1. Convert the files incentives and incentive\_upc excel files into .csv format
2. Convert the above-mentioned files to data frames(spark.read.options..csv)
3. Read the text files transaction\_items and transaction\_items1 using pandas
4. Split the text in both the data frames into columns using delimiter “|”
5. Combine both these data frames into a single data frame.
6. Convert the pandas data frame to spark data frame using CreateDataframe function.
7. Read transactions text file using pandas
8. Split the text in data frame into columns using delimiter “|”
9. Convert the resultant dataframe into spark data frame using createDataFrame function.
10. Convert spark created data frames to temporary tables using registerTempTable function
11. **Joins**

select \*

from incentive\_upc\_table inc\_upc \

inner join incentives\_table inc on inc\_upc.INCENTIVE\_ID = inc.INCENTIVE\_ID \

inner join transactions\_table trans on inc.CLIENTID = trans.CLIENTID \

inner join transactionitems\_table transitems on trans.CLIENTID = transitems.CLIENTID and trans.BASKET\_ID = transitems.BASKET\_ID and trans.DATE = transitems.DATE

1. Targets :
   1. Clientid – Retailer Identification
   2. Active date
   3. Expire date
   4. Date on which transaction was made
   5. Productid
   6. Quantity