HIVE ASSIGNMENT:2

-- CREATION OF CUSTOMERS TABLE

0: jdbc:hive2://localhost:10000> CREATE TABLE CUSTOMERS (cust\_id int, last\_name String, first\_name String, age int, profession String)

. . . . . . . . . . . . . . . .> COMMENT 'CUSTOMERS DETAILS'

. . . . . . . . . . . . . . . .> ROW FORMAT DELIMITED

. . . . . . . . . . . . . . . .> FIELDS TERMINATED BY ','

. . . . . . . . . . . . . . . .> LINES TERMINATED BY '\n'

. . . . . . . . . . . . . . . .> STORED AS TEXTFILE;

-- CREATION OF TRANSACTIONS TABLE

0: jdbc:hive2://localhost:10000> CREATE TABLE TRANSACTIONS (trans\_id int, trans\_date date, cust\_id int, amount double, category String, desc String, city String, state String, pymt\_mode String)

. . . . . . . . . . . . . . . .> COMMENT 'TRANSACTION DETAILS'

. . . . . . . . . . . . . . . .> ROW FORMAT DELIMITED

. . . . . . . . . . . . . . . .> FIELDS TERMINATED BY ','

. . . . . . . . . . . . . . . .> LINES TERMINATED BY '\n'

. . . . . . . . . . . . . . . .> STORED AS TEXTFILE;

-- LOADING INTO CUSTOMERS

0: jdbc:hive2://localhost:10000> LOAD DATA LOCAL INPATH '/home/aditi/futurense\_hadoop-pyspark/labs/dataset/retail/customers.txt' OVERWRITE INTO TABLE customers;

Loading data to table bee\_db1.customers

-- LOADING INTO TRANSACTIONS

0: jdbc:hive2://localhost:10000> LOAD DATA LOCAL INPATH '/home/aditi/futurense\_hadoop-pyspark/labs/dataset/retail/transactions.txt' OVERWRITE INTO TABLE transactions;

Loading data to table bee\_db1.transactions

Q1.

0: jdbc:hive2://localhost:10000> select c.cust\_id,c.first\_name,count(t.trans\_id) from customers c inner join transactions t on c.cust\_id = t.cust\_id group by c.cust\_id;

+------------+---------------+------+

| c.cust\_id | c.first\_name | \_c2 |

+------------+---------------+------+

| 4000001 | Chung | 8 |

| 4000002 | Chen | 6 |

| 4000003 | Melton | 3 |

| 4000004 | Hill | 5 |

| 4000005 | Puckett | 5 |

| 4000006 | Song | 5 |

| 4000007 | Hamilton | 6 |

| 4000008 | Bender | 10 |

| 4000009 | Wagner | 6 |

| 4000010 | McLaughlin | 6 |

+------------+---------------+------+

Q2.

0: jdbc:hive2://localhost:10000> select t.cust\_id,sum(t.amount) from transactions t group by t.cust\_id;

+------------+---------------------+

| t.cust\_id | \_c1 |

+------------+---------------------+

| 4000001 | 651.05 |

| 4000002 | 706.97 |

| 4000003 | 527.5899999999999 |

| 4000004 | 337.06 |

| 4000005 | 325.15 |

| 4000006 | 539.38 |

| 4000007 | 699.5500000000001 |

| 4000008 | 859.42 |

| 4000009 | 457.83 |

| 4000010 | 447.09000000000003 |

+------------+---------------------+

Q3.

0: jdbc:hive2://localhost:10000> select t.cust\_id,sum(t.amount) as total\_amount from transactions t group by t.cust\_id order by total\_amount limit 3 ;

+------------+---------------------+

| t.cust\_id | total\_amount |

+------------+---------------------+

| 4000005 | 325.15 |

| 4000004 | 337.06 |

| 4000010 | 447.09000000000003 |

+------------+---------------------+

Q4.

0: jdbc:hive2://localhost:10000> select cust\_id,pymt\_mode,coUNT(trans\_id) as trans\_count from transactions group by cust\_id, pymt\_mode;

+----------+------------+--------------+

| cust\_id | pymt\_mode | trans\_count |

+----------+------------+--------------+

| 4000001 | cash | 1 |

| 4000001 | credit | 7 |

| 4000002 | cash | 1 |

| 4000002 | credit | 5 |

| 4000003 | credit | 3 |

| 4000004 | cash | 4 |

| 4000004 | credit | 1 |

| 4000005 | cash | 1 |

| 4000005 | credit | 4 |

| 4000006 | credit | 5 |

| 4000007 | credit | 6 |

| 4000008 | credit | 10 |

| 4000009 | credit | 6 |

| 4000010 | credit | 6 |

+----------+------------+--------------+

Q5.

0: jdbc:hive2://localhost:10000> select city,count(trans\_id) as trans\_count from transactions group by city order by trans\_count desc limit 3;

+-------------+--------------+

| city | trans\_count |

+-------------+--------------+

| Honolulu | 3 |

| Columbus | 3 |

| Everett | 2 |

+-------------+--------------+

Q6.

0: jdbc:hive2://localhost:10000> SELECT MONTH(from\_unixtime(unix\_timestamp(trans\_date, 'MM-dd-yyyy'), 'yyyy-MM-dd')) as trans\_month, max(amount) as highest\_transaction from transactions group by MONTH(from\_unixtime(unix\_timestamp(trans\_date, 'MM-dd-yyyy'), 'yyyy-MM-dd'));

+--------------+----------------------+

| trans\_month | highest\_transaction |

+--------------+----------------------+

| 1 | 20.13 |

| 2 | 193.63 |

| 3 | 174.36 |

| 4 | 185.26 |

| 5 | 198.44 |

| 6 | 198.19 |

| 7 | 144.59 |

| 8 | 150.6 |

| 9 | 157.94 |

| 10 | 154.87 |

| 11 | 165.1 |

| 12 | 194.86 |

+--------------+----------------------+

Q7.

0: jdbc:hive2://localhost:10000> SELECT \*

. . . . . . . . . . . . . . . .> FROM transactions

. . . . . . . . . . . . . . . .> TABLESAMPLE(10 PERCENT);