# **Analytics & Employee Satisfaction**

# **in a European Company**

# Data Science Project

IST 687 Team 14th September 2023

**Submitted by:**

Tibaria Al Nouri

Katherine Martinez

Stephen Price

Yesul Song

# Introduction

In the rapidly evolving landscape of today's corporate world, employee satisfaction stands as a linchpin to organizational success. A satisfied workforce not only bolsters productivity but also strengthens brand reputation, ultimately impacting the bottom line. It's with this understanding that our independent data science team was commissioned by the HR department of a notable European company. Tasked with the mission of unravelling the intricacies behind their troublingly low employee satisfaction rates, we embarked on a comprehensive exploration of their expansive dataset. This report chronicles our approach, findings, and the invaluable insights gleaned, which we hope will pave the way for enhanced employee satisfaction and, by extension, company success.

The structure of the given employment data is as follows: the spreadsheet covers 14,999 rows with 18 columns giving our team 26,982 data points to play with.

* **Target satisfaction score**: Goal or benchmark for employee satisfaction.
* **DOJ (date of job)**: Employee's joining date.
* **Department**: Division or segment of the company.
* **Employee Code**: Unique identifier for each employee.
* **Employee Status**: Current working status (e.g., Active, Resigned).
* **Employee Satisfaction score (Prior)**: Previous satisfaction rating.
* **Employee Satisfaction score(Latest)**: Most recent satisfaction rating.
* **Monthly work hours**: Hours worked monthly.
* **No of Yrs spent in company**: Duration of employment in years.
* **Number of Projects undertaken**: Total projects handled by the employee.
* **Promotion in the last 5 yrs**: Indicates promotion in the past 5 years.
* **Reason for Leaving**: Explanation for resignation or termination.
* **Salary bracket**: Range indicating pay scale.
* **Sno**: Serial or unique identifier.
* **Work @ accident**: Incidents at work.
* **City**: Employee's city of work or residence.
* **State**: State of work or residence.
* **Country**: Country of work or residence.

# Business Questions

To guide our analysis, we engaged with stakeholders to come up with some high-level business questions that may be driving retention and satisfaction. There is no limit to the questions that could be asked given this dataset, but the following questions gave our team a jumping-off point. Below the questions are the parameters given to the team to deep dive. The main driving question is how the company can reach their target employee satisfaction score of 65% and above.

1. Is satisfaction linked to departments?

* Calculate average satisfaction scores for each department.
* Visualize the satisfaction scores across departments to identify which departments have higher or lower satisfaction levels.
* Consider conducting statistical tests to determine if there are significant differences in satisfaction between departments.

1. Is there a geographic component to job satisfaction?

* Analyze the distribution of employees across cities, states, and countries.
* Calculate average satisfaction scores for each geographic region.
* Use data visualization (e.g., bar charts, heatmaps) to identify any patterns or regional variations in job satisfaction.

1. Did promotion affect retention?

* Divide employees into two groups: those who received a promotion in the last 5 years and those who did not.
* Calculate retention rates for both groups (percentage of employees who stayed in the company).
* Perform statistical tests (e.g., t-test) to determine if there is a significant difference in retention between the two groups.

1. Did job injury affect retention?

* Separate employees who experienced a work-related accident from those who didn't.
* Compare the retention rates between the two groups.
* Analyze the impact of job injury on employee satisfaction scores.

1. What was their duration of employment before leaving?

* Calculate the tenure (in years) for employees who left the company.
* Compare the average duration of employment for employees who left to those who are still in the company.
* Analyze if there is a relationship between satisfaction scores and the duration of employment before leaving.

# Methodology

To ensure our analysis was robust and comprehensive, we adhered to a methodology that's recognized as a standard across the data analytics landscape. This systematic approach breaks down the analysis into distinct stages, each integral to the success and credibility of the insights derived. The stages are:

**1. Defining the Question:**

* Before diving into the data, we framed our analysis with precise questions. This step is essential for providing direction and narrowing our focus, ensuring the data is viewed through a lens relevant to the company's concerns.

**2. Data Collection:**

* We procured the necessary dataset from the HR department. The quality and comprehensiveness of this data were crucial, as it formed the backbone of our entire analysis.

**3. Data Cleansing:**

* With the data in hand, we embarked on a thorough data cleaning process. This involved:
* Formatting data types for consistency.
* Identifying and addressing any missing or null values, ensuring the dataset's integrity.

**4. Data Analysis:**

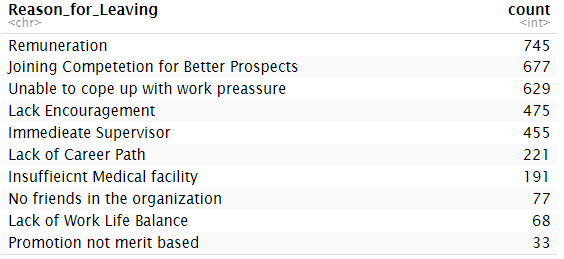
* With a clean dataset, our team then proceeded to the core of the project: the analysis. This encompassed:
* Grouping the data for targeted visualizations, such as bar charts and scatterplots.
* Geographical plotting to visualize employee distribution and satisfaction trends across regions.
* Executing regression analysis to identify potential correlations between variables.

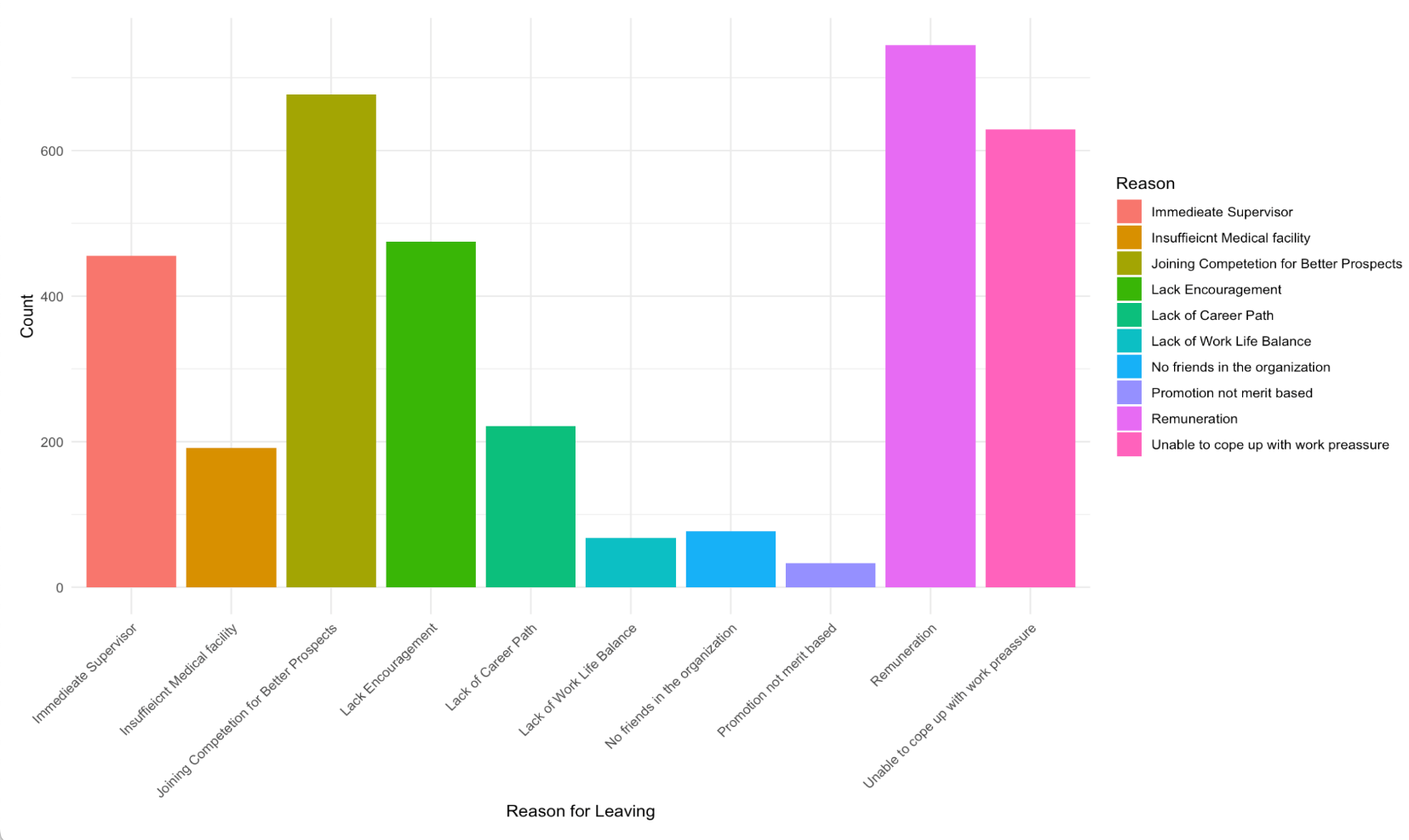
**5. Sharing Results:**

* After analyzing the data and gleaning insights, we compiled our findings into a comprehensive report (like this one). We prioritized clarity and actionable insights, aiming to provide the HR department with tools they need to address their concerns about employee satisfaction.

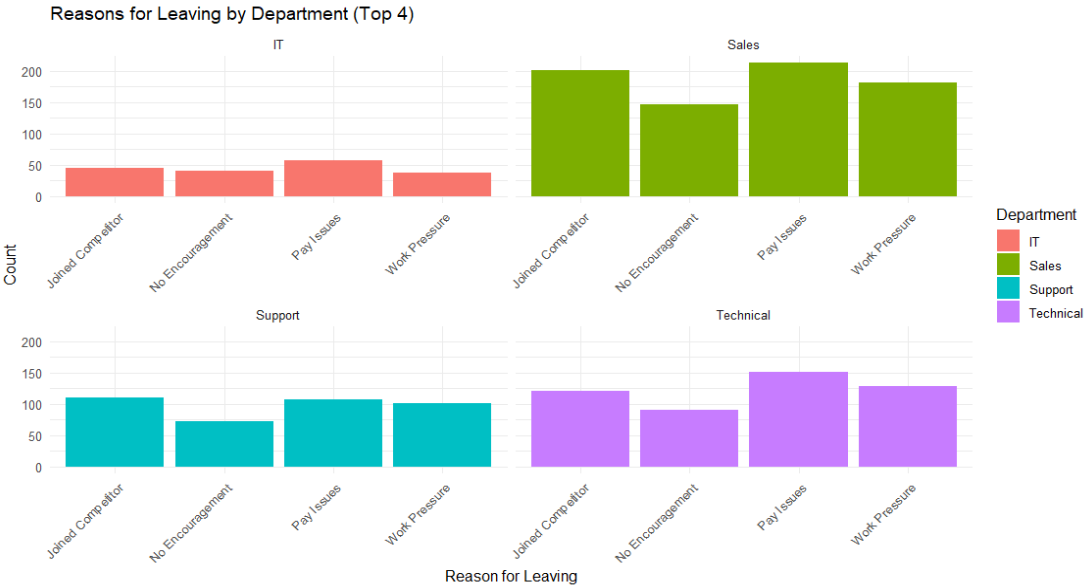
# Analysis

**Reasons for Leaving**

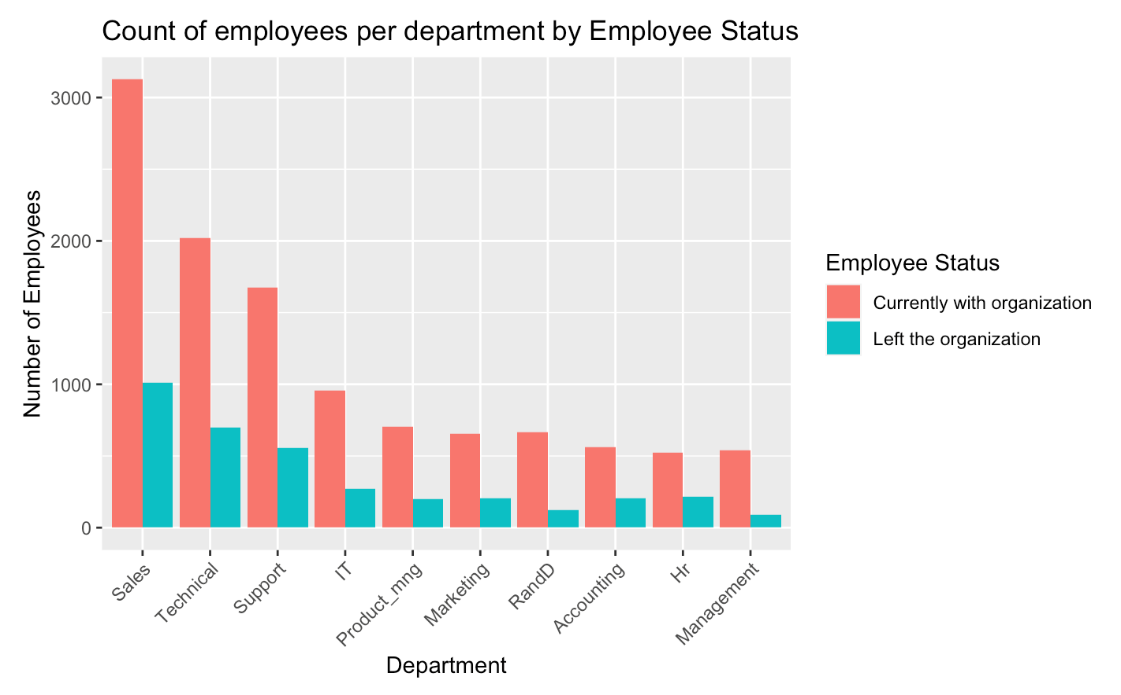
The primary reasons employees cited for departing the company were related to remuneration, the allure of better opportunities at competing firms, and lack of encouragement and their immediate supervisors. Notably, these last two reasons were closely aligned in terms of their frequency of mention.



Among those top 4 reasons for leaving, when looked at by departments with the highest volume of attrition, it seems all departments struggle in the same way to address these concerns. The ranking becomes pay, better prospects elsewhere, work pressure, and lack of encouragement in that order.

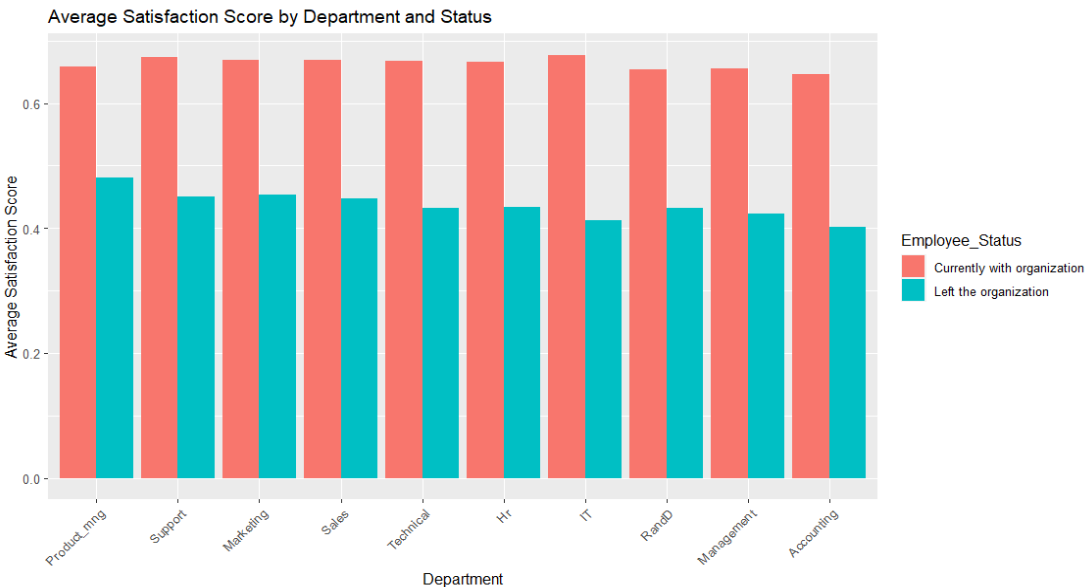


**Attrition by Department**

The departments that exhibited the highest attrition rates were Sales, Technical, Support, and IT. However, when examining departures in relation to the size of each department, the proportion was consistent, with attrition rates ranging between 25-30%.

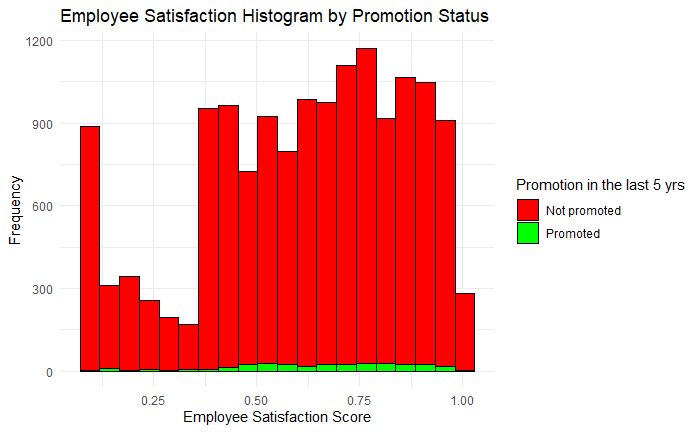
**Is one department more dissatisfied than others?**

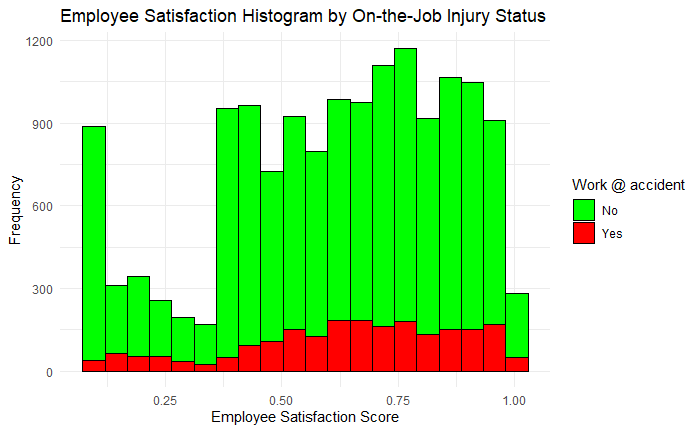
Among those who left the company, the Accounting department surfaced with the lowest satisfaction score, closely trailed by IT.



**Satisfaction linked to promotion or injuries at work?**

Although we found these two factors highly interesting, there is no clear trend linking these variables to employee satisfaction.

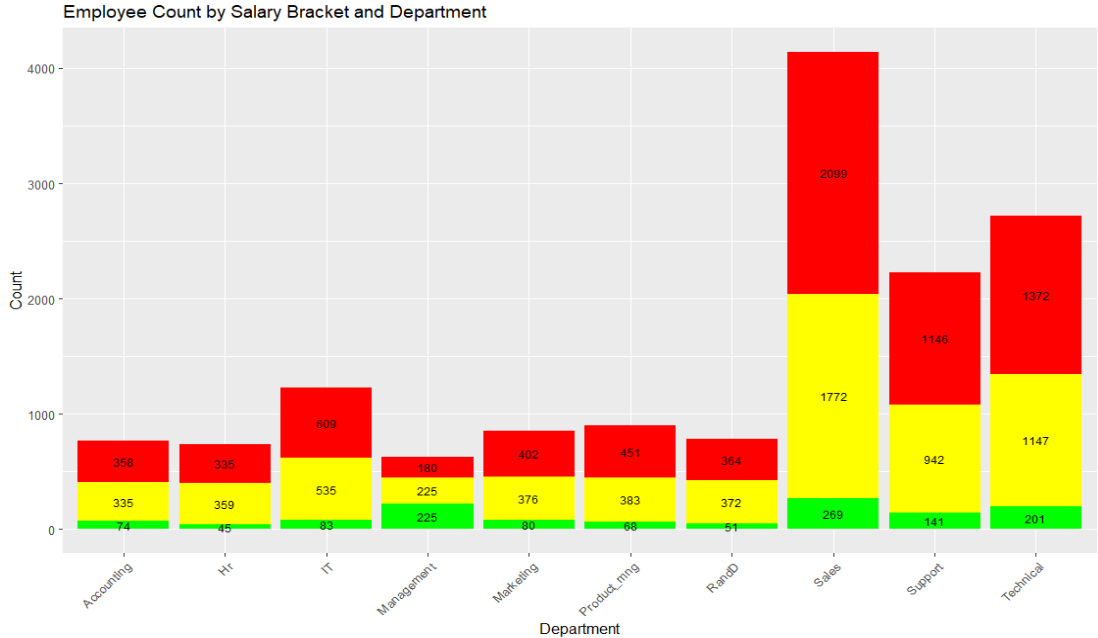


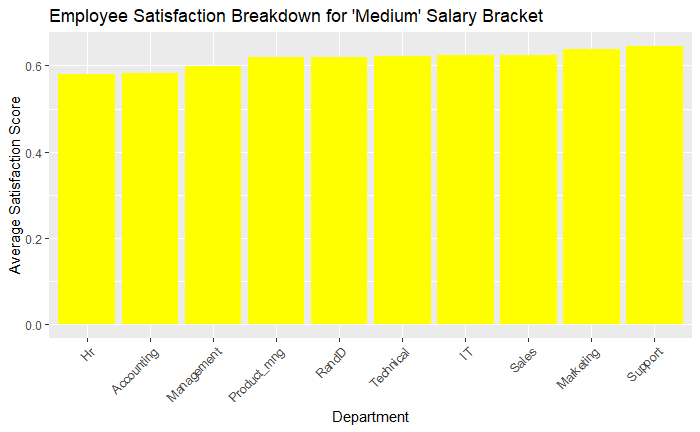
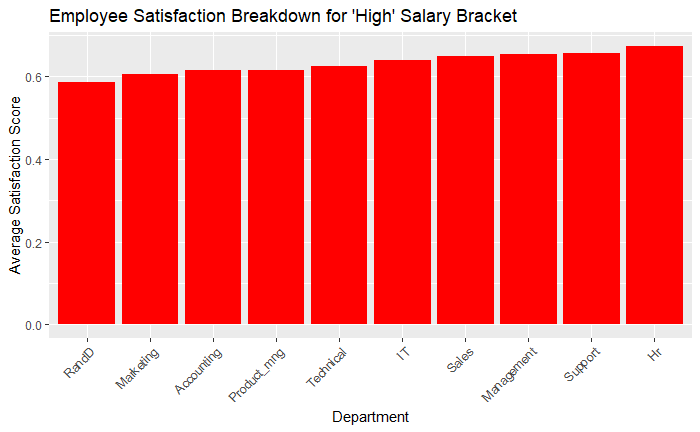


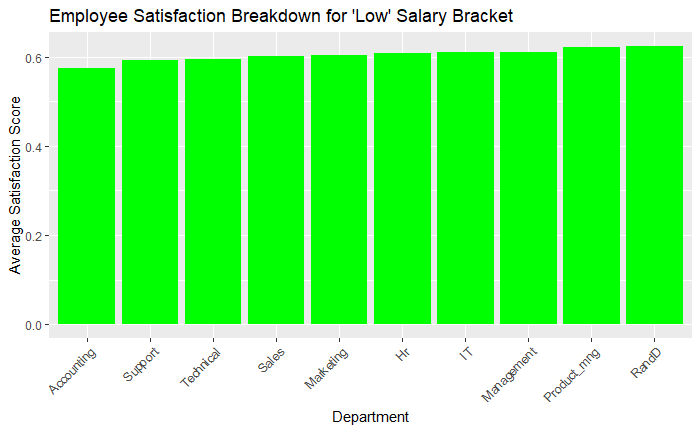
**Satisfaction by Salary Bracket**

Our initial discussions were to dive into the salary brackets to determine if there is a compensation issue within the organization. Furthermore, our regression analysis showed that Low and medium salary brackets did show to be statistically significant. Upon further discussion and analysis, we feel that we need more information and data to confidently state that salary brackets are an issue. We would need to benchmark all job categories and do geographical benchmarking to determine if the organization is paying fair market value for the services being rendered by employees.

Our graphs show normal distribution of ‘High’, ‘Medium’ and ‘Low’ brackets across all departments which could indicate that no department is paying more for talent that other departments. As we have already shown, there is no departmental component to low satisfaction, therefore the breakdown of pay by department is not an issue.







When breaking down each salary bracket by department we can recognize that the distribution of the average satisfaction is similar across salary brackets. While the same department does not show the lowest satisfaction in each bracket, we see that the value of the low and the high department are similar across salary brackets.

# The Regression Analysis

The regression analysis provided a lot of insight on the highest correlated factors impacting employee satisfaction. We build a linear regression model predicting ‘Employee Satisfaction(Latest)’ using the following factors:

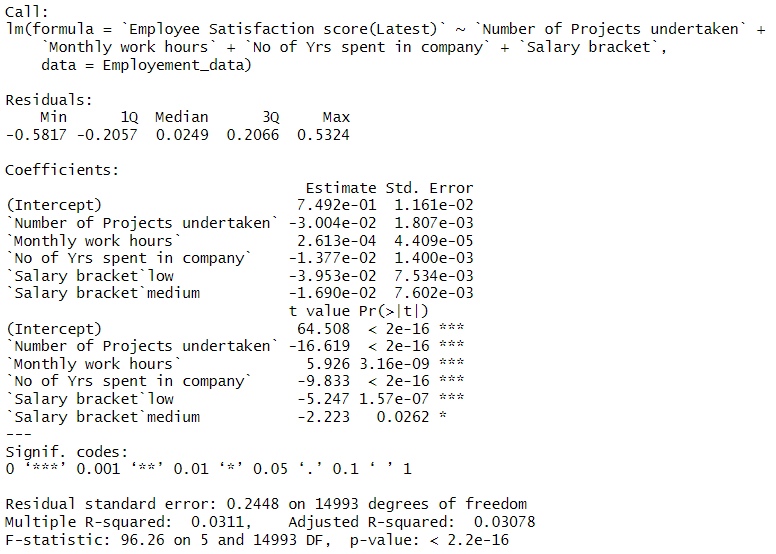
Monthly work hours

No of Years spent in company

Number of Projects undertaken

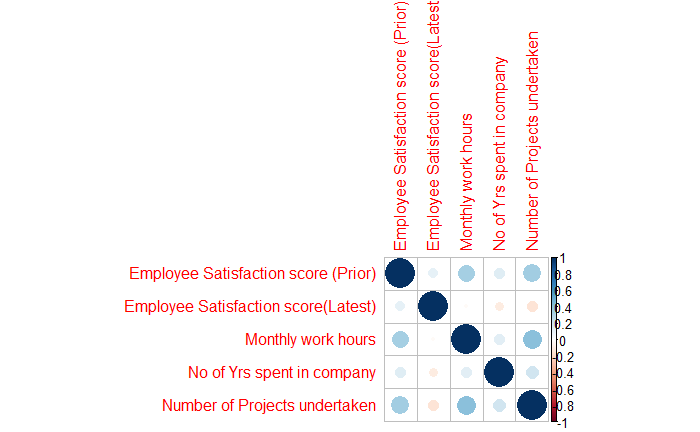
Salary bracket low

Salary bracket medium

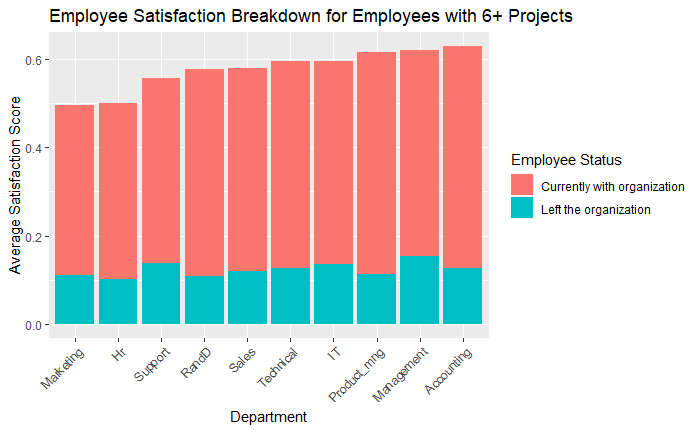


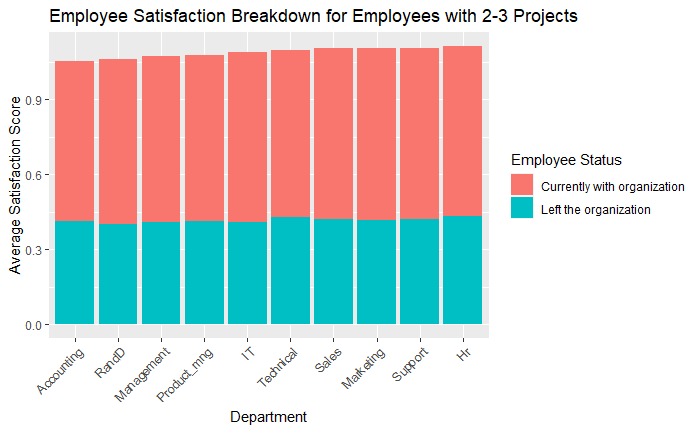
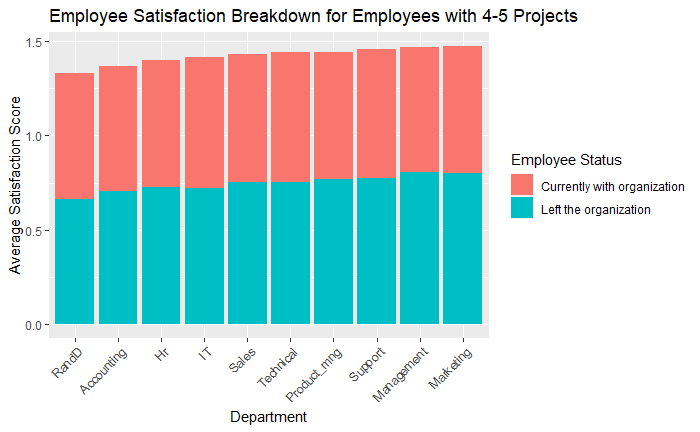
From the results of the model we concluded the following for each factor:

1. Number of Projects undertaken: The coefficient is -0.03004. This suggests that for each additional project an employee undertakes, their satisfaction score decreases by approximately 0.03004 units. Therefore, having a higher number of projects is associated with lower employee satisfaction.
2. Monthly work hours: The coefficient is 0.0002613. This indicates that for each additional hour an employee works per month, their satisfaction score increases by approximately 0.0002613 units. The effect is quite small, suggesting that increased work hours have a minimal positive impact on satisfaction.
3. No of Yrs spent in company: The coefficient is -0.01377. This implies that for each additional year an employee spends in the company, their satisfaction score decreases by approximately 0.01377 units. Longer tenure in the company is associated with slightly lower employee satisfaction.
4. Salary bracket (low): The coefficient is -0.03953. Being in the "low" salary bracket is associated with a decrease of approximately 0.03953 units in employee satisfaction. Lower salary brackets are linked to lower employee satisfaction.
5. Salary bracket (medium): The coefficient is -0.01690. Being in the "medium" salary bracket is associated with a decrease of approximately 0.01690 units in employee satisfaction. Similarly, the "medium" salary bracket is linked to lower employee satisfaction.

In trying to further investigate the data we attempted a correlation analysis shown below. The visual is provided below. However, at our skill level we could not glean any insights from this interesting graphic. Perhaps our Data Science leadership team can guide further analysis in this area.

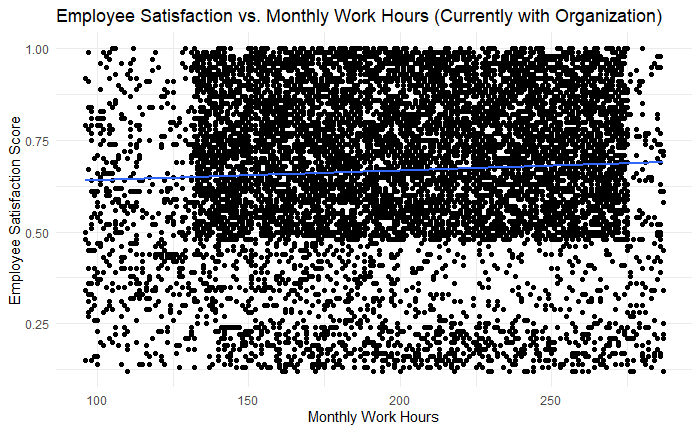
**Is satisfaction linked to projects?**  
The following graphs investigate any correlation between the number of projects employees are assigned, their satisfaction rating, grouped by department and broken up by those employed and those who left. It becomes clear there is a threshold for the number of projects an employee takes on before their satisfaction score begins to diminish.

Loading employees with 6 or more projects seems to leave them feeling overworked and extremely unsatisfied. Those who left the organization had a low satisfaction of 10% while their peers in the company had 40-50% satisfaction – far below the company target and the average.  
  
  
Employees who had 3 or fewer projects and left the organization ranked the company around 42% for satisfaction. In contrast those who left the company but had 4-5 projects ranked around 75% satisfaction. There seems to be a sweet spot between the number of projects assigned and satisfaction. Those under or over-worked at the company show signs of low satisfaction in contrast to those with a healthy workload.

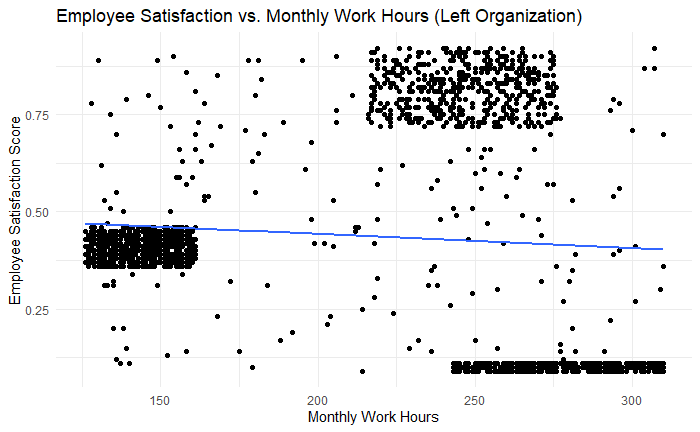


**Satisfaction in relation to hours worked**

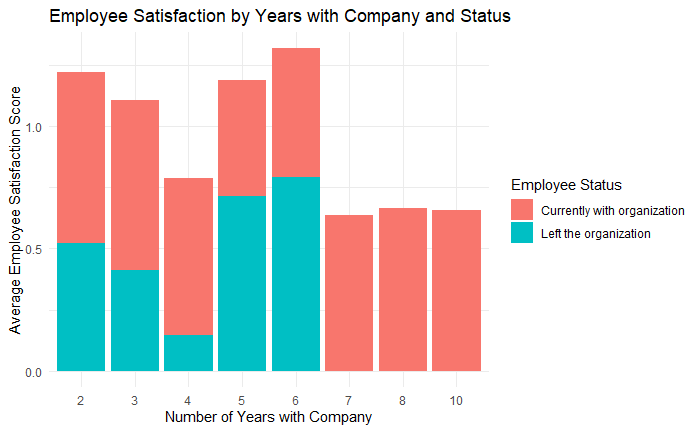
To further iterate the point made above, the number of projects likely correlates to the hours clocked in an organization. While the results of satisfaction and hours worked are a bit varied for employees currently at the organization, there is still a significant cluster in high satisfaction for employees clocking 150-300 hours.



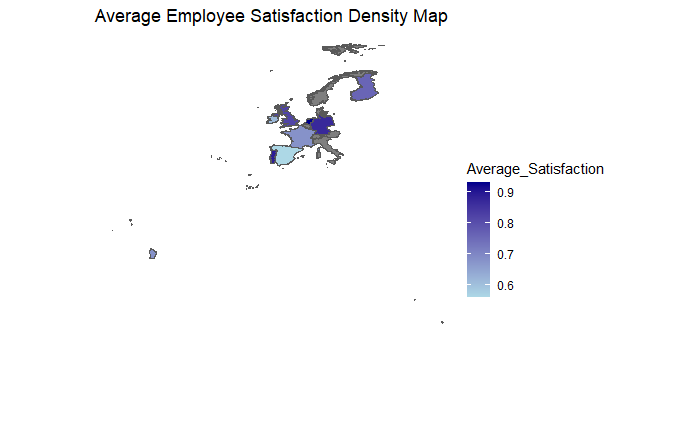
These points are made even more obvious by employees who left the organization. To the left of the graph below, employees who left the company and were underutilized (fewer hours worked) also had a very low satisfaction rating circled in red. The employees in the bottom right who left and were highly utilized (overworked) clocking around 250-350 hours a month, had an extremely low satisfaction rating. Interestingly the cluster in the top right shows employees who left the company with a high satisfaction rating. These employees worked 215-275 hours a month. Once again there is a clear middle spot where being underutilized may leave employees listless and uninspired, but those working over 250 hours a month are even more at risk of dissatisfaction with the company.



**Satisfaction in relation to retention**

Of the 10,000 points, there is no evidence of a single employee leaving after 6 years with the organization. These employees provide key insight into opportunities for higher retention. Interestingly, their satisfaction is not much higher - it hovers around the target 65-70% range. In contrast, employees with a higher satisfaction rating in their 5th and 6th years still left. There is something to be said about overall satisfaction in unison with longevity and retention.

**Does geography play a role in satisfaction?**

Germany and Portugal had the highest satisfaction rating above 90% while Spain had the lowest below 60%. We cannot say anything definitively about geography in relation to satisfaction as there could be many factors at play such as the departments hosted in these countries, taxes, quality of life, etc. But this data does provide us opportunity to deep dive into these different areas for more insights.

# Results and Next Steps

This analysis is meant to provide a high level look for company leaders to make data driven decisions. Compensation, better opportunities at competing firms, lack of encouragement and immediate supervisors are the top cited reasons for leaving. However, the regression analysis showed salary among other factors only predicted 3% of the satisfaction rating.

This analysis highlighted a clear trend in hours worked/projects undertaken and employee satisfaction. The company could address high-risk employees who are either under or over-utilized and bring them in alignment with other more satisfied employees of the company.

Practical next steps include:

* Monitoring the number of projects people are assigned
* Monitoring the overtime or hours people are working
* Focus on opportunities to develop employees

We look forward to partnering with you and discussing these results further.