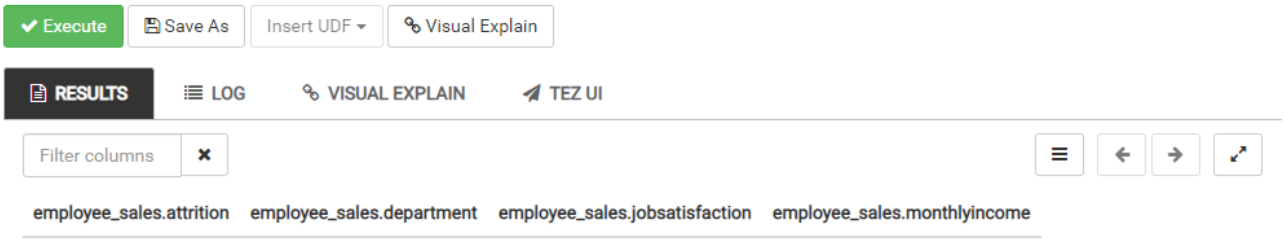
Data Transform Assignment (Human Resources)

Name: Bryan Armstrong
Student ID: 100271977

1a) & c)

Schema:

```
CREATE TABLE employee_sales
(
  Attrition string,
  Department string,
  JobSatisfaction int,
  MonthlyIncome int
  );
```



b) Schema to load employee table into new table:

```
INSERT OVERWRITE TABLE employee_sales
SELECT Attrition, Department, JobSatisfaction, MonthlyIncome
FROM employee;
```

Screenshot showing first 10 rows below

RESULTS | LOG | VISUAL EXPLAIN | TEZ UI

Filter columns

employee_sales.attrition	employee_sales.department	employee_sales.jobsatisfaction	employee_sales.monthlyincome
Yes	Sales	4	5993
No	Research & Development	2	5130
Yes	Research & Development	3	2090
No	Research & Development	3	2909
No	Research & Development	2	3468
No	Research & Development	4	3068
No	Research & Development	1	2670
No	Research & Development	3	2693
No	Research & Development	3	9526
No	Research & Development	3	5237

2) Rounded data in the MonthlyIncome to the nearest 1000 using the following schema:

```
INSERT OVERWRITE TABLE employee_sales
SELECT Attrition, Department, JobSatisfaction, ROUND(MonthlyIncome, -3) AS
MonthlyIncome
FROM employee_sales;
```

Filter columns

employee_sales.attrition	employee_sales.department	employee_sales.jobsatisfaction	employee_sales.monthlyincome
Yes	Sales	4	6000
No	Research & Development	2	5000
Yes	Research & Development	3	2000
No	Research & Development	3	3000
No	Research & Development	2	3000
No	Research & Development	4	3000
No	Research & Development	1	3000
No	Research & Development	3	3000
No	Research & Development	3	10000
No	Research & Development	3	5000

3) Filtered to find those in the sales department with schema:

```
INSERT OVERWRITE TABLE employee_sales
SELECT *
FROM employee_sales
WHERE Department LIKE "%Sales%";
```

employee_sales.attrition	employee_sales.department	employee_sales.jobsatisfaction	employee_sales.monthlyincome
Yes	Sales	4	6000
No	Sales	4	15000
Yes	Sales	1	3000
No	Sales	2	7000
No	Sales	1	19000
Yes	Sales	4	2000
Yes	Sales	3	3000
No	Sales	4	2000
No	Sales	1	5000
No	Sales	3	9000

4) Ordered the data by jobsatisfaction from highest to lowest with schema:

```
INSERT OVERWRITE TABLE employee_sales
SELECT *
FROM employee_sales
ORDER by JobSatisfaction DESC;
```

employee_sales.attrition	employee_sales.department	employee_sales.jobsatisfaction	employee_sales.monthlyincome
No	Sales	4	5000
No	Sales	4	9000
No	Sales	4	18000
No	Sales	4	10000
No	Sales	4	7000
No	Sales	4	8000
Yes	Sales	4	1000
No	Sales	4	5000
No	Sales	4	7000
No	Sales	4	10000