EMPLOYEE ATTRITION PREDICTION: SQL

Creating table after connecting to database

Query | Query History

```
CREATE TABLE Employees (
    EmployeeID SERIAL PRIMARY KEY, JobRole VARCHAR(50), Attrition Varchar (10),
    Department VARCHAR(255), Age INT, BusinessTravel VARCHAR(255),
    DistanceFromHome INT,Education INT, EducationField VARCHAR(255),
    EnvironmentSatisfaction INT, Gender VARCHAR(50), HourlyRate INT,
    JobInvolvement INT,JobLevel INT, JobSatisfaction INT,
    MaritalStatus VARCHAR(255), MonthlyIncome INT, DailyRate INT,
    MonthlyRate INT, NumCompaniesWorked INT, Over18 VARCHAR(5),
    OverTime VARCHAR(5), PercentSalaryHike INT, PerformanceRating INT,
    RelationshipSatisfaction INT, StandardHours INT, StockOptionLevel INT,
    TotalWorkingYears INT, TrainingTimesLastYear INT, WorkLifeBalance INT,
    YearsAtCompany INT,YearsInCurrentRole INT,
    YearsSinceLastPromotion INT, YearsWithCurrManager INT
);
```

First View of Dataset

Data Output | Messages | Notifications

	employeeid [PK] integer	jobrole character varying (50)	attrition character varying (10)	department character varying (255)	age integer	businesstravel character varying (255)	distancefromhome integer	education integer	educationfield character varying (255)
1	1	Sales Executive	Yes	Sales	41	Travel_Rarely	1	2	Life Sciences
2	2	Research Scientist	No	Research & Development	49	Travel_Frequently	8	1	Life Sciences
3	4	Laboratory Technician	Yes	Research & Development	37	Travel_Rarely	2	2	Other
4	5	Research Scientist	No	Research & Development	33	Travel_Frequently	3	4	Life Sciences
5	7	Laboratory Technician	No	Research & Development	27	Travel_Rarely	2	1	Medical
6	8	Laboratory Technician	No	Research & Development	32	Travel_Frequently	2	2	Life Sciences
7	10	Laboratory Technician	No	Research & Development	59	Travel_Rarely	3	3	Medical
8	11	Laboratory Technician	No	Research & Development	30	Travel_Rarely	24	1	Life Sciences
9	12	Manufacturing Director	No	Research & Development	38	Travel_Frequently	23	3	Life Sciences
10	13	Healthcare Representative	No	Research & Development	36	Travel_Rarely	27	3	Medical
11	14	Laboratory Technician	No	Research & Development	35	Travel_Rarely	16	3	Medical
12	15	Laboratory Technician	No	Research & Development	29	Travel_Rarely	15	2	Life Sciences
13	16	Research Scientist	No	Research & Development	31	Travel_Rarely	26	1	Life Sciences
14	18	Laboratory Technician	No	Research & Development	34	Travel_Rarely	19	2	Medical
15	19	Laboratory Technician	Yes	Research & Development	28	Travel_Rarely	24	3	Life Sciences
16	20	Manufacturing Director	No	Research & Development	29	Travel_Rarely	21	4	Life Sciences
17	21	Research Scientist	No	Research & Development	32	Travel_Rarely	5	2	Life Sciences
18	22	Laboratory Technician	No	Research & Development	22	Non-Travel	16	2	Medical
19	23	Manager	No	Sales	53	Travel_Rarely	2	4	Life Sciences
20	24	Research Scientist	No	Research & Development	38	Travel_Rarely	2	3	Life Sciences

Total rows: 1000 of 1470 | Query complete 00:00:00.848

1. Departments with the Highest Attrition

```
SELECT Department, count(attrition) AS attrition_count
FROM Employees
where attrition = 'Yes'
GROUP BY Department
ORDER BY attrition_count desc,Department
```

Data Output | Messages | Notifications

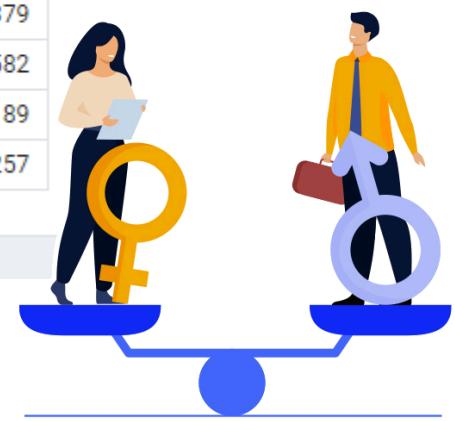
	department character varying (255)	attrition_count bigint
1	Research & Development	133
2	Sales	92
3	Human Resources	12

2. Gender Distribution across Different Departments

```
SELECT Department, Gender, COUNT(*) AS Number_Of_Employees
FROM Employees
GROUP BY Department, Gender
ORDER BY Department, Gender;
```

	department character varying (255)	gender character varying (50)	number_of_employees bigint
1	Human Resources	Female	20
2	Human Resources	Male	43
3	Research & Development	Female	379
4	Research & Development	Male	582
5	Sales	Female	189
6	Sales	Male	257

Total rows: 6 of 6 | Query complete 00:00:00.418



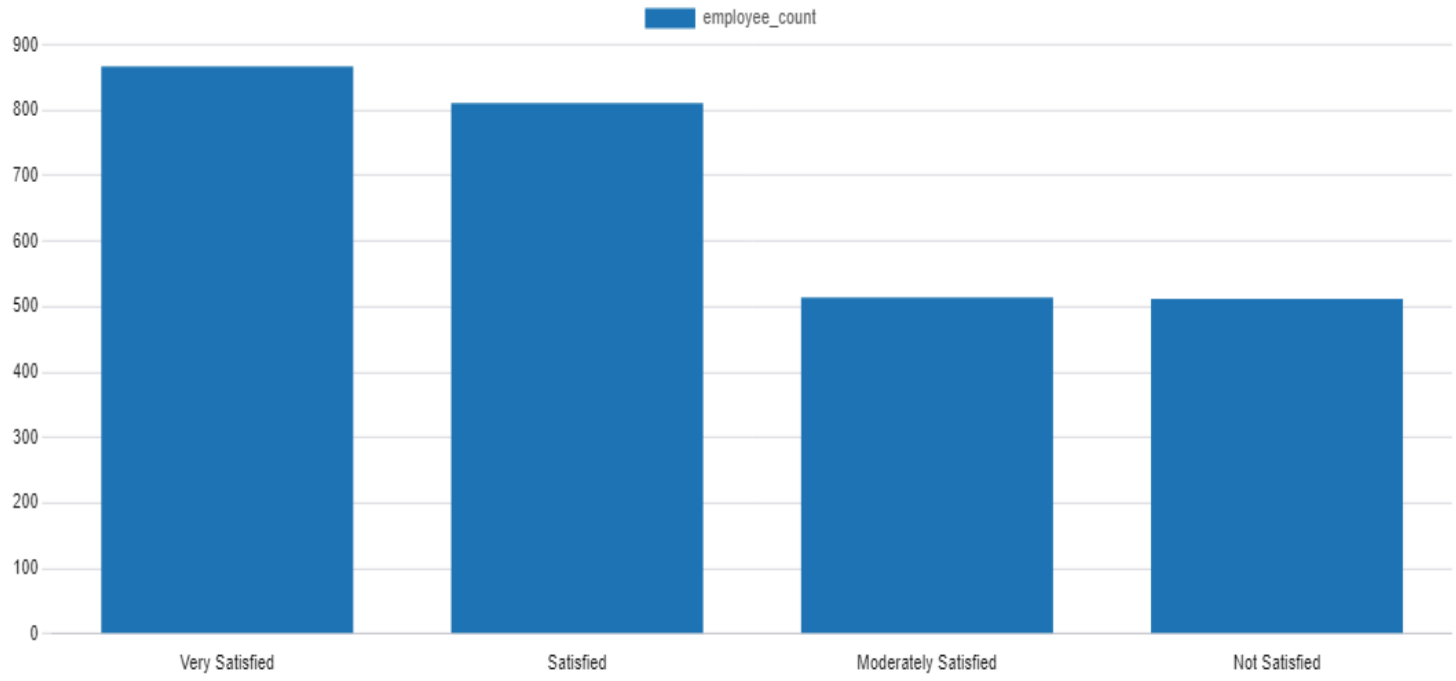
3. Employee Count by Job Satisfaction Level

```
Select case when jobsatisfaction = 1 then 'Not Satisfied'
when jobsatisfaction  =2 then 'Moderately Satisfied'
when jobsatisfaction = 3 then 'Satisfied'
else 'Very Satisfied' end as satisfaction_level , count(*) as Employee_count
from employees
group by satisfaction_level
order by employee_count desc
```

Data Output | Messages | Notifications

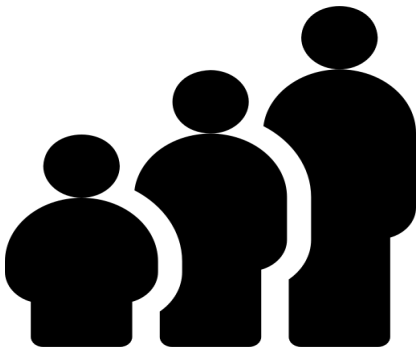
	satisfaction_level text	employee_count bigint
1	Very Satisfied	459
2	Satisfied	442
3	Not Satisfied	289
4	Moderately Satisfied	280

Total rows: 4 of 4 | Query complete 00:00:00.096

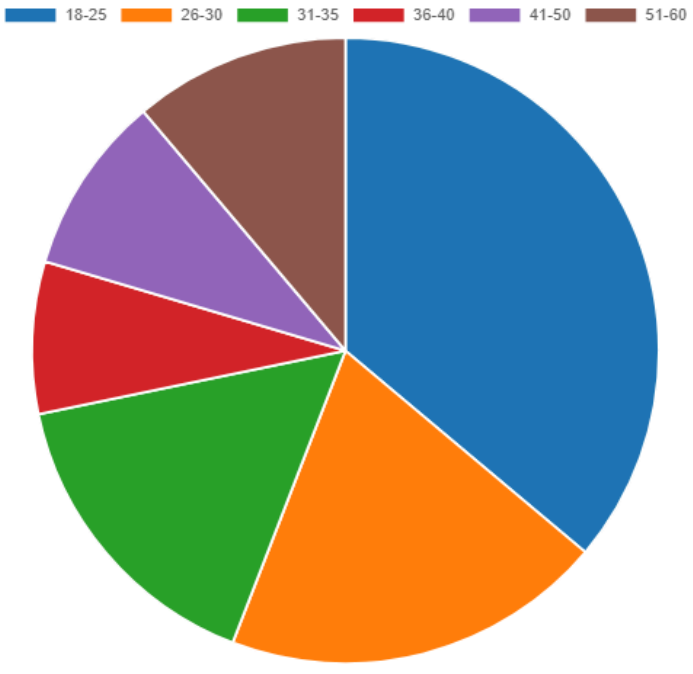


4. Employee Attrition Rate Based on Different Factors

4.a Attrition based on Age-group



```
SELECT
    CASE
        WHEN Age BETWEEN 18 AND 25 THEN '18-25'
        WHEN Age BETWEEN 26 AND 30 THEN '26-30'
        WHEN Age BETWEEN 31 AND 35 THEN '31-35'
        WHEN Age BETWEEN 36 AND 40 THEN '36-40'
        WHEN Age BETWEEN 41 AND 50 THEN '41-50'
        WHEN Age BETWEEN 51 AND 60 THEN '51-60'
    END AS Age_Range,
    ROUND(
        (SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END)* 100.0) / COUNT(*),2) AS Attrition_Rate
FROM
    employees
GROUP BY
    CASE
        WHEN Age BETWEEN 18 AND 25 THEN '18-25'
        WHEN Age BETWEEN 26 AND 30 THEN '26-30'
        WHEN Age BETWEEN 31 AND 35 THEN '31-35'
        WHEN Age BETWEEN 36 AND 40 THEN '36-40'
        WHEN Age BETWEEN 41 AND 50 THEN '41-50'
        WHEN Age BETWEEN 51 AND 60 THEN '51-60' END
ORDER BY
    Age_Range;
```

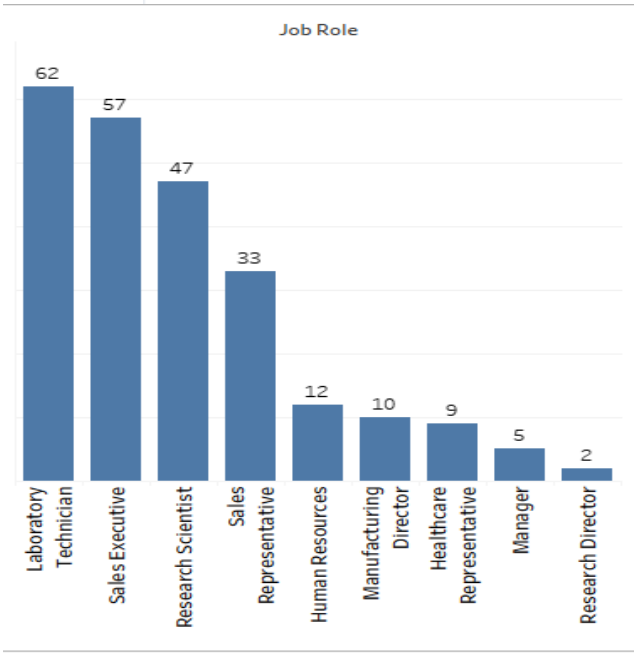


Data Output | Messages | Notifications

	age_range text	attrition_rate numeric
1	18-25	35.77
2	26-30	21.29
3	31-35	17.49
4	36-40	9.06
5	41-50	10.56
6	51-60	12.59

4.b Attrition based on Job Role

```
SELECT JobRole,ROUND((SUM(CASE WHEN Attrition = 'Yes'
                          THEN 1 ELSE 0 END) * 100.0) / COUNT(*),
        2) AS Attrition_Rate
FROM employees
GROUP BY JobRole
```



Data Output | Messages | Notifications

	jobrole character varying (50)	attrition_rate numeric
1	Sales Representative	39.76
2	Laboratory Technician	23.94
3	Human Resources	23.08
4	Sales Executive	17.48
5	Research Scientist	16.10
6	Manufacturing Director	6.90
7	Healthcare Representative	6.87
8	Manager	4.90
9	Research Director	2.50

4.c Attrition rate based on Performance

```
SELECT
    performancerating as Performance_Rating,
    ROUND( (SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0) / COUNT(*),2)
    AS Attrition_Rate
FROM employees
GROUP BY performancerating
ORDER BY Attrition_Rate DESC;
```

Data Output | Messages | Notifications

	performance_rating integer	attrition_rate numeric
1	4	16.37
2	3	16.08

4.d Attrition Rate on Basis of Stock Level

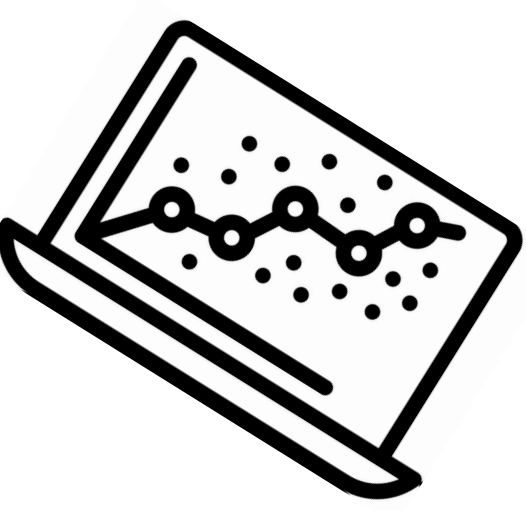
```
Select stockoptionlevel as Stock_level , count(attrition) as employee_count from employees
where attrition  = 'Yes'
group by Stock_level
```

Data Output | Messages | Notifications

	stock_level integer	employee_count bigint
1	0	154
2	1	56
3	3	15
4	2	12

5. Correlation Between Percent Salary Hike and Performance Rating

```
SELECT PerformanceRating, ROUND(AVG(PercentSalaryHike),2) AS Avg_Percent_Salary_Hike
FROM Employees
GROUP BY PerformanceRating
ORDER BY PerformanceRating;
```



Data Output | Messages | Notifications

	performancerating integer	avg_percent_salary_hike numeric
1	3	14.00
2	4	21.85

6. Career Progression Analysis:

What are the average years at the company and total working years for different job roles?

```
SELECT
    JobRole,
    round(AVG(YearsAtCompany),2) AS Avg_Years_At_Company,
    round(AVG(TotalWorkingYears),2) AS Avg_Total_Working_Years
FROM employees
GROUP By JobRole
ORDER BY Avg_Total_Working_Years DESC;
```

Data Output | Messages | Notifications

	jobrole character varying (50)	avg_years_at_company numeric	avg_total_working_years numeric
1	Manager	14.43	24.55
2	Research Director	10.94	21.40
3	Healthcare Representative	8.37	14.07
4	Manufacturing Director	7.60	12.79
5	Sales Executive	7.50	11.10
6	Human Resources	5.33	8.17
7	Research Scientist	5.11	7.72
8	Laboratory Technician	5.02	7.66
9	Sales Representative	2.92	4.67

7. Gender Pay Gap Analysis:

What is the average monthly income of male and female employees?

```
SELECT
    Gender,round(AVG(MonthlyIncome),0) AS Avg_Monthly_Income
FROM employees
GROUP BY Gender;
```

Data Output | Messages | Notifications

	gender character varying (50)	avg_monthly_income numeric
1	Female	6687
2	Male	6381

8. Find the highest earners in each job role

```
SELECT JobRole,EmployeeID, MonthlyIncome
FROM (
    SELECT
        EmployeeID,JobRole, MonthlyIncome,
  RANK() OVER (PARTITION BY JobRole ORDER BY MonthlyIncome DESC) AS IncomeRank
  FROM employees
) subquery
WHERE  IncomeRank =1 ;
```

	jobrole character varying (50)	employeeid [PK] integer	monthlyincome integer
1	Healthcare Representative	1661	13966
2	Human Resources	698	10725
3	Laboratory Technician	944	7403
4	Manager	259	19999
5	Manufacturing Director	1005	13973
6	Research Director	1035	19973
7	Research Scientist	86	9724
8	Sales Executive	131	13872
9	Sales Representative	783	6632

9. Distribution of performance ratings across different job roles and departments

```sql
SELECT
    JobRole,PerformanceRating,COUNT(*) AS Count
FROM employees
GROUP BY JobRole, PerformanceRating
ORDER BY  JobRole, PerformanceRating;
```

	jobrole character varying (50)	performancerating integer	count bigint
1	Healthcare Representative	3	111
2	Healthcare Representative	4	20
3	Human Resources	3	45
4	Human Resources	4	7
5	Laboratory Technician	3	217
6	Laboratory Technician	4	42
7	Manager	3	82
8	Manager	4	20
9	Manufacturing Director	3	118
10	Manufacturing Director	4	27
11	Research Director	3	72
12	Research Director	4	8
13	Research Scientist	3	243
14	Research Scientist	4	49
15	Sales Executive	3	285
16	Sales Executive	4	41
17	Sales Representative	3	71
18	Sales Representative	4	12

10. What is the Employee Retention Rate by Job Role

```sql
SELECT
    JobRole,
    COUNT(*) AS Total_Employees,
    SUM(CASE WHEN Attrition = 'No' THEN 1 ELSE 0 END) AS Retained_Employees,
   round( SUM(CASE WHEN Attrition = 'No' THEN 1 ELSE 0 END) * 100.0 / COUNT(*),2) AS Retention_Rate
FROM
    employees
GROUP BY
    JobRole
ORDER BY Retention_Rate DESC;
```

	jobrole character varying (50)	total_employees bigint	retained_employees bigint	retention_rate numeric
1	Research Director	80	78	97.50
2	Manager	102	97	95.10
3	Healthcare Representative	131	122	93.13
4	Manufacturing Director	145	135	93.10
5	Research Scientist	292	245	83.90
6	Sales Executive	326	269	82.52
7	Human Resources	52	40	76.92
8	Laboratory Technician	259	197	76.06
9	Sales Representative	83	50	60.24

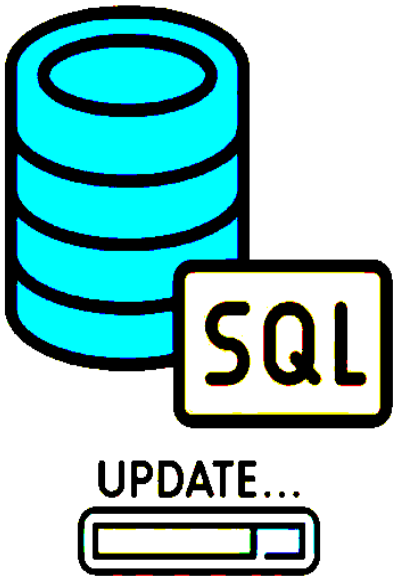
11. UPDATE Monthly income for Sales Department by 20%

```sql
--Start a transaction
BEGIN;

UPDATE Employees
SET MonthlyIncome = MonthlyIncome * 1.20
WHERE Department = 'Sales';

-- Commit the transaction
COMMIT;

select department, round(avg(monthlyincome),0) as Avg_monthly_income from employees
where department  = 'Sales'
group by department
```



Before UPDATE

Data Output | Messages | Notifications

	department character varying (255)	avg_monthly_income numeric
1	Sales	6959

After UPDATE

Data Output | Messages | Notifications

	department character varying (255)	avg_monthly_income numeric
1	Sales	8351

12. Insert a New Employee Record in existing employees table

```
INSERT INTO Employees (EmployeeID, JobRole, Attrition, Department, Age, BusinessTravel, DistanceFromHome,
                       Education, EducationField, EnvironmentSatisfaction, Gender, HourlyRate, JobInvolvement,
                       JobLevel, JobSatisfaction, MaritalStatus, MonthlyIncome, DailyRate, MonthlyRate, NumCompaniesWorked,
                       Over18,OverTime, PercentSalaryHike, PerformanceRating, RelationshipSatisfaction, StandardHours,
                       StockOptionLevel, TotalWorkingYears, TrainingTimesLastYear, WorkLifeBalance, YearsAtCompany)
VALUES (130, 'Data Scientist', 'No', 'Research & Development', 28, 'Travel_Rarely', 5, 4, 'Life Sciences',
        3, 'Female', 50, 3, 2, 4, 'Single', 6000, 1200, 25000, 2, 'Y', 'No', 12, 3, 2, 80, 0, 6, 3, 2, 4);

Select * from Employees where employeeid = 130
```

Data Output | Messages | Notifications

	employeeid [PK] integer	jobrole character varying (50)	attrition character varying (10)	department character varying (255)	age integer	businesstravel character varying (255)	distancefromhome integer	education integer	educationfield character varying (255)
1	130	Data Scientist	No	Research & Development	28	Travel_Rarely	5	4	Life Sciences

13. Delete Records of Employees Who Left the Company

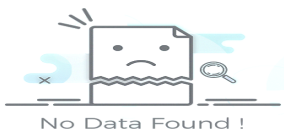
```
DELETE FROM Employees
WHERE Attrition = 'Yes';

Select * from Employees
where Attrition = 'Yes'
```

Data Output | Messages | Notifications

employeeid [PK] integer	jobrole character varying (50)	attrition character varying (10)	department character varying (255)	age integer	businesstravel character varying (255)	distancefromhome integer

An empty result shows a successful query!



14. Create a view to summarize attrition rates by department and job role

```
192
193
194 CREATE VIEW AttritionSummary AS
195 SELECT
196     Department,
197     JobRole,
198     COUNT(*) AS TotalEmployees,
199     SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS AttritionCount,
200     (SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS AttritionRate
201 FROM
202     Employees
203 GROUP BY
204     Department, JobRole;
205
```

Data Output | Messages | Notifications

```
CREATE VIEW

Query returned successfully in 455 msec.
```

VIEW

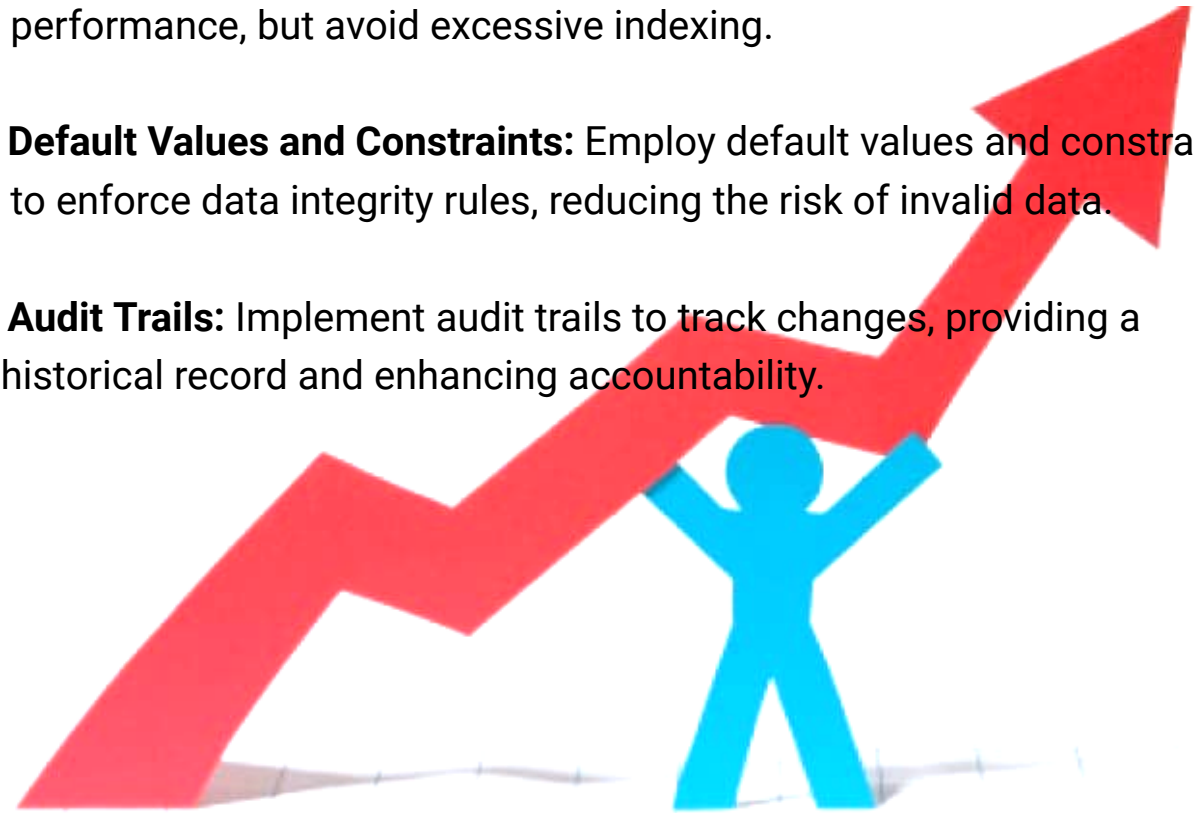
Data Output | Messages | Notifications

	department character varying (255)	jobrole character varying (50)	totalemployees bigint	attritioncount bigint	attritionrate numeric
1	Sales	Sales Executive	595	57	9.5798319327731092
2	Research & Development	Laboratory Technician	456	62	13.5964912280701754
3	Research & Development	Healthcare Representative	253	9	3.5573122529644269
4	Research & Development	Research Director	158	2	1.2658227848101266
5	Research & Development	Manufacturing Director	280	10	3.5714285714285714
6	Sales	Manager	72	2	2.7777777777777778
7	Research & Development	Research Scientist	537	47	8.7523277467411546
8	Human Resources	Human Resources	92	12	13.0434782608695652
9	Research & Development	Manager	105	3	2.8571428571428571
10	Human Resources	Manager	22	0	0.00000000000000000000
11	Sales	Sales Representative	133	33	24.8120300751879699
12	Research & Development	Data Scientist	1	0	0.00000000000000000000

15. Suggest improvements in the database schema to reduce data redundancy and improve data integrity.

Here are suggestions for improving the database schema:

- **Normalization:** Ensure the database follows normalization principles to minimize data redundancy and dependencies.
- **Foreign Keys:** Use foreign keys to establish relationships, ensuring referential integrity and preventing orphaned records.
- **Indexes:** Create indexes on frequently used columns to improve query performance, but avoid excessive indexing.
- **Default Values and Constraints:** Employ default values and constraints to enforce data integrity rules, reducing the risk of invalid data.
- **Audit Trails:** Implement audit trails to track changes, providing a historical record and enhancing accountability.



16. Explain how you can optimize the performance of SQL queries on this dataset

Here are few points for optimizing SQL queries on this dataset:

- **Indexing:** Create indexes on columns frequently used in WHERE clauses or JOIN conditions to enhance query performance.
- **Limit SELECT Columns:** Select only the necessary columns in your queries to reduce data transfer and improve efficiency.
- **Optimize WHERE Clauses:** Ensure efficient WHERE clauses by avoiding functions on indexed columns and optimizing conditions.
- **Use JOINs Efficiently:** Optimize JOIN operations by selecting the appropriate type and ensuring efficient join conditions.
- **Update Statistics Regularly:** Keep table statistics up-to-date to assist the query planner in making informed execution plans.