Mariano Gentile

ABC   
performance analysis

Factors that impact on performance

Contents

[**Workforce Description** 1](#_Toc25436757)

[**Performance** 1](#_Toc25436758)

[**Performance and Worked Hours** 1](#_Toc25436759)

[**Collaboration and Working Hours** 3](#_Toc25436760)

[**Collaboration and Performance** 4](#_Toc25436761)

[**Skills and Performance** 5](#_Toc25436762)

[**Next steps** 7](#_Toc25436763)

[**Additional datasets** 7](#_Toc25436764)

# **Workforce Description**

The *main* (roster table) dataset contains records about 1324 employees from ABC Company employees. The region with most employees is Region 3, with 918 workers. The average tenure in years is 2.94 and the average Seniority is 8.6. Also, ABC employees present an average of 45.25 working hours a week and the mean Utilization of 75.7%.

# **Performance**

The company has a Performance Rating comprised of a 1-5 scale. Approximately 50% of the employees have been ranked with 3 points.

*Table 1. Employees % by Rating.*

|  |  |
| --- | --- |
| Performance Rating | Employees % |
| 1 | 2.04% |
| 2 | 25% |
| 3 | 48.19% |
| 4 | 23.04% |
| 5 | 1.74% |

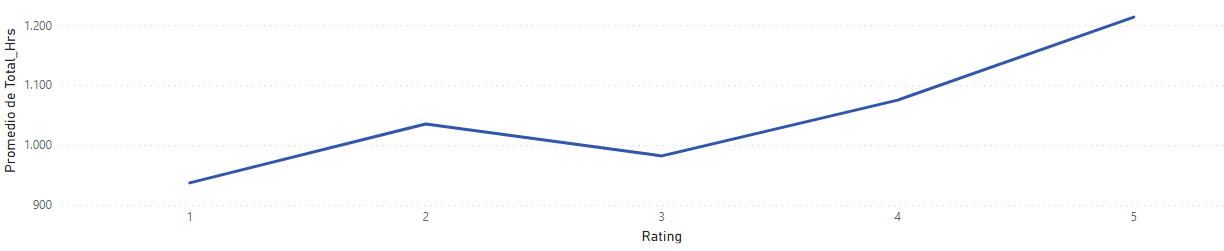
We can point out that regions 0 and 1 do not have employees with a rating of 5.

# **Performance and Worked Hours**

*Observation: for all calculations Total Working Hours (from Hours Table) had been considered, as it represent the actual worked hours by each employee during the period under study.*

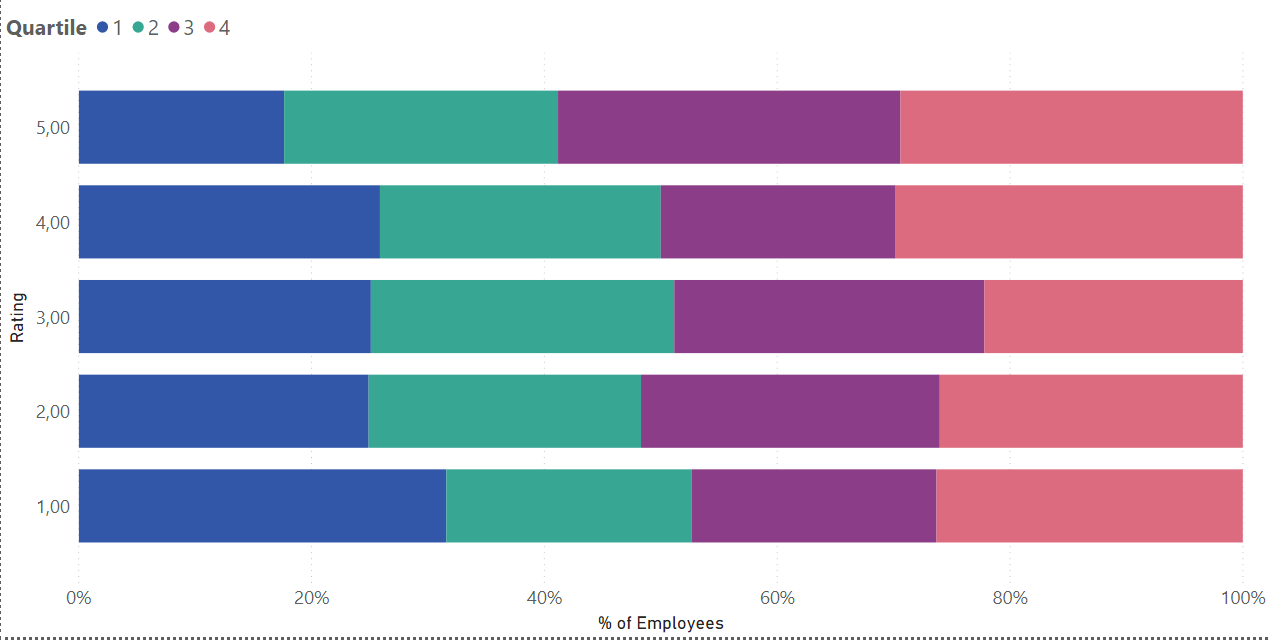
Taking into account the relationship between the mean hours worked throughout the period with the performance rating, a certain positive trend association could be considered, as higher performance ratings tend to have higher average working hours (except for Rating=3 where the mean is lower than for Rating=2).

*Chart 1. Average Total Working Hours by Rating.*



However, analyzing the distribution, that apparent association weakens. Practically all quartiles are almost equivalent among every Rating Level. For example: near 50% of employees in Levels 1, 2, 3 and 4 falls into one half of total working hours distribution, although the percentage drops to 40% in Level 5.

*Chart 2. Working Hs. Quartile Distribution by Performance.*

**

And if we plot the bivariate distribution in a scatter plot the correlation weakness (if there’s any at all) is more evident. Even though there are differences between Ratings 1 and 5, the concentration of cases in the middle diminishes the association.

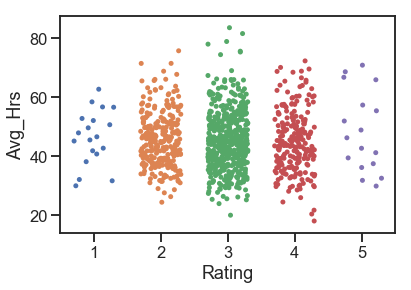
*Chart 3. Avg. Working Hours by Rating. Bivariate distribution. Complete Period (Hours Table).*

A screenshot of a cell phone

Description automatically generated

The relationship does not change considering Average Hours from Roster Table.

*Chart 4. Avg. Working Hours by Rating. Bivariate distribution. Roster Data.*



We could infer that the initial apparent association is due to considering the mean as summary statistic measurement, since this measure of central tendency is particularly susceptible to the influence of outliers. For further analysis of the relationship between these variables we can calculate a Spearman rank-order correlation coefficient (p) and the p-value to test for non-correlation, given one of the variables (Performance Rating) is of an ordinal measurement level.

Using Scipy Python’s Library the calculation would be:

Input:

from scipy.stats import spearmanr

# calculate spearman's correlation

coef, p = spearmanr(df['Rating'], df[‘total hs’])[[1]](#footnote-1)

print('Spearman’s correlation coefficient= %.3f' % coef)

# interpret the significance

alpha = 0.05

if p > alpha:

print('Variables are uncorrelated (fail to reject H0) p=%.3f' % p)

else:

print('Variables are correlated (reject H0) p=%.3f' % p)

Output:

**Spearman’s correlation coefficient= 0.012**

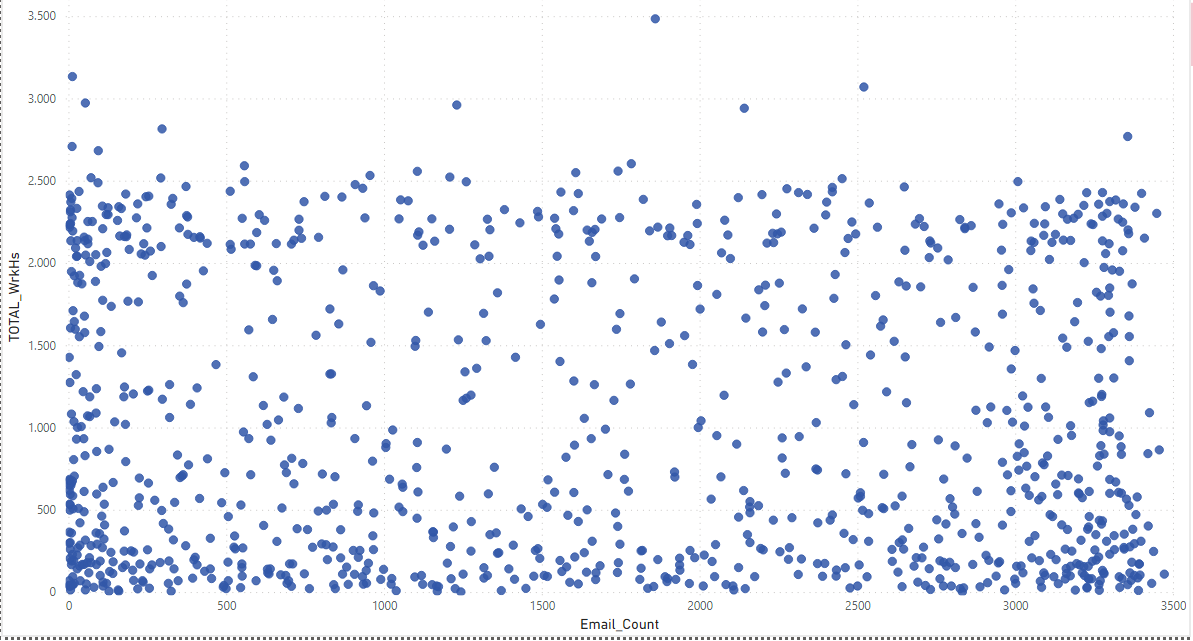
**Variables are uncorrelated (fail to reject H0) p=0.706**

In conclusion, even though there are some differences between working hours across performance ratings, they’re not statistically significant.

# **Collaboration and Working Hours**

The following chart shows there is **no** **correlation** between collaboration (proxy as count of emails sent by each employee to another employee) and total worked hours.

*Chart 5. Relation between collaboration and worked hours.*



*Source: Email Data Table and Hours Table.*

Being both continuous variables we can calculate Pearson’s Correlation Coefficient (r) to measure the strength of the linear relationship.

Input:

from scipy.stats import pearsonr

#Calculate Pearson's correlation

corr, \_ = pearsonr(dfp['Email\_Count'], dfp['TOTAL\_WrkHs'])

print('r= %.3f' % corr)

Output:

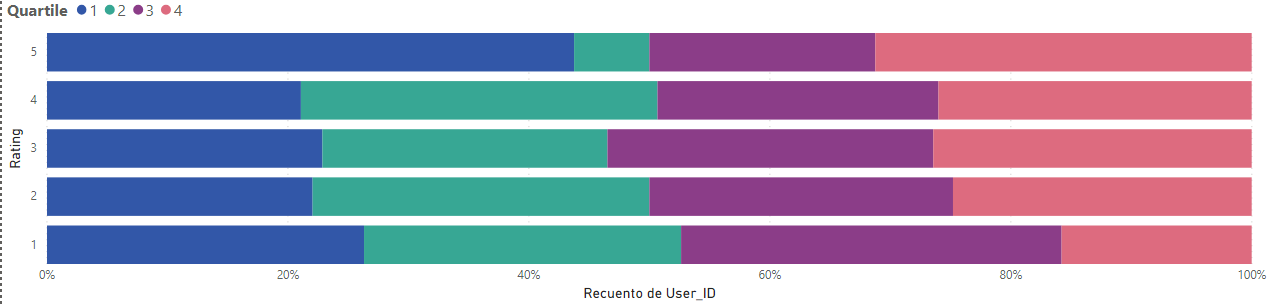
**r = -0.034**

The r = -0.034 value confirms that there is practically no correlation between collaboration and total worked hours.

# **Collaboration and Performance**

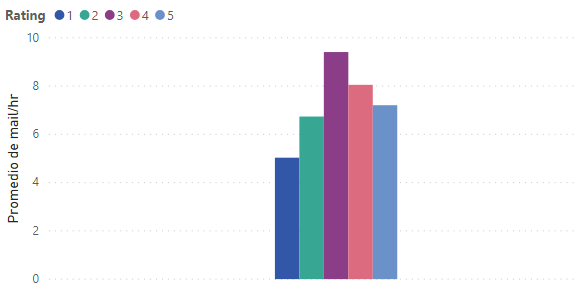
For the following analysis a measurement of Count of Emails / Working Hours has been created to have a synthetic measure of how many e-mails an employee sends by hour. Considering the quartile distribution, we can see that almost half of employees with rating 5 belong to the 25% of the company staff that sends less e-mails by hour, while the first quartile for the other rating levels contains approximately 25% of each level employees.

*Chart 6. Email/Hs Quartile Distribution by Performance.*



The mean of Email/Hs increases from Rating 1 to Rating 3, and then decreases again for Ratings 4 and 5, but stays higher than R1 and R2.

*Chart 7. Avg. Emails/Hs. by Rating.*



However, the Spearman’s coefficient shows no significance for this relationship.

Input:

dfmail = pd.read\_csv('EmailsHs by Rating.csv')

# calculate spearman's correlation

coef, p = stats.spearmanr(dfmail['mail/hr'], dfmail['Rating'])

print('Spearmans correlation coefficient: %.3f' % coef)

# interpret the significance

alpha = 0.05

if p > alpha:

print('Variables are uncorrelated (fail to reject H0) p=%.3f' % p)

else:

print('Variables are correlated (reject H0) p=%.3f' % p)

Output:

**Spearmans correlation coefficient: 0.004**

**Variables are uncorrelated (fail to reject H0) p=0.910**

# **Skills and Performance**

Skills dimension consists of three hierarchical levels: Group, Sub-Group and Attribute (the actual skill), combined with a skill level for Attributes: “Wants to Learn”, “Heavy Supervision”, “Light Supervision”, “Expert”. In the population there were no “Expert” records found.

Considering that there are 296 distinct Attributes I began with an exploratory analysis, highlighting the Top 3 Attributes by Performance Rating.

*Table 2. Top 3 Attributes by Performance Rating.*

|  |  |  |
| --- | --- | --- |
| Performance Rating | TOP 3 Attributes | Frequencies |
| 1 | Client Relations | 9 (1.98%) |
| Client Presentations | 8 (1.76%) |
| English | 8 (1.76%) |
| Public Speaking | 8 (1.76%) |
| 2 | Conducting Interviews | 95 (20.88%) |
| Client Presentations | 90 (19.78%) |
| Cross Office Collaboration | 86 (18.9%) |
| 3 | Client Presentations | 177 (39.9%) |
| Public Speaking | 167 (36.7%) |
| English | 167 (36.7%) |
| 4 | English | 88 (19.34%) |
| Public Speaking | 78 (17.14%) |
| Client Presentations | 78 (17.14%) |
| 5 | English | 7 (1.54%) |
| Public Speaking | 6 (1.32%) |
| Conducting Interviews | 6 (1.32%) |
| Cross Office Collaboration | 6 (1.32%) |

English and Public Speaking skills stand out as the most common skills across ratings.

The chi-square test for independence shows there is no association between Attributes and Performance Rating at 0.05 significance level.

Chi-Square Output

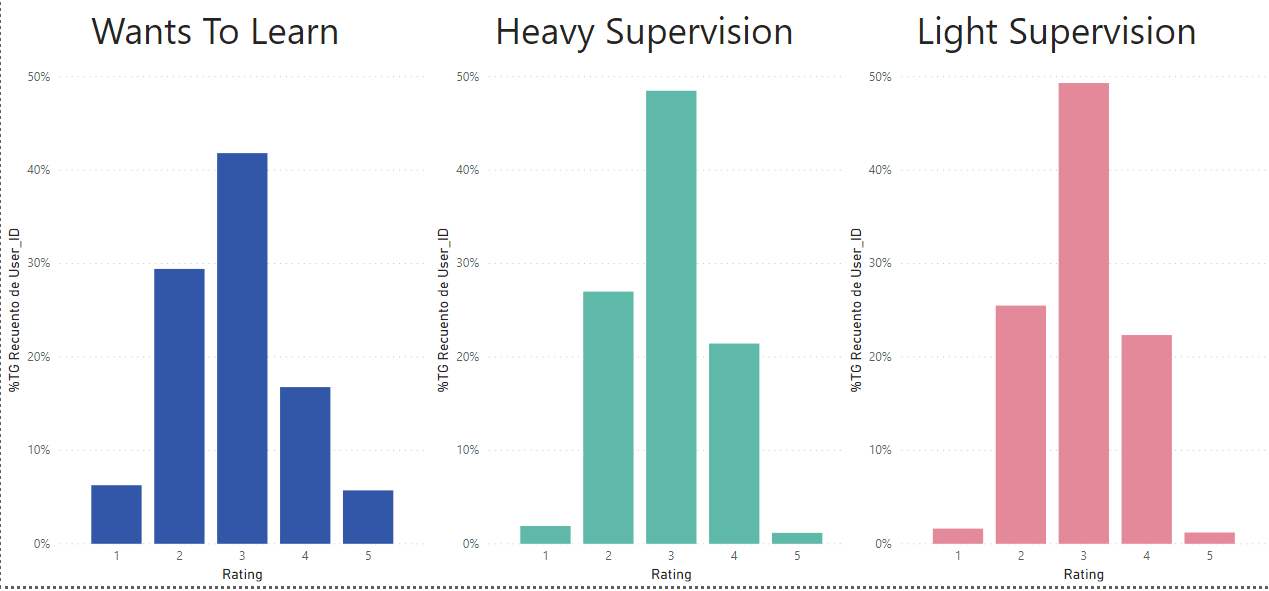
[296 rows x 5 columns]

Degrees of freedom = 1180

**Probability = 0.950, Critical = 1261.027, Stat = 1115.503**

**Independent (fail to reject H0)**

Another relevant dimension for skills is the Attribute Level.



The chart shows that, while there are no great performance differences by skills level in the central positions, as the proportions are higher and similar for rating 3, then for rating 2 and at last rating 4, the effect of skills level is more important within ratings 1 and 5. The proportion of “wanting to learn” employees in those categories is between 4 and 5 times the ones of the remaining skills levels.

In concordance Chi-square returns a significant association between Attribute Level and Performance with a probability of 95%, but Cramer’s V indicates the association is weak.

Chi-square output

[5 rows x 3 columns]

Degrees of freedom = 8

**probability=0.950, critical=15.507, stat=529.444**

**Dependent (reject H0)**

This means that, although higher skill level won’t necessary provide higher performance, and most employees will be rated 3 regardless their skill level, having greater skills reduces, slightly, the chance of a 1 or 5 performance rating.

# **Next steps**

For a more complete understanding of the factors that influence performance a multivariate analysis should be made, since the relationships found are not significant to explain the variations of performance. A machine learning model could be deployed to capture the patterns that impact on performance.

Having such a wide range of total working hours is a factor to analyze, with employees that have worked 4 hs. along the period while the one who worked more time reached 3.487 hs.

Another item to attend is the lack of experts within the company. A complete skills assessment and creation of learning paths shall be implemented.

# **Additional datasets**

An additional dataset that could be useful to understand performance might be comprised of employee satisfaction data. Relationships with peers and leaders, personal development and goals can have an impact on performance.

Also the Email Data table might be enriched with dates for a better correspondence with the worked hours in the estimation of the collaboration.

1. Pandas Dataframe columns containing values from Roster and Hours tables for each employee with available data. [↑](#footnote-ref-1)