



MySQL-Workbench

IBM HR Analytics Employee Attrition & Performance

Relational Data Model

Assumptions/Notes About Data Entities and Relationships and reasons why the data model is in 3NF.

Entities and Attributes

Employee: The central entity representing employees. Attributes include unique identifiers, demographic data, and employment-related characteristics like *EmployeeNumber*, *Age*, *Gender*, etc. All tables appear to have only columns that are directly related to an employee, so they are all functionally dependent on the primary key, **EmployeeNumber**. It is in 3NF.

Departments: Represents different departments within the organization. Contains *DepartmentID* and *Department*. “Department” is functionally dependent only on **DepartmentID** and not on any other non-primary key. It is in 3NF.

Education: Details about education credentials of employees. Contains *EducationID* and *EducationField*. *EducationField* is dependent only on **EducationID**, it satisfies the 3NF.

Jobs: Represents the various job roles within the company. This table has **JobID** and **JobRole** which suggests that **JobRole** is dependent only on **JobID**, which complies with 3NF.

Income: Contains financial details related to employees. It has attributes like *HourlyRate*, *MonthlyIncome*, *PercentSalaryHike*, etc., that are dependent only on the primary key *IncomeID* (*EmployeeNumber* is a foreign key here to link this table to Employee Table). It is in 3NF.

Relationships

Employee - Departments: Has one-to-many relationship. Each employee is associated with one department, and each department can have multiple employees. Linked via *DepartmentID*.

Employee - Education: Has one-to-many relationship. Each employee has one educational record, and each education field can be linked to multiple employees. Linked via *EducationID*.

Employee - Jobs: Has one-to-many relationship. Each employee has one job role, and each job role can be associated with multiple employees. Linked via *JobID*.

Employee - Income: Appears to be a one-to-many relationship. Each employee can have multiple income records associated with them as mentioned in our assumptions earlier (*hourlyRate*, *Monthly* etc.). Linked via *EmployeeNumber*

Physical MySQL Database

Assumptions/Notes About Data Set

- ❖ We have found that columns “EmployeeCount” and “StandardHours” to be constant 1 and 80 respectively. They are not contributing anything or helping with the design of the database. We decided to eliminate them.
- ❖ Also, the column “Over18” provides information that can be obtained from the “age” column. We also eliminate that information.
- ❖ The data also have “DailyRate”, “HourlyRate”, “MonthlyIncome”. Logically, one of these can be used to determine the others. However, since it wasn’t explicitly stated that there is a relationship between them, our assumption is that they are not related to each other. Meaning, the DailyRate cannot be used estimate the HourlyRate or the MonthlyIncome.

Data in the Database

Table Name	Primary Key	Foreign Key	# of Rows in Table
Departments	DepartmentID		3
Education	EducationID		6
Employee	EmployeeNumber	DepartmentID JobID EducationID	1470
Income	IncomeID	EmployeeNumber	1470
Jobs	JobID		9

Screen shot of Physical Database objects.

Departments Table

MySQL Workbench

Local instance MySQL81 (fall2023project) x MySQL Model (Project-ERD.pdf.m... x EER Diagram x EER Diagram1 x Local instance MySQL81 (fall2023project) x

File Edit View Query Database Server Tools Scripting Help

Navigator

Schemas

Filter objects

- fall2023project
 - departments
 - education
 - employee
 - income
 - jobs

Table: departments

Columns: DepartmentID (PK), Department

SQL File 4* SQL File 15*

1. SELECT * FROM fall2023project.departments;

Result Grid

DepartmentID	Department
D1	Sales
D2	Human Resources
D3	Research & Developm...

departments 1 x

Output

Action Output

#	Time	Action	Message	Duration / Fetch
109	14:34:43	Select EmployeeNumber from Employee where DepartmentID = D1 LIMIT 0, 2000	Error Code: 1054. Unknown column 'DT' in 'where clause'	0.000 sec / 0.000 sec
110	14:43:05	SELECT * FROM Departments LIMIT 0, 2000	3 row(s) returned	0.000 sec / 0.000 sec
111	14:52:04	SELECT * FROM Employee WHERE Age > 30 AND DepartmentID = 'D1' LIMIT 0, 2000	330 row(s) returned	0.000 sec / 0.000 sec
112	16:25:24	SELECT * FROM fall2023project.income LIMIT 0, 2000	1470 row(s) returned	0.000 sec / 0.000 sec
113	16:30:37	SELECT * FROM fall2023project.departments LIMIT 0, 2000	3 row(s) returned	0.000 sec / 0.000 sec

Query Completed

48°F Sunny

4:30 PM 11/26/2023

Income Table

MySQL Workbench

Local instance MySQL81 (fall2023project) x MySQL Model (Project-ERD.pdf.m... x EER Diagram x EER Diagram1 x Local instance MySQL81 (fall2023project) x

File Edit View Query Database Server Tools Scripting Help

Navigator

Schemas

Filter objects

- fall2023project
 - departments
 - education
 - employee
 - income
 - jobs

Table: income

Columns: IncomeID (PK), EmployeeNumber, DailyRate, HourlyRate, MonthlyIncome, MonthlyRate, PercentSalaryHike, StockOptionLevel

SQL File 4* SQL File 15*

1. SELECT * FROM fall2023project.income;

Result Grid

IncomeID	EmployeeNumber	DailyRate	HourlyRate	MonthlyIncome	MonthlyRate	PercentSalaryHike	StockOptionLevel
IN1	1	1102	94	5993	19479	11	0
IN10	13	1299	94	5237	16577	13	2
IN100	132	489	67	2042	25043	12	1
IN1000	1408	1147	31	16799	16616	14	1
IN1001	1409	258	54	2950	17363	13	0
IN1002	1411	1462	94	3629	19106	18	0
IN1003	1412	200	60	9362	19944	11	0
IN1004	1415	949	81	3229	4910	11	1
IN1005	1417	652	100	3578	23577	12	0
IN1006	1419	332	51	7988	9769	13	0
IN1007	1420	1475	97	4284	22710	20	0
IN1008	1421	337	84	7553	22930	12	0
IN1009	1422	971	54	17328	5652	19	0
IN101	133	807	63	2073	23648	22	0
IN1010	1423	1055	76	19701	72456	21	1

income 1 x

Output

Action Output

#	Time	Action	Message	Duration / Fetch
110	14:43:05	SELECT * FROM Departments LIMIT 0, 2000	3 row(s) returned	0.000 sec / 0.000 sec
111	14:52:04	SELECT * FROM Employee WHERE Age > 30 AND DepartmentID = 'D1' LIMIT 0, 2000	330 row(s) returned	0.000 sec / 0.000 sec
112	16:25:24	SELECT * FROM fall2023project.income LIMIT 0, 2000	1470 row(s) returned	0.000 sec / 0.000 sec
113	16:30:37	SELECT * FROM fall2023project.departments LIMIT 0, 2000	3 row(s) returned	0.000 sec / 0.000 sec
114	16:31:37	SELECT * FROM fall2023project.income LIMIT 0, 2000	1470 row(s) returned	0.000 sec / 0.000 sec

Query Completed

48°F Sunny

4:32 PM 11/26/2023

Employee Table

The screenshot displays the SQL Server Enterprise interface. The left pane shows the 'Schemas' tree with 'fall2023project' expanded, containing tables 'departments', 'education', 'employee', 'income', and 'jobs'. The 'employee' table is selected, and its structure is shown in the 'Table: employee' pane. The main pane shows the query 'SELECT * FROM fall2023project.employee;' with a result grid containing 20 rows of employee data. The bottom pane shows the 'Output' window with a list of actions and their durations.

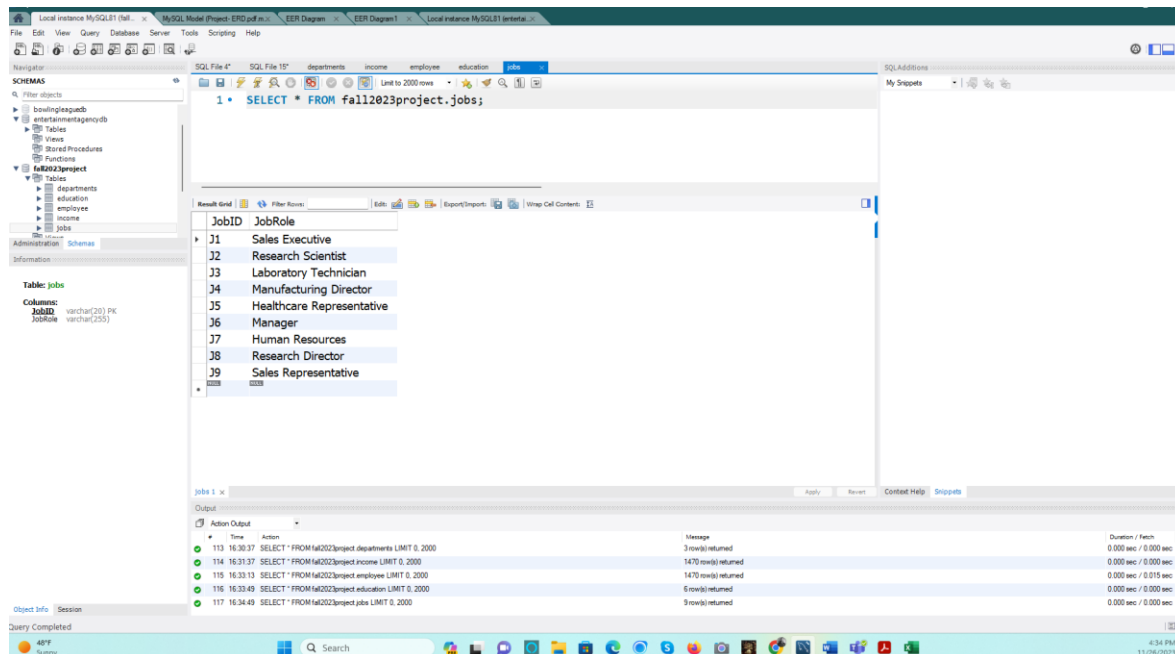
EmployeeNumber	DepartmentID	JobID	EducationID	Age	Attrition	Gender	MaritalStatus	PerformanceRating	RelationshipSatisfaction	EnvironmentSatisfaction
1	D1	J1	E1	41	Yes	Female	Single	3	1	2
2	D3	J2	E1	49	No	Male	Married	4	4	3
4	D3	J3	E3	37	Yes	Male	Single	3	2	4
5	D3	J2	E1	33	No	Female	Married	3	3	4
7	D3	J3	E2	27	No	Male	Married	3	4	1
8	D3	J3	E1	32	No	Male	Single	3	3	4
10	D3	J3	E2	59	No	Female	Married	4	1	3
11	D3	J3	E1	30	No	Male	Divorced	4	2	4
12	D3	J4	E1	38	No	Male	Single	4	2	4
13	D3	J5	E2	36	No	Male	Married	3	2	3
14	D3	J3	E2	35	No	Male	Married	3	3	1
15	D3	J3	E1	29	No	Female	Single	3	4	4
16	D3	J2	E1	31	No	Male	Divorced	3	4	1
18	D3	J3	E2	34	No	Male	Divorced	3	3	2
19	D3	J3	E1	28	Yes	Male	Single	3	2	3
20	D3	J4	E1	29	No	Female	Divorced	3	2	3

Education Table

The screenshot displays the SQL Server Enterprise interface. The left pane shows the 'Schemas' tree with 'fall2023project' expanded, containing tables 'departments', 'education', 'employee', 'income', and 'jobs'. The 'education' table is selected, and its structure is shown in the 'Table: education' pane. The main pane shows the query 'SELECT * FROM fall2023project.education;' with a result grid containing 6 rows of education data. The bottom pane shows the 'Output' window with a list of actions and their durations.

EducationID	EducationField
E1	Life Sciences
E2	Medical
E3	Other
E4	Marketing
E5	Human Resources
E6	Technical Degree

Jobs Table



SQL Queries

1) If the company wants to cut travel costs, which department should the company focus on?

Translation:

Select Department, Count of departments from table Employee joined with table departments matched on Employee's DepartmentID and Departments' DepartmentID where BusinessTravel is 'Travel Frequently' Group by Department.

Clean up:

Select Department, Count(departments) from table Employee joined with table departments matched on Employee' DepartmentID and = Departments' DepartmentID where BusinessTravel is 'Travel Frequently' Group by Department.

Query:

```
SELECT D.Department, COUNT(D.Department) as EmployeesTravelling, E.BusinessTravel
FROM EMPLOYEE E JOIN DEPARTMENTS D ON E.DEPARTMENTID = D.DEPARTMENTID
where E.BusinessTravel = 'Travel_Frequently'
GROUP BY D.DEPARTMENT;
```

Result:

6
7
8 • SELECT D.Department, COUNT(D.Department) as EmployeesTravelling, E.BusinessTravel
9 FROM EMPLOYEE E JOIN DEPARTMENTS D ON E.DEPARTMENTID = D.DEPARTMENTID
10 where E.BusinessTravel = 'Travel_Frequently'
11 GROUP BY D.DEPARTMENT;

Department	EmployeesTravelling	BusinessTravel
Research & Developm...	182	Travel_Frequently
Sales	84	Travel_Frequently
Human Resources	11	Travel_Frequently

From the result we can see that the Research and Development Department has 182 Employees travelling Frequently. Therefore the company should focus on this department in order to reduce travel costs.

2) 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 most?

Translation:

Select Department, Sum of DistanceFromHome as TotalDistanceofAllEmployees from table Employee joined with table departments matched on Employee' DepartmentID = Departments' DepartmentID Group by Departments.

Cleanup:

Select Department, Sum of (DistanceFromHome) as TotalDistanceofAllEmployees from table Employee joined with table departments matched on Employee.DepartmentID = Departments.DepartmentID Group by Departments.

Query:

```
SELECT D.Department, SUM(E.DistanceFromHome) as TotalDistanceOfAllEmployees
FROM EMPLOYEE E JOIN DEPARTMENTS D ON E.DEPARTMENTID = D.DEPARTMENTID
GROUP BY D.DEPARTMENT;
```

Result:

6	
7	
8	• SELECT D.Department,SUM(E.DistanceFromHome) as TotalDistanceOfAllEmployees
9	FROM EMPLOYEE E JOIN DEPARTMENTS D ON E.DEPARTMENTID = D.DEPARTMENTID
10	GROUP BY D.DEPARTMENT;
11	

int PK	Result Grid	Filter Rows:	Export:	Wrap Cell Content:
varchar(20)	Department	TotalDistanceOfAllEmployees		
varchar(20)	Sales	4177		
varchar(20)	Human Resources	548		
int	Research & Developm...	8788		
varchar(20)				
varchar(50)				
varchar(225)				
int				
int				
int				
int				
int				

Here we see that the total distance of all employees in the Research and Development Department is more than the other 2 departments. Hence if there is an increase in per mile compensation , then this department's employees will be benefited the most.

3) 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?

Translation:

Select Count of PerformanceRating , MaritalStatus from table Employee where PerformanceRating >'3' and Gender ='Male' Group by MaritalStatus Order by PerformanceRating.

Clean up:

Select Count of PerformanceRating , MaritalStatus from table Employee where PerformanceRating >'3' and Gender ='Male' Group by MaritalStatus Order by PerformanceRating.

Query:

```
SELECT COUNT(E.PERFORMANCERATING) HighPerformance , E.MaritalStatus
FROM EMPLOYEE E
WHERE E.PERFORMANCERATING >'3'
And E.Gender= 'Male'
GROUP BY E.MARITALSTATUS
ORDER BY HighPerformance ;
```

Result:

6	
7	
8	• SELECT COUNT(E.PERFORMANCERATING) HighPerformance , E.MaritalStatus
9	FROM EMPLOYEE E
10	WHERE E.PERFORMANCERATING >'3'
11	And E.Gender='Male'
12	GROUP BY E.MARITALSTATUS
13	ORDER BY HighPerformance ;
14	

HighPerformance	MaritalStatus
29	Divorced
44	Single
59	Married

Here we see that the 59 Married Men have a high performance rating as compared to the Single and Divorced men who are lesser. Therefore we can agree to the article about married men having higher performance rating from the initial findings.

4) 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?

Translation:

Select Department, Count of JobSatisfaction as NoOfEmployeesSatisfied from table Employee joined with table departments matched on Employee' DepartmentID = Departments' DepartmentID where JobSatisfaction > 2 Group by Department

UNION

Select Department, Count of EnvironmentSatisfaction as NoOfEmployeesSatisfied from table Employee joined with table departments matched on Employee' DepartmentID = Departments' DepartmentID where EnvironmentSatisfaction > 2 Group by Departments

Clean up:

Select Department, CountOf(JobSatisfaction) as NoOfEmployeesSatisfied from table Employee joined with table departments matched on Employee.DepartmentID = Departments.DepartmentID where JobSatisfaction > 2 Group by Department

UNION

Select Department, CountOf(EnvironmentSatisfaction) as NoOfEmployeesSatisfied from table Employee joined with table departments matched on Employee.DepartmentID = Departments.DepartmentID where EnvironmentSatisfaction > 2 Group by Department

Query:

SELECT D.Department,COUNT(E.JobSatisfaction) as NoOfEmployeesSatisfied
FROM EMPLOYEE E JOIN DEPARTMENTS D ON E.DEPARTMENTID = D.DEPARTMENTID
where E.JobSatisfaction > 2
GROUP BY D.DEPARTMENT

UNION

SELECT D.Department,COUNT(E.EnvironmentSatisfaction) as NoOfEmployeesSatisfiedEnvironment

```
FROM EMPLOYEE E JOIN DEPARTMENTS D ON E.DEPARTMENTID = D.DEPARTMENTID
where E.EnvironmentSatisfaction > 2
GROUP BY D.DEPARTMENT;
```

Result:

```

8 • SELECT D.Department,COUNT(E.JobSatisfaction) as NoOfEmployeesSatisfied
9 FROM EMPLOYEE E JOIN DEPARTMENTS D ON E.DEPARTMENTID = D.DEPARTMENTID
10 where E.JobSatisfaction > 2
11 GROUP BY D.DEPARTMENT
12 UNION
13 SELECT D.Department,COUNT(E.EnvironmentSatisfaction) as NoOfEmployeesSatisfiedEnvironment
14 FROM EMPLOYEE E JOIN DEPARTMENTS D ON E.DEPARTMENTID = D.DEPARTMENTID
15 where E.EnvironmentSatisfaction > 2
16 GROUP BY D.DEPARTMENT;
17

```

Department	NoOfEmployeesSatisfied
Sales	274
Human Resources	32
Research & Developm...	595
Sales	262
Human Resources	40
Research & Developm...	597

From the outlook , We can see that There are more number of people with job satisfaction in the other 2 departments when compared to HR department whereas in terms of Environment Satisfaction , Research & Development department have the most. Therefore, from the argument, We can say that the Research & Development department are right.

5) A press article in a business magazine has said that at this company, single women in Sales have worked at the company longer than divorced or married women.
What initial finding can you obtain from the data to help articulate the company's response in this regard?

Translation:

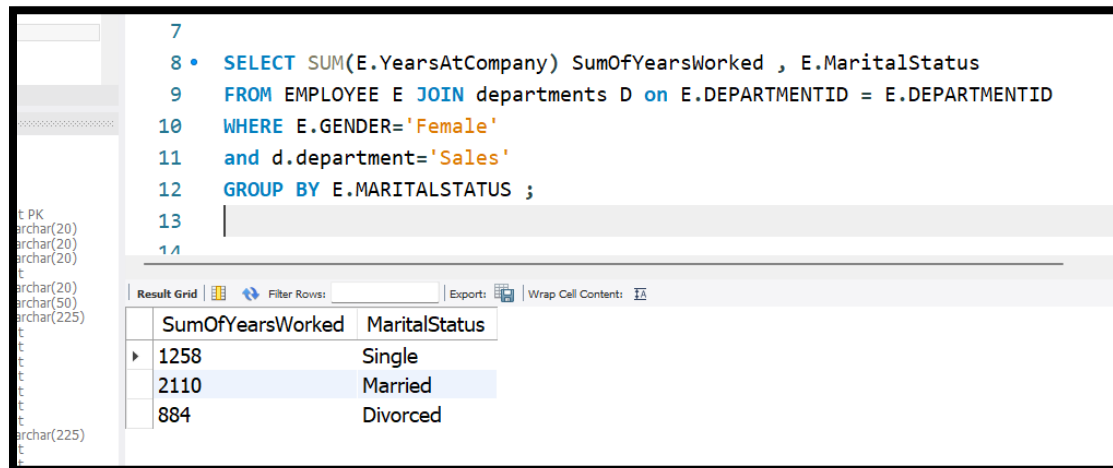
Select Sum(YearsAtCompany) as SumOfYearsWorked , MaritalStatus from table Employee joined with Departments table matched on Employee's DepartmentID = Departments' DepartmentID where Gender is 'female' and department ='Sales' Group by MaritalStatus.

Clean up:

Select Sum(YearsAtCompany) as SumOfYearsWorked , MaritalStatus from table Employee joined with Departments table matched on Employee.DepartmentID = Departments.DepartmentID where Gender is 'female' and department='Sales' Group by MaritalStatus.

Query:

```
SELECT SUM(E.YearsAtCompany) SumOfYearsWorked , E.MaritalStatus
FROM EMPLOYEE E JOIN departments D on E.DEPARTMENTID = E.DEPARTMENTID
WHERE E.GENDER='Female'
and d.department='Sales'
GROUP BY E.MARITALSTATUS ;
```

Result:

The screenshot shows a SQL query editor with a query window on the left and a results grid on the right. The query window contains the following SQL code:

```
7
8 • SELECT SUM(E.YearsAtCompany) SumOfYearsWorked , E.MaritalStatus
9 FROM EMPLOYEE E JOIN departments D on E.DEPARTMENTID = E.DEPARTMENTID
10 WHERE E.GENDER='Female'
11 and d.department='Sales'
12 GROUP BY E.MARITALSTATUS ;
13
14
```

The results grid displays the output of the query, showing the sum of years worked for each marital status of female employees in the Sales department. The grid has two columns: SumOfYearsWorked and MaritalStatus. The data is as follows:

SumOfYearsWorked	MaritalStatus
1258	Single
2110	Married
884	Divorced

From the result we can see that Married women in Sales have the most number of years worked whereas single women in Sales are the second most. We can say that Single Women of Sales department have worked more years than divorced women in Sales but not more than Married women in Sales department.