Data OESON Internship Project 2

# Data Cleaning

From our data cleaning function, we assess the data to look for any null values and duplicate values within the dataset. There are no duplicate values, but for the Manager ID, there are 9 null values. Here are two errors we deal with:

* After quick observation, we realise that the Manager named “Webster Butler” had some missing fields for his Manager ID which we will in with 39
* Another observation is that the Manager “Brandon R. LeBlanc has been assigned to Manager ID’s, 3 and 1. We correct the error by changing the Manager ID to 1. However, note this may be incorrect since there could be two managers with the same name (however this is statistically more unlikely)

# Data Analysis

## Gender Analysis

A graph showing a number of blue rectangular objects

Description automatically generated

A comparison of a graph

Description automatically generatedMales earn slightly more on average than females ($70629.40 in comparison with $67786.73). This isn’t a significant difference, however there may be a cause for concern in gender equality, although this slight difference can be justified by there being 23 males in managerial positions compared to 18 females.

Looking at these histograms, we notice a female has the highest salary at 250,000, and the highest earning male is earning below 180k. However, more men earn over 100k in comparison to females. This data can be supported with the following data frame:

A screenshot of a computer screen

Description automatically generated

The main position “President & CEO” of the company is taken by a female. Furthermore, the CIO (Chief Information Officer) who is responsible for management, implementation and usability of information and computer technologies is also a female earning roughly $220k.

From this data, we can observe the gender split across all professions. Network Engineers get paid the least on average ($51.6k) whereas the President & CEO gets paid the most of $250k. Managerial positions are paid on average higher than regular workers, however software engineer managers get paid less than software engineers (perhaps because managing the workers is easier than doing the work itself). Also, the Principal Data Architect gets paid less than the Data Architect. This could suggest potential errors since Principal Data Architects should be paid more. We also notice this with senior database administrators (Sr. DBA) getting paid less than standard Database Administrators. We notice two female senior accountants getting paid roughly $102k to handle the financial aspects of the company.

We can use this data to support our histogram in justifying the size of each bin as we increase the salary for each gender.

A graph of different sizes and colors

Description automatically generatedA graph of a number of people

Description automatically generatedA graph of a number of positions

Description automatically generatedA graph of a number of positions

Description automatically generatedOverall, from our human resources analysis, this company has a higher number of females to males (176: 135). Production Technician 1 being the most popular career, this is the top position across both genders. The only key difference we can see here is that there are more males than females in Managerial Positions (Area Sales Manager and Production Managers). There are more females than males that work within production technician roles. If we exclude these roles we have 56 females and 61 males.

We observe the IT/IS department and Sales department has more males than females. On the other hand, production, admin and software engineering departments all feature more females than males. The executive office specifically for the CEO has one female. This company shows encouraging gender diversity with higher number of females being present in higher-paying roles.

# Manager Analysis

A graph with numbers and text

Description automatically generatedA graph of employees assigned by each manager

Description automatically generated

We can see the number of employees per each manager. Debra Houlihan has the smallest number of employees to manage with 3. The maximum number of employees managed by a single manager is 22. We also observe the number of employees in each department.

A table with numbers and a number of people

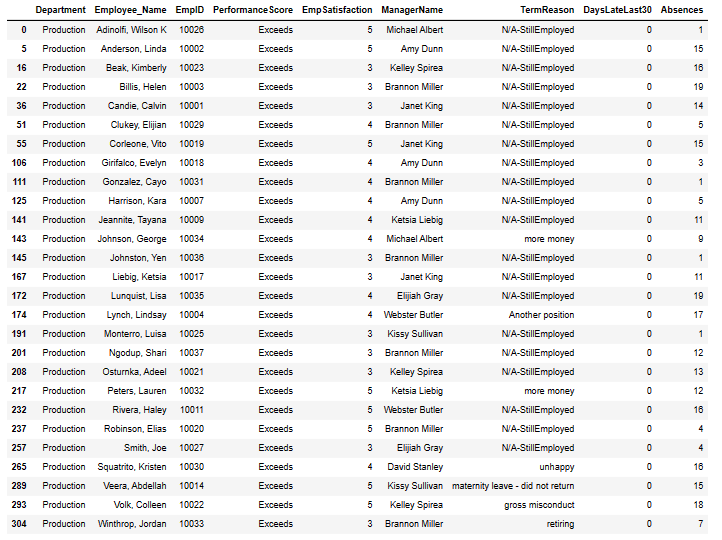
Description automatically generatedA table with numbers and letters

Description automatically generated

We focus on employee engagement under each manager. The Board of Directors manage the most engaged employees, as they are most likely to be in top-paying positions as well. Eric Dougall and and Brandon R. LeBlanc manage the second and third most engaged employees respectively. As per managers with the most employee satisfaction, we observe Debra Houlihan and Alex Sweetwater. Debra Houlihan has employees which do not engage as well, suggesting she is a laid back, nice manager which may favour employees but not the company itself. Alex Sweetwater on the other hand, is possibly the company’s best manager, ranking 2nd on employee satisfaction and 5th on employee engagement managing a total of 9 employees. From our findings, we recommend that new employees within the specific department be managed by Alex Sweetwater.

# A graph of performance scores Description automatically generatedEmployee and Department Analysis

Executive Offices and Admin offices all have good performance scores, fully meeting expectations. The IT/IS department is the best in terms of performance, with only a couple of employees which need improvement and should be fired. Employees which need improvement should be given a notice that if they don’t improve within the next 3 months, they will be fired. The Sales department is the worst, with a high number of employees falling in the PHP category that need to be fired. PHP employees should not be working in the company. Employees with exceeding performance should be rewarded, and given bonuses/ gifts for their hard work.



These are all the employees within the Production department that have are exceeding within their performance. As you can see none of them have been late for the last 30 days. Wilson Adinolfi, has only recorded 1 absence, and also has an employee satisfaction score of 5. This is our best employee within this department. Credit also goes to Luisa Monterro, Cayo Gonzalez and Yen Johnston for low absences and exceptional performance.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

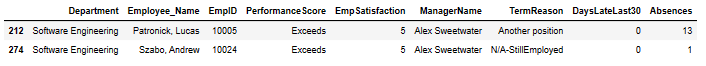
The production department has the highest number of employees which are exceeding performance requirements and not meeting performance requirements. We have already looked at the best employees in the Production department, now we can look at employees performing above expectations in other departments.

A table with text on it

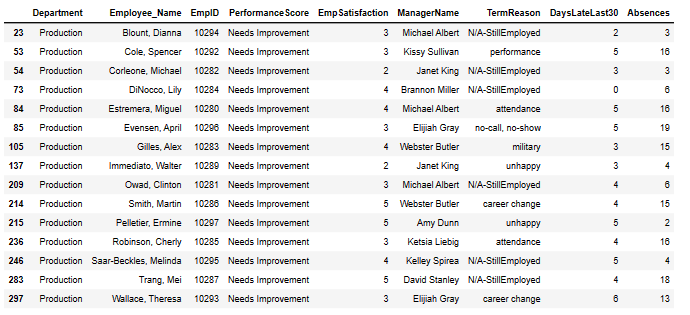
Description automatically generated

Within the IT department, Eric Dougall is our best employee with only 4 absences and an employee satisfaction of 5. He also manages an employee Lindsay Leonara himself!

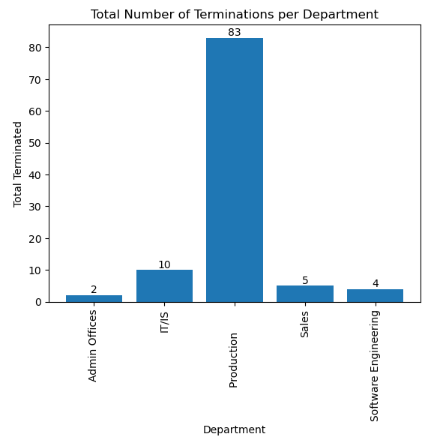
A hand holding a pen

Description automatically generated

For our software engineering departments and sales department, above are the 4 employees performing above expectations. All the employees performing above expectations should be given a bonus for their hard work.

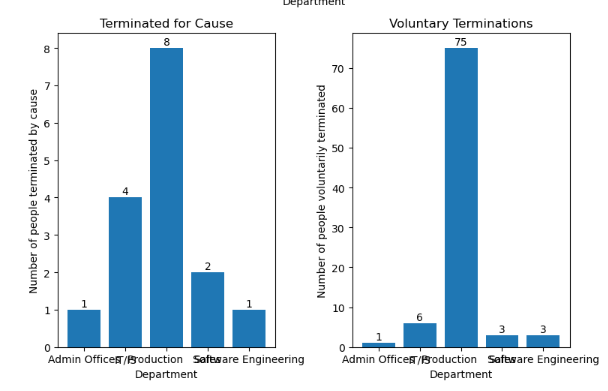


We observe all the employees within the production sector which need to have their performance reviewed. We see that some of these employees are not working anymore and have been terminated. However, Dianna Blount, Michael Corleone, Lily DiNocco, Clinton Owad, Melinda Saar-Beckles and Mei Trang are still employed and need to be monitored over the next few months. Above 5 days of being late and multiple absences should not be tolerated by these employees. The number of late showings needs to go down, and absences should be strictly disallowed unless there is an emergency.

The number of terminations within each department is proportional to the size of the department. IT/IS show the highest number in comparison with their size. Of course, Production will have the greatest number of terminated employees. We analyse the reasons behind these terminations and potential solutions to help the company avoid terminating employees in the future. We look at employees who have been “voluntarily terminated” (have left) and those who have been “terminated for cause”, meaning the company has fired them.

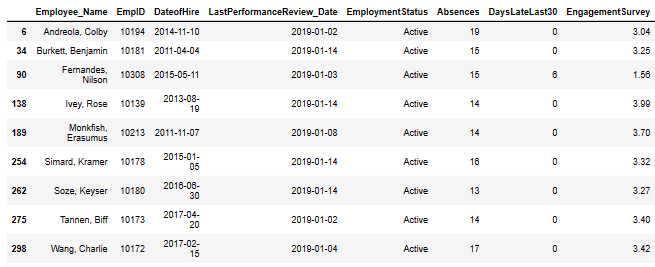
A graph with blue squares

Description automatically generatedA graph of blue bars with white text

Description automatically generatedAs we can see, most employees in the Production Sector have terminated their contracts voluntarily. IT/IS has shown to be the department which has most of their employees terminated for cause compared to other departments. We need to work on improving employee engagement and employee satisfaction within these departments, and taking a look at the managers methods in these areas. Further research can be done to look at the managers who have the most employees which are terminated.

There are two results which need to be considered, with one employee who has been voluntarily terminated for “performance” and another for “attendance”. These should both be in the “terminated for cause” section. This is a potential error; however, we do not correct this since we are past the data cleaning phase. The reasons which are of concern are “unhappy”, “more money” and “hours”. These are all things which we can improve on and potentially change. The other voluntary reasons are out of our control. We can work with managers and employees and design methods that would perhaps improve employee satisfaction. We could look at more pay rises for employees which need more money. Further analysis could be done into the professions of which employees need more money, so we know which positions we need to raise wages for. The idea of bonuses may incentivize workers. 14 workers left because they are unhappy, 11 wanted more money and 8 because the hours were too much. In total, 33 workers left due to conditions set by our company, we can reduce this, and work to increase employee satisfaction.

Analysing the reasons which people are terminated for by cause, background checks can be done on employees to avoid gangsters working for us.



These are the employees which need to be reviewed again. The last time they were reviewed was before the 15th of January 2019, so more than 15 days now, and their engagement is below 4.0. Furthermore, they have more than 10 absences. These are employees we need to monitor and review soon.

# Descriptive Statistics

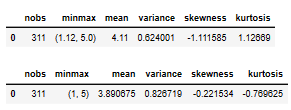
## Descriptive Statistics on Salary

A graph of a salary histogram

Description automatically generated



## Descriptive Statistics on Engagement and Employee Satisfaction



A comparison of a graph

Description automatically generated

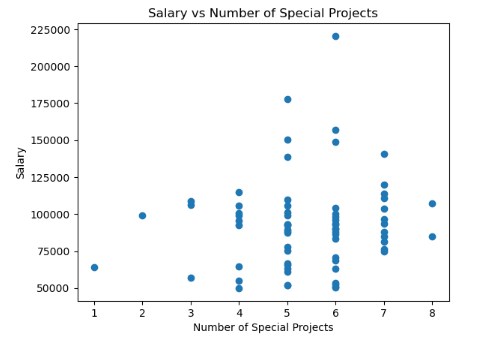
# Correlation Analysis

## Correlation between engagement x performance and salary x performance

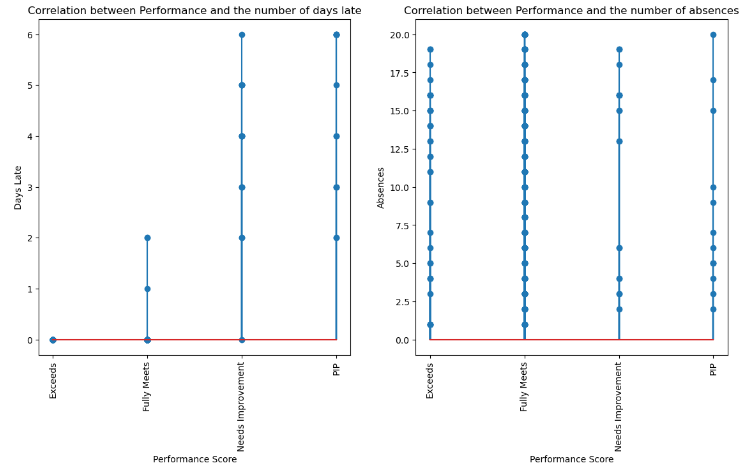
A comparison of a graph

Description automatically generated

## Correlation between number of Special Projects and Salary



## Correlation between Absences, number of days late in the last 30 and Performance Score



## Correlation between Employee Satisfaction and Engagement

A graph of a person with blue dots

Description automatically generated