Employee Churn

Dataset Link :

https://www.kaggle.com/datasets/jacksonchou/hr-data-for-analytics

Description :

The table contains information about the employees in a company and whether they have left the company or not. The dataset includes variables such as age, gender, job role, performance rating, and tenure, among others. We will use PL/SQL to analyze this dataset.

Business Question:

The business question for this project is to identify the factors that contribute to employee churn and to develop a model that can predict which employees are at the highest risk of leaving the company. This will enable the company to take proactive measures to retain these employees, which can lead to increased productivity, higher morale, and improved overall organizational performance.

Sub-Questions:

1) **What are the counts of reasons that employees want to leave the company?**

```
SELECT left, COUNT(*) as cnt
FROM employee_dataset
GROUP BY left
ORDER BY cnt DESC;
```

LEFT	CNT
0	11428
1	3571

2) **What are the demographics of employees who are leaving the company?**

```
SELECT department, salary, COUNT(*)
AS num_of_employees
FROM employee_dataset
WHERE left = 1
```

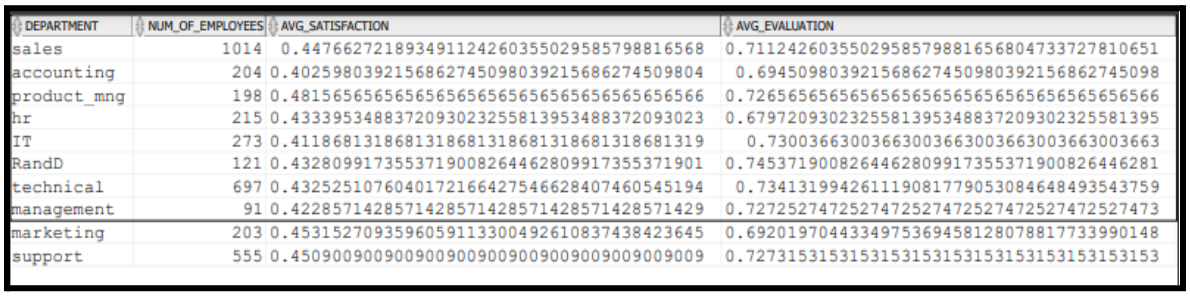
DEPARTMENT	SALARY	NUM_OF_EMPLOYEES
sales	medium	303
technical	low	378
marketing	high	9
sales	high	14
hr	low	92
product_mng	medium	87
management	low	59
marketing	low	126
accounting	medium	100
hr	medium	117
accounting	high	5
RandD	low	55
IT	high	4
IT	medium	97
product_mng	low	105
RandD	medium	62
sales	low	697
support	low	389
IT	low	172
technical	high	25
support	high	8
management	high	1
RandD	high	4
accounting	low	99
hr	high	6
technical	medium	294
support	medium	158
marketing	medium	68
management	medium	31
product_mng	high	6

```
GROUP BY department, salary;
```

This query will group the employees who left by department and salary and count the number of employees in each group. The result will show the distribution of the employees who left by department and salary.

3) **Which departments have the highest and lowest rates of employee turnover, and what factors may be contributing to these rates?**

```
SELECT Department, COUNT(*) AS num_of_employees,
AVG(satisfaction_level) AS avg_satisfaction, AVG(last_evaluation) AS
avg_evaluation
FROM employee_dataset
WHERE left = 1
GROUP BY Department;
```



DEPARTMENT	NUM_OF_EMPLOYEES	AVG_SATISFACTION	AVG_EVALUATION
sales	1014	0.44766272189349112426035502958579881656804733727810651	0.71124260355029585798816568047337278106511
accounting	204	0.40259803921568627450980392156862745098039	0.69450980392156862745098039215686274509804
product_mng	198	0.48156565656565656565656565656565656565656	0.72656565656565656565656565656565656565656
hr	215	0.43339534883720930232558139534883720930232	0.67972093023255813953488372093023255813953
IT	273	0.41186813186813186813186813186813186813191	0.73003663003663003663003663003663003663003
RandD	121	0.43280991735537190082644628099173553719010	0.74537190082644628099173553719008264462810
technical	697	0.43252510760401721664275466284074605451941	0.73413199426111908177905308464848935437590
management	91	0.42285714285714285714285714285714285714290	0.72725274725274725274725274725274725274730
marketing	203	0.45315270935960591133004926108374384236450	0.69201970443349753694581280788177339901480
support	555	0.45090090090090090090090090090090090090090	0.72731531531531531531531531531531531531530

This query will group the results by department, count the number of employees who left in each department, and calculate the average satisfaction level and last evaluation score for those employees who left. This can help provide insights into why employees in certain departments may be leaving.

4) **Are employees leaving due to dissatisfaction with their salaries, and how can we improve retention in lower salary groups?**

```
SELECT salary, COUNT(*) AS num_of_employees, AVG(satisfaction_level)
AS avg_satisfaction, AVG(last_evaluation) AS avg_evaluation
FROM employee_dataset
WHERE left = 1
GROUP BY salary;
```

SALARY	NUM_OF_EMPLOYEES	AVG_SATISFACTION	AVG_EVALUATION
low	2172	0.44124769797421731123388581952117863720 07	0.71699355432780847145488029465930018416 21
high	82	0.43451219512195121951219512195121951	0.64402439024390243902439024390243902
medium	1317	0.43854973424449506454062262718299164768 41	0.72457099468488990129081245254365983295 37

This query will group the results by salary level, count the number of employees who left in each salary group, and calculate the average satisfaction level and last evaluation score for those employees who left. This can help provide insights into whether employees are leaving due to dissatisfaction with their salary or other factors.

5) **Are there any patterns in the performance ratings of employees who are leaving the company?**

```
SELECT AVG(last_evaluation) AS avg_last_evaluation, LEFT
FROM employee_dataset
GROUP BY LEFT;
```

AVG_LAST_EVALUATION	LEFT
0.71811257350882105852702324278913469616 35	1
0.71547339866993349667483374168708435421 77	0

This query will calculate the average last evaluation score for employees who left and employees who stayed. This can help identify if there is a pattern in the performance ratings of employees who left the company.

6) **Is there a correlation between the number of years an employee has been with the company and their likelihood of leaving?**

```
SELECT time_spend_company, AVG(left) AS avg_left
FROM employee_dataset
GROUP BY time_spend_company;
```

TIME_SPEND_COMPANY	AVG_LEFT
6	0.291086350974930362116991643454038997214 5
7	0
2	0.016337854500616522811344019728729963008 6
8	0
4	0.348064137661321861556511536957371920219
5	0.565512559402579769178547182620502376103 2
10	0
3	0.246158621760049666304516529566971907496 5

This query will group employees by the number of years they have spent with the company and calculate the average likelihood of leaving for each group. This can help identify if there is a correlation between the number of years an employee has been with the company and their likelihood of leaving.

7) **What are the number of employees in the department who are leaving the company?**

```
SELECT department, COUNT(*)
AS num_of_employees
FROM employee_dataset
WHERE left = 1
GROUP BY department
ORDER BY num_of_employees
DESC;
```

	DEPARTMENT	NUM_OF_EMPLOYEES
1	sales	1014
2	technical	697
3	support	555
4	IT	273
5	hr	215
6	accounting	204
7	marketing	203
8	product_mng	198
9	RandD	121
10	management	91

8) **Create a package employee_data_pkg that contains a procedure get_employees_by_department which accepts a department name as input and returns the list of employees who work in that**

department. The employee data should include the employee ID, name, department, and salary. Define a custom record type to hold this data and a custom collection type to hold a list of these records. Use the employee_dataset table as the source of employee data.

Creating a Package:

```
CREATE OR REPLACE PACKAGE hr_churn_pkg AS
TYPE hr_churn_rec IS RECORD (
   emp_id            employee_dataset.emp_id%TYPE,
   satisfaction_level employee_dataset.satisfaction_level%TYPE,
   last_evaluation    employee_dataset.last_evaluation%TYPE,
   number_project     employee_dataset.number_project%TYPE,
   average_montly_hours
employee_dataset.average_montly_hours%TYPE,
   time_spend_company employee_dataset.time_spend_company%TYPE,
   Work_accident      employee_dataset.Work_accident%TYPE,
   left               employee_dataset.left%TYPE,
   promotion_last_5years
employee_dataset.promotion_last_5years%TYPE,
   department         employee_dataset.department%TYPE,
   salary             employee_dataset.salary%TYPE
 );
 TYPE hr_churn_tab IS TABLE OF hr_churn_rec INDEX BY PLS_INTEGER;

 PROCEDURE get_hr_churn_data (p_hr_churn_data OUT hr_churn_tab);
END hr_churn_pkg;
/
```

Output:

```
Package HR_CHURN_PKG compiled
```

9) **Create trigger employee_left_trigger that fires before each insert into the employee_dataset table. The trigger should check if the value of the left column in the newly inserted row is equal to 1, indicating that the employee has left the company. If this condition is true, the trigger should output a message to the**

console that says "Employee {employee ID} has left the company", where {employee ID} is the value of the emp_id column in the inserted row.

Creating a Trigger:

```
CREATE OR REPLACE TRIGGER hr_churn_trigger
  BEFORE INSERT ON employee_dataset
  FOR EACH ROW
BEGIN
  IF :NEW.left = 1 THEN
    dbms_output.put_line('Employee ' || :NEW.emp_id || ' has left the
company');
  END IF;
END hr_churn_trigger;
/
```

Output:

```
Trigger HR_CHURN_TRIGGER compiled
```

To Test the trigger:

```
 -- Insert a row where left column is equal to 1
INSERT INTO employee_dataset
VALUES (1234, 0.5, 0.8, 2, 150, 3, 0, 1, 0, 'Sales', 'low');
```

After executing the above code, you should see a message in the console output that says "Employee 1234 has left the company", assuming that there is no other row in the employee_dataset table with the left column equal to 1. If there are multiple rows with the left column equal to 1, you will see the message for each of them.

In the above code, we have created a package called hr_churn_pkg, which contains a record and table type for storing the HR churn data. We have also created a procedure called get_hr_churn_data, which populates the hr_churn_tab with the HR churn data from the HR Data for Analytics dataset.

We have also created a trigger called hr_churn_trigger, which fires before each insert on the HR Data for Analytics table and outputs a message if the employee has left the company.