

SQL-Mongo Project – Employee Attrition

BUAN 6320

Phat Vo

Contents

Relational Data Model	3
Assumptions/Notes About Data Entities and Relationships.....	3
Entity-Relationship Diagram	4
Physical MySQL Database	5
Assumptions/Notes About Data Set	5
Screen shot of Physical Database objects.....	5
Data in the Database.....	10
SQL Queries.....	11
SQL Query 1	11
Question.....	11
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	11
Translation	11
Screen Shot of SQL Query and Results.....	11
SQL Query 2	12
Question.....	12
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	12
Translation	12
Screen Shot of SQL Query and Results.....	12
SQL Query 3	13
Question.....	13
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	13
Translation	13
Screen Shot of SQL Query and Results.....	13
SQL Query 4	14
Question.....	14
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	14
Translation	14
Screen Shot of SQL Query and Results.....	14
SQL Query 5	15
Question.....	15
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	15
Translation	15

Screen Shot of SQL Query and Results.....	15
SQL Query 6	15
Question.....	16
Notes/Comments About SQL Query and Results (Include # of Rows in Result).....	16
Translation	16
Screen Shot of SQL Query and Results.....	16
Data Review for MongoDB.....	17
Assumptions/Notes About Data Collections, Attributes and Relationships between Collections	17
Physical Mongo Database	18
Assumptions/Notes About Data Set	18
Screen shot of Physical Database objects (Database, Collections and Attributes)	18
Data in the Database.....	20
MongoDB Queries/Code	20
Mongo Query 1	20
Question.....	20
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) ...	20
Translation	20
Screen Shot of MongoDB Query/Code and Results.....	20
Mongo Query 5	21
Question.....	21
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) ...	21
Translation	21
Screen Shot of MongoDB Query/Code and Results.....	21
Mongo Query 6	22
Question.....	22
Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) ...	22
Translation	22
Screen Shot of MongoDB Query/Code and Results.....	22

Relational Data Model

Assumptions/Notes About Data Entities and Relationships

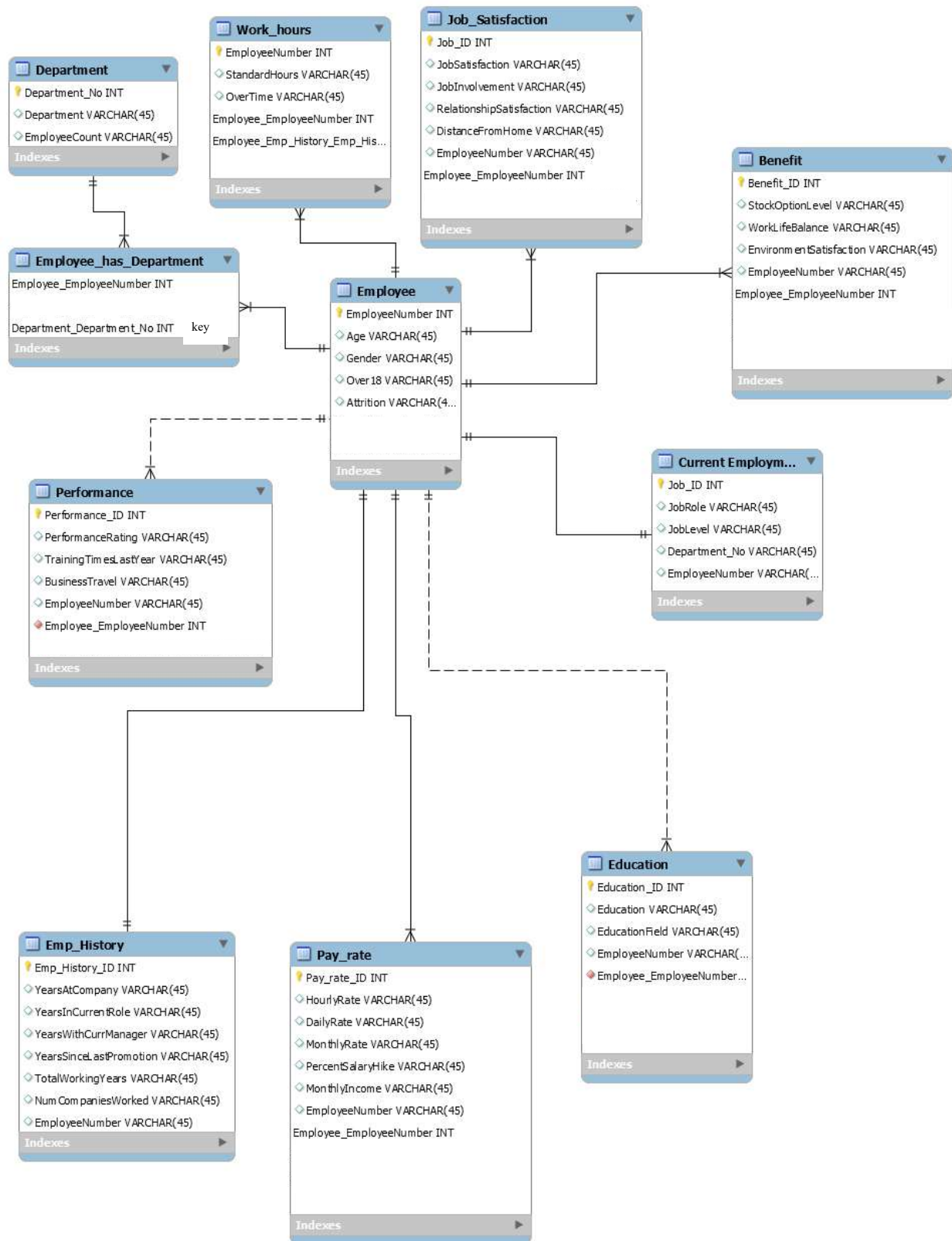
Include assumptions about data entities and their relationships with each other.

1. An employee has one or many education level and education field. Education and education field can belong to one or many employees. Education and Education field are many to many.
2. An employee can only connect to one job level and job role at a time, but job level and job role can be related to one to many employees at the same time, like teams. Job level and Job role is one to many.
3. An employee has one age and over 18, we assume 18 is legal age for employment, but some employees can have the same age. Age and Over18 are one to many.
4. An employee has one hourly rate, standard hours, overtime, daily rate, monthly rate and monthly income, but hourly rate, standard hours, overtime, daily rate, monthly rate and monthly income can be matched with one to many employees.
5. An employee belongs to one and only one department. One department can have one to many employees.
6. An employee can have one to many benefits. A Benefit can be owned by one to many employees.
7. An employee can have one to many employment histories, and employment histories belong to one and only one employee.
8. An employee only has one performance at a time, but performance can be matched with one to many employees.
9. An employee only votes one job satisfaction rates at a time, but job satisfaction rates can be voted by many employees at one time.

Include reasons why the data model is in 3NF.

- It is in atomic form, unique keys and attributes in tables are fully functional dependent
- No missing or Null values
- No multi-part or multi-valued fields

Entity-Relationship Diagram



Physical MySQL Database

Assumptions/Notes About Data Set

Include any assumptions made about data such as empty fields, sparse data, bad data, etc.

- Most data type is factor and numerical, added INT (integer) as primary keys for some tables.
- Drop Employee count since it is related to the tables

Screen shot of Physical Database objects

Employees Table

1 • `SELECT * FROM employee_attrition.employees;`

emp_no	age	gender	marital_status	over_18	attrition
1	41	Female	Single	Y	Yes
2	49	Male	Married	Y	No
4	37	Male	Single	Y	Yes
5	33	Female	Married	Y	No
7	27	Male	Married	Y	No
8	32	Male	Single	Y	No
10	59	Female	Married	Y	No
11	30	Male	Divorced	Y	No
12	38	Male	Single	Y	No

employees 1 x

Output

#	Time	Action	Message
175	13:06:14	SELECT * FROM employee_attrition.employees	1470 row(s) returned
176	13:06:26	SELECT * FROM employee_attrition.employees	1470 row(s) returned

Departments Table

1 • `SELECT * FROM employee_attrition.departments;`

dept_no	dept_name
d003	Human Resources
d002	Research & Development
d001	Sales
NULL	NULL

departments 1 x

Output

#	Time	Action	Message
176	13:06:26	SELECT * FROM employee_attrition.employees	1470 row(s) returned
177	13:07:18	SELECT * FROM employee_attrition.departments	3 row(s) returned

Dept Emp Table

1 • `SELECT * FROM employee_attrition.dept_emp;`

<

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

	department	emp_no	dept_no
▶	Sales	1	d001
	Research & Development	2	d002
	Research & Development	4	d002
	Research & Development	5	d002
	Research & Development	7	d002
	Research & Development	8	d002
	Research & Development	10	d002
	Research & Development	11	d002
	Research & Development	12	d002

dept_emp 1 x

Output

Action Output

#	Time	Action	Message
✓ 177	13:07:18	SELECT * FROM employee_attrition.departments	3 row(s) returned
✓ 178	13:07:43	SELECT * FROM employee_attrition.dept_emp	1470 row(s) returned

Education Table

1 • `SELECT * FROM employee_attrition.education;`

<

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

	edu_id	education	edu_field	emp_no
▶	1	2	Life Sciences	1
	2	1	Life Sciences	2
	3	2	Other	4
	4	4	Life Sciences	5
	5	1	Medical	7
	6	2	Life Sciences	8
	7	3	Medical	10
	8	1	Life Sciences	11
	9	3	Life Sciences	12

education 1 x

Output

Action Output

#	Time	Action	Message
✓ 178	13:07:43	SELECT * FROM employee_attrition.dept_emp	1470 row(s) returned
✓ 179	13:09:50	SELECT * FROM employee_attrition.education	1470 row(s) returned

Current employment Table

1 • `SELECT * FROM employee_attrition.current_employment;`

<

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

	job_id	job_level	job_role	dept_no	emp_no
▶	1	2	Sales Executive	d001	1
	2	2	Research Scientist	d002	2
	3	1	Laboratory Technician	d002	4
	4	1	Research Scientist	d002	5
	5	1	Laboratory Technician	d002	7
	6	1	Laboratory Technician	d002	8

employment 1 x

Output

Action Output

#	Time	Action	Message
✓ 179	13:09:50	SELECT * FROM employee_attrition.education	1470 row(s) returned
✓ 180	13:16:18	SELECT * FROM employee_attrition.current_...	1470 row(s) returned

Emp_hist Table

```
1 • SELECT * FROM employee_attrition.emp_hist;
```

emp_hist 1 x

emp_hist_id	NumCompaniesWorked	TotalWorkingYears	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager	emp_no
1	8	8	6	4	0	5	1
2	1	10	10	7	1	7	2
3	6	7	0	0	0	0	4
4	1	8	8	7	3	0	5
5	9	6	2	2	2	2	7
6	0	8	7	7	3	6	8
7	4	12	1	0	0	0	10
8	1	1	1	0	0	0	11

Output

Action Output

#	Time	Action	Message
✓ 180	13:16:18	SELECT * FROM employee_attrition.current_employment	1470 row(s) returned
✓ 181	13:17:14	SELECT * FROM employee_attrition.emp_hist	1470 row(s) returned

Pay_rate Table

```
1 • SELECT * FROM employee_attrition.pay_rate;
```

pay_rate 1 x

pay_rate_id	hourly_rate	daily_rate	monthly_rate	Percent_salary_hike	monthly_income	emp_no
1	94	1102	19479	11	5993	1
2	61	279	24907	23	5130	2
3	92	1373	2396	15	2090	4
4	56	1392	23159	11	2909	5
5	40	591	16632	12	3468	7
6	79	1005	11864	13	3068	8
7	81	1324	9964	20	2670	10
8	67	1358	13335	22	2693	11

Output

Action Output

#	Time	Action	Message
✓ 181	13:17:14	SELECT * FROM employee_attrition.emp_hist	1470 row(s) returned
✓ 182	13:17:57	SELECT * FROM employee_attrition.pay_rate	1470 row(s) returned

Work hour Table

1 • `SELECT * FROM employee_attrition.work_hour;`

<

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

	work_hour_id	overtime	standard_Hour	emp_no
▶	1	Yes	80	1
	2	No	80	2
	3	Yes	80	4
	4	Yes	80	5
	5	No	80	7
	6	No	80	8
	7	Yes	80	10
	8	No	80	11

work_hour 1 x

Output

Action Output

#	Time	Action	Message
✓ 182	13:17:57	SELECT * FROM employee_attrition.pay_rate	1470 row(s) returned
✓ 183	13:18:26	SELECT * FROM employee_attrition.work_hour	1470 row(s) returned

Performance Table

1 • `SELECT * FROM employee_attrition.performance;`

<

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

	performance_id	business_travel	performance_rating	TrainingTimeLastYear	emp_no
▶	1	Travel_Rarely	3	0	1
	2	Travel_Frequently	4	3	2
	3	Travel_Rarely	3	3	4
	4	Travel_Frequently	3	3	5
	5	Travel_Rarely	3	3	7
	6	Travel_Frequently	3	2	8
	7	Travel_Rarely	4	3	10
	8	Travel_Rarely	4	2	11

performance 1 x

Output

Action Output

#	Time	Action	Message
✓ 183	13:18:26	SELECT * FROM employee_attrition.work_hour	1470 row(s) returned
✓ 184	13:18:50	SELECT * FROM employee_attrition.performance	1470 row(s) returned

Job Satisfaction Table

1 • `SELECT * FROM employee_attrition.job_satisfaction;`

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

	job_id	DistanceFromHome	JobInvolvement	JobSatisfaction	RelationshipSatisfaction	emp_no
▶	1	1	3	4	1	1
	2	8	2	2	4	2
	3	2	2	3	2	4
	4	3	3	3	3	5
	5	2	3	2	4	7
	6	2	3	4	3	8
	7	3	4	1	1	10
	8	24	3	3	2	11

_satisfaction1 x

Output

Action Output

#	Time	Action	Message
✓ 185	13:19:07	SELECT * FROM employee_attrition.benefits	1470 row(s) returned
✓ 186	13:19:33	SELECT * FROM employee_attrition.job_satisfaction	1470 row(s) returned

Benefits Table

1 • `SELECT * FROM employee_attrition.benefits;`

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

	benefit_id	StockOptionLevel	WorkLifeBalance	EnvironmentSatisfaction	emp_no
▶	1	0	1	2	1
	2	1	3	3	2
	3	0	3	4	4
	4	0	3	4	5
	5	1	3	1	7
	6	0	2	4	8
	7	3	2	3	10
	8	1	3	4	11

benefits 1 x

Output

Action Output

#	Time	Action	Message
✓ 184	13:18:50	SELECT * FROM employee_attrition.performance	1470 row(s) returned
✓ 185	13:19:07	SELECT * FROM employee_attrition.benefits	1470 row(s) returned

Data in the Database

Table Name	Primary Key	Foreign Key	# of Rows in Table
Employees Table	emp_no	n/a	1470
Departments Table	dept_no	n/a	3
Dept_Emp Table	emp_no, dept_no	emp_no, dept_no	1470
Education Table	edu_id	emp_no	1470
Current_employment Table	job_id	emp_no, dept_no	1470
Emp_hist Table	emp_hist_id	emp_no	1470
Pay_rate Table	pay_rate_id	emp_no	1470
Work_hour Table	work_hour_id	emp_no	1470
Performance Table	performance_id	emp_no	1470
Job_Satisfaction Table	job_id	emp_no	1470
Benefits Table	benefit_id	emp_no	1470

SQL Queries

SQL Query 1

Question

Assume that the company has several branch offices around the country and employee morale is down. The company feels that not enough employees are traveling frequently between offices and that more employees should travel frequently to the branch offices to improve morale amongst each other. Is the company correct in feeling this way?

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

As we can see in the result, approximately 71% of employees (1043) rarely travel. Only 18.8% of employees are traveling frequently. In conclusion, we would recommend increasing the travel rate for employees so they can gain more skills to improve morale and the company is correct about their feeling.

Translation

(In Screen Shot)

Screen Shot of SQL Query and Results

```
1  /* Assume that the company has several branch offices around the country and employee morale
2     is down. The company feels that not enough employees are traveling frequently between offices
3     and that more employees should travel frequently to the branch offices to improve morale amongst
4     each other. Is the company correct in feeling this way?
5     Translation: Select employee number and business travel from performance table, group by business travel
6     Clean up: Select emp_no, business travel from performance, group by business travel*/
7  • Use employee_attrition;
8  • SELECT COUNT(emp_no) as 'Number of Employees', business_travel
9     FROM performance
10    GROUP BY business_travel;
```

Number of Employees	business_travel
1043	Travel_Rarely
277	Travel_Frequently
150	Non-Travel

Result 2 x

Output

#	Time	Action	Message
✓ 195	19:52:29	Use employee_attrition	0 row(s) affected
✓ 196	19:52:39	SELECT COUNT(emp_no) as 'Number of Employees', business_travel FROM p...	3 row(s) returned

SQL Query 2

Question

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

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

We select all employees who have the min (distance from home) and group by department name with the count of number of employees in each department so we can see the real picture. Therefore, we performed two tasks (two queries) to gain the insight.

Translation

(In Screen Shot)

Screen Shot of SQL Query and Results

- a. Number of employees in each department that has minimal distance form home:

```
1  /* Which department's employee is the most likely to have the shortest
2  commute between home and work?
3  Translation: Select department from dept_emp table, distancefromhome from job_satisfaction table,
4  count number of employees in each department by selecting employeeenumber from dept_emp table,
5  join employeeenumber from dept_emp table, where distancefromhome is 1 and group by department
6
7  Clean up: Select department from dept_emp, distancefromhome from job_satisfaction, COUNT employeeenumber
8  from dept_emp as number of employees, and left join employeeenumber from dept_emp, where distancefromhome = 1,
9  group by department
10 */
11 USE EmployeeAttritionProj;
12 SELECT dept_emp.Department, job_satisfaction.DistanceFromHome, COUNT(dept_emp.EmployeeNumber)
13 as 'Number of Employees'
14 FROM dept_emp
15 JOIN job_satisfaction ON dept_emp.EmployeeNumber = job_satisfaction.EmployeeNumber
16 WHERE DistanceFromHome = 1
17 GROUP BY department;
```

Department	DistanceFromHome	Number of Employees
Sales	1	51
Research & Development	1	147
Human Resources	1	10

- b. List of employees with minimum distance from home:

```
1  /* Which department's employee is the most likely to have the shortest commute between home and work?
2  Translation: Select department from dept_emp table, distancefromhome from job_satisfaction table,
3  count number of employees in each department by selecting employeeenumber from dept_emp table,
4  join employeeenumber from dept_emp table, where distancefromhome is 1 and group by department
5  Clean up: Select department from dept_emp, distancefromhome from job_satisfaction, COUNT employeeenumber
6  from dept_emp as number of employees, and left join employeeenumber from dept_emp, where distancefromhome = 1,
7  group by department*/
8  SELECT dept_emp.Department, job_satisfaction.DistanceFromHome, job_satisfaction.emp_no
9  as 'Employee Number'
10 FROM dept_emp
11 JOIN job_satisfaction ON dept_emp.emp_no = job_satisfaction.emp_no
12 WHERE DistanceFromHome = (select min(DistanceFromHome) from job_satisfaction);
```

Department	DistanceFromHome	Employee Number
Sales	1	1
Research & Development	1	45
Sales	1	52
Research & Development	1	57
Research & Development	1	64
Sales	1	68
Research & Development	1	73
Research & Development	1	77

Result 2 x

Output

Action Output

#	Time	Action	Message	Duration / Fetc
206	20:11:38	SELECT employees.marital_status, AVG(performance performance_rating) as 'A...	3 row(s) returned	0.015 sec / 0.0
207	16:09:06	SELECT dept_emp.Department, job_satisfaction.DistanceFromHome, job_satisf...	208 row(s) returned	0.078 sec / 0.0

SQL Query 3

Question

A new employee from a Medical-related education field 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)

As we can see, there are employees in Sales department that have background in medical or medical related. This new employee is a perfect candidate to apply into Sales department.

Translation

(In Screen Shot)

Screen Shot of SQL Query and Results

```
1  /* A new employee from a Medical-related education field wants to work in Sales.
2  Do you believe the company might be able to give her a chance to work in Sales? Why or Why not?
3  Translation: Select educationfield and employee number from education table, jobrole from current_employment table,
4  and join employeenumber from education table, where jobrole is either sales representative or sales executive, and
5  education field is medical
6  Clean up: Select educationfield from Education, employeenumber from Education, jobrole from current_employment,
7  join employeenumber from education, where jobrole = sales representative or sales executive,
8  and educationfield = medical */

9  • Use employee_attrition;
10 • SELECT Education.edu_Field, Education.emp_no, current_employment.job_role
11 FROM Education
12 JOIN Current_Employment ON Education.emp_no = current_employment.emp_no
13 WHERE (job_role = "Sales Representative" OR job_role = "Sales Executive") AND edu_Field = "Medical";
```

edu_Field	emp_no	job_role
Medical	118	Sales Executive
Medical	121	Sales Executive
Medical	125	Sales Executive
Medical	129	Sales Executive
Medical	131	Sales Executive
Medical	195	Sales Executive
Medical	291	Sales Executive
Medical	302	Sales Executive
Medical	312	Sales Executive
Medical	366	Sales Executive
Medical	378	Sales Executive

Result 1 x

Output

Action Output

#	Time	Action	Message
202	20:00:15	SELECT Education.edu_Field, Education.emp_no, current_employment.job_rol...	Error Code: 1054. Unknown column 'JobRole' in 'where clause'
203	20:00:39	SELECT Education.edu_Field, Education.emp_no, current_employment.job_rol...	83 row(s) returned

SQL Query 4

Question

The HR 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)

In this example, since our data are ordinal, it is not accurate to take average of the data. So, we filter out all high rating (3 or above) and calculate the percentage to tell if the department has higher distribution of high rating comparing to others. As we can see HR department has lowest percent of employees that have 3 or higher job satisfaction rate, 2.18% of 1470 employees. On the other side, Research and Development has highest percentage of employee's Environment satisfaction rate, 40.6% of 1470 employees.

Translation

(In Screen Shot)

Screen Shot of SQL Query and Results

```
1  /* Translation: we select department name from department table. We calculate the percentage of employee who has 3 or greater
2  job satisfaction and environment satisfaction for each department by grouping them.
3  clean up: we select department, count emp.no then divide by 1470 and multiply 100 by joining job_satisfaction and dept_emp. Then we group by department*/
4
5  • SELECT DISTINCT dept_emp.department as 'Department', ((count(job_satisfaction.JobSatisfaction)/1470)*100) as '%Employee for Job Satisfaction'
6  from job_satisfaction
7  JOIN dept_emp on dept_emp.emp_no = job_satisfaction.emp_no
8  WHERE job_satisfaction.JobSatisfaction >=3
9  group by department;
```

Result Grid

Department	%Employee for Job Satisfaction
Sales	18.6395
Research & Development	40.4762
Human Resources	2.1769

```
16
17 • SELECT DISTINCT dept_emp.department as 'Department', ((count(job_satisfaction.emp_no)/1470)*100) as '%Employee for Environment Satisfaction' from job_satisfaction
18 JOIN dept_emp on dept_emp.emp_no = job_satisfaction.emp_no
19 where job_satisfaction.EnvironmentSatisfaction >=3
20 group by department;
```

Result Grid

Department	%Employee for Environment Satisfaction
Research & Development	40.6122
Sales	17.8231
Human Resources	2.7211

SQL Query 5

Question

An employee in Sales department has complained to HR saying that females are paid less than males in the company, in all departments. What insight can you provide to prove or disprove that statement?

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

As we can see the average monthly income for female is higher than male employees in all departments. However, if we just base it on our data, the result is insufficient to answer this question. We need to test to see the significant difference between two groups as well as consider the differences in job position's pay rate, number of hours worked. We can also look at the monthly rate which also show female has higher average pay. In conclusion, the statement from the employee in Sales department is incorrect.

Translation

(In Screen Shot)

Screen Shot of SQL Query and Results

```
1  /* An employee in Sales department has complained to HR saying that females are paid less than males in the company
2  , in all departments. What insight can you provide to prove or disprove that statement?
3  Translation: Select distinct gender, employee number from employees table, average monthly income from pay_rate table,
4  join employee number from dept_emp table and employees table, join employee number from employees table and pay_rate table
5  Clean up: Select distinct gender, employee number from employees, avg monthly income from pay_rate,
6  join employee number from dept_emp and employees, join employee number from employees and pay_rate */
~
6  •  Select DISTINCT employees.gender,
7     AVG(pay_rate.monthly_rate), avg(pay_rate.monthly_income) as 'Average Monthly Income'
8  FROM employees
9     JOIN dept_emp ON employees.emp_no = dept_emp.emp_no
10    JOIN pay_rate ON employees.emp_no = pay_rate.emp_no
11    GROUP BY employees.gender;
12
```

<

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

	gender	AVG(pay_rate.monthly_rate)	Average Monthly Income
▶	Female	14674.6003	6686.5663
	Male	14072.1054	6380.5079

SQL Query 6

Question

A press article in a business magazine has said that at this company, married men have higher performance ratings than divorced or single men. What initial finding can you obtain from the data to help articulate the company's response in this regard?

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

From the result, we can say that married men have slightly lower average performance rating (3.1471) compared to single men (3.1624). Divorced men have the lowest average performance rating (3.1381).

Translation

We selected marital status and took average of the performance rating. Then, we grouped them by marital status to tell which group has the highest.

Screen Shot of SQL Query and Results

The screenshot shows a SQL query editor with a multi-line comment and a SQL query. The comment describes the context of the query: a press article about performance ratings by marital status, and the goal is to find initial findings. The SQL query selects the average performance rating by marital status for male employees.

```
1  /* A press article in a business magazine has said that at this company, married men have higher performance
2  ratings than divorced or single men. What initial finding can you obtain from the data to help articulate
3  the company's response in this regard?
4  Translation: Select marital status from employees table, average performance rating from performance table,
5  join employee number from employees table and performance table, group by marital status
6  Clean up: Select marital status from employees, avg performance rating from performance, join employee number
7  from employees and performance, group by marital status */
8
9  Use EmployeeAttritionProj;
10 SELECT employees.marital_status, AVG(performance.performance_rating) as 'Average Performance Rating'
11 FROM employees
12 JOIN performance ON employees.emp_no = performance.emp_no
13 WHERE gender = 'male'
14 GROUP BY marital_status;
```

Below the query editor, the results are displayed in a table:

marital_status	Average Performance Rating
Married	3.1471
Single	3.1624
Divorced	3.1381

At the bottom, the 'Action Output' pane shows the execution of the query:

#	Time	Action	Message
205	20:04:01	Select DISTINCT employees.gender, COUNT(employees.emp_no) as 'Number...	2 row(s) returned
206	20:11:38	SELECT employees.marital_status, AVG(performance.performance_rating) as 'A...	3 row(s) returned

Data Review for MongoDB

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

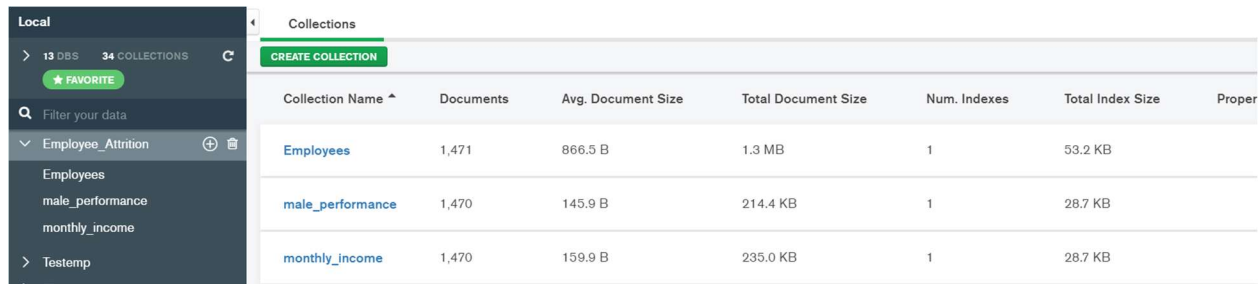
- As mongo dB is not a relational database, we decided to group all data into one collection which can help us reduce the joining task and make our queries become more dynamic and simpler to understand.
- Department number column (in MySQL database) does not exist in mongo database since the dataset does not have it and department number is irrelevant in the question scenario. Thus, we eliminated it.
- Even though it shows 1471 in “Displaying documents”, there are only 1470 records of employee and 1 come from the header in our csv file. This one does not affect the accuracy of our result in queries.

Physical Mongo Database

Assumptions/Notes About Data Set

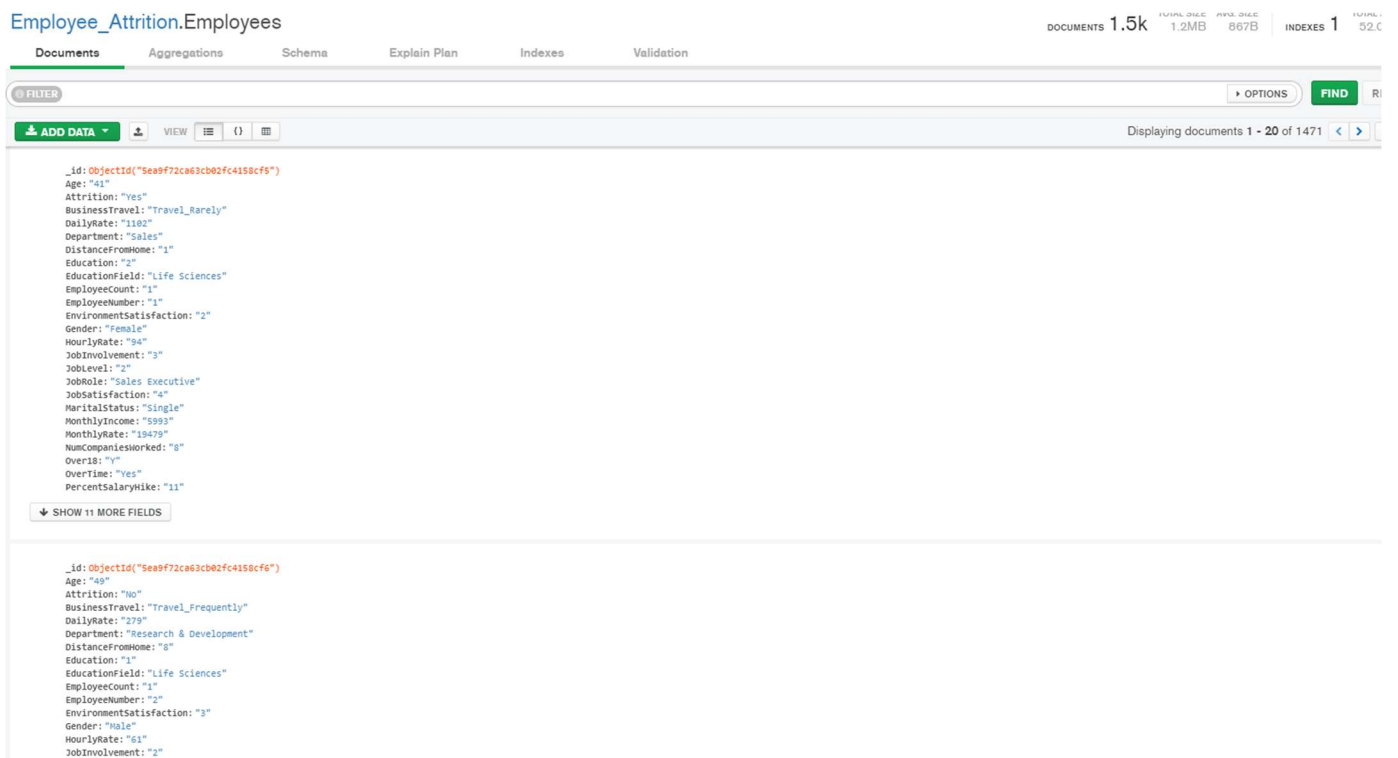
- No null record
- Collection employee has 1 null document from headers

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



Collection Name ^	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size	Proper
Employees	1,471	866.5 B	1.3 MB	1	53.2 KB	
male_performance	1,470	145.9 B	214.4 KB	1	28.7 KB	
monthly_income	1,470	159.9 B	235.0 KB	1	28.7 KB	

Employee collection



Employee_Attrition.Employees

DOCUMENTS 1.5k TOTAL SIZE 1.2MB INDEXES 1

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER

ADD DATA VIEW

Displaying documents 1 - 20 of 1471

```
{
  "_id": "ObjectId('5ea9f72ca63cb02fc4158cf5')",
  "Age": 41,
  "Attrition": "Yes",
  "BusinessTravel": "Travel_Rarely",
  "DailyRate": 1180,
  "Department": "Sales",
  "DistanceFromHome": 1,
  "Education": 2,
  "EducationField": "Life Sciences",
  "EmployeeCount": 1,
  "EmployeeNumber": 1,
  "EnvironmentSatisfaction": 2,
  "Gender": "Female",
  "HourlyRate": 94,
  "JobInvolvement": 3,
  "JobLevel": 2,
  "JobRole": "Sales Executive",
  "JobSatisfaction": 4,
  "MaritalStatus": "Single",
  "MonthlyIncome": 5993,
  "MonthlyRate": 19479,
  "NumCompaniesWorked": 8,
  "Over18": 1,
  "OverTime": "Yes",
  "PercentSalaryHike": 11
}
```

SHOW 11 MORE FIELDS

```
{
  "_id": "ObjectId('5ea9f72ca63cb02fc4158cf6')",
  "Age": 49,
  "Attrition": "No",
  "BusinessTravel": "Travel_Frequently",
  "DailyRate": 1279,
  "Department": "Research & Development",
  "DistanceFromHome": 8,
  "Education": 1,
  "EducationField": "Life Sciences",
  "EmployeeCount": 1,
  "EmployeeNumber": 2,
  "EnvironmentSatisfaction": 3,
  "Gender": "Male",
  "HourlyRate": 61,
  "JobInvolvement": 2
}
```

Male performance collection

Employee_Attrition.male_performance

DOCUMENTS 1.5k TOTAL SIZE 209.4KB AVG. SIZE 146B INDEXES 1

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER **ADD DATA** **VIEW** **OPTIONS** **FIN**

Displaying documents 1 - 20 of 1470

```
{
  "_id": ObjectId("5ea9f99aa63cb02fc41592b4"),
  "emp_no": 1,
  "age": 41,
  "gender": "Female",
  "marital_status": "Single",
  "over_18": "Y",
  "attrition": "Yes",
  "performance_rating": 3
}
```

```
{
  "_id": ObjectId("5ea9f99aa63cb02fc41592b5"),
  "emp_no": 2,
  "age": 49,
  "gender": "Male",
  "marital_status": "Married",
  "over_18": "Y",
  "attrition": "No",
  "performance_rating": 4
}
```

```
{
  "_id": ObjectId("5ea9f99aa63cb02fc41592b6"),
  "emp_no": 4,
  "age": 37,
  "gender": "Male",
  "marital_status": "Single",
  "over_18": "Y",
  "attrition": "Yes",
  "performance_rating": 3
}
```

```
{
  "_id": ObjectId("5ea9f99aa63cb02fc41592b7"),
  "emp_no": 5,
  "age": 33,
  "gender": "Female",
  "marital_status": "Married",
  "over_18": "Y",
  "attrition": "No",
  "performance_rating": 3
}
```

Monthly income collection

Employee_Attrition.monthly_income

DOCUMENTS 1.5k TOTAL SIZE 229.5KB AVG. SIZE 160B INDEXES 1

Documents Aggregations Schema Explain Plan Indexes Validation

FILTER **ADD DATA** **VIEW** **OPTIONS** **FIN**

Displaying documents 1 - 20 of 1470

```
{
  "_id": ObjectId("5ea9ff32a63cb02fc4159e30"),
  "emp_no": 1,
  "age": 41,
  "gender": "Female",
  "marital_status": "Single",
  "over_18": "Y",
  "attrition": "Yes",
  "monthly_income": 5993,
  "monthly_rate": 19479
}
```

```
{
  "_id": ObjectId("5ea9ff32a63cb02fc4159e31"),
  "emp_no": 2,
  "age": 49,
  "gender": "Male",
  "marital_status": "Married",
  "over_18": "Y",
  "attrition": "No",
  "monthly_income": 5130,
  "monthly_rate": 24907
}
```

```
{
  "_id": ObjectId("5ea9ff32a63cb02fc4159e32"),
  "emp_no": 4,
  "age": 37,
  "gender": "Male",
  "marital_status": "Single",
  "over_18": "Y",
  "attrition": "Yes",
  "monthly_income": 2090,
  "monthly_rate": 2396
}
```

```
{
  "_id": ObjectId("5ea9ff32a63cb02fc4159e33"),
  "emp_no": 5,
  "age": 33,
  "gender": "Female",
  "marital_status": "Married",
  "over_18": "Y",
  "attrition": "No",
  "monthly_income": 2909
}
```

Data in the Database

Collection Name	Relationships With Other Collections (if any)	# of Documents in Collection
Employees	na	1470 (not header)
Male_performance	Contain some Employees' variables	1470
Monthly_income	Contain some Employees' variables	1470

MongoDB Queries/Code

Mongo Query 1

Question

Assume that the company has several branch offices around the country and employee morale is down. The company feels that not enough employees are traveling frequently between offices and that more employees should travel frequently to the branch offices to improve morale amongst each other. Is the company correct in feeling this way?

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

From the result, we see the highest (1043) are employees who rarely travel. There are also 150 employees who does not travel which take part 10%. This matches with the assumption of the company that employees should be traveling more to improve morale.

Translation

We performed aggregate and group count employee into group by frequency of traveling.

Screen Shot of MongoDB Query/Code and Results

```
MongoDB Enterprise Cluster0-shard-0:PRIMARY> db.Performance.aggregate(  
...  
...  
... $group:  
...  
... _id : "$BusinessTravel",  
... count:{$sum:1}}})  
...  
... "_id" : "Travel_Rarely", "count" : 1043 }  
... "_id" : "Travel_Frequently", "count" : 277 }  
... "_id" : "Non-Travel", "count" : 150 }
```

Mongo Query 5

Question

An employee in Sales department has complained to HR saying that females are paid less than males in the company, in all 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)

In the result, we have average monthly income for women and men. We can see that women have higher monthly income despite the claim that women are paid less than men in all departments. However, the monthly income may not reflect the accurate picture since women may work more hours or the payrates that are different among the distribution between number of men or women in high pay position. That is why we also calculate the average of monthly rate to compare and female still have higher monthly rate.

Translation

Put employees into gender groups and calculate average value of monthly income and monthly rate to compare among groups.

Screen Shot of MongoDB Query/Code and Results

```
MongoDB Enterprise Cluster0-shard-0:PRIMARY> db.monthly_income.aggregate(
... [ {$group:{
...   _id : "$gender",
...   avgMonthlyIncome:{$avg:"$monthly_income"},
...   avgMonthlyRate:{$avg:"$monthly_rate"}}}]
{ "_id" : "Female", "avgMonthlyIncome" : 6686.566326530612, "avgMonthlyRate" : 14674.600340136054 }
{ "_id" : "Male", "avgMonthlyIncome" : 6380.507936507936, "avgMonthlyRate" : 14072.105442176871 }
MongoDB Enterprise Cluster0-shard-0:PRIMARY>
```

Mongo Query 6

Question

A press article in a business magazine has said that at this company, married men have higher performance ratings than divorced or single men. What initial finding can you obtain from the data to help articulate the company's response in this regard?

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

From the results, we see that single men have the highest average performance rating (3.16) among three type of marital status. We can also see that married men have a higher average performance rating compared to divorced men, but the differences are small. We need to test for the significance of the difference. But to conclude about the claim on the press article, we say that Married men have a higher performance rating than divorced men, but single men have the highest.

Translation

Filter gender in the collection to men and group them base on their marital status. Then we calculate the average of performance rating to compare between groups.

Screen Shot of MongoDB Query/Code and Results

```
MongoDB Enterprise Cluster0-shard-0:PRIMARY> db.male_performance.aggregate(
...   [ {$match:{gender:'Male'}},
...   {
...     $group:
...     {
...       _id : "$marital_status",
...       avgAmount:{$avg:"$performance_rating"}}
...   }]);
...   { "_id" : "Divorced", "avgAmount" : 3.138095238095238 }
...   { "_id" : "Single", "avgAmount" : 3.162361623616236 }
...   { "_id" : "Married", "avgAmount" : 3.14713216957606 }
```