MathWorks Math Modeling Challenge 2022

Watford Grammar School for Boys

Team #15440, Hertfordshire, England

Coach: Ben Eastley

Students: Hamish Starling, Hasan Shahrestani, Luke Powney, James Elcock

M3 Challenge
HONORABLE MENTION—$1,000 (£769) Team Prize

JUDGE COMMENTS

Specifically for Team # 15440—Submitted at the Close of Triage Judging:

COMMENT 1: Paper could be improved by results being explained in the executive summary.

COMMENT 2: Good job noting that the pandemic is causing issues with the data; might also consider major geopolitical events. In Q2 for the percentage of jobs by occupation category, discussed how to apply the model and add a limit based on their result from Q1. Nice that you made the link between Q1 and Q2.

Remote Work: *Fad or Future?*

February 26, 2022

1. Executive Summary

Nobody could doubt that the global coronavirus pandemic has led to radical shifts in work patterns. According to the ONS [2], the proportion of people working from home increased by a phenomenal 10 percentage points from 27% to 37% from 2019 to 2020. But even before this, trends towards "homeworking" were on the rise, from only 22.0% in the US in 2005 to 25.0% in 2013. While working from home offers substantial benefits such as an improved work life balance [2], it can lead to reductions in productivity—undesirable for employers.

In this paper, our team investigates how the trends towards homeworking can be applied into the future, in a post-pandemic society. We apply our models, which determine the percentage of jobs that can be done from home and the percentage that *will* be done from home, to 5 cities—Scranton (PA), Seattle (WA), Omaha (NE), Liverpool (England), and Barry (Wales)—in the hope that this important research can be used to improve sustainable policy making in the US and UK alike.

In Part I of the problem, we are asked to determine the percentage of jobs in the 5 given cities. We divide the labour market within each city into 10 main sectors and make use of an exponential regression model to track the relative proportions of each sector in each city's labour market over time. From this and an estimated proportion of jobs in each sector that can be performed remotely, we achieve results of around 40% of jobs being able to be completed remotely in each of the cities, with the maximum being for Seattle, which we believe to be logical considering Seattle, home of Amazon.com, is one of the most technological cities.

In Part II, we are asked to determine for a given worker who can work from home, whether they will actually work from home, considering the likelihood that they want to work from home and that they are allowed to work remotely. Based on a variety of demographic factors, we make use of a model derived from conditional probability to estimate the percentage chance that a given worker will work from home. This can be compared to the critical value of 0.5 to yield a binary result for a given worker. We give a white male IT Worker aged 40 63% chance of working from home, for example.

In Part III a fusion of the models for jobs (Part I) and for individuals (Part II) in the form of a Monte Carlo simulation is employed based on projected demographics of each town or city to obtain the expected proportion of remote workers. We obtain results ranging from 26.5% for Barry in 2024 (logical, as Barry is not as technologically developed as some regions in the UK, and 2024 is only 2 years away), to 32.5% for Seattle in 2027.

As a result, we make the policy recommendation for the Prime Minister that working from home will remain a substantial part of the UK and US workforces. He must consider it in future labour policies, and the impact of improved technology on the propensity to work from home should be factored into his levelling-up agenda.

Contents

2. Introduction

In this paper we provide an in-depth analysis of post-pandemic working trends in the US and UK, from the proportion of people that are in remote-ready jobs in different towns and cities, to the proportion who will actually take the plunge and leave the office in the near future.

3. Global Assumptions

Throughout this paper we make the following global assumption:

G.1. *Working from home is defined as working from home at least 1 day per week.*
Justification: We have recently seen a rise in hybrid working models in which working from home occurs one day a week, to complement existing models where you either work from home the whole time or not at all. As the problems generally concern themselves with whether a given person "works from home" or a given job can be done "remotely," we are free to choose whether to include these hybrid models. Their inclusion under an umbrella "work from home" term simplifies the scenario to make it more tractable, without reducing the accuracy of the models too much, since exactly how much a person works from home is much easier to change than going fully from the office to home, or vice versa.

4. Part I: Ready or Not

4.1. Problem Statement

Find the percentage of jobs in the following cities that can be accomplished from home in the years 2024 and 2027:

US
- Seattle, WA
- Omaha, NE
- Scranton, PA

UK
- Liverpool, England
- Barry, Wales

4.2. Assumptions and Justifications

1.1. *"Work from home readiness" is only dependent on the specifics of the job in an industry and not an employee's own circumstances.*
Justification: Employees may not be able to work from home due to their own circumstances, but this will not determine whether the job itself is "work from home ready" or not.

1.2. *Proportion of jobs within a given industry which are "work from home ready" will not change in the near future.*
Justification: Any job which is "work from home ready" now will remain "work from home ready," as there will not be any significant technological advancement in robotics or communication technology between now and 2027 which will drastically change how people can work from home.

1.3. *There are no major changes to the trend of industry patterns in a city.*
Justification: It is reasonable to suppose that there will not be any significant government policy or external effect to change the trend of industry patterns in a city significantly, meaning we can use trends to determine how many people in each city will work in each industry.

1.4. *Every candidate for working from home has a device to work from home.*
Justification: Most people in developed countries such as the US and UK own a device and have access to the internet. If someone does not, their employer will likely provide them with a device to work from home

1.5. *After the pandemic, growth in the sectors of the economy we consider will follow pre-pandemic trends, from the lower base of post-pandemic levels.*
Justification: Many employees lost their jobs during the pandemic, which brings down the number of jobs in each sector. However, the trends are likely to return to pre-pandemic levels from this point considering that most countries' policies are to return to the previous "normal."

4.3. Analysing the Problem

As we stated in our assumptions, the effects of wealth and internet connectivity can be considered negligible on the potential for a job to be done from home as these are independent of the job and can be provided by an employer. Hence, the main important variable in determining if a job is "work from home ready" is the type of job, i.e., the industry. We have assumed that the proportion of "work from home ready" jobs within a given industry will remain constant (H_I) between now and 2027. We define an industry pattern to be the proportion of people working in each industry in a specific city at a given time. We start by considering how the industry pattern in a city is changing with time ($P_{I,C}(t)$). Using the $P_{I,C}(t)$ function and the H_I constant, we can find the proportion of jobs in a city which are "work from home ready," which we can do by multiplying them together and summing, as follows:

$$\sum P_{I,C}(t) \cdot H_I \tag{1}$$

4.4. Defining Variables & Constant Parameters

4.4.1 Identifying Variables and Determining Constant Values

The number of people working in an industry will change with time, so we let this be $N_{I,C}(t)$. The workforce of the city will also change over time: $W_C(t)$. Thus the proportion of jobs in an industry in a city will be:

$$P_{I,C}(t) = \frac{N_{I,C}(t)}{W_C(t)} \tag{2}$$

$W_C(t)$ can be expressed as the sum of all $N_{I,C}(t)$ for a given I, C, t.

4.4.2 Table of Variables/Constants

Type	**Symbol**	**Definition**	**Units**
Variable	t	Time since 2000	years
Variable	$N_{I,C}(t)$	The number of people in industry I in city C at time t. This will be determined for each industry	1000 people
Variable	$W_C(t)$	The workforce of the city	1000 people
Function	$P_{I,C}(t)$	Proportion of jobs in a specific industry in a specific city at a specific time.	%
Constant	H_I	Proportion of "work from home ready" jobs in a given industry I.	%

Table 1: Summary of Problem 1 Variables & Constant Parameters

4.5. Developing the Model

We begin by plotting this data on various graphs to look for trends.
When plotting $N_{I,C}(t)$ vs. t, we observed a "wavy pattern":

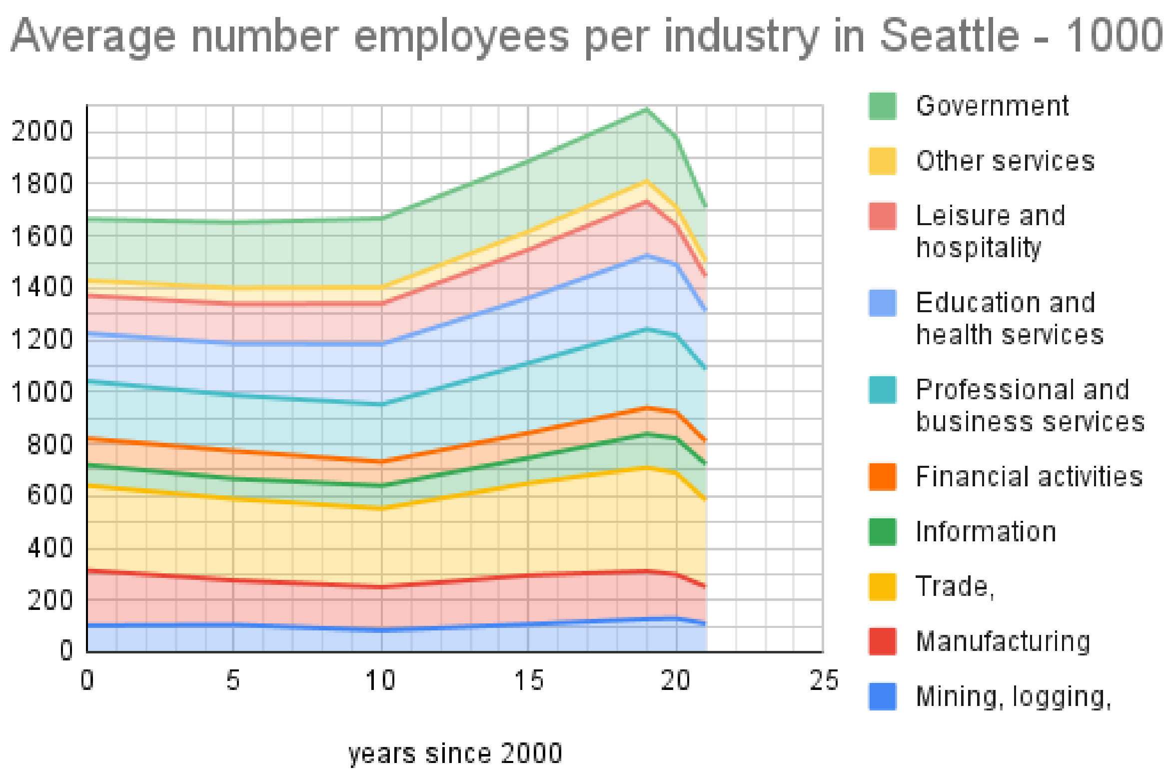

Figure 4.1: stacked chart of $N_{I,C}(t)$ vs. t

As can be seen, the workforce population, which is the total height of Figure 4.1, varies with time, as expected. Plotting $P_{I,C}(t)$ vs. t, we see

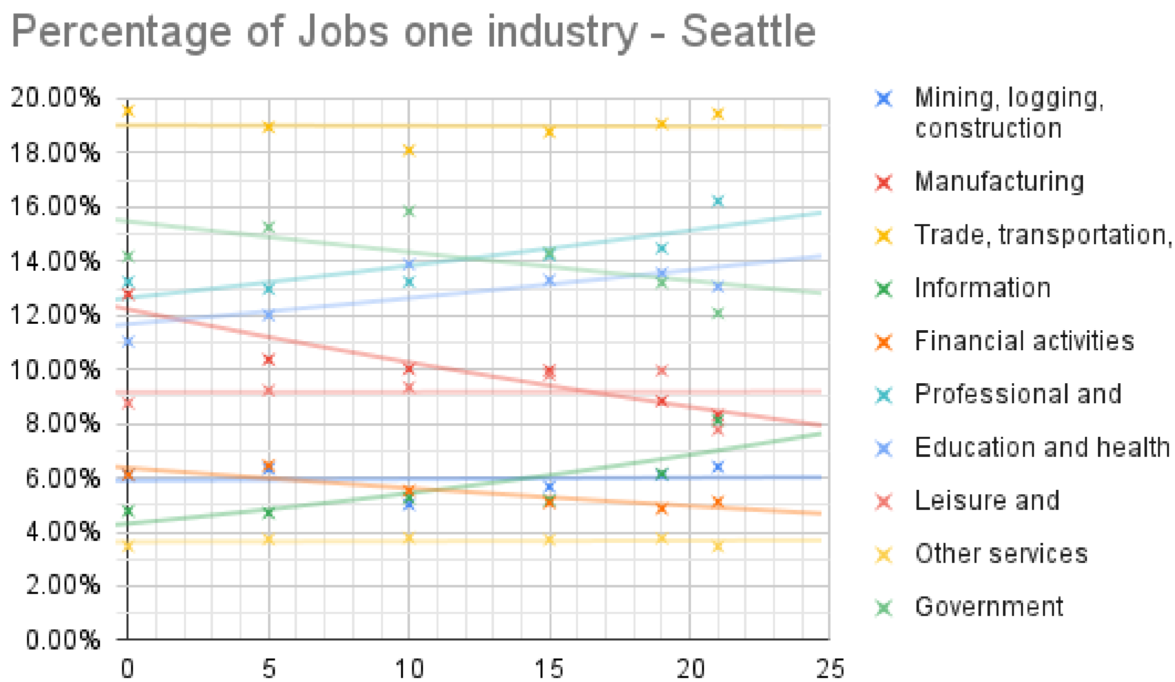

Figure 4.2: Stacked chart of $P_{I,C}(t)$ vs. t

By plotting $P_{I,C}(t)$ we see a clearer trend. We also do not need to account for change in population anymore using this method as $P_{I,C}(t)$ is a percentage of the total workforce population. In addition, we removed the data for 2020 as it seemed anomalous for every industry and thus not representative.

4.5.1 Exponential model

The next task was to find the function $P_{I,C}(t)$ in the form

$$P_{I,C}(t) = a \cdot e^{bt} \tag{3}$$

We are assuming that the proportion of an industry in a city varies as an exponential due to the following reasons:

- An exponential model more accurately represents changes in populations and natural growth over time.
- An exponential model better fits the data.
- An exponential model demonstrates asymptotic behaviour, not allowing for the percentage to fall below a certain level. This is closer to real life as industries will not shrink infinitely to 0.

Using exponential regression, we were able to draw an exponential line of best fit for each $P_{I,C}(t)$.

4.5.2 The H_I constant

The estimated percentage of jobs that can be done at home by occupation category in D3 was the basis for finding a value for H_I. However, it was not as straightforward, as the industries named in D3 were not named the same as the industries in D1, which was the basis for our model of $P_{I,C}(t)$.
So we tried to match each industry in D1 to 1 or more industries in D3.
We then calculated a value for H_I for every industry in D1 based on the the estimated percentage of jobs that can be done at home by occupation category in D3. If there are multiple matches we will calculate a weighted average based on the proportion of the certain occupation in the industry for H_I.

Industry in D1	**Industry in D3**	**weights**	H_I
Mining, logging, construction	Trade, transportation, and utilities	-	0%
Manufacturing	Production	-	1%
Trade, transportation, and utilities	Transportation and material moving	-	30%
Information	Computer and mathematical	-	100%
Financial activities	Business and financial operations	-	88%
Professional and business services	sales-office and administrative-management	0.5-0.4-0.1	49%
Education and health services	Education - health services	0.3-0.7	34%
Leisure and hospitality	Food preparation and service related	-	0%
Other services	-	-	50%
Government	Legal-Office and administrative-Management	0.2-0.7-0.1	74%

Health and Education weights sourced from: [9] [10].

4.6. Applying the Model (Results)

As we have 5 cities and 10 industries, we have 50 $P_{I,C}(t)$ functions. In the interest of saving space, I will give one example of this function:

$$P_{trade,Omaha}(t) = 0.237e^{-0.0117t} \tag{4}$$
$$P_{trade,Omaha}(24) = 17.90\% \tag{5}$$
$$P_{trade,Omaha}(27) = 17.28\% \tag{6}$$

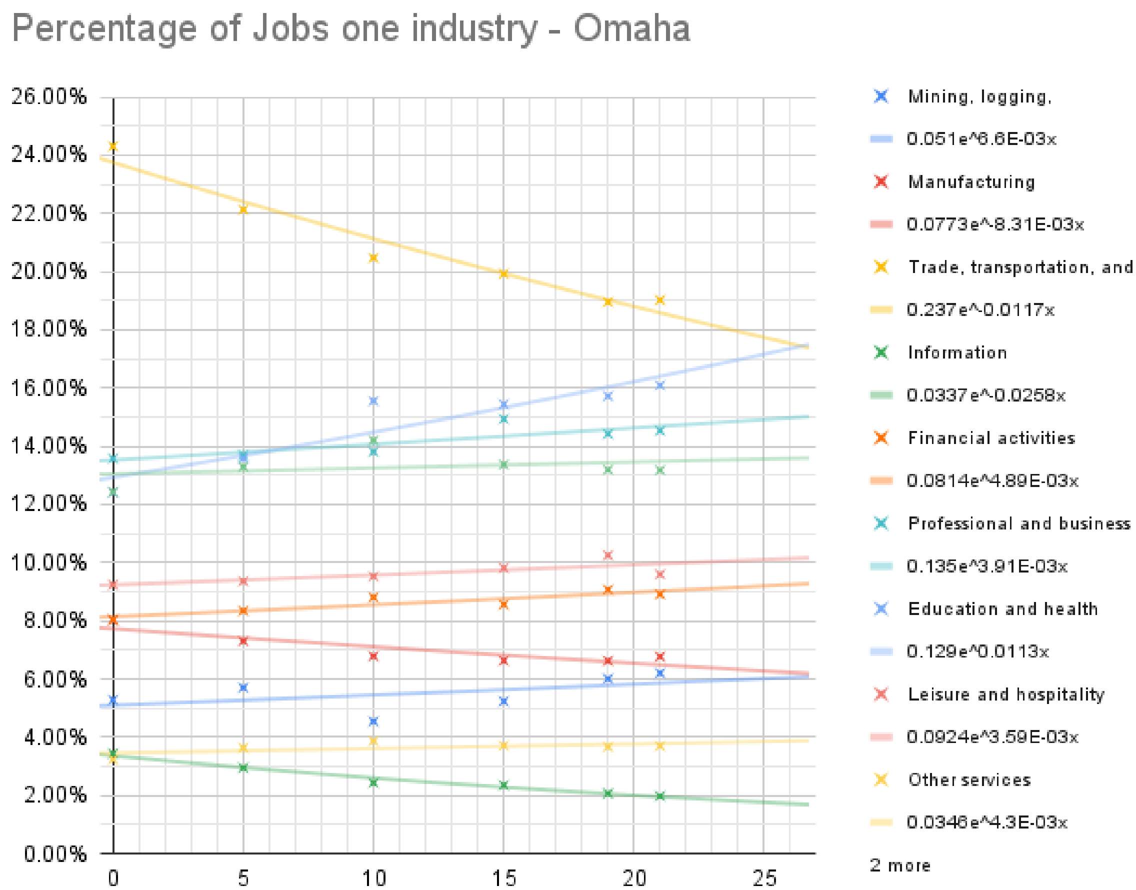

Figure 4.3: $P_{I,C}(t)$ vs. t with exponential trend line

Finally we can find the percentage of "work from home ready" jobs by multiplying the $P_{I,C}(t)$ function by its corresponding H_I value, and summing the results together to get a percentage of all the "work from home ready" jobs as a proportion of all the jobs in the city: $\Sigma P_{I,C}(t) \cdot H_I$ for all industries in the city.

City	**year 2024**	**year 2027**
Seattle	41.31%	41.80%
Omaha	40.22%	40.36%
Scranton	35.91%	36.16%
Liverpool	31.22%	31.04%
Barry	39.07%	38.95%

4.7. Implications

Our results show that approximately 40% of jobs will be "work from home ready." This is not too different from the current rate, because we have assumed technology will not change much in the next 5 years to create a considerable change in the proportion of jobs that can be done from home.

4.8. Evaluating the Model

4.8.1 Validation: Testing for Accuracy

To validate our data we looked online to other predictions for the percentage of jobs that could be done online. One piece of data we came across was a Forbes article [6] that stated that it was about 37%. And as working from home becomes easier as tech is more integrated into work, I would say that our model accurately predicts the proportion of people able to work from home.

4.8.2 Sensitivity Analysis: Testing for Stability and Sensitivity to Assumptions

As an example of how sensitive the model is, we will use the function $P_{trade,Omaha}(25)$ where $P_{trade,Omaha}(t) = 0.237e^{-0.0117t}$.

Δ Time	$P_{trade,Omaha}$
+5%	+0.004%
-5%	-0.003%

Table 2: Sensitivity Analysis for Model $P_{trade,Omaha}$

4.8.3 Model Strengths

- preserves the impact of decreasing industries
- reflects the increasing dominance of certain industries
- is not sensitive to small changes in time
- includes the population of the workforce
- uses the changing the pattern of industries in the model

4.8.4 Model Weaknesses

Over a very large periods of time (100+years) the model will become unsuitable as there is no way to conceivably predict the market share of each industry.

5. Part II: Remote Control

5.1. Problem Statement

In this part of the problem we are asked to "create a model that predicts whether an individual worker whose job is remote-ready will be allowed to and will choose to work from home."

We take this to mean "given the characteristics of an individual worker such as their age, income, and family status, determine the probability that a worker with these characteristics will be allowed to *and* want to work from home, versus in the office. Then we compare this probability to a threshold value to give the binary answer of *yes*, the worker will work from home, or *no*, they won't."

This problem definition is advantageous because it scales more easily to larger numbers of people with the same characteristics, which will be needed when this model is reused in Q3. As human choices are not deterministic, a level of stochasticity can be employed if we use the probabilities, meaning not everyone with the same characteristics behaves in the same way.

5.2. Assumptions and Justifications

2.1. *Whether a worker wants to work from home can be treated as independent of whether they are allowed to, and are in a remote-ready job.*
Justification: This is a logical assumption because it is plausible to want to work from home but not be allowed to, to want to work from home and be allowed to, to not want to work from home and not be allowed to, and to not want to work from home and not be allowed to. In a purely theoretical sense, workers can make a decision on what they want to do separately from what they are allowed to do.

2.2. *2019 ONS Data [3] for the proportion of workers of different ethnicities, sexes, professions, etc., is representative of the pre-pandemic probability that a worker of the given ethnicity/sex, etc., will work from home.*
Justification: The data was collected with a large sample size by a substantial public statistics organization, so it is reasonable to suppose this is the case. We use the most recent pre-pandemic data, as this is likely to be the most representative.

2.3. *We can account for the impact of the pandemic on working from home by applying a pandemic constant correction factor.*
Justification: This reflects how attitudes have changed toward remote working over the course of the pandemic. It would be preferred to determine this on the level of an individual, but this requires data we don't currently have access to. See "Extending the Model."

2.4. *The maximum age of a working person is 80 years and the minimum age is 20 years.*
Justification: This is probably an overestimate but should definitely not be an underestimate, which is important for our model to avoid negative age multipliers.

2.5. *Working age is (apart from the above constraints), normally distributed with mean 35 and standard deviation 10*
Justification: This is likely because we would expect ages to be roughly symmetrical as there are factors which would both prevent younger workers getting into the labour market (education) and remove older workers (retirement). As we don't know the distribution and in the interests of time we make this assumption. By the 68–95–99.7 rule these parameter choices give us 99.7% of the data within the stated min/max, meaning artificially rounding values into this range has next to no effect.

5.3. Analysing the Problem

It is insightful to represent the problem in probability notation. For a given worker, we define the following events.

- W, the event that the worker wants to work from home.
- R, the event that the worker's job enables them to work from home (remote ready job).
- A, the event that the worker's employer allows them to work from home.
- W_p, the event that the worker wanted to work from home pre-pandemic.

We also define a pandemic correction c for the worker which is a multiplying factor so that

$$P(W) = \min\left(1, c \cdot P(W_p)\right) \tag{7}$$

This is designed to reflect how the pandemic has changed attitudes towards remote working. With these events, what the question is asking for becomes clear:

$$P((W \cap A)|R) \equiv \frac{P(W \cap A \cap R)}{P(R)} \tag{8}$$

To simplify the notation we omit conditioning on all of the features of the worker, as we can determine the probabilities of events derived from W, A, R from the worker's features.

Under Assumption 2.1. we can perform the following manipulation:

$$P((W \cap A)|R) \equiv \frac{P(W) \cdot P(A \cap R)}{P(R)} \equiv \frac{c \cdot P(W_p) \cdot P(A \cap R)}{P(R)} \equiv \frac{c \cdot P(W_p \cap A \cap R)}{P(R)} \tag{9}$$

5.4. Defining Variables & Constant Parameters

5.4.1 Identifying Factors Which Affect the Outcome

When it comes to determining which workers work from home, the following factors were considered to be important. These were then condensed into variables for the model below.

- *Age*
  Justification: The age of a worker will affect how "tech-savvy" they are, and remote working requires advanced use of digital technology. It will also affect how much of an active social life the worker has.

- *Commute Time*
  Justification: In [2], the authors determine that "work-life balance was the greatest positive" of homeworking, and time spent commuting is reduced by homeworking, which affects work-life balance.

- *House Features*
  Justification: Working from home is dependent on whether or not a given worker has the space at home to concentrate and work.

- *Income*
  Justification: Working from home requires less transport costs, which saves money, but infrastructure such as a good broadband connection is required.

- *Sex*
  Justification: This may affect the way in which the person socializes with work colleagues, which could be an incentive to come into the office.

- *Industry*
  Justification: This affects their potential productivity when working from home, which may determine whether they are allowed to work from home.

The following features were discounted for the given reasons:

- *Family Characteristics*
  Justification: Workers with young children may need to work from home to look after them, or alternatively may be distracted by them and want to work in the office. It was deemed not worth the time to investigate which of these is most likely. It can also be argued that the worker should not have to see their children if they are working from home; they may use a childminder.

- *Income*
  Justification: Unfortunately we didn't have the required data to perform the below analysis for income as well as the factors used.

- *Size of Team - how many others are working from home?*
  Justification: It is important to maintain a workplace culture, so for large organizations some staff may be required to come into the office, while for small businesses working from home may be preferred to cut costs. However, the data was not available, and so this feature had to be discarded.

5.4.2 Table of Variables/Constants

Type	Symbol	Definition	Value	Units
Constant	σ	Age standard deviation	10	years
Constant	μ	Age mean	35	years
Variable	A	Age of a given worker	$20 < A < 80$	years
Variable	S	Sex of a given worker	Male/Female	years
Variable	E	Ethnicity of a given worker	White/Mixed/Black/Asian/Other	None
Variable	I	Industry in which a given worker works	Selected from the ONS Standard Industrial Classification (SIC) Grouping [1]	None
Variable	L	Level of education of a given worker	No Qualification, GCSE, A Level, etc. American and British qualifications converted to standardise as necessary.	None
Variable	T	Time a worker works	Part Time or Full time	None
Variable	C	Commute Length of a given worker	0 to 10000	m
Variable	H	Whether their house allows work from home or not, i.e., do they have a quiet place to work, based on average house sizes in the region	True/False	None
Constant	c	Pandemic WFH attitudes correction. We choose this value because data from [7], Figure 3, suggests that (cancelling out the hugely improved and worsened reactions) the average tolerance of working from home increased by 40% during the pandemic.	1.4	None

Table 3: Summary of Problem 2 Variables & Constant Parameters

5.5. Developing the Model

From section 5.3, we identify that in order to get the output answer, we need to obtain $P(R)$ and $P(W_p \cap A \cap R)$. c was set as 1.4, as explained above.

- $P(R)$ can simply be determined based on the worker's industry using the given data in D3 [8]. This tells us the proportion of jobs in a given industry being remote-ready. We could also use the model from Part I here; however, we decided not to as that brings in the additional complexity of the city the worker comes from. This should only be brought in at Part III and not Part II.

- $P(W_p \cap A \cap R)$ is the overall probability that this worker worked from home pre-pandemic (as a worker works from home if and only if they want to, are allowed to, and can). We can determine this based on historical work-from-home data for the UK from the ONS [3]. This is the backbone of the model in the determination of $P(W_p \cap A \cap R)$.

The ONS data [3] is provided in the following format. We convert this into a Python dictionary format, enabling easier processing.

Full-time

	Agriculture, forestry and fishing (A)			
	Never	Mainly	Recently	Occasionally
2011	62.90%	3.83%	7.10%	26.17%
2012	69.38%	2.45%	7.19%	20.98%
2013	64.42%	4.21%	5.30%	26.08%
2014	59.32%	3.90%	10.67%	26.11%
2015	66.54%	5.11%	6.41%	21.93%
2016	63.34%	5.62%	8.73%	22.32%
2017	68.70%	4.34%	6.55%	20.41%
2018	67.36%	5.87%	4.48%	22.29%
2019	62.22%	6.95%	6.03%	24.81%
2020	67.63%	7.49%	10.73%	14.14%

Figure 5.1: An example data table from the ONS data. The 2019 data was extracted, as it was the most recent, and converted into Python format for easy usage. See DataONS.py in the appendix.

The model created relies on technical computing through the use of object-oriented programming to represent a person. We made this choice because it is easy and computationally cheap to spin up new instances of an OOP object, meaning it would work well when the model is reused in Part III for the Monte Carlo simulation.

The model works in the following way:

1. Determine the overall proportion of people actually working from home pre-pandemic, and call this the baseRate.

2. For each of the person's characteristics, such as their ethnicity, whether they work full time or part time, etc., we use ONS data to determine the percentage of people working from home within these categories.

3. We then divide these by the baseRate so that if a particular industry is “bang average,” no change is made to the probability of working from home pre-pandemic, whereas if a given industry is more or less prone to working from home, we multiply the baseRate probability by the factor industry rate over baseRate, for example.

4. This process of multiplying the rate is repeated for each of a given person’s characteristics.

5. We then incorporate the continuous (non-categoric) factors affecting working from home of commute distance and age. To incorporate commute distance we take the tanh of the commute distance in km and multiply it by the current probability of working from home for the person. This is justified because tanh goes through the origin, which means that if a person has 0 commute distance their chance of working from home becomes 0 because it is no extra effort to go into the office. As the person gets further away from the office (commute lengthens) the probability of working from home increases as they will not want to have to commute. We use tanh because it peaks at 1, meaning we can’t increase the probability of working from home above 1, and once a worker gets far from their office an increase of 1km has less effect.

6. To incorporate the age of the worker we again multiply the current probability by a new value, which in this case is defined as multiplying factor $= 1.5 - 0.5 \cdot e^{\frac{A-20}{80}}$. This yields the desired behaviour of a steady decline in likelihood of working from home with age; starting at 1 for a 20-year-old due to level of tech savviness.

5.6. Applying & Evaluating the Model (Results)

To evaluate the model we consider some example people. It’s impossible to try all of the many intersections of the different factors, but we think the following demonstrates a range of logical behaviours. The model’s successful integration into Part III leading to reasonable results suggests it is good. See that section for more details.

```
print(PW_n_A_givenR(Person("men","white","InformationCommunication","degreeOrEquiv","Full Time",1000,True,40)))
print(PW_n_A_givenR(Person("men","white","InformationCommunication","degreeOrEquiv","Full Time",1000,True,80)))
print(PW_n_A_givenR(Person("women","white","InformationCommunication","degreeOrEquiv","Full Time",1000,True,40)))
print(PW_n_A_givenR(Person("men","asian","InformationCommunication","noQualifications","Full Time",1000,True,40)))
print(PW_n_A_givenR(Person("men","asian","WholesaleRetailRepairOfVehicles","aLevelOrEquiv","Full Time",1000,True,40)))
```

Figure 5.2: Input to the model to apply and evaluate it

```
0.634931494926454
0.3267207480369347
0.5549085829395297
0.15157197037077277
0.5296584643434392
```

Figure 5.3: Probabilities of the respective people working from home.

6. Part III: Just a Little Home-work

6.1. Problem Statement

In this section of the challenge, we are asked to "synthesize [our] models from the first two questions to create a model which, for a given city, estimates the percentage of workers who will work remotely."

This should work for 2024 and 2027 in the same cities considered previously, namely, Seattle, Omaha, Scranton, Liverpool, and Barry, and then be used to determine a relative ranking for the different cities of the impact of remote working on them.

6.2. Assumptions and Justifications

3.1. *There are no major changes to the population demographic between now and 2027.*
Justification: It is reasonable to assume that there will not be any significant changes to the factors that affect the population demographic. This includes percentage of population that are of working age to be constant, and there will be no net change in migration, as well as things like ethnicity and education levels as we have used current data to make predictions about these values for the future. This is reasonable as 2024 and 2027 are not that far away.

3.2. *The impact of house size on working from home can be modeled as 90% of people having the housing space and bandwidth to work from home.*
Justification: Due to a lack of time and consistent data we make this general assumption for all of the cities considered.

3.3. *Within each sex, for example, each ethnicity is distributed in the same way as it is for the whole population. That is, ethnicity and sex and the other ways the population is to be divided are independent.*
Justification: This is a reasonable assumption to make because we don't have a breakdown of the individual characteristics of each person. As we generate a large random population the impact of this evens out.

6.3. Analysing the Problem

It is valuable to take a moment to understand the characteristics of what we have developed so far and what we need for this part. While Part I of the problem studies, on the level of cities and their industries, the proportion of jobs being "work from home ready," Part II asks on the individual level whether someone who can work from home will do so.

For Part III we require a model which will, on the city level, determine the proportion of people who will actually work from home. Hence this leads us to the plan of using a simple agent-based model to combine the individual and whole city characteristics. We will use results from Part I to identify the number of people in each industry in each city in

2024 and 2027, then use a Monte Carlo simulation to create a profile for each member of the population of each city. Each profile can be run through the model from Part II to produce a probability of the given person working from home, and using a random variable we can count them in or out. Then summing for the whole population allows us to obtain percentages.

6.4. Defining Variables & Constant Parameters

6.4.1 Identifying Variables and Determining Constant Values

As we rely on the model from section 2 to identify the work-from-home preferences of a given individual, we use here the same input data as we used there for each person.

6.4.2 Table of Variables/Constants

Type	**Symbol**	**Definition**	**Value**	**Units**
Variable	$N_{I,C,T}$	The number of people in industry I in city C at time T. As in Part I	-	1000 people

Table 4: Summary of Problem 3 Variables & Constant Parameters

6.5. Developing the Model

We use a Monte Carlo simulation in which we determine the expected population in 2024 and 2027 of each of the cities investigated. For each member of the population, we generate a Person object so that overall the characteristics of the employees match the summary statistics for the town. This generation is performed with the recursiveGen algorithm, which takes a number of people it must generate, a city, and a year. It splits the population in half for each sex, then calls itself to split each of these groups into sections for ethnicity based on the relevant city's ethnicity percentages, and so on for all of the other factors. At the base case, it has generated a value for all of the attributes of the Person class, so it instantiates a number of Person objects equal to amount. It calls the Part II model on these people and sums the results to give a total number of people working from home in 2024 and 2027 for each city. This is then converted to a percentage in each case.

6.6. Applying the Model (Results)

Using this model we achieve the following results for the percentage of people working from home:

Result Year	2024 predictions	2024 ranking	2027 predictions	2027 ranking
Barry	26.51%	4th	26.59%	5th
Liverpool	27.37%	3rd	29.13%	3rd
Omaha	29.07%	2nd	29.26%	2nd
Seattle	30.73%	1st	32.45%	1st
Scranton	26.36%	5th	26.63%	4th

Table 5: Summary of Problem 3 Results for 2024 & 2027

As is to be expected, the most technological city on Earth is to be most revolutionised by remote working, while a small town in Wales where the main industries have always been mining and fishing [4] serves less to benefit. This suggests that as a country if we desire fairness and equality of opportunity we should invest more in remote working capacity for smaller towns like Barry.

6.7. Evaluating the Model

6.7.1 Validation: Testing for Accuracy

The model returns a plausible result. The latest ONS figures suggest that 30% of working adults work from home either partially or exclusively [5]. A modest increase on pre-pandemic levels is to be expected.

6.7.2 Sensitivity Analysis: Testing for Stability and Sensitivity to Assumptions

Due to time constraints we could only perform this analysis on 2024 results. It is desirable

Δ Pandemic Correction Constant	Barry	Liverpool	Omaha	Seattle	Scranton	Average
+10%	+9.7%	+9.8%	+9.9%	+9.8%	+9.7%	+9.78%
+5%	+4.5%	+4.9%	+4.9%	+4.8%	+4.6%	+4.74%
-5%	-4.8%	-4.7%	-5.0%	-4.7%	-5.1%	-4.86%
-10%	-9.7%	-10.0%	-9.9%	-9.9%	-9.9%	-9.88%

Table 6: Sensitivity Analysis for Model 3

for the impact of changing model assumptions to be as low as possible so that if they are wrong, the model is not rendered useless. The most obvious assumption made here was the choice of 1.4 for the pandemic correction constant. We see that varying this changes the percentage of people working from home in the expected way, with change magnitude always less than or equal to the change in the constant c, which is good.

6.7.3 Model Strengths

The model inputs a wide range of factors, including population and job demographics, which means the model can be used for any city with such data values, Furthermore all predicted results are between 26% and 34%, which is similar to current levels and significantly above pre-pandemic levels.

6.7.4 Model Weaknesses

The main weakness of the model is that it is complicated and requires a lot of data on each city. It was time-consuming to format all of this data for each place and sector correctly for the program, so extending the model to other places would be painful.

6.8. Extending the Model

If we had more time we would have

- performed greater validation on this model to ensure that it was accurate, by removing and varying other assumptions and parameters to identify the impact;
- simplified the model or factored out some of the data so as to be more easily extended to other cities.

7. Joining the Dots: Conclusions

In this paper, we investigated how trends towards homeworking can be applied into the future, in a post-pandemic society. In Part I of the problem, we are asked to determined the percentage of jobs that can work remotely in the 5 given cities, achieving results of around 40% of jobs being able to be completed remotely in each city, which is consistent with other data. In Part II, we identified, for a given worker who can work from home, whether they will actually work from home, considering the likelihood that they want to work from home and that they are allowed to work remotely. For example, we gave a white male IT Worker aged 40 63% chance of working from home. In Part III a fusion of the models for jobs (Part I) and for individuals (Part II) in the form of a Monte Carlo simulation was employed based on projected demographics of each town or city to obtain the expected proportion of remote workers. Results ranging from 26.5% for Barry in 2024 to 32.5% for Seattle in 2027 were logical.

As a result of our work, we make the policy recommendation for the Prime Minister that working from home will remain a substantial part of the UK and US workforces. He must consider it in future labour policies, and the impact of improved technology on the propensity to work from home should be factored into his levelling-up agenda.

8. References

[1] Office for National Statistics. UK SIC 2007. https://www.ons.gov.uk/methodology/classificationsandstandards/ukstandardindustrialclassificationofeconomicactivities/uksic2007, 2007. Accessed: 2022-26-02.

[2] Office for National Statistics. Business and individual attitudes towards the future of homeworking, UK Business and individual attitudes towards the future of homeworking, UK: April to May 2021. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/businessandindividualattitudestowardsthefutureofhomeworkinguk/apriltomay2021, Jun 2021. Accessed: 2022-26-02.

[3] Office for National Statistics. Homeworking in the UK, work from home status. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/datasets/homeworkingintheukworkfromhomestatus, Apr 2021. Accessed: 2022-26-02.

[4] Vale of Glamorgan. The Barry Story. https://www.valeofglamorgan.gov.uk/Documents/Working/Regeneration/Economic-Development/The-Barry-Story.pdf. Accessed: 2022-26-02.

[5] Office for National Statistics. Coronavirus (COVID-19) latest insights: Lifestyle. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/lifestyle#homeworking, Feb 2022. Accessed: 2022-26-02.

[6] Forbes Mark Travers. What Percentage Of Workers Can Realistically Work From Home? New Data From Norway Offer Clues. https://www.forbes.com/sites/traversmark/2020/04/24/what-percentage-of-workers-can-realistically-work-from-home-new-data-from-norway-offer-clues/?sh=377f0ebe78fe, Apr 2020. Accessed: 2022-26-02.

[7] Nicholas Bloom Shivani Taneja, Paul Mizen. Working from home is revolutionising the UK labour market. https://voxeu.org/article/working-home-revolutionising-uk-labour-market, Mar 2021. Accessed: 2022-26-02.

[8] SIAM. Remote Work: Fad or Future, MathWorks Math Modeling Challenge 2022. https://m3challenge.siam.org/node/559, Feb 2022. Accessed: 2022-26-02.

[9] UK-Gov. School workforce in England. https://explore-education-statistics.service.gov.uk/find-statistics/school-workforce-in-england, 2020. Accessed: 2022-26-02.

[10] UK-Gov. Number of health and care workers in England and Wales for 2019 and 2020. https://www.ons.gov.uk/

peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/adhocs/13339numberofhealthandcareworkersinenglandandwalesfor2019and2020, 2021. Accessed: 2022-26-02.

9. Appendix: Code Listing for Technical Computing Consideration

9.1. Part II

dataONS.py

```
#General numbers of people working from home (millions, 2019)
kNoWFH_2019 = 23.769735
kWFH_2019 = 8.615113

#Sex (2019) - WFH Percentage
kNoWFH_men_2019 = 12.559057
kWFH_men_2019 =  4.59726200
kNoWFH_women_2019 = 11.210678
kWFH_women_2019 = 4.01785100
sexWFH = {"men":kWFH_men_2019/(kWFH_men_2019+kNoWFH_women_2019),"women": kWFH_women_2019/(kWFH_men_2019+kNoWFH_women_2019)}

#Ethnicity (2019) - WFH Percentage
ethnicityWFH = {"white":1-0.726982116699218,"mixed":1-0.677182853221893,"asian":1-0.786816716194152,"black":1-0.825749695301055,"other":1-0.78287649154663}

#Industry (2019)
#Note: Wanted to use BLS categories but the ONS data was given as percentages making combining difficult without the raw totals.
industryFT_WFH = {"AgricultureForestryFishing":1-0.622196853160858,
"MiningUtilities":1-0.760932683944702,
"Manufacturing":1-0.788643658161163,
"Construction":1-0.751720190048217,
"WholesaleRetailRepairOfVehicles":1-0.822654724121093,
"TransportStorage":1-0.883512794971466,
"AccommodationFoodServices":1-0.856739997863769,
"InformationCommunication":1-0.530150711536407,
"FinancialServicesRealEstate":1-0.638728022575378,
"ProfScientificTechnicalActiv":1-0.590569615364074,
"AdminSupportServices":1-0.747599959373474,
"PublicAdminDefence":1-0.708281695842742,
"Education":1-0.60731154680252,
"HealthSocialWork":1-0.771711289882659,
"OtherServices": 1-0.660328805446624}

industryPT_WFH = {
"AgricultureForestryFishing":1-0.744184255599975,
"MiningUtilities":1-0.653858661651611,
"Manufacturing":1-0.721013069152832,
"Construction":1-0.657320737838745,
"WholesaleRetailRepairOfVehicles":1-0.917843580245971,
"TransportStorage":1-0.865761935710906,
"AccommodationFoodServices":1-0.93391728401184,
"InformationCommunication":1-0.497111767530441,
"FinancialServicesRealEstate":1-0.628497898578643,
"ProfScientificTechnicalActiv":1-0.505080223083496,
"AdminSupportServices":1-0.806695878505706,
"PublicAdminDefence":1-0.744857549667358,
"Education":1-0.728527069091796,
"HealthSocialWork":1-0.824699640274047,
"OtherServices":1-0.748933792114257
}

#Education
educationWFH = {
  "noQualifications":1-0.892920076847076,
```

```
  "entryLevel":1-0.837141454219818,
  "gcseOrEquiv":1-0.780677258968353,
  "aLevelOrEquiv":1-0.722423911094665,
  "degreeOrEquiv":1-0.649750053882598,
  "higherDegree":1-0.573047816753387
}


industryFT_WFH = {"AgricultureForestryFishing":1-0.622196853160858,
"MiningUtilities":1-0.760932683944702,
"Manufacturing":1-0.788643658161163,
"Construction":1-0.751720190048217,
"WholesaleRetailRepairOfVehicles":1-0.822654724121093,
"TransportStorage":1-0.883512794971466,
"AccommodationFoodServices":1-0.856739997863769,
"InformationCommunication":1-0.530150711536407,
"FinancialServicesRealEstate":1-0.638728022575378,
"ProfScientificTechnicalActiv":1-0.590569615364074,
"AdminSupportServices":1-0.747599959373474,
"PublicAdminDefence":1-0.708281695842742,
"Education":1-0.60731154680252,
"HealthSocialWork":1-0.771711289882659,
"OtherServices": 1-0.660328805446624}

#D3 remote work data -> matched up against industries from the ONS.
P_RIndustry = {"AgricultureForestryFishing":0.01,
"MiningUtilities":0.01,
"Manufacturing":0.01,
"Construction":0.01,
"WholesaleRetailRepairOfVehicles":0.28,
"TransportStorage":0.03,
"AccommodationFoodServices":0.05,
"InformationCommunication":1,
"FinancialServicesRealEstate":0.88,
"ProfScientificTechnicalActiv":0.98,
"AdminSupportServices":0.65,
"PublicAdminDefence":0.97,
"Education":0.98,
"HealthSocialWork":0.37,
"OtherServices":0.25}
```

Question2.py

```python
#import necessary modules
from math import exp
from dataONS import * #2019 ONS Data https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/datasets/homeworkingintheukworkfromhomestatus which we have converted to python dictionary format
import random

c = 1.4 #Pandemic attitudes constant

class Person:
  """Defines a single member of the population in the Monte Carlo simulation"""
  def __init__(self,sex,ethnicity,industry,education,time,commute,houseAllowsWFH,age):
    """Take in data about the person and save it to local state"""
    self.sex = sex
    self.ethnicity = ethnicity
    self.industry = industry
    self.education = education
    self.time = time
    self.commute = commute #IN METRES
    self.houseAllowsWFH = houseAllowsWFH
    self.age = age

def tanh(x):
  """the tanh function"""
  return (exp(x) - exp(-x))/(exp(x)+exp(-x))

def PWp_n_A_R(person):
  """Probability that a given person wanted (pre-pandemic) and was allowed to and was ready to WFH"""

  if person.houseAllowsWFH: #if their house does not allow WFH then they cannot WFH at all
    baseRate = kWFH_2019/(kWFH_2019+kNoWFH_2019) #For a generic person, the chance that they WFH pre pandemic

    currentRate = baseRate #Start with this but change it based on the person's characteristics
    currentRate *= sexWFH[person.sex]/baseRate #Do people of their sex WFH more or less than average?
    currentRate *= ethnicityWFH[person.ethnicity]/baseRate #Do people of their ethnicity WFH more or less than average?
    if person.time == "Full Time": #Do people of their full/part time WFH more or less than average?
      currentRate *= industryFT_WFH[person.industry]/baseRate
    else:
      currentRate *= industryPT_WFH[person.industry]/baseRate
    currentRate *= educationWFH[person.education]/baseRate #Do people of their education level WFH more or less than average?
    currentRate *= tanh(person.commute/1000) #If the person lives really close to the office they will be much less likely to work from home. If they live far away we want the impact of moving 1 mile away from already far away to be low so we use tanh.
    currentRate *= 1.5-0.5*exp((person.age-20)/80) #Steady decline in likelihood of WFH with age; starting at 1 for a 20 year old due to level of tech saviness.
    return currentRate
  else:
    return 0

def P_R(person):
  """Based on the given industry, the probability that their job is remote ready"""
  return P_RIndustry[person.industry]

def PW_n_A_givenR(person):
  """Probability that they (post pandemic) want to WFH and are allowed to WFH given that their job is remote ready"""
```

```
    return max(0,min(c*PWp_n_A_R(person)/P_R(person),1))

print(PW_n_A_givenR(Person("men","white","InformationCommunication","degreeOrEquiv","Full Time",1000,True,40)))
print(PW_n_A_givenR(Person("men","white","InformationCommunication","degreeOrEquiv","Full Time",1000,True,80)))
print(PW_n_A_givenR(Person("women","white","InformationCommunication","degreeOrEquiv","Full Time",1000,True,40)))
print(PW_n_A_givenR(Person("men","asian","InformationCommunication","noQualifications","Full Time",1000,True,40)))
print(PW_n_A_givenR(Person("men","asian","WholesaleRetailRepairOfVehicles","aLevelOrEquiv","Full Time",1000,True,40)))
```

9.2. Part III

Question3.py

```python
#import necessary modules
from math import exp
from dataONS import * #2019 ONS Data https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/datasets/homeworkingintheukworkfromhomestatus which we have converted to python dictionary format
from Question2 import * #we reuse question 2 model here in question3.
import numpy as np#for normal distribution
import random

mu, sigma = 35, 10 # age mean and standard deviation
c = 1.4 #Pandemic attitudes constant

def tanh(x):
  """the tanh function"""
  return (exp(x) - exp(-x))/(exp(x)+exp(-x))

#Projected populations based on regression
population = {"2024":{"Barry":56805,"Liverpool":754024,"Omaha":504872,"Seattle":1780424,"Scranton":2461378},"2027":{"Barry":58410,"Liverpool":772777,"Omaha":515231,"Seattle":1846727,"Scranton":2461550}}

#Possible characteristics of a person
years = ["2024","2027"]
cities = ["Barry","Liverpool","Omaha","Seattle","Scranton"]
sexes = ["men","women"]
ethnicities = ["white","mixed","asian","black","other"]
educations = ["noQualifications","entryLevel","gcseOrEquiv","aLevelOrEquiv","degreeOrEquiv","higherDegree"]
industries = ["AgricultureForestryFishing","MiningUtilities","Manufacturing","Construction","WholesaleRetailRepairOfVehicles","TransportStorage","AccommodationFoodServices","InformationCommunication","FinancialServicesRealEstate","ProfScientificTechnicalActiv","AdminSupportServices","PublicAdminDefence","Education","HealthSocialWork","OtherServices"]
times = ["Full Time","Part Time"]
houses = [True,False]

#For each city and sometimes for each year, percentage of people having each tag
sexPercent = {"Barry":{"men":0.5,"women":0.5},"Liverpool":{"men":0.5,"women":0.5},"Omaha":{"men":0.5,"women":0.5},"Seattle":{"men":0.5,"women":0.5},"Scranton":{"men":0.5,"women":0.5}}

ethnicityPercent = {"Barry":{"white":0.961,"mixed":0.016,"asian":0.018,"black":0.004,"other":0.01},"Liverpool":{"white":0.91,"mixed":0.02,"asian":0.041,"black":0.019,"other":0.01},"Omaha":{"white":0.7747,"mixed":0.034,"asian":0.0384,"black":0.1232,"other":0.0297},"Seattle":{"white":0.657,"mixed":0.051,"asian":0.138,"black":0.079,"other":0.126},"Scranton":{"white":0.8309,"mixed":0.0439,"asian":0.0467,"black":0.0585,"other":0.02}}

industryPercent = {"2024":{"Barry":{"AgricultureForestryFishing":0.025,"MiningUtilities":0.025,"Manufacturing":0.083,"Construction":0.025,"WholesaleRetailRepairOfVehicles":0.010,"TransportStorage":0.010,"AccommodationFoodServices":0.194,"InformationCommunication":0.0656,"FinancialServicesRealEstate":0.058,"ProfScientificTechnicalActiv":0.054,"AdminSupportServices":0.114,"PublicAdminDefence":0.114,"Education":0.087,"HealthSocialWork":0.087,"OtherServices":0.0514},"Liverpool":{"AgricultureForestryFishing":0.067,"MiningUtilities":0.067,"Manufacturing":0.135,"Construction":0.067,"WholesaleRetailRepairOfVehicles":0.095,"TransportStorage":0.095,"AccommodationFoodServices":0.084,"InformationCommunication":0.103,"FinancialServicesRealEstate":0.042,"ProfScientificTechnicalActiv":0.029,"AdminSupportServices":0.029,"PublicAdminDefence":0.029,"Education":0.014,"HealthSocialWork":0.014,"OtherServices":0.111},"Omaha":{"AgricultureForestryFishing":0.020,"MiningUtilities":0.020,"Manufacturing":0.063,"Construction":0.020,"WholesaleRetailRepairOfVehicles":0.090,"TransportStorage":0.090,"
```

```
AccommodationFoodServices":0.101,"InformationCommunication":0.018,"↩
FinancialServicesRealEstate":0.091,"ProfScientificTechnicalActiv":0.074,"↩
AdminSupportServices":0.105,"PublicAdminDefence":0.105,"Education":0.085,"↩
HealthSocialWork":0.085,"OtherServices":0.038},"Seattle":{"AgricultureForestryFishing"↩
:0.020,"MiningUtilities":0.020,"Manufacturing":0.080,"Construction":0.020,"↩
WholesaleRetailRepairOfVehicles":0.095,"TransportStorage":0.095,"↩
AccommodationFoodServices":0.092,"InformationCommunication":0.075,"↩
FinancialServicesRealEstate":0.047,"ProfScientificTechnicalActiv":0.078,"↩
AdminSupportServices":0.104,"PublicAdminDefence":0.104,"Education":0.071,"↩
HealthSocialWork":0.071,"OtherServices":0.},"Scranton":{"AgricultureForestryFishing"↩
:0.013,"MiningUtilities":0.013,"Manufacturing":0.093,"Construction":0.013,"↩
WholesaleRetailRepairOfVehicles":0.130,"TransportStorage":0.130,"↩
AccommodationFoodServices":0.086,"InformationCommunication":0.009,"↩
FinancialServicesRealEstate":0.050,"ProfScientificTechnicalActiv":0.058,"↩
AdminSupportServices":0.084,"PublicAdminDefence":0.084,"Education":0.108,"↩
HealthSocialWork":0.108,"OtherServices":0.0301}},"2027":{"Barry":{"↩
AgricultureForestryFishing":0.0252,"MiningUtilities" :0.0252,"Manufacturing" :0.0818,"↩
Construction" :0.0252,"WholesaleRetailRepairOfVehicles" :0.010,"TransportStorage" ↩
:0.010,"AccommodationFoodServices" :0.197,"InformationCommunication" :0.0643,"↩
FinancialServicesRealEstate" :0.0584,"ProfScientificTechnicalActiv" : 0.114,"↩
AdminSupportServices" :0.114,"PublicAdminDefence" :0.027,"Education" :0.0874,"↩
HealthSocialWork" :0.08735,"OtherServices" :0.0507},"Liverpool":{"↩
AgricultureForestryFishing" :0.065,"MiningUtilities": 0.065,"Manufacturing" :0.138,"↩
Construction" :0.065,"WholesaleRetailRepairOfVehicles" :0.095,"TransportStorage" ↩
:0.095,"AccommodationFoodServices" :0.081,"InformationCommunication" :0.103,"↩
FinancialServicesRealEstate" :0.043,"ProfScientificTechnicalActiv" :0.057,"↩
AdminSupportServices" :0.042,"PublicAdminDefence" :0.042,"Education" :0.013,"↩
HealthSocialWork" :0.013,"OtherServices" :0.110},"Omaha":{"AgricultureForestryFishing"↩
 :0.020,"MiningUtilities" :0.020,"Manufacturing" :0.062,"Construction" :0.020,"↩
WholesaleRetailRepairOfVehicles" :0.086,"TransportStorage" :0.086,"↩
AccommodationFoodServices" :0.102,"InformationCommunication" :0.017,"↩
FinancialServicesRealEstate" :0.092,"ProfScientificTechnicalActiv": 0.075,"↩
AdminSupportServices" :0.106,"PublicAdminDefence" :0.106,"Education" :0.088,"↩
HealthSocialWork" :0.088,"OtherServices" :0.039},"Seattle":{"↩
AgricultureForestryFishing" :0.020,"MiningUtilities" :0.020,"Manufacturing":0.076,"↩
Construction" :0.020,"WholesaleRetailRepairOfVehicles" :0.095,"TransportStorage" ↩
:0.095,"AccommodationFoodServices" :0.092,"InformationCommunication" :0.081,"↩
FinancialServicesRealEstate" :0.046,"ProfScientificTechnicalActiv": 0.080,"↩
AdminSupportServices" :0.103,"PublicAdminDefence" :0.103,"Education" :0.072,"↩
HealthSocialWork" :0.072,"OtherServices" :0.037},"Scranton":{"↩
AgricultureForestryFishing":0.013,"MiningUtilities":0.013,"Manufacturing":0.087,"↩
Construction":0.013,"WholesaleRetailRepairOfVehicles":0.133,"TransportStorage":0.133,"↩
AccommodationFoodServices":0.087,"InformationCommunication":0.008,"↩
FinancialServicesRealEstate":0.050,"ProfScientificTechnicalActiv":0.059,"↩
AdminSupportServices":0.057,"PublicAdminDefence":0.057,"Education":0.110,"↩
HealthSocialWork":0.110,"OtherServices":0.110}}}
#(Determined in model 1)

#percentage chance that a person in each city can/cannot work from home due to their house↩
  (not) being suitable
housePercent = {"Barry":{True:0.9,False:0.1},"Liverpool":{True:0.9,False:0.1},"Omaha":{↩
True:0.9,False:0.1},"Seattle":{True:0.9,False:0.1},"Scranton":{True:0.9,False:0.1}}

educationPercent = {"Barry":{"noQualifications":0.063,"entryLevel":0.189,"gcseOrEquiv"↩
:0.393,"aLevelOrEquiv":0.193,"degreeOrEquiv":0.064,"higherDegree":0.098}
,"Liverpool":{"noQualifications":0.064,"entryLevel":0.192,"gcseOrEquiv":0.332,"↩
aLevelOrEquiv":0.077,"degreeOrEquiv":0.232,"higherDegree":0.103}
,"Omaha":{"noQualifications":0.025,"entryLevel":0.075,"gcseOrEquiv":0.206,"aLevelOrEquiv"↩
:0.31,"degreeOrEquiv":0.246,"higherDegree":0.131}
,"Scranton":{"noQualifications":0.046,"entryLevel":0.138,"gcseOrEquiv":0.293,"↩
aLevelOrEquiv":0.277,"degreeOrEquiv":0.149,"higherDegree":0.095}
,"Seattle":{"noQualifications":0.027,"entryLevel":0.082,"gcseOrEquiv":0.154,"aLevelOrEquiv↩
":0.295,"degreeOrEquiv":0.269,"higherDegree":0.173}}

timePercent = {"Barry":{"Full Time":0.75,"Part Time":0.25},"Liverpool":{"Full Time":0.75,"↩
Part Time":0.25}
,"Omaha":{"Full Time":0.75,"Part Time":0.25}
,"Seattle":{"Full Time":0.75,"Part Time":0.25}
```

```python
,"Scranton":{"Full Time":0.75,"Part Time":0.25}}


totalWFH = 0 #total people WFH starts at 0.

def recursiveGen(amount,city,prev,step,year):
  """Recursively generates a population of size amount for city, assuming the existing
     characteristics already determined outlined in prev, for year year.
     The step indicates which characteristic we are currently deciding"""

  global totalWFH

  if step == 0:
    for sex in sexes:
      recursiveGen(sexPercent[city][sex]*amount,city,prev+[sex],step+1,year) #for all possible sexes with probability (by reducing amount)
  elif step == 1:
    for ethnicity in ethnicities:
      recursiveGen(ethnicityPercent[city][ethnicity]*amount,city,prev+[ethnicity],step+1,year) #for all possible ethnicities with probability (by reducing amount)
  elif step == 2:
    for industry in industries:
      recursiveGen(industryPercent[year][city][industry]*amount,city,prev+[industry],step+1,year) #for all possible industries with probability (by reducing amount)
  elif step == 3:
    for education in educations:
      recursiveGen(educationPercent[city][education]*amount,city,prev+[education],step+1,year) #for all possible educations with probability (by reducing amount)
  elif step == 4:
    for time in times:
      recursiveGen(timePercent[city][time]*amount,city,prev+[time],step+1,year)#for all possible times with probability (by reducing amount)
  elif step == 5:
    for house in houses:
      recursiveGen(housePercent[city][house]*amount,city,prev+[house],step+1,year) #for all possible house characteristics with probability (by reducing amount)
  else: #base case
    for person in range(int(amount)): #actually generate all of the people and determine the chance of them working from home post pandemic
      #Here we incorporate the age, post pandemic attitudes constant, and commute distance factors.
      totalWFH += max(0,
                      min(
                        c*PWp_n_A_R(Person(
                          prev[0],prev[1],prev[2],prev[3],prev[4],
                          random.randint(0,10000),
                          prev[5],
                          max(min(
                            np.random.normal(mu, sigma),80),
                              20))),
                      1)) #applying model 2

for year in years: #For both 2024 and 2027
  print("For year: " + str(year))
  for city in cities: #For all cities
    print("For city: " + city)

    totalWFH = 0 #total people WFH starts at 0.
    recursiveGen(population[year][city],city,[],0,year) #generate a representative population and sum the wfh over the population.
    print(totalWFH/population[year][city]*100) #Obtain percentage of people that will now WFH.
```