# Analysis of job satisfaction in the data industry

Date: 2024-01-29

Prepared by: Gabrielė Tamaševičiūtė

# Project Tasks

1. Establishing correlations between job satisfaction and employee characteristics: Examine how various demographic and professional factors of employees, such as age, gender, education, current annual salary, and professional position, correlate with their level of job satisfaction.
2. Impact analysis of career change: Investigate whether transitioning into the data field affects job satisfaction, comparing those who have recently changed their career path to those who have been working in the data analysis field for a long time.

# These objectives will reveal important aspects of job satisfaction and help employers better understand what contributes to employee loyalty and satisfaction at work. The analysis process can utilize various statistical procedures, including Pearson's correlation coefficient, Spearman's rank correlation, Chi-square tests, etc., depending on the type and distribution of the data.

# Data Preparation

Firstly, checking for duplicates. The survey consists of responses from 630 distinct respondents and includes a total of 20 questions, with a mix of quantitative and qualitative types:

* **Quantitative Questions:** There are 7 quantitative questions in the dataset. These questions typically involve numerical responses and include items such as satisfaction ratings (e.g., "How Happy are you in your Current Position with the following? (Salary, Work/Life Balance, Coworkers, Management, Upward Mobility, Learning New Things)") and demographic information (e.g., "Current Age").
* **Qualitative Questions:** The survey contains 16 qualitative questions. These questions are more open-ended and include text or categorical responses, such as job titles ("Which Title Best Fits your Current Role?"), opinions or preferences (e.g., "Did you switch careers into Data?"), and demographic information (e.g., "Which Country do you live in?", "Highest Level of Education").

The dataset also includes additional metadata for each response, such as a unique ID for each respondent, email addresses (anonymized as 'anonymous'), and timestamps for when the survey was taken and completed.

Steps taken for data cleaning and preparation:

1. Identifying Missing Values by Conditional Formatting in Excel.
   1. In the column „Q12 - Highest Level of Education“ all blanks was replaces with „unknown“
   2. Identify respondents who have not completed the quantitative sections and remove them from the dataset.
   3. Locate individual missing entries within quantitative questions and substitute them with the average value of the respective question.
2. Standardization of Data.
   1. Column: Q1 - Which Title Best Fits your Current Role?:
      1. All roles that include the phrase "other: BI..." were updated to "BI Developer/Engineer."
      2. All roles that begin with "Software..." were consolidated under the title "Software Engineer."
      3. Remaining entries classified as "other" were uniformly relabeled as "Other."
   2. Column - Satisfaction Average was created to measure overall satisfaction index between different aspect.

# Descriptive Statistics

Employee satisfaction with aspects of work:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Salary | | Work/Life Balance | | Coworkers | |
| Mean | 4.289176 | Mean | 5.746365 | Mean | 5.8610662 |
| Median | 4 | Median | 6 | Median | 6 |
| Mode | 3 | Mode | 6 | Mode | 5 |
| Skew | 0.139478 | Skew | -0.2783 | Skew | -0.311298 |
| Variance | 7.668669 | Variance | 7.626504 | Variance | 7.307527 |
| St. dev. | 2.769236 | St. dev. | 2.761613 | St. dev. | 2.7032438 |
| CV | 0.645634 | CV | 0.480584 | CV | 0.4612205 |
| Management | | Upward Mobility | | Learning new Things | |
| Mean | 5.321486 | Mean | 4.749596 | Mean | 5.6009693 |
| Median | 5 | Median | 5 | Median | 6 |
| Mode | 5 | Mode | 5 | Mode | 10 |
| Skew | -0.20103 | Skew | -0.06198 | Skew | -0.19397 |
| Variance | 7.474149 | Variance | 7.887035 | Variance | 9.0104198 |
| St. dev. | 2.733889 | St. dev. | 2.808386 | St. dev. | 3.0017361 |
| CV | 0.513745 | CV | 0.59129 | CV | 0.5359315 |
|  |  |  |  |  |  |

**Overall employee satisfaction:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean: | Median: | Mode: | Skew: | Variance: | St. dev. |
| 5.261443 | 5 | 5 | -0.14425 | 8.138739 | 2.852848 |

1. The overall employee satisfaction is above average, with a mean of 5.26 on a 10-point scale. The mean, median, and mode are quite close to each other, indicating that the data distribution is relatively symmetrical.
2. Skewness: The distributions for all categories are slightly skewed, but the skewness values are quite small (all less than |0.3|), indicating no significant asymmetry. A positive skewness indicates more low-value entries, while a negative skewness indicates more high-value entries.
3. Satisfaction with salary: The average satisfaction with salary is the lowest among all categories, with a mean of 4.29. This suggests that salary may be one of the areas where employees feel the least satisfied.
4. Satisfaction with work-life balance and collaboration with colleagues: The mean values for these areas are higher, at 5.75 and 5.86 respectively, showing higher satisfaction in these areas.
5. Rating of management, career opportunities, and learning new things: These areas also have relatively high means (above 5), but "Learning new things" has the highest standard deviation and variance, indicating a greater spread of opinions in this area.
6. Coefficient of Variation (CV): The CV measures the ratio of the standard deviation to the mean and indicates the degree of data dispersion. A lower CV indicates less relative dispersion. In this case, satisfaction with "Colleagues" has the lowest CV, suggesting that ratings in this area are more concentrated around the mean.

|  |  |
| --- | --- |
| Mean | 29.88691 |
| Median | 28 |
| Mode | 25 |
| Standard Deviation | 7.245139 |

Average age:

A graph with numbers and a bar

Description automatically generated

The most frequently job role:

Distribution of educational level:

Gender distribution:

## 1. Determination of the correlation between job satisfaction and employee characteristics

## 1.1 Age

|  |  |
| --- | --- |
|  | Correlation Coefficient |
| Salary | 0.110187547 |
| Work/Life Balance | 0.036331709 |
| Coworkers | 0.041166917 |
| Management | 0.01433742 |
| Upward Milibility | 0.009501082 |
| Learning New Things | -0.010709028 |
| overall satisfacion | 0.039574923 |

* All the provided correlation coefficients are quite close to 0, indicating a weak relationship between age and the listed job satisfaction aspects.
* The highest positive correlation is between age and satisfaction with salary (0.110187547), which suggests greater satisfaction with salary as age increases. This could reflect higher salary expectations or that older employees may earn more due to longer tenure or higher positions.
* The negative covariance with learning opportunities, though weak, may indicate that older employees might be less inclined to value learning and development at work, which is important for organizations to know when planning training and development programs.

## 2.2 Gender

**These correlation coefficients are very small, indicating that gender is not a strong factor in job satisfaction in this dataset.**

|  |  |
| --- | --- |
|  | Correlation Coefficient |
| Salary | 0.00443265 |
| Work/Life | -0.018699326 |
| Coworkers | -0.001438272 |
| Management | -0.033631973 |
| Upward Mobility | -0.021702826 |
| Learning new | -0.061710981 |
| overall satisfacion | -0.027418334 |

Based on the provided data, where women are coded as "1" and men as "0," the correlation coefficients show the relationship between gender and various aspects of job satisfaction. Here's how we can interpret this data:

* Salary: The correlation coefficient is very close to zero (0.00443265), indicating that there is virtually no relationship between gender and satisfaction with salary.
* All other aspects and overall satisfaction: The negative correlation coefficients suggest that women might be slightly less satisfied overall than men, but, as with other areas, the relationship is quite weak.

## 2.3 Education

|  |  |
| --- | --- |
|  | Correlation Coefficient |
| Salary | 0.078810815 |
| Work/Life | 0.039105404 |
| Coworkers | 0.052567614 |
| Management | 0.030431352 |
| Upward Mobility | -0.010299865 |
| Learning new | -0.021019593 |
| overall satisfacion | 0.032982091 |

## From the provided data, we can conclude that there are very weak positive correlations between higher education levels and satisfaction with salary, work-life balance, relationships with colleagues, and supervisors. However, these relationships are extremely weak, indicating that the level of education is not a strong factor in job satisfaction. There are also very weak negative correlations between the level of education and satisfaction with career opportunities and learning new things.

## 2.4 Current annual salary

|  |  |
| --- | --- |
|  | Correlation Coefficient |
| Salary | 0.492502403 |
| Work/Life Balance | 0.227175864 |
| Coworkers | 0.218711564 |
| Management | 0.206913311 |
| Upward Milibility | 0.225830245 |
| Learning New Things | 0.19392293 |
| overall satisfacion | 0.313763952 |

All analyzed aspects of job satisfaction are positively correlated with salary, but the relationships range from weak to moderately strong. Since none of the correlation coefficients are close to 1, it indicates that salary alone is not a decisive factor in job satisfaction. However, a moderately strong correlation coefficient (0.492502403) between salary and employee satisfaction with the salary factor indicates that there is a certain positive relationship – generally, as salary increases, so does satisfaction with the salary.

## 2.5 Professional position

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Data Analyst | Data Engineer | Data Scientist | BI Developer/Engineer | Database Developer |
| Correlation Coefficient | Correlation Coefficient | Correlation Coefficient | Correlation Coefficient | Correlation Coefficient |
| Salary | 0.104072311 | 0.004892039 | 0.103119315 | 0.034564499 | 0.00361431 |
| Work/Life | 0.120724616 | 0.016190265 | 0.027778741 | 0.06235268 | 0.021375899 |
| Coworkers | 0.10955149 | 0.00567966 | 0.068276358 | 0.042953627 | 0.024687221 |
| Management | 0.123378148 | -0.007924483 | 0.059969355 | 0.033657968 | 0.048842331 |
| Upward Mobility | 0.118456624 | 0.037211386 | 0.059247838 | 0.040793782 | 0.014484288 |
| Learning new | 0.123004264 | 0.072170133 | 0.084749554 | -0.008620896 | 0.036075279 |
| overall satisfacion | 0.140685749 | 0.026821471 | 0.081350022 | 0.040598784 | 0.030069844 |

# Some positions, such as "Data Analyst" and "Data Scientist," show stronger positive correlations with satisfaction aspects than others. This may indicate that individuals in these positions experience greater satisfaction with their job aspects.

# Conclusions

Analyzing the provided data on job satisfaction relationships with age, gender, education, annual salary, and professional position, we can draw the following conclusions:

**Age:**

The relationship between age and aspects of job satisfaction is weak, but there is a slightly higher satisfaction with salary as age increases. This may indicate that rising salary expectations with age or higher earnings due to longer tenure and higher positions can positively affect satisfaction with salary.

**Gender:**

The relationship between gender and aspects of job satisfaction is also very weak. Although some coefficients are negative, they are still close to zero, indicating that gender's influence on job satisfaction is minimal. This may mean that in organizations, job satisfaction factors are relatively equal for both men and women.

**Education:**

The relationship between education and job satisfaction is extremely weak. Positive coefficients indicate slightly higher satisfaction in some aspects among those with higher education, but these relationships are too weak to be considered significant. This suggests that education is not a primary factor in job satisfaction.

**Current Annual Salary:**

The relationship between salary and job satisfaction is clearer, especially between salary and satisfaction with salary. A moderately strong relationship indicates that as salary increases, satisfaction with salary usually increases as well. However, the relationship between salary and other aspects of job satisfaction, although positive, is not very strong, indicating that there are other important factors.

**Professional Position:**

The relationships between professional position and job satisfaction vary depending on the specific position. For example, Data Analysts and Data Scientists show slightly stronger positive correlations with satisfaction aspects than representatives of other positions. This may indicate that specialists in these fields experience greater satisfaction with their job aspects, possibly due to the nature of the work or career opportunities.

Overall, the analyzed data show that while some demographic and professional factors have some influence on job satisfaction, their impact is quite limited. This may mean that job satisfaction is more influenced by specific work environments, organizational culture, and individual personality traits than solely demographic or professional characteristics. Therefore, it is important for organizations to consider a wide range of factors to improve employee satisfaction. This includes not only competitive salary policies and career opportunities but also the quality of the work environment, leadership style, employee recognition and inclusion, and work-life balance. Furthermore, given the minimal impact of demographic factors, organizations should individualize their approach to each employee, considering their unique needs, expectations, and motivations. This means it's important not just to create general well-being and satisfaction enhancement programs but also to allow flexibility and personalize strategies to meet the interests of various employee groups and individuals.

## Analysis of the influence of the change of profession

To investigate whether a career change into the data field affects job satisfaction, you should apply the following hypotheses:

* Null Hypothesis (H0): A career change into the field of data analysis does not affect job satisfaction.
* Alternative Hypothesis (H1): A career change into the field of data analysis affects job satisfaction.

Using statistical methods, compare these groups:

* Group 1: People who have recently changed their career to the field of data analysis.
* Group 2: People who have been working in the field of data analysis for a long time.

To accurately determine whether there are statistically significant differences in job satisfaction levels between these groups, we will use a two-sample t-test, assuming equal variances.

|  |  |  |
| --- | --- | --- |
|  | ***swiched*** | ***been long*** |
| Mean | 5.503731343 | 5.013816926 |
| Variance | 5.143140415 | 4.974055783 |
| Observations | 268 | 193 |
| Pooled Variance | 5.072412203 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 459 |  |
| t Stat | 2.304138284 |  |
| P(T<=t) one-tail | 0.010830306 |  |
| t Critical one-tail | 1.648180137 |  |
| P(T<=t) two-tail | 0.021660612 |  |
| t Critical two-tail | 1.965145755 |  |

From the conclusions of the two-sample t-Test with equal variances, the findings are as follows:

* Means: Individuals who switched their careers to the field of data analysis ("switched") have a slightly higher average job satisfaction level (about 5.50) compared to individuals who have been working in this field for a long time ("been long", average about 5.01).
* t-Statistic: The obtained t-statistic is 2.304138284, which is higher than the critical t-values for both one-tailed (1.648180137) and two-tailed (1.965145755) tests.
* p-Values:
  + For the one-tailed test, the p-value is 0.010830306, which is less than the typical 0.05 threshold, hence we reject the null hypothesis.
  + For the two-tailed test, the p-value is 0.021660612, which is also less than 0.05, thus we also reject the null hypothesis here.

Given the p-values and t-statistic, we have sufficient evidence to reject the null hypothesis (H0) that a career change into the field of data analysis does not affect job satisfaction and accept the alternative hypothesis (H1) that a career change into the field of data analysis does affect job satisfaction.

However, it is important to note that while a statistically significant difference is established, the actual difference between the group means is relatively small, meaning that while the effect exists, its practical (real-world) impact may not be as significant. Moreover, we must remember that correlation does not imply causation – we cannot definitively say that career changes caused changes in job satisfaction without further analysis or additional data.