

Assignment 3.4

Note : Since no data is provided, using following dataset

```
[acadgild@localhost ~]$ cat employee_salary.csv
1,Sameer,3215,Sales
2,Anand,1154,IT
3,Juhi,6547,IT
4,Aman,2154,Sales
5,Satya,1254,IT
6,Rahul,3654,IT
7,Rohit,1256,Sales
8,Arjun,3247,Sales
9,Shubh,1398,IT
10,Anjali,3541,IT
11,Rohini,2564,Sales
12,Anusha,2849,Sales
13,Kabir,3357,IT
14,Sana,1547,Sales
15,Sandeep,1657,IT
```

```
CREATE EXTERNAL TABLE IF NOT EXISTS employee_salary (id int, name String, salary int, unit String)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
```

```
LOAD DATA LOCAL INPATH '/home/acadgild/employee_salary.csv' INTO TABLE employee_salary;
```

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS employee_salary (id int, name String, salary int, unit String) ROW FORMAT D
ELIMITED FIELDS TERMINATED BY ',';
OK
Time taken: 0.034 seconds
hive> LOAD DATA LOCAL INPATH '/home/acadgild/employee_salary.csv' INTO TABLE employee_salary;
Loading data to table custom.employee_salary
OK
Time taken: 0.532 seconds
```

1. **Get a list of employees who receive a salary less than 100, compared to their immediate employee with higher salary in the same unit**

```
WITH temp as ( SELECT id, name, salary, unit , LEAD(salary,1) OVER (PARTITION BY unit ORDER BY salary) - salary AS salary_diff FROM employee_salary) SELECT * FROM temp WHERE salary_diff < 100;
```

```
hive> WITH temp as ( SELECT id, name, salary, unit , LEAD(salary,1) OVER (PARTITION BY unit ORDER BY salary) - salary AS salary_diff FROM employee_salary) SELECT * FROM temp WHERE salary_diff < 100;
```

```
OK
1      Sameer  3215    Sales   32
Time taken: 27.918 seconds, Fetched: 1 row(s)
```

2. List of all employees who draw higher salary than the average salary of that department

```
select * from employee_salary a WHERE salary > (SELECT avg(salary) FROM employee_salary b  
where b.unit = a.unit GROUP BY unit);
```

```
hive> select * from employee_salary a WHERE salary > (SELECT avg(salary) FROM employee_salary b where b.unit = a.unit  
GROUP BY unit);
```

```
OK  
1      Sameer  3215    Sales  
3      Juhi    6547    IT  
6      Rahul   3654    IT  
8      Arjun   3247    Sales  
10     Anjali  3541    IT  
11     Rohini  2564    Sales  
12     Anusha  2849    Sales  
13     Kabir   3357    IT  
Time taken: 91.352 seconds, Fetched: 8 row(s)
```