CREATE TABLE employees (id BIGINT, workplace STRING, role STRING, service\_type STRING, full\_or\_part\_time STRING, salary BIGINT) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/cloudera/assessment1’;

**The number of staff that are working full and part-time**

SELECT COUNT (id) FROM employees WHERE full\_or\_part\_time = 'Full-Time'; 2384

SELECT COUNT (id) FROM employees WHERE full\_or\_part\_time = 'Part-Time'; 770

**The number of staff are working in each department**

SELECT service\_type, COUNT(\*) FROM employees GROUP BY service\_type;

**The avg salary for staff working in each department**

SELECT service\_type, CAST(AVG(salary) AS DECIMAL(10,2)) FROM employees GROUP BY service\_type;

**How many staff are earning below university avg salary**

First -> SET hive.auto.convert.join=false;

SELECT COUNT (a.id) FROM (SELECT id, avg(salary) as avg\_salary from employees GROUP BY id) a JOIN (SELECT id, salary FROM employees) b ON a.id = b.id WHERE a.avg\_salary > b.salary;

**The number of staff that are in each job role.**

SELECT role, COUNT(\*) FROM employees GROUP BY role;

**MAIN TASK -> investigation**

SELECT a.role, a.full\_or\_part\_time, b.salary, a.avg\_salary\_in\_department FROM (SELECT role, full\_or\_part\_time, CAST(avg (salary) as DECIMAL(10,2)) AS avg\_salary\_in\_department from employees GROUP BY role, full\_or\_part\_time) a JOIN (SELECT role, salary, full\_or\_part\_time FROM employees) b ON a.role = b.role AND a.full\_or\_part\_time = b.full\_or\_part\_time WHERE 2 \* a.avg\_salary\_in\_department < b.salary;