowmks@gmail.com

Please connect me for any clarification--- 8800686128

Mithilesh Singh