My documentation and insights on the project, “DATA ANALYSIS USING SQL”

Note:

The database name at the workbench is data\_analysis\_using\_sql while the table name that holds the dataset is 'dataset for data analysis using sql'. This table was imported into the MYSQL workbench from Microsoft Excel environment having saved it as a CSV file.

## Analysis using SQL queries:

The database table can be viewed: SELECT \* FROM data\_analysis\_using\_sql.`hr\_analytics`;

**Total number of records in the specified table:**

1. SELECT COUNT(\*) AS total\_count FROM `data\_analysis\_using\_sql`.`hr\_analytics`;

*Insight:*

The total number of records, that is data in rows, were fetched as 1423 in a new column, total\_count because of the use of “AS” statement in the sql query.

**Viewing all the columns in the specified table:**

1. SHOW COLUMNS FROM `data\_analysis\_using\_sql`.`hr\_analytics`;

*Insight:*

This query retrieves information about the columns (fields) in the specified table. It returned a result set with columns like Field, Type, Null, Key, Default and Extra, providing details about each column in the table. The Field column contains the names of the columns in the table, it returned 44 columns.

**Count the Total Number of Employees:**

1. SELECT COUNT(\*) AS total\_employees FROM `data\_analysis\_using\_sql`.`hr\_analytics`;

*Insight:*

The total number of employees returned from the table in the database was 1423.

**Count Employees by Gender:**

1. SELECT gender, COUNT(\*) AS employee\_by\_gender\_count FROM `data\_analysis\_using\_sql`.`hr\_analytics` GROUP BY gender;

*Insight:*

This query returned 'Female', '569' and 'Male', '854'

**NOTE: This correlates with the data I had when I did data analysis on same dataset using excel.**

**Calculate Average Age of Employees:**

1. SELECT AVG(age) AS average\_age FROM `data\_analysis\_using\_sql`.`hr\_analytics`;

*Insight:*

The query returned average age of employees as '36.9284'

**Count Employees by Department:**

1. SELECT department, COUNT(\*) AS num\_employees FROM `data\_analysis\_using\_sql`.`hr\_analytics` GROUP BY department;

*Insight:*

This query returned 'Sales', '431', 'R&D', '924' and 'HR', '62'. Indicating that there are 3 departments.

**NOTE: This correlates with the data I had when I did data analysis on same dataset using excel.**

**Find the Highest Salary:**

1. SELECT MAX(MonthlyIncome) AS highest\_salary FROM `data\_analysis\_using\_sql`.`hr\_analytics`;

SELECT MIN(MonthlyIncome) AS lowest\_salary FROM `data\_analysis\_using\_sql`.`hr\_analytics`;

*Insight:*

The query returned '19999' as highest salary. The reverse which is MIN for lowest salary returned ‘1009’.

**Identify Employees with the Longest Tenure:**

1. SELECT \* FROM `data\_analysis\_using\_sql`.`hr\_analytics` ORDER BY YearsAtCompany DESC LIMIT 3;

**Insight:**

**The query returned 40, 37 and 36 years as first three longest tenures in descending order.**

**Calculate Average Years of Experience:**

1. SELECT AVG(TotalWorkingYears) AS Years\_of\_experience FROM `data\_analysis\_using\_sql`.`hr\_analytics`;

**Insight:**

**The average years of experience of an employee in the said organization was approximately 11years.**

**Count Employees by Education Level:**

1. SELECT EducationField, COUNT(\*) AS num\_employees FROM `data\_analysis\_using\_sql`.`hr\_analytics` GROUP BY EducationField;

**Insight:**

**'Life Sciences', '586', 'Medical', '446', 'Marketing', '155', 'Technical Degree', '129''Other', '81'**

**Count Employees by Marital Status:**

1. SELECT MaritalStatus, COUNT(\*) AS num\_employees FROM `data\_analysis\_using\_sql`.`hr\_analytics` GROUP BY MaritalStatus;

**Insight:**

**'Single', '451', 'Divorced', '317', 'Married', '655'**

**Calculate Employee Turnover Rate (% of Employees Who Left):**

1. SELECT (COUNT(\*) \* 100.0 / (SELECT COUNT(\*) FROM `data\_analysis\_using\_sql`.`hr\_analytics`))

AS turnover\_rate FROM `data\_analysis\_using\_sql`.`hr\_analytics` WHERE attrition = 'Yes';

**Insight:**

**The turnover rate was '16.16304'**

**In this case, the calculated turnover rate was '16.12245', which means that approximately 16.12% of the employees in the dataset have left the company (attrition) based on the data where 'Attrition' is marked as 'Yes'.**

**The benefit of calculating turnover rates can be significant for businesses. Understanding turnover rates helps organizations:**

**Identify Issues:** High turnover rates might indicate underlying problems in the workplace, such as dissatisfaction, lack of engagement, or issues with management.

**Retention Strategies:** By analyzing turnover rates, organizations can implement strategies to retain employees, such as improving work conditions, offering training, or providing better benefits.

**Financial Planning:** High turnover rates can impact financial planning. Hiring and training new employees are costly processes, so reducing turnover can save the company money.

**Employee Satisfaction:** Monitoring turnover rates can gauge the effectiveness of employee satisfaction initiatives. Low turnover rates can indicate a content and stable workforce.

**Predictive Insights:** Analyzing turnover rates over time can provide predictive insights. If rates are rising, it might indicate future challenges in retaining talent.

Understanding turnover rates and acting on this information can lead to better employee satisfaction, lower costs, and a more stable workforce, ultimately benefiting the organization in the long run.

**Note:**

A**ttrition** refers to the gradual reduction in the number of employees or participants in an organization or group as they leave or retire and are not replaced immediately. It's a natural process that occurs over time as people move on to other opportunities, retire, or leave for various reasons like finding a new job, retirement, or personal circumstances.

In the context of a workplace, high attrition might indicate that many employees are leaving the company, which can be a concern for employers. It often prompts organizations to assess their work environment, employee benefits, and job satisfaction levels to understand why employees are leaving and to take steps to improve retention.

**Turnover**, in a business context, refers to the rate at which employees leave a company and are replaced by new employees. It's often expressed as a percentage and is calculated by dividing the number of employees who leave the company (usually within a specific period) by the average number of total employees during the same period.

**Identify Employees with Low Job Satisfaction Given a Benchmark of 2 and below:**

1. SELECT JobSatisfaction FROM data\_analysis\_using\_sql.hr\_analytics;

SELECT JobSatisfaction FROM data\_analysis\_using\_sql.hr\_analytics WHERE JobSatisfaction <= 2;

SELECT EmpID, JobSatisfaction FROM data\_analysis\_using\_sql.hr\_analytics WHERE JobSatisfaction <= 2;

SELECT EmpID, JobSatisfaction, COUNT(\*) AS TotalEmployees FROM data\_analysis\_using\_sql.hr\_analytics WHERE JobSatisfaction <= 2 GROUP BY EmpID, JobSatisfaction;

SELECT EmpID, JobSatisfaction, COUNT(\*) AS TotalEmployees FROM data\_analysis\_using\_sql.hr\_analytics WHERE JobSatisfaction <= 2 GROUP BY EmpID, JobSatisfaction UNION SELECT NULL, NULL, COUNT(\*) AS TotalEmployees FROM data\_analysis\_using\_sql.hr\_analytics WHERE JobSatisfaction <= 2;

**Insight:**

**From line 16 to 27 in the SQL script, we could observe 4 blocks of SQL query in tackling this technical analysis. The reason is that, one line of SQL query will not in any way lead us to what we are looking for. The last block of SQL query deduced that there are 554 employees whose job satisfaction level were at the benchmark of 2 and below. If the organization was to retrench employees, I think 554 would have been retrenched.**

**Identify Employees Who Left the Company (Attrition Analysis):**

1. SELECT \* FROM data\_analysis\_using\_sql.hr\_analytics WHERE attrition = 'Yes';

SELECT COUNT(\*) AS TotalAttrition FROM data\_analysis\_using\_sql.hr\_analytics WHERE attrition = 'Yes';

**Insight:**

**The first query fetches the records of employee that have left the organization, while the second gives us the count as 230.**

**Calculate Average Salary by Department:**

1. SELECT department, AVG(MonthlyIncome) AS avg\_salary FROM data\_analysis\_using\_sql.hr\_analytics GROUP BY department;

**Insight:**

**'Research & Development', '6313.2348', 'Sales', '6942.4348', 'Human Resources', '6659.4677'**

**Identify Employees Who Were Not Promoted:**

1. SELECT \* FROM data\_analysis\_using\_sql.hr\_analytics WHERE YearsSinceLastPromotion = 0;

SELECT \*, (SELECT COUNT(\*) FROM data\_analysis\_using\_sql.hr\_analytics WHERE YearsSinceLastPromotion = 0) AS TotalZeroPromotion

FROM data\_analysis\_using\_sql.hr\_analytics WHERE YearsSinceLastPromotion = 0;

**Insight:**

**The first query return all records of employees that have not being promoted while on the column “**YearsSinceLastPromotion**” if we check manually we will see that some employee were last promoted since 17years ago, others since 12 or 9 years ago, this suggests for the turnover rate of 16.12% we had previously, how can it be since 17 years an employee was last promoted in an organization.**

**The second query above has a subquery that performs a count on all employees whose “**YearsSinceLastPromotion**” equal zero and creates a new column, “**TotalZeroPromotion**” and from the output, we can see 560 employees have not being promoted for about 17 years since the highest in the column “**YearsSinceLastPromotion**” reads 17, so the subquery and its preceeding query works hand in hand.**

**Count Employees by Job Role and Gender:**

1. SELECT job\_role, gender, COUNT(\*) AS num\_employees FROM data\_analysis\_using\_sql.hr\_analytics GROUP BY job\_role, gender;

**Insight:**

**I love this query, see what it did:**

**It shows one can group by more than two columns using SQL queries.**

|  |  |  |
| --- | --- | --- |
| **Job Role** | **Gender** | **Num\_Employees** |
| Laboratory Technician | Male | 171 |
| Sales Representative | Female | 39 |
| Sales Representative | Male | 44 |
| Research Scientist | Male | 167 |
| Laboratory Technician | Female | 79 |
| Research Scientist | Female | 109 |
| Human Resources | Male | 35 |
| Manufacturing Director | Male | 71 |
| Manufacturing Director | Female | 70 |
| Sales Executive | Female | 130 |
| Healthcare Representative | Male | 78 |
| Sales Executive | Male | 188 |
| Human Resources | Female | 16 |
| Healthcare Representative | Female | 49 |
| Research Director | Male | 45 |
| Research Director | Female | 32 |
| Manager | Female | 45 |
| Manager | Male | 55 |

**Identify Employees Who Left with Low Performance Ratings:**

1. Select \* from data\_analysis\_using\_sql.hr\_analytics where Attrition= 'Yes' and PerformanceRating <=2;

**Insight:**

**Zero row was returned reasons being that there was no employee that had** PerformanceRating <=2, I found out that when one criterion does not meet its condition though other criteria met their conditions, the query returns zero result, because of the “And” clause. Same is seen here:

select Attrition, PerformanceRating, Count(\*) from data\_analysis\_using\_sql.hr\_analytics where PerformanceRating <=2 and Attrition = 'Yes'

group by Attrition, PerformanceRating;

**zero results were returned.**

But in this case: select Attrition, PerformanceRating, Count(\*) from data\_analysis\_using\_sql.hr\_analytics where PerformanceRating <=3 and Attrition = 'Yes'

group by Attrition, PerformanceRating; see what was returned:

|  |  |  |
| --- | --- | --- |
| Attrition | PerformanceRating | Count(\*) |
| Yes | 3 | 196 |

**NOTE:**

**There is no employee with low performance rating if performance rating is rated on a scale of 1 to 4 with low performance rated as 2 and below. Yet even at a performance rating of 3, we still had 196 Attrition. This points to several other factors that could be responsible for employees leaving the organization.**