MathWorks Math Modeling Challenge 2024

F.W. Buchholz High School

Team #17497, Gainesville, Florida

Coach: Ziwei Lu

Students: Melissa Li, Sophia Rong, Nathan Wei, Andrew Xing, Luke Xue

M3 Challenge TECHNICAL COMPUTING WINNER—$3,000 Team Prize

JUDGE COMMENTS

Specifically for Team #17497—Submitted at the close of triage judging

COMMENT 1: Models were clearly explained, which shows a thorough understanding of model choices. The entire submission was very well written

COMMENT 2: You presented a nice executive summary with a statement of questions and specific results, as well as nice detail in looking to model seasonal variation in home sales by month. Perhaps the simple increase annually is what might best address our needs when trying to predict the increases or changes of the decades. Perhaps a graph of homelessness over the decades in question would be beneficial.

COMMENT 3: This report was well articulated! They have a very good command of grammar, model generation and its considerations, and finally, the application of the models to solve real issues.

COMMENT 4: The team's submission used two interesting simulation models and one fitted SARIMA. The simulation models are used to study the factors contributing to the homeless population and to explore policy implications for Question 3. The inclusion of the code that implemented the design is a commendable practice, enhancing the transparency and reproducibility of the work. This approach showcases the team's ability to apply theoretical concepts practically.

For the proposed models, more in-depth quantitative analysis of the models' performance and outcomes is desirable. This additional analysis could provide deeper insights into the models' effectiveness and potential limitations.

Your team's ability to break down complex problems into smaller ones is impressive. The use of a simulation model to examine how different factors, such as age and income, is great. This highlights your team's practical skills and innovative thinking. Consider expanding the quantitative analysis of your models. This could involve exploring the sensitivity of your models to different parameters, assessing the models' robustness under various scenarios, or conducting a comparative analysis with other modeling approaches.

The simulation approach involved many details. Iterate on your assumptions and their implications for the applicability and accuracy of your models. While complex models can be powerful, consider first if there are simpler models or methods that could establish a baseline for gaining intuition from the data. Keep up the excellent work and continue to explore the vast potential of mathematical modeling.

Executive Summary

Dear Secretary Fudge,

We are writing to alert you of the imminent danger today's housing and homelessness crises will pose to your constituents without effective government action. In the last decades, growth in home prices has rapidly outpaced real incomes, making the American dream of home ownership less and less likely. Even worse, rising home prices have driven Americans out of homes, and have been linked to a 38% increase in urban homelessness in just this past year. In this report, our team has created three targeted models to examine the long-term changes in the housing market, changes in homeless populations, and how to best assist local governments in creating targeted plans for combating homelessness.

The first section of our report models the growth of the available housing supply within two vibrant metropolises in the US – Albuquerque, New Mexico and Seattle, Washington. We obtained data on the number of new listings, home sales, and construction of new homes per month in both regions. Due to the inherent seasonality of the housing market, we implemented a seasonal autoregressive integrative moving average (SARIMA) model to represent the growth of vacant homes based on the data we collected. Our model predicts that there will be 49, 70, and 113 thousand vacant homes 10, 20, and 50 years from now in Albuquerque, and 31, 56, and 165 thousand in Seattle. The significant growth of available housing suggests that homelessness is not due to a physical lack of housing, but the inability to afford current housing.

In the second section of our report, we created a Monte Carlo simulation that evaluates the economic factors behind homelessness to model the growth of the homeless population in both regions. By comparing a person's income to the cost of their "bare necessities" (food, shelter, transportation, healthcare, and clothing), we determine if a chosen individual would be forced to give up their home in a given year to offset overburdened consumption by removing the largest cost: housing, thereby becoming homeless. Our model predicts a population of 2240, 2379, and 4075 homeless people in Albuquerque 10, 20, and 50 years from now, and that there would be 3813, 4624, and 8052 homeless people in Seattle. The outsized portion of income taken up by housing costs also suggests that it is a prime target for government intervention to help alleviate the financial strains that lead to homelessness.

The final section of our report seeks to assist local city governments in creating a long-term plan for addressing homelessness by evaluating the long-term effects of rent vouchers for decreasing the homeless population. We re-simulated our Monte Carlo simulation with varying levels of rent subsidies to determine their effect on the homeless population in the long-term, using Seattle as our model community. Our results indicate that a $10,000 annual rent voucher could reduce homeless populations by 50.4% within the next 20 years in Seattle. Our model is easily transferable to other populations, and can also account for a variety of unforeseen circumstances, including natural disasters, economic recessions, and increased migrant populations.

The results of our model indicate that without government intervention, the homelessness crisis will continue to plague our cities, subjecting thousands to inhumane living conditions and an endless winter of despair. However, by targeting one of the root causes of homelessness – rising housing prices – local and federal government action can effectively combat this rampant issue.

We strongly urge the federal government to consider these findings in their goal to eliminate homelessness and foster a new spring of hope.

Contents

0.1 Global Assumptions . . . 3

1 Q1: It was the Best of Times **3**
1.1 Problem Restatement . . . 3
1.2 Assumptions . . . 3
1.3 Variables . . . 4
1.4 Our Model . . . 4
1.4.1 Granger Causality . . . 4
1.4.2 Seasonal ARIMA Model . . . 5
1.4.3 Modeling Process . . . 5
1.5 Results . . . 7
1.6 Discussion and Analysis . . . 7
1.6.1 Strengths and Weaknesses . . . 8

2 Q2: It was the Worst of Times **8**
2.1 Problem Restatement . . . 8
2.2 Assumptions . . . 8
2.3 Variables . . . 10
2.4 Our Model . . . 10
2.4.1 Determining the Parameters . . . 10
2.4.2 Distributions of Data . . . 10
2.4.3 Monte Carlo . . . 11
2.5 Results . . . 13
2.6 Discussion and Analysis . . . 13
2.6.1 Strengths and Weaknesses . . . 14

3 Q3: Rising from this Abyss **14**
3.1 Problem Restatement . . . 14
3.2 Assumptions . . . 15
3.3 Variables . . . 15
3.4 Our Model . . . 15
3.5 Results . . . 15
3.6 Discussion and Analysis . . . 16
3.6.1 Strengths and Weaknesses . . . 16

4 Conclusion and Future Directions **17**

5 References **18**

6 Appendix **20**
6.1 Model 1: SARIMA . . . 20
6.2 Model 2: Monte Carlo . . . 23

0.1 Global Assumptions

1. **The birth rate for the next 50 years will remain approximately constant.** Generally, births will be proportional to economic condition, but it is unreasonable to extrapolate the state of the economy years from now. Thus, we will assume that current trends in birthrates will continue into the future.

1 Q1: It was the Best of Times

1.1 Problem Restatement

Prospective homeowners are facing mounting obstacles in finding suitable housing amidst widespread scarcity [1]. In the first section of our report, we predict the changes within the housing supply of two fast-growing US cities — Albuquerque, New Mexico and Seattle, Washington — within the next 10, 20, and 50 years. We will then indicate the level of confidence in our findings.

1.2 Assumptions

1. **The home listings and sales found on Zillow represent the entirety of the housing market within these regions.** Zillow receives many of its listings directly from different realtor companies and multi-listing services [2]. While some houses may be sold through off-the-market deals, this represents a very small proportion of homes, and these homes would likely not have been accessible for many purchasers. There is often some delay from a home's listing to its appearance on Zillow, but this is not a problem when examining historical data.
2. **The construction of new housing will continue at a relatively constant rate.** Due to a growing lack of buildable plots within cities, it is harder to find locations to build new houses, particularly within the urban areas we are examining. However, local governments are taking more action to close this gap and provide more affordable housing. For example, Albuquerque has the Housing Forward act, which pledges to add 5000 housing units by 2025 [3]. The Seattle Housing Levy likewise has built over 2700 housing units in the last 7 years [4]. As the M3 Challenge Problem notes, the construction of new housing is slow and financially expensive, and we do not expect that work by local governments to be able to exponentially grow the housing supply. The newest Seattle Levy plan only calls for 3100 new houses to be built in the next 7 years, which represents just a modest increase from their previous achievement. In the absence of large economic shifts that make construction of new housing significantly more feasible, we will assume that new housing growth is constant due to these limitations. We will also justify this assumption further when developing our model.

1.3 Variables

Variable	Definition	Unit
V	Available or Vacant Housing	# houses
L	New Listings: Houses put on Sale	# houses
C	New Houses Constructed	# houses
S	Houses sold	# houses

1.4 Our Model

Because we are interested in examining the housing market in the context of homelessness, we chose to specifically examine the houses that were currently vacant. This subset of houses is most relevant as they are the ones that are readily available for a new household to live in. To model the change in this supply, we classify the various ways a house can enter or exit this category. Newly constructed or newly listed houses will be entering the supply, while houses being sold or rented out are exiting the market. We obtained monthly data on new listings, sales, and construction of new housing from 2018-2024 via Zillow in the two regions [5].

Using data curated by the Mathworks M3 Challenge, we found the current number of vacant housing units within both of our regions, which we treat as the initial housing supply.

Region	Current Vacant Housing
Albuquerque	15,378
Seattle	27,190

Figure 1: Current Vacant Housing

Assuming that construction of new housing remains consistent into the future, we calculate the number of houses built each month by averaging the monthly construction data from 2018-2024 found on Zillow, rounding to obtain a whole number [5].

Region	New Houses Built per Year
Albuquerque	65
Seattle	596

Figure 2: New Construction per Month

To determine the monthly changes in housing supply $\frac{dV}{dt}$, we will use the following function, where C_t, L_t, and S_t represent the number of houses built, listed, or sold in that month respectively:

$$\frac{dV}{dt} = C_t + L_t - S_t$$

1.4.1 Granger Causality

We initially hypothesized that the the gap between new listings and home sales could influence new construction, as builders might want to start more homes when demand outpaced current supply, or slow down building during periods with a surplus of supply. To further prove that the construction of new housing can be treated as a constant and is not susceptible to these short-term influences, we performed a Granger causality test comparing new construction to the difference between new listings and sales per month. Granger causality is a statistical hypothesis test that determines whether or not one time series is a useful predictor of another time series. It looks for patterns that repeat in both time series after some time lag. However, upon performing Granger causality using the data provided by Zillow, we obtained a p-value of 0.53, indicating we could not reject the null hypothesis that the number of vacant houses does not forecast the number of houses constructed. Thus, for the purpose of this model, we assumed that new construction would remain constant.

1.4.2 Seasonal ARIMA Model

To determine the values of L_t and S_t per month, we decided to utilize a Seasonal Autoregressive Integrated Moving Average (seasonal ARIMA/SARIMA) model. The housing market is known to have extreme seasonal trends [6]. Most houses are sold during the summer months, as people are less likely to move during the holiday season. This makes an ARIMA model effective, as it allows for the preservation of these seasonal trends. While SARIMA is not extremely effective at predicting changes in long time periods, it is well understood that the housing market displays seasonality and we do not expect the human behaviors that cause this to change in this time span. Thus, SARIMA is an excellent model of choice because of its ability to represent longer term trends as well as short term seasonality.

Auto-regressive integrated moving average is a model that analyzes time series data to predict future time-related events. The working principle behind ARIMA models is a combination of an auto-regressive model and a moving average model. An auto-regressive model is a model that predicts current values via a linear combination of previous values in time. The number of prior values used is specified using a parameter p. A moving average model states that current values are equal to the mean of the series plus white-noise error terms, with the number of error terms specified by another parameter q. The final parameter d represents the number of times finite differences are taken between terms to achieve constancy. SARIMA models utilize a similar principle, but also incorporate 4 additional parameters: P, D, Q, and s. P, D, and Q are seasonal analogues of the ARIMA parameters p, d, and q, which allows the SARIMA model to more effectively predict scenarios with significant seasonality, such as housing markets. The final parameter s is used to indicate the length of this seasonal cycle.

1.4.3 Modeling Process

To accurately predict changes in the housing market, we must tune the parameters to minimize our error. Historical analysis of the housing market suggested that it cycles annually. As our data was represented by month, we set $s = 12$ to represent this seasonality. We then used the Augmented Dickey-Fuller (ADF) test to determine d, taking finite differences of the time series data until the ADF test returned a p-value less than our alpha level of 0.05. This occurred after one finite difference, rejecting the null hypothesis that our time series is non-stationary for both Albuquerque and Seattle. To calculate p, we looked at the partial autocorrelation function (PACF), which gives the correlation between series values that are separated by a time lag ignoring the contributions to the autocorrelation by intermediate values. We found that the first lag value was the most significant, and the second lag value was relatively insignificant, meaning $p = 1$ for both Albuquerque and Seattle. To find q, we looked at the autocorrelation function (ACF) in order to find values with a high correlation with respect to the first value. Only one value had an autocorrelation greater than our threshold of 0.3, so we set $q = 3$ for Albuquerque, and $q = 1$ for Seattle. This means that it requires one moving average term to remove the autocorrelation from successive terms. The values of P, Q, and D were set to 4, 1, and 1 respectively to fit the given data the best.

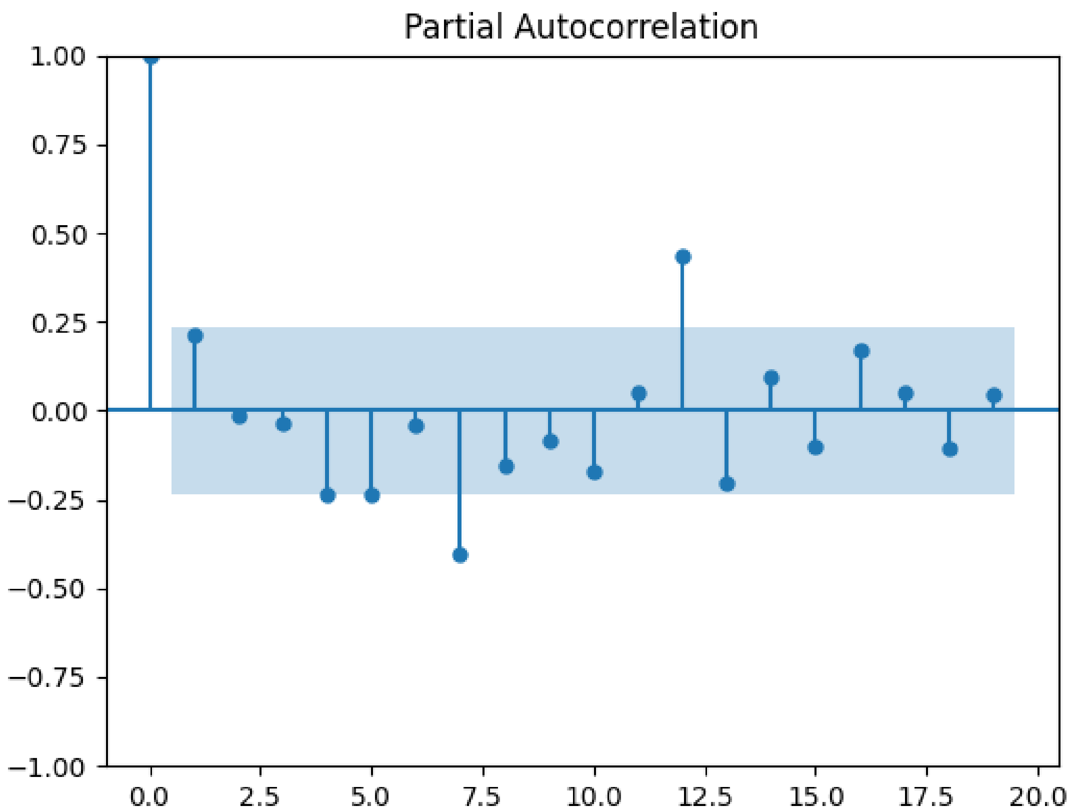

Figure 3: Partial Autocorrelation of House sales in Seattle. The high partial autocorrelation value of the 12th month represents a seasonal variation of 12 months.

In order to obtain accurate results, we adjusted our time series to have a constant mean and variance. Although one finite difference ensured a roughly constant mean, this could not ensure a constant variance. For this, we used the box-cox transformation. The box-cox transformation transforms a data series into a normally distribution by taking the natural logarithm of all data points and raising them to the λ power if λ is nonzero, which is calculated by the skewness of the data. Because we used 0 for λ, the natural logarithm was applied to each data point. After this transformation on the initial dataset, we took a finite difference using the shift() function to have a constant average. This allowed us to finally use the SARIMA model on our transformed dataset. Using our parameters listed above, we trained the model on our dataset of 72 months and used it to predict the next 600 months.

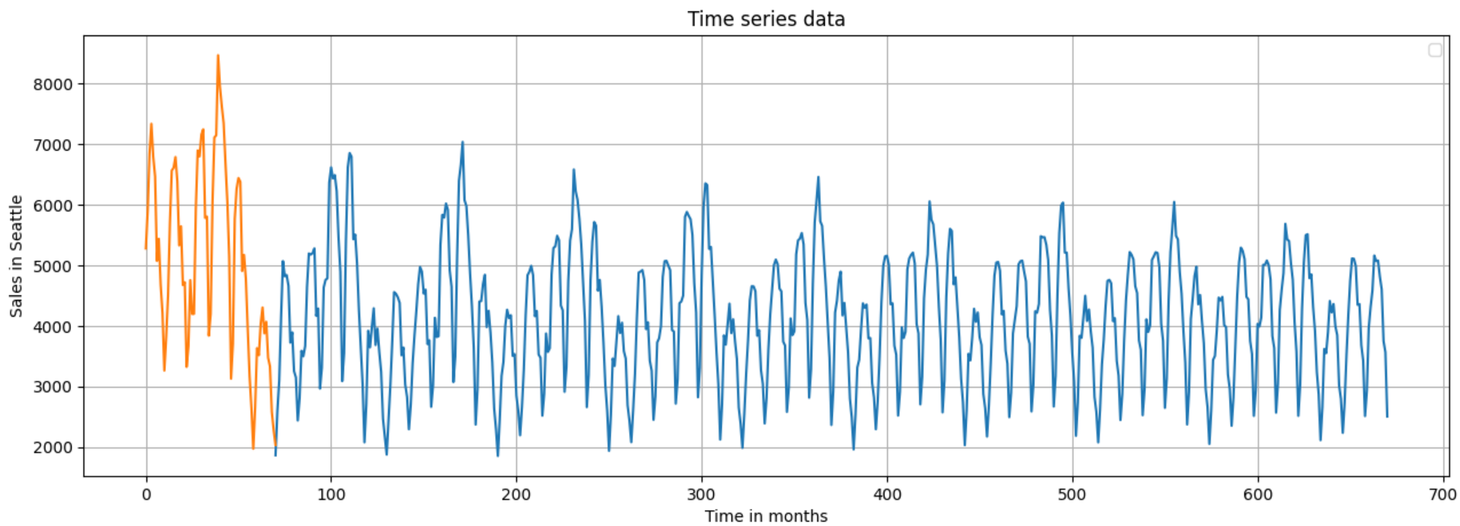

Figure 4: Seattle housing sales over time, with blue representing predicted values and the orange represented given data values

1.5 Results

All margins of errors are considered as the 95% confidence interval.

Years since 2024	Housing Supply	% Change
0	15,378	-
10	48,896 ±2638	218%
20	69,845 ±2654	354%
50	113,255 ±2690	636%

Figure 5: Albuquerque: Change in Housing Supply

Years since 2024	Housing Supply	% Change
0	27,190	-
10	31,148 ±1369	14.5%
20	55,845 ±3769	105%
50	164,536 ±2196	505%

Figure 6: Seattle: Change in Housing Supply

1.6 Discussion and Analysis

Our model shows that the amount of vacant housing available will increase over time. Within the next 10 and 20 years, the results of our model do make sense. Over time, we expect the total amount of housing to increase, which likely means the amount of vacant or available housing will also increase proportionally. Houses have high longevity, meaning we would not expect a sudden decrease in housing stocks without a significant, unpredictable natural disaster.

However, when we use our model to examine results in a significantly longer term future, the results deviate from what would be expected. Our model predicts an 636% increase in vacant housing in Albuquerque and a similar 505% increase in Seattle after 50 years, which deviates from a common sense standpoint. One potential reason for this is high deviation is our assumption of a constant rate of construction of new housing. Eventually, the lack of buildable lots may make potential government subsidies or government sponsored building programs fiscally irresponsible, causing these programs to be scaled down. However, as we have no way to predict the state of the economy 50 years from now, it is unfeasible to model these changes in construction rates.

We have high confidence that our model effectively predicts shorter-term changes in the housing supply. However, it performs less effectively at a 50 year span, making our prediction more unreasonable.

1.6.1 Strengths and Weaknesses

The use of a seasonal ARIMA or SARIMA model allows us to represent the consistent seasonality that is present in the housing market. Through the Augmented Dickey-Fuller test, we were able to objectively determine most of SARIMA's parameters rather than guessing and checking. The box-cox transformation drastically improved our SARIMA result as our input dataset now had a roughly consistent mean and variance.
ARIMA and SARIMA models in general are highly dependent on past data, making it difficult to extrapolate heavily in the future. However, analysis of the housing market in the past shows that it has historically followed a seasonal trend, which gives us higher confidence in the SARIMA model's prediction.
As with any model attempting to predict a lengthy 50 years in the future, there is a heavy amount of variability, since any number of unpredictable events can impact the accuracy of your model. This makes any potential model, including ours, unreliable for this time span, as you must assume a lack of any significant change to the current status quo.

2 Q2: It was the Worst of Times

2.1 Problem Restatement

Homelessness is intimately related to the availability of affordable housing and other economic indicators [7]. In this section, we aim to predict the homeless population in our chosen regions within the next 10, 20 and 50 years based on an individuals income and their cost of various expenditures.

2.2 Assumptions

1. **If an individual's income and remaining savings are not sufficient to account for their basic necessities, they will become homeless.** In order for a person to survive, there are certain necessary expenditures. For the purpose of our model, we will count food, healthcare, transportation, housing, and clothing as our necessary expenses. When a person's income and savings are insufficient to cover these, we argue that they will be forced to become homeless as they will not have enough money to sustain these necessities. For example, food is inherently necessary for survival, and

healthcare expenses are necessary in case of a life-threatening injury. Clothing protects from the elements and transportation is necessary to maintain a job, both of which are critical.

2. **Our distributions of prices and income will remain the same for the foreseeable future.** It will be nearly impossible to predict these macroeconomic changes within our time frame, so we will assume that the current status quo will continue into the future.
3. **The government will not intervene to solve the homelessness crisis.** Government actions in the past have seen success at reducing homelessness [8]. However, it is difficult to predict future motivations for these expensive government actions, especially as the federal government faces record-high deficits.
4. **Out of the homeless population, 22% will stay homeless and 78% will recover in $\frac{2}{3}$ of a year.** "Chronic homelessness" accounts for 22% of all homeless people and includes debilitating mental health disorders, physical disabilities, and substance abuse that will prevent them from being able to find and adjust to permanent shelter [9]. Out of the 78% remaining, the median length of time to re-house is 241 days, which is approximately $\frac{2}{3}$ of the year [10].
5. **Incomes will stay fixed over one's lifetime.** We assume that promotions/change in jobs are not significant enough to impact one's living situation, as these will mostly be caused by inflation rather than true improvement in quality of work.
6. **People retire at age 65.** This is the average age of retirement and also when an individual can collect full Social Security benefits [11].
7. **On average, people stop living with their parents at age 25.** In a national survey, the Bureau of Labor Statistics found that the vast majority of young adults have left their parental households at age 25. When younger, they may still rely on their parents for certain responsibilities such as health insurance, which generally expires at age 26 [12]. Additionally, social stigmas make it less likely for older adults to rely on their parents, making age 25 a good baseline for this event. This means they are now susceptible to homelessness, as they are responsible for their own living arrangements.
8. **People will not die before their retirement age.** Premature deaths from accidents at a young age are rather rare and unpredictable. The majority of people will live through their working life.
9. **Workers will save approximately $\frac{2}{3}$ of their remaining wages after paying for basic necessities.** The money after paying for basic necessities represents a worker's disposable income. While they may choose to spend much of this on consumption, we assume that workers will save some money to act as a safety net in case of unforeseen situations. Additionally, this 2/3 also accounts for the purchase of long-term consumables, such as furniture, finer clothing, and cookware, which in the event of a financial crisis could be sold to prevent homelessness.

2.3 Variables

Variable	**Definition**	**Unit**
N	Price of Necessity N	Dollars
CR_N	Conversion Rate of the prices of necessity N from the national average to each region	Dollars
H	Number of Total Homeless People in a Year	People
H_s	Number of Homeless People at a Specific Time of Observation in the Year	People

2.4 Our Model

2.4.1 Determining the Parameters

An individual at any income level can be at risk of homelessness as a result of sudden price shocks arising from medical emergencies, job losses, or the accumulation of miscellaneous costs. When financial pressures build, they will adjust by searching for methods to minimize expenses. When evaluating the likelihood of an individual being forced into homelessness, we only need to compare their income to the cost of their bare necessities for survival, as it does not make rational sense for a human to give up their home for unnecessary luxury expenses.

After looking over the different categories of spending, we determined that the key necessities for living are food, healthcare, transportation, housing, and clothing. Using data from the U.S. Bureau of Labor Statistics, we found data for the distributions of national average expenditures per household unit for the bottom 20th percentile of the population, since this best represents the cost of these bare necessities. The average household unit in this range contains 1.6 members [13]. Since these expenditures will generally be spent evenly among all members, we will divide this data by 1.6 to find the necessary expenditures for one individual. Although this ignores some finer subtleties such as differences in consumption due to gender or age, we believe these differences are minor, especially when examining the critical necessities for survival.

2.4.2 Distributions of Data

With our data from the Labor Bureau [13], we were able to find distributions for the necessary national expenditure of each essential good of the average adult. We then calculated the distribution of expenditures for each city based upon conversion ratios [14] with

$$N = CR_N \cdot \text{National Average}$$

for each necessity N. Additionally, we were able to find data for the distributions of incomes in both cities from the US Census Bureau [15]. This provided us with a mean income for both regions. However, without a given sample size we were unable to determine a standard deviation from the survey data. We therefore assumed a value of $\$40,000$ for the standard error (SE) of income in Seattle and $\$20,000$ for the standard error of income in Albuquerque. Our results are in the following table:

	Albuquerque Mean	**Albuquerque SE**	**Seattle Mean**	**Seattle SE**
Income	86,268	20,000	167,027	40,000
Food	3,117.63	122.96	3,976.56	156.82
Healthcare [16]	1,908.16	101.06	2,725.94	144.38
Transport	2,864.4	170.47	3,788.4	225.46
Housing	7,356	136.8	17,639.6	326.52
Clothing [17]	565	120.83	678	145

Figure 7: Cost of Living in Albuquerque and Seattle ($)

2.4.3 Monte Carlo

We created a Monte Carlo Simulation using the Numpy package in Python and simulated the individuals in each of the cities. Each simulation represented the tendency of an individual to fall into homelessness over a specific number of years. We ran this Monte Carlo with a number of individuals equal to the populations of Albuquerque, New Mexico and Seattle, Washington with assumed populations of 560,000 and 734,000 respectively.

2.4.3.1 Age Group

We simulated each age group separately, as they will spend different amounts of time in the workforce. For example, someone currently in the 5-9 age group would be 55-59 after 50 years and therefore still be of working age. However, they would be not considered working until they turn 25 for the purpose of our model, yielding a net number of years working of 30. For unborn generations, we used our universal assumption that birth rates will remain relatively constant to use the past 20 years of birth percentages to derive average future birth rates. This yielded an average size of 30,000 for all future age groups. The age group sizes and years working 50 years in the future can be observed in the following graph.

Age Group (in 2022)	% of Population	Total Population	Years Working
(-25) to (-21)	6.33	30000	0
(-20) to (-16)	6.33	30000	5
(-15) to (-11)	6.33	30000	10
(-10) to (-6)	6.33	30000	15
(-5) to (-1)	6.33	30000	20
0 to 4	5.2	29253	25
5 to 9	5.5	30941	30
10 to 14	6.4	36004	35
15 to 19	6.6	37129	40
20 to 24	6.8	38254	45
25 to 29	7.5	42192	40
30 to 34	7.6	42754	35
35 to 39	7.1	39942	30
40 to 44	6.5	36566	25
45 to 49	5.9	33191	20
50 to 54	5.9	33191	15
55 to 59	6.3	35441	10
60 to 64	6.1	34316	5

Figure 8: Simulating 50 Years in the future for Each Age Group in Albuquerque

2.4.3.2 Expenditures

Next, we determined the expenditures on necessities using the previously calculated normal distributions for both cities. Each individual would thus experience varying prices for each of food, healthcare, transportation, housing, and apparel, creating a degree of randomness to reflect real-life differences in necessity costs.

2.4.3.3 Income

To determine whether they became homeless, we assigned each individual with an income taken from a normal distribution of incomes for Seattle and Albuquerque. As one of our assumptions stated, we will ignore income changes, so this annual income will remain constant throughout this individual's life. To account for possible layoffs, we included a 1% chance [22] of the individual losing their job in a given year, as it is often the unpredictable catastrophes that send a person spiraling into homelessness [23]. Using the fact that average unemployment lasts for 7.5 weeks and 80% of people regain a job after 12 weeks [28], we calculated the mean and standard deviation of weeks that someone is unemployed for. If an individual is unemployed for less than 12 weeks, or 3 months, they will still receive income because of unemployment benefits. Once an individual has lost their job for more than three months, they will lose unemployment benefits and will become homeless [24]. As a way to offset expenditures, we allowed a proportion of the money an individual had remaining after subtracting expenditures from income to carry over into the next year. This proportion was based on our assumption that approximately two-thirds of money would be saved from one year to the next.

2.4.3.4 Overview

Based on our first assumption, we deemed an individual homeless if their total necessary expenditure exceeds the sum of their annual income and accrued savings. As we ran different simulations for each age group, the total change in the number of homeless people would be an aggregate of the number of people becoming homeless for each age group. In addition, we considered homeless people to be of two groups: temporarily or permanently homeless. As explained in our 4th assumption, approximately 22% of homeless people stay homeless and 78% will recover in $\frac{2}{3}$ of a year. This means 22% of homeless will be considered permanently homeless and 78% temporarily. In a given year, only $\frac{2}{3}$ of the temporarily homeless will be homeless at the time we observe, yielding the following formula:

$$H_s = 0.22H + 0.78H(\frac{2}{3})$$

Where H_s represents the number of homeless people at a specific time of observation in the year and H the total number of homeless people in a year.

2.5 Results

Below are our Monte Carlo predictions of the homeless populations for the next 10, 20, and 50 years, along with a graph of the projected numbers.

Years	**Albuquerque**	**Seattle**
10	2240	3813
20	2379	4624
30	4075	8052

Figure 9: Predicted Change in Homeless Populations in Albuquerque and Seattle

2.6 Discussion and Analysis

To verify the validity of our results, we can observe the changes in the homeless population in Albuquerque and Seattle that were provided [27]. Although it appears that homelessness in Albuquerque remains relatively constant from 2007 to 2021, we can attribute these changes to the government programs that have been instituted to specifically decrease homelessness. On the other hand, the following graph depicts the change in Seattle's homeless population over the course of 10 years, with the starting year ranging from 2011 to 2015. Seeing how our estimate of an increase in 3,813 homeless people in the next 10 years falls reasonably within the graph, we have a basis to believe our model is accurate.

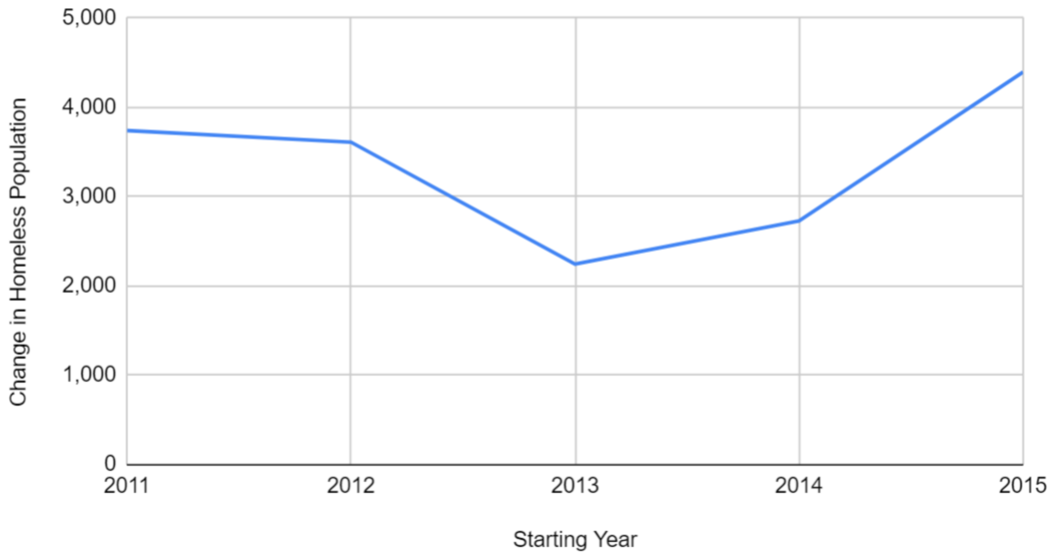

Figure 10: Change in Seattle Homeless Population After 10 Years

Further, assuming that current economic conditions stay relatively similar, our predictions for 20 and 50 years will be fairly precise, as the number of homeless will continue to rise over time. This is mostly due to various economic fluctuations that will hurt the population most at risk of homelessness.

2.6.1 Strengths and Weaknesses

Our Monte Carlo simulation allows us to easily see which individuals are highly susceptible to homelessness, which will allow for more effective government intervention. Additionally, our usage of the prices of necessities enables us to see which expenditures are more impactful upon a consumer.

The Monte Carlo model is also quite robust, since there is a significant amount of built-in uncertainty from generating random samples of probability distributions. This is crucial in a volatile scenario like the economy, where any minute change will drastically affect many households.

Although our model is able to accurately predict homelessness for the next 10 years, it will likely be somewhat inaccurate for values 50 years into the future, as we are unable to account for widespread unpredictable events like technological advancements.

Moreover, our model does not perfectly describe the intricacies behind causes of homelessness, such as addiction or mental illness. However, with more information, we may be able to implement a factor based upon ”chronic homelessness,” as this may change based on cities’ different social environments. This would allow for a more accurate description of how substance abuse or debilitating disorders affect one’s ability to rise out of homelessness.

3 Q3: Rising from this Abyss

3.1 Problem Restatement

The first two sections of our model indicate that there is a significant excess of available housing compared to the homeless population in our chosen regions. Current research also support this, as the United Way shows most cities have a shockingly high ratio of vacant housing to the homeless population [21]. The major problem lies in making existing housing affordable for the homeless. Our final model examines the potential effect of government housing subsidies to optimally prevent "at-risk" individuals from experiencing homelessness within Seattle, Washington. We additionally evaluate how our model could be adjusted to account for unforeseen circumstances such as natural disaster, economic recession, or increased migrant populations.

3.2 Assumptions

1. **The impact of Federal Housing Assistance will be negligible.** The federal government currently has similar voucher programs available to assist those in danger of homelessness. However, these voucher funds would be insufficient to fully tackle the problem for an entire city. Furthermore, these federal programs suffer from extremely long waitlists [26], making them an unrealistic option for many in urgent need. Depending on a person's demographics, these federal funds may also be more difficult to access. In this scenario, it is reasonable to expect that any stimulus from the federal government is negligible, and any impact will solely be a function of the local government's actions.

3.3 Variables

Variable	**Definition**	**Unit**
X	Amount of Rent Subsidy (annual)	Dollars
P_X	Predicted Percent Decrease of Homeless People when X subsidy is applied	N/A

3.4 Our Model

A critical problem driving the homelessness crisis is the un-affordability of purchasing a home. A simple way for local governments to make housing more affordable would be to provide subsidies that decrease the cost of renting or buying a home. Particularly in our model, because housing is generally the highest section of spending, it is most reasonable to target housing prices to alleviate financial pressure on those at risk. Although similar government programs have shown success, such as the Section 8 Housing Choice Voucher Program [25] on the national scale, this success is negligible compared to the number of people we could lift out of poverty with the subsidies our model reflects.

As we determined housing costs to be the largest economic burden on an individual in part 2 and housing subsidies have been effective in the past, it would be well advised for governments to invest more in housing subsidy programs. Using the model developed in Part 2, we simulated the implementation of "housing vouchers" by decreasing the yearly cost of housing by \$1000, \$2500, \$5000, and \$10000 in Seattle, Washington.

3.5 Results

The tables below represent the potential change in the homeless population due to the application of a rent voucher after 10, 20, and 50 years. The data is represented as a percent change in comparison to our results from Model 2, which would represent the scenario where a local government does not intervene to resolve the homelessness crisis.

Subsidy Value(X)	**10 Years**	**20 Years**	**50 Years**
\$10,000	26.0%	50.4%	44.8%
\$5,000	13.7%	30.1%	29.7%
\$2,500	6.71%	15.2%	17.0%
\$1,000	3.20%	5.56%	6.43%

Figure 11: Percent Reduction(P_X) in Homeless Population in Seattle

3.6 Discussion and Analysis

Our results indicate that government subsidies of rent can have a significant influence on the homeless population in a relatively short time scale. The benefit of these subsidies appears to taper off after the 20-year mark, as we see a notably smaller increase from 20 to 50 years compared to 10 to 20 years. We believe this can be attributed to two factors. Firstly, a rent voucher of a set amount is only able to assist a certain number of people, as an individual whose deficit between income and necessary expenses is greater than the provided subsidy. Additionally, Monte Carlo simulations represent a distribution of all possible cases, and have an inherent level of variability. Thus, in a real-world scenario we would be less likely to expect the efficacy of a voucher to decrease, as we see in the \$5,000 and \$10,000 cases.

From an efficiency standpoint, subsidies of value $\$2,500$ appear to be the most worthwhile over a longer period of time, as its percentage decrease per dollar is maximized. However, subsidies of larger values are still fairly successful, and should not be counted out while determining a plan to tackle homelessness.

The incorporation of a job-loss factor makes our model easily adaptable for a variety of unexpected circumstances. For example, a natural disaster could wreck business offices or factories, forcing a company to lay off many workers at least temporarily. Similarly, an economic downturn would also increase the rate of layoffs, as companies attempt to cut costs. Thus, both factors can be incorporated into our model by increasing the probability of job loss, which would simulate the economic impact of these events. While harder to directly account for, an uptick in migrant populations can also be accounted for by observing indirect consequences. For example, temporary housing for migrants could overlap with potential affordable housing for low-income people, momentarily driving up rent prices through increased demand. Further, when evaluating the impact of government-subsidized rent vouchers, voucher amounts may be decreased due to increased local government spending on humanitarian aid for migrants.

3.6.1 Strengths and Weaknesses

Our model can be easily adjusted to specifically target any city, given simple data on the economic status of its population that is at-risk of homelessness, making it widely applicable. Another strong point is its flexibility and facile adaptability for unforeseen events, such as natural disasters, economic downturn, or increased migration.

A larger weakness of government subsidies could be seen through landlords adjusting to this government intervention and subsequently increasing rents. However, this issue could be solved by relevant legislation, such as restrictions on unjustified rent hikes.

In our model, we addressed homelessness from an economic perspective, which does not address more personal issues, like drug use and mental disorders, that may affect the relative urgency of government action. However, we do believe that our model would be of critical concern in helping local governments predict the long-term impact of any potential plans to reduce homelessness.

4 Conclusion and Future Directions

Understanding the root causes of homelessness and developing solutions to them will prove extremely valuable in our modern world, as we work to tackle this growing humanitarian problem. When evaluating these causes, the limited time frame of our report limited us to economic considerations regarding homelessness, which do not account for the influence of additional externalities such as addiction or mental illness.

Our first model analyzes the changes in the supply of available or vacant housing within two regions of the US - Albuquerque, New Mexico and Seattle, Washington. Our model accurately accounts for the seasonality of the housing market, providing a more in-depth view at the trends driving this market.

Our second model predicts long-term changes in the homeless population within our chosen regions. Using a Monte Carlo simulation, we compare a simulated individual's income with the cost of their bare necessities, giving us a good economic evaluation of the causes of homelessness.

Our final model evaluates the potential long-term impacts of rent vouchers for decreasing homeless populations within our chosen model community of Seattle. As the results from our previous model indicated that the heavy cost of housing was a significant financial burden on those at-risk of homelessness, we sought to quantify the impact of an expansive rent voucher on the homeless population. This also aligns with current strategies for reducing homelessness outlined by the federal government, most notably the Section 8 Housing Choice Voucher Program.

Taken together, our model demonstrates the urgency of today's intertwined housing and homeless crises, but also highlights the far-reaching benefits that government intervention could impart.

The limited time constraints presented in this competition did not allow us to fully explore the plethora of possibilities presented by our models. The rate of housing construction likely depends on longer-term economic outlook. While difficult to predict for such long time periods, this would potentially provide our first model with even more realistic estimates. Additionally, the use of more advanced ARIMA models, such as the distributed ARIMA (DARIMA) presented by Wang et. al. [19] could improve the long-term predictability of our model. For part 2, we hope to consider the consequences of substance abuse on basic needs, since addiction drives drug and alcohol use to compulsive tendencies. We can also study its influence on the probability of job loss, as frequent substance use can impair cognition and emotional stability [29], leading to worsened performance in the workplace. In part 3, we can further evaluate the impact of other types of incentives on homelessness, including food subsidies and tax credits. Additionally, it may be beneficial to consider incentivizing the development sector, not just the buyers.

5 References

[1] https://www.npr.org/2022/07/14/1109345201

[2] https://listwithclever.com/real-estate-blog/6-things-to-know-about-houses-not-listed-on-zillow/

[3] https://www.cabq.gov/housing-forward-abq/about-housing-forward-abq

[4] https://www.seattle.gov/housing/levy

[5] https://www.zillow.com/research/data/

[6] https://www.nar.realtor/blogs/economists-outlook/seasonality-in-the-housing-market

[7] https://www.sciencedirect.com/science/article/pii/S0264275121001293

[8] https://www.gao.gov/homelessness

[9] https://endhomelessness.org/homelessness-in-america/who-experiences-homelessness/chronically-homeless/

[10] https://mecklenburghousingdata.org/progress/length-of-time-homeless/

[11] https://www.madisontrust.com/information-center/visualizations/average-retirement-age/

[12] https://www.valuepenguin.com/health-insurance-age-26

[13] https://www.bls.gov/cex/tables/calendar-year/mean-item-share-average-standard-error/cu-income-quintiles-before-taxes-2022.pdf

[14] https://www.payscale.com/cost-of-living-calculator

[15] https://data.census.gov/table/ACSST1Y2022.S1901?q=income%20seattle

[16] https://www.healthsystemtracker.org/chart-collection/health-expenditures-vary-across-population/

[17] https://www.rentcafe.com/cost-of-living-calculator/

[18] https://www.uscis.gov/working-in-the-united-states

[19] https://arxiv.org/pdf/2007.09577.pdf

[20] https://soarworks.samhsa.gov/article/definitions-of-homelessness

[21] https://unitedwaynca.org/blog/vacant-homes-vs-homelessness-by-city/

[22] https://www.zippia.com/advice/layoff-statistics/

[23] https://www.ncbi.nlm.nih.gov/books/NBK218240/

[24] https://fortune.com/2023/03/16/how-long-does-unemployment-last-quitting-great-resignation-3-months-should-i-quit-job/

[25] https://endhomelessness.org/homelessness-in-america/what-causes-homelessness/housing/

[26] https://www.hud.gov/topics/housing_choice_voucher_program_section_8#hcv01

[27] https://m3challenge.siam.org/kdfrldh/

[28] https://www.forbes.com/sites/robinryan/2022/10/18/how-to-quickly-rebound-from-a-layoff/?sh=5c83f9296eab

[29] https://www.mayoclinic.org/diseases-conditions/drug-addiction/symptoms-causes/syc-20365112

6 Appendix

Code

6.1 Model 1: SARIMA

```
import pandas as pd
import matplotlib.pyplot as plt
from pandas.plotting import autocorrelation_plot
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.graphics.tsaplots import plot_pacf
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.statespace import sarimax
from statsmodels.graphics.tsaplots import plot_predict

df = pd.read_excel("../papython/mcmprac/Darima.xlsx", names= ["New
    Listings"], header=0)

result = adfuller(df.values)
print(result[1])
result = adfuller(df.diff().dropna().values)
print(result[1])
result = adfuller(df.diff().diff().dropna().values)
print(result[1])
d = 1

autocorrelation_plot(df.diff().dropna())
plot_pacf(df.diff().dropna())
plt.show() #both cases show p = 1 with the highest autocorrelation
p = 1


plot_acf(df.diff().dropna())
plt.show() #q = 3
q = 3

indices = [i for i in range(70,601)]

from pylab import rcParams
import statsmodels.api as sm
rcParams['figure.figsize'] = 20, 10
decomposition = sm.tsa.seasonal_decompose(df, model='additive', period=
    35) # additive seasonal index
fig = decomposition.plot()
#plt.show()

decomposition = sm.tsa.seasonal_decompose(df, model='multiplicative',
    period= 35) # multiplicative seasonal index
fig = decomposition.plot()
#plt.show()

print(df.head())
df['New Listings'].plot(figsize=(20, 5))
plt.grid()
plt.legend(loc='best')
```

```
plt.title('Listings')

from statsmodels.tsa.stattools import adfuller
adf_test = adfuller(df['New Listings'])

print('ADF Statistic: %f' % adf_test[0])
print('Critical Values @ 0.05: %.2f' % adf_test[4]['5%'])
print('p-value: %f' % adf_test[1])

from statsmodels.tsa.stattools import kpss
kpss_test = kpss(df['New Listings'])

print('KPSS Statistic: %f' % kpss_test[0])
print('Critical Values @ 0.05: %.2f' % kpss_test[3]['5%'])
print('p-value: %f' % kpss_test[1])


from scipy.stats import boxcox
data_boxcox = pd.Series(boxcox(df['New Listings'], lmbda=0), index = df.
    index)

df['New Listings'].plot(figsize=(20, 5))
plt.grid()
plt.plot(data_boxcox, label='After Box Cox tranformation')
plt.legend(loc='best')
plt.title('After Box Cox transform')
#plt.show()

data_boxcox_diff = pd.Series(data_boxcox - data_boxcox.shift(), df.index)
plt.figure(figsize=(20,5))
plt.grid()
plt.legend(loc='best')
plt.title('Time series data')
plt.xlabel("Time in months")
plt.ylabel("Sales in Seattle")
#plt.show()


from statsmodels.tsa.statespace.sarimax import SARIMAX
import numpy as np

model = SARIMAX(data_boxcox, order=(1, 0, 1), seasonal_order=(4, 1, 1, 12)
    )
model_fit = model.fit()
print(model_fit.summary())

y_hat_sarima = model_fit.predict(70, 670)
y_hat_sarima = np.exp(y_hat_sarima)

forecast = pd.DataFrame(model_fit.predict(start=70, end=670), index=
    indices)
forecast.columns = ['Predicted']
plt.plot(y_hat_sarima, label= 'Predicted')
plt.plot(df.values, label='Current')
plt.show()
forecast2 = model_fit.get_forecast(600)
yhat_conf_int = forecast2.conf_int(alpha=0.05)
```

```
yhat_conf_int = np.exp(yhat_conf_int)
yhat_conf_int.to_excel("../papython/mcmprac/Errors.xlsx")

for x in range(len(df.values)):
    print(df.values[x,0])
for x in range(len(y_hat_sarima)):
    print(y_hat_sarima.iloc[x])
```

6.2 Model 2: Monte Carlo

```python
from numpy.random import normal, uniform


def job_loss(income):
    # Mean and SE calcaulated from https://www.forbes.com/sites/robinryan/2022/10/18/how-to-quickly-rebound-from-a-layoff/?sh=695a98846eab
    prob = uniform()
    if prob < 0.01: # probability of being unemployed in a given year
        weeks = normal(7.5, 5.35714) # distribution of unemployment times
        if weeks > 12: # unemployment benefits end after 12 weeks
            return 12 / 52.0 * income
        return weeks / 52.0 * income
    return income


# Taken from https://www.fns.usda.gov/cnpp/thrifty-food-plan-2021 and the payscale calculations
# for seattle and albuquerque
income = {'s_mean': 167027, 's_std': 40000, 'a_mean': 86268, 'a_std': 20000}
food = {'s_mean': 3976.56, 's_std': 156.82, 'a_mean': 3117.63, 'a_std': 122.96}
healthcare = {'s_mean': 2725.94, 's_std': 144.38, 'a_mean': 1908.16, 'a_std': 101.06}
transport = {'s_mean': 3788.4, 's_std': 225.46, 'a_mean': 2864.4, 'a_std': 170.47}
housing = {'s_mean': 17639.6, 's_std': 326.52, 'a_mean': 7356, 'a_std': 136.8}
apparel = {'s_mean': 678, 's_std': 145, 'a_mean': 565, 'a_std': 120.83}


def simulated_indiv_each_year(years):
    money = normal(income.get('a_mean'), income.get('a_std'))
    total = 0
    # data = [homeless?, year homeless]
    data = [False, 0]
    for i in range(years):
        food_c = normal(food.get('a_mean'), food.get('a_std'))
        health_c = normal(healthcare.get('s_mean'), healthcare.get('s_std'))
        transport_c = normal(transport.get('s_mean'), transport.get('s_std'))
        housing_c = normal(housing.get('s_mean') - 10000, housing.get('s_std'))
        apparel_c = normal(apparel.get('s_mean'), apparel.get('s_std'))
        #print(f'food: {food_c}; health: {health_c}; transport: {transport_c}; housing: {housing_c}')
        total = job_loss(money) - food_c - health_c - transport_c - housing_c - apparel_c + total * 0.67
        #print(f'total: {total}')
        if total <= 0:
            data[0] = True
            data[1] = i + 1
    return data

```

```python


homeless = 0 # homeless population
# 10 YEAR SIMULATION
alb_ages_pop = [34316, 35441, 33191, 33191, 36566, 39942, 42754, 42192, 38254,
                37129, 36004, 30941, 29253, 30000, 30000, 30000, 30000, 30000]
for i in range(2):
    for j in range(alb_ages_pop[i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
    for j in range(alb_ages_pop[10 - i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
for i in range(7):
    for j in range(alb_ages_pop[i + 11]):
        temp = simulated_indiv_each_year(10)
        if temp[0]:
            homeless += 1
homeless = int((0.22 * homeless) + (homeless * 0.78 * 2 / 3))
print(f'alb homeless: {homeless}')

sea_ages_pop = [34527, 38200, 43342, 46280, 51423, 66115, 87418, 94764, 59503,
                32323, 27915, 27915, 32323, 30000, 30000, 30000, 30000, 30000]
for i in range(2):
    for j in range(sea_ages_pop[i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
    for j in range(sea_ages_pop[10 - i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
for i in range(7):
    for j in range(sea_ages_pop[i + 11]):
        temp = simulated_indiv_each_year(10)
        if temp[0]:
            homeless += 1
homeless = int((0.22 * homeless) + (homeless * 0.78 * 2 / 3))
print(f'sea homeless: {homeless}')



# 20 YEAR SIMULATION
alb_ages_pop = [34316, 35441, 33191, 33191, 36566, 39942, 42754, 42192, 38254,
                37129, 36004, 30941, 29253, 30000, 30000, 30000, 30000, 30000]
for i in range(4):
    for j in range(alb_ages_pop[i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
```

```
            homeless += 1
    for j in range(alb_ages_pop[10 - i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
for i in range(3):
    for j in range(alb_ages_pop[i + 11]):
        temp = simulated_indiv_each_year(20)
        if temp[0]:
            homeless += 1
homeless = int((0.22 * homeless) + (homeless * 0.78 * 2 / 3))
print(f'alb homeless: {homeless}')

sea_ages_pop = [34527, 38200, 43342, 46280, 51423, 66115, 87418, 94764, 59503,
                32323, 27915, 27915, 32323, 30000, 30000, 30000, 30000, 30000]
for i in range(4):
    for j in range(sea_ages_pop[i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
    for j in range(sea_ages_pop[10 - i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
for i in range(3):
    for j in range(sea_ages_pop[i + 11]):
        temp = simulated_indiv_each_year(20)
        if temp[0]:
            homeless += 1
homeless = int((0.22 * homeless) + (homeless * 0.78 * 2 / 3))
print(f'sea homeless: {homeless}')



# 50 YEAR SIMULATION
alb_ages_pop = [34316, 35441, 33191, 33191, 36566, 39942, 42754, 42192, 38254,
                37129, 36004, 30941, 29253, 30000, 30000, 30000, 30000, 30000]
for i in range(9):
    for j in range(alb_ages_pop[i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
    for j in range(alb_ages_pop[17 - i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
homeless = int((0.22 * homeless) + (homeless * 0.78 * 2 / 3))
print(f'alb homeless: {homeless}')

sea_ages_pop = [34527, 38200, 43342, 46280, 51423, 66115, 87418, 94764, 59503,
                32323, 27915, 27915, 32323, 30000, 30000, 30000, 30000, 30000]
```

```python
for i in range(9):
    for j in range(sea_ages_pop[i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
    for j in range(sea_ages_pop[17 - i]):
        temp = simulated_indiv_each_year(5 * (i + 1))
        if temp[0]:
            homeless += 1
homeless = int((0.22 * homeless) + (homeless * 0.78 * 2 / 3))
print(f'sea homeless: {homeless}')
```