# MathWorks Math Modeling Challenge 2022

## Adlai E. Stevenson High School


Team #15413, Lincolnshire, Illinois

Coach: Paul Kim

Students: Spandan Goel, Joy Qu, Gabriel Visotsky, Andrew Liu, Greycen Ren


## M3 Challenge FINALIST—$5,000 Team Prize

### JUDGE COMMENTS

*Specifically for Team # 15413—Submitted at the Close of Triage Judging:*

**COMMENT 1:** Very nicely done. It was evident that a lot of thought and a lot of work went into your paper. Sometimes the results were a bit overwhelming in the charts and graphs. Overall an extremely good paper.

**COMMENT 2:** The paper has done a great job in address all three questions. It has provided thoughtful and well-reasoned assumptions, easy-to-follow thought process, proficiency with modeling techniques, and creative approach toward each problem. Take Q1 for example, using hourly wage as a dominant indicator for remote readiness is ambitious. Yet, the paper established well-argued connection.  It would have been even better if the paper validates its result with the published remote percentage data in BLS data set 3. The analysis could bring more insight into the effectiveness of the model. Overall, very well done!

# Remote Work: ***Fad or Future***

Team 15413

February 27th, 2022

# 0 Executive Summary

The White House

Washington, USA

For the President:

This briefing paper focuses on the shifting American workforce scene as a result of the Covid-19 pandemic. For your reference, we have also included statistics from the UK to serve as a measure of comparison for a clearer understanding of remote work impact on society.

Part I details the estimation of percentage of workers whose jobs are remote-ready. This will help you understand the current potential of jobs in many industries that have the potential to convert to remote work quickly. Our model uses average hourly wage for each industry from 2006 to 2021 as historical trends to create a variable non-linear regression of wage over time. Then we use a logistical model to relate average hourly wage to remote readiness for the job type of a particular industry. Finally, we apply this model to given cities to predict the percentage of workers with remote-ready jobs in 2024 and 2027.

While Part I focuses on the potential of jobs for turning remote, Part II centers on those already with remote-ready jobs. Mr. President, this is a deeper dive into current public opinion of remote work. Will individuals with remote-ready jobs choose or be allowed to work at home? Part II is a model that evaluates both the chance an employer will select one of the virtual, in-person, and hybrid work models and an individual worker's choice to work from home given the chance to do so. To accomplish the former, we look at factors related to the productivity of working from home and factors related to the propensity that a worker will prefer online work compared to in-person work. In addition, we looked at Costa and McCrae's Five Factor Model for determining personality traits that may evaluate how workers perform in remote and in-person environments. We use a Monte-Carlo simulation to evaluate these metrics from both the worker and the employer perspective, allowing us to see how employers will accommodate the demographics of their worker populations and how individual worker propensities will affect responses to a rapidly changing office scene.

Finally, Part III synthesizes data from Parts I and II to estimate the percentage of workers who will work remotely. Mr. President, this offers you an idea of how the American workforce will respond to changes in remote work in the future by 2024 and 2027. In this model, we consider three factors that affect the magnitude of remote work impact in a particular city: environmental, economic, and happiness indexes. Next, we use those indexes to create a total impact equation that ranks the cities from greatest to least magnitude of remote work impact. This allows you to examine certain areas of the country to look for those that have been affected by remote work, and give you a preliminary idea of how to aid or encourage changes.

We hope that through this briefing paper, you will gain insight into the impact of remote work on the American public and economy. We look forward to your leadership in the coming years in light of our predictions.

Sincerely, Team 15413

## Contents

# 1 Part I: Ready or Not

## 1.1 Overview

What qualifies as remote-ready or not? We define remote-ready as the ability to convert to remote work tomorrow. With this definition in mind, we create a model which, for a given City, estimates the percentage of workers whose jobs are currently remote-ready. We then apply our model to the Cities below to make predictions for the percentage of remote-ready jobs in 2024 and 2027.

## 1.2 Assumptions

1. Remote readiness is dependent on the average wage of any certain Industry in order to accommodate for telework.
   - **Justification**: According to the U.S. Bureau of Labor Statistics, "the average wage is $35.22 in occupations that we deem suitable for telework, compared with $20.31 in occupations that we classify as unsuitable for telework" [21]. As wage plays a role in the remote readiness of any job, even within Industries, we can use wage as a means to determine remote readiness for each Industry. Per the National Bureau of Economic Research, this is especially true since jobs that can be done at home have significantly higher wages than those that cannot [11].
2. There is a current 95% remote readiness of the maximum estimated remote work capacity for each Industry except for government.
   - **Justification**: It is extremely likely that companies which were able to use remote work during the pandemic will be able to convert to remote work "tomorrow," as per our definition of remote-ready. We arbitrarily define our current remote readiness level as 95% as it makes sense that most of those companies that went remote earlier in the pandemic still have the means to do so currently. The other 5% accounts for exceptions to this, such as if a company went bankrupt, or if state governments mandate public school teachers for in-person learning.
3. The government Industry's remote readiness level is at its maximum estimated work capacity in 2021.
   - **Justification**: It is reasonable to assume that the government follows its own remote work guidelines so that government employees who have the capability to go remote all do so immediately when directed by the upper administrations, especially noting the height of the pandemic in 2021.
4. The maximum wage for any individual job is $80 per hour.
   - **Justification**: We expect that the pay for jobs will increase in the future. However, we don't anticipate this increase to exceed 50% of the current average hourly wage [2].
5. Average hourly wage fits a logistic regression.
   - **Justification**: As per Assumption 4, there is a maximum wage for any individual job. Hourly wages will likely approach this maximum asymptotically, making a logistic regression appropriate to model this.
6. The ratio of health care workers to education workers remains the same over time (2006-2021).

   - **Justification**: Both health care workers and education workers serve the overall population of a region. It is intuitive that as the population of an area increases, the number of both health care workers and education workers will do so proportionally.

7. The only Job Industry sectors are those provided in the M3 dataset.
   - **Justification**: These sectors comprise most of the major labor sectors per the Bureau of Labor Statistics (BLS).

8. The effect of wage on remote readiness is the same in the US and the UK.
   - **Justification**: The effect of wage on remote readiness likely comes to greater ability to purchase technology [27]. This is likely proportional to per capita GDP. As wages and prices both vary relatively proportionally, we expect the overall effect to be similar.

## 1.3 Model Development

### 1.3.1 Parameters

1. **Industries ($\mathcal{J}$).** The set of all Industries. We define these as the Industries in the M3 dataset: Logging, Mining, Construction (LM); Manufacturing (MA); Trade, Transportation and Utilities (TT); Information (IN); Financial Activities (FA); Professional and Business Services (PB); Education and Health Services (EH); Leisure and Hospitality Services (LH); Other Services (OS); and Government (GO). For notation purposes, we have $\mathcal{J} = \{LM, MA, TT, IN, FA, PB, EH, LH, OS, GO\}$ [1].
2. **Job Industry ($J$).** Each individual Job Industry $J \in \mathcal{J}$.
3. **Year ($Y$).** Calendar year.
4. **Cities ($\mathcal{C}$).** The set of all Cities we consider: Seattle, WA (SE); Omaha, NE (OM); Scranton, PA (SC); Liverpool, England (LP); Barry, Wales (BA). For notation purposes, we have $\mathcal{C} = \{SE, O, SC, LP, BA\}$ [1].
5. **City ($C$).** Each individual City $C \in \mathcal{C}$.
6. **Proportion of Jobs in Year Y ($p(J,Y)_C$).** The proportion of jobs in each Industry for some City $C$ in some year $Y$.
7. **Hourly Wage ($W_J(Y)$).** The average hourly wage for each Industry in year $Y$.
8. **Remote readiness by Industry in year Y ($r(J(Y))$).** The proportion of jobs in each Industry that are remote-ready by year $Y$.
9. **Workers in each Industry in each City in year Y ($N_C(J,Y)$).** The number of workers in each Industry in each City in year $Y$.

### 1.3.2 Model Derivation

We are tasked to find the proportion of workers in a certain City that are remote-ready. At the most basic level, for each year Y and $C \in \mathcal{C}$, this computation becomes

$$\sum_{J \in \mathcal{J}} p(J,Y)_C \cdot r(J(Y)). \tag{1}$$

We first tackle the remote readiness by job. We compiled data of average hourly wage from 2006-2021 for each Industry $J \in \mathcal{J}$. Using this data, using Assumption 5, we plotted a multi-variable non-linear regression of average hourly wage over time for each Industry [17] [16] [20] [14] [13] [19] [12] [15] [18]. For each Industry, we minimize the sum of squares error to fit the trend to the general equation

$$W_J(Y) = \frac{a}{1 + b \cdot c^{-d(Y-2006)}}. \tag{2}$$

with computed constants $a, b, c, d$ for each individual Industry. As per Assumption 4, we capped the value of $a$ at \$80 as the maximum average hourly salary. We compute $W_J(Y)$ for all $J \in \mathcal{J}$. We use these equations to compute $W_J(2024)$ and $W_J(2027)$ for each $J \in \mathcal{J}$. Our computed constants for our regression can be found in Appendix A.

Using Assumption 2, we then use a logistic model to fit our hourly wage values to remote readiness for each job. Using 95% of the maximum remote readiness as the current level of remote readiness (based on 2021 hourly wage for each job), we create a new set of logistic functions modeling remote readiness for each wage value, which is in turn determined by Industry and year. We utilize the M3 Dataset titled Remote Work Data to provide us with estimated maximum values for remote work by Industry [1].

Some Industries we consider are not given in the Dataset. We account for this in the following ways. For LM, we maintain a maximum value of 0 as it is impossible to do physical labor jobs remotely. For MA, we find that 1% of production jobs can be remote. For TT, we consider sales-related professions (that comprise most of Trades and Utility jobs), thus using a maximum value of 28%. For IN, we consider this equivalent to office and administrative jobs, as information services are primarily required for this. FA and PB are considered as "Business and Financial Operations," which has a maximum of 88% remote work [1]. For EH, we utilize Assumption 6 to compute, as the values for health care workers and education workers are so drastically different. As there are 4,140,800 educators [4] compared to 22,000,000 healthcare workers [3], we weight the remote readiness for EH as such. We find that the readiness among EH workers is

$$\frac{4,140,800 \cdot 0.98 + 22,000,000 \cdot 0.02}{4,140,800 + 22,000,000} = 19.35\%.$$

For LH, we consider jobs as equivalent to personal care and service, using a maximum value of 28%. The BLS defines OS as jobs providing services not otherwise explicitly mentioned, most of which fall under the "Community and Social Service" umbrella [18]. Consequently, we use 37% as our maximum. Finally, for GO, considering government as public administration, we use a similar 65% as with other administrative occupations [1]. We then computed the values of remote readiness for each Industry. These sets of remote readiness values are constant across $C \in \mathcal{C}$ and vary only by Industry and year. These values are shown in Table 1.1. Then, for maximum value $c$, we can write the following general form for $r(J(Y))$ (our computed constants for our regression can be found in Appendix A):

$$r(J(Y)) = \frac{c}{1 + ae^{-b \cdot W_J(Y)}}. \tag{3}$$

Having found the readiness in each Industry in each year, we then need to find the proportion of workers in each Industry in each City for the next few years. Looking at overall trends in the number of workers per Industry over time, the trends are approximately linear. We neglect values for 2020 and 2021 as they are anomalous due to COVID-19; we anticipate that with further reopening efforts, economic trends will eventually mirror those seen before the pandemic.

Constructing linear regressions for each City, we are able to compute the number of individuals working in each Industry $J \in \mathcal{J}$ for each year Y. The calculated constants for each linear regression for

| Remote Readiness by Industry in 2024, 2027 | | |
|---|---|---|
| | $r(J, 2024)(\%)$ | $r(J, 2027)(\%)$ |
| LM | 0 | 0 |
| MA | 0.9952 | 0.9959 |
| TT | 26.6931 | 26.8896 |
| IN | 61.9798 | 62.2528 |
| FA | 83.7216 | 84.0800 |
| PB | 83.7746 | 84.0931 |
| EH | 18.4621 | 18.5949 |
| LH | 18.4621 | 18.5949 |
| OS | 24.7790 | 25.0483 |
| GO | 35.3434 | 35.6778 |

*Table 1.1: Remote readiness by Industry in 2024 and 2027.*

each City can be found in Appendix A. Using these values, for each City $C \in \mathcal{C}$, we can compute

$$p(J, Y)_C = \frac{N_C(J, Y)}{\sum_{J \in \mathcal{J}} N_C(J, Y)}. \tag{4}$$

These values are shown in Tables 1.2 and 1.3.

| Proportion of Workers in Each Industry for Each City in 2024 | | | | | |
|---|---|---|---|---|---|
| | $p(J, 2024)_{SE}(\%)$ | $p(J, 2024)_{OM}(\%)$ | $p(J, 2024)_{SC}(\%)$ | $p(J, 2024)_{LP}(\%)$ | $p(J, 2024)_{BA}(\%)$ |
| LM | 5.8163 | 6.3107 | 3.8247 | 20.0003 | 7.4355 |
| MA | 6.3820 | 7.2771 | 7.6907 | 13.4301 | 5.7694 |
| TT | 18.9676 | 6.8594 | 25.0845 | 22.2339 | 1.2643 |
| IN | 6.2023 | 1.9428 | 0.6417 | 10.1849 | 5.8950 |
| FA | 4.6396 | 10.2899 | 4.7412 | 3.1088 | 6.0160 |
| PB | 14.9336 | 16.6904 | 11.8262 | 5.7734 | 11.1711 |
| EH | 14.6669 | 18.9012 | 21.9171 | 2.7867 | 17.4881 |
| LH | 10.4140 | 11.5737 | 9.5394 | 8.3923 | 21.4362 |
| OS | 3.9213 | 4.3462 | 3.0308 | 11.0300 | 6.0362 |
| GO | 13.9585 | 1.5373 | 10.9311 | 2.7867 | 17.4881 |

*Table 1.2: Proportion of workers in each City for each occupation in 2024.*

| Proportion of Workers in Each Industry for Each City in 2027 | | | | | |
|---|---|---|---|---|---|
| | $p(J, 2027)_{SE}$ (%) | $p(J, 2027)_{OM}$ (%) | $p(J, 2027)_{SC}$ (%) | $p(J, 2027)_{LP}$ (%) | $p(J, 2027)_{BA}$ (%) |
| LM | 5.7870 | 6.3342 | 3.8016 | 19.7196 | 7.4944 |
| MA | 6.0591 | 7.0372 | 6.6675 | 13.6332 | 5.2889 |
| TT | 18.9091 | 6.6298 | 25.5428 | 23.2241 | 1.1141 |
| IN | 6.3471 | 1.7459 | 0.3784 | 10.2265 | 5.6643 |
| FA | 4.4590 | 10.3533 | 4.6881 | 3.0574 | 6.0896 |
| PB | 15.0976 | 16.7718 | 12.2280 | 5.6779 | 11.3089 |
| EH | 14.9450 | 19.2796 | 22.4404 | 2.5708 | 17.4682 |
| LH | 10.5446 | 11.6540 | 9.7806 | 8.0925 | 21.9563 |
| OS | 3.9450 | 4.3760 | 2.9334 | 10.9604 | 6.1472 |
| GO | 13.8115 | 1.5382 | 10.7655 | 2.5708 | 17.4682 |

*Table 1.3: Proportion of workers in each City for each occupation in 2027.*

From here, we can compute the proportion of remote-ready workers in each City. These results are shown in Table 1.4.

| Proportion of Workers Remote-Ready in Each City in 2024, 2027 | | | | | |
|---|---|---|---|---|---|
| | SE | OM | SC | LP | BA |
| 2024 | 41.11% | 43.59% | 35.63% | 28.12% | 40.48% |
| 2027 | 41.13% | 43.82% | 36.06% | 28.18% | 40.80% |

*Table 1.4: Proportion of workers who are remote-ready in each City.*

## 1.4 Results and Discussion

These results show the remote readiness of each City to be between 28% and 44% for the next 5 years with modest increase between 2024 and 2027. This relatively small change is expected as current capacity for remote work has nearly been achieved. Assuming logistic approach towards maximal remote readiness by Industry, following the COVID-19 Pandemic, most Industries have minimal capacity for continued increase in remote readiness. These observations lend credence to our results.

## 1.5 Strengths and Weaknesses

### 1.5.1 Strengths

Our model is based on historical trends from the past decade. Following these trends gives credence to our results based on precedence. Further, we were able to stratify our calculations by year, City, and Industry. Taking these factors separately enables us to minimize the confounding effects between them. Moreover, separating Cities and Industries allows us to avoid overgeneralization between Cities or Industries, enabling us to analyze each City and Industry uniquely.

### 1.5.2 Weaknesses

Though we were able to consider each City and Industry, some of the assumptions we utilized may not be fully applicable in the real world. For instance, logistic growth, though attractive, is not necessarily valid, as it assumes a constant maximal remote readiness for each Industry. It certainly is possible that technological advancement in the coming years will enable greater transition towards remote work in Industries in which it is currently impossible to telework. Further, we assume linear changes in jobs for each City. Individual City growth is related in some manner towards overall population of the nation, which may have similar effects on all Industries. We were unable to account for this complexity in our model.

# 2 Part II: Remote Control

## 2.1 Overview

Part II asks us to create a model to predict the following:

1. Whether or not an individual worker whose job is remote-ready will be allowed to work from home.
2. In the case that such a worker is allowed to work from home, whether or not this worker will choose to work from home.

## 2.2 Assumptions

1. The companies that we consider have three different options for working that they can utilize: having the majority of employees work in-person, having a "hybrid" layout with some working in-person and some working from home, and having a majority of employees work from home.
2. The companies that we consider are remote-ready.
   - **Justification**: Per the prompt, we only need to consider these employers.
3. The primary traits that influence whether or not an employer in a specific Industry will allow employees to work from home, or whether or not an employee himself chooses to work from home, are, for each employee: their age; their number of children; their commuting time between home and workplace; their access to high speed Wi-Fi; and their personality type, which includes Conscientiousness and Neuroticism.
   - **Justification**: We use the Costa and McCrae Five Factor Model as a valid and reliable personality theory model to support the success of an employee. The first factor, Extraversion, determines an individual's sociability. The second factor, Agreeableness, determines an individual's friendliness or tactfulness. The third factor, Conscientiousness, determines an individual's organization skills, and the fourth factor, Neuroticism, determines an individual's anxiety. Lastly, the fifth factor, Openness to Experience, determines an individual's curiosity and open-mindedness. [8].
4. Unless specific data for Job Industries are available, we assume that workers in each particular Industry have Wi-Fi Speeds, Number of Children, Work Travel Times, Conscientiousness, and Neuroticism which are all representative of the overall population of the US.
   - **Justification**: Since these Job Industries have both a wide range of employees and a large number of locations, when a lack of data presents itself, the broader national average values and standard deviations are most likely equal to the averages and standard deviations within the Job Industries.
5. The average American is indifferent with regard to working from home or working in-person.
   - **Justification**: With such a large population and variability in job type across the US, it is reasonable to assume that opinions on remote work are generally split in half, meaning the average American will be approximately indifferent to working remotely or at home.

6. Unless data is available, we assume that random variables for our Monte-Carlo simulation are normally distributed. These random variables (for which no further data is available) include home Wi-Fi speeds, Work Travel Time, Conscientiousness, and Neuroticism.

7. Employers will always seek to optimize the economic productivity of their business ventures.

   - **Justification**: It is only natural that employers wish to increase their own profits.

## 2.3 Model Development

### 2.3.1 Parameters

1. **Industries ($\mathcal{J}$).** See Section 1.3.1.

2. **Job Industry ($J$).** See Section 1.3.1.

3. **The set of all employees working in Industry $J$, ($P_J$).**

4. **The set of WFH-determining traits ($\mathcal{T}$).** Per Assumption 3, we let the elements of $\mathcal{T}$ be Age, Number of Children, Commuting time between home and workplace, Access to high speed Wi-Fi, Conscientiousness, and Neuroticism.

5. **The set of possible trait assignments for each $p \in P_J$, ($\mathcal{L}$).** Every element in $\mathcal{L}$ is a 6-tuple $(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6)$ and represents the numerical value of each of the six traits that the person $p$ possesses. Later in the model, we formalize this definition more rigorously.

6. **The total productivity of an Industry $J$ where the majority of employees work in-person, in dollars of production ($I_J$).**

7. **The total productivity of an Industry $J$ where employees follow a "hybrid" model between working from home and in-person, in dollars of production ($H_J$).**

8. **The total productivity of an Industry $J$ where the majority of employees work from home, in dollars of production ($W_J$).**

9. **The overall probability that any individual worker whose job is remote-ready will be allowed to and will choose to work from home ($P_{WFH}$).**

### 2.3.2 Model Derivation

Generally speaking, we seek to answer the prompt by calculating probabilities stratified by Job Industry. More specifically, for each Job Industry $J$, we compute a probability that any employer in $J$ whose business is remote-ready will allow their employees to work from home. Per Assumptions 1 and 7, this is equivalent to finding the probability that the employers calculate a net economic productivity for when their employees follow a hybrid plan between working from home and working in-person, or follow a plan where the majority of their employees work from home, is greater than the net economic productivity for when the majority of their employees work in-person. We additionally compute the probability that each employee $p \in P_J$ chooses to work from home, given that their employer has given them the option to work from home. Together, multiplying these two probabilities together will give us the final probability that an individual worker whose job is remote-ready will be allowed to **and** will choose to work from home.

For each of the elements in $\mathcal{T}$, index them according to the following definition: Let $T_1$ represent the trait "Age," let $T_2$ represent the trait "Number of Children," let $T_3$ represent the trait "Commuting time between home and workplace," let $T_4$ represent "Access to high speed Wi-Fi," let $T_5$ represent "Conscientiousness," and let $T_6$ represent "Neuroticism."

Now, let $L : P_J \to \mathcal{L}$ be the function which assigns any person $p \in P_J$ a 6-tuple representing the value that person $p$ has of trait $T_i \in \mathcal{T}$ for each $i \in \{1, 2, \dots, 6\}$. We define the "values" for each trait $T_i$ as follows: If $L(p) = (\ell_1, \ell_2, \dots, \ell_6)$, then person $p$ has Age $\ell_1$ in years; they have $\ell_2$ children; the Commuting time between home and workplace is $\ell_3$ minutes; their Wi-Fi speed is $\ell_4$ megabits per second (Mbps); their Conscientiousness score, according to the Big Five Personality Test (per Assumption 4) and normalized out of five, is $\ell_5$; and, their Neuroticism score, according to the Big Five Personality Test and normalized out of five, is $\ell_6$.

We model our desired probabilities using a Monte-Carlo simulation. The algorithm that we use for our simulation generally follows a four-step process:

First, for an arbitrary employee $p_J \in P_J$ we assign them a random set of traits using $L$. Per Assumption 6, we create variability in traits $T_3$, $T_4$, $T_5$, and $T_6$ using normal distributions, with means and standard deviations taken from US National Averages as depicted in the table below:

| National Averages and Standard Deviations per Factors | | |
|---|---|---|
| | National Average | National Standard Deviation |
| Wi-Fi Speed (Mbps) | 127.55 [25] | 35.996 [25] |
| Work Travel Time (Min) | 24.49 | 4.6244 |
| Conscientiousness | 4.2 [9] | 1.08 [9] |
| Neuroticism | 3.58 [9] | 1.18 [9] |

For traits $T_1$ and $T_2$, we create variability using a probability density function (PDF), which we custom-build according to data that stratifies the two traits by each of the ten industries in $\mathcal{J}$. These PDFs are depicted in the table below:

| Proportion of Workers in Age Cohorts per Industry [22] | | | | |
|---|---|---|---|---|
| | 25-34 | 35-44 | 45-54 | 55-64 |
| Mining, logging, construction | 0.2635 | 0.303 | 0.2437 | 0.1898 |
| Manufacturing | 0.256 | 0.249 | 0.26 | 0.235 |
| Trade, transportation, and utilities | 0.311 | 0.24 | 0.231 | 0.218 |
| Information | 0.315 | 0.268 | 0.244 | 0.173 |
| Financial activities | 0.255 | 0.271 | 0.248 | 0.226 |
| Professional and business services | 0.302 | 0.275 | 0.234 | 0.189 |
| Education and health services | 0.227 | 0.225 | 0.219 | 0.175 |
| Leisure and hospitality | 0.213 | 0.16 | 0.13 | 0.101 |
| Religious social and community services | 0.138 | 0.186 | 0.198 | 0.241 |
| Government | 0.206 | 0.235 | 0.247 | 0.21 |

Second, we create a function $g : \mathcal{L} \to \mathbb{R}^{(0,1)}$, where, for any $\mathbf{a} \in \mathcal{L}$, we let $g(\mathbf{a}) = c_{\mathbf{a}}$ mean that the probability that a person $p$ satisfying $L(p) = \mathbf{a}$ will work from home, given that they have the option to do so, is $c_{\mathbf{a}}$. In other words, this function takes in an arbitrary combination of traits, and returns the probability that any person with this combination of traits will work from home (given that they have the choice). We call this probability the *propensity* for person $p$ to work from home. In order to create this function, we weight each of the traits in $\mathcal{T}$ according to how much they impact the propensity for person $p$ to work from home.

We assume that the Age and the two personality traits do not affect this computation. Now consider the time to commute between workplace and home. When this time is small, the propensity to work from home will be relatively low, as a small commuting time does not present any hindrance in the transportation between home and work. However, as this time increases, the propensity to work from home will increase exponentially. Therefore, we model this trait using an exponential function $T_T(\mathbf{a}) = b_T e^{N_T \ell_3}$ for constants $b_T$, $N_T$. We choose these constants such that the national average commuting time gives a propensity of exactly 0.5, per Assumption 5. Similarly, we model the traits "Number of children" and "Access to high speed Wi-Fi" similarly. Using the same line of reasoning as before, we obtain the functions $T_C(\mathbf{a}) = b_C e^{N_C \ell_2}$ (representing the model for the traits Number of children) and $T_W(\mathbf{a}) = 0.9 - b_W e^{N_W \ell_4}$ (representing the model for the traits Accesibility to high speed Wi-Fi). These three functions are depicted in the graphs below. The MATLAB code can be found in Appendix B.

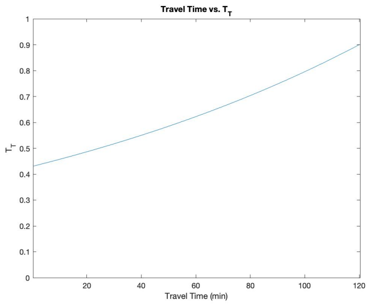


Figure 1: Travel Time vs. $T_T$.

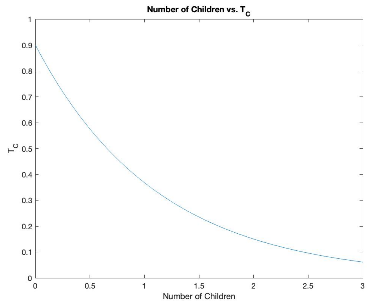


Figure 2: Children vs. $T_C$.

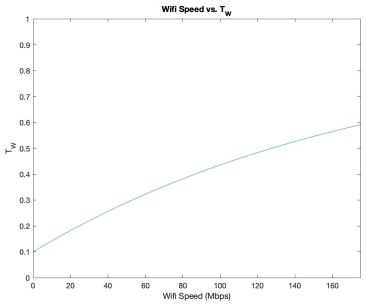


Figure 3: Wi-Fi Speed vs. $T_W$.

Finally, we let $g(\mathbf{a}) = \frac{1}{3}(T_T(\mathbf{a}) + T_C(\mathbf{a}) + T_W(\mathbf{a}))$, which is a linear combination of all three of the factors. Because all of our three functions output a number from 0 to 1 (we fix the domains so that this is true), this ensures that our output for $g(\mathbf{a})$ does indeed output a probability from 0 to 1.

Third, we calculate the relationship between the two functions $h_{WFH} : \mathcal{L} \to \mathbb{R}$ and $h_{IP} : \mathcal{L} \to \mathbb{R}$, which calculates the net productivity of person $p$ when they work from home, and the net productivity of a person $p$ when they work in-person, respectively. More specifically, we compute the ratio $H(\mathbf{a}) = \frac{h_{WFH}(\mathbf{a})}{h_{IP}(\mathbf{a})}$, which represents the ratio of the relative efficiency of person $p$ in both environments.

We construct $H(\mathbf{a})$ according to the fact that higher values of this function should correspond to a higher work-from-home productivity compared to the individual's productivity in-person; also, when running our Monte-Carlo simulation, we want this function to have a mean value of around 1, because this represents the average American worker (per Assumption 5).

We assume that the trait "Commuting time between home and workplace" for each employee does not affect the employer's decision to allow employees to work from home. Now, we consider the traits "Age," "Number of Children," "Access to high speed Wi-Fi," "Conscientiousness," and "Neuroticism." Consider the following equation:

$$k(\mathbf{a}) = \frac{1}{4}\left[\left(1 - \frac{\ell_6 - \mu_{\ell_6}}{0.5}\right) + \left(1 + \frac{\ell_5 - \mu_{\ell_5}}{0.5}\right) + \left(1 - \frac{\ell_2 - \mu_{\ell_2}}{0.45}\right) + \left(1 + \frac{\ell_4 - \mu_{\ell_4}}{36}\right)\right]. \tag{5}$$

This parameter approximately represents the value that we desire out of $H(\mathbf{a})$. For example, consider the term inside of the sum with index 6, which corresponds to the trait "Neuroticism." Whenever the difference $d_6 = \ell_6 - \mu_{\ell_6}$ is positive, where $\mu_{\ell_6}$ represents the national numeric average of the trait "Neuroticism," this indicates that our individual is more neurotic than average. Because more neurotic individuals are defined to be more emotionally unstable, we subtract this value from 1 to indicate that their productivity at home decreases relative to their productivity in the workplace [23]. We divide $d_6$ by half of the standard deviation of $\ell_6$ (across the entire US population) to "normalize" each variable, which is meant to be analogous to a typical $z$-score computation.

To compute $H(\mathbf{a})$ from $k(\mathbf{a})$, we "dampen" the effect of each variable according to a damping function $D(\mathbf{a})$ defined as follows:

$$D(\mathbf{a}) = \begin{cases} 1 + \frac{|k-1|}{e^{\frac{\ell_1 - 20}{100}}} & \text{when } k > 1, \\ 1 - \frac{|k-1|}{e^{\frac{\ell_1 - 20}{100}}} & \text{when } k < 1. \end{cases} \tag{6}$$

We are motivated to construct this damping function by the fact that as the age (recall that this is characterized by the parameter $\ell_1$) of an individual $p$ increases, the effect of the variability of the various traits $T_i$ on that person should decrease, because they are more experienced. We thus "dampen" the effects of variability in these traits by the exponential function $e^{\frac{\ell_1-20}{100}}$, which implies that for every increase in age of 100 years, the quantity $|1 - k(\mathbf{a})|$ (representing the difference between $k(\mathbf{a})$ and 1, which we expect the mean of $H(\mathbf{a})$ to be) is reduced by a factor of $e$.

The fourth and final step in our Monte Carlo process is to generate our final probabilities. By definition,

$$I_J = \sum_{p \in P_J} \sum_{\substack{\mathbf{a} \in \mathcal{L} \\ L(P_J) = \mathbf{a}}} h_{IP}(\mathbf{a}). \tag{7}$$

Similarly,

$$\begin{aligned} H_J &= \sum_{p \in P_J} \sum_{\substack{\mathbf{a} \in \mathcal{L} \\ L(P_J) = \mathbf{a}}} \mathbb{E}[\text{productivity}(P_J)] \\ &= \sum_{p \in P_J} \sum_{\substack{\mathbf{a} \in \mathcal{L} \\ L(P_J) = \mathbf{a}}} g(\mathbf{a}) \cdot h_{WFH}(\mathbf{a}) + (1 - g(\mathbf{a})) \cdot h_{IP}(\mathbf{a}). \end{aligned} \tag{8}$$

Finally,

$$W_J = \sum_{p \in P_J} \sum_{\substack{\mathbf{a} \in \mathcal{L} \\ L(P_J) = \mathbf{a}}} h_{WFH}(\mathbf{a}). \tag{9}$$

Note that $g(\mathbf{a})$ represents the probability that a person $p$ with set of traits $\mathbf{a} \in \mathcal{L}$ will work from home, given the choice to do so (recall that this value is also the propensity that they will work from home). Now, let $A_J$ be the condition that a particular Job Industry $J$ allows workers to work from home. Then,

$$P(A_J) = P((H_J > I_J) \text{ or } (W_J > I_J)). \tag{10}$$

Dividing each of the expressions given by Equations (7), (8), and (9) by quantity $h_{IP}(\mathbf{a})$, this probability statement is equivalent to calculating the probability that either

$$\sum_{p \in P_J} \sum_{\substack{\mathbf{a} \in \mathcal{L} \\ L(P_J) = \mathbf{a}}} g(\mathbf{a}) \cdot H(\mathbf{a}) + (1 - g(\mathbf{a})) > \sum_{p \in P_J} \sum_{\substack{\mathbf{a} \in \mathcal{L} \\ L(P_J) = \mathbf{a}}} 1, \tag{11}$$

corresponding to the inequality $H_J > I_J$, or that

$$\sum_{p \in P_J} \sum_{\substack{\mathbf{a} \in \mathcal{L} \\ L(P_J) = \mathbf{a}}} H(\mathbf{a}) > \sum_{p \in P_J} \sum_{\substack{\mathbf{a} \in \mathcal{L} \\ L(P_J) = \mathbf{a}}} 1, \tag{12}$$

corresponding to the inequality $W_J > I_J$. Our simulation tells us the proportion of employers who have some set of traits $\mathbf{a} \in \mathcal{L}$ that satisfy either Inequality (11) or Inequality (12). To calculate $P(A_J)$, we multiply this value by a pre-assigned job coefficient, determined by the proportion of available jobs in each Industry that are deemed remote-ready (see the table below), according to [1].

| Proportion of Workers in Age Cohorts per Industry [1] | |
|---|---|
| Industry | Job Coefficient |
| LM | 0 |
| MA | 0 |
| TT | 0.03 |
| IN | 0.76 |
| FA | 0.88 |
| PB | 0.88 |
| EH | 0.98 |
| LH | 0.26 |
| OS | 0.37 |
| GO | 0.37 |

Finally, to compute $P_{WFH}$, we sum our calculated value of $P(A_J)$ multiplied by the probability that each individual worker will choose to work from home, across all $p \in P_J$. In other words,

$$P_{WFH} = \frac{\sum_{p \in P_J} \sum_{\substack{\mathbf{a} \in \mathcal{L} \\ L(P_J) = \mathbf{a}}} g(\mathbf{a} \cdot P(A_J))}{|P_J|} \approx \mu_{g(\mathbf{a})} \cdot P(A_J), \tag{13}$$

where $\mu_{g(\mathbf{a})}$ represents the mean value of $g(\mathbf{a})$ across all workers in $P_J$.

### 2.4 Results and Discussion

For the sake of brevity, we don't showcase all of the histograms for all of the industries in $\mathcal{J}$ and instead display just the *Information*, *Professional and Business Services*, and *Leisure and Hospitality* industries. For each of these, we plot $h(\mathbf{a})$, $I_J - H_J$, and $I_J - W_J$ in the figures below. The Python code can be found in Appendix C.

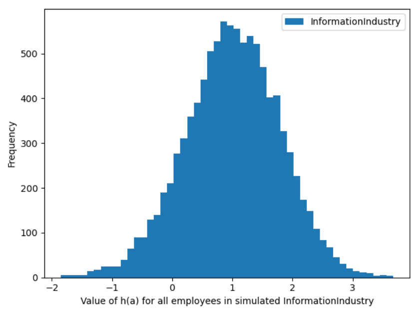


Figure 4: $h(\mathbf{a})$ histogram for the *Information* Industry.

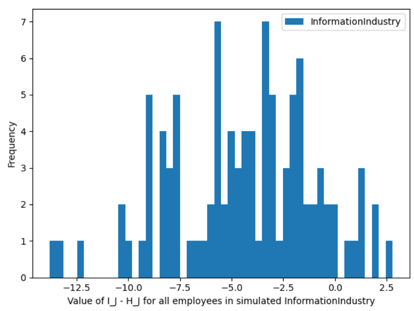


Figure 5: $I_J - H_J$ histogram for the *Information* Industry.

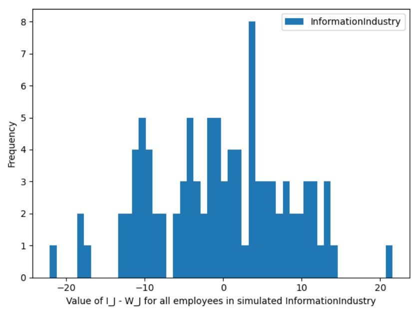


Figure 6: $I_J - W_J$ histogram for the *Information* Industry.

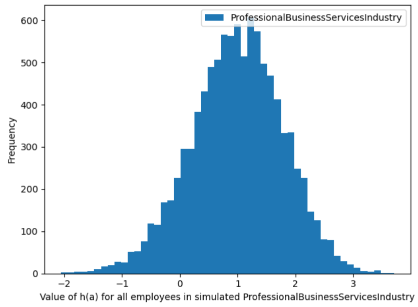


Figure 7: $h(\mathbf{a})$ histogram for the *Professional and Business Services* Industry.

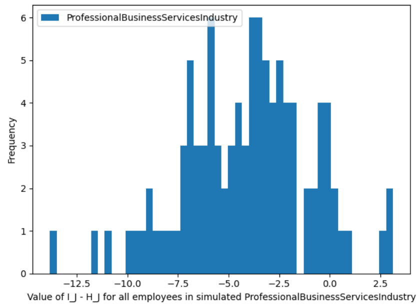


Figure 8: $I_J - H_J$ histogram for the *Professional and Business Services* Industry.

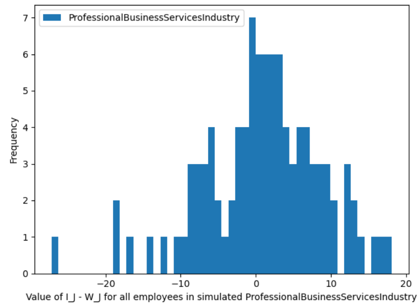


Figure 9: $I_J - W_J$ histogram for the *Professional and Business Services* Industry.

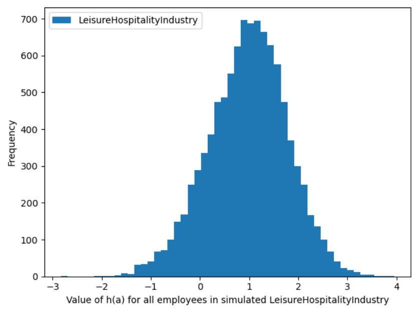


Figure 10: $h(\mathbf{a})$ histogram for the *Leisure and Hospitality* Industry.

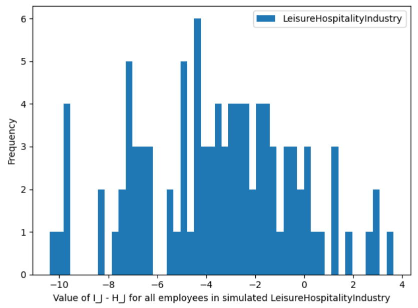


Figure 11: $I_J - H_J$ histogram for the *Leisure and Hospitality* Industry.

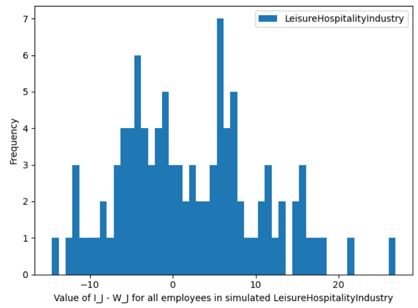


Figure 12: $I_J - W_J$ histogram for the *Leisure and Hospitality* Industry.

The following tables detail our results for all the industries:

| $g(a)$ Mean and Standard Deviation per Industry) | | |
|---|---|---|
| $J$ | $\mu$ | $\sigma$ |
| LM | 0.3996 | 0.1111 |
| MA | 0.4005 | 0.1112 |
| TT | 0.4022 | 0.1105 |
| IN | 0.3992 | 0.1111 |
| FA | 0.4004 | 0.1113 |
| PB | 0.3992 | 0.1112 |
| EH | 0.4003 | 0.1109 |
| LH | 0.3970 | 0.1121 |
| OS | 0.4013 | 0.1109 |
| GO | 0.4004 | 0.1109 |

Figure 13: $g(\mathbf{a})$ values for each of the Industries.

| $H(a)$ Mean, Standard Deviation per Industry) | | |
|---|---|---|
| $J$ | $\mu$ | $\sigma$ |
| LM | 0.9959 | 0.7890 |
| MA | 0.9955 | 0.7787 |
| TT | 0.9961 | 0.7851 |
| IN | 1.0040 | 0.7920 |
| FA | 0.9963 | 0.7915 |
| PB | 0.9908 | 0.7929 |
| EH | 0.9933 | 0.7953 |
| LH | 0.9744 | 0.8039 |
| OS | 0.9962 | 0.7739 |
| GO | 1.0010 | 0.7746 |

Figure 14: $H(\mathbf{a})$ values for each of the Industries.

| Probability employer will allow working from home | |
|---|---|
| $J$ | $P_J(A)$ |
| LM | 0 |
| MA | 0 |
| TT | 0.0160 |
| IN | 0.7269 |
| FA | 0.8264 |
| PB | 0.8446 |
| EH | 0.8800 |
| LH | 0.2009 |
| OS | 0.3235 |
| GO | 0.3235 |

Figure 15: $P(A_J)$ across all $J \in \mathcal{J}$.

| Probability employee chooses work from home | |
|---|---|
| $J$ | $P_{WFH}$ |
| LM | 0 |
| MA | 0 |
| TT | 0.0065 |
| IN | 0.2902 |
| FA | 0.3309 |
| PB | 0.3372 |
| EH | 0.3523 |
| LH | 0.0798 |
| OS | 0.1298 |
| GO | 0.1295 |

Figure 16: $P_{WFH}$ across all $J \in \mathcal{J}$.

Above, we can see that the Industry with the highest probability for an employer to work from home is in Financial Activities. Furthermore, the Industry where an employee will choose to work from home given the choice is in Education and Health Services.

## 2.5 Strengths and Weaknesses

### 2.5.1 Strengths

One large strength of this model is that it accurately considers both the employer side and the employee side of the remote-work dilemma. We are able to fully model the probability that an employer will select an in-person, hybrid, or remote work setting, as we account for all of the major factors employers consider when making such a decision, ranging from emotional health to Wi-Fi speeds. Additionally, on the employee side, by accounting for the main factors that go into an employee's decision to stay at home, given they have the choice, we can make a model that can predicts propensity of an employee to work from home. Furthermore, by simulating all results over a multitude of offices and a multitude of people within those offices, we have created a sensitivity analysis of our own work. In fact, utilizing such randomness in calculating our final results yields much more concrete and usable values.

### 2.5.2 Weaknesses

Many of the data points used within the models are an oversimplification because of a lack of data. Although we make the assumption that each Job Industry is representative of the larger US, this may not always be the case. As such, the results produced by our model may be somewhat skewed towards the general US population and away from the true value that may exist within a Job Industry. We assume that people are independent; that is, the probability that any person works from home on any given day does not influence the probability that another person works from home. Ultimately, we assume that the seamlessness of technology obfuscates the impact of this variable on the overall probability. Additionally, we do not account for the relative importance of each factor in deciding the propensity for a worker to work from home when given the option, instead weighting them equally.

# 3 Part III: Just a Little Home-work

## 3.1 Overview

We use our results from Part I and Part II to estimate the percentage of workers who will work remotely. Next, we consider the three factors of environmental impact, economic impact, and happiness impact to quantify the total impact of remote work on a particular City. This enables us to make predictions for the same Cities considered in Part I for 2024 and 2027; we use these predictions to rank the Cities in terms of the magnitude of impact that remote work will have on the City.

## 3.2 Assumptions

1. Environmental impact of remote work is heavily based on travel time for commuting.
   - **Justification**: It is reasonable to assume that the environmental impact of remote work is mainly based upon the air pollution from vehicle greenhouse gas emissions, including nitrogen oxides, carbon monoxide, and particulate matter [7] [6].
2. The economic impact of remote work is heavily based on the change in productivity as a result of remote work.
   - **Justification**: Remote work affects the work dynamic of the labor force because of the changes in physical surroundings of a worker's "office." This leads to a change in productivity of a worker, and the sum of all these changes in productivity of a workforce in a particular City greatly affects the economic output of that City [26].
3. Happiness impact of remote work is heavily based on the happiness levels self-reported and collected by polls in 2019 and 2021.
   - **Justification**: Happiness as a construct is difficult to objectively measure, so we base our data collection on collected data from polls in the US and UK. Remote work has often been linked to a change in mental health for many people due to drastic changes in lifestyles during the pandemic. Thus, happiness levels of a region are substantially changed as a result of remote work, since a large percentage of a region's population is made up of the workforce.
4. Happiness in a state is constant throughout the state in the US.
   - **Justification**: A state in the US is a small enough region for the variance in happiness level to be relatively minimal. Further, due to unavailability of data at the county or City level within the US, we had to rely on this for meaningful results.

## 3.3 Model Development

### 3.3.1 Parameters

1. **Cities** ($\mathcal{C}$). See Section 1.3.1.
2. **City** ($C$). See Section 1.3.1.
3. **Percentage of remote-ready workers in a City** ($P_{RR}$).
4. **Probability a worker will choose to work from home in a City** ($P_{WFH}$).

5. **Percentage of workers who choose remote work in a City ($P_{RW}$).**
6. **Environmental Index ($E$).** The environmental impact of remote work on a City.
7. **Mean commute time ($M_C$).** The mean commute time for workers in some City $C \in \mathcal{C}$ for in-person work.
8. **Economic Index ($N$).** The economic impact of remote work on a City.
9. **Happiness Index ($H$).** The happiness impact of remote work on a City.
10. **Total Impact ($TI$).** The total impact of remote work for a particular City.

### 3.3.2 Model Derivation

We first used our results from Part I and Part II to determine the percentage of workers who will work remotely. We multiply $P_{RR}$ by $P_{WFH}$ which results in $P_{RW}$, because the proportion of remote-ready workers in a City and the probability a worker will choose to work from home in a City are two independent probabilities. Multiplying the two yields the proportion of workers who will choose remote work in a City.

$$P_{RR} \cdot P_{WFH} = P_{RW}. \tag{14}$$

$P_{RW}$ becomes a key component of the second section of Part III. To quantify the total impact of remote work on a City, we create three indexes that measure the change in environmental, economic, and happiness levels in the City before and after a period of remote work. Thus we look at changes from 2019 to 2021, which is the time when remote work was at its maximum in the peak of the pandemic. Next, we create our total impact equation, incorporating $P_{RW}$ as a weight for our three factors, as the total impact largely depends on the actual number of people who choose remote:

$$TI = E + N + H. \tag{15}$$

To determine E, our environmental index, we use mean travel times to commute to work [1]. The main cause of environmental pollution is due to employee travel vehicle emissions, so the fewer the people who have to travel, the better the environmental result [7]. We determine the environmental effect as the proportion of workers who must travel to the office multiplied by the mean travel time per worker, determining the average minutes traveled every day among all workers. Hence, the average length of commute for all workers is

$$(1 - P_{RW}) \cdot L_C.$$

We scale these such that the least amount of average minutes traveled gains the higher E score of 1.

To determine N, our economic index, we consider the change in productivity from before and after the pandemic [26]. This increase in productivity only occurs for workers who remain in remote work, meaning that $N \propto P_{RW}$. In fact, we can state that the increase in total productivity in each City is $1.45 \cdot P_{RW}$.

To determine H, we utilize the World Happiness Index. The World Happiness Index only measures by country, so we must scale it for each individual City. As per Assumption 4, we use the relative happiness indices for Washington (59.92), Nebraska (59.54), and Pennsylvania (53.18) [10] as the indices for Seattle, Omaha, and Scranton, respectively. We scale these happiness scores relative to 72.94 [10], the rating of Utah (the highest rated state) to determine the relative happiness score of each State based on the World Happiness Index Ranking of the US (7.028) [24]. We use a similar process for England, using the local results for Liverpool (7.23) and the Vale of Glamorgan (7.77) [5] to represent Liverpool and Barry, respectively. These are scaled relative to 8.42 [5], the score for South Northamptonshire, the highest within the UK, and then multiplied to the score for the UK on the World Happiness Index (6.798) [24]. Normalizing these scores relative to the maximum City score as a score of 1, we can determine the Happiness Index for each City.

### 3.4 Results and Discussion

Plugging in our values for $P_{RR}$ and $P_{WFH}$ from Part I and Part II respectively, we can compute the impact of remote work on each City for 2024 and 2027. These results are shown in Tables 3.1 and 3.2.

| Final Rankings for Each City, 2024 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| City | $P_{RR}$ | $P_{WFH}$ | $P_{RW}$ | E | N | H | TI | Rank |
| Seattle | 0.4111 | 0.1729 | 0.07106 | 0.6683 | 0.9518 | 0.9578 | 2.578 | 3 |
| Omaha | 0.4359 | 0.1713 | 0.07466 | 1 | 1 | 1 | 3 | 1 |
| Scranton | 0.3563 | 0.1686 | 0.06001 | 0.8806 | 0.8048 | 0.7188 | 2.4042 | 4 |
| Liverpool | 0.2812 | 0.2024 | 0.05691 | 0.7173 | 0.7623 | 0.7756 | 2.2551 | 5 |
| Barry | 0.4048 | 0.1676 | 0.06784 | 0.8285 | 0.9086 | 0.9935 | 2.7307 | 2 |

*Table 3.1: Final rankings for each City, 2024.*

| Final Rankings for Each City, 2027 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| City | $P_{RR}$ | $P_{WFH}$ | $P_{RW}$ | E | N | H | TI | Rank |
| Seattle | 0.4113 | 0.1729 | 0.07110 | 0.6681 | 0.9473 | 0.8830 | 2.4983 | 3 |
| Omaha | 0.4382 | 0.1713 | 0.07505 | 1 | 1 | 0.9263 | 2.9263 | 1 |
| Scranton | 0.3606 | 0.1686 | 0.06081 | 0.8809 | 0.8102 | 0.6704 | 2.3615 | 4 |
| Liverpool | 0.2818 | 0.2024 | 0.05703 | 0.7170 | 0.7599 | 0.5905 | 2.0675 | 5 |
| Barry | 0.4080 | 0.1676 | 0.06838 | 0.8287 | 0.9110 | 1 | 2.7397 | 2 |

*Table 3.2: Final rankings for each City, 2027.*

Based on our results, the impact of remote work is most positive in Ohama and has the least magnitude in Liverpool.

### 3.5 Sensitivity Analysis

We perform a sensitivity analysis on our results by varying our computed values of $P_{WFH}$ by 5%. These new values are shown in Tables 3.3 and 3.4.

| Sensitivity Analysis for 2024 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| +5% | E | N | H | TI | -5% | E | N | H | TI |
| Seattle | 0.6681 | 0.9518 | 0.9578 | 2.5778 | Seattle | 0.6684 | 0.9518 | 0.9578 | 2.5780 |
| Omaha | 1 | 1 | 1 | 3 | Omaha | 1 | 1 | 1 | 3 |
| Scranton | 0.8799 | 0.8048 | 0.7188 | 2.4035 | Scranton | 0.8814 | 0.8048 | 0.7188 | 2.4050 |
| Liverpool | 0.7165 | 0.7623 | 0.7756 | 2.2544 | Liverpool | 0.7180 | 0.7623 | 0.7756 | 2.2558 |
| Barry | 0.8282 | 0.9086 | 0.9935 | 2.7304 | Barry | 0.8289 | 0.9086 | 0.9935 | 2.7310 |

*Table 3.3: Sensitivity analysis for new values in 2024 with ± 5%.*

| Sensitivity Analysis for 2027 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| +5% | E | N | H | TI | -5% | E | N | H | TI |
| Seattle | 0.6679 | 0.9473 | 0.8830 | 2.4982 | Seattle | 0.6682 | 0.9473 | 0.8830 | 2.4985 |
| Omaha | 1 | 1 | 0.9263 | 2.9263 | Omaha | 1 | 1 | 0.9263 | 2.9263 |
| Scranton | 0.8802 | 0.8102 | 0.6704 | 2.3608 | Scranton | 0.8817 | 0.8102 | 0.6704 | 2.3623 |
| Liverpool | 0.7163 | 0.7599 | 0.5905 | 2.0668 | Liverpool | 0.7178 | 0.7599 | 0.5905 | 2.0682 |
| Barry | 0.8283 | 0.9110 | 1 | 2.7394 | Barry | 0.8290 | 0.9110 | 1 | 2.7400 |

*Table 3.4: Sensitivity analysis for new values in 2027 with ±5%.*

Based on our results, though individual scores have changed, the overall ranking of the Cities has not. Hence, our model is robust and able to account for minute changes associated with sampling variation. We can thus say with greater confidence that the rankings obtained by our model are accurate.

## 3.6 Strengths and Weaknesses

### 3.6.1 Strengths

This model encompasses micro and macro impacts of remote work on a person as a representative of a City. For example, mental health is considered from a micro perspective which is taken into account in the happiness index. In addition, environmental impact is considered on a macro scale as vehicle emissions contribute to global greenhouse gas emissions. The three factors of environmental, economic, and happiness forecast overall remote work impact as a balanced whole. Our consideration of these factors emphasizes different aspects that contribute to the ranking of a City, allowing for a holistic approach that doesn't emphasize one single factor.

### 3.6.2 Weaknesses

This model equally weighs environmental, economic, and happiness as indexes of the total impact equation, which may not accurately reflect reality. In reality, each of these factors of remote work will vary in terms of how much they impact a City because of a City's situational factors, such as industrialization rate, population, and other demographics. For example, an extremely industrialized City already had high vehicle pollution rates before the pandemic, so an extreme reduction in vehicle traffic results in a drastic positive change in air pollution rates. On the other hand, a less industrialized City had low vehicle pollution rates before the pandemic so a reduction in vehicle traffic results in a much smaller or even negligible change in air pollution rates. With further analysis, it may be possible to ascribe additional weights to these factors to create a more sensitive and nuanced model.

Further, while we attempt to account for most major factors, we recognize that our factors may not account for all relevant influences on the magnitude of impact of remote work. However, due to the construction of our model, we anticipate that further new factors can be added without much work, allowing for easy adaptation with future considerations.

# 4 References

## References

[1] Remote Work: Fad or Future, MathWorks Math Modeling Challenge 2022. URL https://m3challenge.siam.org/node/559.

[2] Social Security Administration. National Average Wage Index. URL https://www.ssa.gov/oact/cola/AWI.html.

[3] United States Census Bureau. Who Are Our Health Care Workers? URL https://www.census.gov/library/stories/2021/04/who-are-our-health-care-workers.html.

[4] National Center for Education Statistics. Teacher characteristics and trends. URL https://nces.ed.gov/fastfacts/display.asp?id=28.

[5] Office for National Statistics. Loneliness rates and well-being indicators by local authority. URL https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/datasets/lonelinessratesandwellbeingindicatorsbylocalauthority.

[6] World Wildlife Foundation. Reducing the impact of commuting. URL https://www.worldwildlife.org/magazine/issues/summer-2017/articles/reducing-the-impact-of-commuting#:~:text=It's%20notoriously%20bad%20for%20our,into%20the%20atmosphere%20every%20year.

[7] A. Gaskell. How eco-friendly is remote working? URL https://www.forbes.com/sites/adigaskell/2021/10/21/how-eco-friendly-is-remote-working/#:~:text=The%20study%20found%20that%20working,emissions%2C%20by%20around%2010%25.

[8] L. R. Few, J. D. Miller, J. Q. Morse, K. E. Yaggi, S. R. Reynolds, and P. A. Pilkonis. Examining the reliability and validity of clinician ratings on the five-factor model score sheet. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3682410/.

[9] L. G. Elleman, D. M. Condon, S. E. Russin, and W. Revelle. The personality of U.S. states: Stability from 1999 to 2015. URL https://www.sciencedirect.com/science/article/pii/S009265661630071X.

[10] A. McCann. Happiest States in America. URL https://wallethub.com/edu/happiest-states/6959.

[11] National Bureau of Economic Research. How Many Jobs Can Be Done at Home? URL https://www.nber.org/system/files/working_papers/w26948/w26948.pdf.

[12] U.S. Bureau of Labor Statistics. Education and Health Services. URL https://www.bls.gov/iag/tgs/iag65.htm.

[13] U.S. Bureau of Labor Statistics. Financial Activities. URL https://www.bls.gov/iag/tgs/iag50.htm.

[14] U.S. Bureau of Labor Statistics. Information: NAICS 51. URL https://www.bls.gov/iag/tgs/iag51.htm.

[15] U.S. Bureau of Labor Statistics. Leisure and Hospitality. URL https://www.bls.gov/iag/tgs/iag70.htm.

[16] U.S. Bureau of Labor Statistics. Manufacturing: NAICS 31-33. URL https://www.bls.gov/iag/tgs/iag31-33.htm.

[17] U.S. Bureau of Labor Statistics. Natural Resources and Mining. URL https://www.bls.gov/iag/tgs/iag10.htm.

[18] U.S. Bureau of Labor Statistics. Other Services (except Public Administration): NAICS 81. https://www.bls.gov/iag/tgs/iag81.htm.

[19] U.S. Bureau of Labor Statistics. Professional and Business Services. URL https://www.bls.gov/iag/tgs/iag60.htm.

[20] U.S. Bureau of Labor Statistics. Trade, Transportation, and Utilities. URL https://www.bls.gov/iag/tgs/iag40.htm.

[21] U.S. Bureau of Labor Statistics. Teleworking and lost work during the pandemic: New evidence from the CPS. URL https://www.bls.gov/opub/mlr/2021/article/teleworking-and-lost-work-during-the-pandemic-new-evidence-from-the-cps.htm.

[22] U.S. Bureau of Labor Statistics. Labor Force Statistics from the Current Population Survey. URL https://www.bls.gov/cps/cpsaat18b.htm.

[23] University of Minnesota. Conscientiousness is top personality predictor of positive career and work-related outcomes, has broad benefits. URL https://twin-cities.umn.edu/news-events/conscientiousness-top-personality-predictor-positive-career-and-work-related-outcomes#:~:text=A%20study%20from%20the%20University,that%20best%20predicts%20work%2Drelated.

[24] World Population Review. Happiest Countries in the World 2022. URL https://worldpopulationreview.com/country-rankings/happiest-countries-in-the-world.

[25] World Population Review. Average Internet Speed by State 2022. URL https://worldpopulationreview.com/state-rankings/average-internet-speed-by-state.

[26] Apollo Technical. Surprising Working from Home Productivity Statistics (2022). URL https://www.apollotechnical.com/working-from-home-productivity-statistics/#:~:text=Several%20studies%20over%20the%20past,and%20are%2047%25%20more%20productive.

[27] Toptal. Remote Work Readiness Assessment. URL https://www.toptal.com/remote-readiness.

# Appendix A: Regression Constants

For $W_J(Y) = \dfrac{a}{1 + b \cdot c^{-d(Y-2006)}}$, which is Equation (2):

| Coefficients for Nonlinear Regression for each Industry | | | | | |
|---|---|---|---|---|---|
| | a | b | c | d | $R^2$ |
| LM | 43.2566 | 0.7867 | 1.0813 | 1.0016 | 0.99 |
| MA | 80 | 2.8025 | 1.0317 | 0.9999 | 0.98 |
| TT | 80 | 3.4951 | 1.0327 | 1.0000 | 0.97 |
| IN | 80 | 2.0732 | 1.0627 | 0.9990 | 0.98 |
| FA | 80 | 2.3401 | 9.26 | 0.0227 | 0.97 |
| PB | 80 | 0.2336 | 1.0421 | 0.9997 | 0.98 |
| EH | 80 | 2.9249 | 1.0353 | 0.9999 | 0.99 |
| LH | 80 | 5.8920 | 1.0334 | 0.9997 | 0.85 |
| OS | 80 | 3.586 | 1.0395 | 1.0002 | 0.99 |

For $r(J(Y)) = \dfrac{c}{1 + ae^{-b \cdot W_J(Y)}}$, which is Equation (3):

| Coefficients for Logistic Regression for Remote Readiness | | | |
|---|---|---|---|
| | a | b | c |
| LM | N/A | N/A | 0 |
| MA | 0.0670 | 0.0857 | 0.01 |
| TT | 0.6699 | 0.0968 | 0.28 |
| IN | 0.1788 | 0.0275 | 0.65 |
| FA | 0.0178 | 0.0304 | 0.88 |
| PB | 0.1778 | 0.0333 | 0.88 |
| EH | 0.6705 | 0.0846 | 0.1935 |
| LH | 1.4528 | 0.1806 | 0.26 |
| OS | 1.0484 | 0.1082 | 0.37 |

For $N_C(J, Y) = mY + b$:

| Coefficients for Seattle | | |
|---|---|---|
| | m | b |
| LM | 1076 | -2057600 |
| MA | -887 | 1927000 |
| TT | 3772 | -7242300 |
| IN | 2399 | -4728000 |
| FA | -260 | 620491 |
| PB | 4463 | -8725300 |
| EH | 5217 | -10256000 |
| LH | 3228 | -6318900 |
| OS | 1034 | -2011700 |
| GO | 2037 | -3834200 |

| Coefficients for Omaha | | |
|---|---|---|
| | m | b |
| LM | 283 | -543176 |
| MA | -97 | 228104 |
| TT | -539 | 1182200 |
| IN | -237 | 488009 |
| FA | 517 | -999486 |
| PB | 779 | -1499800 |
| EH | 1337 | -2618000 |
| LH | 578 | -1117000 |
| OS | 217 | -418299 |
| GO | 614 | -1718000 |

| Coefficients for Scranton | | |
|---|---|---|
| | m | b |
| LM | -18 | 47028 |
| MA | -890 | 182070 |
| TT | 415 | -774578 |
| IN | -229 | 466114 |
| FA | -44 | 100642 |
| PB | 358 | -693379 |
| EH | 470 | -894423 |
| LH | 216 | -412782 |
| OS | -83 | 176139 |
| GO | -138 | 308384 |

| Coefficients for Liverpool | | |
|---|---|---|
| | m | b |
| LM | 539 | -942707 |
| MA | 1354 | -2641700 |
| TT | 3900 | -7728700 |
| IN | 742 | -1426700 |
| FA | 64 | -106630 |
| PB | 119 | -198027 |
| EH | -373 | 775442 |
| LH | -235 | 537402 |
| OS | 513 | -956695 |
| GO | -373 | 775442 |

| Coefficients for Barry | | |
|---|---|---|
| | m | b |
| LM | 53 | -101661 |
| MA | -75 | 155405 |
| TT | -26 | 54270 |
| IN | -19 | 42635 |
| FA | 49 | -94307 |
| PB | 90 | -175142 |
| EH | 89 | -169197 |
| LH | 229 | -449921 |
| OS | 57 | -110787 |
| GO | 89 | -169197 |

## Appendix B: MATLAB Code Appendix

```
fplot(@(x) .9*exp(-.894652*x),[0 3])
ylim([0 1])
xlabel('Number of Children')
ylabel('T_C')
title('Number of Children vs. T_C')
fplot(@(x) .430046*exp(.006154*x),[0 120])
ylim([0 1])
xlabel('Travel Time (min)')
ylabel('T_T')
title('Travel Time vs. T_T')
fplot(@(x) .9-.8*exp(-.005434*x),[0 175])
ylim([0 1])
xlabel('Wifi Speed (Mbps)')
ylabel('T_W')
title('Wifi Speed vs. T_W')
```

Listing 1: M3GraphsPart2.m

## Appendix C: Python Code Appendix

Please note that when calculating answers for the different Cities in $\mathcal{C}$ during the modeling of Part III: Just a Little Home-work, we simply made copies of **industryTraits.py** with varying values for the City's demographics, and imported the new file in **model.py** and **classes.py**. For the sake of brevity, we do not include the file modifications in this appendix and just include the generic, non-modified code used in Part II: Remote Control.

```
import numpy as np
from classes import Simulation
from industryTraits import *  # this imports all the classes in industryTraits, which we will use for our simulation
import matplotlib.pyplot as plt  # this imports the plotting library matplotlib

# we create an instance for each industry that we have by using the corresponding child of IndustryTraits
industries = [MiningLoggingConstructionIndustryTraits(), ManufacturingIndustryTraits(),
              TradeTransportationUtilitiesIndustryTraits(), InformationIndustryTraits(),
              FinancialActivitiesIndustryTraits(), ProfessionalBusinessServicesIndustryTraits(),
              EducationHealthServicesIndustryTraits(), LeisureHospitalityIndustryTraits(),
              ReligiousSocialCommunityServicesIndustryTraits(), GovernmentIndustryTraits()]

# a list comprehension is used to create Simulation instances
sims = [Simulation(ind, 100, 100) for ind in industries]
# a list to store the results
results = []
for i in range(len(industries)):
    sims[i].generate_employers()
    sims[i].run_employers()
    results.append(sims[i].determine_p())

# code used for plotting
for i in [3, 5, 7]:
    plt.hist(results[i][2], bins=50, label=f'{str(type(industries[i]))[23:-8]}')
    plt.xlabel(f"Value of h(a) for all employees in simulated {str(type(industries[i]))[23:-8]}")
    plt.ylabel("Frequency")
    plt.legend()
    plt.show()
    plt.clf()
    plt.hist(np.subtract(results[i][3], results[i][4]), bins=50, label=f'{str(type(industries[i]))[23:-8]}')
    plt.xlabel(f"Value of I_J - H_J for all employees in simulated {str(type(industries[i]))[23:-8]}")
    plt.ylabel("Frequency")
    plt.legend()
    plt.show()
    plt.clf()
    plt.hist(np.subtract(results[i][3], results[i][5]), bins=50, label=f'{str(type(industries[i]))[23:-8]}')
```

```
    plt.xlabel(f"Value of I_J - W_J for all employees in simulated {str(type(industries
    [i]))[23:-8]}")
    plt.ylabel("Frequency")
    plt.legend()
    plt.show()

# code used for table creation and data output
for i in range(len(industries)):
    print("yeas")
    print(results[i][0])
    print("G(a)")
    print(np.mean(results[i][1]))
    print(np.std(results[i][1]))
    print("H(a)")
    print(np.mean(results[i][2]))
    print(np.std(results[i][2]))
```

Listing 2: model.py

```
import numpy as np  # numpy is a standard import to help us do mathematical operations
    easily and efficiently
from industryTraits import IndustryTraits  # this imports the IndustryTraits class so
    that we can generate traits for
                                           # the simulation participants
from betadist import beta  # this is the import for the sigmoid-esque beta distribution
     function
import tqdm  # tqdm is a library that helps us visualize the execution of our code and
    the speed of our loops


# the Employee class is defined to represent an average worker with a certain set of
    traits from their respective
# industryTraits child class
class Employee:
    def __init__(self, industryTraits):
        if not isinstance(industryTraits, IndustryTraits):  # sanity check to prevent
    erroneous instances
            raise TypeError("The industry trait set provided is not a valid instance.")

        self.industryTraits = industryTraits
        # these traits are defined as None in self.__init__ but will be populated in
    self.generate_traits
        self.wifiSpeed = None
        self.conscientiousness = None
        self.neuroticism = None
        self.commutingTime = None
        self.numChildren = None
        self.age = None

    def generate_traits(self):
        # for the first four normally distributed traits, we use np.random.normal to
    sample random values from a
        # Gaussian distribution; loc represents the mean and scale represents the
    standard deviation
        self.wifiSpeed = np.random.normal(loc=self.industryTraits.meanWifiSpeed,
                                          scale=self.industryTraits.stdDevWifiSpeed)
```

```
        self.conscientiousness = np.random.normal(loc=self.industryTraits.
meanConscientiousness,
                                                          scale=self.industryTraits.
stdDevConscientiousness)
        self.neuroticism = np.random.normal(loc=self.industryTraits.meanNeuroticism,
                                                   scale=self.industryTraits.stdDevNeuroticism
)
        self.commutingTime = np.random.normal(loc=self.industryTraits.meanCommutingTime
,
                                                     scale=self.industryTraits.
stdDevCommutingTime)

        # a few sanity checks to avoid negative values, also regenerating numbers if
values are negative
        while self.wifiSpeed <= 0:
            self.wifiSpeed = np.random.normal(loc=self.industryTraits.meanWifiSpeed,
                                                     scale=self.industryTraits.stdDevWifiSpeed
)
        while self.conscientiousness <= 0:
            self.conscientiousness = np.random.normal(loc=self.industryTraits.
meanConscientiousness,
                                                              scale=self.industryTraits.
stdDevConscientiousness)

        while self.neuroticism <= 0:
            self.neuroticism = np.random.normal(loc=self.industryTraits.meanNeuroticism
,
                                                       scale=self.industryTraits.
stdDevNeuroticism)

        while self.commutingTime <= 0:
            self.commutingTime = np.random.normal(loc=self.industryTraits.
meanCommutingTime,
                                                         scale=self.industryTraits.
stdDevCommutingTime)

        # for the next two parameters, we use np.random.choice to draw values from a
probability density function; a
        # represents our space, size represents the number of values we need to draw -
we provide None for this
        # argument as we are only drawing a single value, and p represents our PDF
        self.numChildren = np.random.choice(a=self.industryTraits.childrenSpace, size=
None,
                                            p=self.industryTraits.childrenPDF)

        # modifications are required to generate age as our PDF represents an age range
 - thus, after an age range
        # has been determined, there is an equal chance for any integer age to be
selected in that entire range, and we
        # use np.random.randint to do so
        ageRange = np.random.choice(a=self.industryTraits.ageRangeSpace, size=None,
                                    p=self.industryTraits.ageRangePDF)
        self.age = np.random.randint(low=10 * (ageRange + 2) + 5, high=10 * (ageRange +
 3) + 5)

```

```python
    def compute_g(self):
        # this function calculates g, as defined in the rest of the paper; np.exp represents the exponential function
        b_C = 0.9
        N_C = -0.894652
        b_T = 0.430046
        N_T = 0.006154
        b_W = 0.8
        N_W = -0.005434
        return 1/3 * b_C * np.exp(N_C * self.numChildren) + \
               1/3 * b_T * np.exp(N_T * self.commutingTime) + \
               1/3 * (0.9 - b_W) * np.exp(N_W * self.wifiSpeed)

    def compute_h(self):
        # this function calculates h, as defined in the rest of the paper; np.exp represents the exponential function
        averageNumChildren = 0.657102  # this is the constant for the mean number of children from our PDF; this is not
                                       # computed at runtime as it would require excess calls to self.industryTraits
                                       # and excess list comprehensions

        k = 1/4 * (1 - ((self.neuroticism - self.industryTraits.meanNeuroticism) / 0.5) + \
            1 - ((self.numChildren - averageNumChildren) / 0.45) + \
            1 + ((self.conscientiousness - self.industryTraits.meanConscientiousness) / 0.5) + \
            1 + ((self.wifiSpeed - self.industryTraits.meanWifiSpeed) / 36))
        damper = np.abs(k - 1) / np.exp((self.age - 20) / 100)
        kPrime = 1 + damper if k > 1 else 1 - damper
        return kPrime


class Employer:
    def __init__(self, industryTraits, numEmployees):
        if not isinstance(industryTraits, IndustryTraits):  # sanity check to prevent erroneous instances
            raise TypeError("The industry trait set provided is not a valid instance.")
        if type(numEmployees) != int:  # another sanity check
            raise TypeError("The number of employees must be an integer.")

        self.industryTraits = industryTraits
        self.numEmployees = numEmployees
        # self.employees is defined as None in self.__init__ but will be populated in self.generate_employees
        self.employees = None

    def generate_employees(self):
        # we use a list comprehension to wrap a for loop generating numEmployees instances of the Employee class, which
        # are all stored in the list self.people
        self.employees = [Employee(self.industryTraits) for i in range(0, self.numEmployees)]

    def assign_traits(self):
```

```
        # to assign traits to each instance of Employee in self.employee, we iterate over all the people in the list and
        # call generate_traits()
        for employee in self.employees:
            employee.generate_traits()

    def can_work_from_home(self):
        # this determines if an employer will allow their workers to work from home or not, by computing I_J, H_J, and
        # W_J; values of g and h for all employees are computed through a list comprehension iterating through all
        # employees, and so are I_J, H_J, and W_J
        g_values = [employee.compute_g() for employee in self.employees]
        h_values = [employee.compute_h() for employee in self.employees]
        I_J = np.sum([1
                      for i in range(self.numEmployees)])
        H_J = np.sum([g_values[i] * h_values[i] + 1 - g_values[i]
                      for i in range(self.numEmployees)])
        W_J = np.sum([h_values[i]
                      for i in range(self.numEmployees)])

        return [H_J > I_J or W_J > I_J, g_values, h_values, I_J, H_J, W_J]


class Simulation:
    def __init__(self, industryTraits, numEmployers, numEmployees):
        if not isinstance(industryTraits, IndustryTraits):  # sanity check to prevent erroneous instances
            raise TypeError("The industry trait set provided is not a valid instance.")
        if type(numEmployees) != int or type(numEmployers) != int:  # another sanity check
            raise TypeError("The number of employees must be an integer.")

        self.industryTraits = industryTraits
        self.numEmployees = numEmployees
        self.numEmployers = numEmployers
        # self.employers is defined as None in self.__init__ but will be populated in self.generate_employers
        self.employers = None

    def generate_employers(self):
        # we use a list comprehension to wrap a for loop generating numEmployees instances of the Employee class, which
        # are all stored in the list self.people
        self.employers = [Employer(self.industryTraits, self.numEmployees) for i in range(0, self.numEmployers)]

    def run_employers(self):
        # to generate the employer employees and to assign traits to them, we call generate_employees() and
        # assign_traits() on every employer
        for employer in tqdm.tqdm(self.employers): # using tdqm to see progress of the loop during execution
            employer.generate_employees()
            employer.assign_traits()
```

```python

    def determine_p(self):
        # this function determines if the employers in the simulation will allow their workers to work from home,
        # letting us find the proportion of workers who have this privilege
        employer_yeas = 0  # we store the number of employers who answer yes to the million dollar question - do they
                           # allow their workers to work from home?
        employer_aggregate_g = []  # we store all the g_value lists for each employer's employees for analysis
        employer_aggregate_h = []  # more values are stored for analysis
        employer_I = []
        employer_H = []
        employer_W = []
        for employer in tqdm.tqdm(self.employers):  # using tdqm to see progress of the loop during execution
            result = employer.can_work_from_home()
            employer_yeas += 1 if result[0] else 0  # the first value in the list returned by can_work_from_home() to
                                                    # result represents whether the employer agrees to let their workers
                                                    # work from home, therefore we increment employer_yeas if this
                                                    # value is True
            employer_aggregate_g.extend(result[1])  # the second value in the list returned is the list of g_values for
                                                    # all the employer's employees, which are all added to the end of
                                                    # our current aggregate list of g_values
            employer_aggregate_h.extend(result[2])
            employer_I.append(result[3])
            employer_H.append(result[4])
            employer_W.append(result[5])

        return [employer_yeas * beta(self.industryTraits.jobCoefficient, 1.15)/self.numEmployers,
                employer_aggregate_g, employer_aggregate_h,
                employer_I, employer_H, employer_W]  # we return the proportion of employers who let their workers work
                                                     # from home as well as other values for analysis
```

Listing 3: classes.py

```python
class IndustryTraits:  # generic type for traits of an industry that has child classes for each industry
    def __init__(self):
        # these first factors are defined in the parent IndustryTraits class because we assume them to be constant
        # these factors are normally distributed, so we only need a mean and standard deviation for representation
        self.meanWifiSpeed = 127.55  # this is the mean wifi speed across the US in MB/S
        self.stdDevWifiSpeed = 35.996  # this is the standard deviation of the wifi speed across the US in MB/S
```

```

        self.meanConscientiousness = 4.2  # this is the mean conscientiousness score
    for a worker in the US on the
                                          # Costa & McCrae Big 5 Personality test,
    which scores on a scale of 1 - 5
        self.stdDevConscientiousness = 1.08  # this is the standard deviation of the
    conscientiousness score for a
                                             # worker in the US

        self.meanNeuroticism = 3.58  # this is the mean neuroticism score for a worker
    in the US on the Costa & McCrae
                                     # Big 5 Personality test, which scores on a scale
    of 1 - 5
        self.stdDevNeuroticism = 1.18  # this is the standard deviation of the
    neuroticism score for a worker in the US

        self.meanCommutingTime = 26.1  # this is the mean time for a worker's one-way
    commute to their workplace in the
                                       # US as measured in minutes
        self.stdDevCommutingTime = 4.62  # this is the standard deviation of the time
    for a worker's one-way commute to
                                         # their workplace in the US as measured in
    minutes

        # the next factors are distributed with a custom probability density function
        self.childrenSpace = [0, 1, 2, 3]  # these are the potential outcomes for the
    number of children
        self.childrenPDF = [0.6137261597, 0.1738451783, 0.1540290718, 0.05839959026]  #
     this represents the chance for
                                                                                       #
     the corresponding amount of
                                                                                       #
     children in self.childrenSpace

        self.ageRangeSpace = [0, 1, 2, 3]  # these are the representations of the age
    ranges for a particular job,
                                           # where 0 corresponds to an age between
    25-34, 1 to 35-44, 2 to 45-54, and
                                           # 3 to 55-64; for a given person, an exact
    integer for age will be chosen
                                           # after the probabilities in the PDF for the
     corresponding range are divided
                                           # by the range's size

        # the following factors are defined as None as they will be customized in the
    child classes, according to the
        # respective industry
        self.ageRangePDF = None  # this is the probability density function for the age
     ranges defined in
                                 # self.ageRangeSpace, and will vary by industry
        self.jobCoefficient = None  # this is the estimated percent of jobs that can
    theoretically be done at home for
                                    # a given industry


```

```
# for the following child classes of IndustryTraits, we call super().__init__ in self.__init__ so that we can acquire
# all of the information initialized in the parent class' initializer function - as the parent can be retrieved with
# super(), we can call the parent, or IndustryTraits' self.__init__ directly from the child class; variables that change
# for every industry are defined normally in the following lines of the child class' self.__init__ function
class MiningLoggingConstructionIndustryTraits(IndustryTraits):
    def __init__(self):
        super().__init__()
        self.ageRangePDF = [0.2635, 0.303, 0.2437, 0.1898]
        self.jobCoefficient = 0


class ManufacturingIndustryTraits(IndustryTraits):
    def __init__(self):
        super().__init__()
        self.ageRangePDF = [0.256, 0.249, 0.26, 0.235]
        self.jobCoefficient = 0


class TradeTransportationUtilitiesIndustryTraits(IndustryTraits):
    def __init__(self):
        super().__init__()
        self.ageRangePDF = [0.311, 0.24, 0.231, 0.218]
        self.jobCoefficient = 0.03


class InformationIndustryTraits(IndustryTraits):
    def __init__(self):
        super().__init__()
        self.ageRangePDF = [0.315, 0.268, 0.244, 0.173]
        self.jobCoefficient = 0.76


class FinancialActivitiesIndustryTraits(IndustryTraits):
    def __init__(self):
        super().__init__()
        self.ageRangePDF = [0.255, 0.271, 0.248, 0.226]
        self.jobCoefficient = 0.88


class ProfessionalBusinessServicesIndustryTraits(IndustryTraits):
    def __init__(self):
        super().__init__()
        self.ageRangePDF = [0.302, 0.275, 0.234, 0.189]
        self.jobCoefficient = 0.88


class EducationHealthServicesIndustryTraits(IndustryTraits):
    def __init__(self):
        super().__init__()
        self.ageRangePDF = [0.268, 0.266, 0.259, 0.207]
        self.jobCoefficient = 0.98
```

```python


class LeisureHospitalityIndustryTraits(IndustryTraits):
    def __init__(self):
        super().__init__()
        self.ageRangePDF = [0.353, 0.265, 0.215, 0.167]
        self.jobCoefficient = 0.26


class ReligiousSocialCommunityServicesIndustryTraits(IndustryTraits):
    def __init__(self):
        super().__init__()
        self.ageRangePDF = [0.181, 0.244, 0.26, 0.315]
        self.jobCoefficient = 0.37


class GovernmentIndustryTraits(IndustryTraits):
    def __init__(self):
        super().__init__()
        self.ageRangePDF = [0.229, 0.262, 0.275, 0.234]
        self.jobCoefficient = 0.37
```

Listing 4: industryTraits.py

```python
def beta(x, b):
    if x == 0:
        return 0
    return 1 / (1 + ((x / (1 - x)) ** -b))
```

Listing 5: betadist.py