technische universität dortmund

Finance, Risk & Resource Management

The Hybrid Work Premium:
Do remote-friendly firms outperform across sectors?

Project Contributors:

Shriya Madan – Data Science – 268601

Jayasheelan Ravichandran – Data Science – 270257

Nikhil Amala Jerrin James Edward Suresh – Automation & Robotics - 235642

Content

- Objective
- Literature Survey
- Data Preparation
- Methodologies
- Causal Support to the Hypothesis
- Conclusion
- References

Motivation & Research Question

Context:
The rise of hybrid and remote work models in the post-COVID era has reshaped how companies operate, especially in knowledge-driven industries.

Objective:
To evaluate whether workplace flexibility—through hybrid or remote models—has a measurable impact on company performance and stock prices.

Core Question:
Does workplace flexibility create an economic advantage for companies?

Hypothesis (H1):
Companies with a higher proportion of hybrid/remote workers (Hybrid_%) demonstrate greater revenue per employee, particularly in sectors such as Information Technology and Finance.

Literature Survey

- **Employee Expectations Shifted**
  Post-COVID, 77% of employees want more flexibility (Deloitte). 88% prefer hybrid or fully remote work (Harvard Business School).

- **Leadership Resistance**
  70% of CEOs plan full office return (LaSalle Network). Goldman Sachs CEO called remote work an "aberration" (LaSalle).
  78% of HR leaders cite leadership mindset as the key barrier—not tech limitations (Gartner).

- **Productivity & Technology Gains**
  Hybrid work may boost U.S. productivity by 5% (University of Chicago).
  Tech adoption post-pandemic helps counteract productivity stagnation (Bloomberg).

Literature Survey

- **Retention & Satisfaction**
  Flexibility improves performance and retention (Gartner).
  Satisfied employees outperform by 20%; dissatisfaction may cost 1/3 of salary in turnover (Gartner, [7]).

- **Cost & Talent Benefits**
  Hybrid models lower costs (office, travel, perks) and enable global talent recruitment (HBR).
  Culture is sustained via digital tools, not just physical presence (Prodoscore).

FiRRM

Data Preparation

Data Sources

Alpha Vantage

Financial & market metrics per quarter:

Revenue (€M), Operating Cost (€M), Net Profit (€M)
R&D Spending (€M), Stock Price (€M), Market Cap (€B)

Coresignal APIs

Workforce & HR metrics:

Employees, Attrition %, Hiring %, Tenure (Years)
Work-Life Balance (1-5), Job Satisfaction (1-5)

Public Sources

Workforce distribution (by quarter):

Hybrid %, Onsite %, Remote %

Sourced from company websites, official reports, news articles, and Google search results.

Data Preparation

Key Preparation Steps

Developed and used **Python scripts** to:

- Fetch financial and employee data using **API keys** from Alpha Vantage and Coresignal
- Automate data retrieval, reduce manual errors, and ensure consistency

Collected **workforce distribution data** (*Hybrid %, Onsite %, Remote %*) manually:

- Sourced from **company websites**, **official reports**, **news articles**, and **Google search results**

Performed data processing:

- Cleaned missing, incomplete, or inconsistent values
- Standardized formats for **financial metrics (€M, €B)**, **percentages (0-100%)**, and **ratings (1-5)**
- Merged all data sources by **Quarter** and **Company**

Stored the final cleaned and integrated dataset as a **CSV file** for further analysis

Dataset Overview

Data Structure

- **Firm-level panel data** covering multiple companies across:
  - *IT*, *Finance*, *Health*, *Construction*, *Automobile* sectors
- **Period**: Pre-COVID and post-COVID quarters (2017 - 2024)
- **Format**: Integrated dataset saved as CSV for analysis

Key Variables

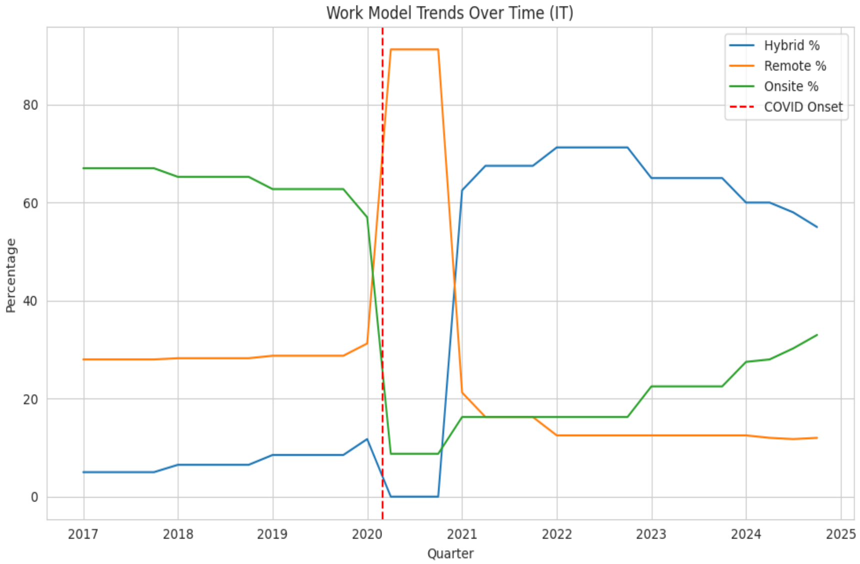
- **Financial & Market Data** *(Alpha Vantage)*
  - *Revenue (€M), Operating Cost (€M), Net Profit (€M)*
  - *R&D Spending (€M), Stock Price (€M), Market Cap (€B)*
- **Workforce & HR Data** *(Coresignal + Public sources)*
  - *Employees, Attrition %, Hiring %, Tenure (Years)*
  - *Work-Life Balance (1-5), Job Satisfaction (1-5)*
  - *Hybrid %, Onsite %, Remote %*

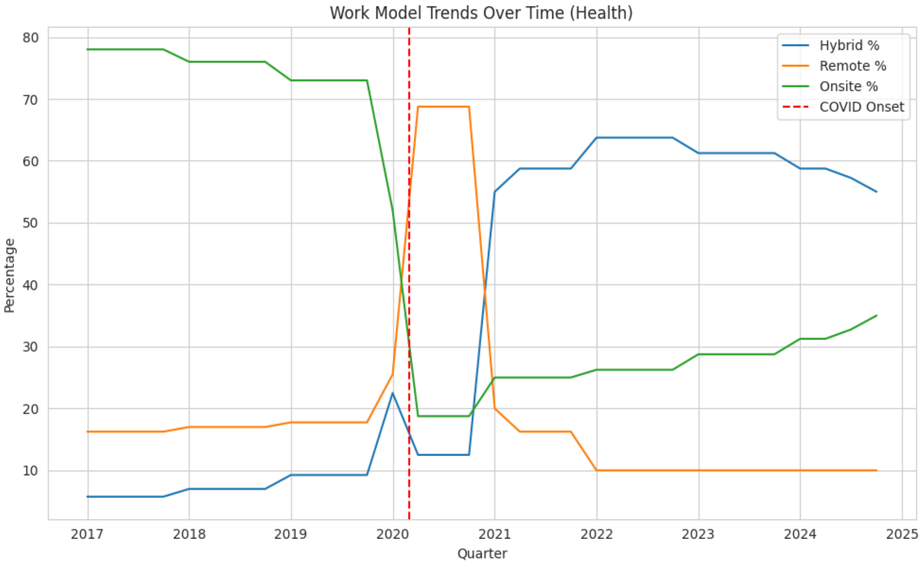
Data Volume

- **20 companies**
- **4 quarters** covered (2017 – 2024)
- ~640 **records** in final CSV file (fill in actual number of rows)

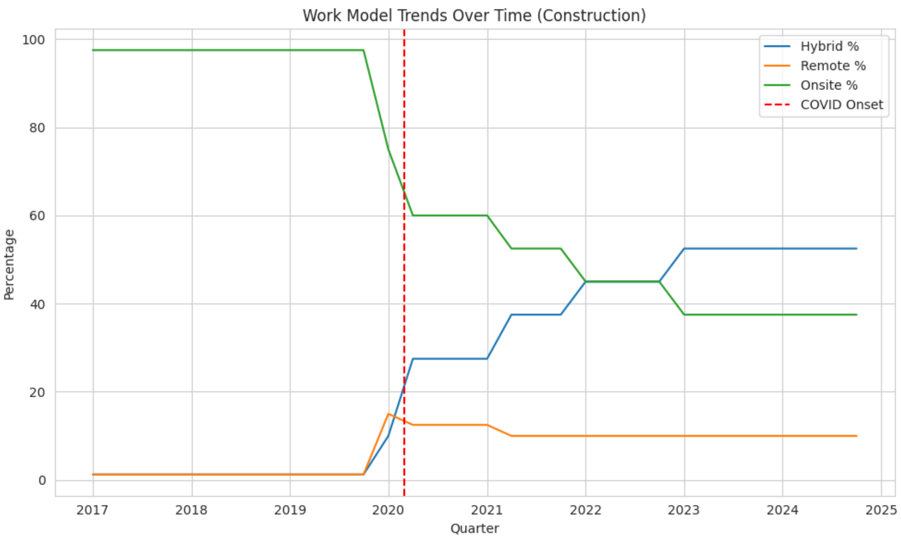
Work Model Trends by Sector

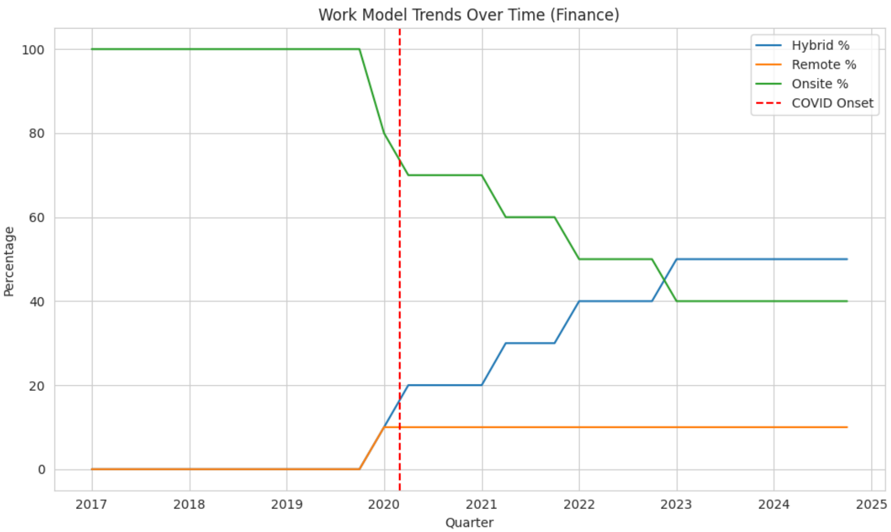
- Each sector responded differently to the pandemic shock.
- IT and Finance rapidly shifted to hybrid post-COVID, while Automobile and Construction stayed mostly onsite.
- These divergent patterns motivate our sector-wise modeling in subsequent modules.

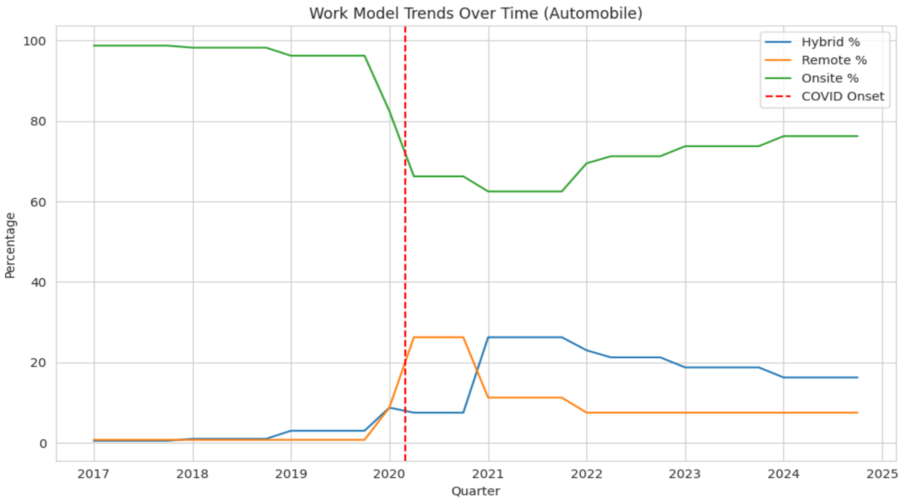




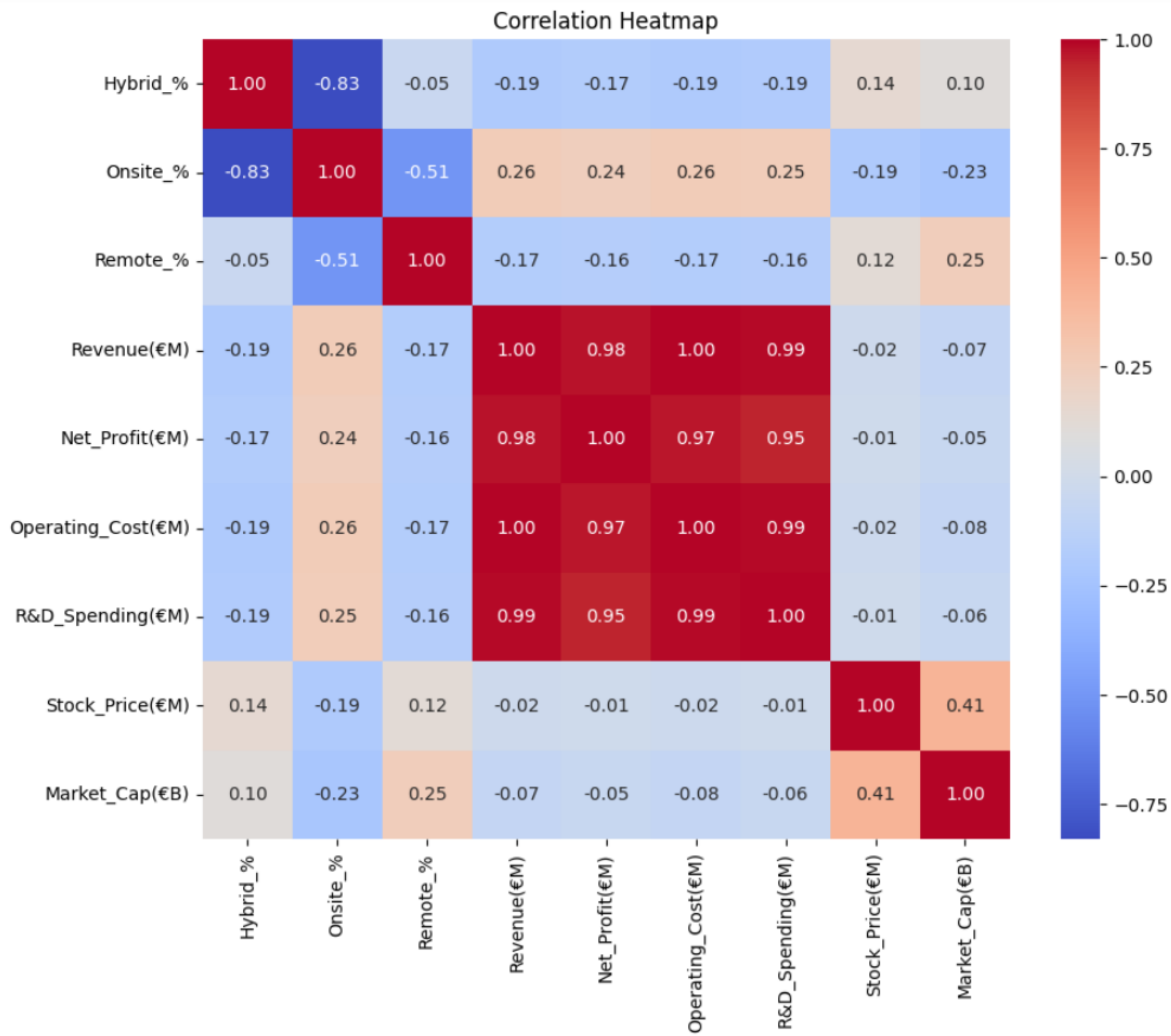
Work Model Trends by Sector







Correlation Matrix



Methodologies

1. Regression Analysis & Productivity Modeling
- OLS to assess effect of work models on **Profit, Revenue, Productivity**
- Sector-wise and economy-wide regressions
- Profit/Revenue per employee as dependent variables
- Measure efficiency impact of work models
- Workforce Cost Risk Index

2. Correlation & Clustering Analysis (KMeans + PCA)
- Track sector-wise changes in Hybrid/Remote/Onsite % over time
- Group firms based on work model and financial profile
- Reveal strategic firm types (e.g., remote-heavy, hybrid-high)

3. Causal Inference
- Difference-in-Differences (DiD):
  Compare pre vs. post-COVID effects of work models on profit
  ➤ Use interaction terms like Hybrid_% x PostCOVID
- Granger Causality Tests:
  Check if changes in work model percentages cause financial outcomes (time series)
- Transfer Entropy:
  Validate causal directionality, especially in nonlinear settings

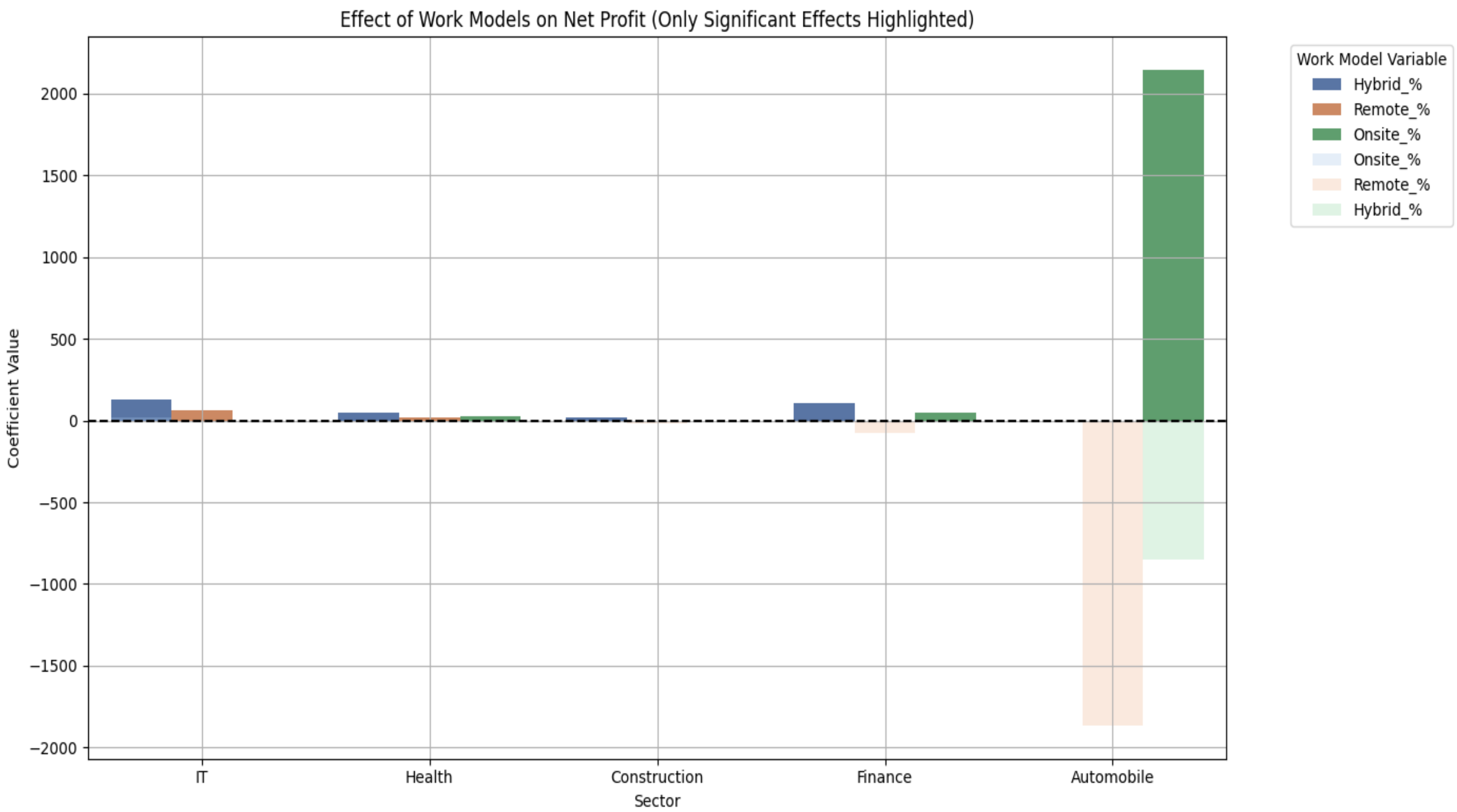
4. Stock Price Prediction (ML Models)
- Random Forest & XGBoost to predict Stock Price(€)

FiRRM

1.1 Impact of Work Models on Firm Financials: Net Profit

Hybrid_% improves Net Profit significantly in IT, Finance.

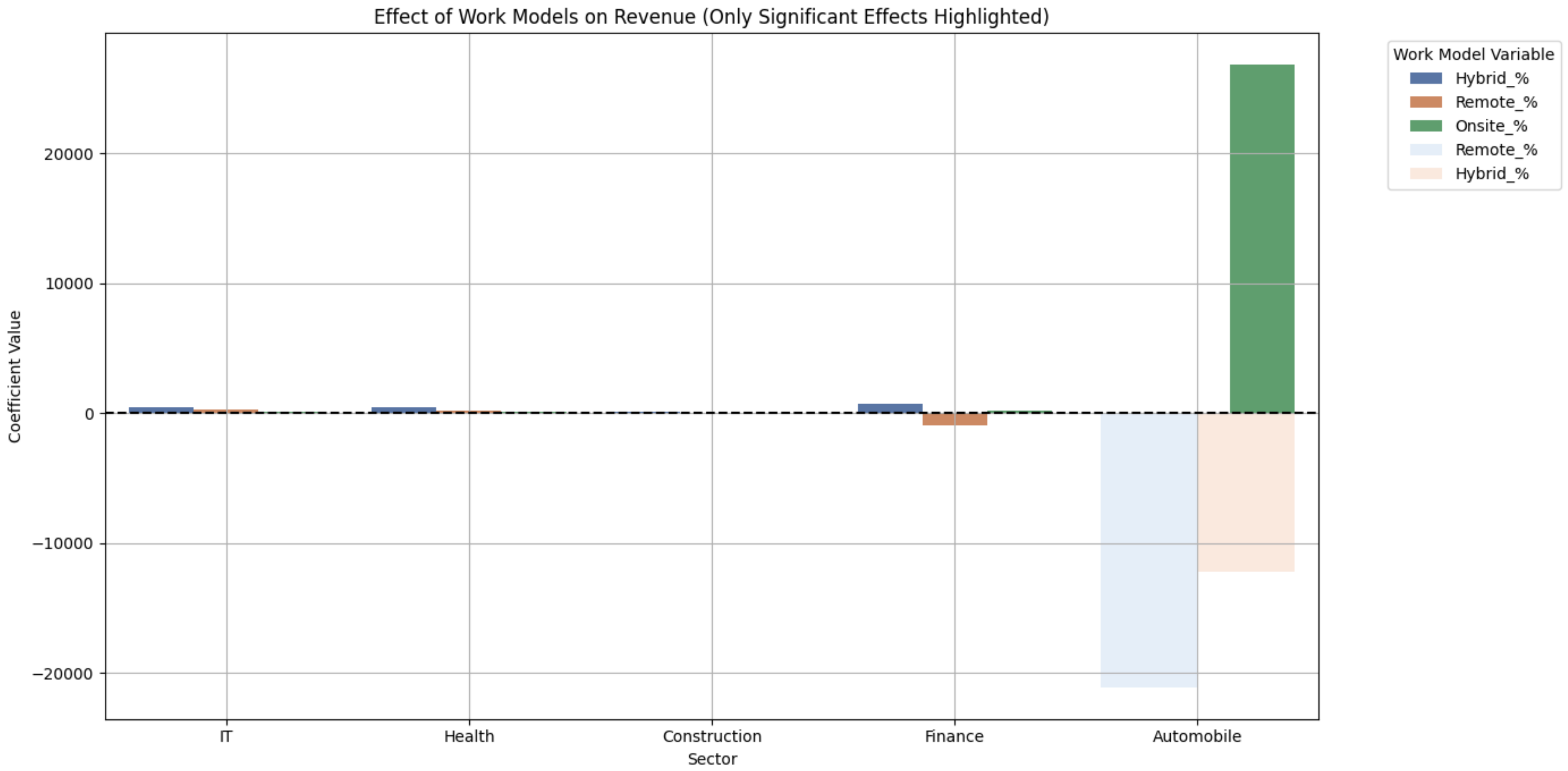
- Remote_% is negative in Finance and neutral in other sectors.
- Onsite_% positively impacts profit in traditional sectors like Automobile.



FiRRM

Impact of Work Models on Firm Financials: Revenue

- Revenue impact aligns with business model flexibility.
- Hybrid_% boosts revenue in IT; negative in Automobile.
- Onsite_% becomes crucial for revenue in post-COVID Automobile sector.



1.2 OLS Regression – Productivity Modeling

- All work models significantly increase revenue per employee.
- **Hybrid_%** has the strongest positive effect.
- Suggests flexible work boosts top-line productivity.

```
                            OLS Regression Results
================================================================================
Dep. Variable:     Revenue_per_Employee   R-squared:                       0.084
Model:                              OLS   Adj. R-squared:                 -0.003
Method:                   Least Squares   F-statistic:                    0.9626
Date:                  Fri, 13 Jun 2025   Prob (F-statistic):              0.398
Time:                          19:03:42   Log-Likelihood:                -236.76
No. Observations:                    24   AIC:                             479.5
Df Residuals:                        21   BIC:                             483.1
Df Model:                             2
Covariance Type:              nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         13.9864      0.308     45.385      0.000      13.346      14.627
Hybrid_%     486.5485     20.064     24.250      0.000     444.824     528.273
Remote_%     483.3741     31.016     15.585      0.000     418.872     547.876
Onsite_%     428.7173     33.223     12.904      0.000     359.626     497.809
==============================================================================
Omnibus:                        0.863   Durbin-Watson:                   0.757
Prob(Omnibus):                  0.650   Jarque-Bera (JB):                0.835
Skew:                          -0.387   Prob(JB):                        0.659
Kurtosis:                       2.514   Cond. No.                     3.11e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 8.5e-31. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

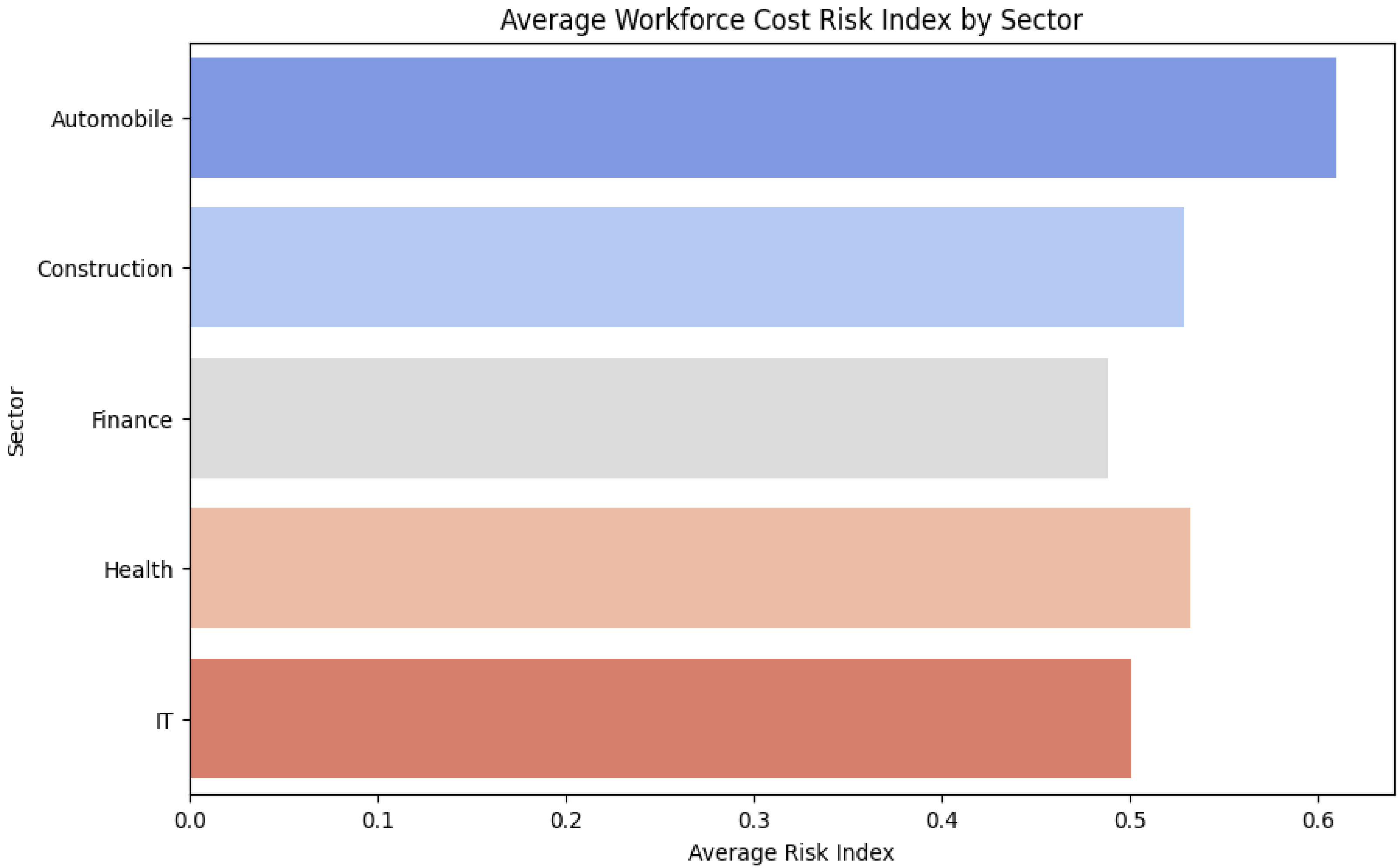
1.2 OLS Regression – Productivity Modeling

- All models show positive, significant profit impact.
- **Onsite_%** has the highest coefficient.
- Hybrid/Remote remain financially effective.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     Profit_per_Employee   R-squared:                       0.322
Model:                             OLS   Adj. R-squared:                  0.257
Method:                  Least Squares   F-statistic:                     4.983
Date:                 Fri, 13 Jun 2025   Prob (F-statistic):             0.0169
Time:                         19:03:52   Log-Likelihood:                -207.76
No. Observations:                   24   AIC:                             421.5
Df Residuals:                       21   BIC:                             425.1
Df Model:                            2
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.4047      0.092     15.258      0.000       1.213       1.596
Hybrid_%      33.6850      5.994      5.620      0.000      21.220      46.150
Remote_%      33.6932      9.266      3.636      0.002      14.424      52.962
Onsite_%      73.0868      9.925      7.364      0.000      52.447      93.727
==============================================================================
Omnibus:                        1.821   Durbin-Watson:                   0.866
Prob(Omnibus):                  0.402   Jarque-Bera (JB):                1.590
Skew:                          -0.571   Prob(JB):                        0.452
Kurtosis:                       2.468   Cond. No.                     3.11e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 8.5e-31. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

1.3 Workforce Cost Risk Index



Workforce Cost Risk Index

- To Calculate Risk Index

$$\text{Workforce Cost Risk Index} = \frac{\text{Attrition\%}}{\text{Tenure\_Yrs}} \times \frac{5 - \text{Job\_Satisfaction}}{5}$$

- The bar plot shows the **average workforce cost risk** across sectors.
- **Higher bars** mean higher risk from attrition, tenure, and job satisfaction factors.
- **Construction, Health, and IT** sectors have noticeably higher risk than Finance.
- Companies with **shorter tenure and lower job satisfaction** face greater risk.
- **Company size doesn't reduce risk**—even large firms experience high workforce cost challenges.
- High risk indicates **hidden costs from turnover, hiring, and lost productivity**.

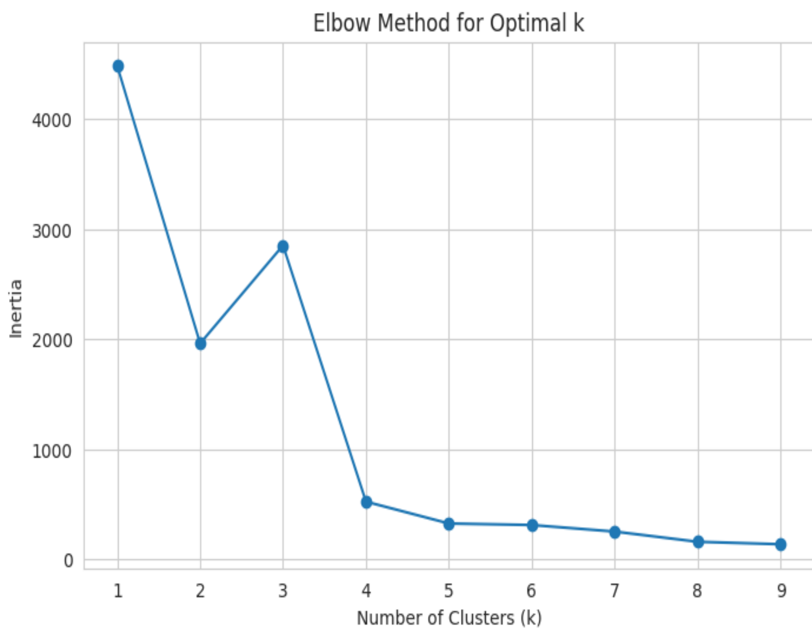
2.1 Clustering Analysis of Work Models & Company Profiles

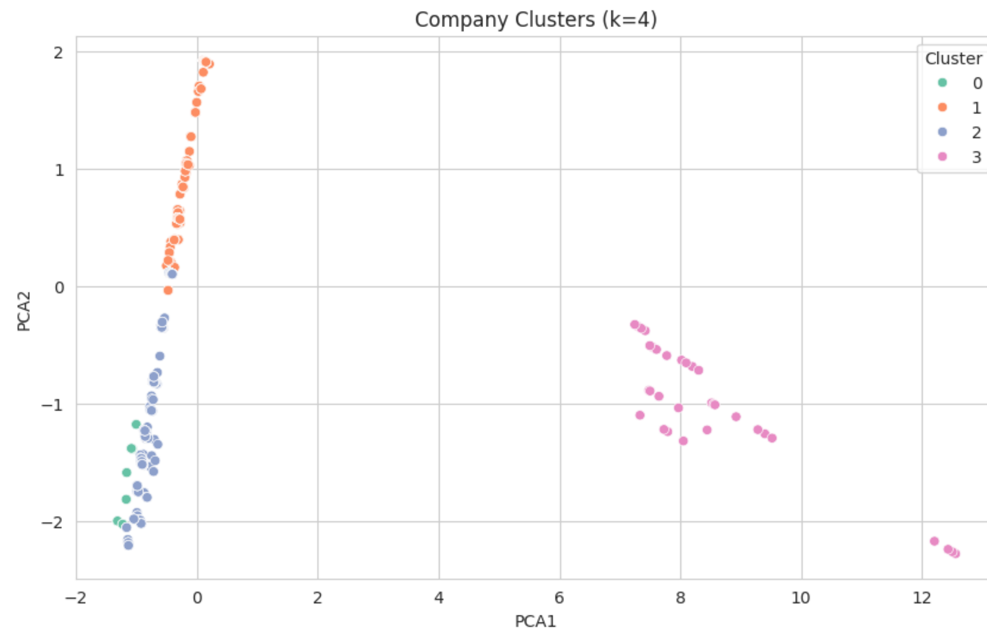
Objective: Group companies based on work model usage and financial indicators to uncover strategic patterns.

Methods Used

- **KMeans Clustering** on:
  - Work model variables: Hybrid_%, Remote_%, Onsite_%
  - Financial metrics: Revenue, Profit, CapEx, R&D
- **PCA (Principal Component Analysis)** for 2D visualization
- **Elbow Method** to choose optimal clusters: **k = 4**

Clustering Analysis of Work Models & Company Profiles





Cluster	Description
Cluster 0	Fully remote companies: 85% Remote, 0% Hybrid, 15% Onsite. High Revenue (€21B), strong Profit (€3.6B), relatively high R&D (€2.9B). Likely large IT/digital companies.
Cluster 1	Traditional Onsite firms: 85% Onsite, very low Hybrid & Remote. Lower Revenue (€17B), Profit (€2.4B), low R&D. Possibly Finance, Construction, or Manufacturing firms that resist hybrid models.
Cluster 2	Strong Hybrid companies: 52% Hybrid, 12% Remote, 35% Onsite. Highest average Revenue (€27B) and Profit (€4.7B). Likely top-performing hybrid companies — IT, Health, and some Finance.
Cluster 3	Extremely large industrial companies: 94% Onsite, almost no Remote/Hybrid. Huge Revenue (€7.4 trillion), Profit (€596B), CapEx extremely high (€946B), R&D also enormous (€256B). Likely Automobile or mega-manufacturing sector.

3.1 Causal Impact of Work Models (Difference-in-Differences)

Objective

- To check **how the impact of work models (Hybrid, Remote, Onsite)** on company profit **changed after COVID**, using **Difference-in-Differences**.

Method

- OLS regression with interaction terms (e.g., Hybrid_% × PostCOVID)
- Captures pre vs. post-COVID performance impact of each work model

Insights

- Hybrid and Remote effects do not show any statistically significant shift after COVID in this model.
- Onsite work became even more important for profit post-COVID.
- This makes sense especially for sectors like Automobile, Construction etc.
- Hybrid/remote likely vary too much across sectors — your earlier sector-wise models captured these differences better.

Causal Impact of Work Models (Difference-in-Differences)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            Net_Profit(€M)   R-squared:                       0.163
Model:                               OLS   Adj. R-squared:                  0.156
Method:                    Least Squares   F-statistic:                     24.68
Date:                   Fri, 13 Jun 2025   Prob (F-statistic):           9.37e-23
Time:                           19:12:45   Log-Likelihood:                -8425.9
No. Observations:                    640   AIC:                         1.686e+04
Df Residuals:                        634   BIC:                         1.689e+04
Df Model:                              5
Covariance Type:               nonrobust
================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const              -1.9673     14.366     -0.137      0.891     -30.178      26.244
Hybrid_%           620.9485   1962.483      0.316      0.752   -3232.805    4474.702
Remote_%         -1239.8463    850.917     -1.457      0.146   -2910.802     431.110
Onsite_%           422.1679    106.115      3.978      0.000     213.789     630.547
Hybrid_PostCOVID -1636.1788   1971.003     -0.830      0.407   -5506.663    2234.306
Remote_PostCOVID   650.1748    902.220      0.721      0.471   -1121.527    2421.876
Onsite_PostCOVID  1670.2749    201.802      8.277      0.000    1273.993    2066.557
==============================================================================
Omnibus:                     568.946   Durbin-Watson:                   0.080
Prob(Omnibus):                 0.000   Jarque-Bera (JB):            14671.355
Skew:                          3.982   Prob(JB):                         0.00
Kurtosis:                     25.062   Cond. No.                     2.19e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 8.06e-29. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

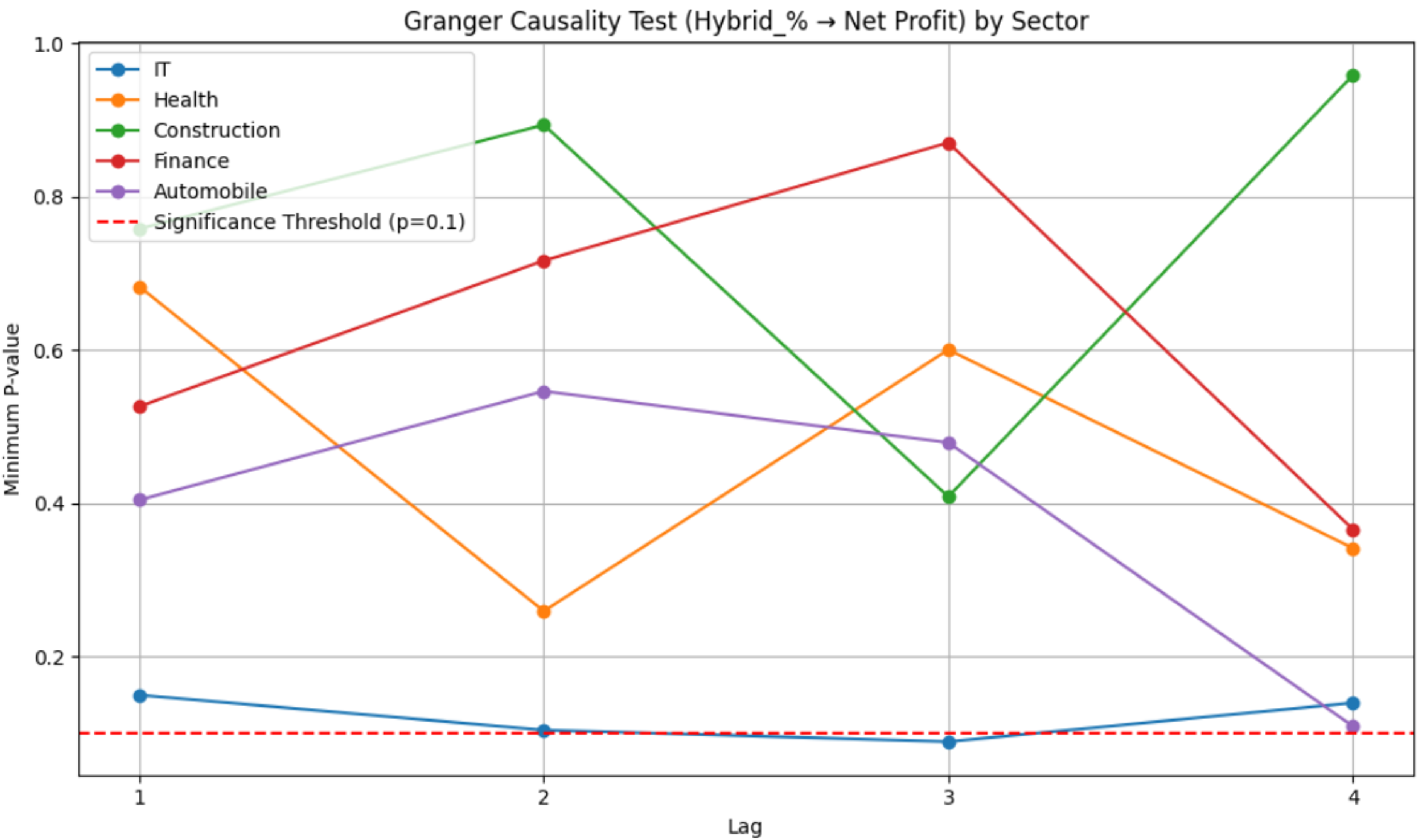
Variable	Coefficient	p-value	Interpretation
Hybrid_%	620.95	0.752	Not significant → hybrid share (pre-COVID) has no significant profit effect alone.
Remote_%	-1239.85	0.146	Not significant → remote share (pre-COVID) not significantly related.
Onsite_%	422.17	0.000	✅ Significant → onsite work consistently drives higher profit (pre-COVID).
Hybrid_PostCOVID	-1636.17	0.407	Not significant → no evidence that hybrid became better/worse after COVID.
Remote_PostCOVID	650.17	0.471	Not significant → no meaningful change for remote after COVID.
Onsite_PostCOVID	1670.27	0.000	✅ Significant → post-COVID onsite work shows stronger profit impact.

3.2 Advanced Causal Analysis - Granger Causality

- **Granger Causality Test**:
    - Overall, sector-wise, company-wise

Sector	Min p-values (Lag 1–4)	Interpretation
IT	0.15, 0.10, **0.089**, 0.14	Lag 3 shows causality (p=0.089)
Health	0.68, 0.26, 0.60, 0.34	No causality
Construction	0.76, 0.89, 0.41, 0.96	No causality
Finance	0.53, 0.72, 0.87, 0.37	No causality
Automobile	0.40, 0.55, 0.48, **0.11**	No causality

3.2 Advanced Causal Analysis - Granger



3.2 Advanced Causal Analysis - Granger Causality

- **In the IT sector**, there is **statistically significant evidence** that **past values of Hybrid_% (specifically with a 3-quarter lag)** Granger-cause Net Profit.
  This suggests that **increased or strategic hybrid work adoption** may lead to **improved profitability** in the IT sector after several quarters.

- For all other sectors — Health, Construction, Finance, and Automobile — **no significant causality** was found.
  This indicates that hybrid work trends in these sectors do **not have a measurable predictive impact on profits** within the time lags considered.

- **Conclusion:**
  - **Hybrid work positively influences profitability in the IT sector, but not in others**, possibly due to the IT sector's flexibility and lower reliance on physical presence.

3.2 Advanced Causal Analysis - Transfer Entropy

Transfer Entropy:

What is Transfer Entropy?

- Transfer Entropy measures **how much information** the past of one variable (here, Hybrid_%) provides about the **future** of another (here, Net Profit) — capturing both **linear and nonlinear dependencies**.

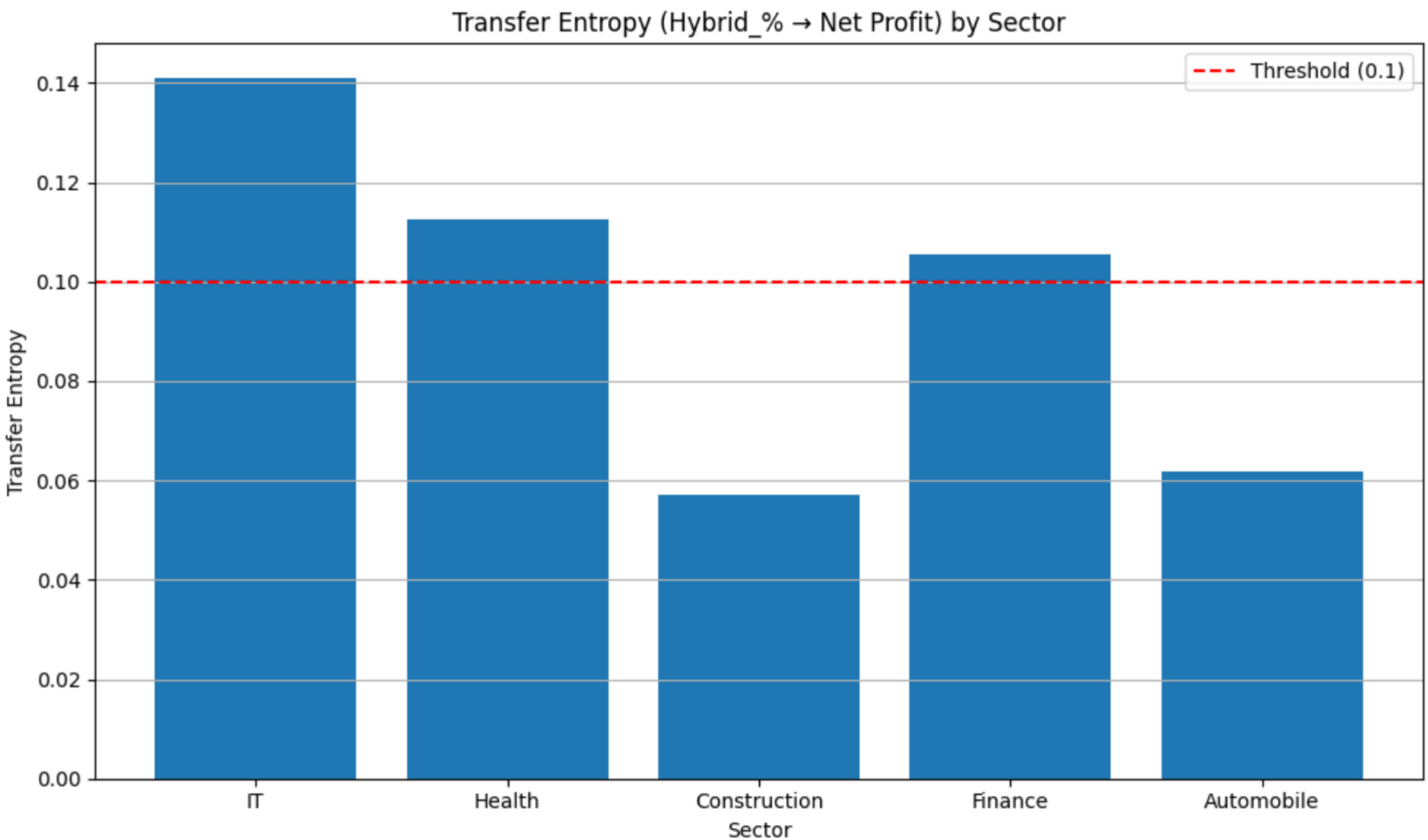
- TE = 0: No information transfer
- TE > 0: Some predictive influence
- TE > 0.1: Moderate or meaningful influence (interpretation depends on data discretization)

3.2 Advanced Causal Analysis - Transfer Entropy

Sector	Transfer Entropy	Interpretation
IT	0.141	Strongest influence
Health	0.113	Moderate influence
Construction	0.057	Weak influence
Finance	0.106	Moderate influence
Automobile	0.062	Weak influence

- **IT Sector** shows **clear evidence** that changes in hybrid work strategy carry **predictive information** about future profits.
- **Health and Finance sectors** show **moderate TE values**, indicating some influence — though weaker than IT.
- **Construction and Automobile** show **low TE**, suggesting hybrid work has **little to no effect on profit trends** in these sectors.

3.2 Advanced Causal Analysis - Transfer



4.1 Stock Price Prediction

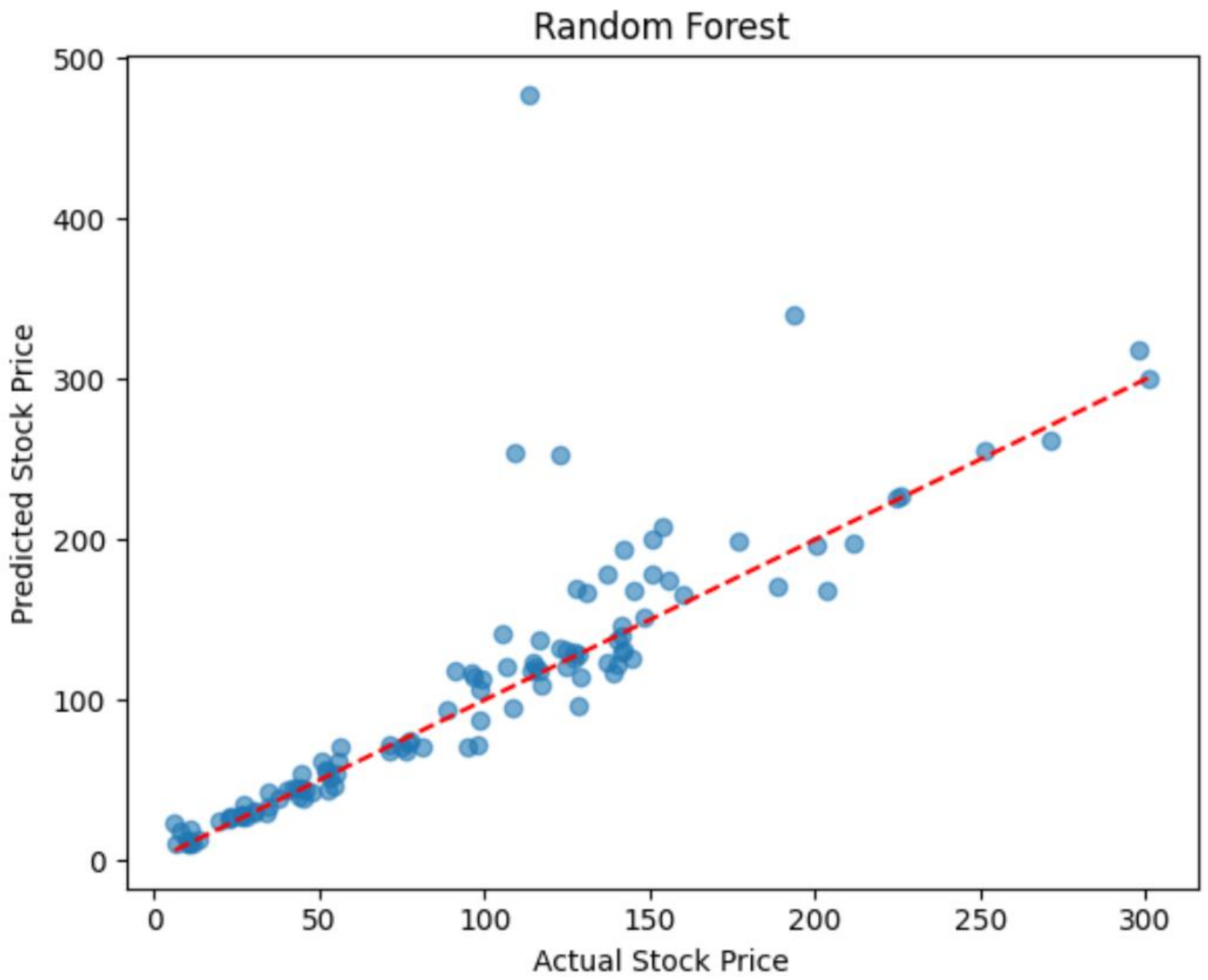
Objective:

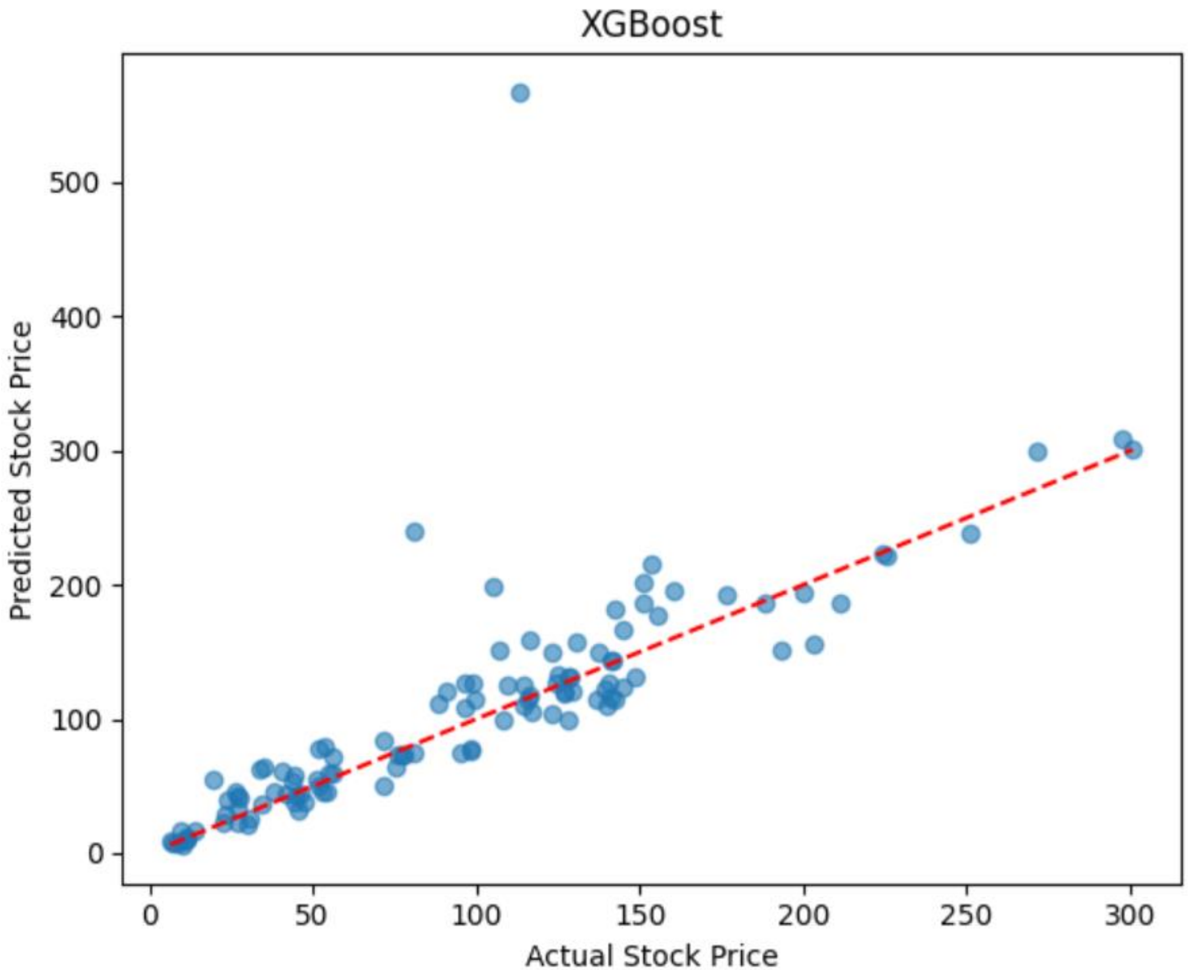
To predict company stock prices using machine learning models and examine how these predictions reflect broader economic performance.

Key Insights:

- **Random Forest** outperformed XGBoost with:
- **Higher R^2**
- **Lower RMSE**
- **XGBoost** showed slightly weaker results on this dataset.
- **Conclusion:** Random Forest captured patterns more effectively—likely due to its robustness against noise and overfitting.

4.1 Stock Price Prediction





Model	RMSE (↓ better)	R^2 Score (↑ better)
Random Forest	166.6135	0.8039
XGBoost	222.0428	0.6517

Key takeaways..

- Our analysis reveals that **hybrid and remote work models can positively influence company performance**, particularly in **knowledge-driven sectors** like **IT and Finance**. Using methods like **Granger Causality**, **Transfer Entropy**, and **machine learning models**, we found that:

- **Hybrid% shows significant correlation with revenue per employee** and, in some cases, stock performance—especially in IT.

- **Workplace flexibility** appears to offer an **economic premium** in sectors where output is not tightly bound to physical presence.

- **Machine learning models**, especially **Random Forest**, can effectively predict stock prices, highlighting their potential in understanding market dynamics.

- **Workforce risk** varies across sectors, with **Construction and Automobile** facing higher hidden costs due to turnover and dissatisfaction.

- Overall, the findings support the hypothesis that **workplace flexibility is not just a cultural shift—but an economically impactful one**, reshaping how companies perform and are valued in the market.

Bridging Firm Performance & Employee Experience

To deepen our understanding of hybrid work's impact, we now shift from firm-level metrics to employee-level evidence using survey data from 3,000+ professionals.

This enables us to triangulate our findings across multiple dimensions of productivity and policy.

Part II – The Hybrid Work Premium: Uncovering Productivity, Preferences, and Policy Insights

We drew on a two-wave remote-work survey administered in 2020 and 2021 by the NSW Productivity Commission (via Maven Analytics). Each wave interviewed ~1,500 “remoteable” employees in New South Wales across diverse industries and roles. Key survey domains included:

- **Attitudes & Preferences**: How workers feel about remote versus in-office work, and their ideal mix post-pandemic
- **Actual Work Patterns**: Share of hours spent remotely in Q4 2020 and time-use breakdown (work, commuting, family, domestic)
- **Organizational Context**: Company size, policies, leadership attitudes, and barriers to hybrid arrangements
- **Outcomes**: Self-reported productivity comparisons and perceptions of remote work’s impact on retention, recruiting, collaboration, and management

This rich, multi-year dataset lets us quantify the so-called “hybrid work premium” — how varying remote-work shares affect perceived productivity and inform strategic policy

INTIAL EXPLORATION

- **Survey Variability**
  - Question wording and available answers changed between 2020 and 2021 (2021 added 30+ new items), leading to inconsistent labels and punctuation.
- **Data Types & Cleaning Needs**
  - Almost all columns are stored as strings (categorical responses), requiring careful parsing and recoding.
  - Some time-use fields contain implausible outliers (e.g. >24 hours in a day), which we flagged for filtering or adjustment.
- **Sample Demographics**
  - The two waves together yield an even gender and age split, and balance managers vs. non-managers.
  - Slight over-representation of metro workers, large firms (>200 employees), and long-tenured staff (>5 years).
  - While typical remote-friendly sectors (e.g. professional/technical and finance) appear, the survey intentionally sampled across 19 industries to capture broader workforce perspectives.
  - No race or compensation data were collected.
- **Built-in Remote Bias**
  - All respondents are "remoteable" by design, and the distribution of answers skews heavily toward high–remote-share schedules and strong remote-work preferences.

These observations guided our cleaning strategy (standardizing categories, capping time estimates, recoding numeric fields) and shaped how we structured the subsequent econometric analysis.

DATA CLEANING

- **Objective**: Combine the 2020 and 2021 survey waves into a single, harmonized dataset.
- **Challenge**: Most question wordings—and thus column names—changed between years, even when asking the same thing.
- **Approach**:
- Reviewed both data frames side by side to identify matching concepts.
- Built a “column key” mapping each original question text to a standardized label.
- Selected 30 core variables common to both years for the merged file.
- **Implementation**: Used a Python notebook to extract and rename those columns, then concatenated the two waves.
- **Result**: A unified data frame with **3,019 rows** (1,507 + 1,512) and **30 clean, consistent columns**—ready for further data cleaning and econometric analysis.

```
columns = [
"birth_year",
"gender",
"industry_desc",
"occupation_desc",
"organization_size",
"manage_others",
"household",
"years_at_job",
"metro_or_regional",
"rw_percentage_2020",
"org_encouraged_rw",
"org_prepared_for_rw",
"rw_is_common_at_org",
"rw_permission_is_attainable",
"rw_collaboration_easy",
"preferred_rw_percentage_2020",
"preferred_rw_percentage_future",
"if_no_covid_employer_encourage_rw",
"if_no_covid_employer_support_rw",
"if_no_covid_i_would_have_choice_about_rw",
"productivity_remote_vs_office",
"inperson_hours__commuting",
"inperson_hours__working",
"inperson_hours__personal_family_time",
"inperson_hours_domestic_responsibilities",
"remote_hours_commuting",
"remote_hours_working",
"remote_hours_personal_family_time",
"remote_hours_domestic_responsibilities",
]
```

Cleaning Methodology & Transformation

Standardize Column Names

- Stripped punctuation (spaces, question marks, colons) and replaced with underscores.
- Introduced a unique responder_id field to distinguish rows across both years.

Handle Missing Values

- Counted nulls per column.
- For categorical fields, filled blanks with “No response” to preserve patterns of skipped questions.
- For numeric fields, imputed missing entries using the column median to mitigate outlier effects.

Detect & Cap Outliers

- Converted open-ended hour estimates to numeric and visualized via boxplots.
- Flagged any values > 3 σ above the mean (e.g. respondents entering weekly rather than daily hours) and replaced them with the median.

Bucket & Encode Categories

- Collapsed rare industry / occupation labels into an “Other” group to reduce sparsity.
- Re-mapped heterogeneous remote-share responses (percentages, text) into five discrete “days per week” bins (0–1, 1–2, …, 4–5 days).

Create Engineered Features

- Computed commute_time_difference = (in-person commute hours) – (remote commute hours).
- Added any further derived metrics (e.g. total work hours, remote share) to support downstream modeling.

Machine Learning Analysis

We recoded productivity into three levels—“less,” “same,” or “more” productive—and then combined “same” and “more” into a single “non-loss” category. The rationale: employees whose output does not decline remotely still bolster the business case for hybrid work.

Modeling Approaches

- **Ensemble Classifiers:** Random Forest, EasyEnsemble (a balanced-bagging AdaBoost), and standard AdaBoost.
- **Resampling Techniques:**
  - **SMOTE Oversampling** — synthetically balance the minority class.
  - **Random Over-Sampling** — duplicate minority examples.
  - **Cluster-Centroid Under-Sampling** — reduce majority class by clustering.
  - **SMOTEENN** — combine SMOTE oversampling with Edited Nearest Neighbors cleaning.
  - **Initial Results**
- **Random Forest** achieved the best overall performance for detecting the “more productive” group (precision ≈ 0.61, accuracy ≈ 0.93).
- However, **all models** struggled to accurately identify the “less productive” class—reflecting persistent class-imbalance challenges and overlapping feature patterns.

Hyperparameter Tuning

- To squeeze more performance out of our top model, we applied grid-search hyperparameter optimization on Random Forest settings (e.g., number of trees, max depth, min samples split/leaf, feature sampling).
- This tuning ensures we’re capturing genuine signal rather than noise, especially critical given the skew in productivity labels.
- Along the way, feature-importance rankings from the tuned Random Forest offer a transparent, interpretable ordering of which factors—commute hours saved, self-reported remote-share preference, industry, generation, organizational support, etc.—matter most in predicting remote productivity.

Before

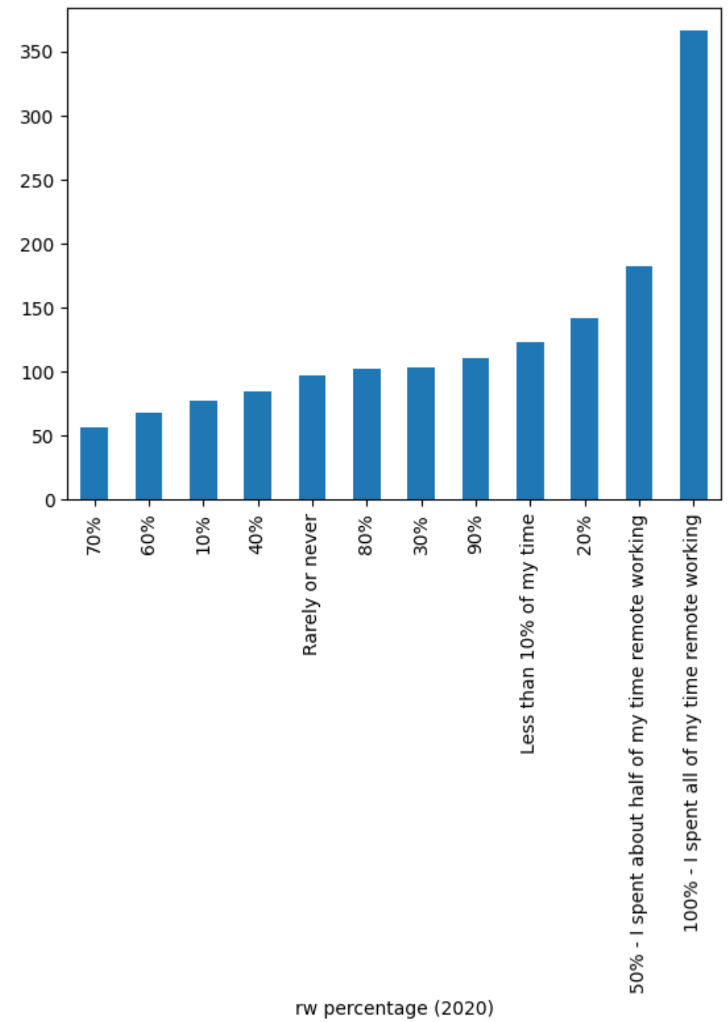
	pre	rec	spe	f1	geo	iba	sup
less_productive	0.45	0.15	0.97	0.23	0.38	0.13	112
more_productive	0.61	0.94	0.17	0.74	0.39	0.17	440
same_productive	0.46	0.09	0.96	0.15	0.29	0.08	203
avg / total	0.55	0.60	0.50	0.51	0.37	0.14	755

After

	pre	rec	spe	f1	geo	iba	sup
less_productive	0.77	0.93	0.95	0.84	0.94	0.88	112
more_productive	0.95	0.89	0.94	0.92	0.92	0.84	440
same_productive	0.90	0.92	0.96	0.91	0.94	0.88	203
avg / total	0.91	0.91	0.95	0.91	0.93	0.85	755

RW SHARE

Fully remote work dominated: **>12%** of your sample went 100 % remote in Q4 2020, and another **6%** were at exactly a half-remote schedule.
Purely in-person is very rare—fewer than **100** respondents "Rarely or never" worked from home.
Together, these numbers confirm that **most employees** in your study had at least **some** remote days, with a heavy tilt toward fully remote or strong hybrid (2–3 days+ per week).
100 % remote was the most common arrangement—**367** respondents reported being fully remote.
The next largest group (**183**) spent about **50 %** of their time remote (roughly half their week).
About **142** people were at **20 %** remote (one day per week), and **123** at **"More than 200"**… sorry that's a different chart—here it's actually **123** at **20 %** remote.
Mid-range hybrid schedules (30 %, 40 %, 60 %, 70 %, 80 %, 90 %) each attracted between **56** and **110** respondents.
Only **97** people said they were "Rarely or never" remote.



FiRRM

Using our tuned Random Forest classifier, we ranked which survey features best predict whether employees feel more productive when working remotely. Strikingly, **age** emerged as the single strongest predictor (importance ≈ 0.0468), suggesting that generational differences shape remote-work effectiveness. Close behind, the **amount of time spent working remotely in late 2020** (0.0361) and the **difference in commuting time** between in-office and remote days (0.0333) both strongly forecast self-reported productivity gains—underscoring the role of saved commute hours. Pre-pandemic **in-person commute hours** also register high importance (0.0310), pointing to the productivity premium for those who formerly endured longer commutes. Finally, the **hours devoted to domestic responsibilities while remote** (0.0299) complete the top five

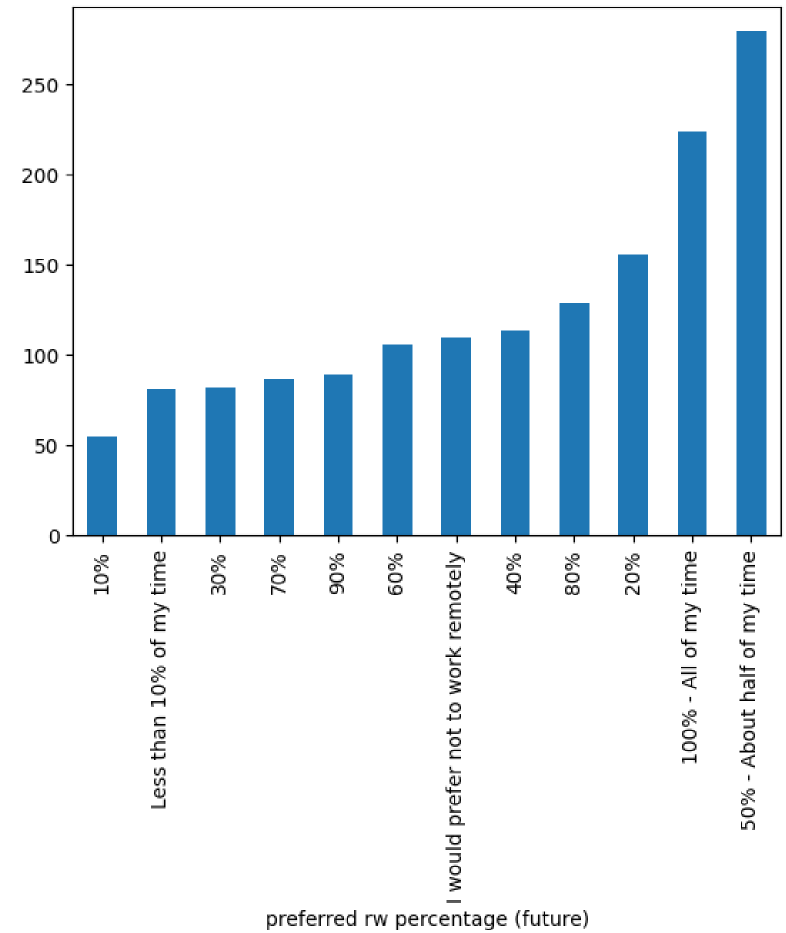
```
[(np.float64(0.04982832710783562), 'responder_id'),
 (np.float64(0.046800091937189696), 'birth_year'),
 (np.float64(0.036107191075112635), 'remote_hours_working'),
 (np.float64(0.033263094107517294), 'commute_time_difference'),
 (np.float64(0.031033122180365684), 'inperson_hours__commuting'),
 (np.float64(0.029399891786084497), 'remote_hours_domestic_responsibilities'),
 (np.float64(0.028991159876625806), 'remote_hours_personal_family_time'),
 (np.float64(0.027146203324834774), 'inperson_hours__personal_family_time'),
 (np.float64(0.02659352018896037), 'remote_hours_commuting'),
 (np.float64(0.025204955895868107), 'inperson_hours__working'),
 (np.float64(0.025081851592572296),
  'inperson_hours_domestic_responsibilities'),
 (np.float64(0.022800990159933577), 'preferred_rw_percentage_future_0-1 days'),
 (np.float64(0.02262150323857354), 'preferred_rw_percentage_2020_0-1 days'),
 (np.float64(0.020798723803885682), 'preferred_rw_percentage_2020_4-5 days'),
 (np.float64(0.02065046693907222), 'rw_collaboration_easy_Strongly agree'),
 (np.float64(0.01843570513797154), 'preferred_rw_percentage_future_4-5 days'),
 (np.float64(0.01618843892607781), 'rw_collaboration_easy_Somewhat disagree'),
 (np.float64(0.012669840278495582), 'manage_others_Yes'),
 (np.float64(0.012422770316405516), 'org_prepared_for_rw_Strongly agree'),
 (np.float64(0.01234024767280948), 'manage_others_No'),
 (np.float64(0.011759219159405131), 'organization_size_More than 200'),
 (np.float64(0.011254066707564212),
  'household_Couple with dependent children')
```

FiRRM

RW TIME SHARE

This bar chart shows respondents' **preferred share of remote work** once COVID concerns have passed:

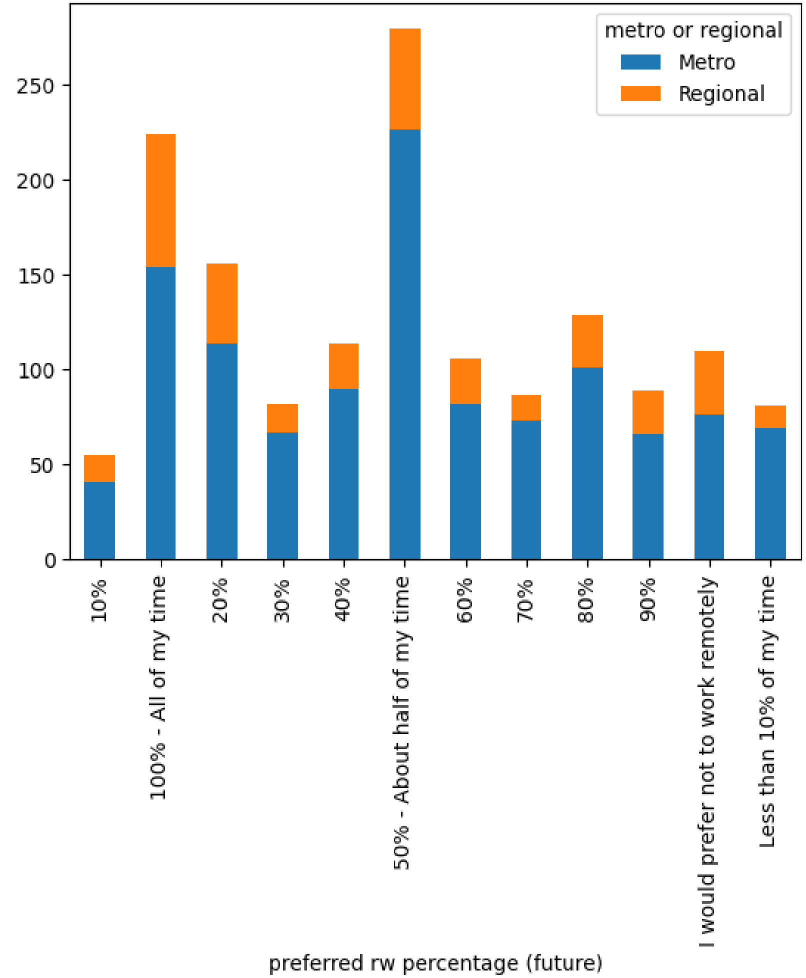
- The single largest group (~280 respondents) wants to work **about half their time remotely** (2–3 days/week).
- **Fully remote** ("100 % – All of my time") is next (~225 people).
- Roughly 155 prefer **20 % remote** (one day/week), and 130 **80 % remote** (four days/week).
- **Mid-range hybrids** — 30 % (1–2 days), 40 %, 60 %, 70 %, 90 % — each draw between 80–110 responses.
- Only about 90 respondents say they **"would prefer not to work remotely"** at all, and ~55 want just **10 % remote**.



RW PREFERENCE

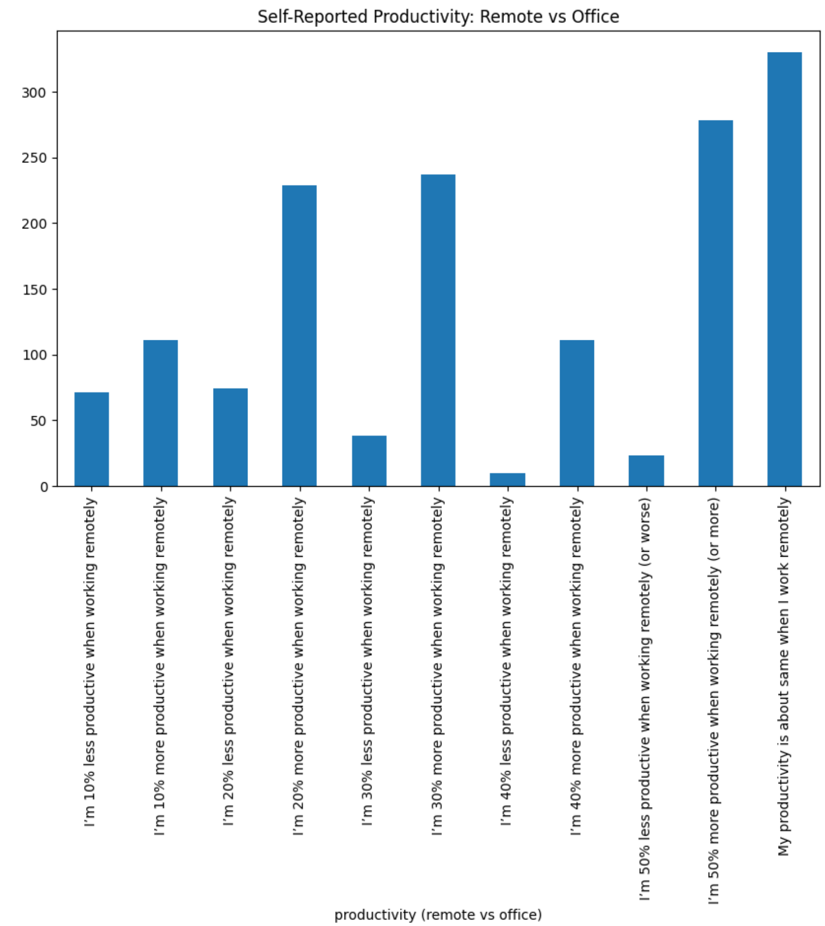
This bar chart breaks down **future remote-work preferences** by whether respondents live in **metro** (blue) or **regional** (orange) areas:

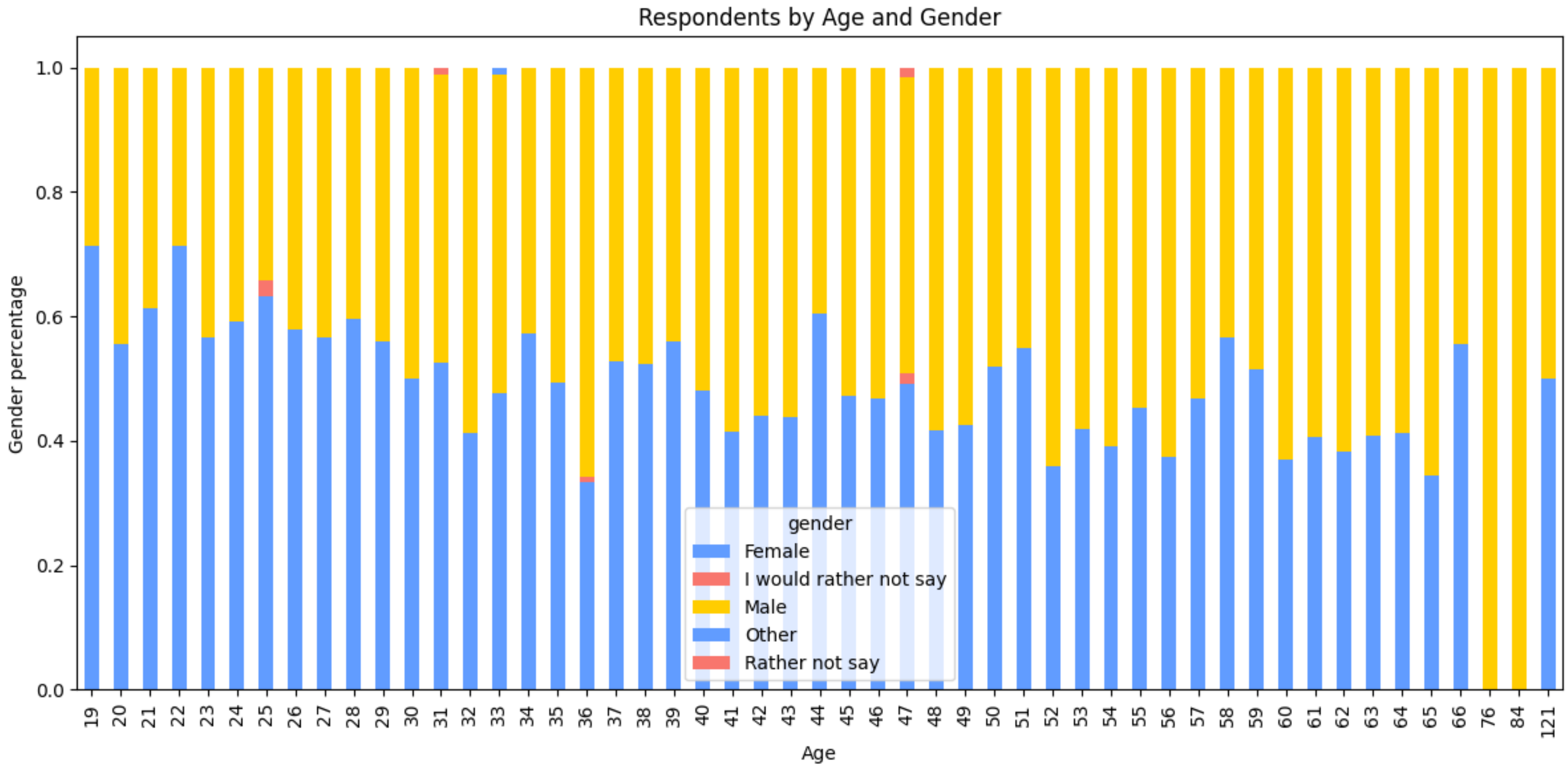
- **Half-time remote (50%)** is again the top choice in both metro and regional zones—around **225** metro and **55** regional respondents.
- **Fully remote (100%)** is the second-most popular: about **155** metro vs. **70** regional.
- For lighter remote schedules (20%, 30%, 40%), metro residents consistently outnumber regional ones by roughly 2:1.
- Regional employees make up a slightly larger share of the “would prefer not to work remotely” group—and also of the **10%** remote group—suggesting a small tilt toward in-person work in those areas.
- Overall, both cohorts strongly favor hybrid and full-remote arrangements, but **metro respondents** show even stronger demand, especially for heavier-remote schedules.

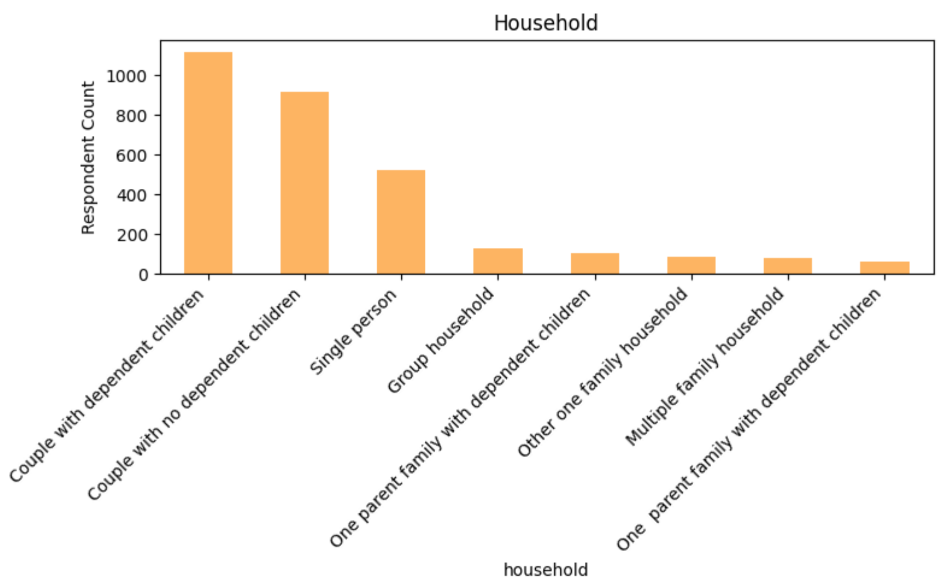


PRODUCTIVITY

- The **single largest** response is **“My productivity is about same when I work remotely”** (~330 respondents), indicating many feel neutral about the change in location.
- The next biggest group reports **“I’m 50% more productive when working remotely”** (~280 people), and then **“I’m 20% more productive…”** (~230).
- Smaller but still substantial numbers say **“I’m 30% more productive…”** (~240) and **“I’m 10% more productive…”** (~110).
- Fewer respondents report **productivity declines**:
  - “I’m 10% less productive…” (~70)
  - “I’m 20% less productive…” (~75)
  - “I’m 30% less productive…” (~40)
  - “I’m 40% less productive…” (~10)
  - “I’m 50% less productive…” (~25)
- **Overall takeaway**:
  - A majority feel **as productive or more productive** when remote (over 800 total), while only a small minority feel significantly less productive.




Respondents by Age and Gender
Gender percentage
1.0
0.8
0.6
0.4
0.2
0.0
gender
Female
I would rather not say
Male
Other
Rather not say
19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 76 84 121
Age


Household
Respondent Count
1000
800
600
400
200
0
Couple with dependent children
Couple with no dependent children
Single person
Group household
One parent family with dependent children
Other one family household
Multiple family household
One parent family with dependent children
household

FiRRM

Gender by Age

- Younger remote-capable respondents (mid-20s to early-30s) skew **female**, while older cohorts (late-30s and up) skew **male**.
- The balance flips around age **35**, marking the point where male share begins to exceed female.
- Very few participants chose “Other” or “Rather not say,” indicating most were comfortable disclosing gender.

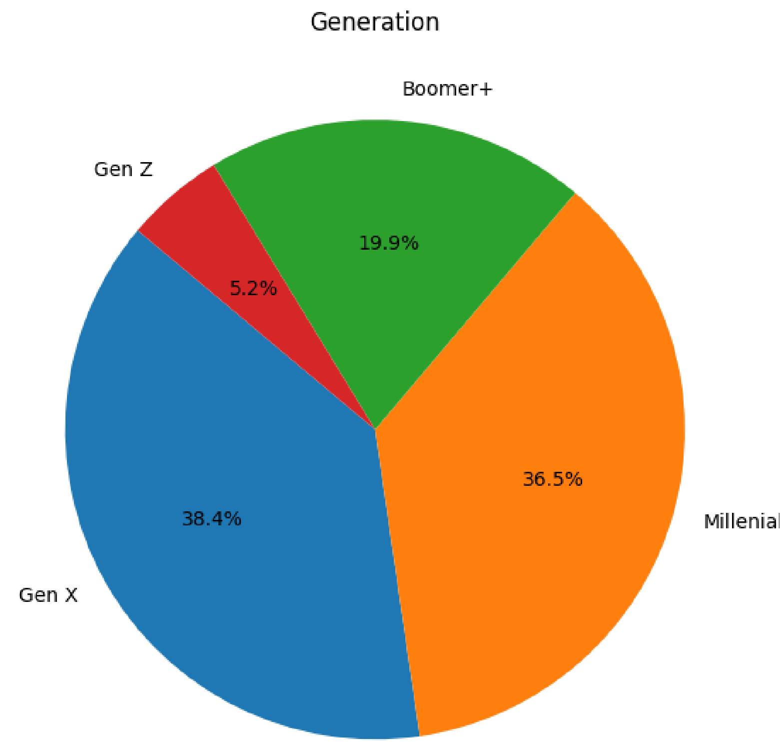
Generational Breakdown

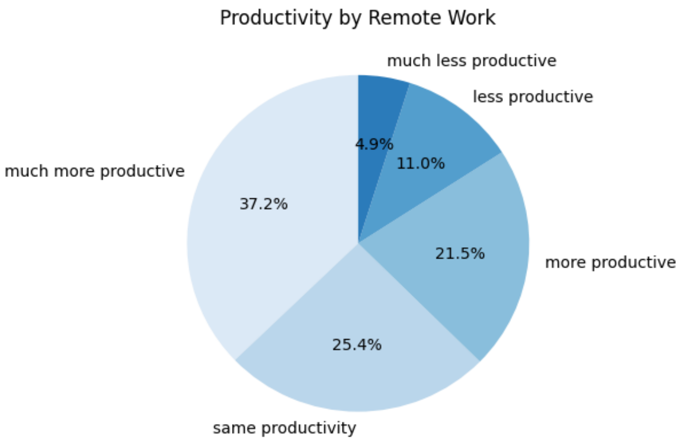
- **Gen Z (born ≥1997; ages ~18–24)**: 5.2% of respondents
- **Millennials (born 1981–1996; ages ~25–40)**: 36.5%
- **Gen X (born 1965–1980; ages ~41–56)**: 38.4%
- **Boomer+ (born ≤1964; ages 57+)**: 19.9%

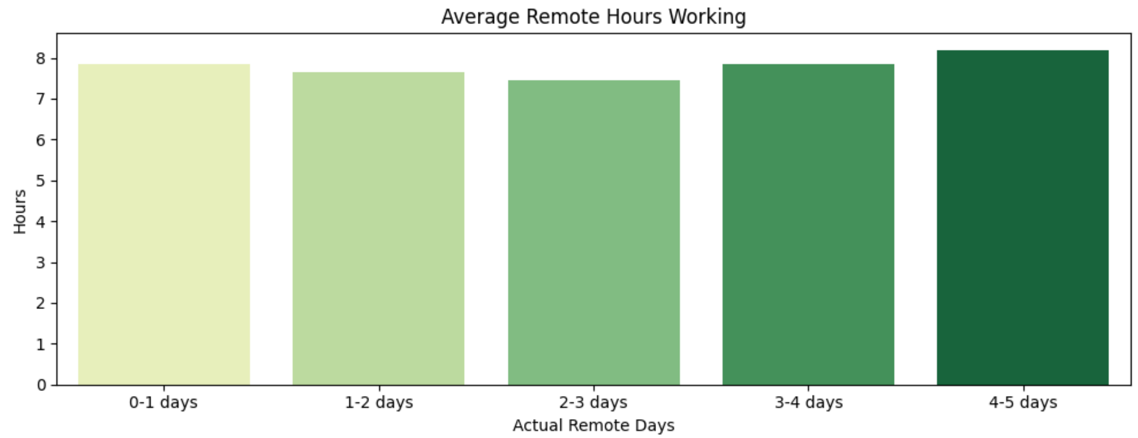
Together, Gen X and Millennials account for nearly three-quarters of the remote-capable workforce, with Boomer+ and Gen Z making up the tails on either end of the age spectrum.

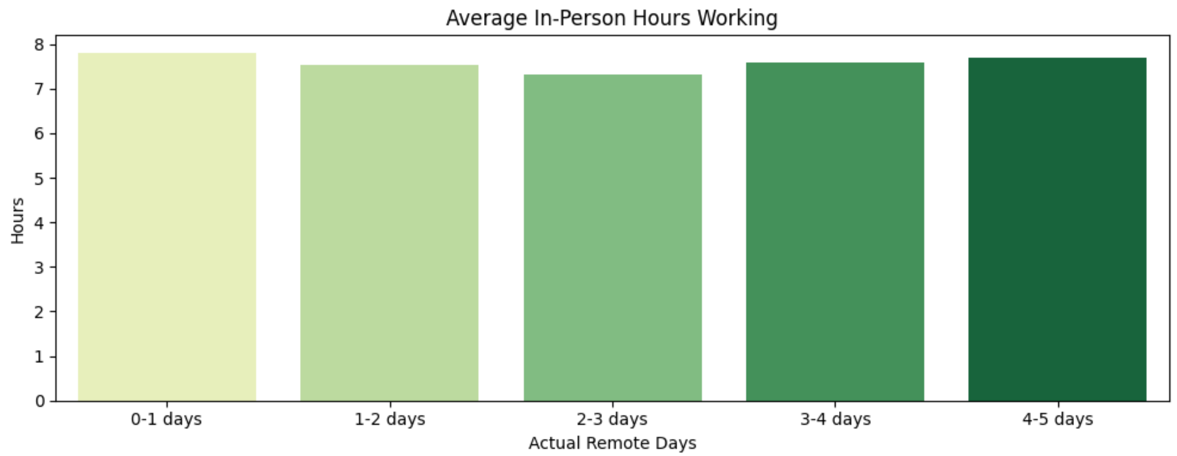
Household Composition

- **Couple with dependent children** is the largest group (≈1,100 respondents), followed by **couple with no dependent children** (≈920).
- **Single-person households** are next (≈530).
- **Group households**, **one-parent**, and **multi-family households** each fall below 150 respondents.





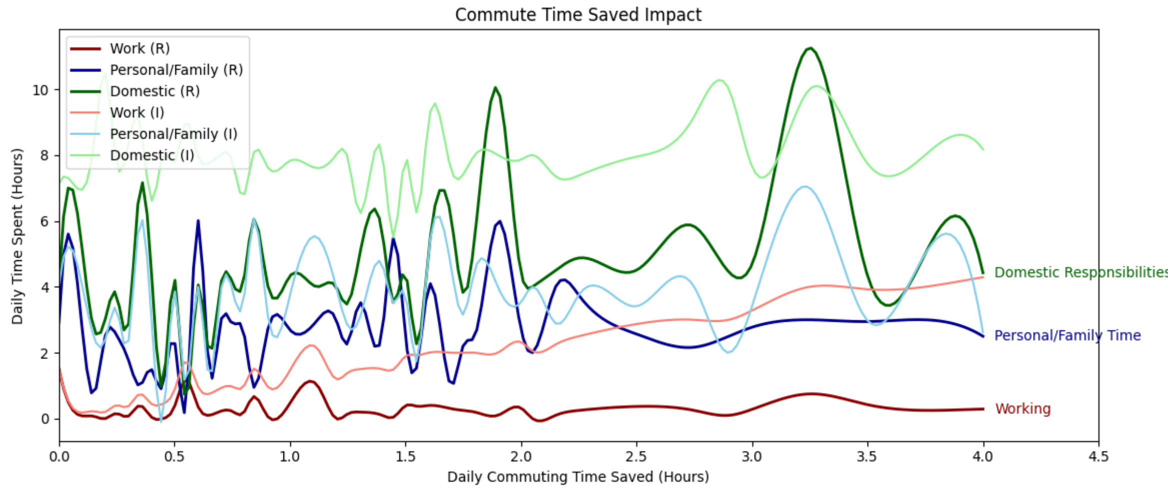




- **Productivity:** 37% report being "much more productive" remotely, and 85% say they're as productive or better off-site.
- **Remote Hours:** Average remote-day length rises from ~7.8 h (0–1 remote days/week) to ~8.2 h (4–5 days/week).
- **In-Person Hours:** On-site days dip from ~7.8 h to ~7.3 h around a 2–3-day remote split, then rebound to ~7.7 h for heavy hybrid users.

Takeaway: Employees overwhelmingly maintain or boost output when remote, and hybrid schedules shift longer, more focused remote days alongside slightly compressed office days.

•Remote days eliminate nearly all commute time (dropping from 1.7 h to 0.3 h on average), freeing up about 1.4 h per day that workers devote mostly to personal/family (61%) and domestic tasks (35%), while preserving roughly 7.4 h of focused work. In contrast, office days still consume 13.2 h—including 1.7 h commuting—highlighting how hybrid schedules both reclaim time for life and maintain total work output.



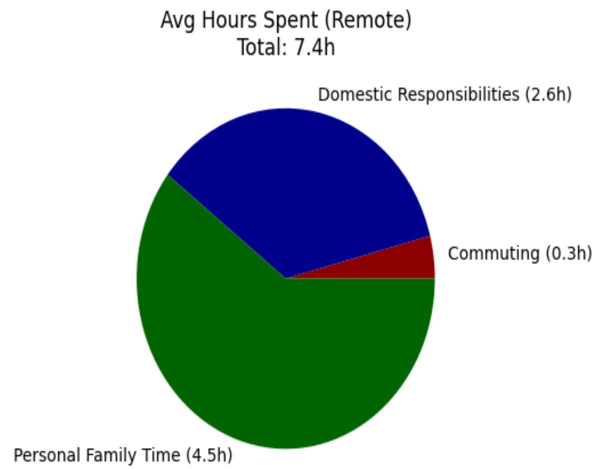
Average Hours Spent: Remote vs. In-Person

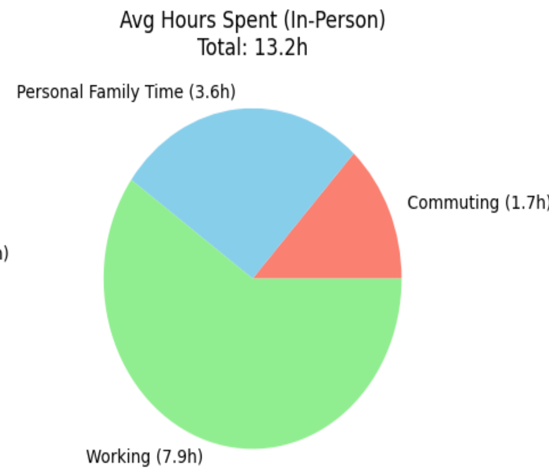
Remote days average **7.4 hours** split roughly into:

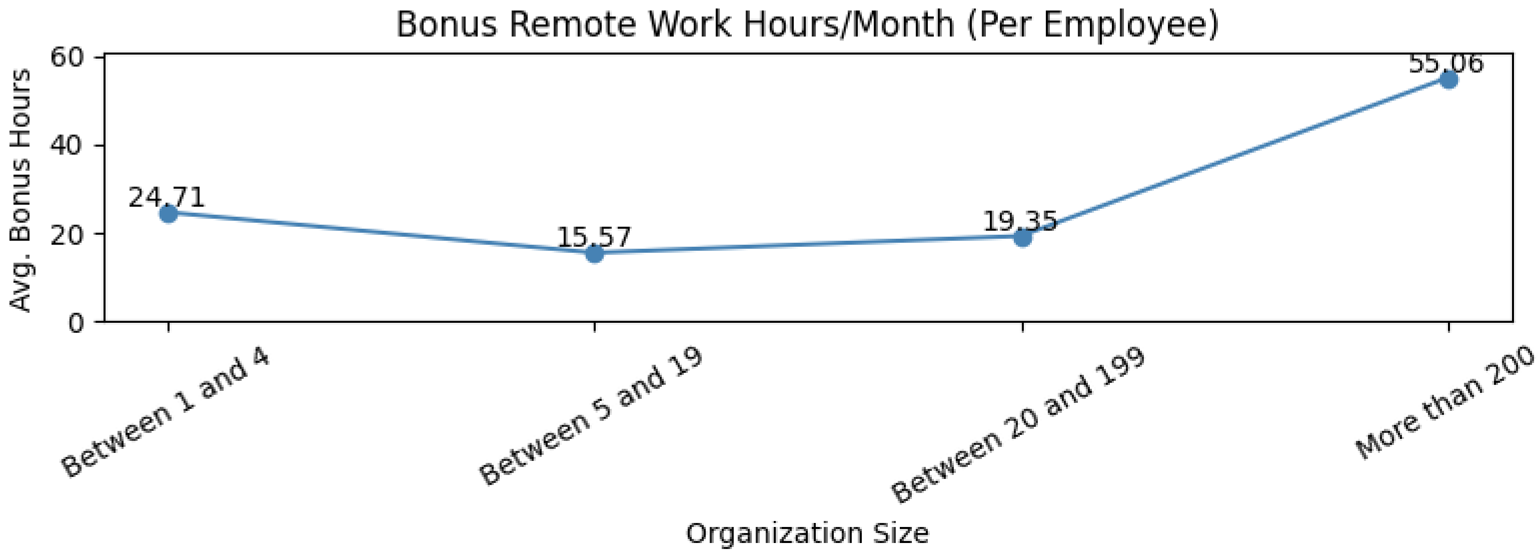
- **4.5 h** personal/family time (61%)
- **2.6 h** domestic responsibilities (35%)
- **0.3 h** commuting (4%)

In-person days average **13.2 hours** broken into:

- **7.9 h** working (60%)
- **3.6 h** personal/family time (27%)
- **1.7 h** commuting (13%)



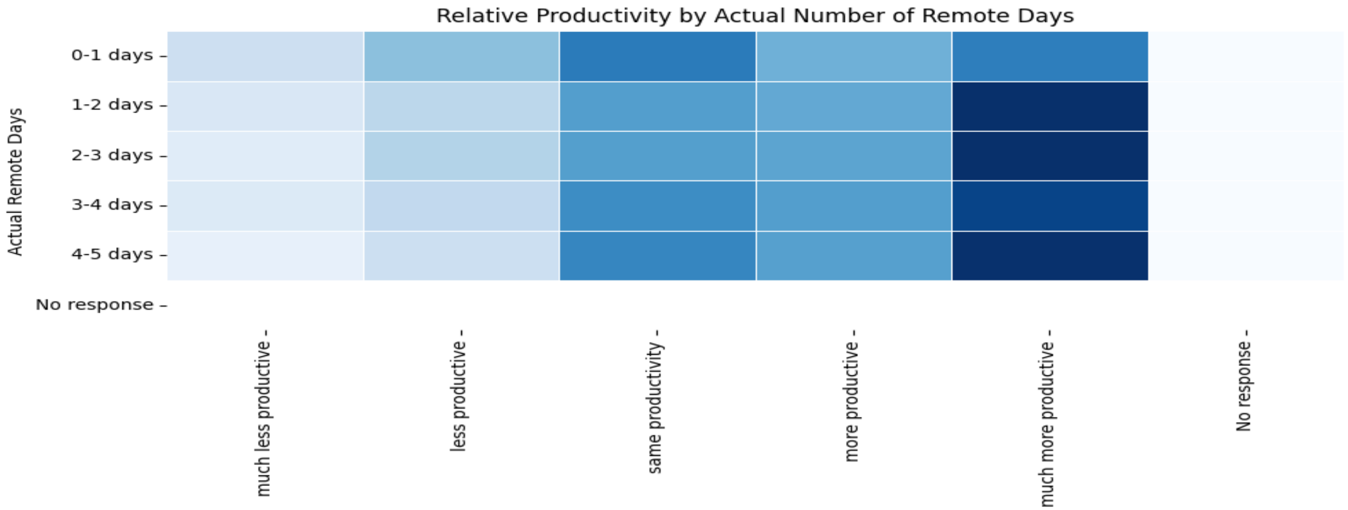


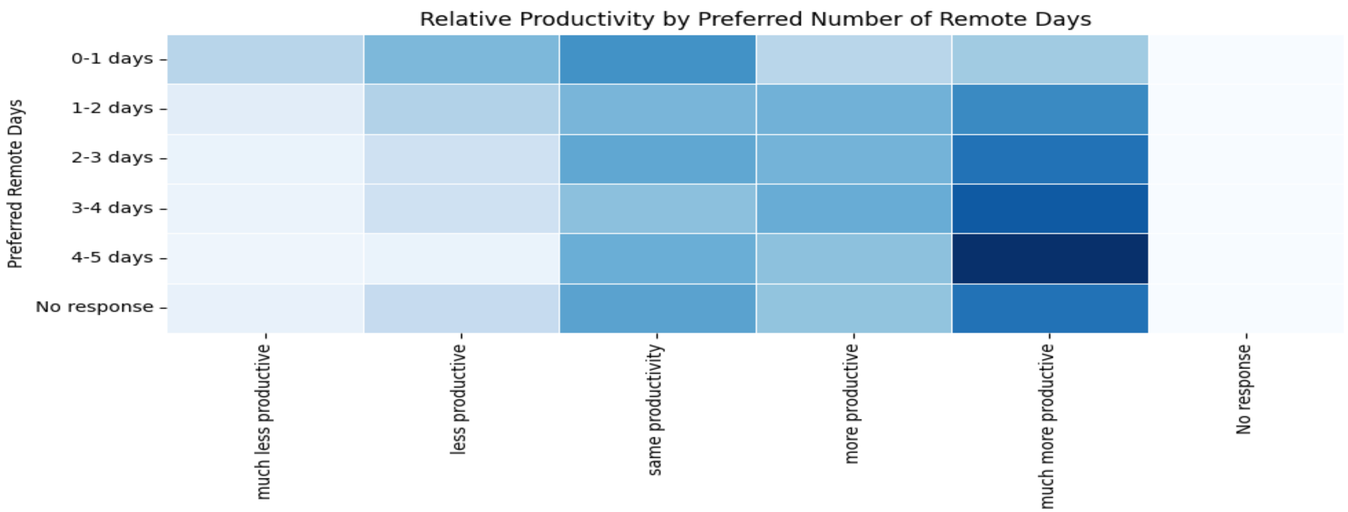


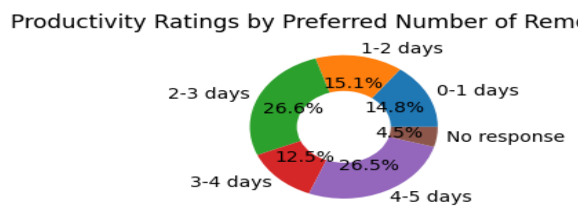
As company size grows, the "bonus" hours gained from remote work (i.e., extra remote-logged hours minus in-person hours) first dips slightly in small firms and then rises sharply in large organizations:

•**Small outfits (1–4 employees)** see about **25 extra hours/month** per employee.
•**Medium–small (5–19)** drop to **≈16 h/month**, suggesting less remote efficiency at that scale.
•**Mid-sized (20–199)** rebound to **≈19 h/month**.
•**Large enterprises (200+ employees)** unlock by far the largest gain—**≈55 bonus hours/month** per person—underscoring that bigger organizations realize the greatest productivity uplift from remote work.

How Productivity changes with actual and preferred remote days







FiRRM

Heatmap 1: Productivity vs. Actual Remote Days

- Across **all actual remote-day bins** (0–1 up to 4–5 days/week), the darkest cells lie under **"much more productive,"** showing that employees who *do* work remotely—even just 1–2 days—are likeliest to report a strong productivity boost.
- The **"same productivity"** column is also fairly dark in every row, confirming that most remote-capable workers feel no drop in output when off-site.
- Only a small fraction in each bin select **"less productive"** or **"much less productive."**

<u>Heatmap 2: Productivity vs. Preferred Remote Days</u>

- Preferences for **4–5 days/week remote** correspond to the deepest color in **"much more productive,"** indicating those who want nearly full-time remote work anticipate the largest gain.
- Even among those preferring **2–3 days/week**, "much more productive" dominates, while the "same productivity" column remains consistently strong across all preference levels.
- Those who prefer **0–1 days** show lighter shading overall, but still lean toward "same" or "more productive" rather than losses.

<u>Pie Chart: Distribution of Preferred Remote-Day Bins</u>

- **4–5 days** is the single largest slice at **26.5%**, closely followed by **2–3 days** at **26.6%**—together, over half the sample wants at least a "half-week" remote schedule.
- **1–2 days** accounts for **15.1%**, and **0–1 days** just **14.8%**, showing a clear skew toward multi-day remote working.

- **Strong Productivity Gains:** Even minimal remote schedules (1–2 days) yield substantial self-reported boosts, with "much more productive" the modal response across bins.
- **Preference Aligns with Performance:** Those desiring heavier remote schedules (2+ days) anticipate—and experience—the largest productivity increases.
- **Majority Lean Hybrid-Heavy:** Over half the workforce prefers 2–5 days remote, reinforcing that hybrid models aren't niche but central to sustaining high productivity.

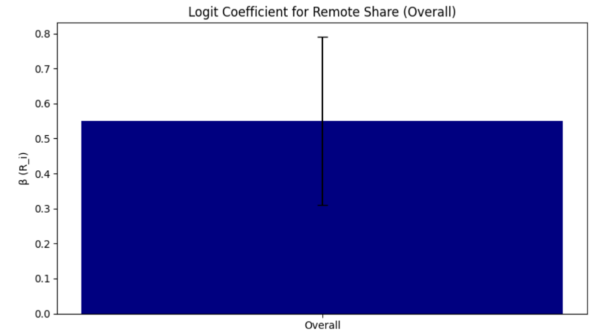
Cobb–Douglas Analysis

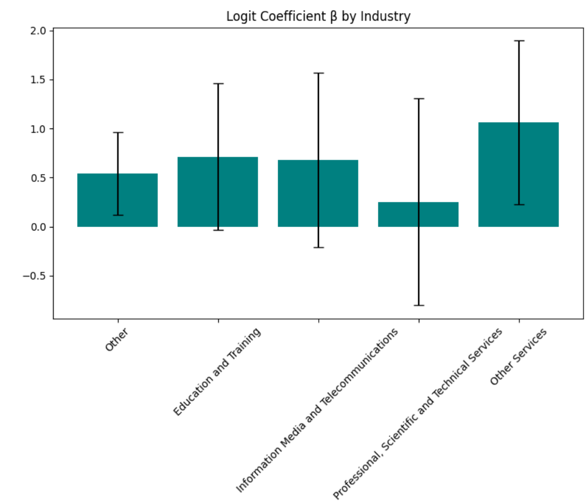
Y = A*L^(alpha)*K(beta)

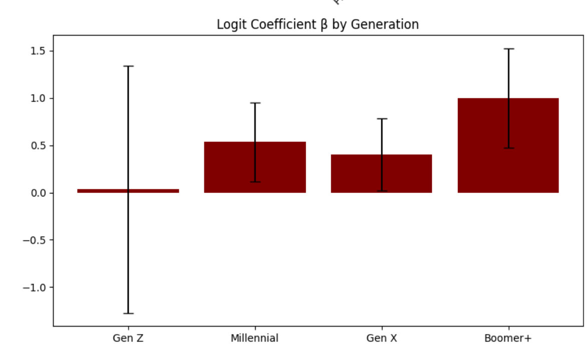
- Y is output (say, total value produced)
- L is labor input
- K is capital input
- A is a technology parameter
- α,β are elasticities showing the percent-change in output from a 1 % change in each input.

TOP CHART:

- $\beta \approx 0.55$ (95 % CI [0.31, 0.79])
- This is the log-odds coefficient on R_i , the fraction of each worker’s time spent remotely.
- Interpretation: Holding total hours worked (labor) and firm size (capital) fixed, a one-unit increase in remote share (i.e. going from 0 % to 100 % remote) multiplies the odds of reporting “more productive” so about a 73 % higher odds of saying “I’m more productive.”







Heterogeneity by Industry

- We ran separate logistic regressions for the five largest industry groups.
- **Professional, Scientific & Technical Services**:
- $\beta \approx 1.06 \rightarrow$ going fully remote more than **doubles** the odds of reporting higher productivity.
- **Education & Training** and **Information Media & Telecom**:
- $\beta \approx 0.7$ with wide confidence intervals → positive but less precise estimates.
- **Other Services** (e.g. hospitality, personal services):
- $\beta \approx 1.06 \rightarrow$ similarly large productivity gains from remote work.
- **Other (catch-all) industries**:
- $\beta \approx 0.54 \rightarrow$ moderate but still significant productivity boost.

Heterogeneity by Generation

- We segmented respondents into Gen Z, Millennials, Gen X, and Boomers+.
- **Boomers+**: $\beta \approx 1.00$
- Fully remote work **triples** their odds of feeling more productive (odds ratio $\approx e^1 \approx 2.7$).
- **Millennials**: $\beta \approx 0.55 \rightarrow \sim 1.7\times$ higher odds.
- **Gen X**: $\beta \approx 0.40 \rightarrow \sim 1.5\times$ higher odds.
- **Gen Z**: $\beta \approx 0.05$ (not statistically distinguishable from zero)
- Likely driven by small sample size in “remoteable” roles.

CONCLUSION

Remote share (percentage of days worked remotely) is a **strong, positive predictor** of self-reported productivity.

- A 100 pp increase in remote share multiplies the odds of reporting “more productive” by ≈ 1.73 (e^0.55).

The productivity uplift is **largest** in:
- **Professional, Scientific & Technical** and **Other Services** industries
- Among **Baby Boomer** respondents

These coefficients quantify the individual-level “hybrid work premium”—showing that more remote days translate into materially higher productivity odds across most groups.

References

[1] Directory. (2020). *Re-architecting work models*. https://www.deloitte.com/content/dam/assets-zone1/au/en/docs/about/2023/deloitte-au-hc-re-architecting-work-models.pdf
[2] LaSalle Network. (2024, April 18). *2021: Office Re-Entry Index*. https://www.thelasallenetwork.com/resources/2021-office-re-entry-index/
[3] Becker Friedman Institute for Economics at UChicago. (2024, July 19). *Why working from home will stick | Becker Friedman Institute*. Becker Friedman Institute. https://bfi.uchicago.edu/working-paper/why-working-from-home-will-stick/
[4] Andre, L. (2025, April 25). Employee turnover, while completely normal for any business, can become problematic if left unchecked. A fast. *Financesonline.com*. https://financesonline.com/employee-turnover-statistics/
[5] Do, J. (n.d.). *Why 100% of Fortune 500 Companies Embrace Hybrid Work (with Sodexo's Henrik Järleskog)*. https://www.flexos.work/learn/deeper-dive-into-hybrid-work-with-henrik-jarleskog
[6] Redmond, S. W. P. a. J. J. (2024, October 31). *The rise in remote work since the pandemic and its impact on productivity*. Bureau of Labor Statistics. https://www.bls.gov/opub/btn/volume-13/remote-work-productivity.htm
[7] Garton, E., & Mankins, M. (2020, December 1). *The pandemic is widening a corporate productivity gap*. Harvard Business Review. https://hbr.org/2020/12/the-pandemic-is-widening-a-corporate-productivity-gap
[8] J. Bisack, “The Human Element in Hybrid Teams,” Performance Improvement Possible, 2021.

References

[9] Accenture. (2022). *The future of work: productive anywhere*. https://www.bruegel.org/sites/default/files/wp-content/uploads/2022/04/Accenture_Future_of_Work_Key-Findings_April28_MK.pdf
[10] Garton, E., & Mankins, M. (2020b, December 1). *The pandemic is widening a corporate productivity gap*. Harvard Business Review. https://hbr.org/2020/12/the-pandemic-is-widening-a-corporate-productivity-gap
https://www.ibm.com/investor/annualreport
https://www.ibm.com/investor/financials
https://www.statista.com/statistics/265007/number-of-employees-at-ibm-since-2000/
https://www.alphavantage.co
https://coresignal.com
https://press.siemens.com
https://www.sap.com/investors/en/financial-reports.html
https://www.reuters.com
https://www.bloomberg.com

Q&A

Thank You for your Attention