

Research Paper

Uncovering the Impact of Working from Home on Employees' Collaboration, Happiness, and Promotion Chances Using Machine Learning

Mehdi Imani*

Department of Computer and System Sciences, Stockholm University, Stockholm, Sweden.
Email ID: m.imani@gmail.com

Received: 16/10/2023,

Revised: 07/11/2023,

Accepted: 19/12/2023

Published: 03/01/2024

Abstract: - This research explores the impact of working from home on employees' happiness, collaboration, and promotion prospects using machine learning techniques. The study is guided by three research questions aiming to investigate the correlation between working from home and employees' collaboration and promotion prospects. Moreover, the research aims to find a relationship between the number of households and employees' happiness levels while working from home. The data is collected from ICT engineers working at a technology company in Sweden through a questionnaire-based survey. Probability sampling was selected for data collection to reduce bias and enhance the generalizability of the findings. The data is pre-processed and then analysed in Jupyter Notebook using the Python programming language. Various libraries and models, including Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn, were employed for data analysis. Both Pearson correlation and p-values in Pearson correlation were used in this study to analyse the relationships between different variables. However, based on the results, this study did not find any significant relationship between working from home and employees' promotion prospects or collaboration issues. Additionally, the results did not provide evidence of a significant relationship between the number of households of employees and their happiness levels while working from home.

Keywords- Machine learning, COVID-19, Working from home, Employees' collaboration, Employees' happiness, Employees' promotion chances.

1. Introduction

The pandemic of COVID-19 has caused a significant rise in the number of people working from home [WFH], with WFH participation levels increasing substantially since March 2020 [1-3]. Businesses and policymakers have actively supported telecommuting as a way to prevent the spread of COVID-19 [4-6].

1.1. Background

While working remotely has been popular in organizations with advancements in information and communication technology, prior to the pandemic, there was already a significant increase in remote work [7]. Although working from home has its benefits, such as increased temporal flexibility, autonomy, and productivity, scholars have also noted potential threats to employee well-being, such as physical and temporal separation,

social isolation, and blurred work-home boundaries [8-10]. Despite this, research has shown that both local worksites and teleworkers can experience increased performance and decreased absenteeism when offered alternatives to the current work environment, and key factors such as comfortable workplaces, trust in superiors, and positive work relationships with colleagues can contribute to this [11,12]. In Sweden, the pandemic lockdown required all employees to work remotely, and this study aims to investigate the viability of continuing remote work initiatives by assessing work satisfaction, performance, and productivity levels in the ICT industry [13].

1.2. Research Problem

Due to the COVID-19 pandemic, many employees have been required to work from their homes, which has raised concerns about maintaining their job performance and job satisfaction in this uncertain situation. Therefore,

there is an urgent need for research that identifies effective approaches and interventions to support remote workers during the COVID-19 era. This will assist both practitioners and scholars in better understanding how to assist remote workers. Workers can feel that WFH makes a positive impact on their work-life balance and consequently helps to minimise workplace costs. When WFH, employees can feel that they have to prolong the working day or put in extra time to do their personal jobs, which could decrease job satisfaction. This is an example of when a considerable portion of WFH employees are likely to suffer from negative outcomes such as work-family conflicts and health issues [14,15]. Recent research shows that many factors can influence both performance and motivation. Identifying the correlations between these factors is the main goal of this research.

As a result of the lockdown, the internet has become the primary mode of communication for most people. Consequently, there is a question about whether employees should be allowed to work from home after the lockdown. Online transactions have increased substantially due to employees' teleworking or remote work situations, as demonstrated in some developed countries, where public sector or government employees have been teleworking for years due to the perceived benefits [16]. However, it is unknown whether the perceived benefits extend to developing countries' employees during the lockdown. Regardless, job satisfaction is critical for both employees and administrators of the organization. Thus, this study's goal is to provide empirical evidence on the feasibility of remote work for an ICT company in Sweden.

Several research investigated the relationship between working from home, job performance, and happiness. For example, the relationship between the number of households and employees' job satisfaction while working remotely was investigated by Sridhar and Bhattacharya [17], but the authors did not investigate the impacts of the number of households on employees' happiness. For another example, authors [18] investigated the impact of working from home on a team's performance, and authors [19] investigated the impact of working from home on employees' salary. However, no prior research investigated the impact of working from home on employees' collaboration and chance of promotion.

1.3. Aim and Research Question

This study aims to investigate the correlation between working from home and employees' collaboration and promotion chances. Moreover, the study aims to find a relation between the number of households and the level of employees' happiness while they work from home for ICT engineers employed at a technology company based in

Sweden. To achieve these objectives, the study was guided by the following research questions:

- Is there any correlation between working from home and employees' collaboration?
- Is there any correlation between the number of households of employees and employees' happiness while working from home?
- Is there any correlation between working from home and employees' promotion chances while working from home?

1.4. Delimitations of the Study

This study is limited to examining the impact of WFH on employee satisfaction and happiness. Due to time and resource constraints, a comprehensive and in-depth investigation is not feasible. To narrow the scope of the study population, only engineers from the ICT department of the company are included in this study. The study aims to identify the factors that contribute to collaboration between WFH and employee satisfaction, using a quantitative approach. Moreover, the study will only investigate the impact of WFH on employee performance and happiness in Sweden. The decision to limit the study to Sweden was based on the accessibility of the company engineers in Sweden and the availability of a reasonable empirical base to analyse and draw conclusions. Additionally, the ability to reach the company engineers in Sweden via email and Slack further justifies this geographical delimitation.

1.5. Hypothesis

The pandemic of COVID-19 has led to a significant increase in remote work, and there is a need to understand the impact of WFH on employee productivity and job satisfaction. This study presents two hypotheses that aim to examine the relationship between WFH and employee outcomes. The hypotheses are as follows:

Hypothesis 1: Working from home has an impact on collaboration.

H₀: There is no significant correlation between working from home and collaboration.

H₁: There is a significant correlation between working from home and collaboration.

Hypothesis 2: The number of households of employees has an impact on employees' happiness while WFH.

H₀: There is no significant correlation between the number of households of employees and Employees' happiness.

H₁: There is a significant correlation between the number of households of employees and Employees' happiness.

Hypothesis 3: Working from home has an impact on employees' promotion chances while WFH.

H₀: There is no significant correlation between working from home and employees' promotion chances.

H₁: There is a significant correlation between working from home and employees' promotion chances.

Understanding these relationships will help organizations develop effective WFH policies and practices that balance productivity and employee well-being.

2. Methodology

2.1. Research Strategy

Denscombe [20] outlines several research strategies that can be employed to conduct successful research, including case studies, grounded theory, surveys, and action research. The selection of an appropriate research strategy depends on various factors, such as the research aim, ethical considerations, and feasibility in light of constraints [20]. Although the case study approach is suitable for small-scale, in-depth studies that aim to explain why specific outcomes might occur [20], it was not considered for the present study as the research aim is to describe how certain factors influence the population on a broader scale quantitatively. Similarly, grounded theory, which involves generating theories from participants' experiences and is reliant on qualitative data, was also rejected due to the uncertainty in planning the project's duration and the aim of the present study to capture quantitative data [20]. Action research, which is typically employed in real-world settings to investigate practical problems and generate possibilities for change, was not deemed suitable for answering the research question or fulfilling the purpose of the study [20]. Ethnography, which investigates individual cultures or social relations [20], and phenomenology, which is concerned with interpretation rather than measurement [20], were also ruled out as they do not align with the research aim. After considering the various options, a survey was deemed the best fitting strategy for this study.

2.2. Data Collection Method

The present study collected individual data from engineers in a Swedish ICT company to investigate the research questions. A web-based questionnaire designed on Google Form was used as the instrument to collect primary data. This method was deemed the most suitable for the study due to its cost-effective distribution and the ability to reach respondents regardless of their location. According to Denscombe [20], web-based questionnaires are ideal for research projects with limited resources and enable the collection of standardized data from participants electronically, thereby facilitating quantitative analysis and hypothesis testing, which is necessary to investigate the research question. The questionnaire allowed for the

calculation of descriptive and correlational analyses for all variables used in the study. Alternative data collection approaches, such as interviews and face-to-face surveys, were considered but ultimately deemed unsuitable for the research question. Interviews were rejected due to their resource-intensive nature, and Denscombe [20] notes that they are better suited for studies requiring in-depth insights about specific topics or phenomena. Face-to-face surveys were also deemed impractical due to the time and scheduling required to conduct them successfully. The use of web-based questionnaires was preferred due to their ease of distribution and the ability to reach respondents regardless of location or time constraints, allowing for a higher response rate and more successful data collection. Several questions are used in the questionnaire and some of them are inspired by the research, which is done by Sridhar and Bhattacharya [17]. Due to the fact that a significant number of engineers in the company operate in a hybrid capacity and often work remotely, as well as being located in various offices, including Stockholm, Gothenburg, and Lund, it was determined that conducting a web-based survey would be the most efficient method of communication. The engineers were contacted via the company's Slack channel and provided with a link to an online questionnaire. The resulting dataset includes responses from 67 employees, of which 59.7% were male and 40.3% were female, and the average range of participants is 47.9 [SD = 9.32], from 28 to 64 years old.

2.3. Sampling

Probability sampling and convenience sampling are two widely used methods of selecting participants for research studies. Probability sampling refers to the selection of some samples from a population in which every one of the members has the same probability of being chosen [21]. This method of sampling helps to reduce bias and increase the generalizability of the findings. According to Bryman [22], there are different methods for achieving probability sampling, including systematic sampling, simple random sampling, stratified sampling, and cluster sampling. In contrast, the convenience sampling technique is a non-probability sampling technique that involves choosing members based on accessibility and availability [23]. Researchers often use this method when time and resources are limited or when the population under study is difficult to access. Convenience sampling is often prone to selection bias and may lead to results that are not generalizable to the larger population [24]. This method is often used in research studies due to its practicality and low cost.

However, there are some potential drawbacks to using convenience sampling. One of the main issues is selection bias, which means that the sample of participants may not

be representative of the larger population being studied. This can occur because the participants in the sample may not be a random selection and may have certain characteristics that differ from those in the larger population. As a result, the conclusions drawn from a convenience sample may not be generalizable to the larger population, and the findings may not accurately reflect what would be observed if the entire population were studied. Therefore, researchers should be cautious when using convenience sampling and consider other sampling methods that are more likely to produce representative samples.

The difference between probability sampling and convenience sampling is that probability sampling involves the use of random selection techniques to ensure that every population member has an equal probability of being chosen. In contrast, convenience sampling relies on the ease of access to potential participants, and researchers have little control over who is selected [22,25]. As a result, probability sampling is often preferred when a representative sample is necessary, while convenience sampling may be more practical when the study is exploratory or when the population under study is difficult to access [24]. Both probability sampling and convenience sampling are useful methods of selecting participants for research studies. Probability sampling provides a more rigorous and representative sample, while convenience sampling may be more practical when time and resources are limited. However, researchers need to be aware of the potential biases associated with convenience sampling and adjust their study design accordingly.

For this study, based on all the above explanations, probability sampling was chosen for collecting data to reduce bias and increase the generalizability of the findings. In addition, based on the research questions, the target audience for the survey was engineers who work in the IT section. The reason for selecting IT engineers is that the study aims to find out the correlation between working from home and collaboration and promotion chances and working from home is much more popular in the section of IT, compared to other sections such as factory workers. Moreover, the study aims to investigate the relationship between the number of households and employees' happiness while working from home. To have better and more precise insights, the target audience for the survey was people in different age groups, ranging from 28 to 64 years old.

2.4. Data Analysis Method

Several different variables in this study using a five-point Likert scale ranging from 1 [strongly negative/disagree] to 5 [strongly positive/agree], including collaboration problems, reduced promotion chance, the

impact of WFH on the employer, feeling left out, Feeling better with WFH, Feeling more active with WFH. In this study, all analyses were conducted with Python programming language in the Jupyter Notebook environment using NumPy, Pandas, Matplotlib, Seaborn, and Scikit-learn. Both Pearson correlation and p-values in Pearson correlation were used in this study to analyse the correlation between different variables. The type of data collected in this study was either integer or string. In this study, the categorical string variables are converted to numeric variables using Pandas `get_dummies()` method in Python [26,27]. Pearson's coefficients and P-values are used to investigate the research questions in this study. Similarly, Al Jassmi et al. [28] examined the correlation between the level of happiness exhibited by construction workers and their productivity. This investigation involved the utilization of statistical measures such as p-values and Pearson's coefficients.

2.4.1 Pearson Correlation

Pearson correlation is one of the statistical tools that we use to evaluate both the magnitude and the direction of the linear relationship between two quantitative variables. It computes a correlation coefficient [r] that varies between -1 and +1, where values closer to -1 indicate a robust negative correlation, values closer to +1 indicate a strong positive correlation, and values near zero signify negligible correlation. This technique is widely used in research fields such as medicine, psychology, and business, among others. According to Field [29], Pearson correlation is "one of the most frequently used statistical techniques in data analysis" [p. 33]. Pearson correlation is useful for exploring relationships between variables and can help researchers identify patterns and trends that might not be immediately obvious. In addition, Pearson correlation can be used to make predictions about the value of one variable based on the value of another variable. However, it is worth mentioning that correlation does not imply causation and additional research is often necessary to establish causal relationships between variables. Overall, Pearson correlation is a valuable tool for researchers who seek to better understand the relationships between different variables in their data.

2.4.2 P-values

The p-value in Pearson correlation is an important measure of statistical significance that helps researchers determine the probability of observing a correlation coefficient between two variables due to chance. According to Sullivan and Artino [30], "a p-value less than .05 is generally considered statistically significant" [p. 53]. For example, in a study examining the relationship between performance and WFH, a Pearson correlation coefficient equal to 0.65 with a p-value equal to 0.025

would indicate a significant positive correlation between the two variables. This would suggest that participants who work from home are more likely to achieve better performance. On the other hand, a Pearson correlation coefficient equal to 0.35 with a p-value equal to 0.35 would suggest that the observed correlation is not statistically significant, and any relationship between the two variables is probably accidental. Therefore, understanding the p-value in Pearson correlation is crucial for making informed decisions about the relationship between different variables.

2.5 Research Ethics

Research ethics are principles and guidelines that govern the conduct of research to ensure that it is conducted in a responsible and ethical manner. Ethics in research involves ensuring the protection of human subjects, animal welfare, and the integrity of the research process [20]. Researchers are expected to adhere to ethical standards and regulations to ensure that their research is conducted with integrity, transparency, and accountability. The use of human subjects in research, in particular, requires researchers to obtain informed consent from participants, protect their confidentiality and privacy, and ensure that their rights and welfare are respected [31]. Ethical considerations in research also extend to the dissemination of research findings, which should be accurate and transparent, and avoid misleading or biased reporting [32].

Research ethics play a critical role in ensuring that research is conducted in a responsible and ethical manner. Adherence to ethical principles and guidelines is essential for protecting the rights and welfare of human subjects, maintaining the integrity of the research process, and promoting transparency and accountability in research. In this study, all the participants were informed about the purpose, the procedure, and the confidentiality of the survey. The informed consent form can be found in Appendix A

3. Results

3.1 Data Collection and Analysis

The questionnaire was designed to include 12 factor-related questions, with all questions and associated references provided in Appendix B. The survey was available for participants to complete for one month in February 2023, with the survey link distributed via email and a Slack channel. Participants' responses were automatically collected and stored in a spreadsheet for later analysis. Prior to analysis, data pre-processing steps were performed in Jupyter Notebook to ensure that all data

were complete and free of errors. The questionnaire was structured in a closed-answer format, requiring participants to provide numerical [integer] or text [string] responses for each question. In this study, a pre-processing phase is used to prepare the data for analysis. The pre-processing phase includes the following steps:

Data Cleaning: This step includes several tasks such as handling missing values, handling duplicates, and finding and correcting typos. For handling missing values, there are several methods that can be used depending on the concepts and variables including 1- Dropping missing values which remove all the rows and columns that contain missing values using `dropna[]` method from Pandas library. 2- Replacing missing values which replace the missing values with the mean, median, or mode of the values of the variables using `fillna[]` method from Pandas library. 3- Interpolation which estimates the missing values based on other variables' values using `interpolate[]` method from Pandas library. 4- Imputation which predicts the missing values using some machine learning techniques such as Random Forest, Support Vector Machines [SVM], and K-Nearest Neighbors [KNN]. All these algorithms are available in scikit-learn library for Python. In this study, based on the concept and needs, replacing missing values using `fillna[]` method in Pandas library is used to handle the missing values. In this study, we employed the identical methodology as in our prior research [33].

Handling Outliers: For handling outliers, we wrote the below function that uses the Interquartile Range [IQR] method to remove outliers from a dataset. The IQR is a measure of the spread of the data, defined as the difference between the third quartile [Q3] and the first quartile [Q1].

```
#function for removing outliers
def remove_outliers[train,labels]:
    for label in labels:
        q1 = train[label].quantile[0.25]
        q3 = train[label].quantile[0.75]
        iqr = q3 - q1
        upper_bound = q3 + 1.5 * iqr
        lower_bound = q1 - 1.5 * iqr
        train[label] = train[label].mask[train[label]<
lower_bound, train[label].median[],axis=0]
        train[label] = train[label].mask[train[label]>
upper_bound, train[label].median[],axis=0]

    return train
```

The function first calculates the IQR for each label [i.e., column] in the dataset by using the `quantile[]` method to find the 25th and 75th percentiles. Next, the function identifies outliers by finding values that are more than 1.5 times the IQR away from the upper or lower quartiles.

These are considered extreme values that are unlikely to be representative of the underlying population and are typically removed. To remove the outliers, the function replaces values that are less than the lower bound with the median value of the column and values that are greater than the upper bound with the median value of the column. This is done using the mask[] method, which replaces values in the specified range with the median value of the column. By using the IQR method to identify and remove outliers, this function provides a quick and simple way to clean datasets and improve the accuracy of statistical analyses.

Data Integration: This step aims to form a unified dataset from different sources using several tasks such as handling inconsistent data structures, handling data format differences, and handling naming conflicts. This study did not use this step as all the data were collected in a single source using a web-based survey.

```
sns.scatterplot(data=df3, x="Feeling_better_with_wfh", y="household")
<matplotlib.axes._subplots.AxesSubplot at 0x26638e4af98>
```

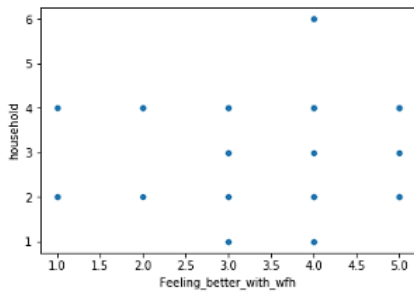


Figure 1 The scatterplot between the two variables: "working from home percentage" and "number of employees' households"

```
sns.scatterplot(data=df3, x="wfh%", y="collaboration_problem")
<matplotlib.axes._subplots.AxesSubplot at 0x26639e67828>
```

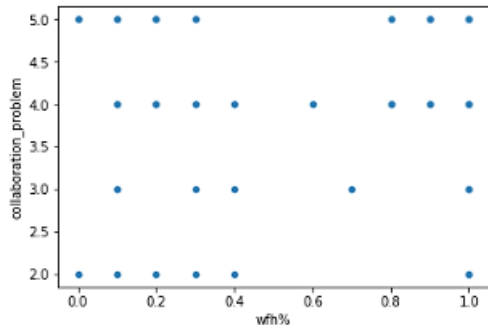


Figure 2 The scatterplot between the two variables: "working from home percentage" and "employees' collaboration problem"

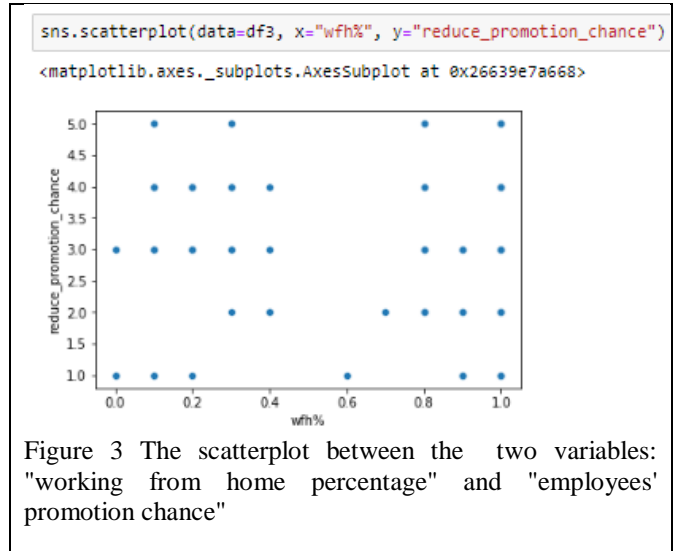


Figure 3 The scatterplot between the two variables: "working from home percentage" and "employees' promotion chance"

Data Transformation: This step aims to convert the type of data into a suitable format for analysis. As mentioned earlier, in this study, the categorical string variables are converted to numeric variables using Pandas get_dummies[] method in Python [26,27] which is more suitable for quantitative analysis.

Data Reduction: This step includes several tasks to reduce the amount of data by removing noisy data, irrelevant data, and redundant data. To achieve this goal several dimensionality reduction techniques such as feature selection and principal component analysis [PCA] can be used. As all the questions in the survey were designed based on the specific purpose, there were no noisy and irrelevant data and therefore this study did not need to use this step for reducing the data.

Data Discretization: This step includes converting continuous data into discrete categories using some techniques such as binning, equal-width binning, and equal-frequency binning. This study used binning technique to form discrete categories for three variables such as year of experience, age, and the number of households.

Linearity check: This step refers to the relationship between two variables being analysed. Specifically, it refers to whether the relationship between the variables is a straight line or whether it exhibits a curved pattern. When conducting a Pearson correlation analysis, it is important to check for linearity to ensure that the relationship between the variables is accurately captured. A scatterplot can be used to visualize the relationship between two variables. If the points on the scatterplot form a straight line, then the relationship between the variables is linear. Figures 1 to 3 show the scatter plot between the variables.

Normality check: This step refers to whether the data being analysed follows a normal distribution or bell curve. In the context of a Pearson correlation analysis, it is important to check for normality to ensure that the data being analysed is appropriate for the analysis. A histogram can be used to check for normality. In a histogram, the frequency distribution of the data is displayed as bars. If the bars form a bell-shaped curve, then the data follows a normal distribution.

Homoscedasticity check: This step refers to whether the variability of the data is consistent across all levels of the variables being analysed. When conducting a Pearson correlation analysis, it is important to check for homoscedasticity to ensure that the results accurately reflect the relationship between the variables being analysed. A scatterplot can be used to check for homoscedasticity. If the variability of the data is consistent across all levels of the variables being analysed, then the scatterplot will show the points evenly scattered around the line of best fit. Alternatively, a residual plot can be used to check for homoscedasticity. In a residual plot, the residuals are plotted against the predicted values. If the points on the residual plot are evenly scattered around the horizontal line at zero, then the data is homoscedastic. As we do not aim to predict a variable based on other variables in the dataset, then calculating the residuals using the best-fit line is not the case in our study. In summary, to conduct a Pearson correlation analysis, it is important to check for linearity, normality, and homoscedasticity to ensure that the relationship between the variables is accurately captured and the results are reliable.

3.2 Findings

Figure 4, describes the statistics of the variables used in this study. This table contains statistics for a sample of 67 individuals who participated in the survey and answered several questions focusing on various aspects of working from home [WFH] and their experience with it. The variables included in the table are id [identification number], age, gender [0=Female, 1=Male], years of experience, household [number of people living in their household], WFH% [percentage of time spent working from home], hybrid [whether they work in a hybrid model or not], collaboration problem [how often they face problems collaborating with colleagues while WFH], reduce promotion chance [how much WFH reduces their chances of promotion], impact of WFH on employer [how much WFH impacts their employer], feeling left out [how often they feel left out while WFH], feeling better with WFH [how much they feel better while WFH], and feeling more active with WFH [how much they feel more active while WFH]. The "count" row shows that there were 67 individuals surveyed for each variable. The "mean" row

shows the average value for each variable across the sample. The "std" row shows the standard deviation of the values for each variable, indicating how much variation there is within the sample. The "min" and "max" rows show the minimum and maximum values for each variable in the sample. Finally, the "25%", "50%", and "75%" rows show the quartiles for each variable, which can be used to determine the range and distribution of values for each variable. Based on the below figure, the average age of participants is 47.9, ranging from 28 to 64 with a standard deviation of 9.32.

```
In [79]: format_vertical_headers(df3.describe())
```

	id	age	gender	years_of_experience	household	WFH%	hybrid	collaboration_problem	reduce_promotion_chance	impact_of_wfh_on_employer	feeling_left_out	feeling_better_with_wfh	feeling_more_active_with_wfh
count	67.000000	67.000000	67.000000	67.000000	67.000000	67.000000	67.000000	67.000000	67.000000	67.000000	67.000000	67.000000	67.000000
mean	34.000000	47.885522	0.587015	3.597015	3.313433	0.588507	0.641791	4.000000	2.778119	3.880597	3.000000	3.805970	3.718418
std	19.485037	9.327597	0.494200	1.588493	1.144274	0.378391	0.483393	0.937437	1.185422	1.080437	0.904534	1.090438	1.058593
min	1.000000	28.000000	0.000000	0.500000	1.000000	0.000000	0.000000	2.000000	1.000000	1.000000	1.000000	1.000000	1.000000
25%	17.500000	42.000000	0.000000	2.500000	2.000000	0.200000	0.000000	4.000000	2.000000	3.000000	3.000000	3.000000	3.000000
50%	34.000000	47.000000	1.000000	5.000000	4.000000	0.800000	1.000000	4.000000	3.000000	4.000000	3.000000	4.000000	4.000000
75%	50.500000	56.000000	1.000000	5.000000	4.000000	1.000000	1.000000	5.000000	4.000000	5.000000	3.000000	5.000000	4.000000
max	67.000000	64.000000	1.000000	5.000000	6.000000	1.000000	1.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Figure 4 Description of the statistics of different variables

There are several useful and insightful diagrams that can be created based on the above table. Here are some examples:

Histogram of Age Distribution:

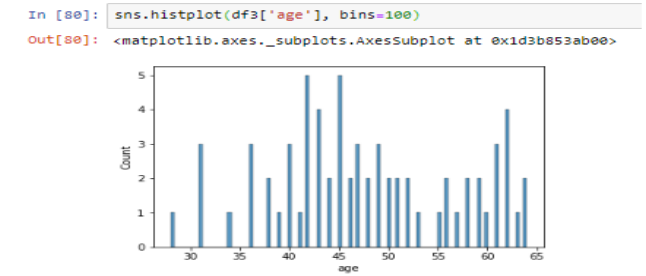


Figure 5 Histogram of Age Distribution

Figure 5 shows the distribution of ages of the participants. The x-axis represents age, and the y-axis represents the frequency of individuals at each age. This can help us understand the age range of the participants and identify any patterns or trends.

Histogram of Gender Distribution:


```
In [81]: sns.histplot(df3['gender'], bins=2)
Out[81]: <matplotlib.axes._subplots.AxesSubplot at 0x1d3b89fe278>
```

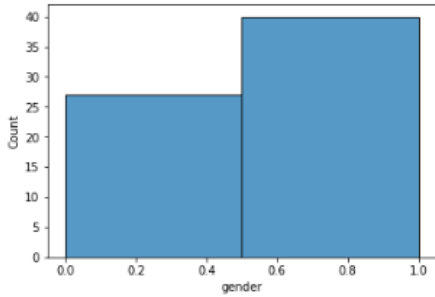


Figure 6 Histogram of Gender Distribution

Figure 6 shows the distribution of genders of the participants. The x-axis represents gender, and the y-axis represents the number of participants of each gender. This can help us understand the gender range of the participants and identify any patterns or trends.

Histogram of years of experience distribution:

```
In [46]: sns.histplot(df3['years_of_experience'], bins=100)
Out[46]: <matplotlib.axes._subplots.AxesSubplot at 0x1d9084e9828>
```

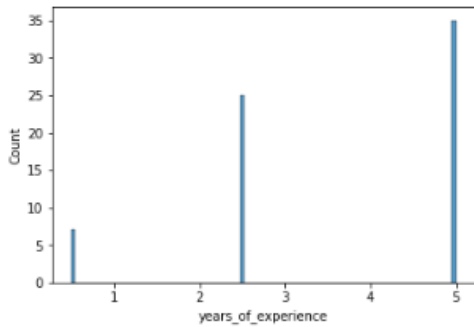


Figure 7 Histogram of years of experience distribution

Figure 7 shows the distribution of years of experience of the participants. The x-axis represents the number of years of experience, and the y-axis represents the number of participants. This can help us understand the experience range of the participants and identify any patterns or trends. Figure 8 shows the distribution of the number of households of the participants. The x-axis represents the number of households, and the y-axis represents the number of participants. This can help us understand the number of households of the participants.

```
sns.histplot(df3['household'], bins=5)
<matplotlib.axes._subplots.AxesSubplot at 0x26639e7a6a0>
```

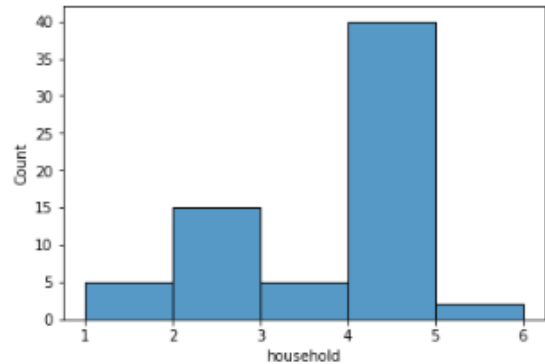


Figure 8 The histogram of the number of employees' households

Figure 9 shows that the correlations between the variables in the study are generally weak. As the figure shows, a correlation coefficient of 0.14 suggests a weak positive correlation between working from home and collaboration problems, meaning that as the frequency of working from home increases, the collaboration problem also slightly increases. Similarly, a correlation coefficient of -0.08 suggests a weak negative correlation between working from home and reducing the chance of promotion, meaning that as the frequency of working from home increases, the chance of promotion slightly decreases. Finally, a correlation coefficient of 0.11 suggests a weak positive correlation between the number of households and the happiness of the employees while working from home, meaning that as the number of households increases, the happiness of the employees while working from home also slightly increases. It is important to note that while these correlations are statistically significant; the strength of the correlations is weak, indicating that the relationship between the variables is not very strong.

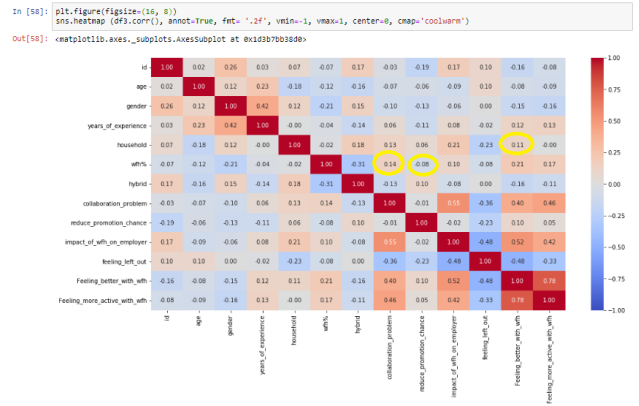


Figure 9 the correlations of the different variables

3.3 Hypotheses Analysis:

Figure 10 shows the p-values of the different variables.

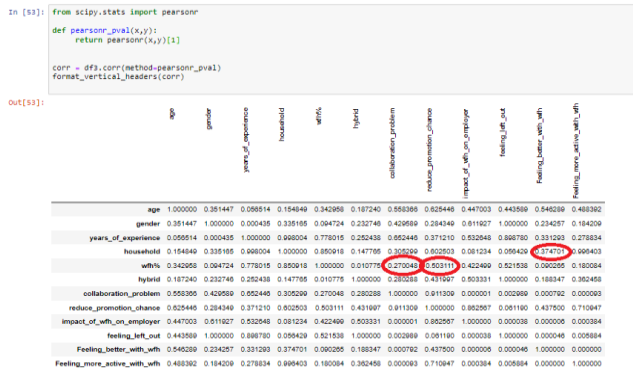


Figure 10 the p-values of the different variables

Hypothesis 1:

The null hypothesis [H0] states that there is no correlation between working from home and collaboration, while the alternative hypothesis [H1] states that there is a correlation. The p-value for this hypothesis is 0.27.

If the significance level [α] is set to 0.05, this means that we would reject the null hypothesis if the p-value is less than 0.05, and fail to reject it if the p-value is greater than or equal to 0.05.

Since the p-value for this hypothesis is greater than α [0.27 > 0.05], we fail to reject the null hypothesis. Therefore, we cannot conclude that there is a significant correlation between working from home and collaboration.

Hypothesis 2:

The null hypothesis [H0] states that there is no correlation between the number of households of employees and employees' happiness while WFH, while the alternative hypothesis [H1] states that there is a correlation. The p-value for this hypothesis is 0.374. Again, if we set α to 0.05, we would reject the null hypothesis if the p-value is less than 0.05, and fail to reject it if the p-value is greater than or equal to 0.05. Since the p-value for this hypothesis is greater than α [0.374 > 0.05], we fail to reject the null hypothesis. Therefore, we cannot conclude that there is a significant correlation between the number of households of employees and employees' happiness while WFH.

Hypothesis 3:

The null hypothesis [H0] states that there is no correlation between working from home and employees' promotion chance, while the alternative hypothesis [H1] states that there is a correlation. The p-value for this

hypothesis is 0.503. Once again, if we set α to 0.05, we would reject the null hypothesis if the p-value is less than 0.05, and fail to reject it if the p-value is greater than or equal to 0.05. Since the p-value for this hypothesis is greater than α [0.503 > 0.05], we fail to reject the null hypothesis. Therefore, we cannot conclude that there is a significant correlation between working from home and employees' promotion chances while WFH. In conclusion, all three hypotheses failed to provide evidence to reject their respective null hypotheses. Therefore, we cannot conclude that there is a significant correlation between working from home and collaboration, employees' happiness while WFH, and employees' promotion chances while WFH, respectively.

4. Discussion

4.1 Analysis of the results

The aim of this study was to find out 1- Is there any correlation between working from home and employees' collaboration?, 2- Is there any correlation between the number of households of employees and employees' happiness while working from home?, and 3- Is there any correlation between working from home and employees' promotion chances while working from home?. After analysing the data as explained above, the results and findings of this study to answer the above-mentioned questions are summarized as follows:

To answer question 1 "Is there any correlation between working from home and employees' collaboration?" the study analysed the Pearson correlation coefficient and P-Values between the variable "working from home" and the variable "collaboration problem". The correlation coefficient of 0.14 suggests a weak positive correlation between "working from home" and "collaboration problem", meaning that as the frequency of working from home increases, the collaboration problem also slightly increases. In addition, the p-value between "working from home" and "collaboration problem" is 0.27 and the p-value is greater than α [0.27 > 0.05]. Therefore, we cannot conclude that there is a significant correlation between working from home and collaboration.

To answer question 2 "Is there any correlation between the number of households of employees and employees' happiness while working from home?" the study analysed the Pearson correlation coefficient and P-Values between the variable "the number of households of employees" and the variable "employees' happiness while working from home". The correlation coefficient of 0.11 suggests a weak positive correlation between "the number of households of employees" and "employees' happiness while working from home", meaning that as the number of households of employees increases, the employees'

happiness while working from home also slightly increases. In addition, the p-value between “the number of households of employees” and “employees' happiness while working from home” is 0.374 and the p-value is greater than α [0.374 > 0.05]. Therefore, we cannot conclude that there is a significant correlation between the number of households of employees and employees' happiness while working from home.

To answer question 3 “Is there any correlation between working from home and employees' promotion chances while working from home?” the study analysed the Pearson correlation coefficient and P-Values between the variable “working from home” and the variable “reducing the chance of promotion”. The correlation coefficient of -0.08 suggests a weak negative correlation between “working from home” and “reducing the chance of promotion”, meaning that as the frequency of working from home increases, the reducing the chance of promotion slightly decreases. In addition, the p-value between “working from home” and “reducing the chance of promotion” is 0.503 and the p-value is greater than α [0.503 > 0.05]. Therefore, we cannot conclude that there is a significant correlation between working from home and reducing the chance of promotion.

4.2 Future research

The study encountered a number of limitations, such as a small participant pool and a homogeneous group of engineers from the same company and field, which could potentially introduce bias into the results. Future research in this area could focus on addressing the limitations identified in this study. One approach could be to conduct a larger-scale study that includes a more diverse sample of participants from different industries and domains to investigate whether the results are consistent across different populations. Additionally, future studies could also explore whether the findings of this study generalize to other types of professionals or individuals who have different levels of expertise in the same domain. Another direction for future research could be to investigate the potential sources of bias in this study and to explore methods to mitigate these biases. For example, researchers could use different sampling methods to recruit participants from different companies or industries, or use a multi-method approach to data collection, such as combining self-report measures with behavioural or physiological measures, to provide a more comprehensive assessment of the participants' attitudes or behaviours.

5. Conclusion

This study examines the viability of continuing remote work initiatives by assessing work satisfaction, performance, and productivity levels in the ICT industry in

Sweden. The COVID-19 pandemic has led to a significant increase in remote work, which has raised concerns about maintaining employees' job satisfaction and performance. The study aims to identify effective approaches and interventions to support remote workers determine the level of job satisfaction among the engineers and identify the pertinent performance indicators for their work. In this study, three questions were developed to examine 1- the relationship between working from home and collaboration problems, 2- the relationship between the number of households of employees and employees' happiness while working from home, and 3- the relation between working from home and employees' promotion chance, which will help organizations develop effective remote work policies and practices that balance productivity and employee well-being. Regarding question 1, the study finds a weak positive correlation [correlation coefficient of 0.14] between working from home and collaboration problems. However, the p-value is greater than α [0.27 > 0.05], indicating that there is no significant correlation between these variables. Therefore, the study cannot conclude that there is a significant correlation between working from home and collaboration problems. Regarding question 2, the study finds a weak positive correlation [correlation coefficient of 0.11] between the number of households of employees and employees' happiness while working from home. However, the p-value is greater than α [0.374 > 0.05], indicating that there is no significant correlation between these variables. Therefore, the study cannot conclude that there is a significant correlation between the number of households of employees and employees' happiness while working from home. Regarding question 3, the study finds a weak negative correlation [correlation coefficient of -0.08] between them, meaning that as the frequency of working from home increases, the reducing the chance of promotion slightly decreases. However, the p-value is greater than α [0.503 > 0.05], which means there is no significant correlation between them. Therefore, no conclusive evidence suggests a relationship between working from home and reducing the chance of promotion.

In overall, based on the correlation coefficients and p-values between variables in the study, we can conclude there is no strong relationship between variables in this study.

References

- [1] Raišienė AG, Rapuano V, Varkulevičiūtė K, Stachová K. Working from home—Who is happy? A survey of Lithuania's employees during the COVID-19 quarantine period. Sustainability [Switzerland]. 2020;12[13]. doi:10.3390/su12135332.

- [2] Vyas L, Butakhieo N. The impact of working from home during COVID-19 on work and life domains: An exploratory study on Hong Kong. *Policy Design and Practice*. 2020;4[1]:1–18. doi:10.1080/25741292.2020.1863.
- [3] Oakman J, Kinsman N, Stuckey R, Graham M, Weale V. A rapid review of mental and physical health effects of working at home: How do we optimise health? *BMC Public Health*. 2020;20[1]:1–13. doi:10.1186/s12889-020-09875-z.
- [4] International Labour Organization. *An Employers' Guide on Working from Home in Response to the Outbreak of COVID-19*. International Labour Organization; 2020. Available from: https://www.ilo.org/actemp/publications/WCMS_745024/lang--en/index.htm.
- [5] Hashim R, Bakar A, Noh I, Mahyudin HA. Employees' Job Satisfaction and Performance through working from Home during the Pandemic Lockdown. *Environment-Behaviour Proceedings Journal*. 2020;5[15]:461–467. doi:10.21834/ebpj.v5i15.2515.
- [6] Jamal MT, Anwar I, Khan NA, Saleem I. Work during COVID-19: Assessing the influence of job demands and resources on practical and psychological outcomes for employees. *Asia-Pacific Journal of Business Administration*. 2021. doi:10.1108/APJBA-05-2020-0149.
- [7] Hering B. Remote work statistics: Shifting norms and expectations. *Flexjobs*. 2020. Available from: <https://www.flexjobs.com/blog/post/remote-work-statistics/>.
- [8] Dimitrova D. Controlling teleworkers: Supervision and flexibility revisited. *New Technology, Work and Employment*. 2003;18[3]:181–195. doi:10.1111/1468-005X.00120.
- [9] Golden TD, Veiga JF, Dino RN. The impact of professional isolation on teleworker job performance and turnover intentions: Does time spent teleworking, interacting face-to-face, or having access to communication-enhancing technology matter? *Journal of Applied Psychology*. 2008;93[6]:1412–1421. doi:10.1037/a0012722.
- [10] Grant CA, Wallace LM, Spurgeon PC. An exploration of the psychological factors affecting remote e-worker's job effectiveness, well-being and work-life balance. *Employee Relations*. 2013;35[5]:527–546. doi:10.1108/ER-08-2012-0059.
- [11] Timsal A, Awais M. Flexibility or ethical dilemma: An overview of the work from home policies in modern organizations around the world. *Human Resource Management International Digest*. 2016;24[7]:12–15. doi:10.1108/HRMID-03-2016-0027.
- [12] Anderson D, Kelliher C. Enforced remote working and the work-life interface during lockdown. *Gender in Management*. 2020;35[7]:677–683. doi:10.1108/GM-07-2020-0224.
- [13] Baker E, Avery GC, Crawford J. Satisfaction and perceived productivity when professionals work from home. *Research & Practice in Human Resource Management*. 2007.
- [14] Sahni DJ. Impact of COVID-19 on Employee Behavior: Stress and Coping Mechanism During WFH [Work From Home] Among Service Industry Employees. *International Journal of Operations Management*. 2020;1[1]:35–48. doi:10.18775//ijom.2757-0509.2020.11.4004.
- [15] Bernarto I, Bachtiar D, Sudibjo N, Suryawan IN, Purwanto A, Asbari M. Effect of transformational leadership, perceived organizational support, job satisfaction toward life satisfaction: Evidences from Indonesian teachers. *International Journal of Advanced Science and Technology*. 2020;29[3]:5495–5503.
- [16] Kwon M, Jeon SH. Do Leadership Commitment and Performance-Oriented Culture Matter for Federal Teleworker Satisfaction With Telework Programs? *Review of Public Personnel Administration*. 2020;40[1]:36-55. doi:10.1177/0734371x18776049.
- [17] Sridhar V, Bhattacharya S. Significant household factors that influence an IT employees' job effectiveness while on work from home. *International journal of innovation science*. 2021;13[1]:105-117.
- [18] Van Der Lippe T, Lippényi Z. Co-workers working from home and individual and team performance. *New technology, work and employment*. 2020;35[1]:60-79.
- [19] Song Y, Gao J. Does telework stress employees out? A study on working at home and subjective well-being for wage/salary workers. *Journal of Happiness Studies*. 2020;21[7]:2649-2668.
- [20] Denscombe M. *The good research guide: for small-scale research projects*. Maidenhead, Berkshire: McGraw-Hill Education; 2014.
- [21] Babbie E. *The practice of social research*. Cengage Learning. 2010;15.
- [22] Bryman A. *Social research methods*. Oxford university press; 2016.
- [23] Creswell JW. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications; 2013.
- [24] Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences*. 2011;4[1]:8–11. doi:10.4103/0974-1208.82352.
- [25] Neuman WL. *Social research methods: Qualitative and quantitative approaches*. Pearson Education; 2013.

- [26] Cerda P, Varoquaux G, Kégl B. Similarity encoding for learning with dirty categorical variables. *Machine Learning*. 2018;107[8-10]:1477-1494.
- [27] Sahoo K, Samal AK, Pramanik J, Pani SK. Exploratory data analysis using Python. *International Journal of Innovative Technology and Exploring Engineering [IJITEE]*. 2019;8[12].
- [28] Al Jassmi H, Ahmed S, Philip B, Al Mughairbi F, Al Ahmad M. E-happiness physiological indicators of construction workers' productivity: A machine learning approach. *Journal of Asian Architecture and Building Engineering*. 2019;18[6]:517-526.
- [29] Field A. *Discovering statistics using IBM SPSS Statistics*. 4th ed. Sage Publications; 2013.
- [30] Sullivan GM, Artino AR Jr. Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education*. 2013;5[4]:541-542. doi:10.4300/JGME-D-13-00154.1.
- [31] American Psychological Association. *Ethical principles of psychologists and code of conduct*. 2017. Available from: <https://www.apa.org/ethics/code/>.
- [32] Fisher CB, Fried AL, Goodman SJ. Ethical issues in dissemination and implementation research. In: Brownson RC, Colditz GA, Proctor EK, editors. *Dissemination and implementation research in health: Translating science to practice*. 2nd ed. Oxford University Press; 2017. p. 571–588.
- [33] Imani M, Arabnia HR. Hyperparameter Optimization and Combined Data Sampling Techniques in Machine Learning for Customer Churn Prediction: A Comparative Analysis. *Technologies*. 2023;11[6]:167.

Appendix A: Informed Consent Form

Informed Consent Form

Title of Study: The impacts of working from home on employees' collaboration, happiness, and promotion chances.

Investigator: Mehdi Imani

Introduction:

You are invited to participate in a survey that is being conducted by Mehdi Imani to investigate the correlation between working from home and employee's collaboration, happiness, and promotion chances.

Purpose:

The purpose of this study is to investigate the correlation between working from home and employee's collaboration, happiness, and promotion chances. Participation in this study is voluntary. You may choose to participate or not. If you choose to participate, you may withdraw at any time during the survey.

Procedure:

If you agree to participate, you will be asked to complete an online survey. The survey will take approximately five minutes to complete. The survey consists of 12. Your responses will be confidential and anonymous. No identifying information will be collected during the survey.

Confidentiality:

Your survey responses will be kept confidential and anonymous. Your identity will not be revealed in any reports or publications resulting from this study. All data will be kept secure and only accessible to the research team.

Voluntary Participation:

Your participation in this study is voluntary. If you choose to participate, you may withdraw at any time during the survey. You are free to refuse to answer any questions that you do not wish to answer without penalty or negative consequences.

Contact Information:

If you have any questions or concerns about this survey, please contact the Principal Investigator, Mehdi Imani, m.imani@gmail.com.

Statement of Consent:

By clicking the "I agree" button below, you acknowledge that you have read and understood this Informed Consent Form and agree to participate in the survey.

Appendix B: Questionnaire

1- What year were you born? [Please enter year of birth]

2- What is your gender?

Male

Female

Other

3- How long have you been working in your current job?

Less than 1 year

1-3 years

4-6 years

7-10 years

More than 10 years

4- Which of the following descriptions best fits your household?

Single

Single parent with 1 children

Single parent with 2 children

Single parent with 3+ children

Married/Partnered with no children

Married/Partnered with 1 children

Married/Partnered with 2 children

Married/Partnered with 3+ children

5- How many days have you worked remotely this year? Please provide a number between 1 and 5.

6- Have you ever worked remotely and on-site on the same day at your employer's workplace?

Yes

No

7- To what extent do you agree or disagree with the statement: "It is effortless for me to work together with my colleagues while working remotely."

Strongly agree [1]

Agree [2]

Neither agree nor disagree [3]

Disagree [4]

Strongly disagree [5]

8- To what extent do you agree or disagree with the statement: "Remote work decreases my prospects of getting promoted."

Strongly agree [1]

Agree [2]

Neither agree nor disagree [3]

Disagree [4]

Strongly disagree [5]

9- Do you believe remote working is a positive or negative aspect for your employer?

Strongly Positive [1]

Positive [2]

Neither Positive nor Negative [3]

Negative [4]

Strongly Negative [5]

10- Have the following barriers to remote work improved or worsened for you: feeling left out and/or isolated?

Improved

Stayed the same

Worsened

11- To what extent do you agree or disagree with the statement: "I feel better on the days I work remotely."

Strongly agree [1]

Agree [2]

Neither agree nor disagree [3]

Disagree [4]

Strongly disagree [5]

12- To what extent do you agree or disagree with the statement: "I am more active on the days I work remotely."

Strongly agree [1]

Agree [2]

Neither agree nor disagree [3]

Disagree [4]

Strongly disagree [5]