Retain Top Talent
In The Future Of Work

Improve employee engagement in the 'new normal' with machine learning insights

Daisy Uy
Business Analytics & Insights

duy6@my.centennialcollege.ca

August 17, 2022

Meet Daisy

- **A data-driven corporate communications expert**
- 10 years of practical experience
- Connects the dots with data to create compelling stories

A bumpy return to the workplace

Sources:
Gartner, 2022.
Future Forum, 2021.
Gallup, 2019.

Executive-employee disconnect

A wide disparity in executives' and employees' work expectations is leading to deteriorating employee engagement and workplace dissatisfaction.

Turnover will increase and is costly

52% say flexible work policies will affect the decision to stay at their organization.

63% are open to getting a new job in 2022.

50-200% of employee's annual salary is the hard cost of replacing one employee

Reinventing the Future of Work

Harness employee insights

Discover key drivers in employees' work style preferences.

Iterate post-pandemic work policies

Incorporate employee insights learned into Future of Work decision-making.

Engage employees

Increase employee satisfaction with workplace arrangements and retain talent.

What drives employees' expectations around workplace arrangements?

Key insights uncovered by the machine learning model

Note:
- WFH (work from home)
- WBP (work from business premises/office)

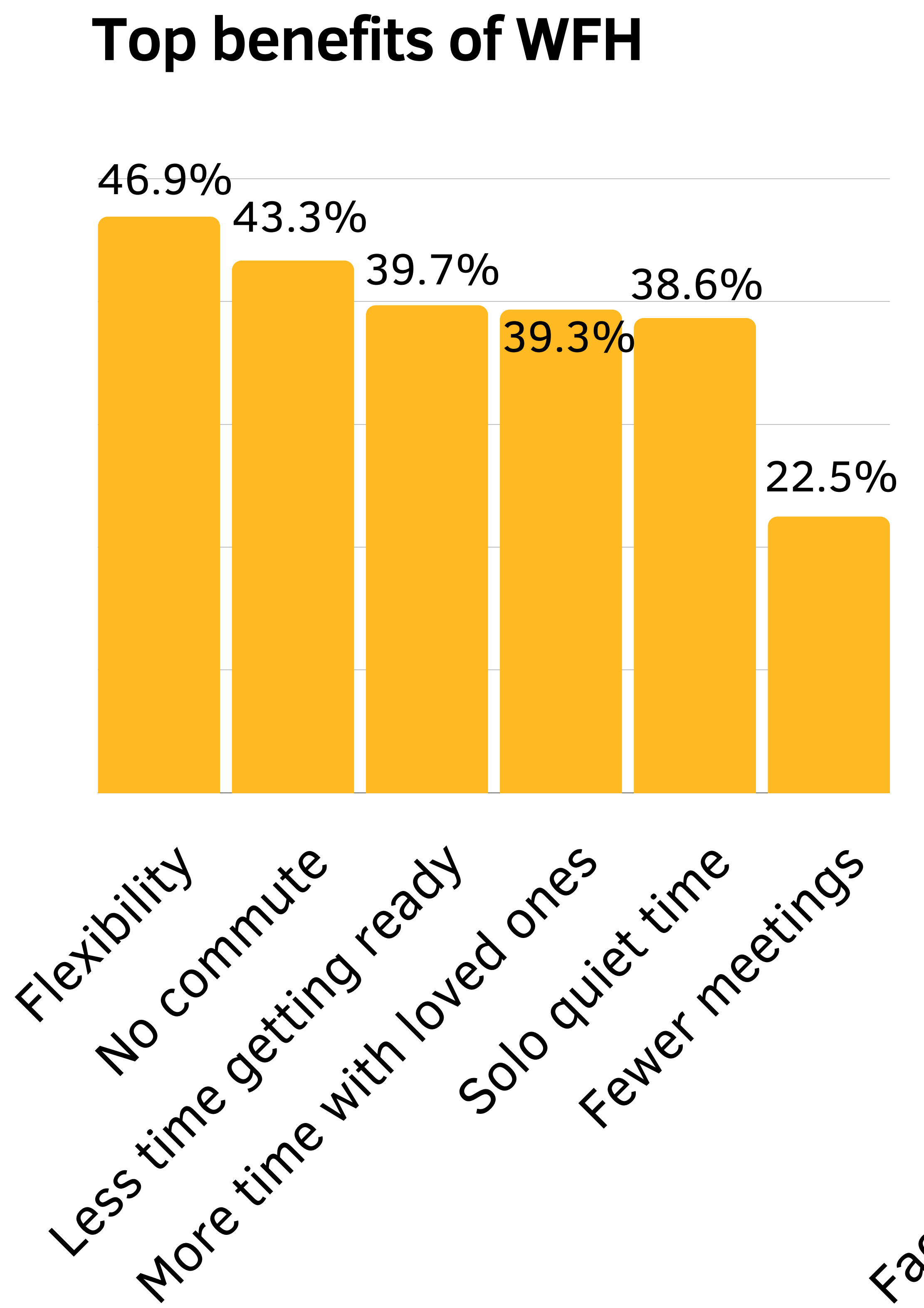
Your business location and size matters.
If your office is in a city or is large or medium in size, WFH or hybrid is a more likely arrangement.

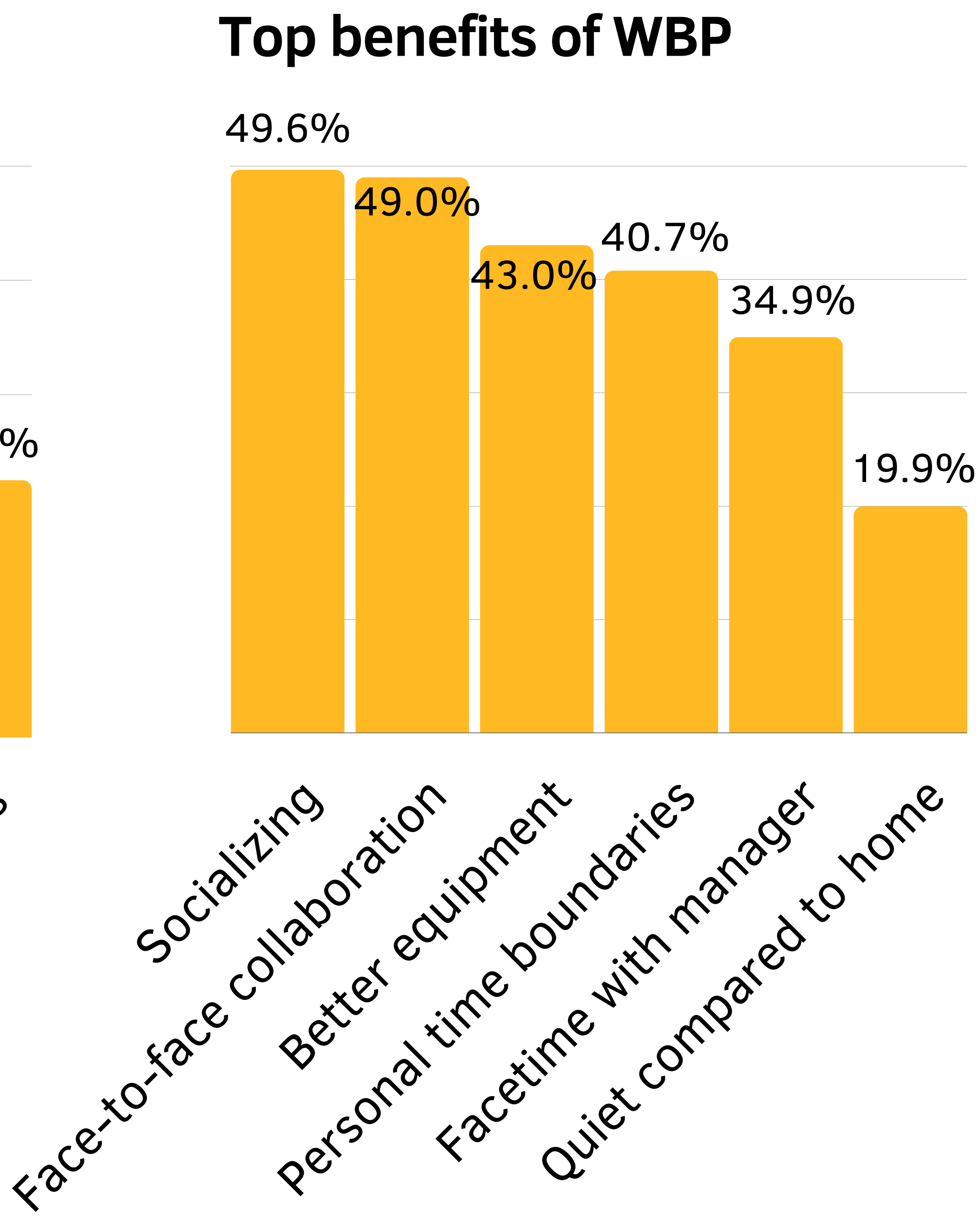
Worked from home during COVID?
If you WFH and felt substantially more efficient than pre-Covid, you're more likely to want to WFH.

Was extra time saved on commuting spent at work?
If employees spent time working that would have otherwise been commute time, it affects their preferred workplace arrangement.

Employees weigh in on the benefits of workplace arrangements

Key insights uncovered by the machine learning model





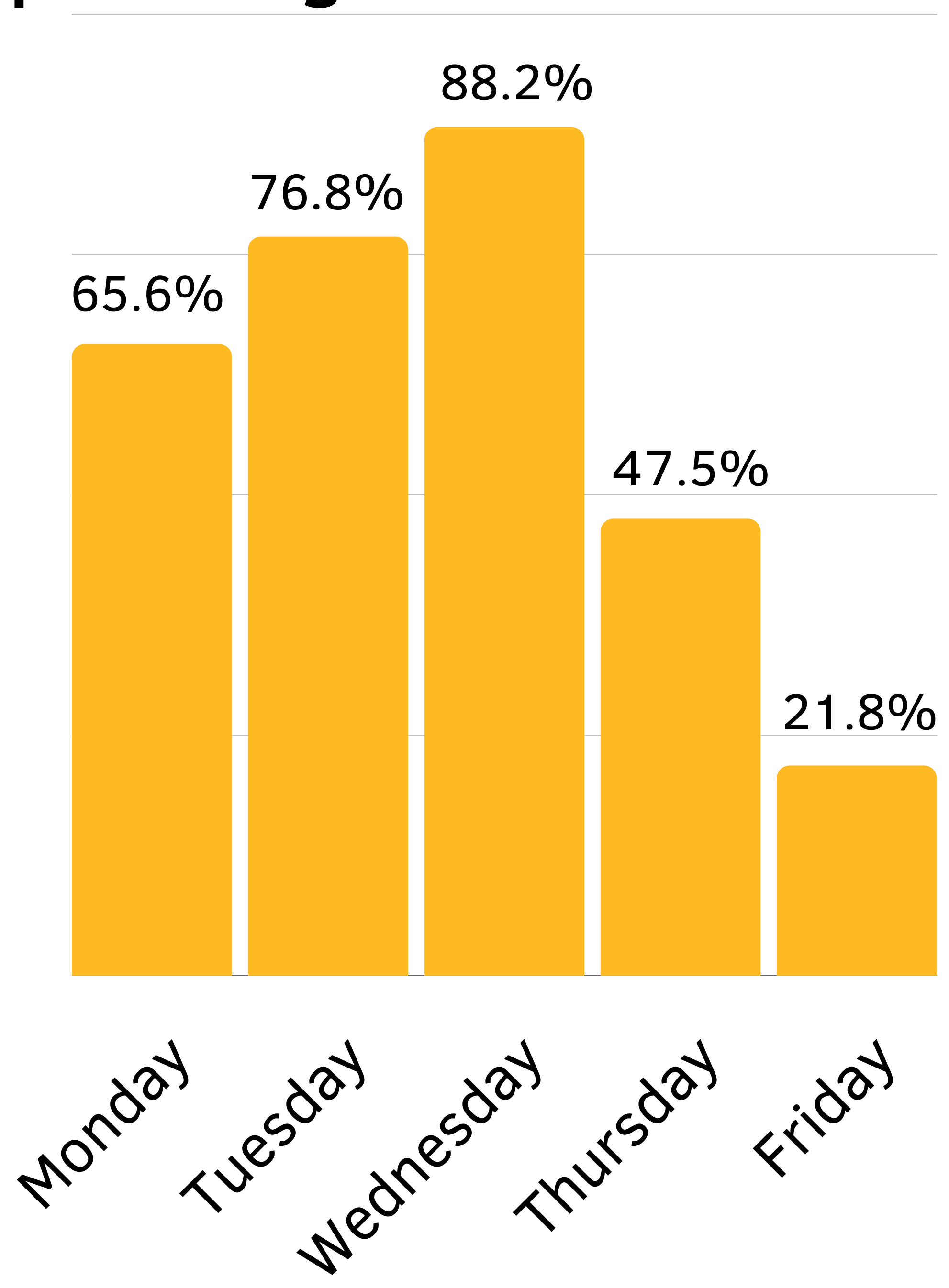
Note:
- WFH (work from home)
- WBP (work from business premises

When going to work, employees prefer...

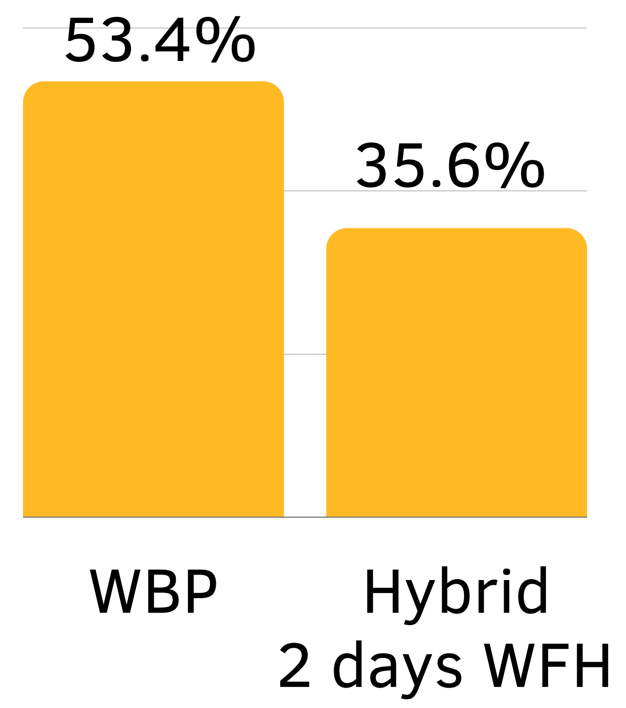
Key insights uncovered by the machine learning model

Note:
- WFH (work from home)
- WBP (work from business premises

Day of the week employees prefer to go to office



After COVID, the work arrangement I prefer is...



82.3%

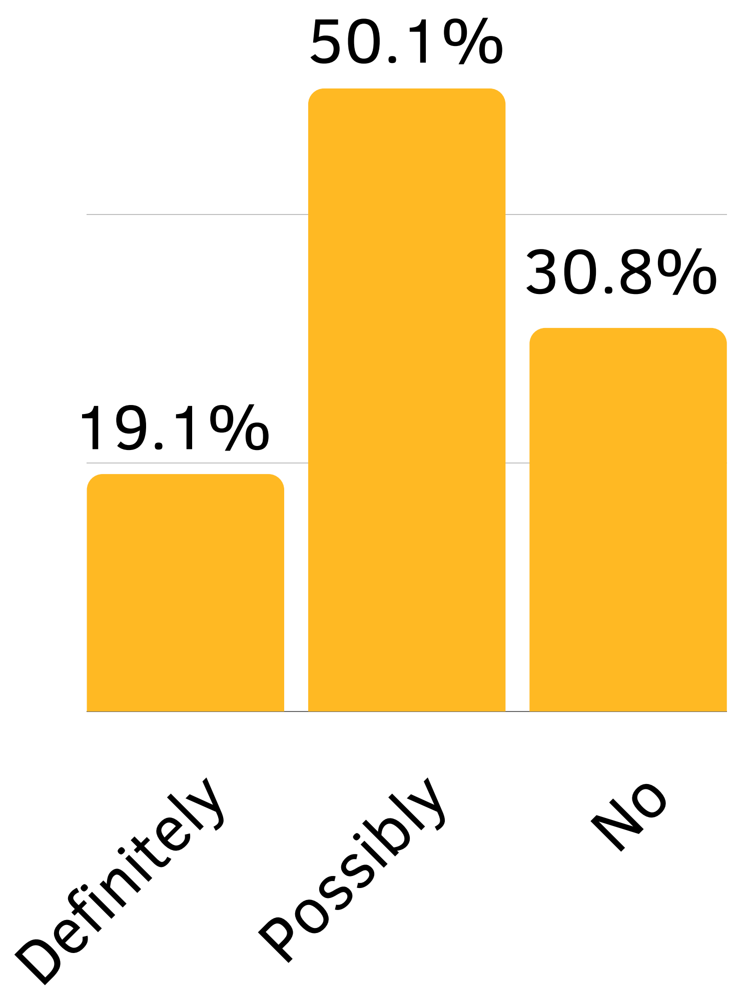
prefer going to work the same day as their coworkers

Would employees quit to WFH?

Key insights uncovered by the machine learning model

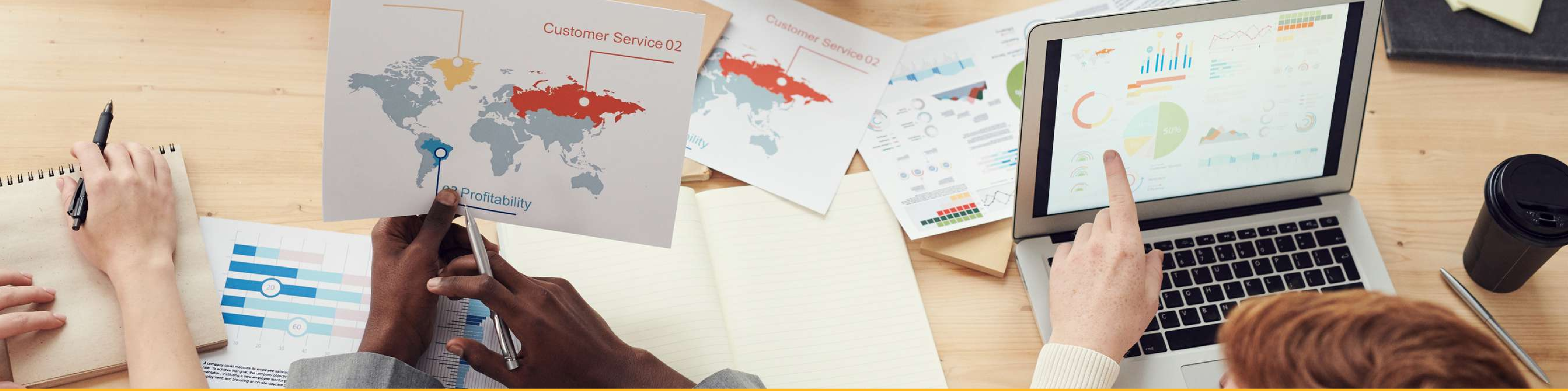
Note:
- WFH (work from home)
- WBP (work from business premises

Would you start looking for another job if you were guaranteed to get a WFH job?



21.5%

quit their job in the last 6 months



Explore data on employee work arrangement expectations with machine learning

A machine learning model analyzes proprietary survey data from American employees, leading to key insights for workforce planning and decision-making.

Advantages of using a machine learning model

Over traditional methods of studying employee engagement

More effective

Data at this scale corresponds to a diverse workforce.

More efficient

This approach uses less time and resources.

More objective

This model balances the natural subjectivity and bias inherent in decision-making.

How was this model created?

A multi-step approach towards our insights

1. **UNDERSTAND** our employee survey data.
2. **PREPARE** our data for the model.
3. **DISCOVER** the key drivers of U.S. employee work arrangement expectations.
4. **CLASSIFY** employees into their current work arrangements using the model.
5. **EVALUATE** how accurate our model is.
6. **ITERATE** on our model to improve performance.

Where is this employee data from?

Proprietary data from the Survey of Working Arrangements and Attitudes

56,000+ respondents

U.S. residents aged 20 to 64 years old who made ≥$10,000 in 2019

July 2021-June 2022

When respondents answered the survey (3,000 - 5,000 respondents monthly)

Featured by news outlets

Covered by the Wall Street Journal, New York Times, Financial Times, NPR, Bloomberg, and The Times (to name a few)

How do we mitigate our model's risks?

Taking into account our data source and model's assumptions and limitations

Potential errors in survey answers are reduced.

Respondents' attention was monitored during the survey.

Respondents are not duplicates of each other.

Though respondents answer in batches, they are different respondents each month.

Employee attitudes to work arrangements are likely to evolve.

Changing attitudes to the Future of Work must be monitored to adjust the model accordingly.

Insights represent American employees overall.

Some questions were not asked to every respondent, but we assume the trends observed apply to all respondents.

Key takeaways

- Employee turnover is expected to increase as we shift to the Future of Work.
- Incorporating employee input in workforce planning mitigates this.
- Key drivers behind employees' workplace arrangements include efficiency experienced during WFH and employer location and size.
- This machine learning model approach is an efficient and effective way to gain insight on employees for decision-making.

Appendix

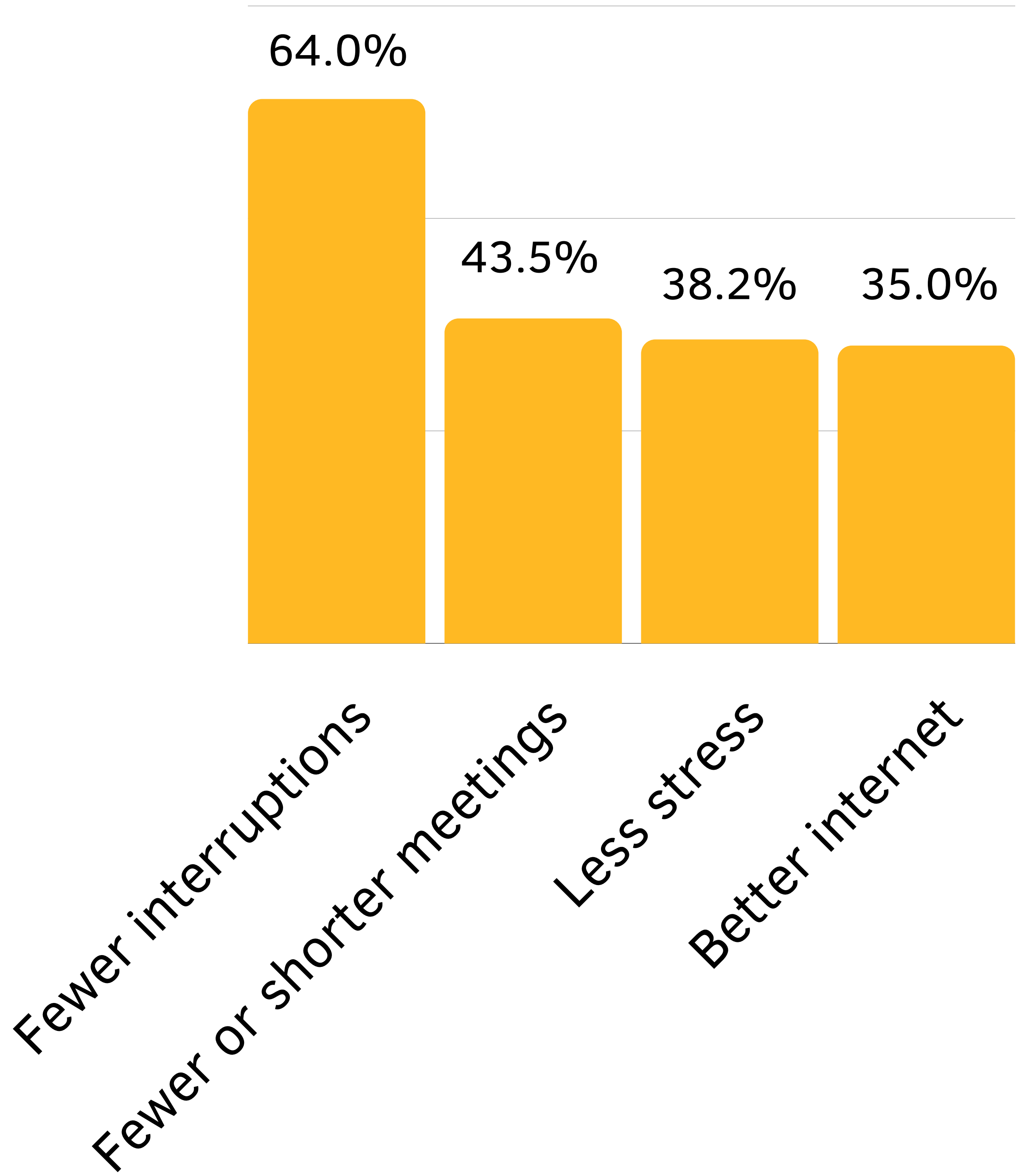
Technical specifications of the model build, detailed comparison of various model options, and additional statistics

Efficiency of WFH

Key insights uncovered by the machine learning model

Note:
- WFH (work from home)
- WBP (work from business premises

Apart from commuting, employees are more efficient WFH because...



24.0%

say they are at least **35% more efficient** WFH during COVID than pre-pandemic in office

50.4%

say they are less efficient WFH because many tasks cannot be done remotely

Objectives answered by the model

My aims when creating this model

What are the biggest contributing drivers behind U.S. employees' work arrangement expectations, as supported by data?

This will be answered by exploring the data collected by economists for the SWAA survey (Barrero et al., 2022).

Who is back in office, who isn't?

Based on the drivers discovered, a model will be created to classify if an American resident is working in business premises (in-office), working from home, or not currently working.

Key Drivers of The Model

By descending order of importance with description of the corresponding survey question

logpop_den_job_current	Log(Population density of the ZIP code of current residence)
workhours_duringCOVID	Hours worked per week at the time of the survey (during COVID) -- if currently working, otherwise missing
wfhcovid_ever	100 x 1(Ever WFH during COVID)
work_computer_pct	When working, what percentage of the time are you using a laptop or desktop computer?
date	Date when respondents answered the survey (Month and Year)
workhours_preCOVID	Hours worked per week pre-COVID
wfh_eff_COVID_quant	How efficient are you WFH during COVID, relative to on business premises before COVID (%)
hourly_wage	Hourly wage = (2019 income)/(pre-COVID weekly work hours * 50 weeks per year)
commutetime_quant	Length of commute (in minutes)
drivealone_current_pct	Driving alone: percent of commuting trips currently
nocommute_current_pct	Do not commute currently (0 to 100)
employer_sizecat	Counting all locations where your primary employer operates, what is the total number of persons who work for your employer?
uploadspeed	Internet upload speed from speed test. Winsorized at the 1st and 90th percentiles within each category from `internet_quality' variable

Key Drivers of The Model

By order of importance with corresponding survey question

downloadspeed	Internet download speed from speed test. Winsorized at the 1st and 90th percentiles within each category from `internet_quality' variable
work_industry	Industry of their current or most recent job
worktime_nonremoteable_pct	What percentage of your total working time do you usually spend on tasks that cannot be done remotely?
worktime_remoteable_pct	What percentage of your total working time do you usually spend on tasks that can be done remotely?
occupation_clean	Occupation of respondent (prepared for data use)
extratime_1stjob	Percent of commute time savings spent working on primary or current job
wfh_interviewing	Has working from home made it harder or easier to interview for prospective jobs?
groomtime_commute	How much time do you spend on grooming and getting ready for work when you commute to your employer's or client's worksite?
self_employment	Which of the following best describes your employment situation?
extratime_2ndjob	Percent of commute time savings spent on a second or new secondary job
extratime_indoorleisure	Percent of commute time savings spent on leisure indoors (e.g. reading, watching TV and movies)

Information On The Data (19 variables)

Build statistics of the dataset with 19 variables

Build Statistics

	logpop_den_job_current	workhours_duringCOVID	wfhcovid_ever
count	37842.000000	43910.000000	56062.000000
mean	7.480030	32.631200	67.619778
std	2.020219	14.880027	46.792975
min	2.316244	0.000000	0.000000
25%	6.150221	25.000000	0.000000
50%	7.701000	36.000000	100.000000
75%	8.717586	40.000000	100.000000
max	11.453425	70.000000	100.000000

	work_computer_pct	workhours_preCOVID	wfh_eff_COVID_quant
count	10304.000000	56062.000000	37596.000000
mean	59.202340	32.464183	10.812054
std	33.078866	16.329965	17.871907
min	0.000000	0.000000	-40.000000
25%	30.000000	25.000000	0.000000
50%	60.000000	40.000000	7.500000
75%	90.000000	40.000000	20.000000
max	100.000000	69.000000	40.000000

	hourly_wage	commutetime_quant	drivealone_current_pct
count	52584.000000	56062.000000	33587.000000
mean	143.003253	25.942439	52.290976
std	556.173427	23.914240	43.701732
min	1.000000	-20.000000	0.000000
25%	18.000000	10.000000	10.000000
50%	34.375000	20.000000	50.000000
75%	81.286337	32.500000	100.000000
max	20000.000000	120.000000	100.000000

	nocommute_current_pct	employer_sizecat	uploadspeed	downloadspeed
count	33587.000000	38339.000000	47230.000000	49926.000000
mean	14.472862	3.699418	79.695911	104.835637
std	35.183225	1.308189	99.370402	120.072471
min	0.000000	1.000000	0.000000	0.000000
25%	0.000000	3.000000	10.000000	19.219999
50%	0.000000	4.000000	36.674999	50.000000
75%	0.000000	5.000000	100.000000	147.884993
max	100.000000	5.000000	467.000000	460.000000

	work_industry	worktime_nonremoteable_pct	worktime_remoteable_pct
count	54775.000000	7011.000000	17654.000000
mean	8.282136	53.448153	50.339980
std	4.305827	37.208408	36.563067
min	1.000000	0.000000	0.000000
25%	5.000000	20.000000	16.000000
50%	7.000000	50.000000	50.000000
75%	11.000000	100.000000	85.000000
max	18.000000	100.000000	100.000000

	occupation_clean	extratime_1stjob	workstatus_current
count	54190.000000	37909.000000	56062.000000
mean	7.028954	25.260624	1.800810
std	2.518534	26.534263	0.770094
min	1.000000	0.000000	1.000000
25%	5.000000	9.000000	1.000000
50%	8.000000	20.000000	2.000000
75%	9.000000	35.000000	2.000000
max	12.000000	100.000000	3.000000

Median of the variables

	date	logpop_den_job_current	workhours_duringCOVID	wfhcovid_ever
0	2022-06-01	10.364245	40.0	100

	work_computer_pct	workhours_preCOVID	wfh_eff_COVID_quant	hourly_wage
0	100.0	40	0.0	17.5

	commutetime_quant	drivealone_current_pct	nocommute_current_pct
0	30.0	100.0	0.0

	employer_sizecat	uploadspeed	downloadspeed	work_industry
0	5.0	100.0	100.0	6.0

	worktime_nonremoteable_pct	worktime_remoteable_pct	occupation_clean
0	100.0	100.0	8.0

	extratime_1stjob	workstatus_current
0	0.0	1

Information On The Data (24 variables)

Build statistics of the dataset with 24 variables

Build Statistics

	logpop_den_job_current	workhours_duringCOVID	wfhcovid_ever
count	37842.000000	43910.000000	56062.000000
mean	7.480030	32.631200	67.619778
std	2.020219	14.880027	46.792975
min	2.316244	0.000000	0.000000
25%	6.150221	25.000000	0.000000
50%	7.701000	36.000000	100.000000
75%	8.717586	40.000000	100.000000
max	11.453425	70.000000	100.000000

	work_computer_pct	workhours_preCOVID	wfh_eff_COVID_quant
count	10304.000000	56062.000000	37596.000000
mean	59.202340	32.464183	10.812054
std	33.078866	16.329965	17.871907
min	0.000000	0.000000	-40.000000
25%	30.000000	25.000000	0.000000
50%	60.000000	40.000000	7.500000
75%	90.000000	40.000000	20.000000
max	100.000000	69.000000	40.000000

	hourly_wage	commutetime_quant	drivealone_current_pct
count	52584.000000	56062.000000	33587.000000
mean	143.003253	25.942439	52.290976
std	556.173427	23.914240	43.701732
min	1.000000	-20.000000	0.000000
25%	18.000000	10.000000	10.000000
50%	34.375000	20.000000	50.000000
75%	81.286337	32.500000	100.000000
max	20000.000000	120.000000	100.000000

	nocommute_current_pct	employer_sizecat	uploadspeed	downloadspeed
count	33587.000000	38339.000000	47230.000000	49926.000000
mean	14.472862	3.699418	79.695911	104.835637
std	35.183225	1.308189	99.370402	120.072471
min	0.000000	1.000000	0.000000	0.000000
25%	0.000000	3.000000	10.000000	19.219999
50%	0.000000	4.000000	36.674999	50.000000
75%	0.000000	5.000000	100.000000	147.884993
max	100.000000	5.000000	467.000000	460.000000

	work_industry	worktime_nonremoteable_pct	worktime_remoteable_pct
count	54775.000000	7011.000000	17654.000000
mean	8.282136	53.448153	50.339980
std	4.305827	37.208408	36.563067
min	1.000000	0.000000	0.000000
25%	5.000000	20.000000	16.000000
50%	7.000000	50.000000	50.000000
75%	11.000000	100.000000	85.000000
max	18.000000	100.000000	100.000000

	occupation_clean	extratime_1stjob	wfh_interviewing
count	54190.000000	37909.000000	8917.000000
mean	7.028954	25.260624	1.919928
std	2.518534	26.534263	1.084726
min	1.000000	0.000000	1.000000
25%	5.000000	9.000000	1.000000
50%	8.000000	20.000000	1.000000
75%	9.000000	35.000000	3.000000
max	12.000000	100.000000	4.000000

	groomtime_commute	self_employment	extratime_2ndjob
count	39158.000000	46694.000000	37909.000000
mean	26.432300	1.367263	9.264502
std	20.901037	0.761533	13.910620
min	0.000000	1.000000	0.000000
25%	10.000000	1.000000	0.000000
50%	23.000000	1.000000	5.000000
75%	35.000000	1.000000	15.000000
max	90.000000	4.000000	100.000000

	extratime_indoorleisure	workstatus_current
count	37909.000000	56062.000000
mean	16.479332	1.800810
std	18.193605	0.770094
min	0.000000	1.000000
25%	5.000000	1.000000
50%	10.000000	2.000000
75%	20.000000	2.000000
max	100.000000	3.000000

Information On The Data (24 variables)

Build statistics of the dataset with 24 variables

```
Median of the variables
        date  logpop_den_job_current  workhours_duringCOVID  wfhcovid_ever  \
0 2022-06-01               10.364245                   40.0            100

   work_computer_pct  workhours_preCOVID  wfh_eff_COVID_quant  hourly_wage  \
0              100.0                  40                  0.0         17.5

   commutetime_quant  drivealone_current_pct  nocommute_current_pct  \
0               30.0                   100.0                    0.0

   employer_sizecat  uploadspeed  downloadspeed  work_industry  \
0               5.0        100.0          100.0            6.0

   worktime_nonremoteable_pct  worktime_remoteable_pct  occupation_clean  \
0                       100.0                    100.0               8.0

   extratime_1stjob  wfh_interviewing  groomtime_commute  self_employment  \
0               0.0               1.0               30.0              1.0

   extratime_2ndjob  extratime_indoorleisure  workstatus_current
0               0.0                     10.0                   1
```

Model Results: Random Forest

Important variables and statistics measuring accuracy

Note: This model used 19 variables.

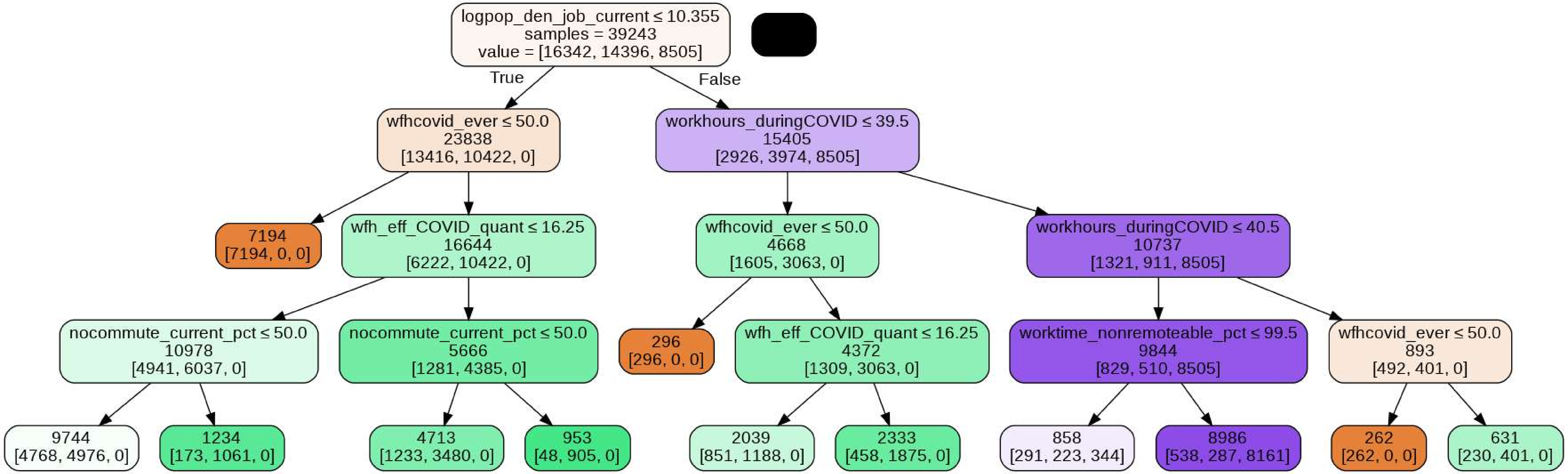
	feature	importance	std
2	day	0.000000	0.000000
0	year	0.005246	0.001647
12	nocommute_current_pct	0.013089	0.008381
17	worktime_nonremoteable_pct	0.015221	0.005037
6	work_computer_pct	0.021484	0.006795
18	worktime_remoteable_pct	0.025612	0.008610
19	occupation_clean	0.026264	0.002812
1	month	0.027081	0.004048
11	drivealone_current_pct	0.030892	0.013315
16	work_industry	0.030896	0.002115
13	employer_sizecat	0.034346	0.014609
14	uploadspeed	0.037007	0.002155
15	downloadspeed	0.038049	0.002365
10	commutetime_quant	0.052709	0.017991
9	hourly_wage	0.055734	0.015613
7	workhours_preCOVID	0.055838	0.027336
8	wfh_eff_COVID_quant	0.057470	0.029265
20	extratime_1stjob	0.059618	0.034027
5	wfhcovid_ever	0.089665	0.047310
4	workhours_duringCOVID	0.109852	0.026754
3	logpop_den_job_current	0.213928	0.033583

```
Multi-label Confusion Matrix:
[[[ 8847   995]
  [ 2076  4901]]

 [[ 8636  1988]
  [ 1062  5133]]

 [[12963   209]
  [   54  3593]]]
Classification Report:
              precision    recall  f1-score   support

           1       0.83      0.70      0.76      6977
           2       0.72      0.83      0.77      6195
           3       0.95      0.99      0.96      3647

    accuracy                           0.81     16819
   macro avg       0.83      0.84      0.83     16819
weighted avg       0.82      0.81      0.81     16819

Accuracy: 0.81021463820679
Mean Absolute Error: 0.19822819430406088
Mean Squared Error: 0.21511385932576252
Root Mean Squared Error: 0.4638036861925124
```

Model Results: Decision Tree

The first three splits show the main deciding variables.

Note: This model used 19 variables.



Model Results: Decision Tree

Important variables and statistics measuring accuracy

Note: This model used 19 variables.

```
Multi-label Confusion Matrix:
[[[ 9842     0]
  [ 3691  3286]]

 [[ 7261  3363]
  [  235  5960]]

 [[12609   563]
  [    0  3647]]]
Classification Report:
              precision    recall  f1-score   support

           1       1.00      0.47      0.64      6977
           2       0.64      0.96      0.77      6195
           3       0.87      1.00      0.93      3647

    accuracy                           0.77     16819
   macro avg       0.84      0.81      0.78     16819
weighted avg       0.84      0.77      0.75     16819

Accuracy: 0.7665735180450681
Mean Absolute Error: 0.25292823592365776
Mean Squared Error: 0.2919317438611095
Root Mean Squared Error: 0.5403070829270236
```

Model Results: Random Forest

Important variables and statistics measuring accuracy

Note: This model used 24 variables.

	feature	importance	std
2	day	0.000000	0.000000
0	year	0.005246	0.001647
12	nocommute_current_pct	0.013089	0.008381
17	worktime_nonremoteable_pct	0.015221	0.005037
6	work_computer_pct	0.021484	0.006795
18	worktime_remoteable_pct	0.025612	0.008610
19	occupation_clean	0.026264	0.002812
1	month	0.027081	0.004048
11	drivealone_current_pct	0.030892	0.013315
16	work_industry	0.030896	0.002115
13	employer_sizecat	0.034346	0.014609
14	uploadspeed	0.037007	0.002155
15	downloadspeed	0.038049	0.002365
10	commutetime_quant	0.052709	0.017991
9	hourly_wage	0.055734	0.015613
7	workhours_preCOVID	0.055838	0.027336
8	wfh_eff_COVID_quant	0.057470	0.029265
20	extratime_1stjob	0.059618	0.034027
5	wfhcovid_ever	0.089665	0.047310
4	workhours_duringCOVID	0.109852	0.026754
3	logpop_den_job_current	0.213928	0.033583

```
Multi-label Confusion Matrix:
[[[ 8887   955]
  [ 2094  4883]]

 [[ 8608  2016]
  [ 1010  5185]]

 [[12986   186]
  [   53  3594]]]
Classification Report:
```

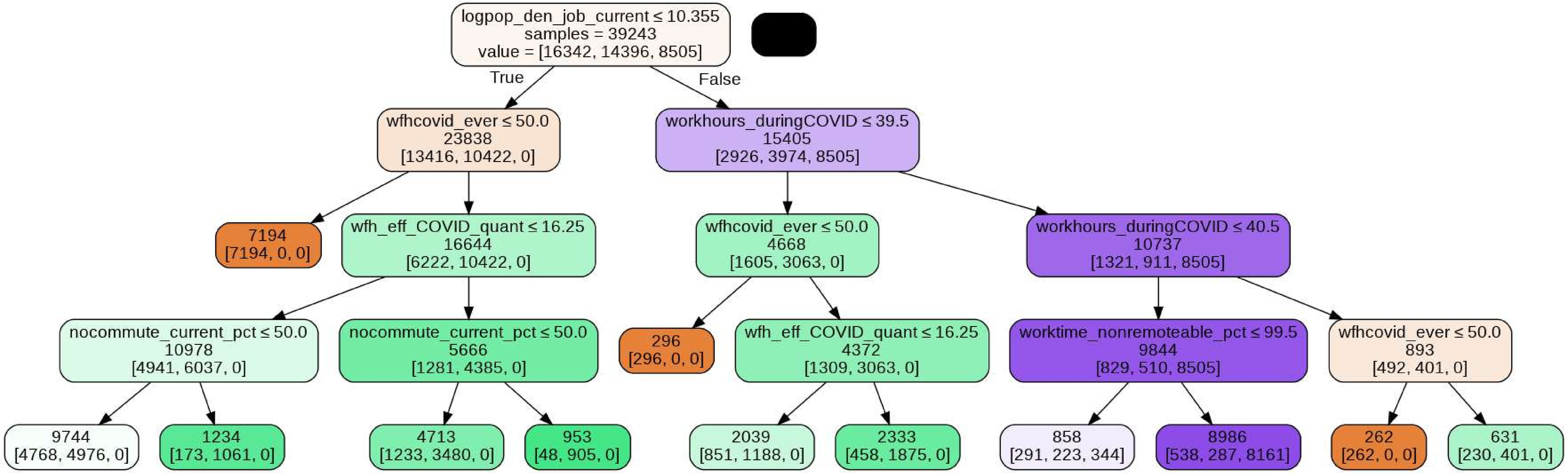
	precision	recall	f1-score	support
1	0.84	0.70	0.76	6977
2	0.72	0.84	0.77	6195
3	0.95	0.99	0.97	3647
accuracy			0.81	16819
macro avg	0.84	0.84	0.83	16819
weighted avg	0.82	0.81	0.81	16819

```
Accuracy: 0.8122956180510137
Mean Absolute Error: 0.19549319222308104
Mean Squared Error: 0.2110708127712706
Root Mean Squared Error: 0.4594244364106796
```

Model Results: Decision Tree

The first three splits show the main deciding variables.

Note: This model used 24 variables.



Model Results: Decision Tree

Important variables and statistics measuring accuracy

Note: This model used 24 variables, but accuracy is the same with the tree with 19 variables.

```
Multi-label Confusion Matrix:
[[[ 9842     0]
  [ 3691  3286]]

 [[ 7261  3363]
  [  235  5960]]

 [[12609   563]
  [    0  3647]]]
Classification Report:
              precision    recall  f1-score   support

           1       1.00      0.47      0.64      6977
           2       0.64      0.96      0.77      6195
           3       0.87      1.00      0.93      3647

    accuracy                           0.77     16819
   macro avg       0.84      0.81      0.78     16819
weighted avg       0.84      0.77      0.75     16819

Accuracy: 0.7665735180450681
Mean Absolute Error: 0.25292823592365776
Mean Squared Error: 0.2919317438611095
Root Mean Squared Error: 0.5403070829270236
```