Advanced Database management lab

Name: Rahul Rajan Sawant

Roll no: 24MCA49

Q.A) Create an ADT details_type for name, age and designation. Then create a table INFORMATION with attributes id, centre and details

1. Display the number of persons in uppercase.

2. Display the centre of persons , if their age is greater than 25

3. Display all the details of persons if centre is "Kolkata"

4. Display centre if their destination is "Analyst"

Q.B) create a dataset employee.csv with the attributes id, name, address, salary, deduction, age, job .

 Implement the following preprocessing

1. dealing with Missing values

2. Add a new variable net_price by calculating Salary-deductions.

3. Rename the "job" to "designation"

4. Display the first five records from the dataset

5. Sort the data on salary field

6. If the salary is Greater than 50000 display the Status as 'Greater than 50000 ' in status column and status as " Less than or equal to 50000" for others

Q.C) Create an Excel file with name User(userid, firstname, lastname, emails, address).

1.Transformation: a) merge the name into one field.

b) Sort the data on ascending order of email and descending

order of address.

c) load the data to an XML file.

Output Of A)

```
mysql> create database ADBMS_Lab;
Query OK, 1 row affected (0.00 sec)

mysql> use ADBMS_Lab
Database changed
mysql> CREATE TABLE INFORMATION (
    ->     id INT PRIMARY KEY,
    ->     centre VARCHAR(50),
    ->     name VARCHAR(100),
    ->     age INT,
    ->     designation VARCHAR(50)
    -> );
Query OK, 0 rows affected (0.01 sec)
```

```
mysql> INSERT INTO INFORMATION (id, centre, name, age, designation)
    -> VALUES
    -> (1, 'Kolkata', 'John Doe', 30, 'Analyst'),
    -> (2, 'Mumbai', 'Alice Smith', 28, 'Developer'),
    -> (3, 'Delhi', 'Bob Brown', 35, 'Manager'),
    -> (4, 'Kolkata', 'Jane White', 24, 'Tester');
Query OK, 4 rows affected (0.04 sec)
Records: 4  Duplicates: 0  Warnings: 0
```

Output of Q.A.1)

```
mysql> SELECT
    ->     UPPER(name) AS name
    -> FROM INFORMATION;
+-------------+
| name        |
+-------------+
| JOHN DOE    |
| ALICE SMITH |
| BOB BROWN   |
| JANE WHITE  |
+-------------+
4 rows in set (0.00 sec)

mysql>
```

Output of Q.A.2)

```
mysql> SELECT centre
    -> FROM INFORMATION
    -> WHERE age > 25;
+---------+
| centre  |
+---------+
| Kolkata |
| Mumbai  |
| Delhi   |
+---------+
3 rows in set (0.00 sec)
```

Output of Q.A.3)

```
mysql> SELECT name, age, designation
    -> FROM INFORMATION
    -> WHERE centre = 'Kolkata';
+------------+------+-------------+
| name       | age  | designation |
+------------+------+-------------+
| John Doe   |   30 | Analyst     |
| Jane White |   24 | Tester      |
+------------+------+-------------+
2 rows in set (0.00 sec)
```

Output of Q.A.4)

```
mysql> SELECT centre
    -> FROM INFORMATION
    -> WHERE designation = 'Analyst';
+---------+
| centre  |
+---------+
| Kolkata |
+---------+
1 row in set (0.00 sec)

mysql> |
```

Implementating the Following Preprocessing:

```r
library(dplyr)
# Load the dataset
employee <- read.csv("employee.csv")


# Dealing with missing salary values
employee$salary[is.na(employee$salary)] <- mean(employee$salary, na.rm = TRUE)
employee$deduction[is.na(employee$deduction)] <- mean(employee$deduction, na.rm = TRUE)
print(employee)


# adding new variable net_price
employee$net_price <- employee$salary - employee$deduction
print(employee)


#rename the job to destination
employee <- rename(employee, designation = job)
print(employee)


# displaying first 5 rcords
head(employee, 5)



#Sort data in salary feild
employee <- employee %>%
  arrange(desc(salary))
print(employee)
```

Salary Comparisons

employee$status <- ifelse(employee$salary > 50000,

"Greater than 50000",

"Less than or equal to 50000")

print(employee)

Output Of the B)

```
>
> print(employee)
   id         name        address  salary deduction age designation net_price                       status
1   9   Tom Harris  606 Spruce St 70000.00 6000.000  45     Manager  64000.00           Greater than 50000
2   4  Emily Davis  101 Maple St 60000.00 4500.000  38     Manager  55500.00           Greater than 50000
3   1    John cena    123 Elm St 55000.00 5000.000  29     Manager  50000.00           Greater than 50000
4  10   Anna Green   707 Fir St 53000.00 4000.000  30  Technician  49000.00           Greater than 50000
5   7    David Lee 404 Redwood St 51444.44 2500.000  32  Technician  48944.44           Greater than 50000
6   8   Lisa Black  505 Willow St 51000.00 3888.889  28    Engineer  47111.11           Greater than 50000
7   3 Robert Browney   789 Pine St 48000.00 3000.000  41  Technician  45000.00 Less than or equal to 50000
8   5 Michael Johnson  202 Birch St 45000.00 3500.000  25    Engineer  41500.00 Less than or equal to 50000
9   2   Jade Smith    456 Oak St 42000.00 4000.000  34    Engineer  38000.00 Less than or equal to 50000
10  6  Alice White  303 Cedar St 39000.00 2500.000  27  Technician  36500.00 Less than or equal to 50000
> library(dplyr)
>
```

1.Missing Values

```
 salary deduction a
55000.00  5000.000
42000.00  4000.000
48000.00  3000.000
60000.00  4500.000
45000.00  3500.000
39000.00  2500.000
51444.44  2500.000
51000.00  3888.889
70000.00  6000.000
53000.00  4000.000
```

3.Rename the job as Destination

```
designation r
    Manager
   Engineer
 Technician
    Manager
   Engineer
 Technician
 Technician
   Engineer
    Manager
 Technician
```

2Adding a new variable net_price

```
net_price
 50000.00
 38000.00
 45000.00
 55500.00
 41500.00
 36500.00
 48944.44
 47111.11
 64000.00
 49000.00
```

4.Displaying the first 5 records from the dataset

```
>
> # displaying first 5 rcords
> head(employee, 5)
  id            name        address salary deduction age designation net_price
1  1       John cena   123 Elm St  55000      5000  29     Manager      50000
2  2      Jade Smith   456 Oak St  42000      4000  34     Engineer     38000
3  3   Robert Browney  789 Pine St 48000      3000  41   Technician     45000
4  4      Emily Davis 101 Maple St 60000      4500  38     Manager      55500
5  5  Michael Johnson 202 Birch St 45000      3500  25     Engineer     41500
>
> #Sort data in salary feild
```

5,sorting data on salary field

```
  salary
70000.00
60000.00
55000.00
53000.00
51444.44
51000.00
48000.00
45000.00
42000.00
39000.00
```