**HR dataset documentation**

**Project: Employees Data Analysis**

**Date:** 1/10/2024  
**Prepared by:** Team 3 (Youssef, Nadeem, Rana, Islam)  
**Objective:** Clean and transform the dataset to ensure data consistency and accuracy for analysis. Analize dataset to get insights and create dynamic dashboard to get data-driven decisions.

**Introduction:**

* **Overview:**
* Company information about employees describes each employee, his education level, name, gender, salary, performance, years at company, satisfaction, attrition……
* five tables (EducationLevel, Employee, PerformanceRating, RatingLevel, SatisfiedLevel)

**Tables Description:**

* **EducationLevel**:

|  |  |  |
| --- | --- | --- |
| **Column name** | **Data type** | **Description** |
| **EducationLevel** | String | The level of education an employee has attained. |
| **EducationLevelID** | Integer | A unique identifier for each education level. |

* **Employee:**

|  |  |  |
| --- | --- | --- |
| **Column name** | **Data type** | **Description** |
| **EmployeeID** | Text | A unique identifier for each employee. |
| **FirstName** | Text | The first name of employee. |
| **LastName** | Text | The last name of the employee. |
| **Gender** | Text | The gender of the employee. |
| **Age** | Integer | The age of the employee. |
| **BusinessTravel** | Text | Indicates the business travels for each employee. |
| **Department** | Text | The department the employee works in. |
| **DistanceFromHome (KM)** | Integer | The distance in kilometers between the employee’s home and the workplace. |
| **State** | Text | The state in which the employee works. |
| **Ethnicity** | Text | The employee’s reported ethnicity (e.g., White, Asian, African American). |
| **EducationLevelID** | Integer | A numeric representation of the employee’s education level. |
| **EducationField** | Text | The field of study for the employee’s highest level of education. |
| **JobRole** | Text | The specific job role or title of the employee. |
| **MaritalStatus** | Text | The marital status of the employee. |
| **Salary** | Integer | The employee’s salary in the given time period. |
| **StockOptionLevel** | Integer | Indicates the employee’s level of stock options. |
| **OverTime** | Text | Whether or not the employee works overtime. |
| **HireDate** | Date | The date the employee was hired. |
| **Attrition** | Text | Whether the employee has left the company. |
| **Actual years at company** | Decimal number | The actual number of years the employee has worked at the company. |
| **YearsAtCompany** | Integer | The number of years the employee has worked at the company. |
| **YearsInMostRecentRole** | Integer | The number of years the employee has been in their current role. |
| **YearsSinceLastPromotion** | Integer | The number of years since the employee’s last promotion. |
| **YearsWithCurrManager** | Integer | The number of years the employee has been working with their current manager. |

* **PerformanceRating:-**

Typically contains structured information related to employee performance evaluations. This table is essential for tracking and analyzing how employees are performing in their roles over time.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Column Type** | **Description** |
| **Performance ID** | Text | is a unique identifier assigned to each performance rating for employees |
| **Employee ID** | Text | is a unique identifier assigned to each employee within the organization. |
| **Review Date** | Date | refers to the date on which an employee's performance evaluation or review takes place. |
| **Environment Satisfaction** | Number | measures employees' contentment with their workplace environment. |
| **Job Satisfaction** | Number | measures employees' overall contentment and happiness with their job roles. |
| **Relationship Satisfaction** | Number | measures employees' contentment with their interpersonal relationships at work. |
| **Training Opportunities Within Year** | Number | typically tracks the availability and participation of employees in training and development programs within a specific year. |
| **Training Opportunities Taken** | Number | typically indicates whether employees have participated in any training programs or professional development opportunities offered by the company. |
| **Work Life Balance** | Number | typically reflects employees' perceptions and experiences regarding their ability to balance work responsibilities with personal life |
| **Self Rating** | Number | typically captures employees' assessments of their own performance, skills, or competencies. |
| **Manager Rating** | Number | typically reflects the assessments made by managers regarding their employees' performance, skills, and contributions. |

* **RatingLevel:-**

Typically defines the criteria and categories used to evaluate employee performance. This table serves as a reference for how performance ratings are assigned and helps ensure consistency in evaluations across the organization.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Column Type** | **Description** |
| **Rating ID** | Number | typically serves as a unique identifier for each performance rating record. |
| **Rating Level** | Text | typically categorizes the performance ratings assigned to employees. |

* **SatisfiedLevel:-**

Typically contains information about employee satisfaction or engagement levels within the organization. This table helps HR departments understand how employees feel about their work environment, job roles, and overall organizational culture.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Column Type** | **Description** |
| **Satisfaction ID** | Number | typically serves as a unique identifier for each record related to employee satisfaction or engagement assessments. |
| **Satisfaction Level** | Text | typically captures employees' assessments of their overall job satisfaction or engagement within the organization. |

**Cleaning overview**

This part of documentation outlines the steps taken to clean and transform the HR dataset. The focus was on standardizing values in the **EducationField** column, correctly transforming the **HireDate** column to the appropriate date format, and creating a calculated column, **Actual years at company**, with decimal precision. Each transformation step is described in detail below.

**1. Standardizing the "Marketing" Values in the EducationField Column**

* **Column:** EducationField
* **Issue Identified:** The **EducationField** column contained two similar but inconsistent values for "Marketing"—one with an extra trailing space.
  + Example: "Marketing " and "Marketing"



* **Cleaning Action:**
  + The value "Marketing " (with an extra space) was replaced with "Marketing" to ensure consistency across the dataset.
* **Approach:**
  + Applied a transformation to replace "Marketing " with "Marketing".

A screenshot of a computer

Description automatically generated

* **Code**:

= Table.ReplaceValue(#"Previous Step", "Marketing ", "Marketing", Replacer.ReplaceText, {"EducationField"})

**2. Correcting the "HireDate" Format**

* **Column:** HireDate
* **Issue Identified:** The **HireDate** column had dates in the **MM/DD/YYYY** format, which could not be directly converted to the **DD/MM/YYYY** format needed for the analysis.
* **Cleaning Action:**
  + The **HireDate** column was split into **three separate columns** for the month, day, and year using the “/” delimiter.
  + After splitting, the **order** of the columns was rearranged to **DD/MM/YYYY** format.
  + Finally, the three columns were **merged back** together into a single date column in the correct format.

A screenshot of a computer

Description automatically generated

* **Steps Taken:**
  + **Split the HireDate** column by the / delimiter to create three separate columns: Month, Day, and Year.
  + **Rearranged the columns** from Month-Day-Year to Day-Month-Year.
  + **Merged the columns back** into one column in the DD/MM/YYYY format.
  + Transformed the merged column to the **Date** data type.
* **Code**:

// Step 1: Split the column

Table.SplitColumn(#"Previous Step", "HireDate", Splitter.SplitTextByDelimiter("/", QuoteStyle.Csv), {"Month", "Day", "Year"})

// Step 2: Reorder and merge the columns

Table.CombineColumns(#"Previous Step", {"Day", "Month", "Year"}, Combiner.CombineTextByDelimiter("/", QuoteStyle.Csv), "HireDate")

// Step 3: Convert to Date type

Table.TransformColumnTypes(#"Previous Step", {{"HireDate", type date}})

**3. Creating the "Actual years at company" Column**

* **Calculated Column:** Actual years at company
* **Objective:** To calculate the exact tenure (in decimal years) for employees with no attrition, while retaining the value from the **YearsAtCompany** column for employees with attrition.

A screenshot of a computer

Description automatically generated

* **Steps Taken:**
  1. Created a new calculated column, **"Actual years at company"**, based on the following logic:
     + If **Attrition** = "No", calculate the number of years from **HireDate** to the current date (31/12/2022).
     + If **Attrition** = "Yes", use the existing value from the **YearsAtCompany** column.
  2. Formula used for this calculation:

Table.AddColumn(#"Change Type", "Actual years at company", each if [Attrition] = "No" then Duration.Days(#date(2022, 12, 31) - [HireDate]) / 365 else [YearsAtCompany])

**4. Rounding "Actual years at company" and Changing Data Type**

* **Objective:** Round the **Actual years at company** column to **2 decimal places** for consistency and readability, and then change its type to **Decimal Number**.
* **Steps Taken:**
  1. Applied rounding to the **Actual years at company** column to limit the values to 2 decimal places.
  2. Changed the column type to **Decimal Number** for accurate numerical representation and calculations.

**Code:**

Table.TransformColumns(#"Previous Step", {{"Actual years at company", each Number.Round(\_, 2), type number}})

A screenshot of a computer

Description automatically generated

**Summary of Transformations:**

* **Standardized** the "Marketing" values in the **EducationField** column by replacing "Marketing " with "Marketing".
* **Corrected** the **HireDate** format by splitting, rearranging, and merging date components to follow the **DD/MM/YYYY** structure.
* **Created** a new calculated column, **"Actual years at company"**, to represent employees' tenure in decimal years.
* **Rounded** the calculated values to 2 decimal places and **converted** the column type to **Decimal Number** for precision.

**Data modeling:A screenshot of a computer

Description automatically generated**

**Next Steps:**

With the dataset now cleaned and standardized, further analysis and insights can be conducted based on accurate and reliable data.

**Analysis and insights:**

1. **Performance and productivity vs salary and promotions.**

* **Is there a correlation between salary, performance ratings and years at company?**
  + Understand if higher salaries lead to higher performance, and if longer employment periods lead to higher salaries.
* **Are employees with higher performance closer to have promotions?**
  + Search the relation between performance rating and promotions.
* **What is the effect of marital status on employee performance, travelling and overtime?**
  + See how marital status affect employee performance, travelling and overtime.

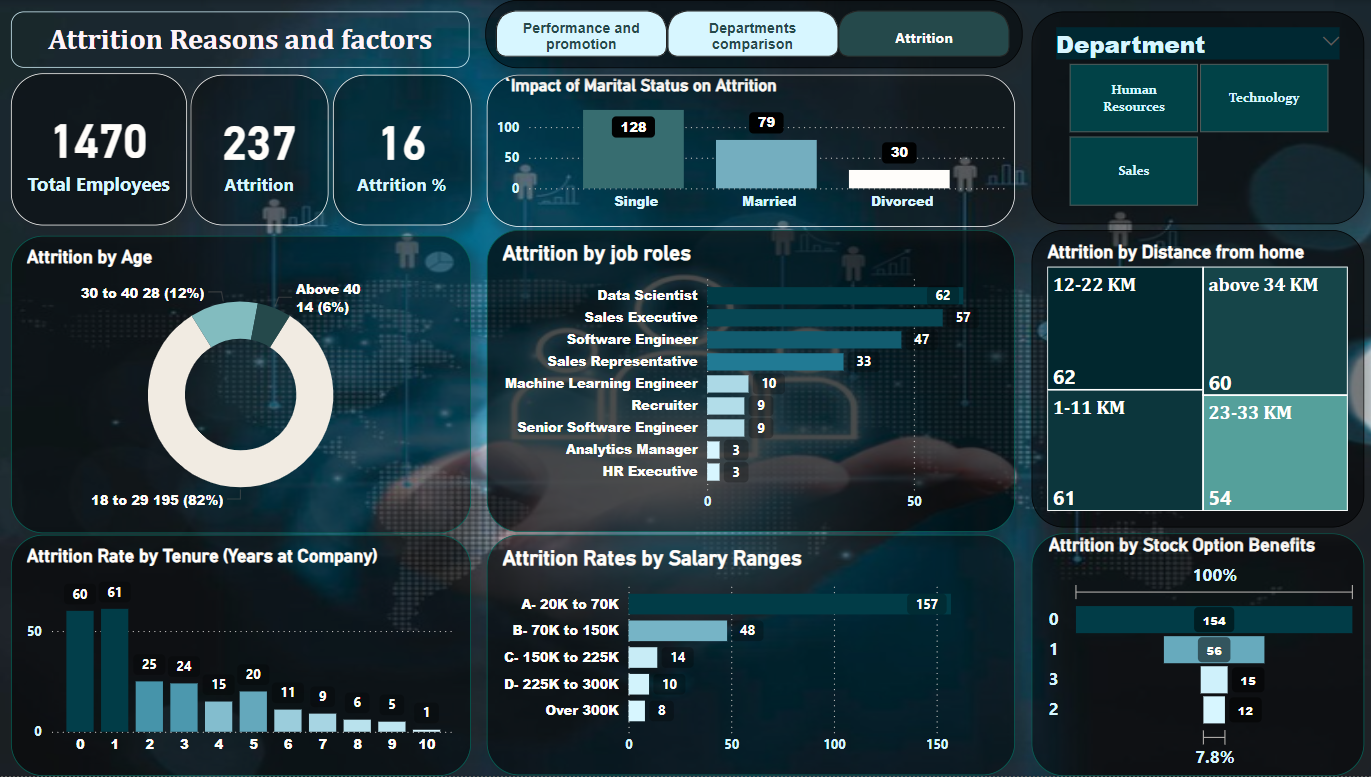
1. **Departments comparison**

* **What are the differences in job satisfaction and attrition across different departments?**
  + Compare satisfaction levels and attrition to identify areas needing attention.
* **What are the differences between departments across average salaries, average employees rating and education levels?**

1. **Attrition factors and reasons.**

* **What is the attrition rate by age, gender, and marital status?**
  + Understanding which areas have higher turnover can help target retention efforts.
* **What are the key factors (e.g., stoke option levels, distance from home, salary) influencing employee attrition?**
* **Is there a correlation between tenure (years at company) and attrition?**
  + Investigate whether longer tenure is associated with lower attrition rates.
* **Do employees with higher salaries have lower attrition rates?**
  + Examine if competitive compensation reduces turnover.

**Visualization and dashboards:**

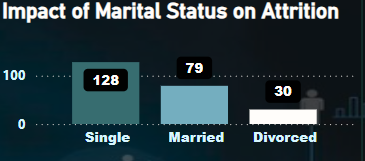


* The **main goal** of this dashboard is to analyze and understand the **factors contributing to employee attrition** within the company. By visualizing different dimensions such as **age, gender, tenure, salary, marital status, distance from home**, and **stock option benefits**, the dashboard aims to:
* **Identify patterns and trends** in employee turnover.
* **Pinpoint key factors** influencing why employees leave.
* **Compare attrition across different departments** to see if specific areas of the organization are more affected.
* **Support decision-making** for HR and management teams to implement targeted retention strategies based on the insights gained.
* Overall, the dashboard helps stakeholders **reduce employee attrition** by providing actionable data and identifying areas for improvement in terms of employee satisfaction, compensation, and other organizational factors.



**The KPIs:**

1. **Total Employees (1470)**: This is the total number of employees in the company.
2. **Attrition (237)**: The number of employees who have left the company.
3. **Attrition % (16%):** This percentage represents the proportion of employees who have left the company relative to the total workforce.



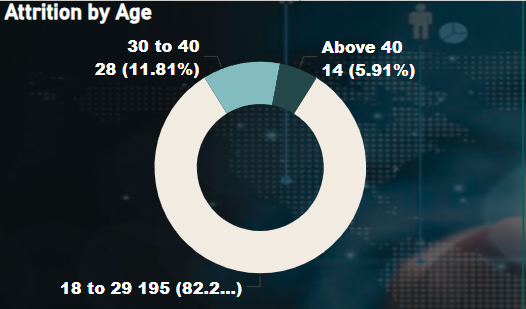
**Impact of Marital Status on Attrition**

* This bar chart highlights how marital status affects attrition:
  + **Single (128)**: The highest number of people leaving were single.
  + **Married (79)**: Fewer married people left compared to single employees.
  + **Divorced (30)**: The smallest group among those who left.

A screenshot of a computer screen

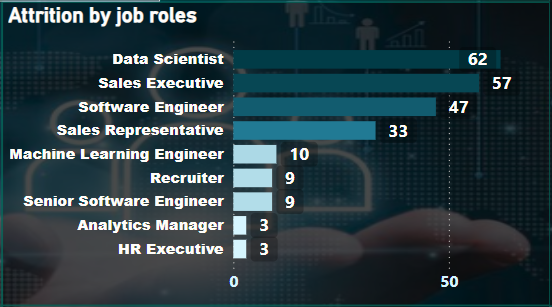
Description automatically generated

We used Department as a filter



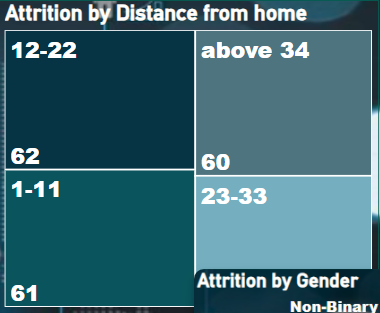
**Attrition by Age**

* This donut chart breaks down the employees who left by age group:
  + **18 to 29 (82.2%)**: The largest group of employees leaving the company.
  + **30 to 40 (11.81%)**: A smaller but still significant portion.
  + **Above 40 (5.91%)**: The smallest group contributing to the attrition.

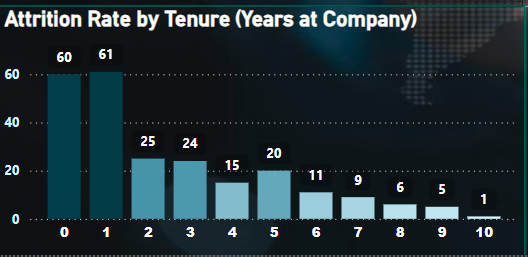


**Attrition by job roles**

* This horizontal bar chart shows attrition rates by job roles:
  + **Data Scientist (62)**: Highest attrition rate and then **Sales Excutive (57), Software Engineer (47), Sales Representative (33).**
  + **medium attrition rates:** Nearly the same **Machine Learning Engineer(10), Recruiter(9)** and **Senior software Engineer(9)**.
  + **Low attrition rates:** lowest of attrition is in job role **Analytics Manager(3)** and **HR Executive(3).**

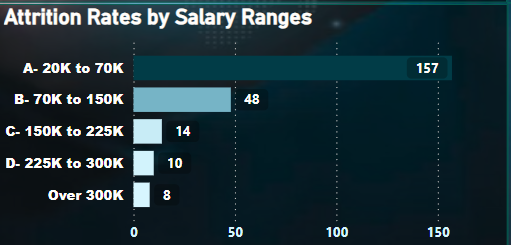
  
**Attrition by Distance from Home**

* This treemap visualizes attrition based on the employee's distance from home:
  + **12-22 (62)** and **1-11 (61)**: A significant number of employees leaving live within these distance ranges.
  + **Above 34 (60)** and **23-33 (54)**: The remaining employees live further away, and there’s still considerable attrition from these groups.



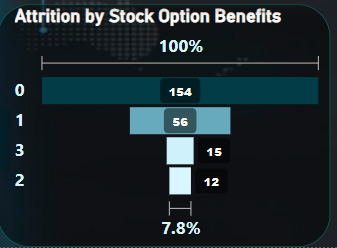
**Attrition by Tenure (Years at Company)**

* A bar chart showing how many employees leave based on how long they've been with the company:
  + **0 Years (60)** and **1 Year (61)**: The highest attrition occurs within the first two years of employment.
  + The numbers gradually decrease as tenure increases, though there is a spike in the 5-year mark.



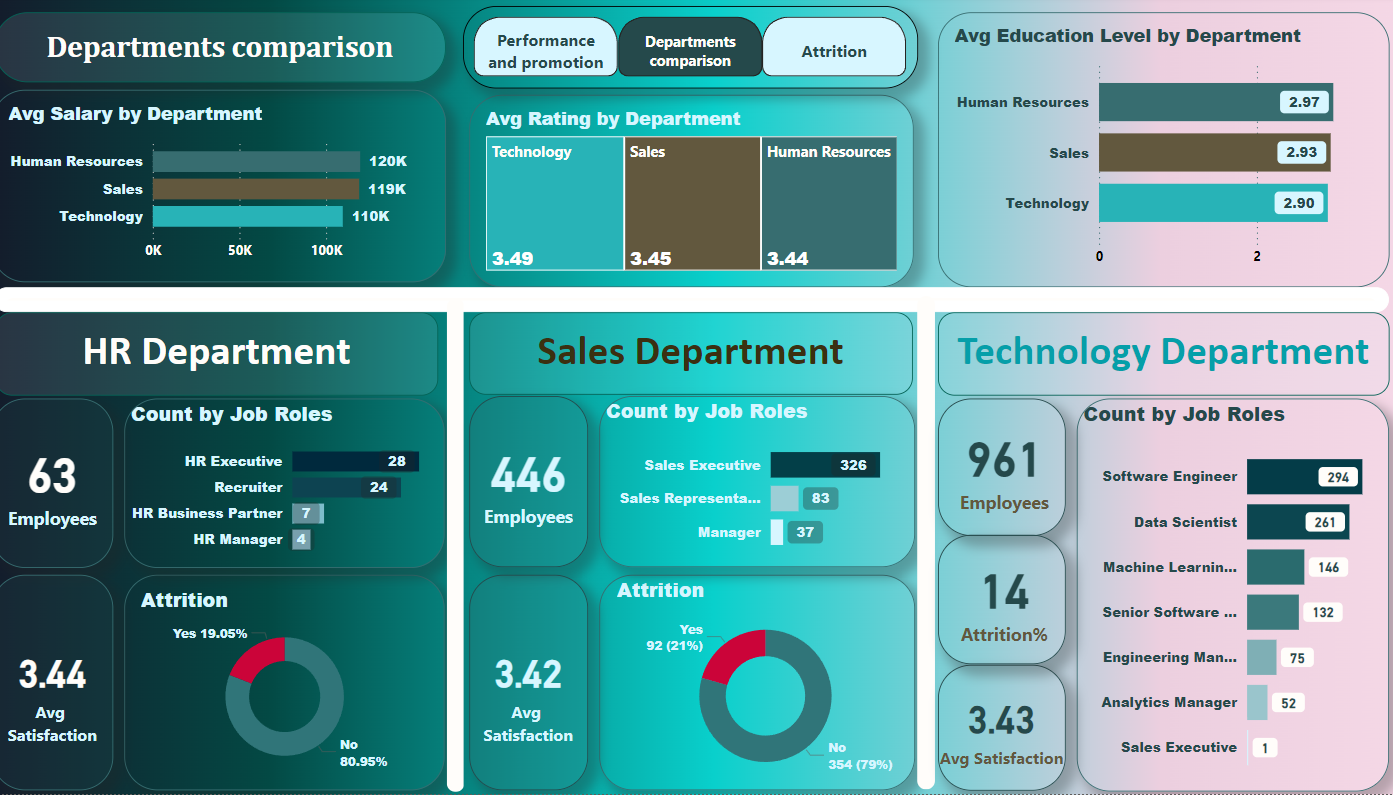
**Attrition Rates by Salary Ranges**

* This horizontal bar chart categorizes attrition by salary range:
  + **A: 20K to 70K (157)**: The vast majority of those leaving fall within the lower salary range.
  + **B: 70K to 150K (48)**: A moderate number leave in the middle range.
  + **C: 150K to 225K (14)** and **D: 225K to 300K (6)**: Very few employees earning higher salaries leave.
  + **Over 300K (1)**: Only one employee earning above 300K left.



**Attrition by Stock Option Benefits**

* This waterfall chart shows the impact of stock options on attrition:
  + **0 (154)**: The majority of employees leaving have no stock option benefits.
  + **1 (56)** and **2 (12)**: Some employees leave even if they have stock options.

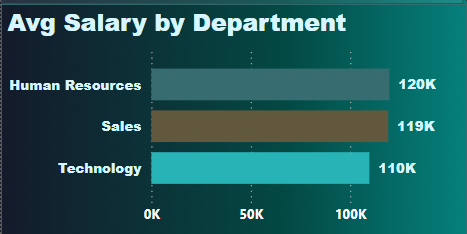


**1. Departments Comparison Dashboard :**

**Main Goal**: The goal of this dashboard is to **compare key metrics across departments**, such as average salary, employee satisfaction, and attrition rates. It helps management:

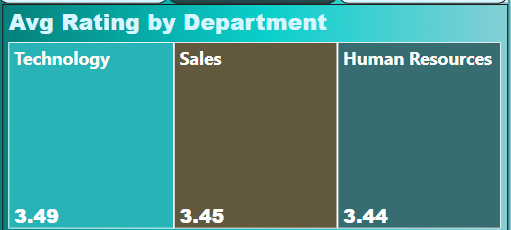
* Understand how each department performs relative to others.
* Identify discrepancies in salary, satisfaction, and attrition across departments.
* Focus on areas needing improvement, support or further investigation (e.g., high attrition in a specific department).

This comparative analysis provides insights for decision-making regarding **resource allocation, employee benefits, and retention strategies** specific to each department.



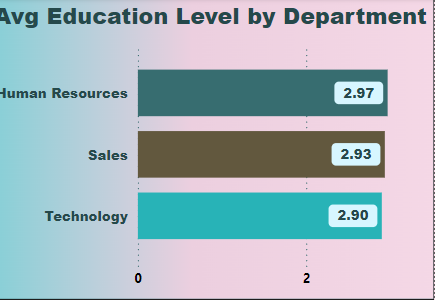
**Average Salary by Department**:

* Compares the average salary of employees across departments.
* **Human Resources (120K)** has the highest salary, followed by **Sales (119K)** and **Technology (110K)**.



**Average Rating by Department**:

* Employees in **Technology** rate the highest at **3.49**, followed by **Sales (3.45)** and **Human Resources (3.44)**.



**Average Education Level by Department**:

* **Human Resources (2.97)** has the highest average education level, followed by **Sales (2.93)** and **Technology (2.90).**



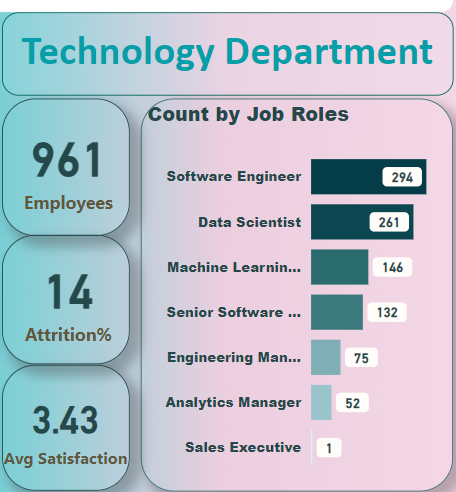
**HR Department**:

* **63 employees**, primarily aged **18 to 29 (42 employees)**.
* **Attrition Rate**: 12 employees have left the department, with an **attrition rate of 19%**.
* **Average Satisfaction**: 3.44.



**Sales Department**:

* **446 employees**, mostly aged **18 to 29 (282 employees)**.
* **Attrition Rate**: 92 employees have left, with an **attrition rate of 20.6%**.
* **Average Satisfaction**: 3.42.

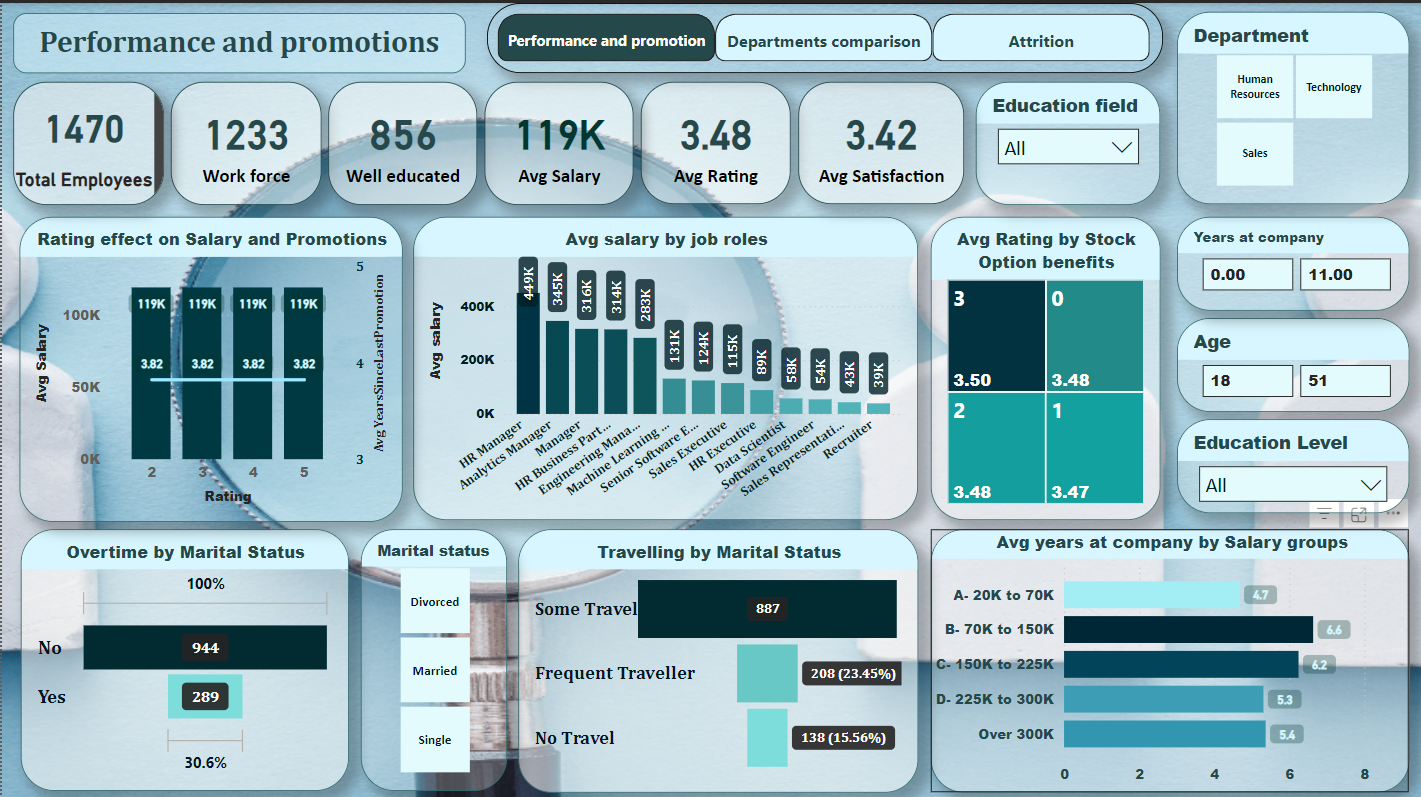


**Technology Department**:

* **961 Employees** total.
* **14% Attrition Rate** (employees leaving).
* **3.43 Average Satisfaction** score.

It also shows the **number of employees by job role**:

* **294 Software Engineers** (largest group).
* **261 Data Scientists**.
* **146 Machine Learning Engineers**.
* Other roles include Senior Software Engineers, Engineering Managers, Analytics Managers, and 1 Sales Executive.

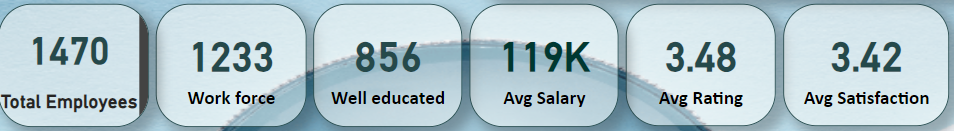


**Performance and Promotions Dashboard.**

**Main Goal**: The goal of this **Performance and Promotions** dashboard is to provide a comprehensive view of employee performance, compensation, and promotion-related insights. It helps decision-makers understand the relationship between key factors such as:

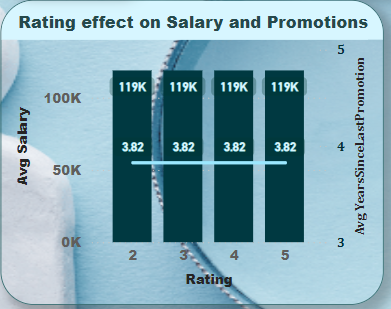
* **Employee Performance Ratings** and their influence on **salary** and **promotions**.
* How **job roles** correlate with different **salary levels** across the organization.
* The effect of **stock option benefits** on employee performance and satisfaction.
* The distribution of **overtime** work across different **marital statuses** and how that might impact performance or job satisfaction.
* **Employee retention** based on salary ranges, showing how long employees typically stay at the company depending on how much they earn.
* Insights into **employee travel behavior** based on marital status, which could influence work-life balance and productivity.

In summary, this dashboard helps management make informed decisions about promotions, compensation strategies, and overall workforce management. It enables the identification of patterns in performance and satisfaction across different departments, roles, and demographics, which can be used to improve employee retention and job satisfaction.



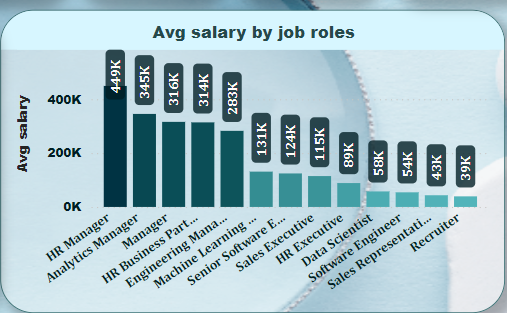
**The KPIs :**

* **Total Employees (1470)**: Shows the total workforce in the company.
* **Workforce (1233)**: Highlights the number of employees actively working.
* **Well Educated (856)**: Displays the number of employees with higher educational qualifications.
* **Average Salary (119K)**: Represents the average salary across the workforce.
* **Average Rating (3.48)**: Displays the average performance rating.
* **Average Satisfaction (3.42)**: Shows the average employee satisfaction score.

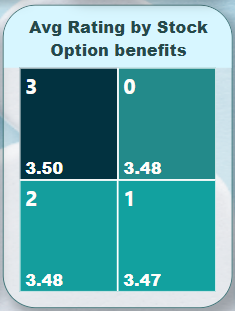


**Rating Effect on Salary and Promotions**:

* Displays the **average salary** by **performance rating (2 to 5)**, with all ratings resulting in a salary of 119K.
* The corresponding **years since last promotion** is 3.82 years for all ratings.

  
**Average Salary by Job Roles:**

* **This bar chart shows the average salary per job role.**
* **The highest-paid role is HR Manager (449K), followed by Analytics Manager (345K) and HR Business Partner (316K).**



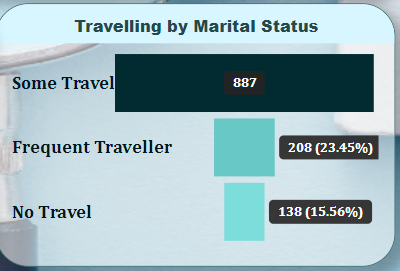
**Average Rating by Stock Option Benefits:**

* **Compares the average rating based on the number of stock option benefits employees receive.**
* **Ratings are slightly higher for those receiving 3 stock options (3.50) and 2 stock options (3.48).**



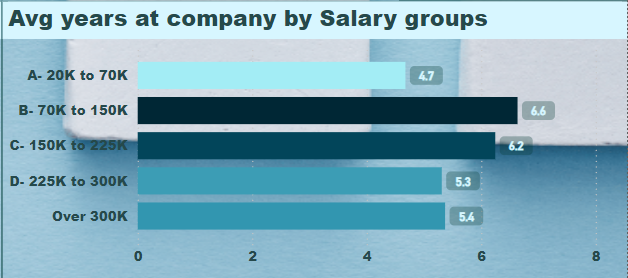
**Overtime by Marital Status:**

* **Illustrates the number of employees working overtime, split by marital status.**
* **Most employees (944) don’t work overtime, while 30.6% (289) do.**



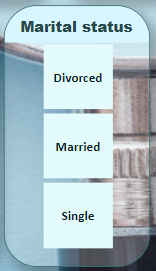
**Travelling by Marital Status:**

* **A funnel chart that shows how marital status influences business travel.**
* **Single employees tend to travel frequently (74) while married (90)**
* **Married employees tend to travel the most (400 employees with some travel), while Single employees dominate in some travel (303).**



**Average Years at Company by Salary Groups**:

* Shows the average tenure of employees within specific salary ranges.
* Employees earning **70K to 150K** tend to stay the longest, with an average tenure of **6.6 years**, while those earning **20K to 70K** stay for **4.7 years**.



We used Age, Years at company, Department, Education level, and Marital status as a filters

**Conclusion:**

**First dashboard (Performance of workforce):**

* **Salaries** and **promotions** are **independent** ofthe employee **rating which we must consider.**
* **Salary Progression:** There’s a **clear pattern** of **higher salaries** for **more senior job roles,** which **makes sense** and **aligns** with **responsibility levels.**
* **Stock option benefits** have **little impact** onemployees **rating.**
* **Marital status** has **obvious effect** onemployees **traveling frequency** and **over time.**

1. **Divorced employees** are the **highest** to have **overtime** work **then married employees** andthe **least percentage** is for the **single employees. (as percentages)**
2. **Single employees** havethe **highest percentage** of **frequent traveling** and **married employees** havethe **lowest percentage. Divorced employees** have the **highest percentage** of **no traveling** and **married employees** havethe **lowest** percentage of **no traveling.**

* The **average salaries** ofemployeesare **independent** of **years at company (employment period) which we must consider.** The **lowest** average **employment period** isbetween **225k** and **300k salary** on the other hand **the highest** average **employment period** is between **70k** and **150k.**

**Second dashboard (Departments comparison):**

* The **HR department has** the **highest average salary, yet** the **lowest average rating.**
* **Conversely,** the **technology department** hasthe **highest ratings** butthe **lowest salaries, pointing to a potential mismatch** between **compensation** and **performance.**
* The **HR department boasts** the **highest education levels,** whilethe **technology** department has the **lowest.** This **aligns** with **salary differences, indicating education might influence salary scales.**
* The **technology department** has the **highest** number of **employees** with the **lowest attrition rate** while the **human resources** department has the **lowest** number of **employees,** and the **sales department** hasthe **highest attrition rate.**
* The **average satisfaction rate** is **almost** the **same** in all **departments.**
* The **attrition** reasons should be searched to **reduce the attrition rate,** which we are discussing in the third dashboard.

**Third dashboard (Attrition reasons and factors):**

* **Overall attrition rate** is approximately **(16%) 237** employees **out of 1470** employees
* The **Single employees experience** the **highest attrition rates, followed by married employees,** with **divorced employees** showingthe **lowest rates.**
* **Age** is a critical factor: **82%** of those who are leaving are in the 18–29 age group, while only 6% are aged **over 40**.
* The **highest job roles** in **attrition** are **data scientist (62), Sales executive (57), software engineer (47), sales representative (33). The reasons of attrition in** these **job roles** should be **searched** and theemployeesthat **left** these **roles** shouldbe **compensated.**
* **Distances from home** have **little impact** onemployee’s attrition. (But not strong factor)
* The **longer** the **employment period** the **lower** the **attrition rate,** which is an **indication** of **good company work environment.**
* The **highest attrition rate** is in the **lowest salary range** from **20K to 70K (157)** andthe **higher** the **salary range** the **lower** the **attrition rate.**
* **Stock option benefits** are a very **strong factor** to **reduce attrition rate** as the **higher stock option benefits** the **lower** the **attrition rates.**

**Summary:**

**Increase Minimum Salaries and Stock Benefits: Raising base salaries** and **offering more stock options** can help **reduce** employee **turnover**, especially for lower-paid employees. This may also **improve** employee **performance.**

**Focus on High-Turnover Roles:** Jobs like **data scientist, sales executive, and software engineer need special attention**. The company should figure out why employees in these roles are leaving and offer better compensation or growth opportunities.

**Link Performance to Pay:** Employee performance **ratings** should **influence** **salary** **increases** and **promotions**. Rewarding high performers will motivate employees to work harder.

**Reward Long-Term Employees:** Giving **salary** **increases** based on how long employees stay at the company can **encourage** them to **remain** for **longer** periods.

**Address Attrition in Sales and HR:** The sales and HR departments need attention due to **high turnover**. Finding out why employees are leaving and making improvements will help reduce attrition in these departments. Also, **compensating** the **candidates who have left is important.**