**Group 8**

Abhishek Rai

Jack Dayton

Himanshu Jagtap

**SQL-Mongo Project**

**Employee Satisfaction Analysis**

BUAN 6320

|  |  |  |  |
| --- | --- | --- | --- |
| **Activity** | **Abhishek** | **Himanshu** | **Jack** |
| Prepared Data Model and Created Physical DB | ✔ | ✔ | ✔ |
| Loaded Data into Database | ✔ | ✔ | ✔ |
| Wrote SQL Queries | ✔ | ✔ | ✔ |
| Prepared Mongo Database | ✔ | ✔ | ✔ |
| Loaded data into Mongo DB | ✔ | ✔ | ✔ |
| Wrote Mongo Queries | ✔ | ✔ | ✔ |
| Prepared Report | ✔ | ✔ | ✔ |
| Reviewed Report | ✔ | ✔ | ✔ |

Contents

Relational Data Model 4

Assumptions/Notes About Data Entities and Relationships 4

Entity-Relationship Diagram 6

Physical MySQL Database 7

Assumptions/Notes About Data Set 7

Screen shot of Physical Database objects 7

Data in the Database 14

SQL Queries 15

SQL Query 1 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 2 16

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

SQL Query 3 17

Question 17

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 17

Translation 17

Screen Shot of SQL Query and Results 17

SQL Query 4 18

Question 18

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 18

Translation 18

Screen Shot of SQL Query and Results 18

SQL Query 5 19

Question 19

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 19

Translation 19

Screen Shot of SQL Query and Results 19

SQL Query 6 20

Question 20

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 20

Translation 20

Screen Shot of SQL Query and Results 21

Data Review for MongoDB 22

Assumptions/Notes About Data Collections, Attributes and Relationships between Collections 22

Physical Mongo Database 23

Assumptions/Notes About Data Set 23

Screen shot of Physical Database objects (Database, Collections and Attributes) 23

Data in the Database 27

MongoDB Queries/Code 28

Mongo Query 1 28

Question 28

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 28

Translation 28

Screen Shot of MongoDB Query/Code and Results 28

Mongo Query 2 29

Question 29

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 29

Translation 29

Screen Shot of MongoDB Query/Code and Results 29

Mongo Query 3 30

Question 30

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 30

Translation 30

Screen Shot of MongoDB Query/Code and Results 31

# Relational Data Model

## Assumptions/Notes About Data Entities and Relationships

We have made the **Employee\_Details** table the main table in our model because all of the other tables should relate back to an individual employee and thus the EmployeeNumber, which is the primary key for this table. Each other table in the model relates to the **Employee\_Details** table, so the primary key for each other table is also in the **Employee\_Details** table as the foreign keys. Every relationship between tables is also 1:1, as each entry is one employee who has one job and who takes one survey.

* **The model is 1NF** because all rows have the same number of columns and all values are singular.
* **The model is 2NF** because each column is fully functionally dependent on the primary key for that table. We created individual ‘ID’ primary keys for each table that all relate to the **Employee\_Details** table.
* **The model is 3NF** because there is no transitive dependency for non-prime attributes as well as it is in the second normal form. Each column of the table depends on its primary key only.

We have created the primary keys columns for all the tables except for the **Employee\_Details** table in which EmployeeNumber is the primary key column.

Assumptions:

1. We have assumed that each Employee has Age over 18 and this is not going to change in the future.
2. We have assumed that each employee has Standard working hours equal to 80 hrs which is not going to change in the future.
3. We have assumed that employee Count remains 1 for each employee.
4. We have also assumed that Each employee has its own daily and monthly Rate other than Monthly Income and these rates are used to bill clients of the company providing service to them.
5. We also assume that Distance from home is the distinct relative quantity and higher the value means farther is the distance from home.
6. We are also assuming that Education value denotes the level of education. This means that higher the value of an education higher the employee education is.
7. We assume that one employee will be from one education field only.
8. We assume that StockOptionLevel for each employee denotes the amount of stock owned by each employee on a relative scale. E.g. An employee with StockOptionLevel of 3 owns bigger amount of stocks compared with that with a StockOptionLevel of 1.
9. We assume Overtime attributes give the idea that employee has done overtime for that month meaning worked more than standard hours.
10. We assume that hourly rate is used to calculate the overtime income with whatever hours of overtime is done by an employee.
11. We are assuming for each department that there are multiple Job roles, and for each role, they have multiple Job levels. Job role defines the work of the person whereas Job level denotes that employee’ position in the hierarchy structure of that role.
12. We assumed that one employee did not take multiple surveys.
13. We assumed that one person was not working multiple different jobs and thus worked for two different companies or took surveys for each job.
14. We assumed that there were not multiple entries for the same employees.

## Entity-Relationship Diagram

A screenshot of a computer

Description automatically generated

# Physical MySQL Database

## Assumptions/Notes About Data Set

There are no empty fields or sparse data in the data set.

All employees are above age of 18 and have 80 Working hours.

## Screen shot of Physical Database objects

A screenshot of a cell phone

Description automatically generated

**Employee\_Details tableA screenshot of a cell phone

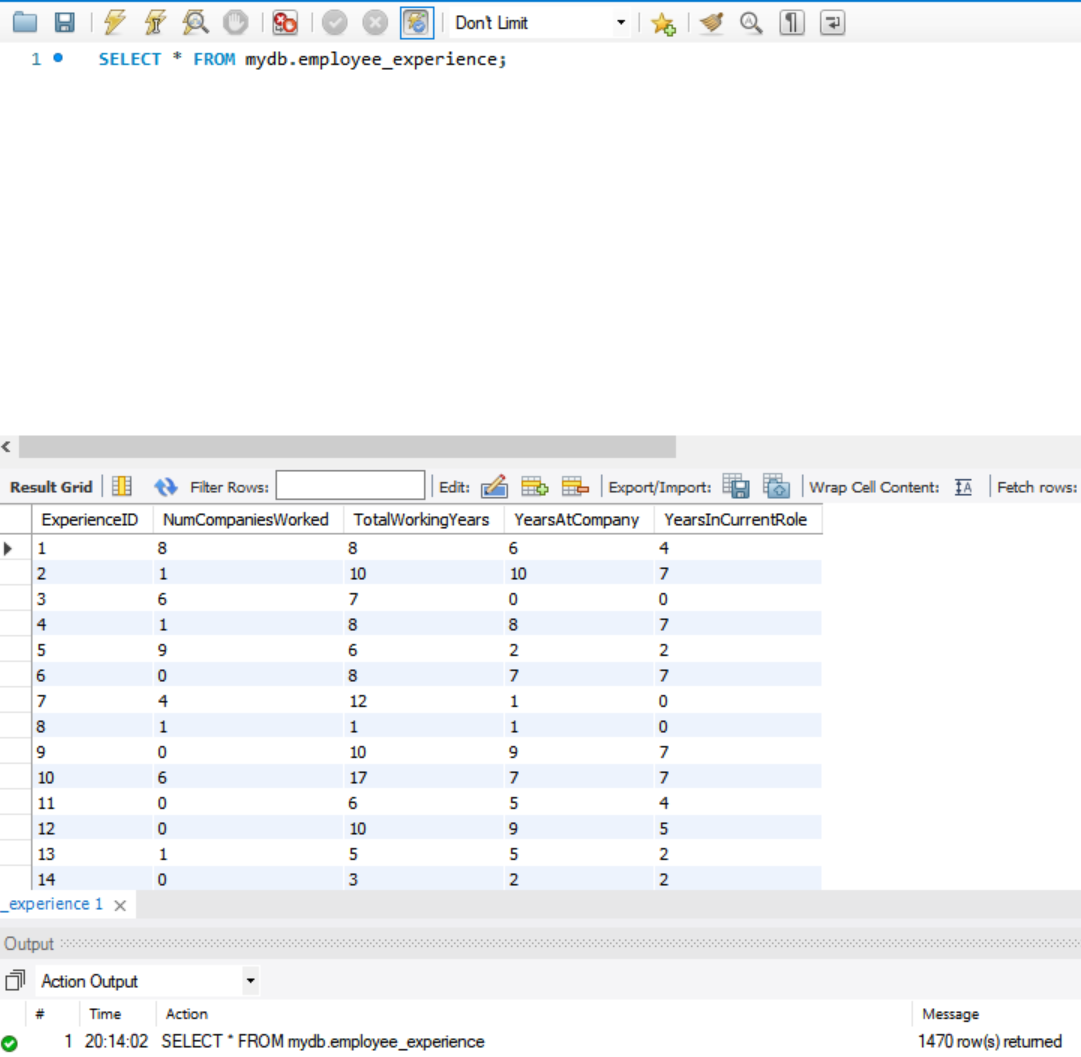
Description automatically generated**

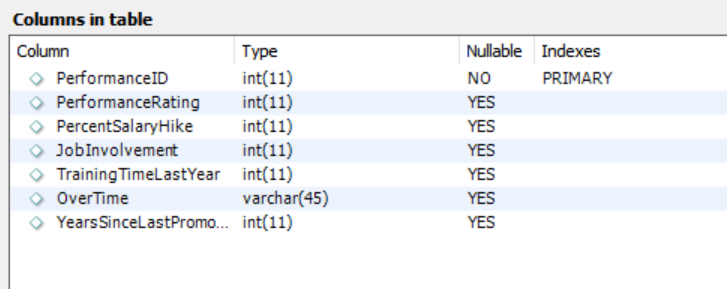
**A screenshot of a social media post

Description automatically generated**

**Employee\_Experience Table**A screenshot of a cell phone

Description automatically generated

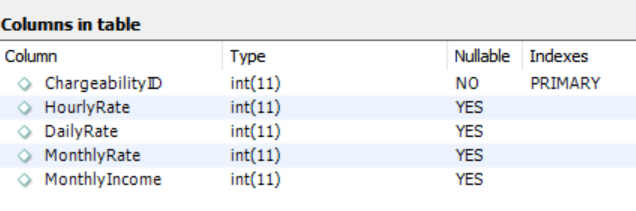


**Employee\_Performance Table**

**A screenshot of a social media post

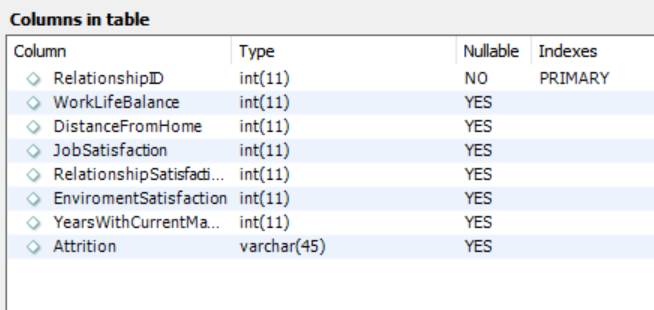
Description automatically generated**

**Employee\_Chargeability Table**



**A screenshot of a social media post

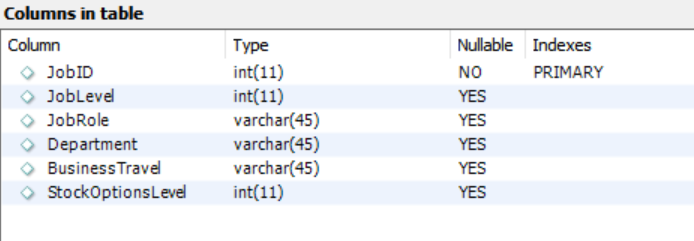
Description automatically generated**

**Employee\_Relationship Table**

A screenshot of a social media post

Description automatically generated

**Job\_Details table**



A screenshot of a social media post

Description automatically generated

## Data in the Database

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** | **# of Rows in Table** |
| Job\_details | JobID | No Foreign Key | 1470 |
| Employee\_Details | EmployeeNumber | 1. employee\_Experience\_ExperienceID 2. employee\_Performance\_PerformanceID 3. Employee\_chargeability\_ChargeabilityID 4. employee\_Relationship\_RelatioshipID 5. Job\_Details\_JobID | 1470 |
| employee\_performance | PerformanceID | No Foreign Key | 1470 |
| employee\_chargeability | ChargeabilityID | No Foreign Key | 1470 |
| employee\_experience | ExperienceID | No Foreign Key | 1470 |
| employee\_relationship | RelationshipID | No Foreign Key | 1470 |

# SQL Queries

## SQL Query 1

### Question

A new research scientist who loves to travel joins the firm and is told by HR that his job role is one of the top two roles in terms of employees that travel frequently. Is HR right in saying so? Why or Why not?

### Insights About SQL Query and Results (Include # of Rows in Result)

According to the data, HR is right in saying so, as we see that Sales Executive and Research Scientist are the top two roles in terms of employees that travel frequently.

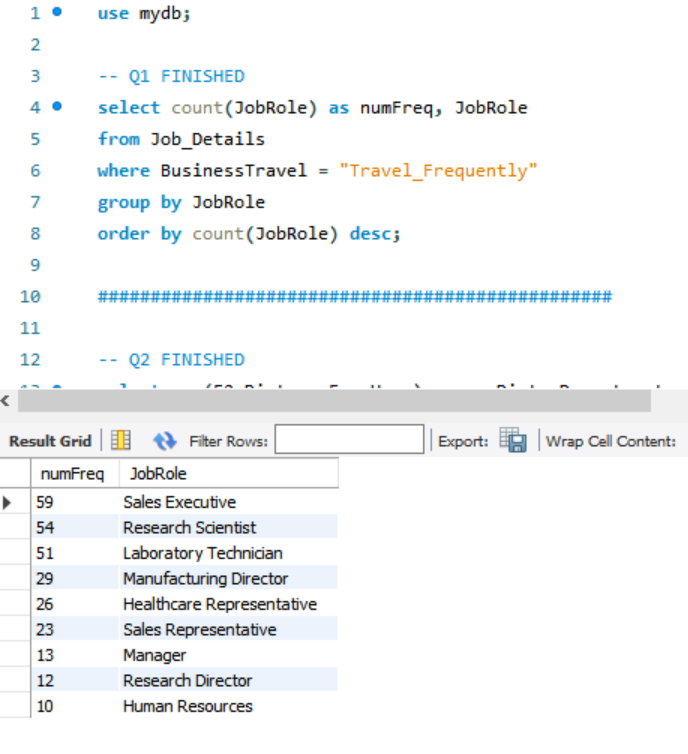
# of Rows = 9.

### Translation

Select the count ofJob Role and Job Role from Job Details table where Business Travel is equal to = “Travel\_Frequenctly”, group by JobRole and order by the count of Job Roles in a descending order.

CLEANUP: Select ~~the~~ count ~~of~~JobRole as numFreq ~~and~~ JobRole from Job\_Details ~~table~~ where BusinessTravel ~~is~~ ~~equal to~~ = “Travel\_Frequently”, group by JobRole ~~and~~ order by ~~the~~ count(JobRole) ~~in a~~ desc~~ending order.~~

### Screen Shot of SQL Query and Results



## SQL Query 2

### Question

The company has been paying gas expenses for miles traveled by employees between their home and work. If they want to increase the per mile compensation, which department's employees will gain the least?

### Insights About SQL Query and Results (Include # of Rows in Result)

Total distance traveled to work from home is least for the employees of the Human Resources Department hence they will receive the least gain by the increase of compensation.

# of Rows = 3.

### Translation

Select the sum of Distance from Home, and Department from Job Details table joined with Employee Details table on JobID matching Job\_Details\_JobID, then joined with employee Relationship table on RelationshipID matching employee\_Relationship\_RelationshipID, grouped by Department and then ordered by the total of distance from home.

CLEANUP: Select ~~the~~ sum ~~of~~ DistanceFromHome as TotalDist, ~~and~~ Department from Job\_Details ~~table~~ join~~ed~~ ~~with~~ Employee\_Details ~~table~~ on JobID ~~matching~~ Job\_Details\_JobID, ~~then~~ join~~ed~~ ~~with~~ employee\_Relationship ~~table~~ on RelationshipID ~~matching~~ employee\_Relationship\_RelationshipID, group~~ed~~ by Department ~~and then~~ order~~ed~~ by TotalDist.

### Screen Shot of SQL Query and Results

A screenshot of a social media post

Description automatically generated

## SQL Query 3

### Question

A new employee with a Marketing degree wants to work in HR. Do you believe the company might be able to give him a chance to work in HR? Why or Why not?

### Insights About SQL Query and Results (Include # of Rows in Result)

The company won't give the new employee with a Marketing degree a chance to work in HR because no previous HR employees have an Education Field in Marketing.

# of Rows in Result = 1

### Translation

Select count of EmployeeNumber from Job Details table joined with Employee Details table on Job ID matched with Job\_Details\_JobID where Education Field is equal to = “Marketing” and Department in Job Details table is equal to = “Human Resources”

CLEANUP: Select count ~~of~~ EmployeeNumber as employeeCount from Job\_Details ~~table~~ join~~ed with~~ Employee\_Details ~~table~~ on Job\_ID ~~matched with~~ Job\_Details\_JobID where EducationField ~~is equal to~~ = “Marketing” and Department ~~in Job\_Details table is equal to~~ = “Human Resources”

### Screen Shot of SQL Query and Results

A screenshot of a cell phone

Description automatically generated

## SQL Query 4

### Question

Sales feels that their environment satisfaction score is higher than HR but HR job satisfaction score is lower than Research & Development. Are they right?

### Insights About SQL Query and Results (Include # of Rows in Result)

We have compared the average Job Satisfaction & Environment Satisfaction scores of each department and we conclude that Sales is WRONG in saying that its Environment Satisfaction score is higher than HR, but it is RIGHT in saying that the Job Satisfaction score of HR is less than R&D.

# of Rows in Result = 3

### Translation

Select average of Job Satisfaction, average of Environment Satisfaction, and department from Employee Relationship table joined with employee details table on relationship ID matched with employee\_Relationship\_RelationshipID, and then joined with job details table on job ID matched with Job\_Details\_JobID, grouped by Department

CLEANUP: Select avg ~~of~~ JobSatisfaction as avgJobSatisfaction, avg ~~of~~ EnvironmentSatisfaction as avgEnvSatisfaction, ~~and~~ department from Employee\_Relationship ~~table~~ join~~ed~~ ~~with~~ employee\_details table on relationshipID ~~matched with~~ employee\_Relationship\_RelatioshipID, ~~and then~~ join~~ed~~ ~~with~~ job\_details ~~table~~ on jobID ~~matched with~~ Job\_Details\_JobID, group~~ed~~ by Department

### Screen Shot of SQL Query and Results

A screenshot of a social media post

Description automatically generated

## SQL Query 5

### Question

An employee from Life Sciences education field working in Research & Development department has complained to HR that employees with her educational background are paid more in Sales department than in Research & Development. What insight can you provide to prove or disprove that statement?

### Insights About SQL Query and Results (Include # of Rows in Result)

We have calculated the average incomes of employees having Life Sciences in Education Field working in Sales, R&D, and HR departments and it looks like the employee is correct as the average income of employees having Education Field as Life Sciences working in the Sales department is higher than those in the R&D department. It provides support to the prove the statement by the employee is Correct.

# of Rows in Result = 3

### Translation

Select average of monthly income, and department from employee chargeability table joined with employee details table on ChargeabilityID matched with Employee\_chargeability\_ChargeabilityID, and then joined with Job Details table on JobID matched with Job\_Details\_JobID, where Education Field is equal to = “Life Sciences”, grouped by Department.

CLEANUP: Select avg ~~of~~ monthlyincome as avgInc, ~~and~~ department from employee\_chargeability ~~table~~ join~~ed~~ ~~with~~ employee\_details ~~table~~ on ChargeabilityID ~~matched with~~ Employee\_chargeability\_ChargeabilityID, ~~and then~~ join~~ed with~~ Job\_Details ~~table~~ on JobID ~~matched with~~ Job\_Details\_JobID, where EducationField ~~is equal to~~ = “Life Sciences”, group~~ed~~ by Department.

### Screen Shot of SQL Query and Results

## A screenshot of a cell phone Description automatically generatedSQL Query 6

### Question

A press article in a business magazine has said that at this company, Marital status of women in Research & Development has severely affected their promotion rates. What initial finding can you obtain from the data to help articulate the company's response in this regard?

### Insights About SQL Query and Results (Include # of Rows in Result)

As seen in the two tables that we've queried the magazine claims of discrimination on the basis of Marital Status of women in Research & Development Department turns out to be complete false.

|  |  |  |
| --- | --- | --- |
| **Research & Development Department**  (Average time for promotion in years) | **Marital Status** | **Overall Department**  (Average time for promotion in years) |
| 2.11 | Married | 2.31 |
| 2.29 | Single | 2.09 |
| 2.62 | Divorced | 2.57 |

We can compare the Research & Development Department result with the overall Company Data and we can see its completely Random as Married Women employees are averaging good in Research & Development Department.

Hence the Company can counter the magazines claims with the company’s overall average for Women employees.

# of Rows in Result = 3

### Translation

Part 1.

TRANSLATION: Select average of Years since last promotion and Marital Status from Employee Performance table joined with employee details table on PerformanceID matched with employee\_Performance\_PerformanceID, and then joined with job details table on JobID matched with Job\_Details\_JobID, where Department is equal to = “Research & Development” and Gender is equal to = “Female”, grouped by Marital Status.

CLEANUP: Select avg YearsSinceLastPromotion as avgLP ~~and~~ MaritalStatus from Employee\_Performance ~~table~~ join~~ed with~~ employee\_details ~~table~~ on PerformanceID ~~matched with~~ employee\_Performance\_PerformanceID, ~~and then~~ join~~ed~~ ~~with~~ job\_details ~~table~~ on JobID ~~matched with~~ Job\_Details\_JobID, where Department ~~is equal to~~ = “Research & Development” and Gender ~~is equal to~~ = “Female”, group~~ed~~ by MaritalStatus.

Part 2.

TRANSLATION: Select average of Years since last promotion and Marital Status from Employee Performance table joined with employee details table on PerformanceID matched with employee\_Performance\_PerformanceID, where gender is equal to = “Female”, grouped by Marital Status.

CLEANUP: Select avg YearsSinceLastPromotion as avgLP ~~and~~ MaritalStatus from Employee\_Performance ~~table~~ join~~ed with~~ employee\_details ~~table~~ on PerformanceID ~~matched with~~ employee\_Performance\_PerformanceID, where gender ~~is equal to~~ = “Female”, group~~ed~~ by MaritalStatus.

Screen Shot of SQL Query and Results

Part 1.

A screenshot of a cell phone

Description automatically generated

Part 2. A screenshot of a cell phone

Description automatically generated

# Data Review for MongoDB

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

In the NoSQL database, we have used similar tables as those used in our SQL database to keep the consistency of the report. In this, we have used the lookup command in the shell to overcome the inability of the NoSQL database to deal with Relation between Collections.

Assumptions:

1. We have assumed that each Employee has Age over 18 So didn’t include that in the table but mongo can be restructured with each document added so we can add it in future if required.
2. We have assumed that each employee has Standard working hours equal to 80 hrs which if changed can be added at a later very easily with NoSql.
3. We have assumed that employee Count remains 1 for each employee.
4. We have also assumed that Each employee has its own daily and monthly Rate other than Monthly Income and these rates are used to bill clients of the company providing service to them.
5. We also assume that Distance from home is the distinct relative quantity and higher the value means farther is the distance from home.
6. We are also assuming that Education value denotes the level of education. This means that higher the value of education higher the employee education is.
7. We assume that one employee will be from one education field only.
8. We assume that StockOptionLevel for each employee denotes the amount of stock owned by each employee on a relative scale. E.g. An employee with StockOptionLevel of 3 owns a bigger amount of stocks compared with that with a StockOptionLevel of 1.
9. We assume Overtime attributes give the idea that employee has done overtime for that month meaning worked more than standard hours.
10. We assume that the hourly rate is used to calculate the overtime income with whatever hours of overtime is done by an employee.
11. We are assuming for each department that there are multiple job roles, and for each role, they have multiple Job levels. Job role defines the work of the person whereas Job level denotes that employee’s position in the hierarchy structure of that role.
12. We assumed that one employee did not take multiple surveys.
13. We assumed that one person was not working multiple different jobs and thus worked for two different companies or took surveys for each job.
14. We assumed that there were not multiple entries for the same employees.

# Physical Mongo Database

## Assumptions/Notes About Data Set

There are no empty fields or sparse data in the data set.

All employees are above the age of 18 and have 80 Standard Working hours.

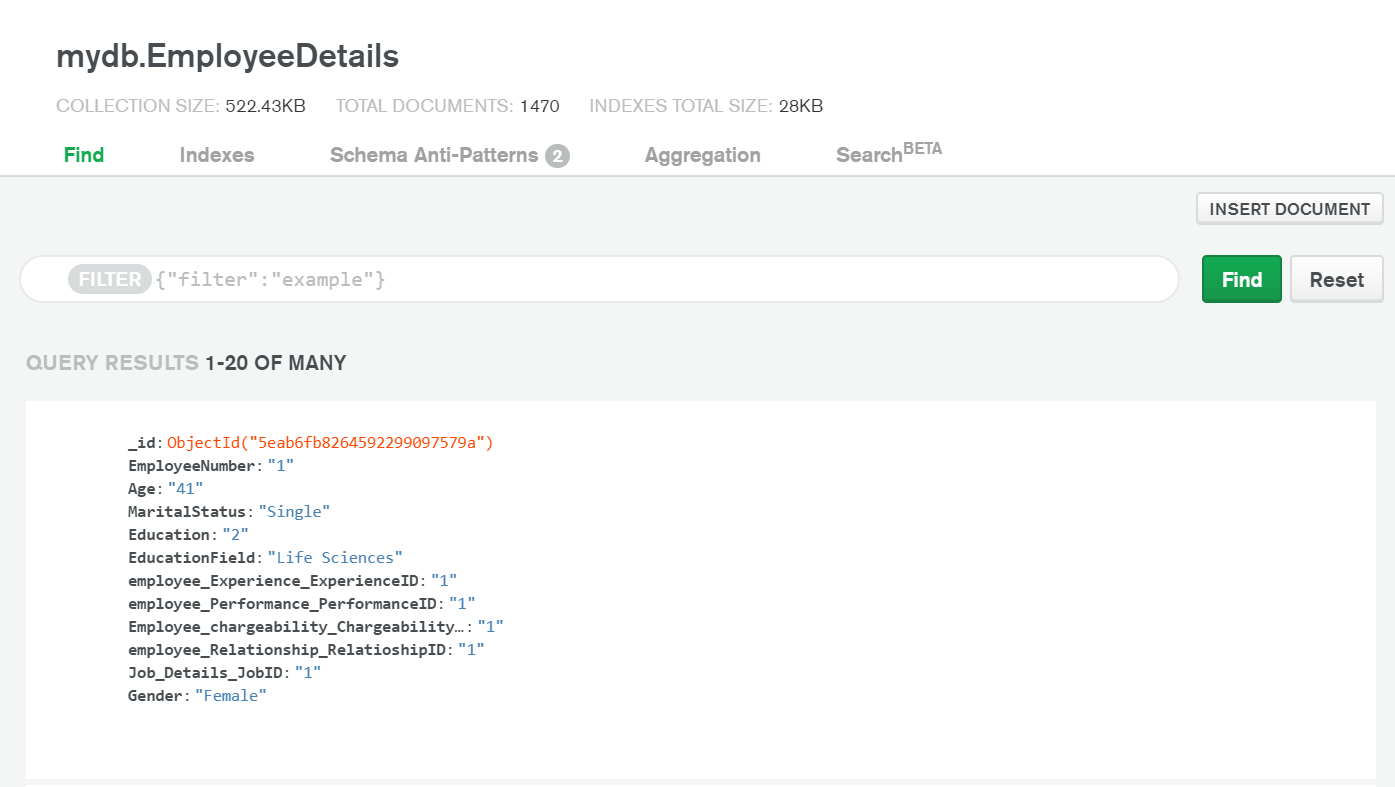
## Screen shot of Physical Database objects (Database, Collections and Attributes)

Model From Shell:

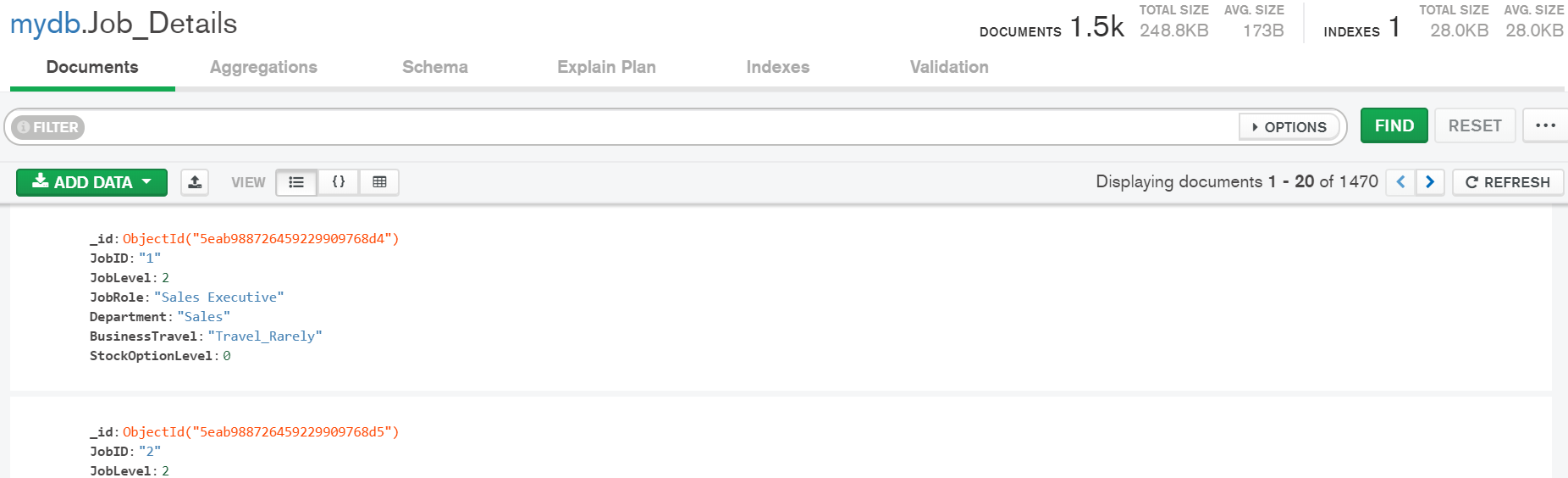


All the collections screenshots from ATLAS and Compass combined:

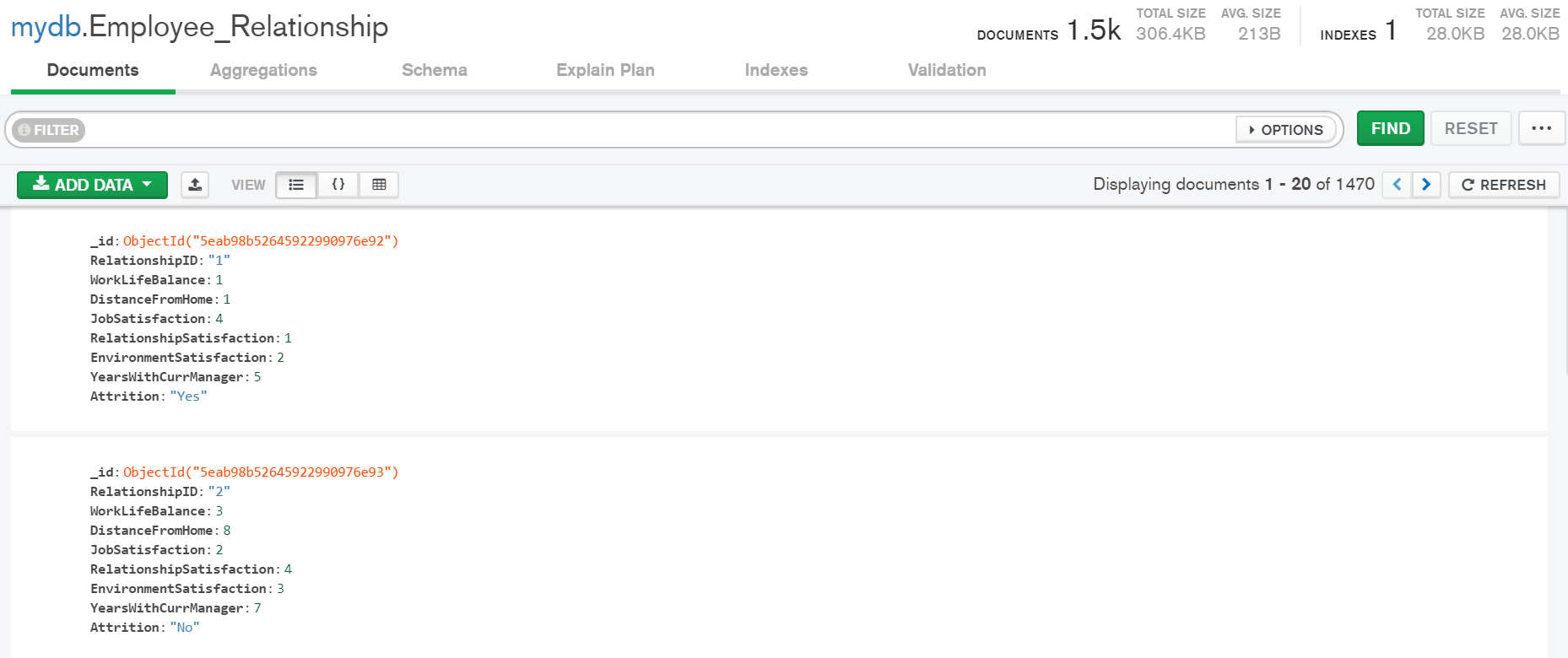
EmployeeDetails Collection:



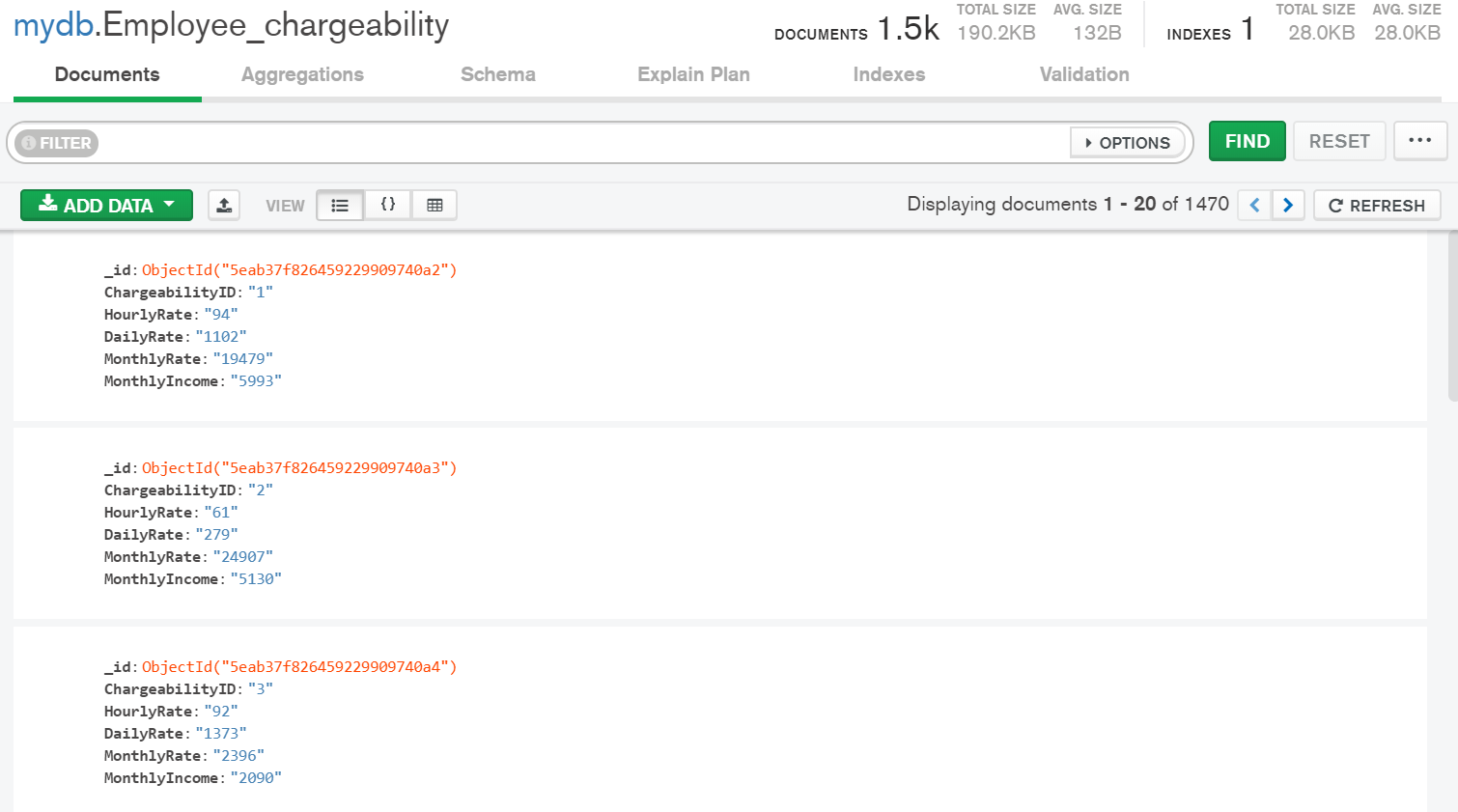
Job\_Details Collection:



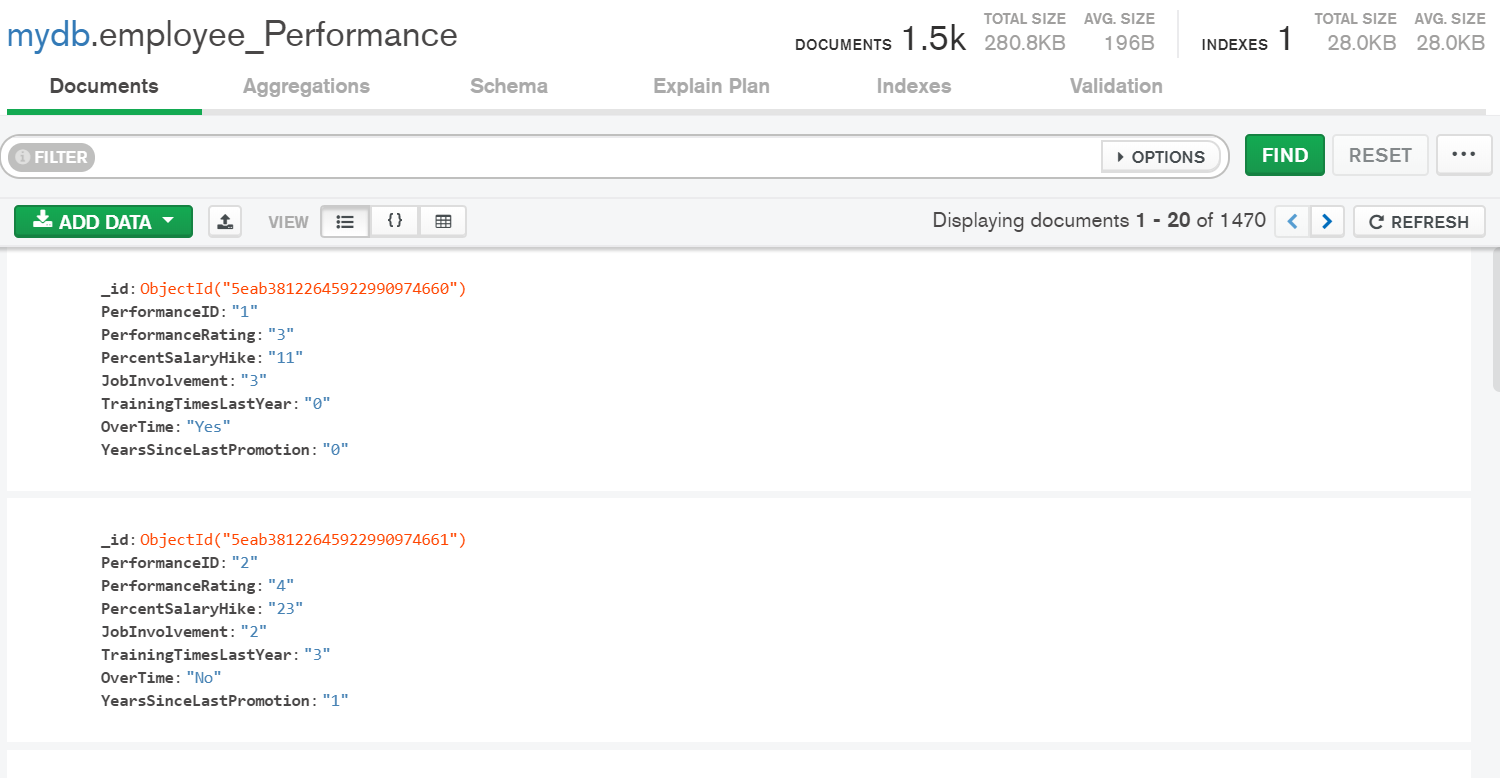
Employee\_Relationship Collection:



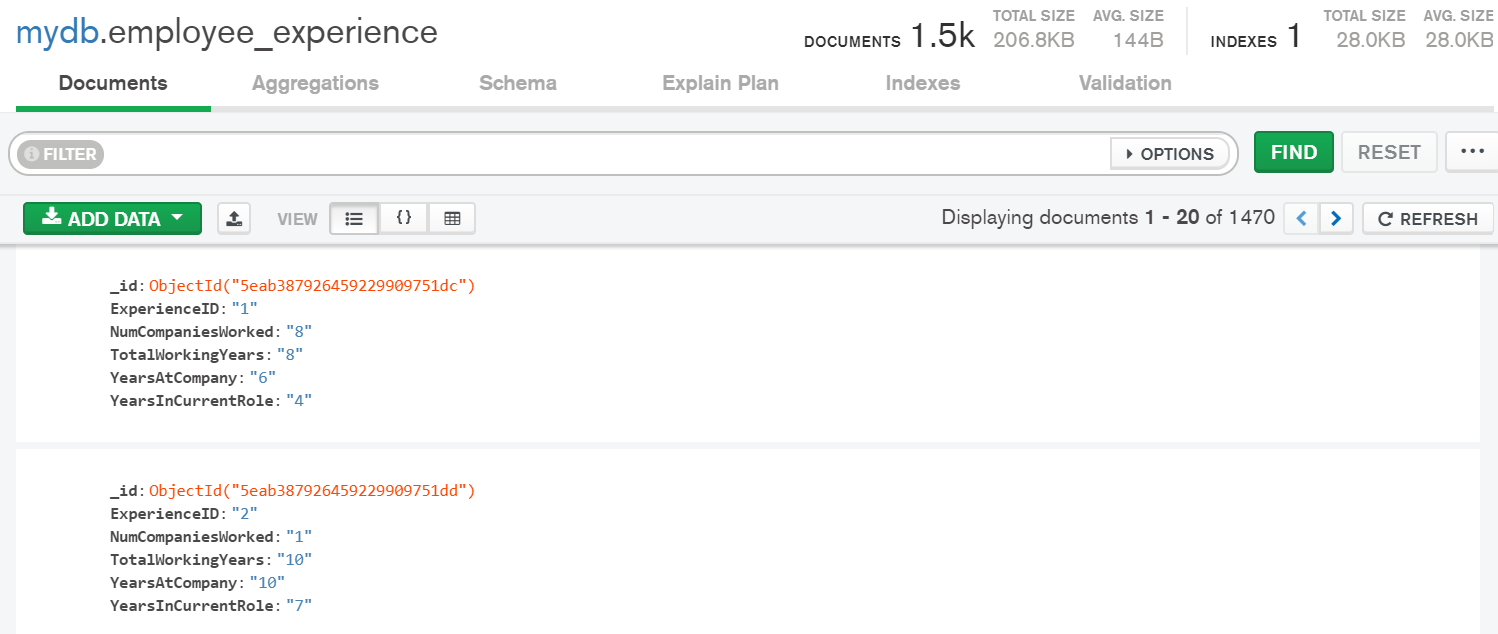
Employee\_chargeability Collection:



Employee\_Performance Collection:



Employee\_experience Collection:



## 

## Data in the Database

|  |  |  |
| --- | --- | --- |
| **Collection Name** | **Relationships With Other Collections (if any)** | **# of Documents in Collection** |
| Employee\_experience |  | 1470 |
| Employee\_Performance |  | 1470 |
| Employee\_chargeability |  | 1470 |
| Employee\_Relationship |  | 1470 |
| Job\_Details |  | 1470 |
| EmployeeDetails |  | 1470 |

# MongoDB Queries/Code

Pick 3 SQL queries and write them in MongoDB

## Mongo Query 1

### Question

A new research scientist who loves to travel joins the firm and is told by HR that his job role is one of the top two roles in terms of employees that travel frequently. Is HR right in saying so? Why or Why not?

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

No of Documents = 9

Results:

According to the data, HR is right in saying so as we see that Sales Executive and Research Scientist are the top two roles in terms of employees that travel frequently.

**Code:**

db.Job\_Details.aggregate([

{

$match: {"BusinessTravel":'Travel\_Frequently'}},

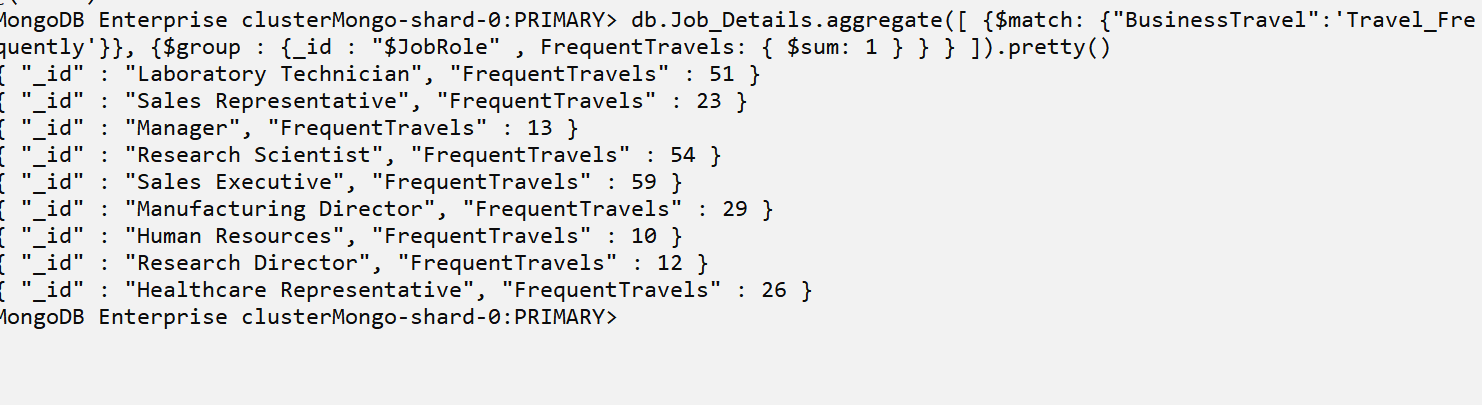
{$group : {\_id : "$JobRole" , FrequentTravels: { $sum: 1 } } }

]).pretty()

### Translation

On mydb aggregate Job\_Details, match BusinessTravel is equal to Travel\_Frequently and group them by JobRole and count it

### Screen Shot of MongoDB Query/Code and Results



## Mongo Query 2

### Question

The company has been paying gas expenses for miles traveled by employees between their home and work. If they want to increase the per mile compensation, which department's employees will gain the least?

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

No of Documents = 3

Result:

Total distance traveled to work from home is least for the employees of the Human Resources Department hence they will receive the least gain by the increase of compensation.

**Code:**

db.EmployeeDetails.aggregate([

{ $lookup: { from: "Job\_Details", localField: " Job\_Details\_JobID", foreignField: "JobID", as: "e1" } },{ $unwind:"$e1" },

{ $lookup: { from: "Employee\_Relationship", localField: " employee\_Relationship\_RelatioshipID", foreignField: "RelationshipID",as: "e2"}},

{ $unwind:"$e2" },

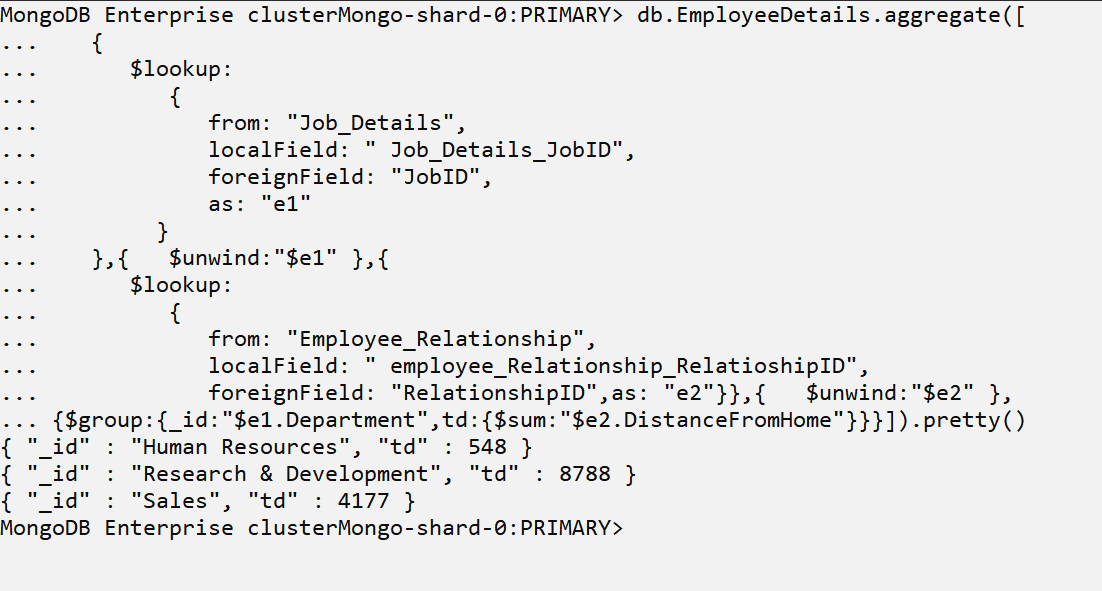
{$group:{\_id:"$e1.Department",td:{$sum:"$e2.DistanceFromHome"}}}

]).pretty()

### Translation

On mydb aggregate EmployeeDetails, lookup for Job\_Details collection using Job\_Details\_JobID as localfield and JobID as foreignField and grouped them by Department and then Sum the DistanceFromHome.

### Screen Shot of MongoDB Query/Code and Results



## Mongo Query 3

### Question

Sales feels that their environment satisfaction score is higher than HR but HR job satisfaction score is lower than Research & Development. Are they right?

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

No of documents = 3

Result:

We have compared the average Job Satisfaction & Environment Satisfaction scores of each department and we conclude that Sales is WRONG in saying that its Environment Satisfaction score is higher than HR, but it is RIGHT in saying that the Job Satisfaction score of HR is less than R&D.

**Code:**

db.EmployeeDetails.aggregate([

{

$lookup:

{

from: "Job\_Details",

localField: " Job\_Details\_JobID",

foreignField: "JobID",

as: "e1"

}

},{ $unwind:"$e1" },{

$lookup:

{

from: "Employee\_Relationship",

localField: " employee\_Relationship\_RelatioshipID",

foreignField: "RelationshipID",as: "e2"}},{ $unwind:"$e2" },

{

$group:{\_id:"$e1.Department",avgJobSatisfaction:{$avg:"$e2.JobSatisfaction"},

averageEnvSatisfaction:{$avg:"$e2.EnvironmentSatisfaction"}}}]).pretty()

### Translation

On mydb aggregate EmployeeDetails, lookup for Job\_Details collection using Job\_Details\_JobID as localfields and JobID as foreignField then lookup for Employee\_Relationship collection using employee\_Relationship\_RelatioshipID as localfield and RelationshipID as foreignField and grouped them by Department and then calculate average EnvironmentSatisfaction and average JobSatisfaction.

### Screen Shot of MongoDB Query/Code and Results

