

|  |
| --- |
| **ePROJECT DATA SCIENCE**  **HR Analytics** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Developed by**  **Members:**   |  |  |  | | --- | --- | --- | | **No.** | **Student Name** | **Student ID** | | 1 | Trần Trường Thọ | Student1503313 | | 2 |  |  | | 3 |  |  | | 4 |  |  |  * **Class No.: DT2307L-NK** * **Start Date:** **26/06/2024** * **End Date: 31/07/2024** * **Name of the Coordinator: Hồ Nhựt Minh** * **Date of Submission: 31/07/2024** |
| **1. GET THE DATA** |

Source :

<https://conferences.shrm.org/presenter/carla-patalano-dba-shrm-scp>

https://www.kaggle.com/datasets/rhuebner/human-resources-data-set

* Dr. Carla Patalano created an HR-related dataset used in one of our graduate MSHRM courses called HR Metrics and Analytics, at the University of New England School of Business
* The dataset consists of 36 columns and 312 rows :

Data Dictionary

| **Feature** | **Description** | **DataType** |
| --- | --- | --- |
| Employee Name | Employee’s full name | Text |
| EmpID | Employee ID is unique to each employee | Text |
| MarriedID | Is the person married (1 or 0 for yes or no) | Binary |
| MaritalStatusID | Marital status code that matches the text field MaritalDesc | Integer |
| EmpStatusID | Employment status code that matches text field EmploymentStatus | Integer |
| DeptID | Department ID code that matches the department the employee works in | Integer |
| PerfScoreID | Performance Score code that matches the employee’s most recent performance score | Integer |
| FromDiversityJobFairID | Was the employee sourced from the Diversity job fair? 1 or 0 for yes or no | Binary |
| Salary | The person’s yearly salary. $ U.S. Dollars | Float |
| Termd | Has this employee been terminated - 1 or 0 | Binary |
| PositionID | An integer indicating the person’s position | Integer |
| Position | The text name/title of the position the person has | Text |
| State | The state that the person lives in | Text |
| Zip | The zip code for the employee | Text |
| DOB | Date of Birth for the employee | Date |
| Sex | Sex - M or F | Text |
| MaritalDesc | The marital status of the person (divorced, single, widowed, separated, etc) | Text |
| CitizenDesc | Label for whether the person is a Citizen or Eligible NonCitizen | Text |
| HispanicLatino | Yes or No field for whether the employee is Hispanic/Latino | Text |
| RaceDesc | Description/text of the race the person identifies with | Text |
| DateofHire | Date the person was hired | Date |
| DateofTermination | Date the person was terminated, only populated if, in fact, Termd = 1 | Date |
| TermReason | A text reason / description for why the person was terminated | Text |
| EmploymentStatus | A description/category of the person’s employment status. Anyone currently working full time = Active | Text |
| Department | Name of the department that the person works in | Text |
| ManagerName | The name of the person’s immediate manager | Text |
| ManagerID | A unique identifier for each manager. | Integer |
| RecruitmentSource | The name of the recruitment source where the employee was recruited from | Text |
| PerformanceScore | Performance Score text/category (Fully Meets, Partially Meets, PIP, Exceeds) | Text |
| EngagementSurvey | Results from the last engagement survey, managed by our external partner | Float |
| EmpSatisfaction | A basic satisfaction score between 1 and 5, as reported on a recent employee satisfaction survey | Integer |
| SpecialProjectsCount | The number of special projects that the employee worked on during the last 6 months | Integer |
| LastPerformanceReviewDate | The most recent date of the person’s last performance review. | Date |
| DaysLateLast30 | The number of times that the employee was late to work during the last 30 days | Integer |
| Absences | The number of times the employee was absent from work. | Integer |
|  |  |  |

* This dataset can be used to analyze employee behavior.
* By design, there are some issues that are present. It is primarily designed as a teaching data set - to teach human resources professionals how to work with data and analytics.

**Example**

|  |
| --- |
| **2. PREPARE THE DATA** |

* In the dataset, there are 2 columns [' DateofTermination'],[' ManagerID'] with null values, this makes sense because:

- [' DateofTermination'] is the termination date, if that employee is still working, there will be no data about the termination date

- [' ManagerID'] is the Unique Identifier for each manager, if that employee does not have their manager, this column will be null

* in the dataset there are columns with invalid data type, we converted it:

- Column ['DOB'] has object data type which we converted to datetime

- Column ['DateofHire'] has object data type which we converted to datetime

- Column ['DateofTermination'] has object data type which we converted to datetime

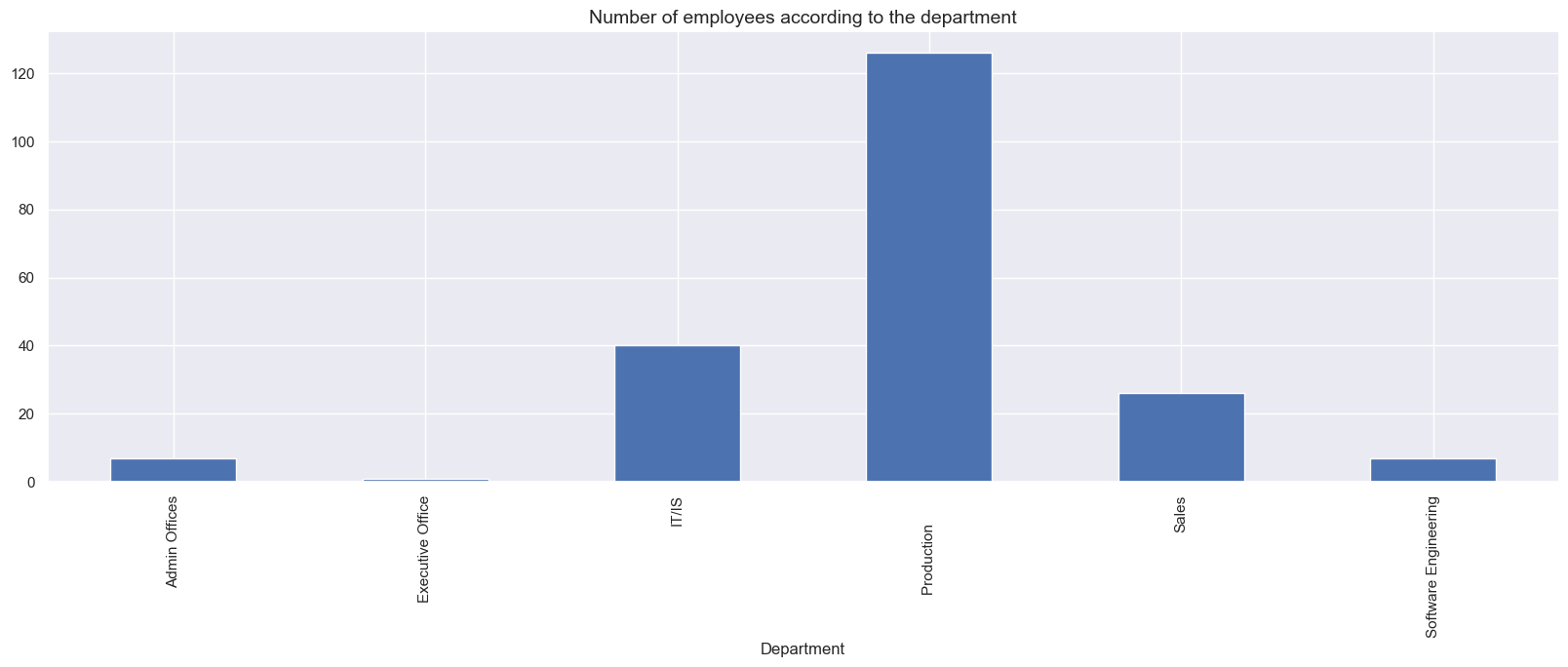
- Column ['LastPerformanceReview\_Date'] has object data type which we converted to datetime

* Now the data is ready to explore.

|  |
| --- |
| **3. EXPLORE THE DATA (EXPLORATORY DATA ANALYSIS, OR EDA)** |

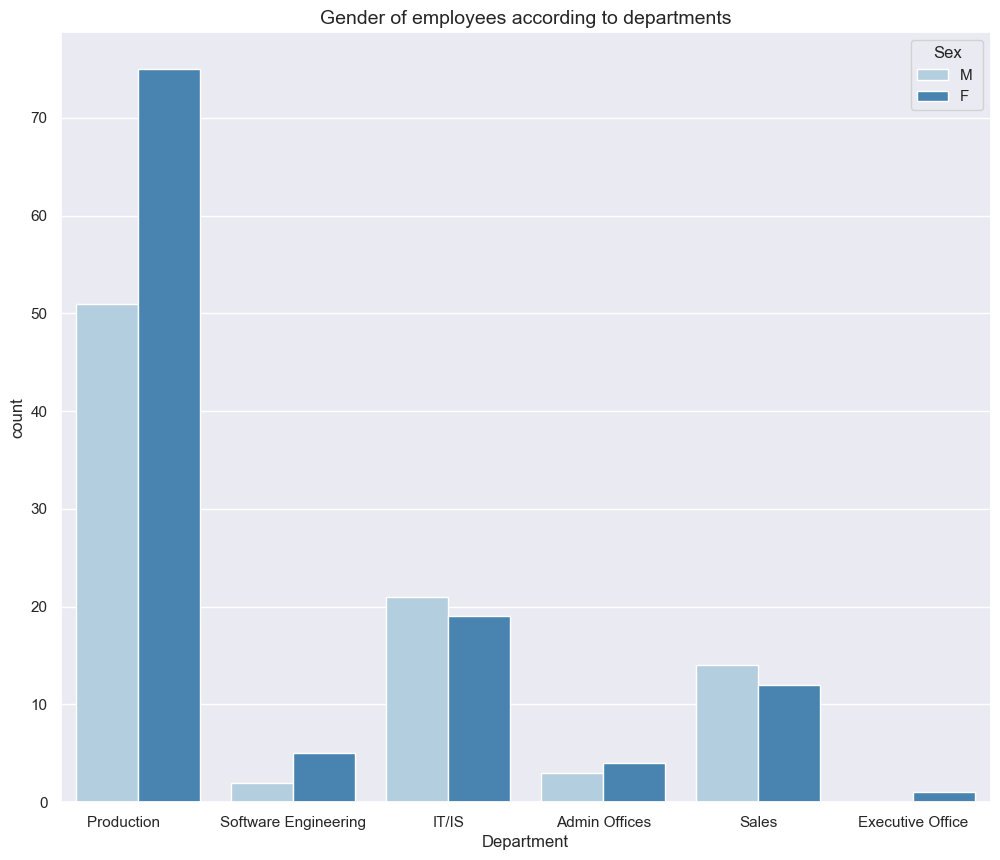
When looking at the aggregated data, we discovered a number of points that can be used to analyze this data .

* Analyze relationships between departments and other parameters
* Number of employees according to the department :



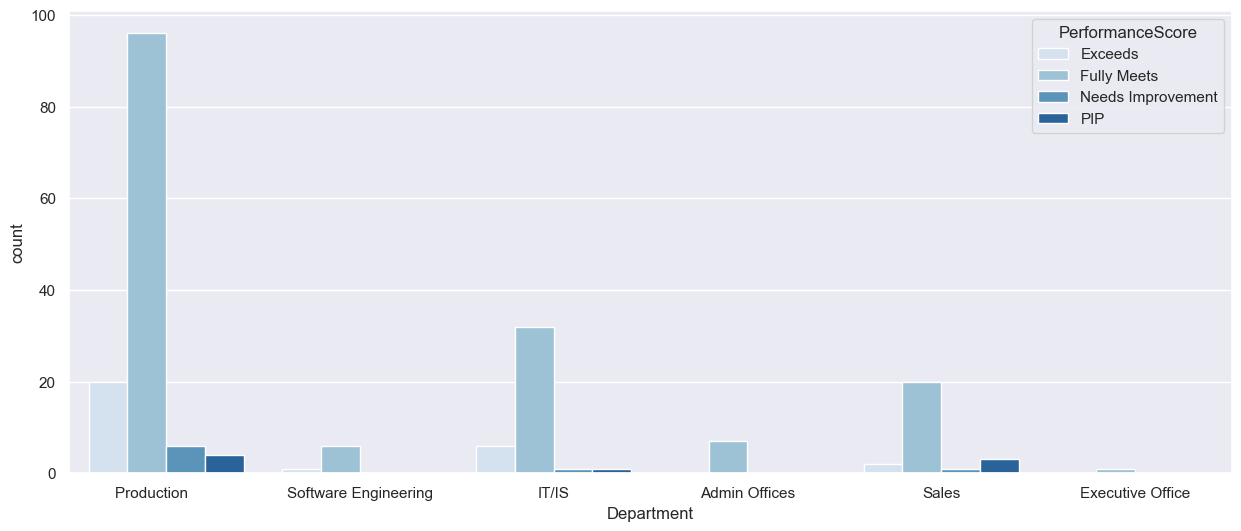
**Conclusion** : The company has 5 departments and the production department has the largest number of employees in the company.

* Gender of employees according to departments :



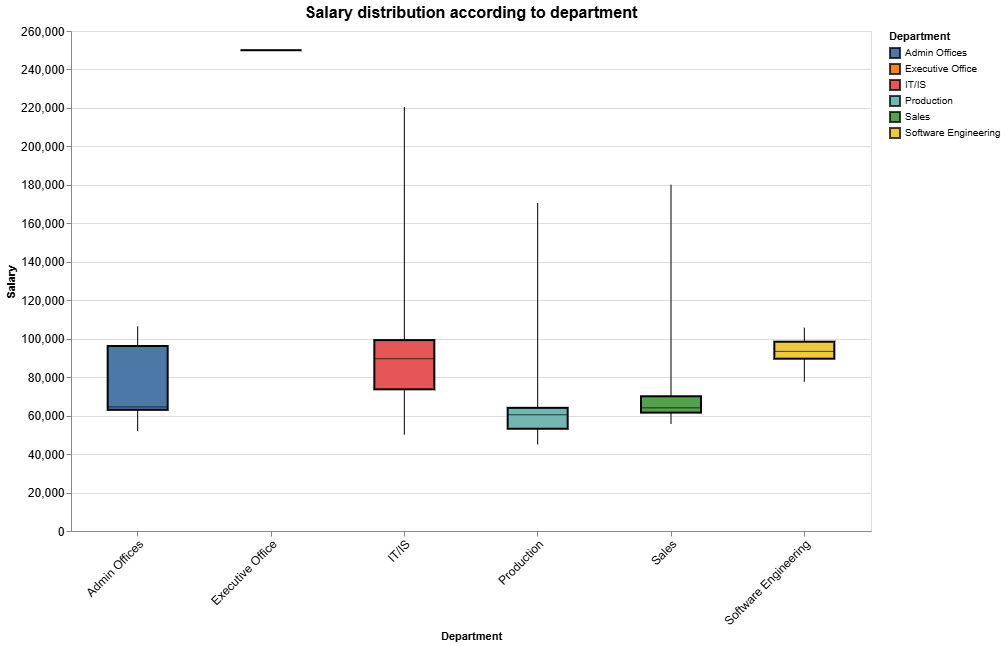
**Conclusion** : The production , software engineering and admin office departments have more female employees than male employees.

* Working performance according to departments:



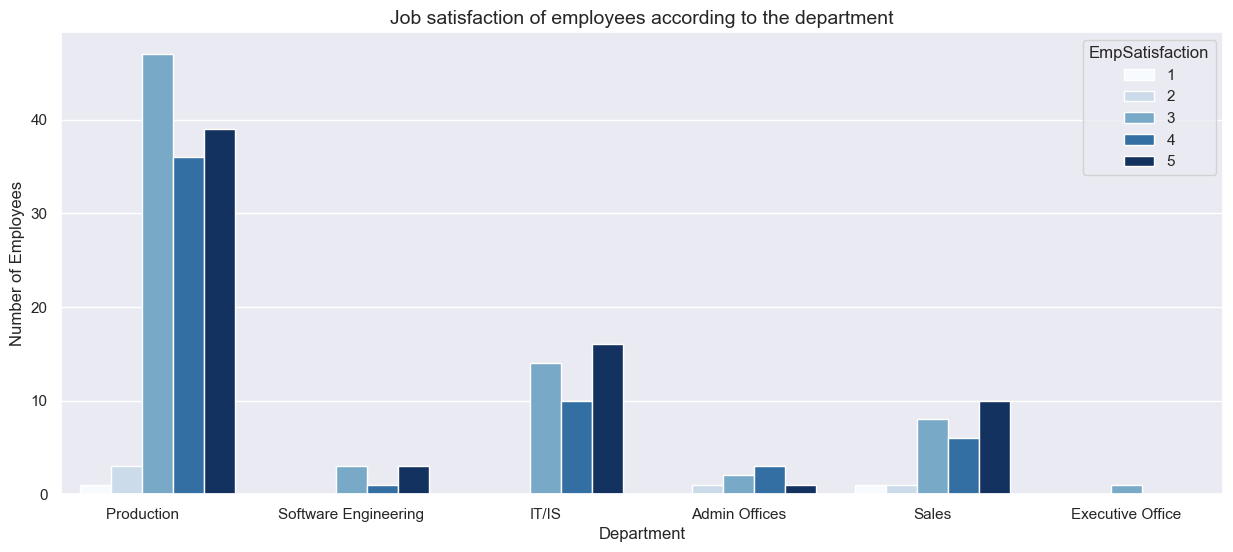
**Conclusion :** Most departments achieved their goals.

* Salary distribution according to department :



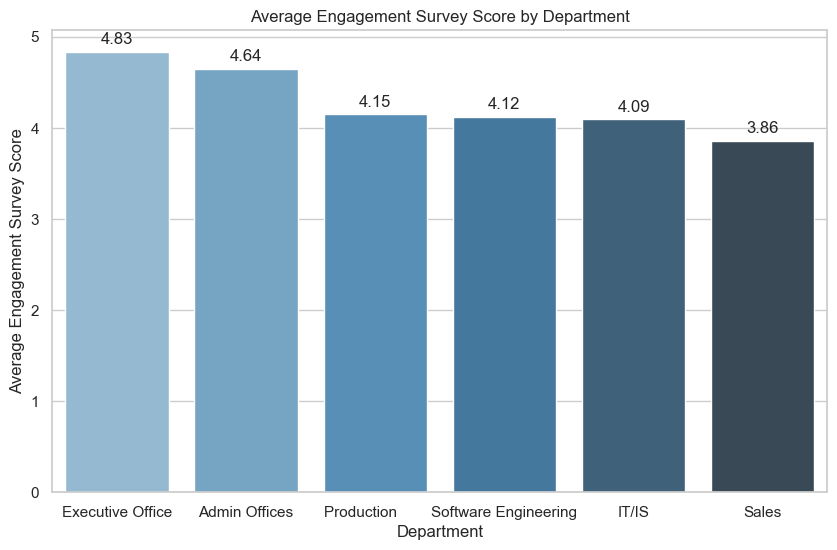
**Conclusion** : The production department has an average salary lower than the company average.

* Job satisfaction of employees according to the department :



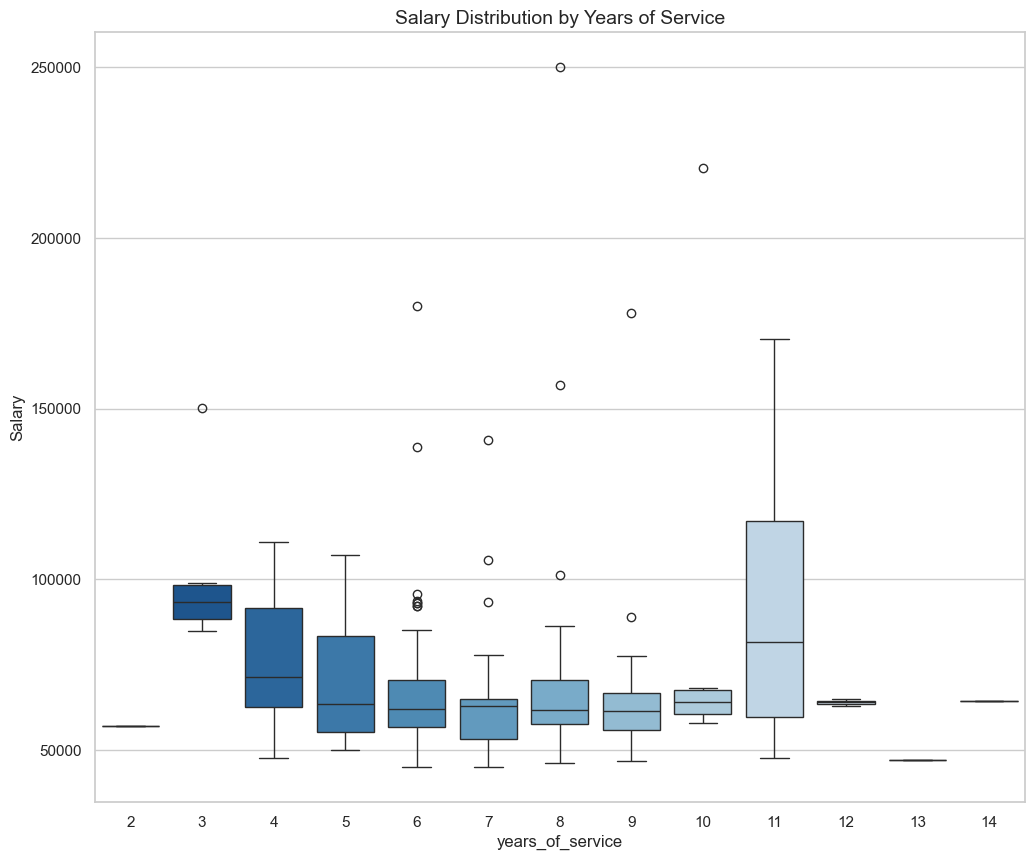
**Conclusion** : Most employees are satisfied with their jobs.

* Average Engagement Survey Score by Department :

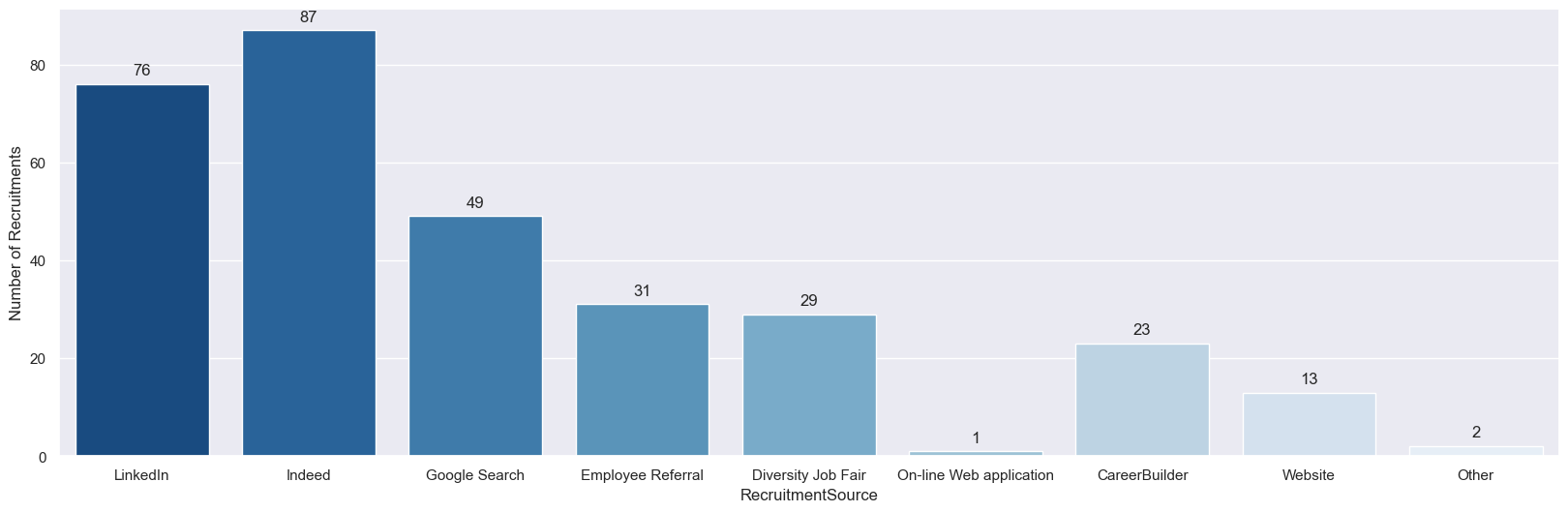


**Conclusion** : Sales and IT/IS have lower levels of engagement than other departments

* Salary Distribution by Years of Service :

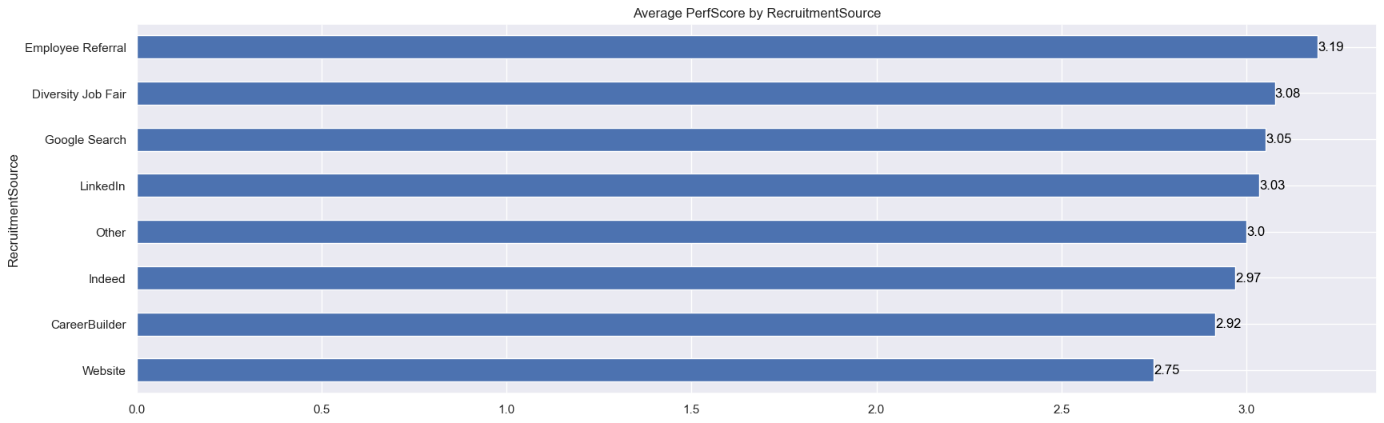


* Recruitment source :



**Conclusion** : Most of the company's employees are recruited from Indeed, LinkedIn, Google Search

* Average PerfScore by RecruitmentSource :



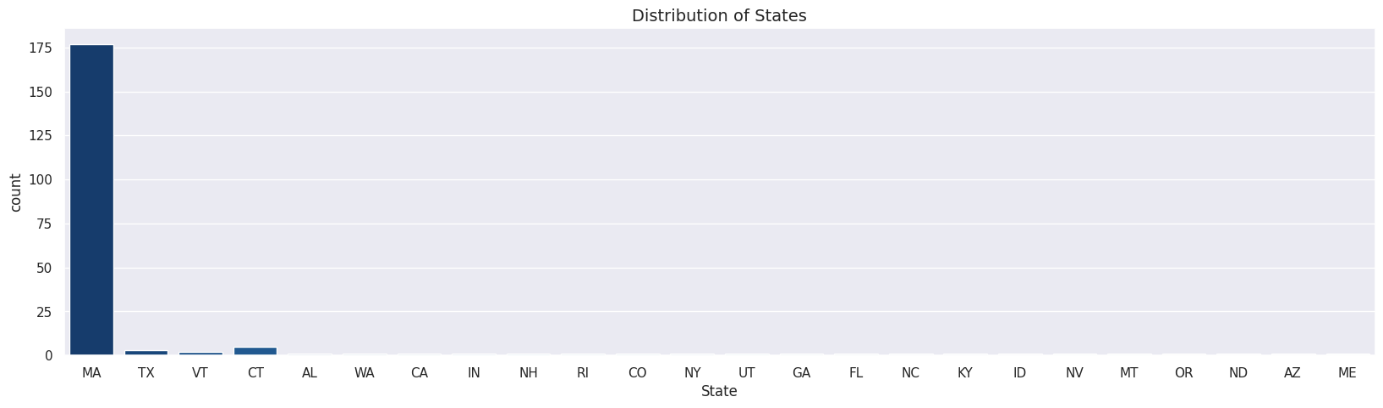
Employees from the "Employee Referral" recruitment source have higher performance scores than other recruitment sources.

* Analyze employee data in the company
* Distribution of Marital Status :



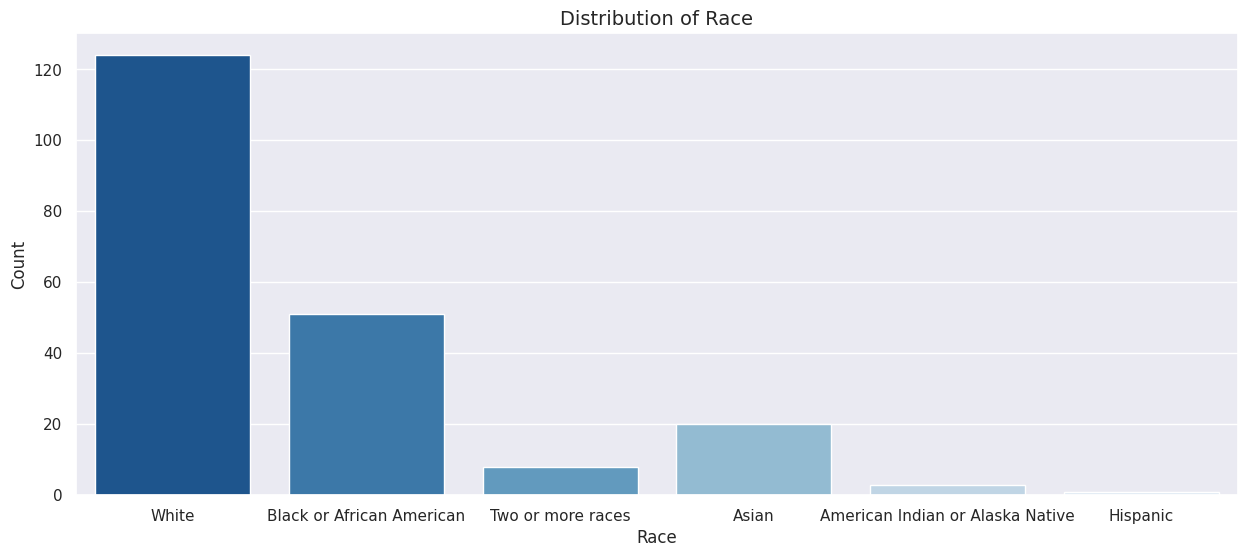
**Conclusion** : Most of the employees in the company are single.

* Distribution of States :



**Conclusion** : Most employees are located in the state of Massachusetts.

* Racial distribution



**Conclusion** : Most of the employees in the company are white.

|  |
| --- |
| **4. ANALYZE THE DATA** |

* **Overview of the personnel situation :**
* Company personnel chart
* Statistics on the percentage of employees leaving the organization

A pie chart with text on it

Description automatically generated

**Conclusion** : The turnover rate is 33.4%, indicating a fairly high employee turnover.

* **Income of employees**

**A graph with colored squares and lines

Description automatically generated**

**Conclusion** : Avg salary of all working employees is 70.694,033 USD.

* **Employee Performance Assessments :**
* **Recuitment**

A graph with blue squares

Description automatically generated

**A blue and white striped graph

Description automatically generated**

**Conclusion** :The Employee Referral is the best source base on performance but indeed, linkedin and google search still the best source to recruit base on number

* **Average PerfScore by Department**

**A graph of a person with a bar graph

Description automatically generated with medium confidence**

**Conclusion** : The chart shows that the work performance is very good, most of the employees completed their work on time.

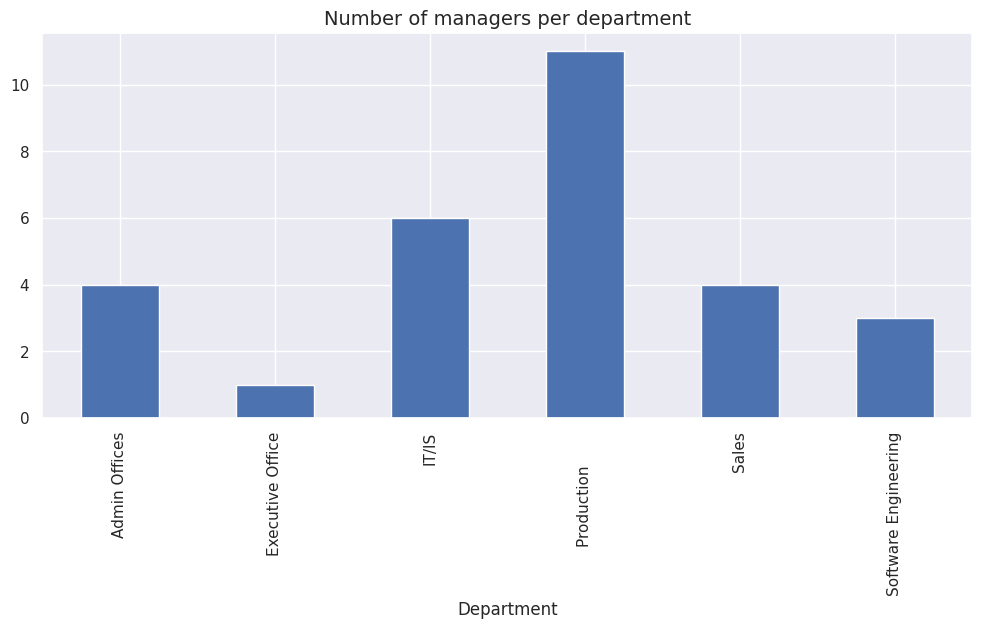
* **Employee Positive Percentage by Department**

**A graph of a bar graph

Description automatically generated with medium confidence**

**Conclusion** : Operational support department have high efficiency than product and business operations

* **Managers and department :**
* Managers/employees ratio and distribute of managers



**A blue and red circle with a red triangle

Description automatically generated**

* Management performance evaluation

A chart with numbers and a number of positive percentages

Description automatically generated with medium confidence**Conclusion** : 5 managers achieve maximum work performance

* **What impacts work efficiency? :**
* Correlation between salary, satisfaction, commitment, and performance

A screenshot of a color palette

Description automatically generated

**Conclusion** : We can find the stronger correlation between Engagement and Perfromance than Employees Satisfaction and Performance

* Histogram of Performance Scores

A blue and white graph

Description automatically generated

**Conclusion** : high employee completion rate

* Distribution of salary vs performance

A graph of a graph showing a number of objects

Description automatically generated with medium confidence

**Conclusion** : The chart shows that employees who complete their work will have higher salaries than employees who do not complete their work.

* EngagementSurvey BY MANAGER

A graph with blue and white lines

Description automatically generated

* THE INFLUENCE OF MANAGERS ON EMPLOYEES

A graph showing the difference between employees

Description automatically generated A graph showing a salary gap

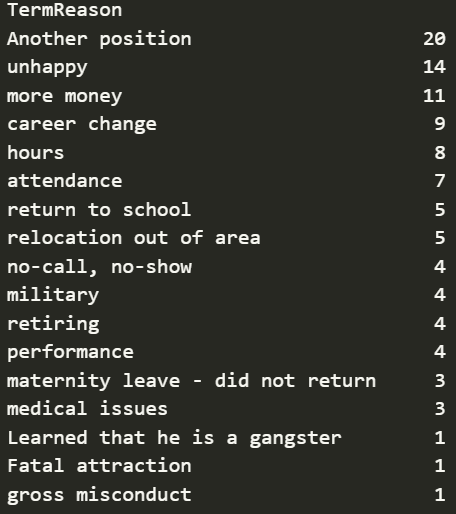
Description automatically generated

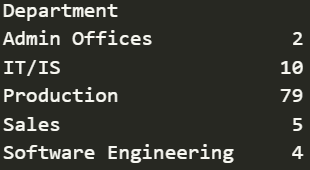
A graph of a survey gap

Description automatically generated with medium confidence

**Conclusion** : It seems that good managers have more engaged employees and higher paid employees, but dissatisfied employees. So we can say that Engagement is more important than Satisfaction.

* **Why employees quit their jobs**
* Employee Resignation Overview
* QUIT JOB REASON AND ACCORDING TO DEPARTMENT



 A graph showing a positive and negative result

Description automatically generated

**Conclusion** : These 3 tables show that the production department has a lot of employee turnover and the reasons for turnover are often negative.

* Average Performance Score Comparison

A blue and white graph

Description automatically generated

**Conclusion** : The performance of employees who quit was no worse than that of those who remained employed, suggesting that they did not perform poorly and quit.

* Average Salary Comparison

A graph showing a salary comparison

Description automatically generated

A graph of a number of employees

Description automatically generated



**Conclusion** : Most of the employees are in MA state we took the average salary of MA in 2020 to compare we found that the salary of the employees who quit were all higher than the average salary of the state so we found that the employees quit due to low salary seems wrong

* **The impact of management on employees**

**A graph with a red bar

Description automatically generated**

**A graph with red squares

Description automatically generated**

**A graph showing the average quitted employees performance

Description automatically generated**

**Conclusion** : After we analyzed the management levels, we found that good managers have higher employee turnover rates and higher work performance than other managers, so we found that management style has a major impact on employee turnover.

|  |
| --- |
| **5. COMMUNICATE THE RESULTS** |

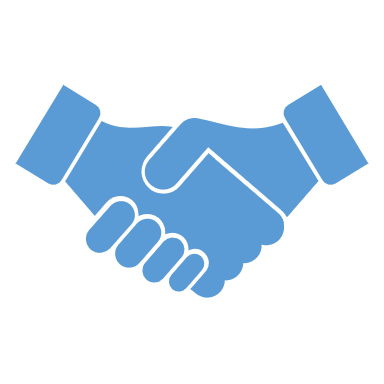
* **Some concluding remarks :**
* HOW TO IMPROVE WORK EFFICIENCY AND RETAIN EMPLOYEES
* Building company culture



Creating a culture of engagement requires more than completing an annual employee survey and then leaving managers on their own, hoping they will learn something from the survey results that will change the way they manage. It requires an organization to take a close look at how critical engagement elements align with their performance development and human capital strategies.

Highly engaged organizations also hold their managers accountable - not just for their team's measured engagement level, but also for how it relates to their team's overall performance. They ensure that managers are engaging employees from the first minute of their first day at work.

* Building a friendly and happy working environment



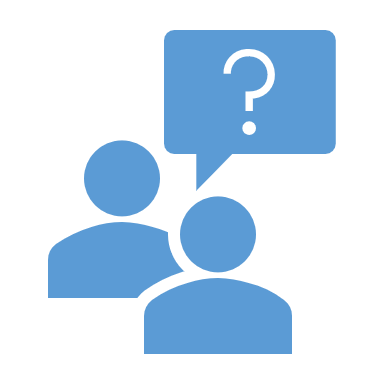
Improved workspace

Providing the right facilities

Freedom

Training, recognize achievements and rewards appropriately

* Create an efficient workflow
* Managers and Employees communicate regularly and continuously



Engage with the company's general direction.

Work with employees to build goals clear and measurable.

Evaluate employees ‘s ability continuously, regularly and exactly.