Code Implementation

**Task 1: User Session per page per day.**

**Step 1 :** Create Input Table to hold the User session transaction log.

*CREATE*

*TABLE “PAGE\_VIEWS"*

*(*

*"ID" VARCHAR2(5 BYTE),*

*"USER\_ID" VARCHAR2(7 BYTE),*

*"PAGE\_ID" VARCHAR2(7 BYTE),*

*"VISIT\_DATE" DATE,*

*"VISIT\_TIME" VARCHAR2(10 BYTE)*

*);*

**Step 2 :** Import the input csv file

**Step 3:** Calculate User Sessions per page per day.

*select PAGE\_ID,VISIT\_DATE,COUNT(1) TOTAL\_USER\_SESSIONS from*

*PAGE\_VIEWS*

*group by*

*PAGE\_ID,VISIT\_DATE*

*Order by PAGE\_ID ;*

**Task 2: Market Basket Analysis.**

**Step 1:** Create table to import Order transactions

*CREATE*

*“ORDER\_HISTORY"*

*(*

*"ROW\_ID" VARCHAR2(10 BYTE),*

*"ORDERID" VARCHAR2(20 BYTE),*

*"ORDERDATE" DATE,*

*"CUSTOMERID" VARCHAR2(20 BYTE),*

*"PRODUCTID" VARCHAR2(20 BYTE),*

*"PRODUCTNAME" VARCHAR2(200 BYTE),*

*"SALES" VARCHAR2(15 BYTE),*

*"QUANTITY" VARCHAR2(15 BYTE)*

*);*

**Step 2:** Import the Order csv file.

**Step 3:** Find top 10 Bestselling products.

*CREATE table TOP10\_PRODUCTS as*

*select B.\* from (*

*select a.\*,*

*ROW\_NUMBER() over (order by TOTAL\_ORDERS desc) as ROWN*

*from (*

*select COUNT(DISTINCT(ORDERID)) TOTAL\_ORDERS,PRODUCTNAME*

*from ORDER\_HISTORY group by PRODUCTNAME)A ) B WHERE B.ROWN <11;*

**Step 4:** Find Product Combinations and the number of occurrences.

*create table LHSRHS\_FACT as*

*select C.\* from (*

*select LHS,RHS,SUM(NUMLHS) SUMNUMLHS,SUM(NUMRHS) SUMNUMRHS,COUNT(\*) NUMLHSRHS from (*

*(select PRODUCTNAME LHS,COUNT(distinct ORDERID)NUMLHS ,ORDERID from ORDER\_HISTORY group by PRODUCTNAME,ORDERID )SUMLHS*

*join*

*(select PRODUCTNAME RHS,COUNT(distinct ORDERID) NUMRHS ,ORDERID from ORDER\_HISTORY group by PRODUCTNAME,ORDERID ) SUMRHS*

*on SUMLHS.ORDERID=SUMRHS.ORDERID*

*) group by LHS, RHS)C;*

*--SUMNUMLHS – No of Left Hand Side Products in the transactions*

*--SUMNUMRHS – No of Right Hand Side Products in the transactions*

*--NUMLHSRHS – No of Occurrences of LHS and RHS combination*

**Step 5:** Find the Number of Transactions for Main Product (Left Hand Side Product)

*create table LHS\_TXN as*

*select D.LHs,SUM(SUMNUMLHS) LHSNUMTXN from (*

*select C.\* from (*

*select LHS,RHS,SUM(NUMLHS) SUMNUMLHS,SUM(NUMRHS) SUMNUMRHS,COUNT(\*) NUMLHSRHS from (*

*(select PRODUCTNAME LHS,COUNT(distinct ORDERID)NUMLHS ,ORDERID from ORDER\_HISTORY group by PRODUCTNAME,ORDERID )SUMLHS*

*join*

*(select PRODUCTNAME RHS,COUNT(distinct ORDERID) NUMRHS ,ORDERID from ORDER\_HISTORY group by PRODUCTNAME,ORDERID ) SUMRHS*

*on SUMLHS.ORDERID=SUMRHS.ORDERID*

*) group by LHS, RHS)C ) D*

*group by D.LHs*

*;*

*--LHSNUMTXN – Total no of transactions containing the Best Selling Product.*

**Step 6:**Calculate Support, Confidence and Lift.

*select*

*a.LHS,a.RHS,A.NUMLHSRHS NO\_OF\_OCCURENCES,*

*ROUND(NUMLHSRHS/B.LHSNUMTXN,3) as SUPPORT,*

*ROUND(TOTAL\_ORDERS/B.LHSNUMTXN,3) as CONFIDENCE,*

*ROUND(ROUND(TOTAL\_ORDERS/B.LHSNUMTXN,3)/ROUND(SUMNUMRHS/B.LHSNUMTXN,2),2) as LIFT*

*from LHSRHS\_FACT a,*

*LHS\_TXN B,*

*TOP10\_PRODUCTS C*

*where*

*a.LHS= B.LHS and*

*a.LHS = C.PRODUCTNAME and*

*a.LHS != a.RHS*

*ORDER BY A.NUMLHSRHS desc;*

*--LHS – Best Selling Product*

*--RHS – Product bought frequently with the Best Selling Product*

**Support** – No of occurrences of a pair of combination / Total no of Transactions

**Confidence** – No of Occurrence of Antecedent / Total no of Orders which contains Antecedent

**Lift** – confidence / No of Occurrence of Decedent / Total no of Orders which contains Decedent

*Note: I was not able to fetch any output which had*

*- Support >= 0.2*

*- Confidence >= 0.6*

*- Lift ratio > 1*

*So I have left the final query ordered by the maximum occurrences of a pair with the best selling product.*