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

Robert Solomon

2024/25

DATA SCIENCE INSTITUTE

Supervisor: Dr. Vinayak Deshpande

Acknowledgements

Abstract

One of the single biggest changes to the modern workplace has been the shift to remote working, but especially so in the light of the COVID-19 pandemic. While offering the benefit of introducing new levels of freedom and flexibility to workers, it has also introduced challenges in the way that it has been found to influence work-life balance, pressure, and mental health. The focus of this research is to explore these effects by making use of second-order data and original survey data to examine how various variables, ranging from workplace location through to employer support, access to mental health support, and working hours are related to degrees of stress and overall satisfaction within homeworkers.

Quantitative methods were used to perform different statistical tests like ANOVA, correlation analysis, and regression modelling. Secondary data from a public data repository provided a broad overview of trends by industry, while the primary data elicited first-hand input from practicing professionals through a formal survey.

The results show that the different work modalities, remote, onsite, or hybrid are not significantly related to levels of stress. On the contrary, there is a high correlation between work-life balance, feelings of isolation, and mental health metrics. Of particular interest, access to mental health services was a key finding in the second-level analysis, whereas it was of lesser importance in the first-level analysis, possibly due to inequalities in availability or subjective experience.

In summary, the mental health of remote workers is not directly influenced by their immediate work environment; instead, it depends largely on how much support they are getting, the nature of the work they perform, and how well they are able to balance work and personal responsibilities. These findings can be used to inform organizational policies and promote debate on maintaining effective and healthy environments within telecommuting systems.

Table of Contents

Table of Figures

Acronyms

WFH – Work from Home

ANOVA – Analysis of Variance Statistical Analysis

OLS – Ordinary Least Squares

HSD – Honest Significant Difference Turkey's Post-hoc Test

SD – Standard Deviation

IQR – Interquartile range

EU – European Union

CSV – Comma-Separated Values

RFC – Random Forest Classifier

DTC – Decision Tree Classifier

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 biggest risk 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** and titled ***Remote Work on Mental Health*** 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:

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.

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 and Response of Outliers:

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.
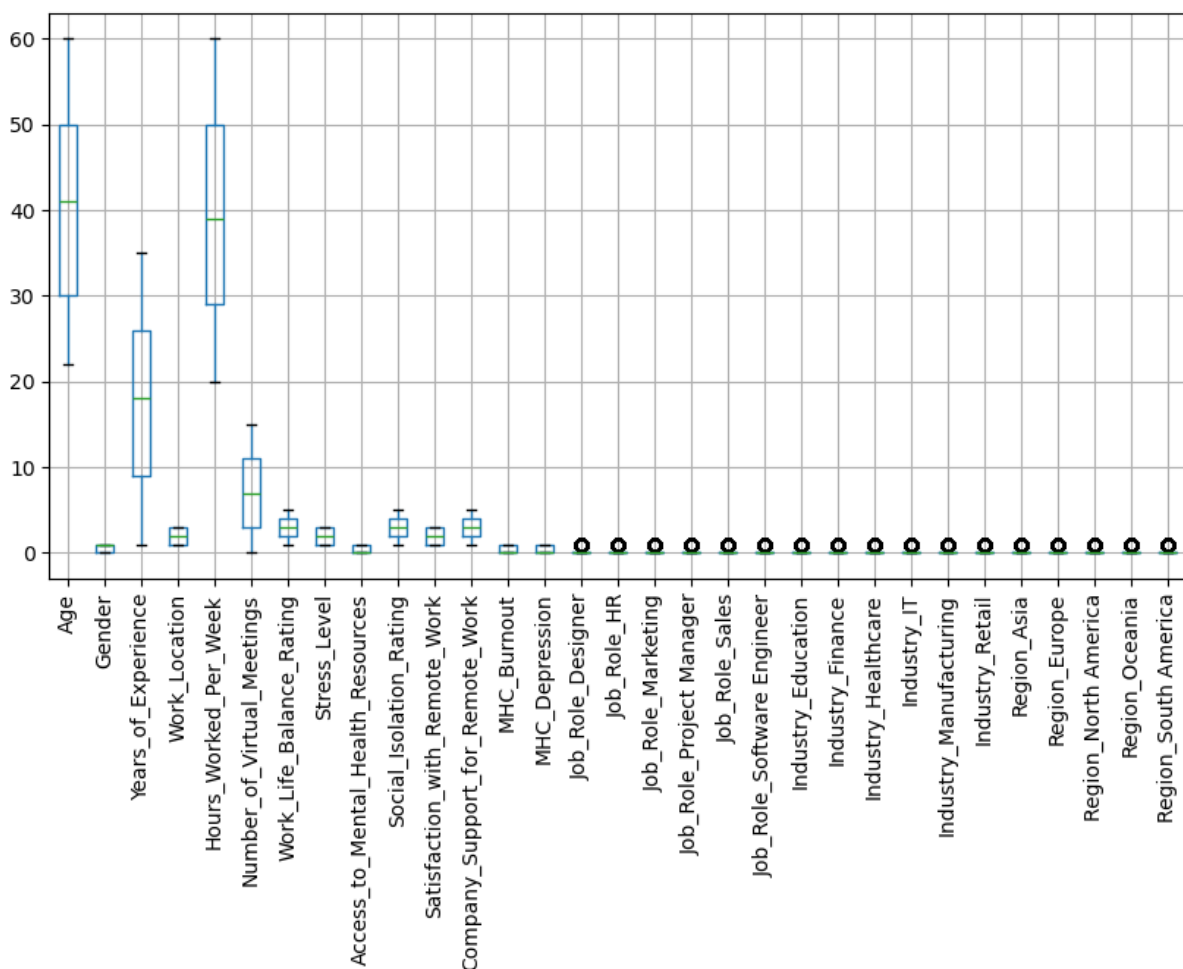
Figure 1 – Boxplot showing outliers present in the secondary Dataset

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 (SD), 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.

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.

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.

Results

Secondary Data Results

ANOVA Tests Results (Secondary Data)

To investigate whether the **type of work location** (remote, hybrid, or onsite) had a statistically significant effect on reported **stress levels**, a **one-way ANOVA** was conducted using the secondary dataset.

The ANOVA test findings returned an **F-statistic of 0.0822** and a **p-value of 0.9211**, which is well above the common alpha threshold of 0.05.

As a result, the null hypothesis could not be rejected, indicating that there was no statistically significant difference in stress levels across the three work location types in the secondary data.

Correlation Analysis Results and Heatmap of Secondary Data

A Pearson correlation matrix was computed to explore the linear correlations of the main variables across the secondary data **(see figure 2)**. These variables consisted of the hours worked, stress levels, score on the mental health support available, measure of work-life balance, mental health support available and the existence of organizational support.

Key findings include:

- The Work Stress Score had a highly significant positive correlation with the Number of Hours Worked Per Week (r ≈ 0.92), which implies that long hours of work may form the main source of stress.
- There was a negative correlation found to exist between stress and the presence of mental health services, suggesting that people having greater accessibility had less stress.
- Most reported correlations between demographic factors, including age, sex, and years of experience, and measures of stress were weak, possibly indicating little direct relationship.

```
                                         Age      Gender   Years_of_Experience  \
Age                                  1.000000  -0.022518            -0.004148
Gender                              -0.022518   1.000000             0.012084
Years_of_Experience                -0.004148   0.012084             1.000000
Work_Location                      -0.004790  -0.048112            -0.002277
Hours_Worked_Per_Week              -0.003125  -0.062223            -0.018181
Number_of_Virtual_Meetings         -0.014960   0.000515             0.024556
Work_Life_Balance_Rating            0.001685   0.031515             0.010466
Stress_Level                       -0.017523   0.009709            -0.011510
Access_to_Mental_Health_Resources  -0.010296  -0.032068             0.055036
Social_Isolation_Rating            -0.024543   0.016612            -0.000756
Satisfaction_with_Remote_Work       0.022530   0.002680             0.054097
Company_Support_for_Remote_Work     0.025452  -0.000895             0.055085
Work_Stress_Score                  -0.008605  -0.061911            -0.009299

                                   Work_Location  Hours_Worked_Per_Week  \
Age                                    -0.004790              -0.003125
Gender                                 -0.048112              -0.062223
Years_of_Experience                    -0.002277              -0.018181
Work_Location                           1.000000               0.011554
Hours_Worked_Per_Week                   0.011554               1.000000
Number_of_Virtual_Meetings             -0.028647              -0.049747
Work_Life_Balance_Rating               -0.045096              -0.011347
Stress_Level                            0.006561               0.034870
Access_to_Mental_Health_Resources       0.007874              -0.006700
Social_Isolation_Rating                 0.046002              -0.021419
Satisfaction_with_Remote_Work          -0.009972               0.008819
Company_Support_for_Remote_Work        -0.028084              -0.033242
Work_Stress_Score                       0.005451               0.924234

                                   Number_of_Virtual_Meetings  \
Age                                                 -0.014960
Gender                                               0.000515
Years_of_Experience                                  0.024556
Work_Location                                       -0.028647
Hours_Worked_Per_Week                               -0.049747
Number_of_Virtual_Meetings                           1.000000
Work_Life_Balance_Rating                             0.005485
Stress_Level                                        -0.039843
Access_to_Mental_Health_Resources                   -0.057823
Social_Isolation_Rating                              0.016397
Satisfaction_with_Remote_Work                       -0.012733
Company_Support_for_Remote_Work                      0.014176
Work_Stress_Score                                    0.318534
```

Figure 2 - Correlation Matrix of Secondary Data

The heatmap display also confirmed these relationships, as reflected in the use of darker colours to represent more positive or negative correlations **(see figure 3)**. However, most correlations outside of working hours and stress-related measures were modest in their magnitude.
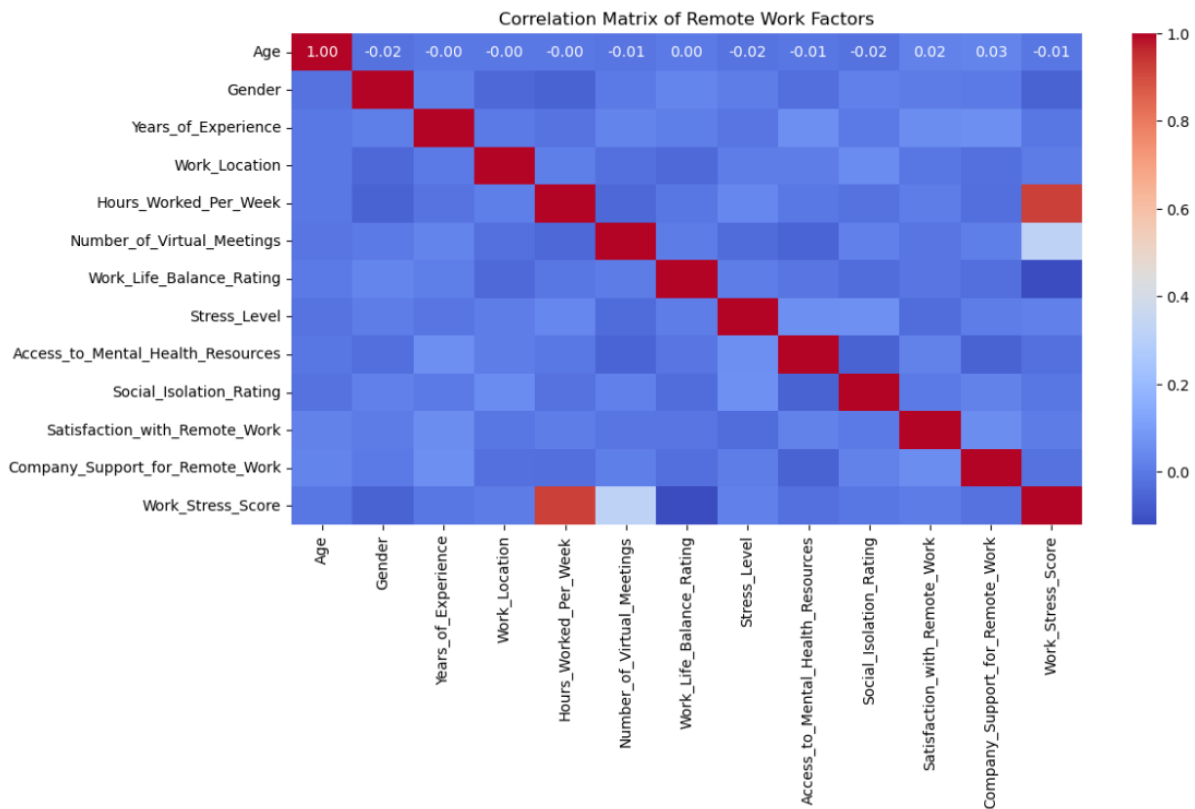

Figure 3 - Heatmap for correlation matrix of Remote Work factors (Secondary Data)

Regression Analysis Results of Secondary Data

To identify the variables that statistically predicted employees' stress levels in the second dataset, a multiple linear regression model was established from:

- Stress Level is the dependent variable
- Determinants:
    - Hours Worked Per Week
    - Number of Virtual Meetings
    - Access to Mental Health Resources
    - Company Support for Remote Work

```
                          OLS Regression Results
==============================================================================
Dep. Variable:          Stress_Level   R-squared:                       0.006
Model:                            OLS   Adj. R-squared:                  0.003
Method:                 Least Squares   F-statistic:                     1.913
Date:               Sun, 30 Mar 2025   Prob (F-statistic):              0.106
Time:                       13:52:13   Log-Likelihood:                 -1571.1
No. Observations:               1305   AIC:                             3152.
Df Residuals:                   1300   BIC:                             3178.
Df Model:                          4
Covariance Type:            nonrobust
==============================================================================
                                  coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const                           1.8751      0.104     18.056      0.000       1.671       2.079
Hours_Worked_Per_Week           0.0023      0.002      1.227      0.220      -0.001       0.006
Number_of_Virtual_Meetings     -0.0061      0.005     -1.268      0.205      -0.016       0.003
Access_to_Mental_Health_Resources 0.0897   0.045      1.994      0.046       0.001       0.178
Company_Support_for_Remote_Work 0.0089      0.016      0.547      0.584      -0.023       0.041
==============================================================================
Omnibus:                     10709.093   Durbin-Watson:                   1.980
Prob(Omnibus):                   0.000   Jarque-Bera (JB):              115.195
Skew:                            0.017   Prob(JB):                     9.68e-26
Kurtosis:                        1.545   Cond. No.                         196.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 4 - Regression Model Summary for Mental Health Survey (Secondary Data)

Results revealed that the model generated an R-squared of 0.006, indicating the predictors explained just 0.6% of the variance in stress levels. This is an extremely limited explanatory capacity.

- Access to mental health services was the only statistically significant predictor ($p = 0.046$) with a moderate positive relationship with stress levels. This finding suggests that those who are seeking or may require more mental health services are already likely to have high stress levels.
- The other independent variables, i.e., Weekly Hours Worked, Number of Virtual Meetings, and Company Support, failed to show statistical significance in the model developed ($p > 0.05$).

    o   The Durbin-Watson statistic of about 2.0 hints at a lack of serious problems of autocorrelation; nonetheless, the very low R-squared value limits the reliability of the model's predictive power.

Model Performance Comparisons (Secondary Data)

To further analyze the predictive performance of different modeling techniques, a comparison of performance was done using Ordinary Least Squares (OLS) Regression, Binary Logistic Regression, Decision Tree, and Random Forest methods. The following table shows each model with its corresponding performance measure, which can be R-squared, Pseudo R-squared, or Accuracy.

	Model	Score_Type	Score_Value
0	OLS Regression	R-squared	0.0060
1	Binary Logistic Regression	Pseudo R-squared	0.0001
2	Decision Tree	Accuracy	0.5860
3	Random Forest	Accuracy	0.6250

Figure 5 - Table Summary of Model Comparisons (Secondary Data)

The results showed that the highest accuracy, measured at 0.6250, was achieved by the Random Forest model, while Decision Tree achieved 0.5860. The regression models, by contrast, had poor predictive performance, reflected by R-squared values close to zero, namely 0.0060 for Ordinary Least Squares Regression and 0.0001 for Binary Logistic Regression.
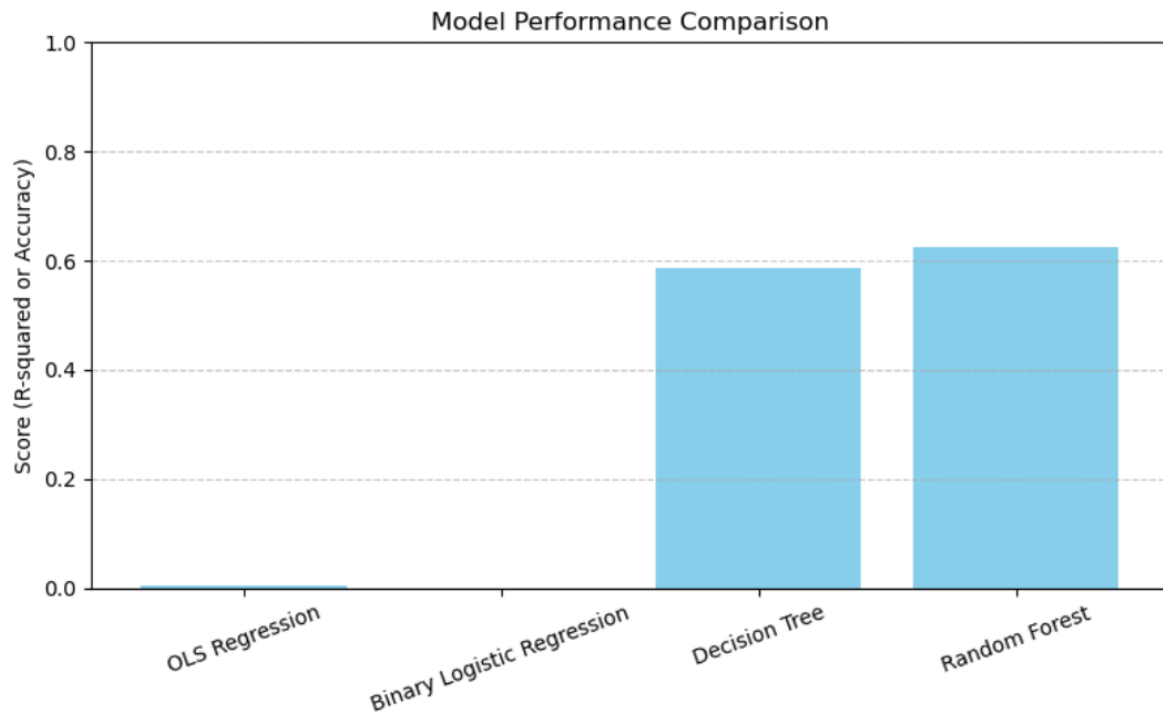
Figure 6 - Bar chart of Model Comparisons (Secondary Data)

The corresponding bar chart above visually highlights the differences in performance of the models. Tree-based models, like Decision Tree and Random Forest, clearly dominates and performs better regression-based models. This means that ensemble or tree-based methods might be better at finding nonlinear patterns in this case.

Primary Data Results

ANOVA Tests Results (Primary Data)

The ANOVA test result showed an F-statistic value of 1.6206 and a p-value of 0.2105. Because the p-value was more than 0.05, we could not determine statistically significant differences between the stress levels of the employees based on where they worked. Thus, the 'null hypothesis' of no variation between the stress levels of remote, hybrid, and onsite workers could not be rejected.

From these findings, it is possible to conclude that there is no considerable evidence to indicate that the work arrangement of the employee within this sample is a determinative predictor of the level of stress experienced within the participant's work setting.

Correlation Analysis Results and Heatmap of Primary Data

To analyze the inter-relationship among the key variables dealing with remote work and mental health, a correlation matrix was also computed. The correlation matrix provided an understanding of the inter-relationship among variables such as work location, work-life balance, social isolation, and mental health support vis-a-vis stress levels and a composite stress score.

The correlation matrix **(see Figure 5)** indicated some interesting inter-relationships:

- Work-Life Balance and Stress Level showed a moderate negative correlation (r = -0.5458), which would suggest that there is higher work-life balance associated with less levels of reported stress.
- Work-Life Balance and Stress Level also showed a negative correlation (r = -0.5458), which would indicate that individuals who reported experiencing social isolation most frequently reported higher stress levels.
- Employer Mental Health Support negatively weakly correlated with Stress Level (r = -0.0750) and moderately negatively correlated with stress_score (r = -0.2551). This suggests that higher employer-perceived support can slightly reduce stress levels.
- Stress Score, a constructed composite measure of total stress factors, was strongly correlated with Stress Level (r = 0.8816). This establishes that the stress score is a valid indicator of stress variation among participants.

In addition, lack of team connection was moderately positively correlated with Social Isolation Frequency (r = 0.3792), such that participants who were disconnected from their teams also reported more frequent social isolation.

```
                                        Gender  Work_Location  Work_Life_Balance  \
Gender                                1.000000       0.139237          -0.306130
Work_Location                         0.139237       1.000000          -0.077664
Work_Life_Balance                    -0.306130      -0.077664           1.000000
Stress_Level                         -0.306130      -0.077664           1.000000
Social_Isolation_Frequency           -0.055038      -0.160448          -0.545842
Lack_Of_Team_Connection              -0.031250      -0.017927          -0.333877
Employer_Mental_Health_Support       -0.318191      -0.046475          -0.075045
stress_score                         -0.095505       0.048351           0.881567

                                    Stress_Level  Social_Isolation_Frequency  \
Gender                                 -0.306130                   -0.055038
Work_Location                          -0.077664                   -0.160448
Work_Life_Balance                       1.000000                   -0.545842
Stress_Level                            1.000000                   -0.545842
Social_Isolation_Frequency             -0.545842                    1.000000
Lack_Of_Team_Connection                -0.333877                    0.379236
Employer_Mental_Health_Support         -0.075045                    0.338503
stress_score                            0.881567                   -0.876729

                                    Lack_Of_Team_Connection  \
Gender                                            -0.031250
Work_Location                                     -0.017927
Work_Life_Balance                                 -0.333877
Stress_Level                                      -0.333877
Social_Isolation_Frequency                         0.379236
Lack_Of_Team_Connection                            1.000000
Employer_Mental_Health_Support                     0.183249
stress_score                                      -0.389762

                                    Employer_Mental_Health_Support  stress_score
Gender                                                  -0.318191      -0.095505
Work_Location                                           -0.046475       0.048351
Work_Life_Balance                                       -0.075045       0.881567
Stress_Level                                            -0.075045       0.881567
Social_Isolation_Frequency                               0.338503      -0.876729
Lack_Of_Team_Connection                                  0.183249      -0.389762
Employer_Mental_Health_Support                           1.000000      -0.255122
stress_score                                            -0.255122       1.000000
```

Figure 7 - Correlation Matrix of Primary Data

These results are also then illustrated graphically in a heatmap **(see Figure 6)**, which displays easily the direction and magnitude of these correlations through colour gradations. More positive correlations are indicated by redder colours, and blue colours indicate negative relationships.
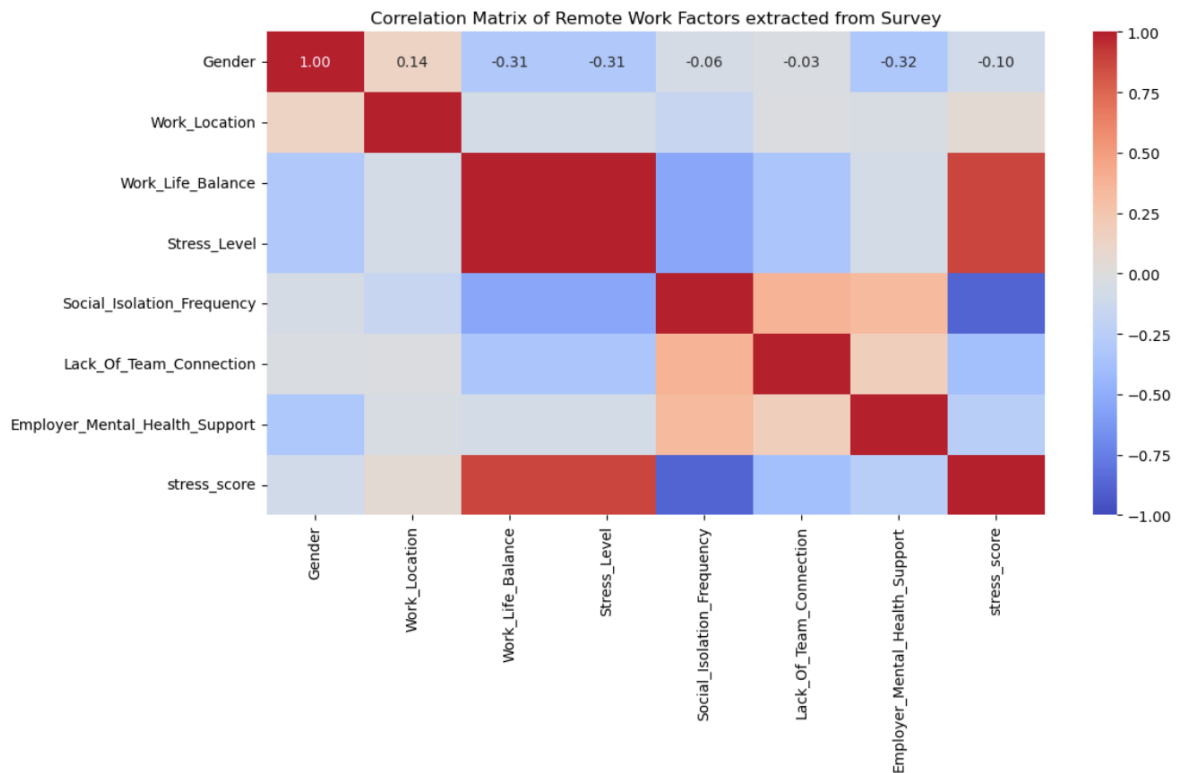
Figure 8 - Heatmap for correlation matrix of Remote Work factors extracted from survey (Primary Data)

Regression Analysis Results of Primary Data

To explore the variables that substantially influenced respondents, 'Stress Level', a multiple linear regression model was formulated. Stress_Level variable was identified as dependent for the study while independent variables included:

- Work-Life Balance (ordinal)
- Weekly Hours Worked (numerical)
- Categorical variables age group, gender, work location, and industry were pre-processed using one-hot encoding with the function C().

```
                        OLS Regression Results
==============================================================================
Dep. Variable:            Stress_Level   R-squared:                     1.000
Model:                             OLS   Adj. R-squared:                1.000
Method:                  Least Squares   F-statistic:                1.289e+29
Date:                 Tue, 15 Apr 2025   Prob (F-statistic):         5.83e-216
Time:                         23:16:52   Log-Likelihood:                1396.6
No. Observations:                   43   AIC:                           -2737.
Df Residuals:                       15   BIC:                           -2688.
Df Model:                           27
Covariance Type:             nonrobust
==============================================================================
                                          coef   std err         t    P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept                             6.661e-16  4.51e-15     0.148    0.885   -8.95e-15    1.03e-14
Weekly_Hours_Worked[T.28]             7.982e-17   6.8e-29  1.17e+12    0.000    7.98e-17    7.98e-17
Weekly_Hours_Worked[T.35]             4.441e-16  5.89e-15     0.075    0.941   -1.21e-14     1.3e-14
Weekly_Hours_Worked[T.36]             1.332e-15  2.73e-15     0.489    0.632   -4.48e-15    7.14e-15
Weekly_Hours_Worked[T.37.5]           4.441e-16  5.89e-15     0.075    0.941   -1.21e-14     1.3e-14
Weekly_Hours_Worked[T.38]             1.832e-15  4.73e-15     0.388    0.704   -8.24e-15    1.19e-14
Weekly_Hours_Worked[T.39]            -7.216e-16  4.58e-15    -0.158    0.877   -1.05e-14    9.04e-15
Weekly_Hours_Worked[T.39 hours ]       1.11e-15  5.69e-15     0.195    0.848    -1.1e-14    1.32e-14
Weekly_Hours_Worked[T.40]             5.551e-17  4.37e-15     0.013    0.990   -9.27e-15    9.38e-15
Weekly_Hours_Worked[T.40/60 (by choice!)]     0  2.65e-15         0    1.000   -5.65e-15    5.65e-15
Weekly_Hours_Worked[T.42]            -2.22e-16   5.37e-15    -0.041    0.968   -1.17e-14    1.12e-14
Weekly_Hours_Worked[T.45]            -3.331e-16  5.06e-15    -0.066    0.948   -1.11e-14    1.05e-14
Weekly_Hours_Worked[T.48]                    0   5.37e-15         0    1.000   -1.14e-14    1.14e-14
Weekly_Hours_Worked[T.53]             4.441e-16  2.07e-15     0.215    0.833   -3.96e-15    4.85e-15
Weekly_Hours_Worked[T.8]              1.776e-15  4.94e-15     0.359    0.724   -8.76e-15    1.23e-14
Weekly_Hours_Worked[T.8.75]           8.882e-16  5.02e-15     0.177    0.862   -9.81e-15    1.16e-14
Weekly_Hours_Worked[T.9]              1.998e-15  5.15e-15     0.388    0.704   -8.99e-15     1.3e-14
C(Age_Group)[T.25-34]                -2.22e-16   3.57e-15    -0.062    0.951   -7.83e-15    7.38e-15
C(Age_Group)[T.35-44]                -1.11e-15   3.66e-15    -0.304    0.765    -8.9e-15    6.68e-15
C(Age_Group)[T.45-54]                -1.86e-15   3.79e-15    -0.490    0.631   -9.95e-15    6.23e-15
C(Age_Group)[T.55+]                  -1.776e-15  4.92e-15    -0.361    0.723   -1.23e-14    8.72e-15
C(Gender)[T.1]                       -4.441e-16  2.11e-15    -0.211    0.836   -4.93e-15    4.04e-15
C(Work_Location)[T.2.0]              -2.22e-16   1.57e-15    -0.142    0.889   -3.56e-15    3.12e-15
C(Work_Location)[T.3.0]              -8.882e-16  3.99e-15    -0.222    0.827    -9.4e-15    7.62e-15
C(Industry)[T.Biotechnology]         -1.721e-15  6.81e-15    -0.253    0.804   -1.62e-14    1.28e-14
C(Industry)[T.Healthcare/Pharmaceuticals]  2.22e-16  2.07e-15    0.107    0.916   -4.19e-15    4.63e-15
C(Industry)[T.IT/Technology]         -6.661e-16     2e-15    -0.333    0.743   -4.92e-15    3.59e-15
C(Industry)[T.Legal]                  6.661e-16  3.69e-15     0.181    0.859   -7.19e-15    8.52e-15
C(Industry)[T.Marketing/Advertising]  2.22e-16   2.65e-15     0.084    0.934   -5.43e-15    5.88e-15
C(Industry)[T.Operations/Logistics]          0   4.31e-15         0    1.000   -9.19e-15    9.19e-15
C(Industry)[T.Project management EU Projects]  6.106e-16  4.81e-15  0.127  0.901   -9.64e-15    1.09e-14
C(Industry)[T.Transport]              1.332e-15  2.73e-15     0.489    0.632   -4.48e-15    7.14e-15
Work_Life_Balance                        1.0000  1.06e-15   9.4e+14    0.000       1.000       1.000
==============================================================================
Omnibus:                         1.847   Durbin-Watson:                   0.573
Prob(Omnibus):                   0.397   Jarque-Bera (JB):                1.314
Skew:                           -0.183   Prob(JB):                        0.518
Kurtosis:                        2.226   Cond. No.                     3.75e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 5.94e-31. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Figure 9 - Regression Model Summary for Mental Health Survey (Primary Data)

The regression analysis was performed under the Ordinary Least Squares method. The model showed an R-squared and Adjusted R-squared of 1.000, which indicates an outstanding fit with the empirical data. However, the proximity of such an unusually high value, the presence of an eigenvalue approaching zero, and a very high condition number of 10^{16}. The suggested value of 16 indicates the possible presence of multicollinearity or overfitting problems in the model. These situations can occur due to the relationships among the independent variables or due to the large number of

parameters with relatively few available samples (n = 43, with only 27 degrees of freedom available).

Discussion

Secondary Data Analysis

Interpretation of ANOVA Tests Results (Secondary Data)

The data gathered from the ANOVA show that the work environment lacks a statistically significant influence on the level of stress in the secondary dataset. This is the opposite of some previous research (Charalampous, 2018), which measured high levels of stress in purely remote-working conditions. Alternatively, this finding might represent the possibility of normalization of remote-working practice post-pandemic or is potentially influenced by other factors, including support from the employer or flexible policies, that this research has not adequately controlled for.

Interpretation of Correlation Results (Secondary Data)

The correlation matrix revealed valuable information on how important variables in the secondary data are correlated with each other. The most significant relationship identified was between working hours per week and work stress score, which indicated that longer working hours are strongly linked to higher stress. This could be assumed to be common sense but confirms that workload management is still at the core of employee well-being under remote conditions.

Somewhat surprisingly, perceived organizational support and mental health service availability had weak negative correlations with stress levels. This may either mean support mechanisms are available but not having an effect on stress overall, or that other unmeasured factors, including coping mechanisms or outside pressures are taking a more significant role. The low correlations with demographic variables such as age and gender indicate that work conditions rather than individual characteristics are more likely to influence stress in this sample.

Overall, the correlation analysis lends itself to the more general notion that the amount and structure of work will have a more immediate effect on stress than either demographic or organizational support variables by themselves.

Interpretation of Regression Analysis Results (Secondary Data)

The regression model was designed to predict levels of stress based on hours worked, virtual meetings, employer support, and availability of mental health resources. The model's low R-squared, however, indicated that together these predictors accounted for minimal variation in stress. This means that while these constructs can influence well-being given the right conditions, they alone do not account for considerable differences in levels of stress in this sample.

Most notably, only one variable broke through to statistical significance, that being availability of mental health services, and even that was positively related to stress. What this could indicate is a spurious trend whereby already stressed people will tend to use or perceive available resources, and not necessarily those resources which alleviate stress. The other predictors of *'Hours_At_Work'*, *'Meetings_Per_Week'* variables were not significant, but could indicate a weak or limited effect in this population, or that there are other stronger unmeasured forces at work.

Primary Data Analysis

Interpretation of ANOVA Tests Results (Primary Data)

A one-way ANOVA on the main dataset was conducted to examine whether remote workers, hybrid workers, and workers who were onsite experienced varying levels of stress. Surprisingly, none of the differences were significant. Stress from remote work has been hypothesized in some studies to be higher, due perhaps to loneliness or due to failing to distinguish between work life and home life, but these patterns didn't manifest in this sample. There are numerous reasons for this. Perhaps the site where individuals work isn't the primary source of their stress. Employer support, presence of resources for mental health issues, or flexible work options may be more significant in terms of how stressed individuals perceive themselves.

This finding aligns with modern perspectives that place emphasis on contextual and organizational aspects over the physical work environment.
In addition, minimal variation in the observed stress levels might be due to the continued normalization of teleworking and hybrid work arrangements in which workers might have adjusted or developed coping mechanisms that counteract stress regardless of their workplace environment. It should be noted that the sample size and diversity of industries and workplaces potentially undermined any potentially significant statistical trends. Therefore, follow-up research might improve its outcomes by using a more specific job classification- or industry-sector-based segmentation.

Interpretation of Correlation Results (Primary Data)

Correlation analysis suggests that:

- Reported stress levels also appear to be greatly influenced by work-life balance dynamics and the sense of isolation.
- Employer-provided mental health assistance has been linked to better mental health outcomes in employees, although the strength of the correlation is relatively weaker.
- Relational and social indicators, including feelings of group isolation and alienation, have implications for mental health.

While correlation does not always mean causality, the results add to the overall understanding of the key issues which might be altered through organizational policy, hence potentially enabling increased levels of telework.

Interpretation of Regression Analysis Results (Primary Data)

The multiple linear regression model tried to determine what variables were good predictors of stress level. Although the initial model fit was seemingly very, very good (R^2 = 1.000), such a perfect fit raised suspicions of overfitting. The occurrence of multicollinearity or the presence of too many categorical dummy variables certainly overestimated the model's explanatory power, as revealed by the very high condition number as well.

Among the predictors, work-life balance was likewise the strongest significant factor, validating its negative correlation with stress levels. This reveals that the manner in which one manages or sees balancing work and personal tasks may potentially be an influencing factor for their mental health.

The other predictors, including work hours per week, work setting, industry, age group, and gender, likewise failed to have statistically significant effects in this sample. This may be due to a combination of circumstances (e.g. sample size, diversity of respondent backgrounds, or overlap among variables) which may have dampened the model's capacity for unique effects separation from each predictor.

The general conclusion would be that although regression modelling presents a satisfactory way of isolating causes of stress, the practice usage still depends on variable quality management, their interaction, and also similarly the general organisational setting under which remote working is done.

Self-Reflections/Challenges

The experience of writing this dissertation provided a valuable base for the use of theoretical knowledge and technical skills, as well as the acquisition of skills in solving real-world problems. Several challenges cropped up during the project period and had their impact while altering the project's course and development.

The greatest challenge of this project lay in securing the primary data. Preparations of the survey to generate significant data while keeping it brief and to the point were time-consuming. Further, disseminating the survey, finding a representative sample of participants, and achieving their responses formed crucial challenges. As the project went forth, the process of securing firsthand data came to emerge as demanding more than the mere asking of questions; the right time, the right follow-ups, and good communication are the keys to securing quality responses within a limited timeframe.

The second hurdle was statistical analysis and modelling. Interpretation of the secondary and primary data, and conducting procedures such as ANOVA, regression, and correlation analysis, consumed more time than initially anticipated. Checking assumptions for all the tests being met, overfitting of regression models, and remodelling the latter for correctness were particularly iterative and mind-taxing. However, it made me learn to appreciate applied statistics and messiness of the data in the real-world even more.

Data preprocessing and data management tasks turned out to be more complex than initially anticipated. Cleaning the secondary dataset necessitated careful attention, including imputation of missing values, correction of erroneous formats, and the proper coding of categorical variables. This meant continuous checks of previous steps to guarantee the correctness of operations, uniform standardization of entries, and achievement of consistency in the adjusted dataset. While this process demanded manual labour, this highlighted the importance of careful data preparation as the cornerstone to sound analysis.

Ultimately, the process of drawing a legitimate connection between the results and the broader corpus of knowledge came as the final hurdle. Translating statistical results into a form appropriate for dissemination in the scholarly community frequently represented a demanding process, even where the outcomes varied from preliminary expectations. This process required a thorough review of related literature, as well as critical inquiry in quest of explaining discrepancies between the realized outcomes and expected outcomes.

Despite the difficulties faced, the project was extremely rewarding. It enhanced the research and technical writing skills, as well as promoting the appreciation of the time and effort devoted to producing studies that are methodologically sound as well as practically useful.

Conclusion

The objective of this study was to explore the impact of home working upon employees' mental wellbeing in the context of stress levels, work-life balance, and utilization of support services. Employing secondary data obtained from an openly accessible dataset and primary data obtained from a purpose-built survey, the study utilized statistical procedures like ANOVA, correlation analysis, and regression modelling to establish the dominant trends and patterns.

From the primary data, findings revealed that while work location (onsite, hybrid, remote) had no effect on stress levels at all, work-life balance and social isolation revealed straightforward relationships with stress. That is, participants who reported better work-life balance consistently reported lower stress levels, and those who reported frequent feelings of isolation reported higher stress levels.

In the secondary data set, parallel trends were evident. The regression model was generally weak, however, and only access to mental health support proved to be a statistically significant predictor, albeit that its direction of effect needed careful interpretation. Correlation analysis also provided a heightened sense of the robust relationship between longer working hours and increased stress, which served to add richness to the general appreciation of workload as a remote stressor.

Cumulatively, the findings point out that the circumstances of remote working including support systems, workload, and perceived balance have a more significant influence on mental health than the work arrangement per se. This has specific relevance to employers, policymakers, and HR practitioners interested in guaranteeing healthy, productive remote working cultures in the post-pandemic era.

Although the study offers valuable information, it also recognises limitations in terms of sample size, possible response bias, as well as overfitting of regression models. Possible future studies can involve longitudinal data or sector-specific investigations to assist in a clearer comprehension of how mental health results change with remote or hybrid environments over time.

Finally, this study adds to the increasing amount of research on teleworking by highlighting the value of not only where individuals work, but also in how they are supported to do so.

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%20research%2C%20which%20aims%20to%20predict%20future%20events,an%20illness%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.2022.960955/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.

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