Impact of working remotely on social wellbeing and productivity

Amaury Chartier Valentine Salvat Thomas Blondel
Elisa Rigazzio

2026-10-12

This study explores how remote work affects productivity, social well-being, and job satisfaction. Using the 2020 NSW Remote Working Survey, we applied descriptive statistics and correlation analysis to quantify attitudes toward remote work. Findings show mixed productivity outcomes: most employees report gains up to 50%, fewer others declines. Collaboration indicators remain neutral, while reduced commuting significantly increases personal and family time, supporting better work–life balance. Most respondents prefer more remote work, suggesting higher satisfaction overall. Limitations include self-reported data and cross-sectional design. Results indicate organizations should maintain flexible work arrangements and support social engagement strategies.

Introduction

Project Goals

Our objective in this project is to understand how remote work, following the COVID-19 pandemic, has changed firms' work practices. During this period, many companies had no choice but to adopt remote working practices, reshaping traditional work environments. The main goal is to analyze how working remotely influences people's mental health, job satisfaction, and performance levels. Using accurate data from the Remote Working Survey (2020), we want to explore factors such as communication satisfaction, work-life balance, social connection, and productivity outcomes, in order to understand whether remote work affects employees positively or negatively. By converting survey responses into numerical values, we can quantify attitudes toward remote work and compare groups such as gender, age, or employment type.

As we began working with the dataset, we refined our focus on the social and psychological dimensions of remote work. To make the analysis easier to interpret, we also transformed qualitative survey responses into numerical values. And, in addition, we removed categories with very small sample sizes to keep the dataset consistent and avoid unreliable comparisons.

Research Questions

- How does remote work influence employees' overall productivity?
- How does remote working affect employees' ability to maintain social connections and avoid feelings of isolation?
- Has job satisfaction decreased or increased as a result of remote working conditions?
- Which factors contribute the most to employees' overall satisfaction with remote work?

Related Work

In this section, we discuss the main academic studies, methods, and resources that guided our analysis of remote work and well-being.

Domain literature

A lot of studies have looked at how remote work affects people's productivity, well-being, and social life, which helps us understand the context of our own project.

For example, a systematic review by Oakman et al. (2022) showed that working from home can improve productivity and work–life balance, but it can also lead to more isolation, communication problems, and mental stress. These points match the variables we analyze in our dataset, such as job satisfaction, social connection, and communication quality. Another study by Correia et al. (2024) investigated how research on remote workers has evolved over the years. They found that psychological factors like loneliness, emotional pressure, and reduced social interaction are becoming increasingly important in academic discussions. This supports our decision to focus mainly on the social and psychological side of remote work instead of technical aspects.

Finally, a well-known experiment by Bloom et al. (2013) showed that remote work can significantly improve productivity when employees have clear structures and good communication with their team. This connects directly to our analysis, since we also look at collaboration, communication, and satisfaction using the Remote Working Survey (2020). Overall, these studies confirm that the social, psychological, and performance-related impacts of remote work are important topics to explore, and they strongly support the research questions we chose for our project.

Methodological references

For our project, we used a few common data-science methods that are usually applied in survey analysis. We started with simple descriptive statistics to get a first idea of the patterns in the data, things like averages, proportions, and basic comparisons between groups. We also looked at correlations to see how different variables are related, for example whether good communication or social well-being is linked to higher productivity. Since many of our questions were answered with text options (like “strongly agree”), we converted these answers into numbers using standard Likert-scale coding so that we could analyze them properly.

On the technical side, we followed the Python methods shown in Azizi (2025), which helped us clean the data and structure our exploratory analysis in Google Colab. We mainly used pandas for organizing and transforming the dataset (following McKinney, 2022), and matplotlib/seaborn to create the visualizations with the help of VanderPlas, 2016. These tools and references guided our process and helped us follow good data-science practices while analyzing the Remote Working Survey (2020).

Course material

The structure of our project is inspired by the material used in class, especially the DSAS notes from Azizi (2025). These resources helped us understand how to organize our data analysis, clean our dataset, and apply basic Python techniques in Google Colab. With that said, we used the same approach shown in the course for tasks like loading data, creating new variables, handling missing values, and running exploratory data analysis. The examples provided in the course made it easier for us to use tools like pandas, matplotlib, and seaborn in a consistent way throughout the project.

Technical resources

On the technical side, we mainly relied on a few well-known Python resources to help us structure and run our analysis. McKinney’s book (2022) guided us with all the pandas-related tasks, such as cleaning the dataset, creating new variables, and handling missing values. We also used matplotlib and seaborn to create readable plots for our visualizations by following VanderPlas’s (2016) explanations.

Data

Sources

The dataset used in this project is the 2020 Remote Working Survey, part of the NSW Remote Working Survey series, publicly available through the New South Wales Government’s Open Data Portal. “Data source: Remote working survey 2020” The 2020 survey was conducted during August and September 2020 and aimed to understand workers experiences and attitudes toward remote and hybrid working following the first phase of the COVID-19 pandemic to study the impact of remote work on professional and personal well-being.

To be eligible, respondents had to:

- be employed NSW residents

- have experience of remote working in their current job. After excluding unemployed individuals and those whose occupations cannot be performed remotely (dentists, cashiers, cleaners), the sample represents approximately 59% of NSW workers.

This dataset corresponds to the 2020 wave of the NSW Remote Working Survey. Although a second wave of the survey was collected in March and April 2021, it is not included in the present analysis. The 2021 dataset differs substantially in terms of survey structure, variable definitions, and measurement scales, which would require extensive re-harmonisation and could introduce inconsistencies across waves. To preserve internal consistency and ensure methodological robustness, the analysis therefore focuses exclusively on the 2020 dataset.

To provide additional contextual insight, an external dataset sourced from Kaggle is introduced at the exploratory data analysis stage. This dataset contains country-level COVID-19 indicators, including total cases and deaths, and is used for descriptive cross-country comparison. It is not merged with the survey data and serves only to contextualize the remote working analysis.

Description

The cleaned 2020 NSW Remote Working Survey dataset contains four main types of variables: categorical, ordinal, numeric, and binary.

- The categorical variables describe qualitative characteristics of respondents and their employment context, such as Industry, Job_type, Organisation_Size, Household, and Years_in_job. These variables are stored as text and provide context for grouping and comparison across sectors or demographic profiles.
- The ordinal variables represent ordered responses on Likert scales, reflecting opinions and perceptions about remote working. Variables such as Org_encouraged_remote_last_year, Collaboration_remote_last_year, Org_encouraged_remote_3_months, and Collaboration_remote_3_months are encoded numerically from 1 (“Strongly disagree”) to 5 (“Strongly agree”), allowing for quantitative analysis of attitudes.
- The numeric variables capture measurable quantities including time allocation, remote work proportions, and productivity. Examples include Age, Remote_pct_last_year, Preferred_remote_last_year, Productivity_remote_vs_workplace, and several variables representing hours spent on commuting, working, and personal or domestic activities. These are stored as integers or floats, making them suitable for descriptive statistics and correlation analysis.
- The binary variables (Gender, Managing_position) indicate Male/Femal or Yes/No conditions, encoded as 0 and 1. These variables enable comparisons between distinct groups.

The dataset also contains several important pieces of metadata that ensure its quality and usability for analysis. Each observation is identified by a unique respondent code (Response_ID), which guarantees traceability and prevents duplication during data processing.

In addition, all variable names were standardized to concise, descriptive identifiers (Org_encouraged_remote_last_year) to make the variables easier to read, reuse, and analyze in statistical software. This naming convention makes the analysis process simpler and clearer throughout the project.

ℹ Dataset Overview Template

- **File used:** 2020_rws-updated.csv
- **Format:** CSV (comma-separated values)
- **Encoding:** latin-1 (ISO-8859-1) (Remark: when loading the dataset in Python (Google Colab), attempting to read with utf-8 caused a UnicodeDecodeError due to typographic apostrophes and special characters. The correct parameter encoding latin1 was required to successfully load the data.)
- **Memory usage:** approximately 7.46 MB
- **Number of observations:** 1507 respondents (1370 rows after cleaning)
- **Number of variables:** 73 columns (25 after cleaning)
- **Time period:** August and September 2020
- **Geographic coverage:** New South Wales, Australia
- **Key variables:** Gender, Age, Job_type, Organisation_Size, Managing_position, Remote_pct_last_year, Preferred_remote_last_year, Org_encouraged_remote_last_year, Collaboration_remote_last_year, Productivity_remote_vs_workplace

Loading Data

Following best practices, the file is loaded using a relative path via project_root, ensuring that the document remains fully reproducible regardless of the execution location. Because the original file contained specific typographic characters that caused decoding issues during import, the dataset is read using the latin-1 encoding.

After loading the file, several checks were performed to confirm correct import: verification of the dataset dimensions, inspection of column names and data types (df.info()), preview of the first rows to ensure values were properly formatted. These steps guarantee that the dataset is correctly imported and ready for subsequent exploratory analysis.

```
Dataset shape: 1507 rows × 73 columns

Data types:
Response ID
What year were you born?
What is your gender?
Which of the following best describes your industry?
Which of the following best describes your industry? (Detailed)

Compare remote working to working at your employer s workplace. Select the worst aspect of re
life balance ; My on-the-job learning opportunities ; Managing my personal commitments ; My c
Compare remote working to working at your employer s workplace. Select the best aspect of rem
life balance ; My on-the-job learning opportunities ; My daily expenses ; My personal relatio
Compare remote working to working at your employer s workplace. Select the worst aspect of re
life balance ; My on-the-job learning opportunities ; My daily expenses ; My personal relatio
Compare remote working to working at your employer s workplace. Select the best aspect of rem
Compare remote working to working at your employer s workplace. Select the worst aspect of re
Length: 73, dtype: object

First 5 rows:
```

	Response ID	What year were you born?	What is your gender?	Which of the following best describes y
0	1	1972	Female	Manufacturing
1	2	1972	Male	Wholesale Trade
2	3	1982	Male	Electricity, Gas, Water and Waste Serv
3	4	1987	Female	Professional, Scientific and Technical S
4	5	1991	Male	Transport, Postal and Warehousing

Wrangling

General Transformations

Several preprocessing and wrangling steps were performed to prepare the dataset for analysis. Before detailing each transformation.

All preprocessing steps described below are fully reproducible and validated. Each transformation relies on deterministic Python operations (such as replace(), rename(), astype(), or simple arithmetic), meaning that re-running the same code on the raw dataset will always produce identical results. After every transformation, checks such as head(), value_counts(), or describe() were used to validate that the modifications were correctly applied and that the resulting values were coherent (age ranges, Likert scales, or percentage conversions).

Variable Duplicated and Variable Reduction

Before any transformation, it is essential to verify that each observation in the dataset is unique. Duplicate rows can bias the analysis by over representing certain respondents or records. Identifying and removing them ensures data integrity.

This operation can be rerun on the raw dataset at any stage of the workflow, guaranteeing consistent detection of duplicated records. The command returned an empty DataFrame, confirming that no duplicate rows were present. Therefore, no observations were removed in this step.

In parallel, a large number of variables from the initial dataset were removed during preprocessing in order to improve readability, interpretability, and analytical relevance. This included variables such as industry, current occupation, share of time spent remote working, perceived best aspects of remote work (working hours, work-life balance, job satisfaction), and perceived barriers to remote working (organisational systems, workspace conditions, or management discouragement). This was done by dropping selected columns as shown below:

This step allowed the dataset to be reduced to a smaller and more meaningful set of variables, while preserving all observations.

Rename Columns

Renaming the original survey questions into shorter and clearer variable names was necessary to improve readability and make the dataset easier to work with.

Standard Name Organisation Size

The Organisation_Size variable was recoded by replacing the long original text categories with shorter, standardised. To make grouping and comparison more intuitive.

Binary Variables

The variables Gender and Managing_position, the original text responses were converted into binary categories to make them suitable for statistical analysis and group comparisons. Respondents who selected “Rather not say” for gender were removed, as this category represented only two individuals and could not form a meaningful subgroup. Gender was then encoded as 0 = Male and 1 = Female, while Managing_position was encoded as 0 = No and 1 = Yes, indicating whether the respondent supervises others. This transformation produces clean, consistent, and analysis-ready binary variables that can be easily used in descriptive statistics, visualisations, and modelling.

Variable Ages

The variable Age, the dataset originally reported the respondent's year of birth. This information was converted into a more interpretable age variable by subtracting the birth year from 2020, the year the survey was conducted. For example, a respondent born in 1985 becomes 2020 − 1985 = 35 years old. Expressing this information directly as age is more intuitive and easier to interpret in descriptive statistics, comparisons, and visualizations.

Variable Years in Job

The variable Years_in_job, the original responses were long text categories describing tenure intervals. These were simplified into shorter and more readable labels: "5+", "5-", and "1-". This transformation keeps the original meaning while making the variable easier to interpret, compare, and visualise in tables and plots.

Variable Likert Scale Mapping

The Likert-scale variables, this transformation allows these subjective perceptions to be analysed quantitatively. The four variables related to organisational support and collaboration were mapped using this scale. Any missing responses were replaced with the neutral value 3, corresponding to "Neither agree nor disagree", to keep these observations in the dataset while avoiding bias from missing attitudes.

Variables Working Remotely

The variables describing the percentage of time spent or preferred working remotely, the original responses ("Less than 10% of my time"…"100% - All of my time"). These were first converted into numeric percentage values using a mapping dictionary. All four percentage-related variables were processed in the same way. Before applying the mapping, non-breaking spaces () and extra whitespace were removed from the strings to avoid parsing issues. After the text values were mapped to numeric percentages ("80%" → 80), the values were converted into proportions between 0 and 1 by dividing by 100.

For example: 80 becomes 0.80 50 becomes 0.50 0 becomes 0.00

This final step creates clean numerical variables that can be easily averaged, compared, or visualised in the analysis.

Variables Productivity Cleaning

The variable Productivity_remote_vs_workplace, the survey responses (I'm 20% more productive when I work remotely" or "I'm 10% less productive"). These text responses were converted into clean numeric values using a custom parsing function. The mapping works as follows: Statements indicating more productive remotely return a positive percentage ("20% more productive" → +20). Statements indicating less productive remotely return a negative percentage ("10% less productive" → –10). Statements indicating no difference return 0. This transformation results in a numeric scale where positive values mean higher productivity when working remotely, negative values reflect lower productivity, and zero indicates no change. This allows the variable to be analysed quantitatively, averaged across groups, or used in visualisations.

Spotting Mistakes and Missing Data

Before conducting the analysis, the dataset was reviewed to identify missing values, inconsistencies, and unusually small categories. This ensures that only reliable and interpretable data are used in the following steps.

Identified missing data

- Most variables contained very few missing values, mainly in attitudinal questions where some respondents simply did not answer.
- Additional missingness appeared when converting textual inputs (percentages or productivity statements) into numeric formats, entries that could not be parsed were intentionally converted to NaN.
- The inspection of category sizes showed that some groups were extremely small, such as the “Rather not say” gender category (2 respondents), which was removed because it cannot support meaningful analysis.

Approach to handling missing data

- Deletion was applied when missingness or category size was extremely small and analytically useless.
- Imputation with a neutral value (3 = Neither agree nor disagree) was used for missing Likert-scale answers to preserve observations without creating bias.
- Conversion to numeric with errors=“coerce” was used for percentage and productivity variables, producing valid NaN values when entries could not be interpreted.

Because missingness was limited and mostly isolated to subjective questions, more complex methods were not necessary.

Future handling of small or irrelevant categories

- If additional categories, during the analysis progress, are found to be too small to contribute meaningfully, they will be removed, flagged, or when conceptually appropriate grouped together with similar categories to preserve statistical power.

Future variable selection

- Some variables will ultimately explain the effects of remote work on health, productivity, or work life balance better than others. Variables that show no meaningful correlation or explanatory power in later stages of the project will also be removed to keep the analysis focused, interpretable, and relevant.

Listing Anomalies and Outliers

A detailed inspection of the numeric variables, using both summary statistics and histogram visualisations, revealed several anomalies and potential outliers in the dataset.

Detected anomalies

- Age variable: Two respondents appear with an age of 120 years, which is biologically impossible and indicates a clear data entry error. This type of anomaly commonly occurs when the birth year is mistyped—for example, entering 1900 instead of 2000, or 2005 instead of 1920—which produces unrealistic age values. These observations will therefore be removed, while other extreme but plausible ages (such as 75 or 83) are retained at this stage. The minimum value (19) is plausible for entry-level workers.
- Domestic_hours_workplace: A single observation of –1 hour is impossible and indicates a recording error.
- Working and commuting time variables: Extremely high records were observed, such as 23 hours of work in a day or 10–12 hours of commuting.

While unlikely, these may represent exceptional cases (long-distance travel, extended shifts). They are retained unless later analysis shows they distort results.

- Commute_hours_remote: Values up to 12 hours on remote days are implausible and likely due to misinterpretation or input mistakes.
- Productivity variable: Extreme values (+50%, –50%) appear in the data but remain plausible since the question explicitly asked respondents to report percentage differences.

Approach to handling outliers

- Outliers were evaluated using:
  - Visual inspection (histograms from univariate EDA)
  - Summary statistics (min/max, interquartile ranges)
  - Domain knowledge (negative hours, impossible ages)
- Following best practices
  - Impossible values (age = 120, domestic hours = –1) will be removed before modelling.
  - Extreme but plausible behaviours (very long workdays) are kept unless they later bias model results.
  - Additional outliers identified during the analysis phase may be removed, flagged, or grouped depending on their relevance and impact.

Outliers are not always errors; some reveal meaningful variability in remote work habits. The chosen approach maintains data integrity while ensuring that the analysis focuses on realistic, interpretable patterns.

The following EDA sections will further analyse these variables to understand their patterns and relationships.

EDA

Univariate Analysis

Examine each variable individually to understand its distribution, central tendency, and spread.

From this code we got the full results of all our variables in terms of distribution sorted by occurrence. The relevant information we gathered from these are the followings:

```
--- Age ---
Age
35      47
56      47
60      45
50      44
30      43
59      43
42      43
49      40
40      39
55      38
43      38
54      37
33      37
58      37
57      36
51      36
32      36
31      36
47      35
45      35
52      34
48      34
```

```
63     33
36     33
61     32
34     32
44     30
53     30
39     28
37     28
46     27
38     26
64     26
62     26
29     24
41     24
25     21
28     18
26     13
27     12
24     11
65      9
20      8
23      7
22      6
120     2
21      1
83      1
75      1
19      1
Name: count, dtype: int64
```

- Age: we have some outliers for the age variable with 2 people 120 years old that needs to be treated

```
--- Gender ---
Gender
0    766
1    604
Name: count, dtype: int64
```

- Gender: good balance (604 / (604+766) = 44% of women)

```
--- Years_in_job ---
Years_in_job
5+    725
5-    494
1-    151
Name: count, dtype: int64
```

- Years in job: sample is quite experienced with only 11% with less than 1 year of experience.

```
--- Productivity_remote_vs_workplace ---
Productivity_remote_vs_workplace
 0     400
 50    199
 20    191
 30    173
 10     91
-20     85
-10     85
 40     74
-30     40
-50     26
-40      6
Name: count, dtype: int64
```

- Productivity remote vs workplace: we can see that for most of them productivity is either equivalent, or they even gained in productivity (46% gained between 20 to 50% of productivity).

Distribution Plot

Concretely, if we look at the distribution of the change in productivity with this code :

```
Unable to display output for mime type(s): text/html
```

```
Unable to display output for mime type(s): text/html
```

```
Skewness change in productivity:
-0.1398603974975266
```

The histogram we obtain from the code shows an important part of 0 meaning for a lot of people working remotely doesn't affect their productivity. Looking at the skewness (-0.13), we can deduce that data is significantly more on the right of the graph, meaning that people are more productive being at home for work.

Bivariate Analysis

```
Unable to display output for mime type(s): text/html
```

In this Violin plot, we can understand that the change in productivity is not really influenced by the size of the company. Nevertheless, we can observe for small enterprises that a negative change in productivity is even rarer. We can guess that people in small businesses don't depend too much on working with peers and with teams, so going to the workplace is less necessary than in big companies and communications facilitated.

```
Unable to display output for mime type(s): text/html
```

We wanted to look at the distribution of the managing positions based on the gender and we could see that almost 60% of men are managers compared to only 40% for women so this is factored to take into consideration when giving conclusions.

Correlation/Multivariate Analysis

```
Unable to display output for mime type(s): text/html
```

```
Unable to display output for mime type(s): text/html
```

can see on those 2 pie charts representing the distribution of the day based on the time in hours working, commuting, for personnal and family, and domestic. We can see that the time of work is sensibly the same for remote and workplace (50% compared to 48.8%). What is important here is to see that time to commute is divided by 3 when remote working so it gives more time for domesting and personnal/family time going from 39.6% of your day to 46%. This difference shows a better work/life balance for employees when remotely working. This could affect mental health positively.

Key Findings

Univariate Analysis - The sample is professionally experienced: only 11% have experience of less than 1 year. - There is a correct balance between genders: 44% of the sample is composed by women. - Productivity is either not affected or positively impacted by remote work: 46% improved their productivity through remote work.

Bivariate Analysis Working on place or remotely does not affect small businesses productivity, from which we can conclude small-structure workers does not depend too much on working with peers and teams, and going physically to workplace is less necessary than in big companies and communications facilitated.

Manager jobs are composed of men at a 60% rate whereas only 40% are women, a factor we must take into consideration when giving conclusions.

Correlation/Multivariate Analysis According to the distribution of the day based on the time in hours working, commuting, for personal and family, and domestic, we can observe that the time of work is sensibly the same remotely and physically at work (50% compared to 48.8%). Furthermore, time to commute is divided by 3 when working remotely, which gives more time for domestic activities and personal/family time going from 39.6% of daytime to 46%. This difference shows a better work/life balance for employees through remote work, an element that could positively affect mental health.
There is a positive correlation of 0.14 between productivity and working remotely. However, this correlation coefficient remains positively low.

Findings and Discussion

Perceived productivity (Productivity_remote_vs_workplace)

The distribution remains wide but not excessively dispersed. Both positive and negative self-reported productivity changes appear, with no clear dominant direction.

Remote work share (Remote_pct_*)

Clear multimodal distribution: - a large cluster at 0 (never remote), - a midpoint cluster around 0.5 (hybrid), - and a peak at 1.0 (fully remote).

The regression plot shows a weak positive trend between remote work percentage and perceived productivity.

Remote_pct_*: meaning all the columns that start with Remote_pct_ – in our context, remote_pct_last_year and remote_pct_last_3_months variables.

Hours commute / work / personal / domestic

These variables exhibit high variability, and under the classical IQR test, some show large numbers of outliers (Working_hours_workplace: 271, Commute_hours_remote: 142).

Commute_hours_remote is essentially zero for most respondents, whereas workplace commute shows very high values for a subgroup.

Satisfaction / Preferences (Preferred_remote_*)

A significant proportion of respondents indicate a preference for higher levels of remote work — suggesting overall satisfaction with remote arrangements.

Demographics

Age is concentrated within typical working-age ranges but includes two implausible values around 120, considered invalid extremes under both IQR methods. Gender distribution shows moderate imbalance.

Outliers & cleaning (3×IQR method)

Outlier counts highlight substantial structural variability in the dataset. Using the classical 1.5×IQR rule would remove more than 40% of the sample, which risks discarding legitimate behavioral diversity. Using the 3×IQR rule instead focuses only on filtering clear data errors or implausible extreme values (ages >100, commute hours >10, negative preferred remote percentages). With this more cautious threshold and a tolerance of zero outliers per row, the dataset is reduced from 1370 to 1196 rows, meaning only obviously invalid extreme records were removed. This approach preserves representativity while still improving data quality.

Interpretation (per finding)

Perceived productivity Histograms and boxplots show the distribution of Productivity_remote_vs_workplace is still wide after cleaning.

The regression line suggests a slight positive correlation between remote work percentage and productivity. The mean appears small relative to the variance, indicating high heterogeneity in productivity experiences. Some individuals report substantial productivity gains, others substantial losses.

Through these observations, we directly answer this question: "How does remote work influence employees' overall productivity?", and the answer is: effects differ strongly across individuals; no universal impact exists. It is consistent with Bloom et al. (2015) and COVID-era studies showing mixed or context-dependent productivity outcomes.

The limitations are composed of biases and complexity-lacking interpretation such as Self-reported measure bias and univariate-only interpretation.

Thus, organisations should adopt flexible and individualized policies, and further multivariate analysis should identify moderators (job type, household structure, management role…).

Job satisfaction / preferences

Preferred_remote_* shows many respondents preferring more remote work. Preference likely reflects higher satisfaction with remote work conditions relative to on-site conditions.

It directly informs remote work increases job satisfaction for many workers and addresses: “Has job satisfaction increased or decreased due to remote working?”: univariate evidence points to increased satisfaction for many workers. Preference does not equal measured job satisfaction; there is a potential selection bias. Hybrid working models and flexible remote-work options should be explored.

Social connection / collaboration Likert variables for collaboration (Collaboration_remote_*) are centered around neutral values (Likert value: 3). At an aggregate level, respondents do not report a clear decline or improvement in collaboration while working remotely.

It addresses the question: “How does remote working affect employees’ ability to maintain social connections and avoid isolation?”. Digital collaboration tools may offset loss of informal interactions and individual variability likely remains. Likert responses are proxies; no direct psychological isolation is measured.

However, a preference-based bias might be taken into consideration: as remote work seems more desirable for most respondents, they might answer in a way that avoids remote work discreditation. Organizations should focus interventions on subgroups reporting lower collaboration.

Structural factors (commute and personal time) Remote commute is almost zero; remote work yields increase in personal, family, and domestic hours. Savings in commute time and reallocation toward personal activities are plausible drivers of increased satisfaction with remote work. We can extrapolate this job and free-time balance to an enhanced mental health and less stressed psychology.

To “Which factors contribute most to remote-work satisfaction?”, we can thus conclude that Time savings and increased autonomy are strong candidates, which aligns with research showing commute reduction increases well-being and perceived control over time. However, the trends are univariate only; further multivariate modeling might be needed. Policies should preserve gains in personal time and support work-life balance.

Overall answers to research questions

Productivity

Remote work has a mixed and highly heterogeneous effect on productivity; the average effect is not clearly positive or negative.

Social connection / isolation

Aggregate responses are neutral and no evidence of widespread isolation or deterioration of collaboration is detected.

Job satisfaction

Preference patterns imply that for most workers, job satisfaction has increased under remote arrangements.

Drivers of satisfaction

The strongest univariate candidates are: - reduced commute, - increased personal/family time, - greater schedule flexibility.

Analysis

Methods

Remote Work Intensity and Productivity

This section examines the relationship between remote work intensity and self-reported productivity change using both graphical analysis and linear regression models.

Exploratory relationship

Figure X plots productivity change against the share of time spent working remotely, together with a fitted linear regression line. The slope of the line is positive, and the Pearson correlation coefficient is approximately 0.14, indicating a weak positive association between remote work intensity and productivity change. Individuals who work remotely more frequently tend, on average, to report slightly higher productivity. However, the magnitude of this association is limited and should not be interpreted as causal.

Baseline regression model

To formalise this relationship, productivity change is first modelled using a simple ordinary least squares (OLS) regression with remote work intensity as the only explanatory variable:

$$\text{Productivity}_i = \beta_0 + \beta_1 \, \text{RemoteShare}_i + \varepsilon_i$$

```
Unable to display output for mime type(s): text/html
```

This baseline model captures the raw relationship between remote work intensity and productivity without accounting for any additional individual or organisational characteristics. The estimation results show that the coefficient on remote work intensity is positive and statistically significant. Moving from no remote work to full remote work is associated with an increase of approximately 11–12 percentage points in reported productivity. Despite this statistically significant effect, the explanatory power of the model remains very low, with an R-squared of around 0.02. This indicates that remote work intensity alone explains only a very small share of the variation in productivity changes across individuals.

Productivity by remote work intensity groups

To complement the continuous analysis, productivity changes are also examined by grouping individuals into remote work intensity categories (0–20%, 20–50%, 50–80%, and 80–100%). This binned representation allows for a clearer comparison of productivity distributions across levels of remote work.

The results show that median productivity gains increase once remote work exceeds 20% of working time and stabilise around +20% for medium to high levels of remote work intensity. Mean productivity changes follow a similar pattern, rising from approximately +10% in the lowest group to nearly +20% in the highest group.

However, the distributions overlap substantially across groups, with wide dispersion in all categories and both positive and negative productivity changes observed at every level of remote work intensity. This highlights considerable individual heterogeneity and reinforces the conclusion that remote work intensity alone is insufficient to explain productivity outcomes.

```
Unable to display output for mime type(s): text/html
```

Extended regression model with controls

The analysis is then extended by including additional explanatory variables that capture individual and organisational characteristics:

$$\text{Productivity}_i = \beta_0 + \beta_1\,\text{RemoteShare}_i + \beta_2\,\text{Age}_i + \beta_3\,\text{OrganisationSize}_i + \beta_4\,\text{ManagingPosition}_i + \beta_5\,\text{OrgEncourage}$$

These variables were selected based on exploratory analysis and correlation checks, as they showed meaningful associations with productivity change and are directly related to working conditions under remote work. After including these controls, the coefficient on remote work intensity remains positive and statistically significant, although slightly smaller, at around 9.6. Importantly, the explanatory power of the model improves: the R-squared increases from approximately 0.02 in the baseline model to about 0.05 in the extended specification. This increase indicates that organisational context and collaboration quality account for an additional share of the variation in productivity changes. Among the control variables, perceived collaboration quality when working remotely shows a strong positive association with productivity, while organisational encouragement of remote work is negatively associated. Age and managing position do not exhibit statistically significant effects once other factors are controlled for.

Interpretation

Overall, the results convey a consistent message across graphical analysis and regression models. Remote work intensity is positively associated with productivity change, but the relationship is weak in magnitude and explains only a limited fraction of productivity differences. The increase in R-squared when adding organisational and collaboration-related variables confirms that productivity outcomes under remote work are influenced by multiple factors. While remote work intensity plays a role, individual heterogeneity and organisational conditions appear to be more important drivers of productivity changes.

Conclusion

Summary

In this project, we analyzed how remote work affects employees' productivity, social well-being, and job satisfaction using the Remote Working Survey (2020), which includes 1,507 respondents. We investigated our four research questions using descriptive statistics, visualizations, and correlation analysis after cleaning the dataset and translating qualitative responses into numerical values. Our results show that the effect of remote work on productivity is highly heterogeneous. While many respondents reported no change, the productivity distribution displayed a slight negative skew (–0.13), meaning productivity gains were more common than losses. Productivity did not vary significantly by organisation size, and the correlation between remote-work percentage and productivity change was positive but weak (r = 0.14), suggesting that working remotely more often is associated with a small increase in perceived productivity.

The clearest pattern we found was the improvement in work–life balance. Remote work nearly eliminated commuting time, and our comparison showed that personal and family time increased from 39.6% to 46% of the day on average when employees worked from home. This extra time probably improved mental health and could be responsible for the overall high desire for remote work. Indicators of collaboration and connection on the social side were centered around neutral responses, showing that employees' feeling of social connection was neither greatly improved nor harmed by distant work. Overall, our research indicates that working remotely can boost well-being and productivity, mainly through time savings and better daily time management, but its effects on social interaction seem to be less clear. These findings help us better understand how employees' lives are shaped by remote work and directly address the objectives of our project.

Limitations

Although our project provides meaningful insights, it also has several limitations that should be acknowledged. First, the dataset comes from a single region (New South Wales, Australia) and only reflects the year 2020, when remote work was still relatively new due to the COVID-19 pandemic. This context may not represent long-term or post-pandemic habits, which limits how generalizable our findings are to other geographic areas, industries, or years. The data also contained several extreme outliers (for example, two respondents aged 120), as well as categories with very small sample sizes such as “rather not say” for gender. We removed these categories to avoid inconsistent comparisons, but doing so slightly reduces the diversity of our sample.

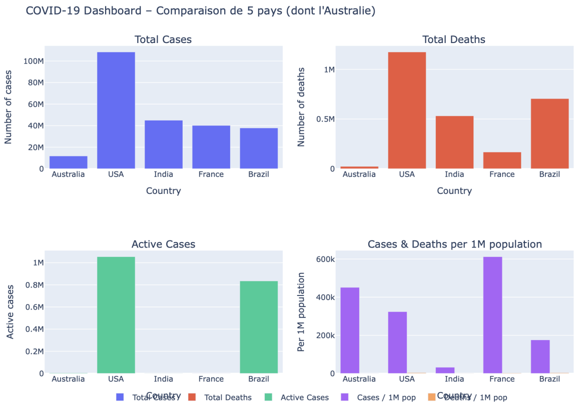
Comparative Analysis of the Impact of COVID-19 (United States, India, France, Brazil, Australia)

The comparative analysis highlights substantial disparities in the evolution and impact of the COVID-19 pandemic across the five selected countries. The United States records the highest number of total cases, followed by India, France, and Brazil, whereas Australia clearly stands out with a significantly lower number of reported infections. With respect to mortality, a similar pattern is observed. The United States and Brazil exhibit the highest death tolls among the sample, while France occupies an intermediate position. In contrast, Australia displays a markedly lower mortality level, indicating a comparatively limited health impact of the pandemic. The analysis of active cases, defined as the number of individuals currently infected (total cases minus recoveries and deaths), reveals that active infections remained relatively high in the United States and Brazil at the time of data collection.

Interpretation of the Australian Case

Based on our interpretation of the graphical results, Australia appears to have been relatively less affected by the COVID-19 pandemic, particularly in terms of mortality and active infections. Several structural and geographical factors may help explain this outcome. First, Australia has a comparatively younger population, and younger age groups were statistically less exposed to severe forms of the disease. Furthermore, the country is characterized by a large territory combined with a low population density, which naturally limits close interpersonal contact and reduces the speed of viral transmission. In addition, Australia’s geographical isolation likely delayed the initial introduction of the virus and facilitated the implementation of strict border control measures, which played a critical role during the early stages of the pandemic. Finally, early public health interventions and the overall efficiency of the healthcare system may also have contributed to limiting severe cases and fatalities.

From a methodological standpoint, our study mostly uses correlation analysis and descriptive statistics, which help in finding patterns but cannot allow us to prove causation. For example, even if we discovered a slight positive correlation ($r = 0.14$) between the percentage of remote work and productivity change, this does not show that remote work raises productivity. These results could be impacted by additional unobserved factors including home environment, digital technologies, management assistance, or individual preferences. In addition, a number of important factors such as productivity, teamwork, and satisfaction are self-reported, which may create bias because workers may exaggerate or understate their experiences.



(a) comparison

Figure 1: The emotional journey of a data scientist debugging their code

Future Work

- **Extend dataset:** "Include the 2021 wave of the Remote Working Survey and perform longitudinal comparisons to track changes over time."
- **Enhance analytical approach:** "Use advanced statistical models like multiple regression to identify key predictors of productivity and satisfaction."
- **Broaden contextual factors:** "Add variables on digital tools, communication habits, and mental health indicators for richer insights."
- **Integrate mixed data types:** "Combine quantitative survey data with qualitative responses to capture nuanced experiences of remote work in other regions not only Australia."

References

- NSW Government. (2020). Remote Working Survey 2020 Dataset. Retrieved from https://data.nsw.gov.au/data/dataset/nsw-remote-working-survey (Used to describe data source, survey methodology, and metadata.)
- Pandas Documentation. (n.d.). Pandas User Guide. Retrieved from https://pandas.pydata.org/docs/ (Referenced for operations such as rename(), replace(), map(), astype(), errors="coerce", value_counts(), info().)
- Seaborn Documentation. (n.d.). Seaborn: Statistical Data Visualization. Retrieved from https://seaborn.pydata.org/ Matplotlib Documentation. (n.d.). Matplotlib: Visualization with Python. Retrieved from https://matplotlib.org/stable/ (Used for histograms, boxplots, and univariate analysis.)
- Quarto Documentation. (n.d.). Quarto Guide. Retrieved from https://quarto.org/ (Referenced for callouts, section structure, code blocks, and PDF/HTML formatting best

practices.)

- Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert Scale: What it is & How to Use It. Retrieved from https://www.researchgate.net/publication/281874183_Likert_Scale_What_it_is_and_How_to_Use_It (Implicitly used to justify converting textual modalities to numeric values 1–5.)
- Charalampous, M., Grant, C. A., Tramontano, C., & Michailidis, E. (2022). 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, 19(19), 12373. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC9566387/
- García-Sánchez, E., & García-Sánchez, I. M. (2024). Remote workers' well-being: Are innovative organizations really concerned? A bibliometrics analysis. Journal of Innovation & Knowledge, 9(4), 100313. Retrieved from https://www.sciencedirect.com/science/article/pii/S2444569X24001343
- Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2013). Does Working from Home Work? Evidence from a Chinese Experiment. National Bureau of Economic Research Working Paper No. 18871. Retrieved from https://www.nber.org/papers/w18871