# SQL-Mongo Project – HR Employee Attrition

BUAN 6320

| Activity | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Prepared Data Model and Created Physical DB | x | | | |
| Loaded Data into Database | x | x | x | x |
| Wrote SQL Queries | x | x | | |
| Prepared Mongo Database | | | x | x |
| Loaded data into Mongo DB | | | x | x |
| Wrote Mongo Queries | | | | x |
| Prepared Report | x | x | x | x |
| Reviewed Report | x | x | x | x |

# Contents


Relational Data Model....4
Assumptions/Notes About Data Entities and Relationships....4
Entity-Relationship Diagram....5
Physical MySQL Database....6
Assumptions/Notes About Data Set....6
Screen shot of Physical Database objects....7
Data in the Database....20
SQL Queries....22
SQL Query 1....22
Question....22
Notes/Comments About SQL Query and Results (Include # of Rows in Result)....22
Translation....23
Screen Shot of SQL Query and Results....23
SQL Query 2....26
Question....26
Notes/Comments About SQL Query and Results (Include # of Rows in Result)....26
Translation....26
Screen Shot of SQL Query and Results....26
SQL Query 3....27
Question....27
Notes/Comments About SQL Query and Results (Include # of Rows in Result)....27
Translation....27
Screen Shot of SQL Query and Results....28
SQL Query 4....28
Question....29
Notes/Comments About SQL Query and Results (Include # of Rows in Result)....29
Translation....29
Screen Shot of SQL Query and Results....30
SQL Query 5....33
Question....33
Notes/Comments About SQL Query and Results (Include # of Rows in Result)....33

Translation....33
Screen Shot of SQL Query and Results....34
SQL Query 6....35
Question....35
Notes/Comments About SQL Query and Results (Include # of Rows in Result)....35
Translation....35
Screen Shot of SQL Query and Results....36
Data Review for MongoDB....38
Assumptions/Notes About Data Collections, Attributes and Relationships between Collections....38
Physical Mongo Database....39
Assumptions/Notes About Data Set....39
Screen shot of Physical Database objects (Database, Collections and Attributes)....39
Data in the Database....40
MongoDB Queries/Code....41
Mongo Query 1....41
Question....41
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)....41
Translation....41
Screen Shot of MongoDB Query/Code and Results....41
Mongo Query 2....42
Question....42
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)....42
Translation....42
Screen Shot of MongoDB Query/Code and Results....42
Mongo Query 3....43
Question....43
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)....43
Translation....43
Screen Shot of MongoDB Query/Code and Results....43

## Relational Data Model

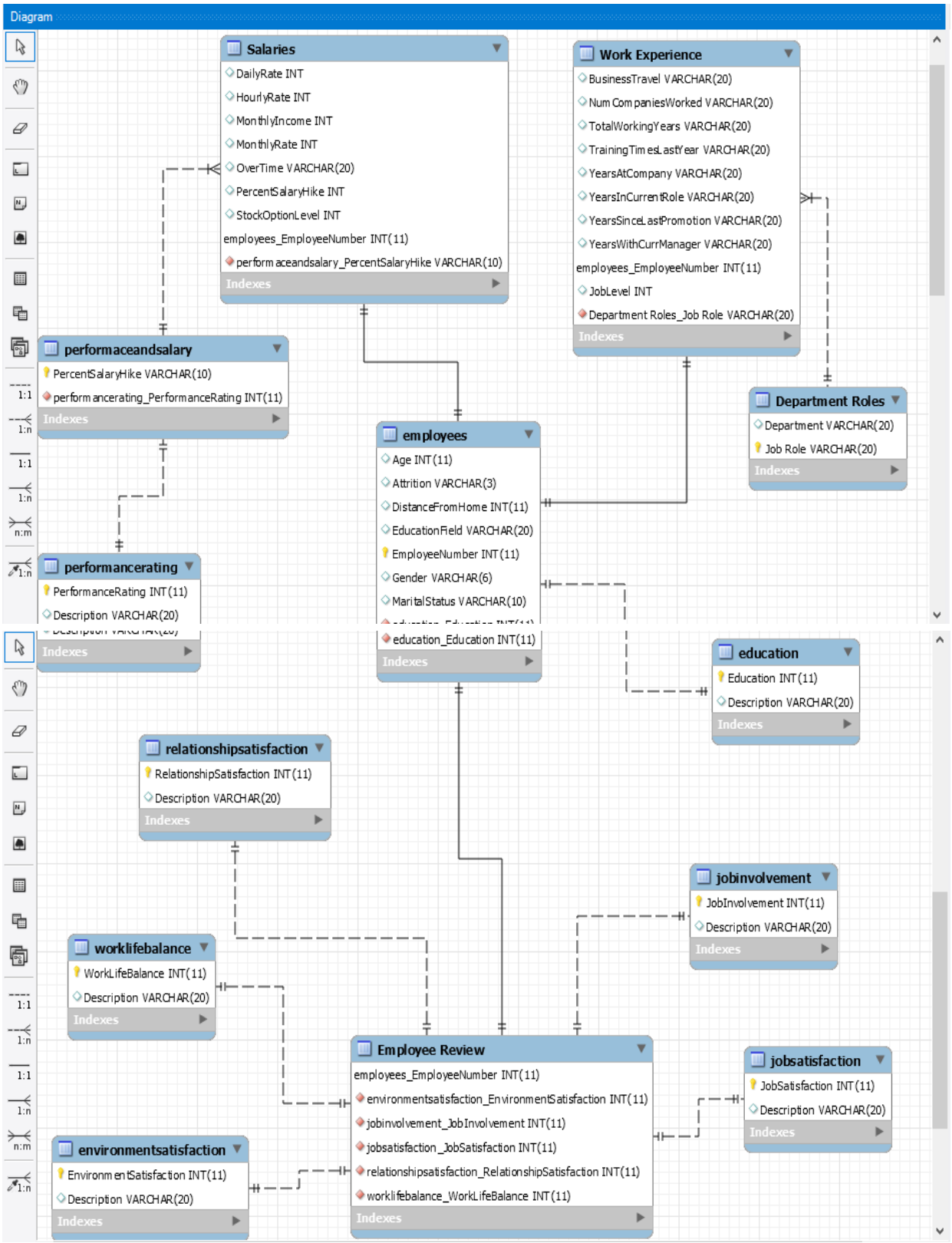
### Assumptions/Notes About Data Entities and Relationships:

The database is divided into 14 tables: some directly partitioning data and others describing it.

The employees table is one to one related to salaries, work experience, education, department roles and employee review. Salaries table is many to one related to performanceandsalary table. Every PerformanceRating from performanceandsalary has many PercentSalaryHike numerical values in the Salaries table. The performanceandsalary table is one to one related to performancerating. The Employee Review table is one to one related to jobinvolvement, jobsatisfaction, environmentsatisfaction, worklifebalance and relationshipsatisfaction.

The column Over18 has been removed because whether an employee is aged over 18 or not can be determined directly using the Age column. Also, the column EmployeeCount has been omitted from the database because the row-wise count of employee will be 1 (the entry in all rows of this column).

## Entity-Relationship Diagram



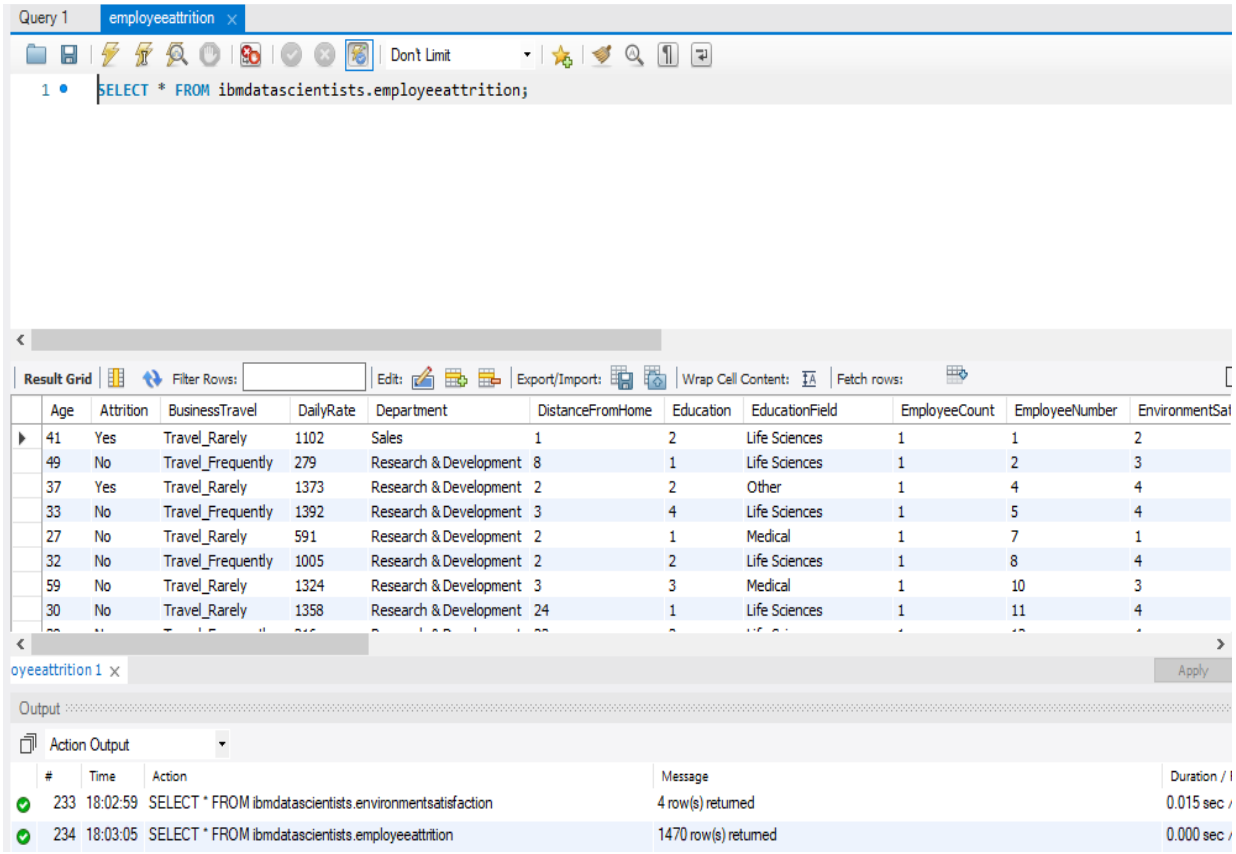
## Physical MySQL Database

### Assumptions/Notes About Data Set

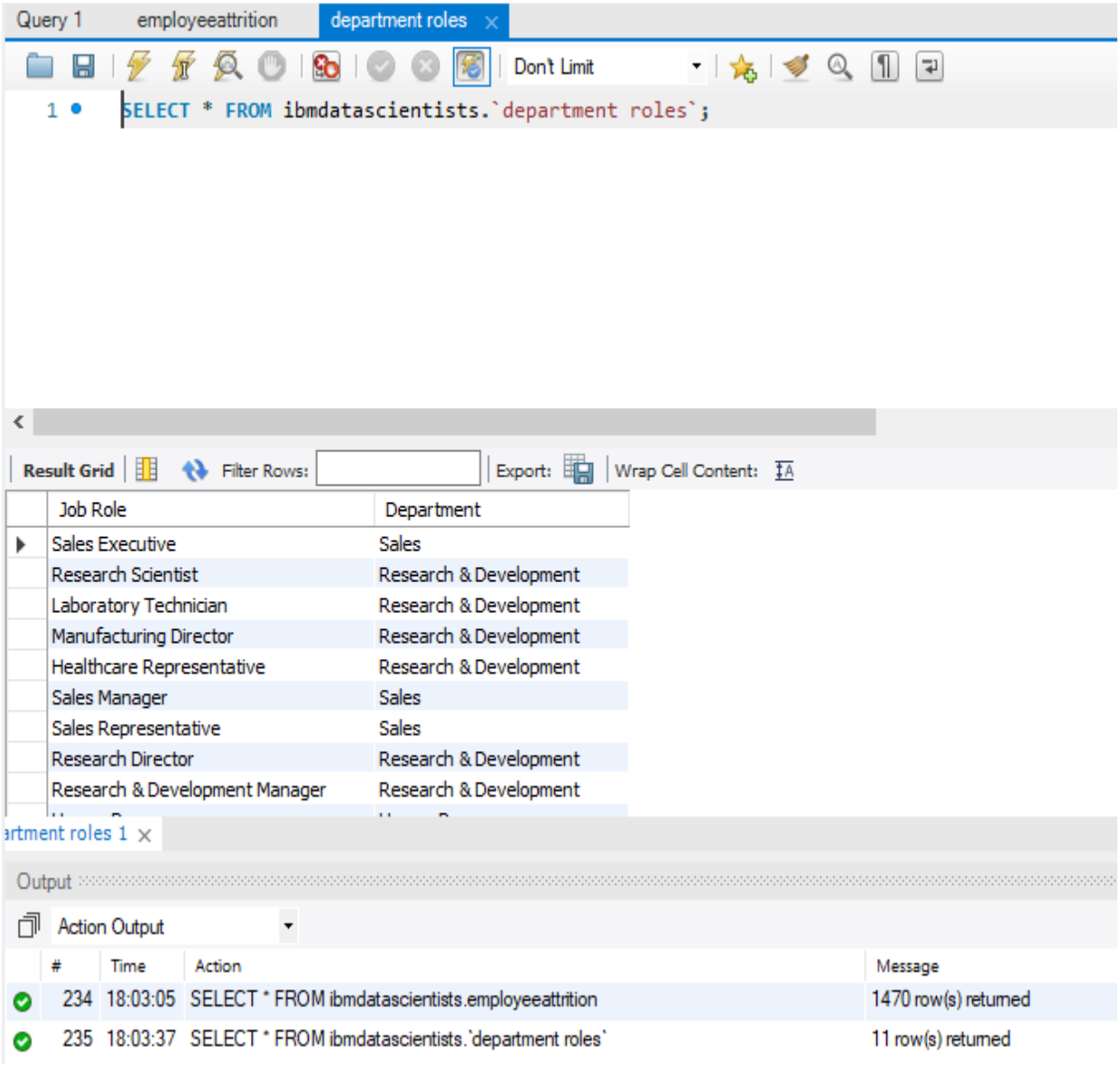
1. All static columns pertaining to an employee belong to employees table
2. Attrition implies that the particular person is no more employed
3. Any person rejoining the company would have a different EmployeeNumber and his/his details would be listed as any new employee's would, i.e., YearsAtCompany and YearsWithCurrManager will be calculated with respect to this position and this EmployeeNumber. NumCompaniesWorked will also include this company.
4. TrainingTimesLastYear are the number of trainings the employee has undertaken with this company in the last year
5. All salary details are present in Salaries table (including StockOptionLevel, which is part of salary for some employees)
6. All employees are paid monthly salaries. The hourly rate, daily rate and monthly rate are the rates that were paid to contractors that are not on the company payroll but recruited on the same day. The contract employees are not listed here.
7. All job-related experience, including job level and role in the current company, can be varying and is hence considered as part of Work Experience table
8. Every department has specific roles in it, which do not overlap with any other department
9. For employees with rating = 3 salary hike is < 20% while those with rating = 4 get a hike of ≥ 20%
10. Employees have been asked to rate their experiences at work and in personal life which are considered as Employee Review.

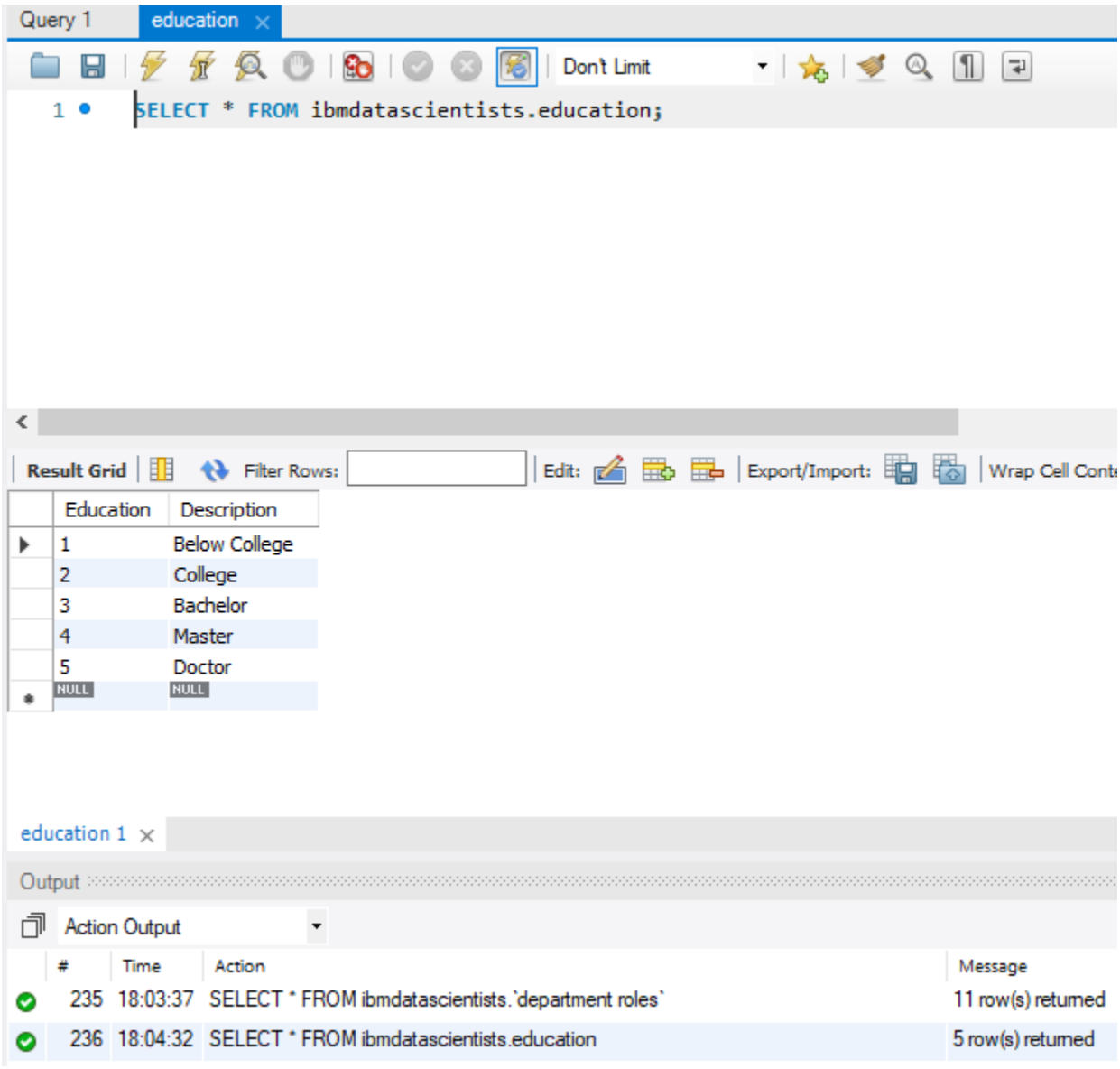
## Screen shot of Physical Database objects:

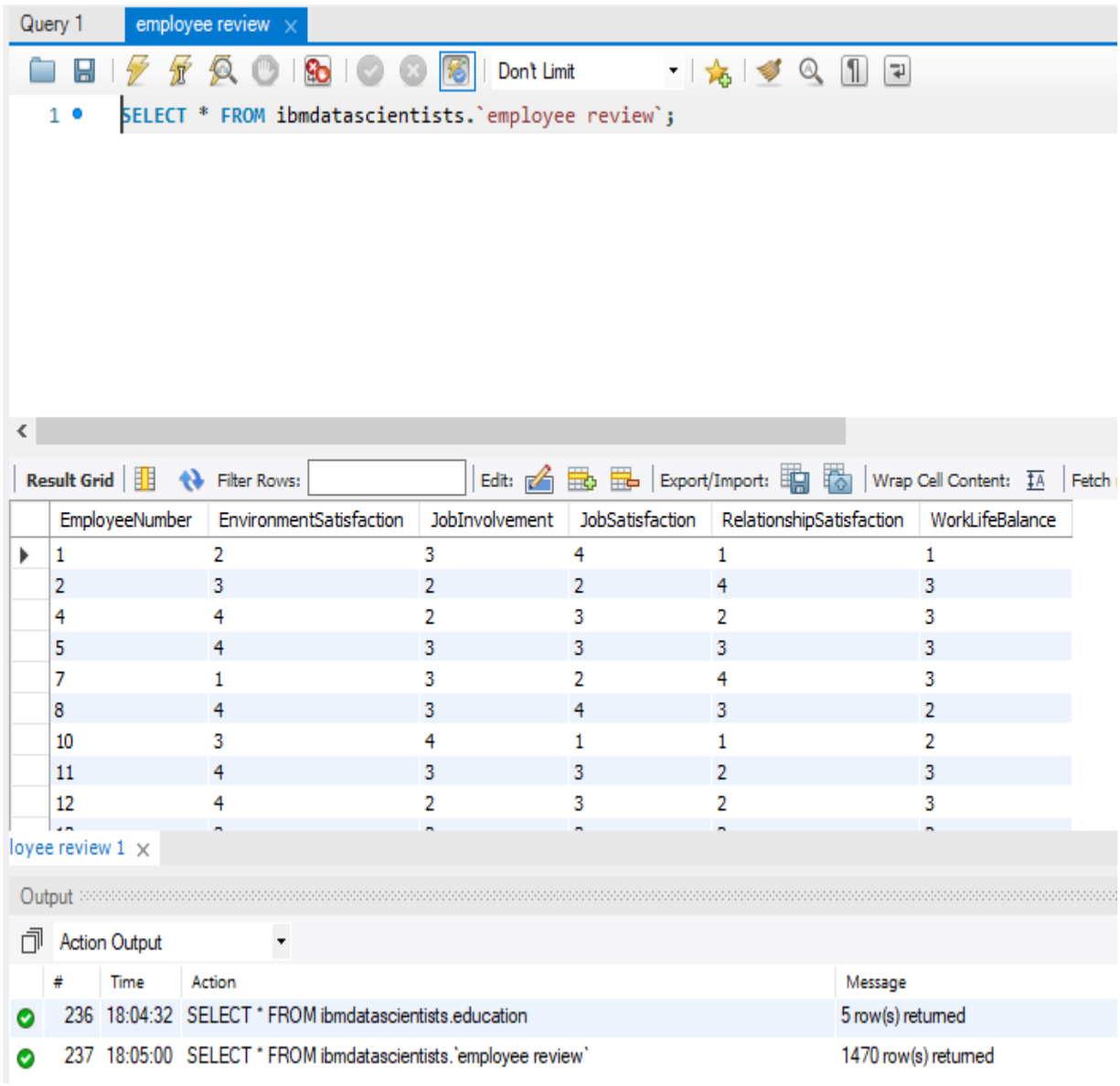
### Loaded entire dataset into MySQL Workbench:



Screenshots of individual tables:




Query 1
education
Don't Limit
1 SELECT * FROM ibmdatascientists.education;
Result Grid
Filter Rows:
Edit:
Export/Import:
Wrap Cell Cont
Education
Description
1
Below College
2
College
3
Bachelor
4
Master
5
Doctor
NULL
NULL
education 1
Output
Action Output
#
Time
Action
Message
235
18:03:37
SELECT * FROM ibmdatascientists.`department roles`
11 row(s) returned
236
18:04:32
SELECT * FROM ibmdatascientists.education
5 row(s) returned


Query 1
employee review
Don't Limit
1 SELECT * FROM ibmdatascientists.`employee review`;
Result Grid
Filter Rows:
Edit:
Export/Import:
Wrap Cell Content:
Fetch
EmployeeNumber EnvironmentSatisfaction JobInvolvement JobSatisfaction RelationshipSatisfaction WorkLifeBalance
1 2 3 4 1 1
2 3 2 2 4 3
4 4 2 3 2 3
5 4 3 3 3 3
7 1 3 2 4 3
8 4 3 4 3 2
10 3 4 1 1 2
11 4 3 3 2 3
12 4 2 3 2 3
loyee review 1
Output
Action Output
# Time Action Message
236 18:04:32 SELECT * FROM ibmdatascientists.education 5 row(s) returned
237 18:05:00 SELECT * FROM ibmdatascientists.`employee review` 1470 row(s) returned

Query 1 employees

Don't Limit

```
1  SELECT * FROM ibmdatascientists.employees;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

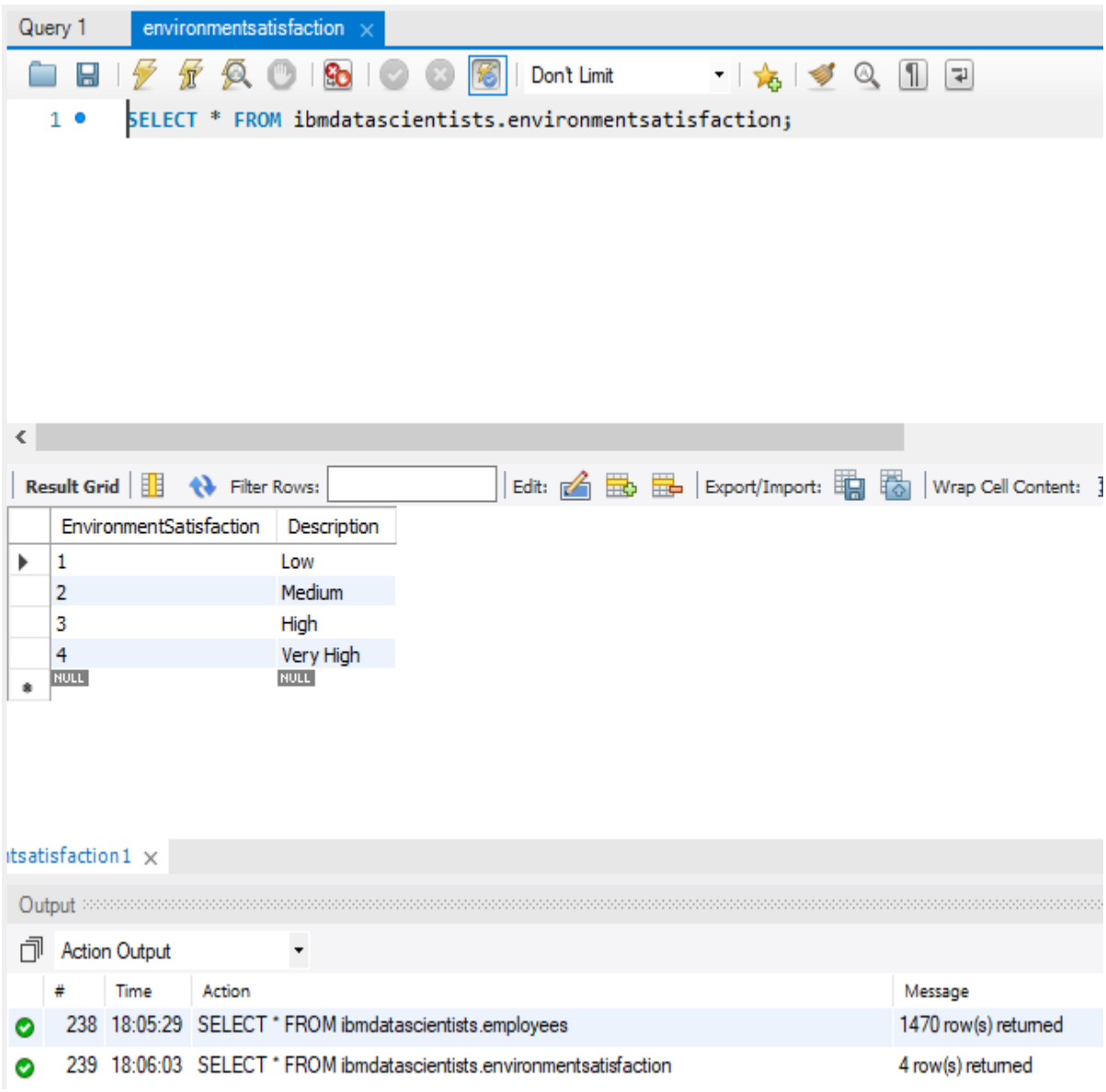
| Age | Attrition | DistanceFromHome | Education | EducationField | EmployeeNumber | Gender | MaritalStatus |
|---|---|---|---|---|---|---|---|
| 41 | Yes | 1 | 2 | Life Sciences | 1 | Female | Single |
| 49 | No | 8 | 1 | Life Sciences | 2 | Male | Married |
| 37 | Yes | 2 | 2 | Other | 4 | Male | Single |
| 33 | No | 3 | 4 | Life Sciences | 5 | Female | Married |
| 27 | No | 2 | 1 | Medical | 7 | Male | Married |
| 32 | No | 2 | 2 | Life Sciences | 8 | Male | Single |
| 59 | No | 3 | 3 | Medical | 10 | Female | Married |
| 30 | No | 24 | 1 | Life Sciences | 11 | Male | Divorced |
| 38 | No | 23 | 3 | Life Sciences | 12 | Male | Single |

employees 1

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 237 | 18:05:00 | SELECT * FROM ibmdatascientists.`employee review` | 1470 row(s) returned |
| 238 | 18:05:29 | SELECT * FROM ibmdatascientists.employees | 1470 row(s) returned |


Query 1
environmentsatisfaction
Don't Limit
1
SELECT * FROM ibmdatascientists.environmentsatisfaction;
Result Grid
Filter Rows:
Edit:
Export/Import:
Wrap Cell Content:
EnvironmentSatisfaction
Description
1
Low
2
Medium
3
High
4
Very High
NULL
NULL
itsatisfaction 1
Output
Action Output
#
Time
Action
Message
238
18:05:29
SELECT * FROM ibmdatascientists.employees
1470 row(s) returned
239
18:06:03
SELECT * FROM ibmdatascientists.environmentsatisfaction
4 row(s) returned

Query 1 | jobinvolvement

Don't Limit

```
1 • SELECT * FROM ibmdatascientists.jobinvolvement;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Conte

| JobInvolvement | Description |
|---|---|
| 1 | Low |
| 2 | Medium |
| 3 | High |
| 4 | Very High |
| NULL | NULL |

jinvolvement 1

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 239 | 18:06:03 | SELECT * FROM ibmdatascientists.environmentsatisfaction | 4 row(s) returned |
| 240 | 18:06:24 | SELECT * FROM ibmdatascientists.jobinvolvement | 4 row(s) returned |

Query 1 jobsatisfaction

Don't Limit

```
1  SELECT * FROM ibmdatascientists.jobsatisfaction;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Cont

| JobSatisfaction | Description |
|---|---|
| 1 | Low |
| 2 | Medium |
| 3 | High |
| 4 | Very High |
| NULL | NULL |

bsatisfaction 1

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 240 | 18:06:24 | SELECT * FROM ibmdatascientists.jobinvolvement | 4 row(s) returned |
| 241 | 18:06:50 | SELECT * FROM ibmdatascientists.jobsatisfaction | 4 row(s) returned |

Query 1 performaceandsalary

Don't Limit

```
1  SELECT * FROM ibmdatascientists.performaceandsalary;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Cont

| PercentSalaryHike | PerformanceRating |
|---|---|
| <20 | 3 |
| >=20 | 4 |
| NULL | NULL |

aceandsalary 1

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 241 | 18:06:50 | SELECT * FROM ibmdatascientists.jobsatisfaction | 4 row(s) returned |
| 242 | 18:07:12 | SELECT * FROM ibmdatascientists.performaceandsalary | 2 row(s) returned |

Query 1 | performance rating

Don't Limit

```
1  SELECT * FROM ibmdatascientists.`performance rating`;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

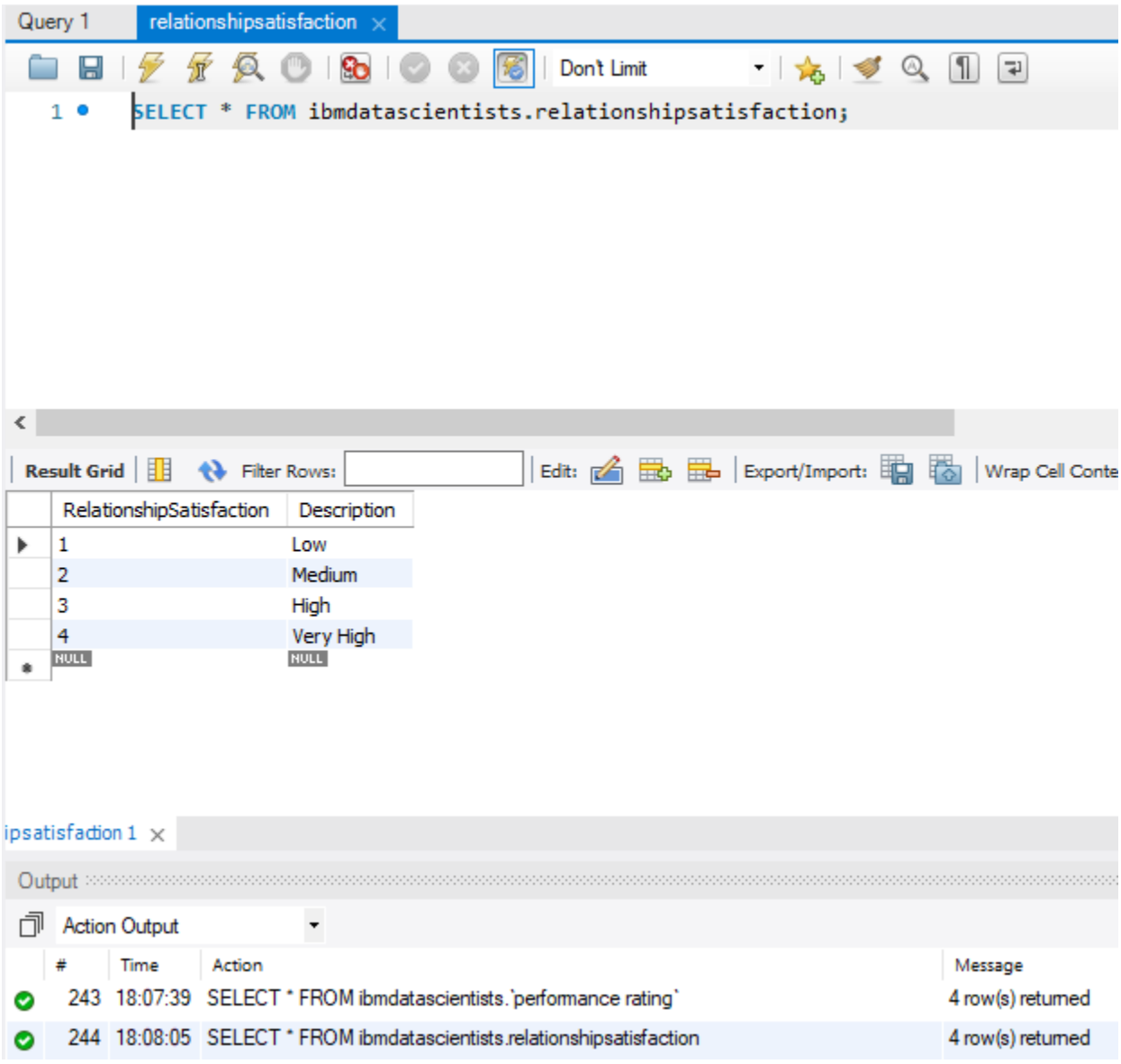
| PerformanceRating | Description |
|---|---|
| 1 | Low |
| 2 | Good |
| 3 | Excellent |
| 4 | Outstanding |

mance rating 1

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 242 | 18:07:12 | SELECT * FROM ibmdatascientists.performaceandsalary | 2 row(s) returned |
| 243 | 18:07:39 | SELECT * FROM ibmdatascientists.`performance rating` | 4 row(s) returned |


Query 1
relationshipsatisfaction
Don't Limit
1 SELECT * FROM ibmdatascientists.relationshipsatisfaction;
Result Grid
Filter Rows:
Edit:
Export/Import:
Wrap Cell Conte
RelationshipSatisfaction
Description
1
Low
2
Medium
3
High
4
Very High
NULL
NULL
ipsatisfaction 1
Output
Action Output
#
Time
Action
Message
243
18:07:39
SELECT * FROM ibmdatascientists.`performance rating`
4 row(s) returned
244
18:08:05
SELECT * FROM ibmdatascientists.relationshipsatisfaction
4 row(s) returned

Query 1 | salaries

Don't Limit

```
1 • SELECT * FROM ibmdatascientists.salaries;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: | Fetch r

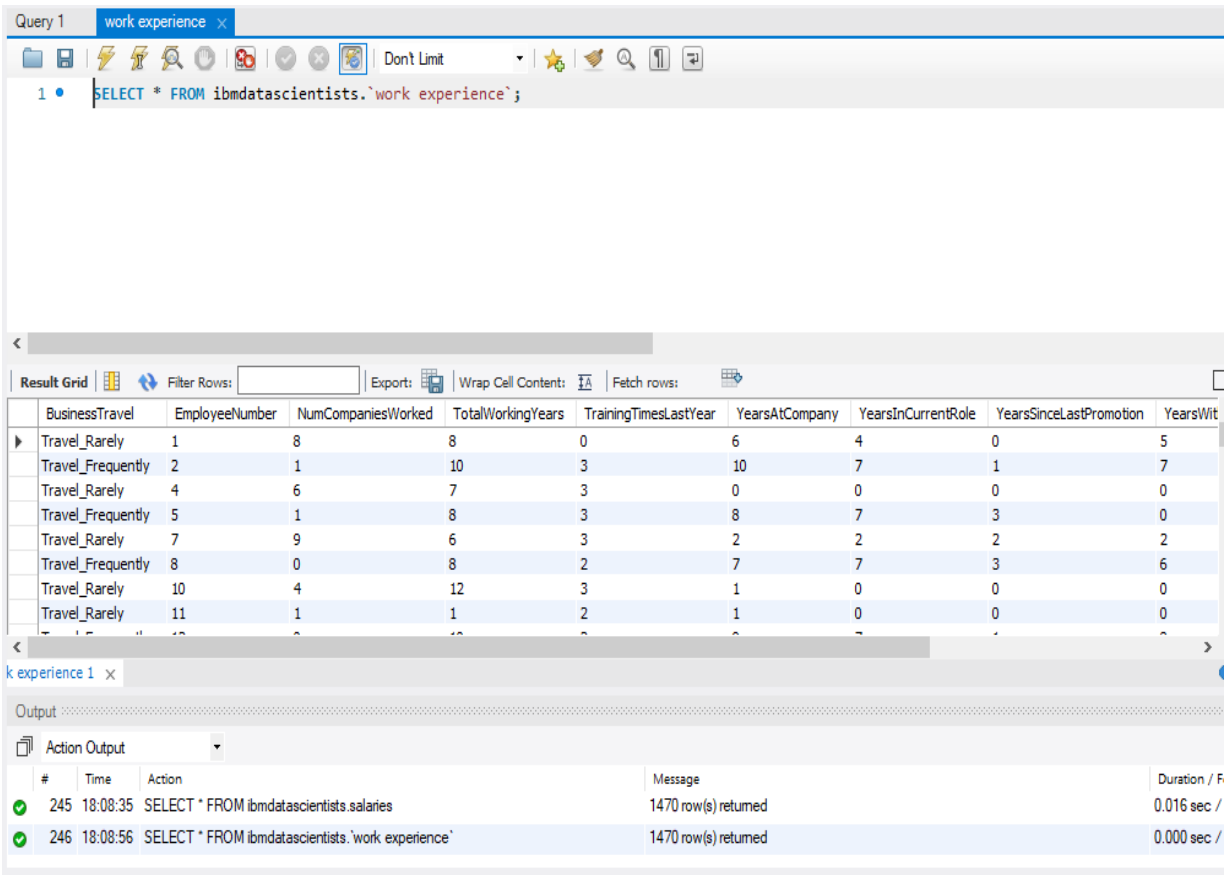
| DailyRate | EmployeeNumber | HourlyRate | MonthlyIncome | MonthlyRate | OverTime | PercentSalaryHike | StockOptionLevel |
|---|---|---|---|---|---|---|---|
| 1102 | 1 | 94 | 5993 | 19479 | Yes | 11 | 0 |
| 279 | 2 | 61 | 5130 | 24907 | No | 23 | 1 |
| 1373 | 4 | 92 | 2090 | 2396 | Yes | 15 | 0 |
| 1392 | 5 | 56 | 2909 | 23159 | Yes | 11 | 0 |
| 591 | 7 | 40 | 3468 | 16632 | No | 12 | 1 |
| 1005 | 8 | 79 | 3068 | 11864 | No | 13 | 0 |
| 1324 | 10 | 81 | 2670 | 9964 | Yes | 20 | 3 |
| 1358 | 11 | 67 | 2693 | 13335 | No | 22 | 1 |
| 216 | 12 | 44 | 9526 | 8787 | No | 21 | 0 |

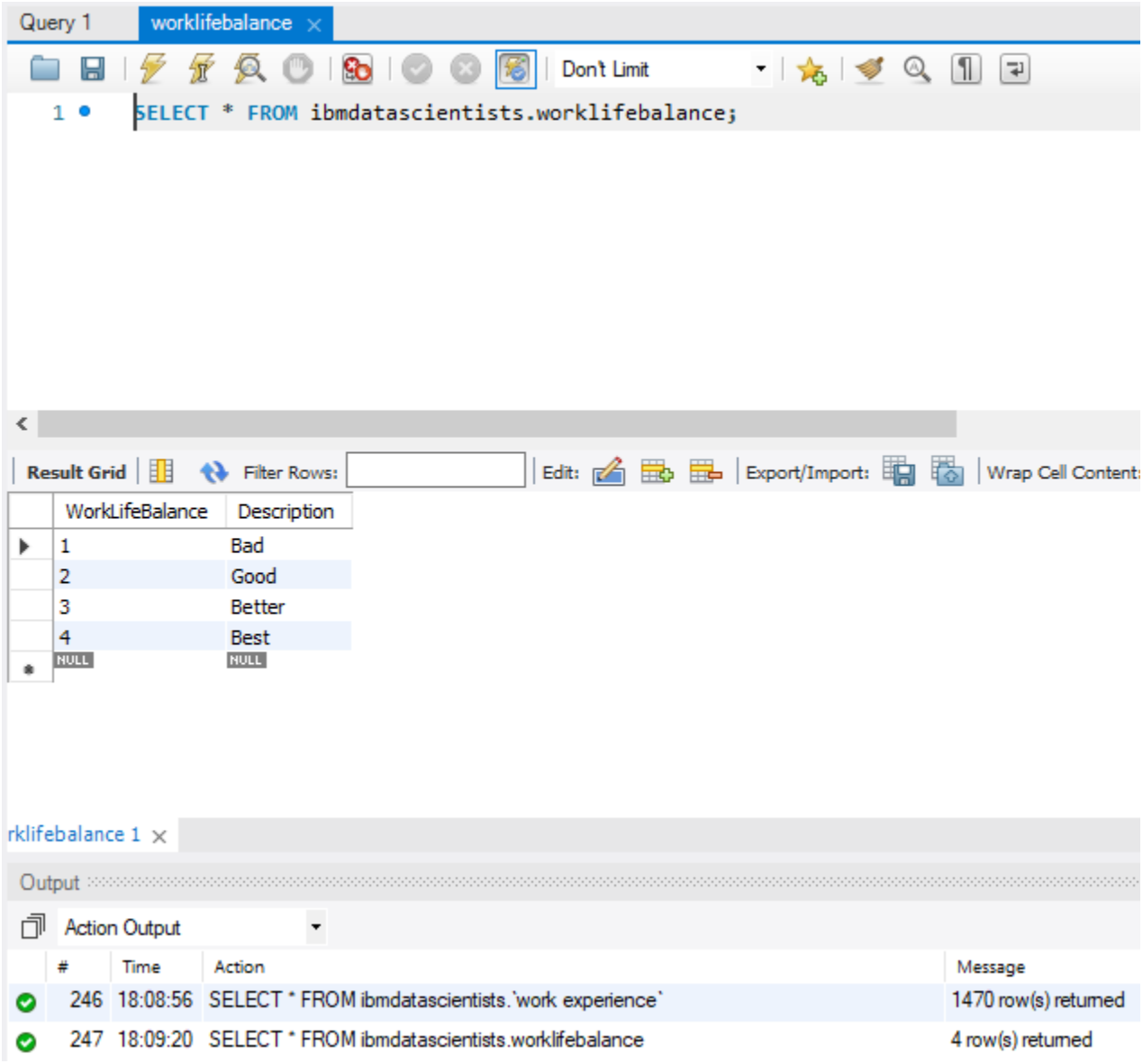
salaries 1

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 244 | 18:08:05 | SELECT * FROM ibmdatascientists.relationshipsatisfaction | 4 row(s) returned |
| 245 | 18:08:35 | SELECT * FROM ibmdatascientists.salaries | 1470 row(s) returned |


Query 1
work experience
Don't Limit
1 • SELECT * FROM ibmdatascientists.`work experience`;
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
Fetch rows:
BusinessTravel | EmployeeNumber | NumCompaniesWorked | TotalWorkingYears | TrainingTimesLastYear | YearsAtCompany | YearsInCurrentRole | YearsSinceLastPromotion | YearsWit
Travel_Rarely | 1 | 8 | 8 | 0 | 6 | 4 | 0 | 5
Travel_Frequently | 2 | 1 | 10 | 3 | 10 | 7 | 1 | 7
Travel_Rarely | 4 | 6 | 7 | 3 | 0 | 0 | 0 | 0
Travel_Frequently | 5 | 1 | 8 | 3 | 8 | 7 | 3 | 0
Travel_Rarely | 7 | 9 | 6 | 3 | 2 | 2 | 2 | 2
Travel_Frequently | 8 | 0 | 8 | 2 | 7 | 7 | 3 | 6
Travel_Rarely | 10 | 4 | 12 | 3 | 1 | 0 | 0 | 0
Travel_Rarely | 11 | 1 | 1 | 2 | 1 | 0 | 0 | 0
k experience 1
Output
Action Output
# | Time | Action | Message | Duration / F
245 | 18:08:35 | SELECT * FROM ibmdatascientists.salaries | 1470 row(s) returned | 0.016 sec /
246 | 18:08:56 | SELECT * FROM ibmdatascientists.`work experience` | 1470 row(s) returned | 0.000 sec /


Query 1
worklifebalance
Don't Limit
1 SELECT * FROM ibmdatascientists.worklifebalance;
Result Grid
Filter Rows:
Edit:
Export/Import:
Wrap Cell Content:
WorkLifeBalance
Description
1
Bad
2
Good
3
Better
4
Best
NULL
NULL
rklifebalance 1
Output
Action Output
#
Time
Action
Message
246
18:08:56
SELECT * FROM ibmdatascientists.`work experience`
1470 row(s) returned
247
18:09:20
SELECT * FROM ibmdatascientists.worklifebalance
4 row(s) returned

## Data in the Database

| Table Name | Primary Key | Foreign Key | # of Rows in Table |
|---|---|---|---|
| Department roles | Job Role | | 11 |
| Education | Education | | 5 |
| Employee review | EmployeeNumber | EmployeeNumber<br>EmployeeSatisfaction<br>JobInvolvement<br>JobSatisfaction<br>RelationshipSatisfaction<br>WorkLifeBalance | 1470 |
| Employees | EmployeeNumber | Education | 1470 |
| Environmentsatisfaction | EnvironmentSatisfaction | | 4 |
| Jobinvolvement | JobInvolvement | | 4 |
| Jobsatisfaction | JobSatisfaction | | 4 |
| performanceandsalary | PercentSalaryHike | PerformanceRating | 2 |
| Performance rating | PerformanceRating | | 4 |
| Relationshipsatisfaction | RelationshipSatisfaction | | 4 |
| Salaries | EmployeeNumber | EmployeeNumber<br>PercentSalaryHike | 1470 |
| Work experience | EmployeeNumber | EmployeeNumber<br>Job Role | 1470 |
| worklifebalance | WorkLifeBalance | | 4 |

## SQL Queries

### SQL Query 1

Question: The company has decided to focus on cost cutting to improve their bottom line in a difficult economy. Should the company focus on Business Travel? Why or Why not?

Notes/Comments About SQL Query and Results (Include # of Rows in Result):

1. We consider just the currently employed people for this analysis because cost cutting can only be done on the currently employed
2. Assuming that the budget for every employee's travel is the same ($100)
3. Assuming that the employees travelling frequently travel once a month and employees travelling rarely travel once every quarter
4. Assuming that only frequent and rare travelers like traveling for business purpose and non-travelers don't like to
5. The output of the query contains 3 rows, one for each category of business travelers. We are populating the total number of employees, percentage of these employees that are highly satisfied and percentage of these highly satisfied employees that have high performance.
6. We find that frequent travelers have highest satisfaction and non-travelers have highest performance
7. For the sake of simplicity if we consider that every employee spends $100 on every trip then the frequent travelers will be spending 208 x 3 x 100 = $62,400 in a quarter while the rare travelers spend 887 x 100 = $88,700 in a quarter. That means, we spend $151,100 per quarter on travel.
8. If frequent travelers are made to travel just once a quarter, their satisfaction level may drop but there may not be a lot of difference in their performance because that is directly related to the percent of salary hike they will receive. Additionally, travel expenditure may drop to (887 + 208) x 100 = $109,500 per quarter, saving 151100 – 109500 = $41,600 per quarter
9. On the flip side, the satisfaction level of non-travelers will increase, if they
10. So, we suggest that any employee traveling for business purpose can do so only once per quarter and yes, business travel is an important factor that the company must focus on for cost cutting

Translation:
Select business travel categories, count of employees from work experience table where employees belong to Attrition = ’No‘ joined with select list of percentage of employees from work experience and employee review table for those who have job satisfaction = 4 and have Attrition = ’No‘ grouped by business travel categoories joined with select list of percentage of employees from work experience, employee review and salaries tables for those who have job satisfaction = 4, Attrition = ’No‘ and PercentSalaryHike >= 20 (signifying performance = 4 which is high) grouped by the business travel categories

Screen Shot of SQL Query and Results:
SQL Query:

```
select p.BusinessTravel, `Count Employees`, `Percent Satisfied employees`,
`Percent High Performance Satisfied employees`

from

(select BusinessTravel, count(EmployeeNumber) as `Count Employees`

from `work experience`

where EmployeeNumber in (select EmployeeNumber from employees where
Attrition = 'No')

group by BusinessTravel) as p

join

(select x.BusinessTravel, round((`Satisfied employees`/`Total Satisfied
employees`)*100, 2) as `Percent Satisfied employees`

from

(select BusinessTravel, count(we.EmployeeNumber) as `Satisfied employees`

from `work experience` as we

join `employee review` as er

on we.EmployeeNumber = er.EmployeeNumber

where we.EmployeeNumber in (select EmployeeNumber from employees
where Attrition = 'No') and JobSatisfaction = 4

group by BusinessTravel) as x
```

```
join
(select BusinessTravel, count(we.EmployeeNumber) as `Total Satisfied employees`
from `work experience` as we
join `employee review` as er
on we.EmployeeNumber = er.EmployeeNumber
where we.EmployeeNumber in (select EmployeeNumber from employees where Attrition = 'No')
group by BusinessTravel) as y
on x.BusinessTravel = y.BusinessTravel) as q
on p.BusinessTravel = q.BusinessTravel
join
(select x.BusinessTravel, round((`High Performance Satisfied employees`/`Satisfied employees`)*100, 2) as `Percent High Performance Satisfied employees`
from
(select BusinessTravel, count(we.EmployeeNumber) as `High Performance Satisfied employees`
from `work experience` as we
join `employee review` as er
on we.EmployeeNumber = er.EmployeeNumber
join salaries as s
on we.EmployeeNumber = s.EmployeeNumber
where we.EmployeeNumber in (select EmployeeNumber from employees where Attrition = 'No') and JobSatisfaction = 4 and s.PercentSalaryHike >= 20
group by BusinessTravel) as x
join
(select BusinessTravel, count(we.EmployeeNumber) as `Satisfied employees`
from `work experience` as we
```

join `employee review` as er

on we.EmployeeNumber = er.EmployeeNumber

join salaries as s

on we.EmployeeNumber = s.EmployeeNumber

where we.EmployeeNumber in (select EmployeeNumber from employees where Attrition = 'No') and JobSatisfaction = 4

group by BusinessTravel) as y

on x.BusinessTravel = y.BusinessTravel) as r

on r.BusinessTravel = p.BusinessTravel;

```
222 •  select p.BusinessTravel, `Count Employees`, `Percent Satisfied employees`, `Percent High Performance Satisfied employees`
223    from
224    (select BusinessTravel, count(EmployeeNumber) as `Count Employees`
225    from `work experience`
226    where EmployeeNumber in (select EmployeeNumber from employees where Attrition = 'No')
227    group by BusinessTravel) as p
228    join
229    (select x.BusinessTravel, round((`Satisfied employees`/`Total Satisfied employees`)*100, 2) as `Percent Satisfied employees`
230    from
231    (select BusinessTravel, count(we.EmployeeNumber) as `Satisfied employees`
232    from `work experience` as we
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| BusinessTravel | Count Employees | Percent Satisfied employees | Percent High Performance Satisfied employees |
|---|---|---|---|
| Travel_Frequently | 208 | 37.02 | 16.88 |
| Travel_Rarely | 887 | 32.36 | 17.77 |
| Non-Travel | 138 | 31.16 | 18.60 |

Result 33

Output

Action Output

| # | Time | Action | Message | D |
|---|---|---|---|---|
| 39 | 20:55:25 | select p.BusinessTravel, `Count Employees`, `Percent Satisfied employees`, `Percent... | 3 row(s) returned | 0. |
| 40 | 20:57:30 | select p.BusinessTravel, `Count Employees`, `Percent Satisfied employees`, `Percent... | 3 row(s) returned | 0. |

## SQL Query 2

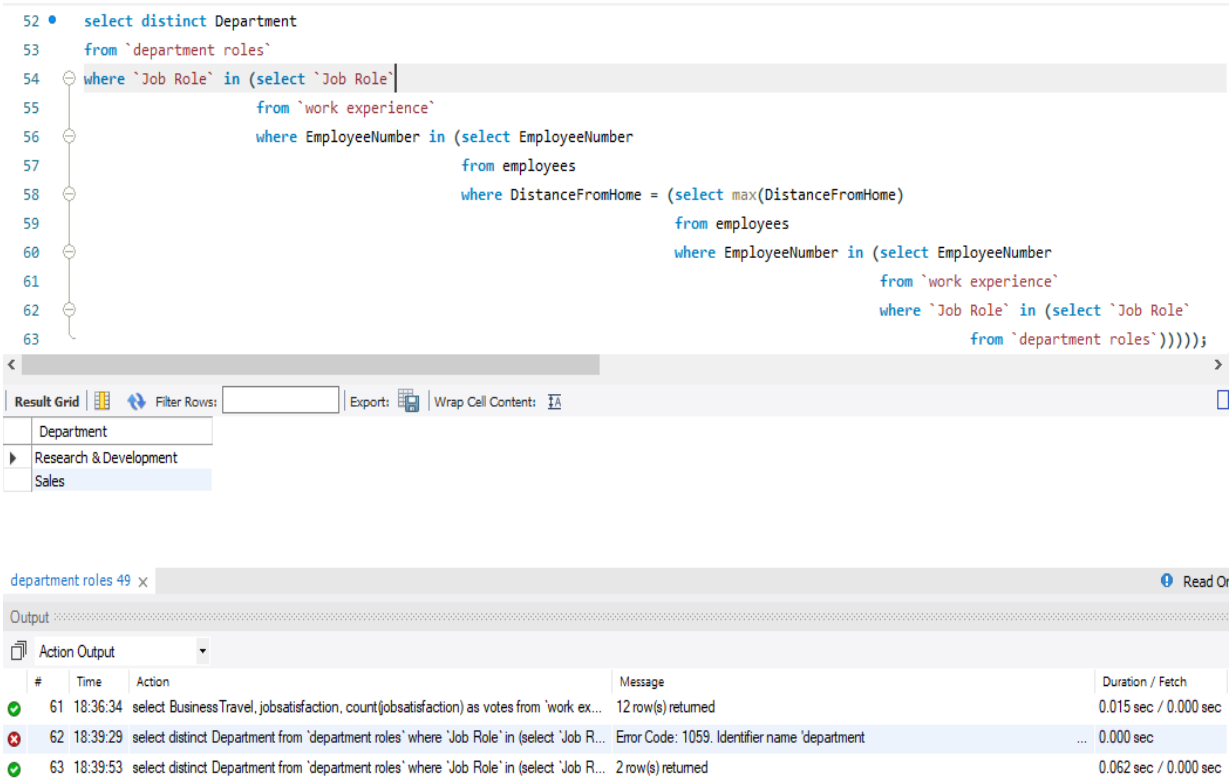
Question: Which department's employee is the most likely to have the longest commute between home and work?

Notes/Comments About SQL Query and Results (Include # of Rows in Result):

1. The result set contains 2 rows
2. Employees of the departments ‘Research & Development’ and ‘Sales’ are most likely to have longest commute between home and work

Translation: select department name from department roles table for all the job roles that are in the work experience table whose employee number column is in empoloyees table and distance from home is maximum

Screen Shot of SQL Query and Results:



## SQL Query 3

Question: A new employee with a Technical Degree wants to work in Sales. Do you believe the company might be able to give her a chance to work in Sales? Why or Why not?

Notes/Comments About SQL Query and Results (Include # of Rows in Result):

1. Assuming that the new employee is fully capable to working in the sales team. What we want to verify is if there have been female employees with technical degrees who have worked in Sales in the past. If such employees indeed exist, this particular employee will be able to work in Sales.
2. The following query tells us the names of departments in which there have been female employees with technical degrees. It contains 3 rows, one of them being Sales.
3. Yes, the company will be able to give the new employee a chance to work in Sales.

Translation: select Department name from department roles table where job role in work experience table where employee number in employees table where education field is technical degree and gender is female

## Screen Shot of SQL Query and Results:

```
71  select distinct Department
72  from `department roles`
73  where `Job Role` in (select `Job Role`
74                       from `work experience`
75                       where EmployeeNumber in (select EmployeeNumber
76                                                from employees
77                                                where EducationField = 'Technical Degree' and Gender = 'Female'));
78
79
80
81
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Department |
|---|
| Research & Development |
| Sales |
| Human Resources |

department roles 50

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 62 | 18:39:29 | select distinct Department from `department roles` where `Job Role` in (select `Job R... | Error Code: 1059. Identifier name 'department |
| 63 | 18:39:53 | select distinct Department from `department roles` where `Job Role` in (select `Job R... | 2 row(s) returned |
| 64 | 18:42:56 | select distinct Department from `department roles` where `Job Role` in (select `Job R... | 3 row(s) returned |

## SQL Query 4

Question: The Sales department feels they have the highest job satisfaction while Research & Development department feels their department has the highest environment satisfaction. Who is right?

Notes/Comments About SQL Query and Results (Include # of Rows in Result):

1. We are going to find the proportion of employees who have ‘very high’ job satisfaction and ‘very high’ environment satisfaction in each department and get 3 rows in the result set
2. Sales department indeed has highest job satisfaction while Research & Development department has highest environment satisfaction

Translation: select department name from department roles table join with work experience table on job role join with employee review table on employee number join with job satisfaction table on employee number where job satisfaction description is very high. Also display count of job satisfaction grouped by department name. Join the result with the result of the following query. select department name from department roles table join with work experience table on job role join with employee review table on employee number join with environment satisfaction table on employee number where environment satisfaction description is very high. Also display count of environment satisfaction grouped by department name. Repeat the same to find the total number of employees belonging to each department and divide the counts with the totals.

Screen Shot of SQL Query and Results:

```
select x.Department, x.`Job satisfaction`/y.`Total Job satisfaction` as `Job satisfaction`,
x.`Environment satisfaction`/y.`Total Environment satisfaction` as `Environment satisfaction`
from
(select a.Department, `Job satisfaction`, `Environment satisfaction`
from
(select Department, count(jobsatisfaction.Description) as `Job satisfaction`
from `department roles`
join `work experience`
on `department roles`.`Job Role` = `work experience`.`Job Role`
join `employee review`
on `work experience`.EmployeeNumber = `employee review`.EmployeeNumber
join Jobsatisfaction
on `employee review`.JobSatisfaction = Jobsatisfaction.JobSatisfaction
where jobsatisfaction.Description = 'Very High'
group by Department)a
join
(select Department, count(environmentsatisfaction.Description) as `Environment satisfaction`
from `department roles`
join `work experience`
on `department roles`.`Job Role` = `work experience`.`Job Role`
join `employee review`
on `work experience`.EmployeeNumber = `employee review`.EmployeeNumber
join environmentsatisfaction
on `employee review`.EnvironmentSatisfaction =
environmentsatisfaction.EnvironmentSatisfaction
where environmentsatisfaction.Description = 'Very High'
group by Department)b
on a.Department = b.Department) as x
join
(select a.Department, `Total Job satisfaction`, `Total Environment satisfaction`
from
(select Department, count(jobsatisfaction.Description) as `Total Job satisfaction`
from `department roles`
join `work experience`
on `department roles`.`Job Role` = `work experience`.`Job Role`
```

```
join `employee review`
on `work experience`.EmployeeNumber = `employee review`.EmployeeNumber
join Jobsatisfaction
on `employee review`.JobSatisfaction = Jobsatisfaction.JobSatisfaction
group by Department)a
join
(select Department, count(environmentsatisfaction.Description) as `Total Environment
satisfaction`
from `department roles`
join `work experience`
on `department roles`.`Job Role` = `work experience`.`Job Role`
join `employee review`
on `work experience`.EmployeeNumber = `employee review`.EmployeeNumber
join environmentsatisfaction
on `employee review`.EnvironmentSatisfaction =
environmentsatisfaction.EnvironmentSatisfaction
group by Department)b
on a.Department = b.Department) as y
on x.Department = y.Department;
```

```
97  select x.Department, x.`Job satisfaction`/y.`Total Job satisfaction` as `Job satisfaction`,
98  x.`Environment satisfaction`/y.`Total Environment satisfaction` as `Environment satisfaction`
99  from
100 (select a.Department, `Job satisfaction`, `Environment satisfaction`
101 from
102 (select Department, count(jobsatisfaction.Description) as `Job satisfaction`
103 from `department roles`
104 join `work experience`
105 on `department roles`.`Job Role` = `work experience`.`Job Role`
106 join `employee review`
107 on `work experience`.EmployeeNumber = `employee review`.EmployeeNumber
108 join Jobsatisfaction
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Department | Job satisfaction | Environment satisfaction |
|---|---|---|
| Sales | 0.3296 | 0.2848 |
| Research & Development | 0.3070 | 0.3174 |
| Human Resources | 0.2698 | 0.2222 |

Result 58

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 71 | 23:06:20 | select a.Department, 'Job satisfaction', 'Environment satisfaction' from (select Depa... | 3 row(s) returned |
| 72 | 23:08:15 | select a.Department, 'Total Job satisfaction', 'Total Environment satisfaction' from (s... | 3 row(s) returned |
| 73 | 23:09:00 | select x.Department, x.'Job satisfaction'/y.'Total Job satisfaction' as 'Job satisfactio... | 3 row(s) returned |

## SQL Query 5

Question: Company has put out a public statement saying that they have no gender gap when it comes to compensation, in all the departments. What insight can you provide to prove or disprove that statement?

Notes/Comments About SQL Query and Results (Include # of Rows in Result):

1. The result set consists of 3 rows, one corresponding to each department
2. Females receive higher monthly income and stocks than male employees in both Human Resources and Sales
3. Female employees receive more monthly income than male employees in Research & Development, but they receive less stocks essentially balancing out the monthly income in this department

Translation: select department name, gender of employee, average of stock option level and average monthly income from department roles table joined with work experience table on job role joined with employees table on employee number joined with salaries table on employee number

## Screen Shot of SQL Query and Results:

```
152 •  select Department, Gender, avg(StockOptionLevel) as AvgStockOptionLevel, avg(MonthlyIncome) as AverageMonthlyIncome
153    from `department roles`
154    join `work experience`
155    on `department roles`.`Job Role` = `work experience`.`Job Role`
156    join employees
157    on `work experience`.EmployeeNumber = employees.EmployeeNumber
158    join salaries
159    on employees.EmployeeNumber = salaries.EmployeeNumber
160    group by Department, Gender
161    order by Department, Gender;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Department | Gender | AvgStockOptionLevel | AverageMonthlyIncome |
|---|---|---|---|
| Human Resources | Female | 0.8500 | 7264.0000 |
| Human Resources | Male | 0.7442 | 6371.0233 |
| Research & Development | Female | 0.7678 | 6513.6913 |
| Research & Development | Male | 0.8282 | 6129.8883 |
| Sales | Female | 0.7989 | 6972.1270 |
| Sales | Male | 0.7549 | 6949.6459 |

Result 59

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 72 | 23:08:15 | select a.Department, `Total Job satisfaction`, `Total Environment satisfaction` from (s... | 3 row(s) returned |
| 73 | 23:09:00 | select x.Department, x.`Job satisfaction`/y.`Total Job satisfaction` as `Job satisfactio... | 3 row(s) returned |

## SQL Query 6

Question: HR is trying to determine whether gender and marital status affect performance ratings of employees in each department. What initial finding can you obtain from the data to help in this regard?

Notes/Comments About SQL Query and Results (Include # of Rows in Result):

1. The result set consists of 18 rows, each giving information on department, gender, marital status and employees' average performance rating
2. Married females and single males perform well in Human Resources
3. Performances of both, male and female employees is comparable irrespective of their marital statuses
4. Female employees perform better than male employees in Sales

Translation: select department name, gender, marital status and average performance rating for every department name, gender and marital statusfrom employees table joined with salaries on employee number joined with performance and salary on percent salary hike joined with performance rating on performance rating column joined with work experience on employee number and joined with department roles on job role

Screen Shot of SQL Query and Results:

```
select Department, Gender, MaritalStatus, round(avg(performaceandsalary.PerformanceRating),
2) as PerformanceRating
from employees
join (select *, case when PercentSalaryHike >= 20 then '>=20'
                                                  else '<20' end as PSH
            from salaries) as a
on employees.EmployeeNumber = a.EmployeeNumber
join performaceandsalary
on a.PSH = performaceandsalary.PercentSalaryHike
join `performance rating`
on performaceandsalary.PerformanceRating = `performance rating`.PerformanceRating
join `work experience`
on employees.EmployeeNumber = `work experience`.EmployeeNumber
join `department roles`
on `work experience`.`Job Role` = `department roles`.`Job Role`
group by Department, Gender, MaritalStatus
order by Department, Gender, MaritalStatus;
```

```
166 •  select Department, Gender, MaritalStatus, round(avg(performaceandsalary.PerformanceRating), 2) as PerformanceRating
167    from employees
168    join (select *, case when PercentSalaryHike >= 20 then '>=20'
169                        else '<20' end as PSH
170          from salaries) as a
171    on employees.EmployeeNumber = a.EmployeeNumber
172    join performaceandsalary
173    on a.PSH = performaceandsalary.PercentSalaryHike
174    join `performance rating`
175    on performaceandsalary.PerformanceRating = `performance rating`.PerformanceRating
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Department | Gender | MaritalStatus | PerformanceRating |
|---|---|---|---|
| Human Resources | Female | Divorced | 3.00 |
| Human Resources | Female | Married | 3.27 |
| Human Resources | Female | Single | 3.00 |
| Human Resources | Male | Divorced | 3.25 |
| Human Resources | Male | Married | 3.04 |
| Human Resources | Male | Single | 3.25 |
| Research & Development | Female | Divorced | 3.16 |
| Research & Development | Female | Married | 3.17 |

Result 60

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 74 | 23:28:20 | select Department, Gender, avg(StockOptionLevel) as AvgStockOptionLevel, avg(... | 6 row(s) returned |
| 75 | 23:44:16 | select Department, Gender, MaritalStatus, round(avg(performaceandsalary.Performa... | 18 row(s) returned |

## Data Review for MongoDB

Assumptions/Notes About Data Collections, Attributes and Relationships between Collections

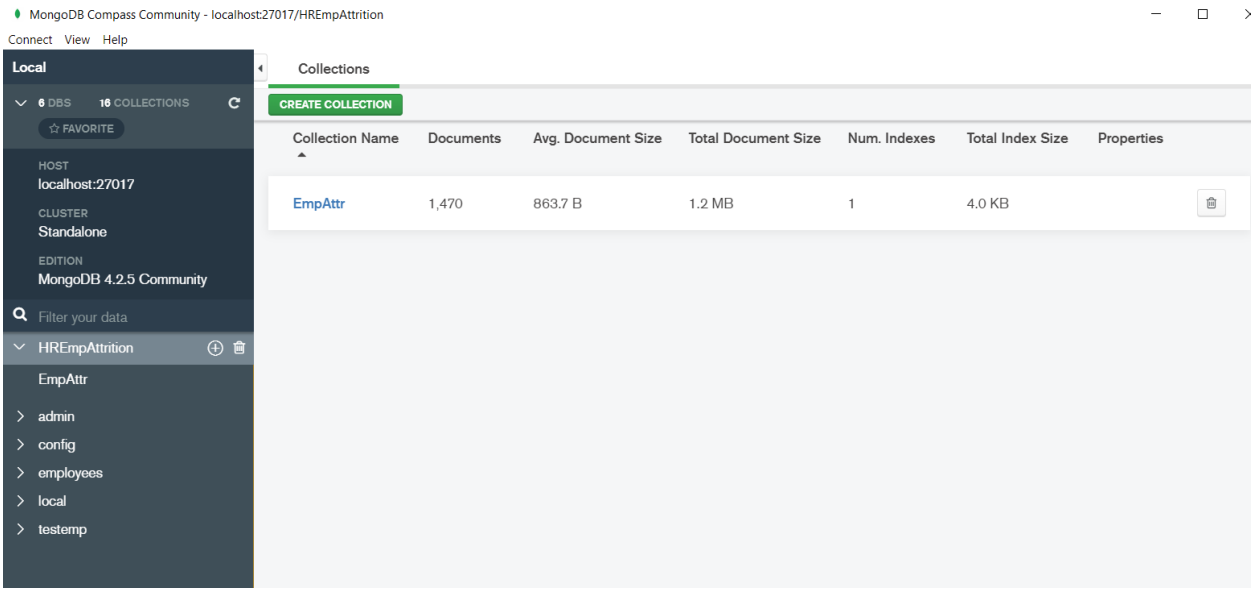
- The raw dataset is loaded into Mongo Server because it is capable of handling unstructured data too. This is now also used to compare result sets fetched from Mongo with SQL result sets
- Data is loaded into Mongo using Compass and visualized using Compass and Mongo Shell
- HREmpAttrition Database has been created into which EmpAttr collection has been created. Documents are loaded into EmpAttr collection.
- Execution of queries is done through the Mongo Shell. The results from Mongo DB and SQL are verified and validated.

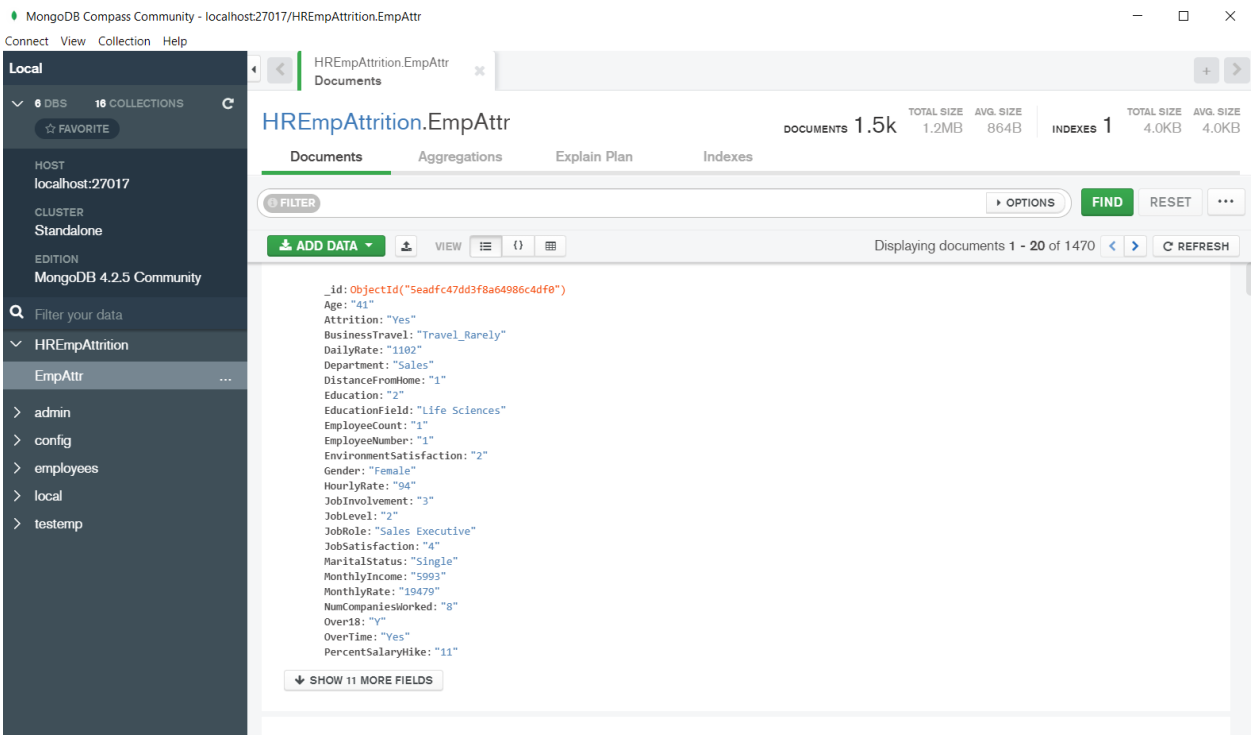
# Physical Mongo Database

## Assumptions/Notes About Data Set

- We have imported raw csv file given into MongoDB server through Mongo compass and there are no duplicate values and each record is identified by unique id '_id' column.
- Attrition implies that the particular person is no more employed
- Any person rejoining the company would have a different EmployeeNumber and his/his details would be listed as any new employee's would, i.e., YearsAtCompany and YearsWithCurrManager will be calculated with respect to this position and this EmployeeNumber. NumCompaniesWorked will also include this company.
- TrainingTimesLastYear are the number of trainings the employee has undertaken with this company in the last year
- StockOptionLevel, which is part of salary for some employees
- All employees are paid monthly salaries. The hourly rate, daily rate and monthly rate are the rates that were paid to contractors that are not on the company payroll but recruited on the same day. The contract employees are not listed here.
- Every department has specific roles in it, which do not overlap with any other department
- For employees with rating = 3 salary hike is < 20% while those with rating = 4 get a hike of ≥ 20%

## Screen shot of Physical Database objects (Database, Collections and Attributes)






## Data in the Database

| Collection Name | Relationships With Other Collections (if any) | # of Documents in Collection |
|---|---|---|
| EmpAttr | NA | 1470 |

# MongoDB Queries/Code

## Mongo Query 1

Question 2: Which department's employee is the most likely to have the longest commute between home and work?

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

1. The result set contains 3 documents
2. Maximum distance between work and home for employees in each department has been calculated using 'DistanceFromHome'
3. The departments with longest commute are "Sales" and "Research & Development"

Translation

Query the 'EmpAttr' collection in the 'HREmpAttrition' database to get the maximum distance travelled by employee for each department by grouping the employees based on department field and show the count in the result set.

Screen Shot of MongoDB Query/Code and Results

```
> db.EmpAttr.aggregate([{"$group": {"_id": "$Department", "Max Distance": {"$max": { $toInt: "$DistanceFromHome" }}}}])
{ "_id" : "Sales", "Max Distance" : 29 }
{ "_id" : "Research & Development", "Max Distance" : 29 }
{ "_id" : "Human Resources", "Max Distance" : 26 }
>
```

## Mongo Query 2

Question 5: Company has put out a publlc statement saying that they have no gender gap when it comes to compensation, in all the departments. What insight can you provide to prove or disprove that statement?

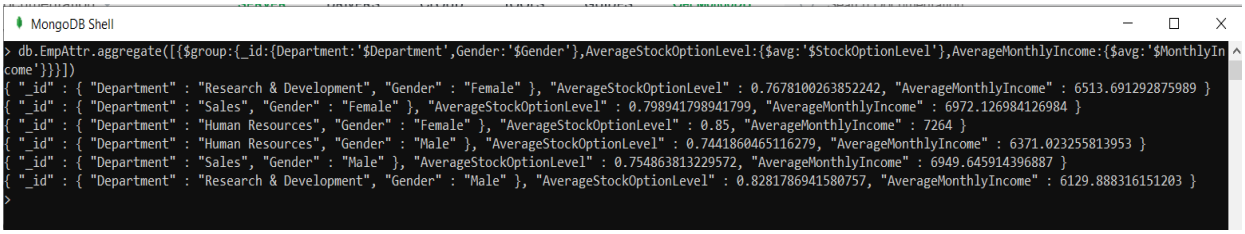
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

1. The result set contains 6 documents
2. Females receive higher monthly income and stocks than male employees in both Human Resources and Sales
3. Female employees receive more monthly income than male employees in Research & Development, but they receive less stocks essentially balancing out the monthly income in this department

Translation

Using EmpAttr collection, utilize aggregation framework to group by Department and Gender fields and find the Average Stock Option Level and Average Monthly Income for each of the Department grouped by gender

Screen Shot of MongoDB Query/Code and Results



## Mongo Query 3

Question 6: HR is trying to determine whether gender and marital status affect performance ratings of employees in each department. What initial finding can you obtain from the data to help in this regard?

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

1. The result set consists of 18 documents, each giving information on department, gender, marital status and employees' average performance rating
2. Married females and single males perform well in Human Resources
3. Performances of both, male and female employees is comparable irrespective of their marital statuses
4. Female employees perform better than male employees in Sales

Translation

Using EmpAttr collection utilize aggregation framework group the documents by multiple fields which includes Department, Gender and Marital Status and find the average performance rating of them

Screen Shot of MongoDB Query/Code and Results