Submissions:

Submitted two .sql files which can be opened using notepad editors with queries and answers.

As the raw data was provided in CSV format, the data is loaded into MySQL for processing into three separate tables: Sales1, Sales2 and Returns_tbl

All the queries were written in MySQL, there might be slight changes when the evaluator uses another SQL engine.

Assumptions:

Sales1 and Returns_tbl are used for the Question set 1, and Sales2 is used for the Question set 2

There are multiple entries in the returns table where the returnsales value is 0. They were considered, and an annotation of code is provided wherever necessary in case they must be removed.

Answer 5 from Question 1 set:

Sales1 and returns_tbl were considered.

The idea that I considered is that the most valuable customer is the one who makes large number of sales with minimum number of transactions and minimum number of returns. To achieve that, I created a new variable to calculate the ration of sales/returnsales higher the ratio, better the customer is.

In case there are no returns for a customer, then the ratio is just sales.

Assumptions: There are a few customers where the salesreturns is greater than the sales. I believe that this is because of the data coming from the sales2 table. Also, there are a couple of outliers which I removed.

Rating Model:

Once the ratio is ready and after removing the outliers and errors due to data spill, the net ratio's are scaled from a range of 0 to 5 and 5 being the best customer, shown below.

customerid	sales	return	ord_c	ret_count	sales_ratio	rating
POLLF17399	28606.45	0.00	2	0	28606.450000	5.00
BISSP26380	25659.67	0.00	1	0	25659.670000	4.48
JUSUS89531	21581.55	0.00	3	0	21581.550000	3.77
KEENP37511	20413.45	0.00	4	0	20413.450000	3.57
ABDAH58684	19394.99	0.00	2	0	19394.990000	3.39
MADTE80584	19024.51	0.00	1	0	19024.510000	3.32

Using the Model: Once the user is rated, it's up to the business owner to take an action like providing premium membership for free or giving offers for certain period. Also, the users with moderate rating can be encouraged by giving some heads up with less offers and increase their rating to the upper brackets.

Also, the business owner can now not concentrate on the low-rating users and stop sending them the advertisements/product updates as they are not going to bring any revenue according to the data.

Further Improvements:

If the accurate loss occurred due to returns were given, then the number of returns & orders could also be considered in developing the model by giving more weightage to the orders placed, penalizing the number of returns and a better model could be achieved.