MSC. DISSERTATION IN DATA SCIENCE: IMPACT OF REMOTE WORK ON MENTAL HEALTH

Robert Solomon

MSc. Dissertation In Data Science (Impact of Remote Work on Mental Health)

Table of Contents

MSc. Dissertation In Data Science (Impact of Remote Work on Mental Health)

Table of Figures

Table of Figures

Introduction

The emerging trend for remote working has reshaped workspaces and has brought benefits, along with problems. One of the central concerns for remote workers is the psychological impact, with particular reference to feelings of isolation and decreased work-life balance. The abrupt transition towards remote working in the context of the COVID-19 pandemic has drawn a great deal of interest among scholars; yet there has been little research on investigating how these factors unfold to influence mental health under long-term remote working conditions.

Research Questions

- What is the relationship between work location (remote, hybrid, onsite) and mental health outcomes such as stress and social isolation?
- How does access to mental health resources impact productivity and satisfaction levels in remote workers?
- Are certain job roles or industries more susceptible to mental health challenges in remote settings?
- Which factors (e.g., hours worked, work-life balance, physical activity) most strongly correlate with stress levels in remote workers?

Objectives

This research aims to investigate how different aspects of remote work, such as work-life balance, hours worked, and virtual interactions, impact mental health outcomes, including stress levels, mental health conditions (like depression and anxiety), and social isolation. Understanding these relationships can guide employers and policymakers in creating healthier work environments.

- Identify key factors influencing mental health among remote workers.
- Assess the role of company support and access to mental health resources on employee productivity.
- Investigate demographic and job-specific differences in mental health responses to remote work.

Literature Review

Introduction

This expanded literature review delves deeper into existing research to contextualize the benefits, challenges, and underexplored areas of remote work's impact on mental health.

The rise of remote work, particularly in the post-pandemic era has transformed the modern workforce, bringing both opportunities and challenges. While it provides flexibility and eliminates commuting time, there are also other psychological effects related to isolation, burnout, and work-life balance that remote work has been linked to, making understanding the same an essentiality.

This review synthesizes current findings, identifies gaps in the literature, and explores theoretical frameworks relevant to the psychological effects, contributing factors, and potential interventions in understanding the interplay between remote work and mental health.

Psychological Effects of Remote Work

Isolation and Loneliness

Multiple studies highlight the increased risk of social isolation among remote working colleagues and the workplace. According to (Becker, 2022), employees working remotely report feeling disconnected from colleagues, leading to loneliness and reduced job satisfaction and emotional exhaustion, minor counterproductive work behaviours, among other negative effects. Remote employees often miss the casual social interactions found in office environments, which contribute to a sense of belonging and professional advancement opportunities.

Burnout and Work-Life Balance

Burnout is the riskiest factor for remote workers. The blurring of boundaries between professional and personal life increases stress levels, with workers often struggling to "switch off" after working hours. Without clear separation between workspaces and personal spaces, many remote workers report working longer hours than they would in a traditional office setting. For example, a systematic review by (Shaholli, Manai, Iantorno, Di Giampaolo, & Nieto, 2024) showed that the bridging of personal and working life because of teleworking can result in increased stress and burnout.

Productivity and Job Satisfaction

While some studies suggest that working from home can enrich productivity, others stress the challenges regarding motivation and keeping workers engaged. The level of managerial support is an important factor in how the variation in the impact on job

satisfaction and mental well-being is moderated. A systematic review by (Guidarini, 2023) noted that telecommuting is associated with higher levels of job satisfaction, but the relationship is moderated by factors such as autonomy and support.

Contributing Factors to Mental Health Challenges in Remote Work

Digital Communication Overload

Continuous exposure to digital communication tools can result in cognitive overload and stress. The "always-on" culture associated with remote work is adding to the rising expectations of immediate responses, making it hard for employees to detach themselves from work-related activities. According to a study by (Hall, 2023), excessive use of communication technology is related to increased stress and reduced well-being.

Home Environment and Workspace Design

The home environment is a determinant of working effectiveness from home. Individuals who have home offices tend to have
less stress than others whose workspace is shared or setup in non-ergonomic ways. Poor workstations lead to physical strains and mental exhaustion. A study by (Felstead & Reuschke, 2020) indicated that the quality of the home workspace
is one of the strongest predictors of job satisfaction for remote workers.

Managerial and Organizational Support

Management, therefore, plays a vital role in mitigating mental health challenges. Conversely, frequent virtual check-ins, mental health resources, and clear performance expectations within an organization lead to reduced stress among remote workers. The supportive leadership and policies, such as flexible scheduling and mental health days, will go a long way in contributing to the overall well-being of individuals. A cross-referenced study by (Philips, 2020) established that organizational support is integral in mitigating the negative impacts of telework on mental health.

Potential Interventions and Solutions

Technology-Based Wellness Solutions

Recent studies have shown that digital mental health platforms support remote workers in maintaining good mental health. The integration of technology into corporate wellness programs has so far helped reduce feelings of loneliness. A 2025 study by (Carraro Elisabetta, 2025) indicated that ***"social networks play a crucial role in promoting mental health, suggesting that strong and meaningful relationships can serve as a buffer against anxiety and depression"*** which would include remote delivery of psychotherapy, including telephone, video, and online modalities, that is found to be just as effective as effective as face-to-face therapy in treating anxiety and

depression. These methods offer accessibility and convenience, making them suitable alternatives for those unable to access in-person care.

Work-Life Balance Strategies

Encouragement of structured work schedules and "right to disconnect" policies helps alleviate burnout. Countries like France and Ireland have created legal requirements ensuring that workers are not compelled to participate in work communications during periods outside of work. Therefore, such initiatives will go a long way toward assuring better mental health outcomes for remote employees. A report by (Carvalho VS, 2021) suggests that by integrating an overall inter-role valuation of congruity between work and family domains contributes to reducing burnout and increasing flourishing.

Hybrid Work Models

Various studies also praise hybrid models of work as a balance between flexibility and face-to-face interactions. Employees who split their time between working from home and in the office tend to be more satisfied with their jobs and have less stress compared to those who work fully at home. Hybrid models allow them to collaborate with others in person and maintain some of the autonomy of working remotely. A study by (Ashish Sarangi MD, 2022) establish that telecommuting is related positively to job satisfaction, especially when combined with periodic work at the office.

Methodology

Secondary Research Data & Design

Data Source

This secondary research sought to analyze the relationship between mental health and telecommuting by using quantitative secondary data analysis. An available public dataset scraped from **Kaggle** titled ***"Remote_Work_on_Mental_Health.csv"*** was utilized. It contains extensive information regarding employees from various sectors and regions, such as their job location, balance between personal and professional life, stress, loneliness, and availability of mental health resources. Using previously collected data sets for this study makes sense because it is possible to identify trends on a larger scale without the complications that come with collecting primary data. As the dataset is organized, it enables comparison with primary survey data that was gathered first-hand, hence cross checking and expanding the meaning of the results is easier.

Secondary Data Collection

The secondary dataset consists of 5000 entries of from employees regarding telework and mental health outcomes. It includes demographic data (e.g., gender, age, industry, job title, and years of experience), work variables (e.g., work arrangement, number of hours worked weekly, and number of virtual meetings), and psychological variables (e.g., level of stress, productivity change, social isolation rating, and remote work satisfaction). The data collection spans a diversity of industries, such as healthcare, IT, finance, education, and consultancy, with workers' feedback selected from a variety of different countries such as North America, Europe, and Asia. The information was originally compiled for an autonomous study that had quantified the impact of home working on wellness and work effectiveness, and so it is extremely pertinent to this dissertation's research objectives.

Since the dataset was publicly available, ethical guidelines were taken into account to fulfil research needs. The data was anonymized, with no identifiable information except for general demographic categories. The application of different employees from diverse professional backgrounds also increases the generalizability of the findings.

Data Preprocessing

Before performing any statistical analysis, rigorous preprocessing and cleaning of data were carried out in order to arrange the dataset properly for analysis. This included various important steps:

MSc. Dissertation In Data Science (Impact of Remote Work on Mental Health)

Handling Missing Values:

Missing values were found in several variables, including Mental Health Condition, Company Remote Work Support, and Physical Activity. In order to preserve data integrity, categorical missing values were filled with mode, while numerical missing values were filled in with the median to prevent the data from becoming biased. Entries with an excessive number of missing values (i.e., missing more than three critical attributes) were dropped in order to make the dataset more reliable.

Encoding Categorical Variables:

For convenience of statistical analysis, **categorical** variables were translated into numbers:

- *Binary Encoding:*
    - Mental Health Resources Access was encoded as 1 (Yes) and 0 (No).
    - Productivity Change was assigned -1 (Decrease), 0 (No Change), and 1 (Increase) to maintain ordinal relationships.

- *Ordinal Encoding:*
    - Work Location was encoded as 1 (Remote), 2 (Hybrid), and 3 (Onsite).
    - Stress Level was recoded to 1 (Low), 2 (Medium), and 3 (High).
    - Job Satisfaction with Remote Work was measured on a scale of 1 (Unsatisfied), 2 (Neutral), and 3 (Satisfied).

- *One-Hot Encoding:*
    - Variables such as Job Role, Industry, Region, and Mental Health Condition were re-coded to dummy variables in order to facilitate independent categorical contrasts.

To further validate the dataset after encoding categorical variables, there was a minor verification process carried out. The encoded dataset was temporarily exported using a simple script:

wfh_mentalHealth_data.to_csv("./Secondary_Research/SR_Dataset/post_Encoded_Remote_Work_on_Mental_Health.csv", index=False)

This gave room for a manual verification of the encoded file to confirm all transformations had been properly applied before moving on to data standardization and outlier handling. This validation process ensured data integrity and consistency throughout the preprocessing phase.

MSc. Dissertation In Data Science (Impact of Remote Work on Mental Health)

Standardizing Numerical Data:

For consistency in numeric variables, all the work-life balance scores, stress ratings, and social isolation measurements were formatted accordingly. Continuous variables such as work hours per week and virtual meetings held were not changed from their original numerical format to ensure employee work conditions accuracy.

Identification of Outliers and Handling Them:

Outliers for age, years of experience, working hours per week, work/life balance rating, stress levels and social isolation ratings were examined with the use of the Interquartile Range (IQR) method. Variables such as *'Age'*, *'Hours_of_Experience'* and *'Hours_Worked_Per_Week'* were found to have outliers beyond 2.5 times the IQR. It was decided that these outliers would be removed as it caused a significant amount of data loss.
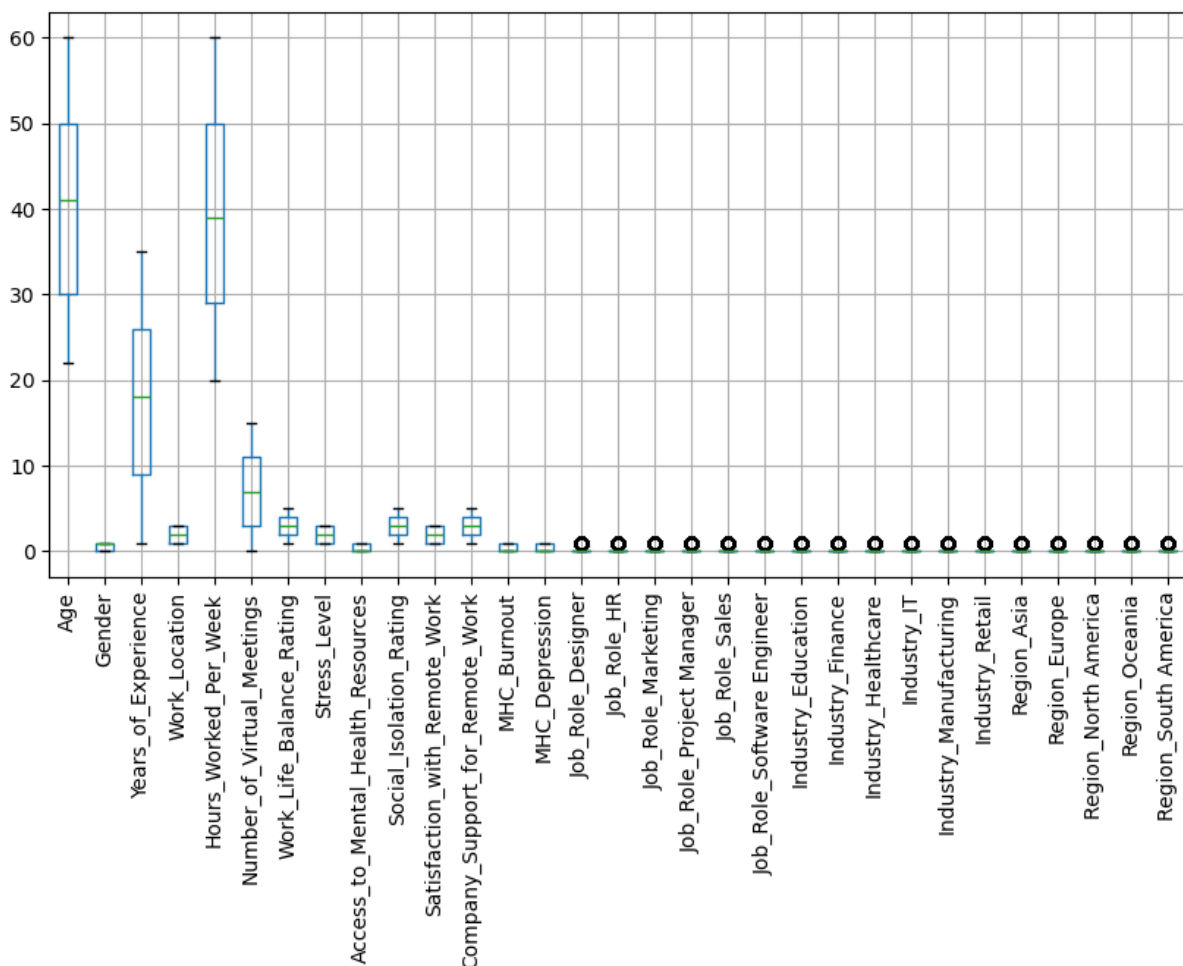

Figure 1 – Boxplot showing outliers present in the secondary Dataset

Feature Engineering

After the data cleaning, feature engineering was performed on the dataset to create new variables that would strengthen the analysis. Some of the most significant engineered

features included a composite Work Stress Score, Hours worked per week, number of virtual meetings, work life balance rating, each of which was generated by logically combining related variables. These new features facilitated stronger correlation analysis, regression modelling, and clustering, with deeper insight into the effect of remote work on mental health.

Saving Pre-processed Secondary Dataset to New CSV File:

Once the cleaning process was complete, the final version of the dataset was saved separately as the *"cleaned_Remote_Work_on_Mental_Healt.csv"* cleaned secondary dataset, ready for further analysis.

Analytical Methods Carried Out for Secondary Dataset

Upon the preprocessing of the data, machine learning algorithms and statistical techniques were utilized to evaluate the relationship of work location to stress levels, work-life balance, and mental health outcomes within the secondary dataset. The method employed was selected in an attempt to gain complete understanding of the data and ensure that trends displayed may also occur within the primary dataset.

To begin with, descriptive statistics were estimated to generate summaries of significant variables such as work-life balance scores, stress levels, productivity variations, and working hours. Statistics such as mean, median, standard deviation, and interquartile range (IQR) were predicted in order to examine data distribution and identify potential outliers. Visualization techniques such as histograms and box plots were also employed to seek patterns in employee well-being across job roles and work arrangements.

Correlation Analysis

This was then followed by correlation analysis to determine inter-relations between key work-related variables. Pearson's correlation coefficient was used to test linear relations between quantitative variables, for example, the correlation between number of working hours per week and stress. In addition, Spearman's correlation was used for ordinal variables, for example, the impact of work-life balance ratings on employee satisfaction with telework.

MSc. Dissertation In Data Science (Impact of Remote Work on Mental Health)

One-Way ANOVA and Post-Hoc Test (Tukey's HSD)

One-way Analysis of Variance (ANOVA) was used to determine whether the work location (remote, hybrid, or onsite) was a strong predictor of the level of stress reported and whether the mean stress levels differed significantly across the various work arrangements. If the ANOVA test was statistically significant, a Tukey's Honest Significant Difference (HSD) post-hoc test was used to identify the specific groups that differed from one another. This helped with the overall analyses conducted on the secondary dataset by providing a better understanding of how different work environments could be linked to variations in employee stress.

Regression Modelling (Linear)

Furthermore, multiple linear regression was used to predict stress levels and work-life balance outcomes from independent variables such as working hours per week, number of virtual meetings, access to employer-provided mental health resources, and job title. The regression models provided insight into the workplace factors that most impact the well-being of employees.

MSc. Dissertation In Data Science (Impact of Remote Work on Mental Health)

Primary Research Data & Design

Survey Design

In order to support and complement the analysis of secondary data, a primary research study was conducted to gather firsthand data regarding the impact of remote work on employees' well-being, work-life balance, and mental health. A quantitative cross-sectional survey design was employed to enable structured responses from a range of participants at one time. This architecture facilitates the research objectives in relation to conducting statistical analysis of relations and trends among top variables in the secondary data set.

The principal research not only aimed to validate patterns identified in the secondary data but also to explore other dimensions of remote work experiences that may not have been evidenced in the existing dataset.

Primary Data Collection

The data collection process involved developing and distributing an **online survey** designed via **Google Forms** to capture self-reported data on stress, work-life balance, and isolation among remote workers. The survey consisted of **12 questions** including a **Likert scale** to assess subjective experiences (work-life balance, stress levels). The survey was shared through my professional networks on **LinkedIn, Remote-working WhatsApp groups, internal communication channels within my working organisations** and open for **3 weeks**. **46 responses** were collected which is not the amount I was hoping to gather.

Data Cleaning & Preprocessing

Similarly to the Secondary Dataset, before performing any statistical analysis on the secondary dataset, rigorous preprocessing and cleaning of data were also carried out in order to ensure the dataset was properly organised for analysis. This included various important steps:

Handling Missing Data/Values:

Before continuing to the analysis, the primary dataset was well cleaned and prepared to be in the proper form for statistical testing. In doing so, missing data were found, most prominently in fields such as "stress factors" and "mental health recommendations," primarily due to partially filled-in survey returns. Where missing data were small such as for Stress levels, imputation was employed to replace these missing data points however, similarly where records had large gaps, imputations were also employed to fill those datapoints due to the dataset not containing many records and avoiding loss of insights.

Encoding Categorical Variables:

For **categorical** variables such as gender, label encoding was used and converted to numerical values for simplicity. For other **categorical** variables like Industry, and Region, one-hot encoding was used.

For convenience of statistical analysis, categorical variables were translated into numbers:

- *Label Encoding:*
    - Label encoding was applied to categorical variable gender and encoded as 1 (Female) and 0 (Male).

- *One-Hot Encoding:*
    - **Categorical** variables like Industry, and Region, one-hot encoding was used.

- *Ordinal Encoding:*
    - **Likert-type** responses such as for age group, gender, work location, job role/industry and stress level were also treated as ordinal data, without assuming equal spacing between ratings while preserving the natural rating order.
    - Ordinal variables for **ordinal categories** such as work location, social isolation frequency and employer mental health support columns or 'yes/no' responses such as for lack of team connection, were converted to numerical values (1 for Yes and 0 for No) for ease of handling them statistically.

Standardizing Numerical Data:

For consistency in numeric variables, all the work-life balance scores, stress ratings, and social isolation measurements were formatted accordingly. Continuous variables such as work hours per week and virtual meetings held were not changed from their original numerical format to ensure employee work conditions accuracy.

Saving Pre-processed Primary Dataset to New CSV File:

Similarly, the cleaning process was complete, the final version of the dataset was saved separately as the ***"cleaned_WFH-Mental_Health_(Survey).csv"*** cleaned primary dataset, ready for further analysis.

Analytical Methods Carried Out for Primary Dataset

Analysis for the primary dataset was carried out after pre-processing and cleaning of the main dataset. Its initial phase was to reveal associations between key variables using inferential statistical methods. Rather than calculating common descriptive

summaries of data such as means, medians, or standard deviations for the variables, focus was laid on addressing main research questions.

One-Way ANOVA

In order to determine if there were statistically significant differences in stress levels among various work arrangements (remote, hybrid, and onsite), One-way ANOVA was applied. This test answered the overarching question posed in the main study of differences in remote work and mental health.

Correlation Analysis

This was then followed by correlation analysis to determine inter-relations between key work-related variables. Pearson's correlation coefficient was used to test linear relations between quantitative variables, for example, the correlation between number of working hours per week and stress. In addition, Spearman's correlation was used for ordinal variables, for example, the impact of work-life balance ratings on employee satisfaction with telework.

Acronym

<Will include any acronyms used throughout the research here>

References

Akbar K. Waljee, M. M. (2013, December 26). A Primer on Predictive Models. *Clinical and Translational Gastroenterology*, 4. Retrieved November 2024, from pmc.ncbi.nlm.nih.gov: https://pmc.ncbi.nlm.nih.gov/articles/PMC3912317/#:~:text=Prediction%20rese arch%2C%20which%20aims%20to%20predict%20future%20events,an%20illne ss%20or%20the%20risk%20of%20developing%20an

Ashish Sarangi MD, D. K. (2022, September 9). *The mental health impact of work from home: A literature review.* doi:10.12746/swrccc.v10i45.1085

Becker, W. B. (2022). Surviving remotely: How job control and loneliness during a forced shift to remote work impacted employee work behaviors and well-being. *Human Resource Management*. Retrieved from https://doi.org/10.1002/hrm.22102

Carraro Elisabetta, R. P. (2025, February 06). Remote workers' life quality and stress during COVID-19: a systematic review. *European Journal of Public Health*. doi:https://doi.org/10.1093/eurpub/ckae167

Carvalho VS, S. A. (2021, June 30). Please, Do Not Interrupt Me: Work–Family Balance and Segmentation Behavior as Mediators of Boundary Violations and Teleworkers' Burnout and Flourishing. (M. S. Pérez, & E. Cifre, Eds.) *Sustainability*, 5. doi:https://doi.org/10.3390/su13137339

Charalampous, M. G. (2018, November 01). Systematically reviewing remote e-workers' well-being at work: a multidimensional approach. *European Journal of Work and Organizational Psychology*, 51 - 73. Retrieved from https://doi.org/10.1080/1359432X.2018.1541886

Efimov, I., Rohwer, E., Harth, V., & Mache, S. (2022). *Virtual leadership in relation to employees' mental health, job satisfaction and perceptions of isolation: A scoping review.* Sec. Health Psychology. Retrieved October 12, 2024, from https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.96 0955/full

Felstead, A., & Reuschke, D. (2020). HOMEWORKING IN THE UK: BEFORE AND DURING THE 2020. *WISERD Report*, 17. Retrieved January 2025, from https://eprints.soton.ac.uk/444076/1/Homeworking_in_the_UK_Report_Final_3_ 3.pdf

Ferrara, B., Pansini , M., Vincenzi, C. D., Buonomo, I., & Benevene, P. (2022, September 28). Investigating the Role of Remote Working on Employees' Performance and Well-Being: An Evidence-Based Systematic Review. *International Journal of Environmental Research and Public Health*, 12.

MSc. Dissertation In Data Science (Impact of Remote Work on Mental Health)

Guidarini, C. H. (2023, December). A Systematic Review of How Remote Work Affects Workplace Stress and Mental Health. In C. H. Guidarini, *Human-Automation Interaction* (Vol. 12, pp. 79-96). Springer, Cham. doi:https://doi.org/10.1007/978-3-031-10788-7_5

Hall, C. D. (2023). The relationship between homeworking during COVID-19 and both, mental health, and productivity: a systematic review. *BMC Psychology*, 2, 8. doi:https://doi.org/10.1186/s40359-023-01221-3

Makowski, P. (. (2023, July 18). Remote Leadership and Work Engagement: A Critical Review and Future Directions. . *European Journal of Business and Management Research.*, 7. Retrieved November 2024, from https://www.ejbmr.org/index.php/ejbmr/article/view/1835

Nia Sarinastiti, A. B. (2022). RELATIONS OF REMOTE WORKING TO MENTAL HEALTH. . *ASPIRATION Journal*. Retrieved from https://doi.org/10.56353/aspiration.v2i2.40.

Pandey, D. S. (2020, April 17). Principles of Correlation and Regression Analysis. *Journal of the Principles of Correlation and Regression Analysis*, 5. Retrieved November 2024, from www.j-pcs.org

Philips, S. (2020). Working through the pandemic: Accelerating the transition to remote working. Business Information Review. *SAGE Journals*, 129–134. doi:https://doi.org/10.1177/0266382120953087

Shaholli, D., Manai, M., Iantorno, F., Di Giampaolo, L., & Nieto. (2024). Teleworking and Mental Well-Being: A Systematic Review on Health Effects and Preventive Measures. Retrieved January 2025, from https://doi.org/10.3390/su16188278

Sharma, R. (2024, May 14). *What is Clustering in Machine Learning and Different Types of Clustering Methods*. Retrieved November 2024, from Upgrad.com: https://www.upgrad.com/blog/clustering-and-types-of-clustering-methods/

Yokoi, K., Shimura, A., Ishibashi, Y., Akatsuka, Y., & Inoue, T. (2021, September 30). Remote Work Decreases Psychological and Physical Stress Responses, but Full-Remote Work Increases Presenteeism. (U. o. Merce Mach, Ed.) Retrieved October 2024, from Frontiers in Psychology: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.730969/full