Analysis of HR Employee Attrition using SQL

- PRADEEP K I

About Dataset

Dataset Used : HR Employee Attrition Datasets (kaggle.com)

1. **Employee Office Survey (Feedbacks):**

- Captures employee feedback data for each office location from the years 2017 to 2022.
- Includes ratings or feedback scores for employees, shedding light on their performance and satisfaction levels.
- Provides insights into the evolving trends in employee feedback over the specified timeframe.

2. **Job Position Structure:**

- Details the organizational hierarchy and job roles within different departments.
- Outlines the structure with information on departments, job levels, and specific job roles.
- Serves as a reference for understanding the organization's job hierarchy and the diversity of roles available.

3. **Office Locations (Canada and US):**

- Encompasses information about office locations, particularly highlighting 5 offices in Canada and 3 offices in the United States.
- Includes details such as office codes, city locations, provinces (or states), and countries.
- Offers a geographical perspective on the distribution of offices across North America.

4. **Employee Attrition Information:**

- Provides insights into employee attrition, including details such as leaving years, reasons for leaving, and relieving statuses.
- Helps in understanding patterns and factors contributing to employee turnover within the organization.
- Acts as a valuable resource for analyzing workforce dynamics and making informed HR decisions.

Objectives

To conduct a comprehensive case study on HR employee attrition using SQL, aiming to analyze various aspects of employee turnover within the organization. By addressing key questions such as average age of departing employees, factors influencing attrition rates, departmental variations, and demographic correlations, the study seeks to provide actionable insights for HR decision-making. Through SQL queries and data analysis, the objective is to uncover patterns, trends, and potential areas for improvement in retention strategies, ultimately aiming to enhance organizational stability and employee satisfaction.

Questionnaire:

1. Find the average age of employees who left the organization in each department.
2. List the top 5 departments with the highest average job satisfaction among employees who have not received a promotion in the last 3 years.
3. Determine the percentage of employees who left the organization due to various reasons.
4. Find the top three office location with the highest average distance from home for employees who work overtime
5. Find the employees with the highest monthly income in each department.
6. Which department has the highest and lowest attrition rates?
7. Do married employees have a lower attrition rate compared to single or divorced employees?
8. Identify the top three office locations with the highest average performance ratings for employees who joined the organization in the past three years.
9. What is the average monthly income of employees, grouped by their job level, within each department?
10. How many employees hold a degree different city, categorized by their education field

1. Find the average age of employees who left the organization in each department.

SQL QUERY

```
SELECT
	department,
	ROUND(AVG(age)) as Average_age
FROM
    hr_employee
GROUP BY
    department
```

OUTPUT

	department character varying (50)	average_age numeric
1	Delivery	36
2	HR	36
3	Sales	36
4	Corporate Functions	37
5	Product	36
6	Marketing	36

CONCLUSION: The analysis reveals a consistent average age of departure across departments, with employees leaving at an average age of 36 to 37 years. This suggests that age may not be a significant factor influencing attrition rates across different departments within the organization.

2. List the top 5 departments with the highest average job satisfaction among employees who have not received a promotion in the last 3 years

---

SQL QUERY

```
SELECT Department,
       ROUND(AVG(JobSatisfaction),2) AS AvgJobSatisfaction
FROM
     HR_Employee
WHERE
     YearsSinceLastPromotion > 3
GROUP BY
     Department
ORDER BY
     AvgJobSatisfaction DESC
LIMIT 5;
```

OUTPUT

	department character varying (50)	avgjobsatisfaction numeric
1	Product	2.35
2	HR	2.34
3	Delivery	2.29
4	Sales	2.28
5	Marketing	2.28

CONCLUSION : The top 5 departments with the highest average job satisfaction among employees who haven't received a promotion in the last 3 years are Product, HR, Delivery, Sales, and Marketing. This suggests that these departments have been successful in maintaining employee satisfaction despite promotion disparities.

3. Determine the percentage of employees who left the organization due to various reasons.

SQL QUERY

```sql
SELECT
    Reason,
    ROUND(100.0 * COUNT(*) / (SELECT COUNT(*) FROM HR_Employee WHERE Attrition = 'Yes'),2) AS Percentage
FROM
    HR_Employee
WHERE
    Attrition = 'Yes'
GROUP BY
    Reason
ORDER BY
    Percentage DESC;
```

OUTPUT

	reason character varying (100)	percentage numeric
1	retiring	9.76
2	medical issues	9.72
3	offered more money	9.38
4	return to school	9.29
5	long hours	9.26
6	reallocation out of the area	8.79
7	unhappy	8.79
8	career change	8.32
9	poor work performance	7.13
10	cost cutting	5.78
11	Layoffs	5.75
12	restructuring	5.16
13	Misconduct	1.66
14	maternity leave - did not return	0.91
15	death	0.31

CONCLUSION : From the analysis, it's evident that retiring and medical issues are the primary reasons for employee departure, each accounting for approximately 9-10% of the total attrition. Other significant factors include being offered more money, returning to school, and long hours, each contributing to roughly 9-9.5% of employee exits. Additionally, reasons such as reallocation out of the area, unhappiness, and career changes also play substantial roles, each representing around 8-9% of departures. Conversely, reasons like misconduct, maternity leave without return, and unfortunate events like death contribute to a significantly lower percentage of attrition. This breakdown highlights the multifaceted nature of employee turnover and underscores the importance of addressing various factors to mitigate attrition effectively.

4. **Find the top three office location with the highest average distance from home for employees who work overtime.**

SQL QUERY

```
SELECT
    oc.city,
    oc.province,
    oc.country,
    ROUND(AVG(e.DistanceFromHome),2) AS AvgDistanceFromHome
FROM
    HR_Employee e
INNER JOIN
    Office_code oc ON e.office_code = oc.office_code
WHERE
    e.OverTime = 'Yes'
GROUP BY
    oc.city,
    oc.province,
    oc.country
ORDER BY
    AvgDistanceFromHome DESC
LIMIT 3;
```

OUTPUT

	city character varying (50)	province character varying (50)	country character varying (50)	avgdistancefromhome numeric
1	Ottawa	ON	Canada	9.65
2	Boston	MA	US	9.22
3	Philadelphia	PA	US	9.20

CONCLUSION : Employees working overtime in Ottawa, Boston, and Philadelphia face longer average commuting distances, indicating potential impacts on work-life balance and employee well-being. This insight underscores the importance of considering geographical factors in workforce planning to optimize employee satisfaction and productivity.

5. Find the employees with the highest monthly income in each department

SQL QUERY

```sql
WITH RankedEmployees AS (
    SELECT
        EmployeeID,
        Department,
        MonthlyIncome,
        ROW_NUMBER() OVER (PARTITION BY Department ORDER BY MonthlyIncome DESC) AS Rank
    FROM
        HR_Employee
)
SELECT
    EmployeeID,
    Department,
    MonthlyIncome
FROM
    RankedEmployees
WHERE
    Rank <= 1;
```

OUTPUT

	employeeid [PK] integer	department character varying (50)	monthlyincome integer
1	100059	Corporate Functions	19973
2	100559	Delivery	19841
3	100064	HR	19953
4	100055	Marketing	19999
5	104449	Product	19871
6	100592	Sales	19938

CONCLUSION: The analysis identifies top earners within each department, showcasing variations in monthly income across organizational roles. This insight can inform salary benchmarking and compensation strategies to ensure competitiveness and equity within the workforce.

6. Which department has the highest and lowest attrition rates?

SQL QUERY

```
SELECT
    Department,
    COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) AS Attrition_Count,
    COUNT(*) AS Total_Count,
    (100 * COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) / COUNT(*))  AS Attrition_Rate
FROM
    HR_Employee
GROUP BY
    Department
ORDER BY
    Attrition_Rate DESC;
```

OUTPUT

	department character varying (50)	attrition_count bigint	total_count bigint	attrition_rate bigint
1	Sales	573	2277	25
2	Corporate Functions	577	2321	24
3	Delivery	523	2218	23
4	Product	537	2239	23
5	HR	496	2187	22
6	Marketing	492	2181	22

CONCLUSION : The department with the highest attrition rate is Sales with 25%, while HR and Marketing departments tie for the lowest attrition rate at 22%. This data suggests a potential need for targeted retention strategies in Sales to mitigate turnover, while HR and Marketing may benefit from analyzing factors contributing to their comparatively lower attrition rates for potential replication across other departments.

7. Do married employees have a lower attrition rate compared to single or divorced employees?

SQL QUERY

```
SELECT
    MaritalStatus,
    COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) AS Attrition_Count,
    COUNT(*) AS Total_Count,
    (COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) * 100 / COUNT(*))  AS Attrition_Rate
FROM
    HR_Employee
GROUP BY
    MaritalStatus
ORDER BY
   Attrition_Rate DESC;
```

OUTPUT

	maritalstatus character varying (20)	attrition_count bigint	total_count bigint	attrition_rate bigint
1	Single	1622	4137	39
2	Married	1276	7043	18
3	Divorced	300	2243	13

CONCLUSION : Married employees exhibit a notably lower attrition rate at 18%, compared to single employees at 39% and divorced employees at 13%. This suggests that marital status may indeed correlate with attrition rates, with married individuals demonstrating higher retention within the organization.

8. **Identify the top three office locations with the highest average performance ratings for employees who joined the organization in the past three years.**

SQL QUERY

```sql
WITH Employee_Performance AS (
    SELECT
        e.emp_id,
        e.off_cde,
        e.rated_year,
        e.rating,
        ROW_NUMBER() OVER (PARTITION BY e.emp_id ORDER BY e.rated_year DESC) AS row_num
    FROM
        Employee_office_survey e
    WHERE
        e.rated_year BETWEEN 2020 AND 2022
)
SELECT
    oc.city,
    oc.province,
    oc.country,
    ROUND(AVG(ep.rating)::numeric,2) AS avg_performance_rating
FROM
    Office_code oc
JOIN
    Employee_Performance ep ON oc.office_code = ep.off_cde
WHERE
    ep.row_num = 1
GROUP BY
    oc.city,
    oc.province,
    oc.country
ORDER BY
    avg_performance_rating DESC
LIMIT 3;
```

OUTPUT

	city character varying (50)	province character varying (50)	country character varying (50)	avg_performance_rating numeric
1	Ottawa	ON	Canada	8.94
2	Philadelphia	PA	US	8.93
3	Markham	ON	Canada	8.89

CONCLUSION : The top three office locations with the highest average performance ratings for employees who joined the organization in the past three years are Ottawa, ON, Canada with an average rating of 8.94, followed by Philadelphia, PA, US with an average rating of 8.93, and Markham, ON, Canada with an average rating of 8.89. This indicates that employees in these locations, particularly those who recently joined, tend to demonstrate higher performance levels compared to other office locations.

9. What is the average monthly income of employees, grouped by their job level, within each department?

SQL QUERY

```
SELECT
    department,
    ROUND(AVG(CASE WHEN joblevel_updated = 'L1' THEN monthlyincome ELSE 0 END ),2) AS Joblevel_L1,
    ROUND(AVG(CASE WHEN joblevel_updated = 'L2' THEN monthlyincome ELSE 0 END ),2)AS Joblevel_L2,
    ROUND(AVG(CASE WHEN joblevel_updated = 'L3' THEN monthlyincome ELSE 0 END ),2) AS Joblevel_L3,
    ROUND(AVG(CASE WHEN joblevel_updated = 'L4' THEN monthlyincome ELSE 0 END ),2) AS Joblevel_L4,
    ROUND(AVG(CASE WHEN joblevel_updated = 'L5' THEN monthlyincome ELSE 0 END ),2) AS Joblevel_L5,
    ROUND(AVG(CASE WHEN joblevel_updated = 'L6' THEN monthlyincome ELSE 0 END ),2) AS Joblevel_L6,
    ROUND(AVG(CASE WHEN joblevel_updated = 'L7' THEN monthlyincome ELSE 0 END ),2) AS Joblevel_L7
FROM
   Hr_employee
GROUP BY
   department
```

OUTPUT

department character varying (50)	joblevel_l1 numeric	joblevel_l2 numeric	joblevel_l3 numeric	joblevel_l4 numeric	joblevel_l5 numeric	joblevel_l6 numeric	joblevel_l7 numeric
Delivery	1286.10	1967.20	1292.67	1027.43	667.47	0.00	0.00
HR	1333.63	2045.98	1287.48	826.78	667.02	0.00	0.00
Sales	1291.68	1994.38	1341.96	912.02	712.64	0.00	0.00
Corporate Functions	1210.64	1886.88	1137.81	846.19	446.64	1335.09	49.16
Product	1318.89	1938.09	1305.93	1003.57	642.41	0.00	0.00
Marketing	1250.80	2069.78	1316.48	866.59	726.40	0.00	0.00

CONCLUSION : The analysis reveals variations in average monthly income across different job levels within each department. Generally, higher job levels correspond to higher average incomes, as seen in the incremental pattern from L1 to L5 across departments. However, there are exceptions, such as in Corporate Functions, where L6 appears to have a notably higher average income compared to L4 and L5, suggesting unique compensation structures or roles within that department.

10. How many employees hold a degree different city, categorized by their education field

SQL QUERY

```
SELECT
   oc.city,
   oc.country,
   COUNT(CASE WHEN hr.educationfield = 'Diploma' THEN 1 ELSE 0 END ) AS Diploma,
   COUNT(CASE WHEN hr.educationfield = 'Doctorate' THEN 1 ELSE 0 END ) AS Doctorate,
   COUNT(CASE WHEN hr.educationfield = 'Bachelors' THEN 1 ELSE 0 END ) AS Bachelors,
   COUNT(CASE WHEN hr.educationfield = 'Masters' THEN 1 ELSE 0 END ) AS Masters
FROM
   HR_employee hr
INNER JOIN
   office_code oc ON hr.office_code = oc.office_code
GROUP BY
   oc.city,
   oc.country
ORDER BY
   oc.country ASC
```

OUTPUT

city character varying (50)	country character varying (50)	diploma bigint	doctorate bigint	bachelors bigint	masters bigint
Toronto	Canada	1576	1576	1576	1576
Markham	Canada	1608	1608	1608	1608
Ottawa	Canada	1473	1473	1473	1473
Vancouver	Canada	1648	1648	1648	1648
Calgary	Canada	1699	1699	1699	1699
NewYork	US	1794	1794	1794	1794
Boston	US	1852	1852	1852	1852
Philadelphia	US	1773	1773	1773	1773

CONCLUSION: The output indicates the count of employees holding different degrees categorized by their education field in various cities and countries. Each city has the same count of employees across all degree categories (diploma, doctorate, bachelors, and masters). This suggests uniformity in educational qualifications among employees across different cities and countries, reflecting a consistent hiring or educational attainment pattern within the organization.