MathWorks Math Modeling Challenge 2024

Livingston High School

Team # 17400, Livingston, New Jersey

Coach: Cheryl Coursen

Students: Harsh Akunuri, Olivia Mei, Timothy Wu, Nathan Zhang, Kevin Zhang

M3 Challenge THIRD PLACE—$10,000 Team Prize

JUDGE COMMENTS

Specifically for Team #17400—Submitted at the close of triage judging

COMMENT 1: Very well-written summary and a nice job of making a compelling case for the conclusions of your work while still making clear they are projections. Good presentation and explanation of your first model. Your discussion of strengths/ weaknesses of your model choice was appreciated. You might have tested for correlation of your factors rather than assuming they are correlated. You might also have commented on the R values in your regressions to see if other choices might have been better for the data shape. Nice idea for your second model and you again explained it well. Good job connecting your final model to your previous work. Some of the values you used, such as 15% of people recovering, would have been better if you had cited external sources for them or made it clear that you were simply making a guess. Particularly for the second case you might have also considered sensitivity of your model to these values. Impressive work overall!

COMMENT 2: Great executive summary. For Q1, this could have been stronger with more justification for the logistic model (log may have worked here also). Q2 is well done, good consideration of potential variables. Q3 could have used more details about other potential causes for homelessness and other ways to home them, other than only treat/care for them. Good work overall!

COMMENT 3: You've answered these questions with lots of details. Well done.

COMMENT 4: The paper stands out for its innovative use of sophisticated mathematical modeling techniques, particularly random forest regression models for housing predictions and a state machine approach for homelessness, to tackle the pressing issues of housing shortage and homelessness in Albuquerque and Seattle. The team's effort to integrate critical variables such as population, unemployment rate, and median income showcases a commendable attempt to capture the complexities of urban planning and provides detailed projections that are invaluable for policymakers. It is particularly impressive how the models not only offer strong predictive capabilities but also maintain a high level of methodological transparency, making these findings both reliable and reproducible. While the paper already makes reasonable contribution, it has the potential to offer even richer insights by exploring nonlinear dynamics and broadening its data sources.

COMMENT 5: Your paper was neat, organized and very well written, and the mathematical models were clearly explained.

M3 Challenge 2024:

A Tale of Two Crises: *The Housing Shortage and Homelessness*

TEAM #17400

March 1st, 2024

Executive Summary

To the Secretary of the U.S. Department of Housing and Urban Development,

As the United States rapidly evolves into a hub of technology and progress, its urban centers are expanding to accommodate the increasing population. As of 2023, Seattle boasted a population exceeding 3.5 million, with thousands of new homes constructed each year to meet the demand.[1][2] Predicting the trajectory of housing in urban regions is essential for the ongoing development of cities.

By implementing two forest regression models, we predicted the number of housing units in Seattle, Washington, and Albuquerque, New Mexico, at intervals of 10, 20, and 50 years. By factoring in population size, unemployment rate, and median income, we effectively trained our models. Our predictions indicate that Albuquerque is projected to reach 287,075; 312,593; and 382,767 housing units, while Seattle is expected to reach 423,980; 512,389; and 653,023 housing units in 10, 20, and 50 years, respectively.

With a state machine approach and considering four significant factors—homelessness, drug use, serious mental illness, domestic violence, and poverty—we projected the future of homelessness in Seattle, and Albuquerque. Our model incorporates the potential for change in these factors, such as rehabilitation from drug use, overcoming poverty, recovery from mental illness, among others. In Albuquerque, we estimate the homeless population to be 8,530 in 10 years, 9.735 in 20 years, and 13,214 in 50 years. For Seattle, our predictions indicate 13,082 homeless individuals in 10 years, 12,094 in 20 years, and 7,918 in 50 years.

Instead of providing temporary aid, comprehensive reforms must be implemented through government policies and initiatives aimed at alleviating the root causes of homelessness. By focusing on key factors such as mental health and substance abuse, we can significantly reduce the number of people experiencing homelessness. With our aforementioned state machine approach, we projected that with government initiatives to combat drug abuse the homeless population in Albuquerque could decrease to 9,621 individuals within 50 years. [3] Similarly, by implementing government-funded mental health treatments over a prolonged period, we anticipate a significant reduction in homelessness. Our projections suggest that such initiatives could decrease the homeless population in Albuquerque to 9,283 individuals over the same time frame. However, it's essential to acknowledge that homelessness can be influenced by factors beyond individual choices, such as natural disasters. Taking into account the impact of natural disasters, we estimate that without intervention, the homeless population in 50 years could reach 13,947 individuals. Nevertheless, by implementing mental health and drug prevention policies, we project that the homeless population could be reduced to 11,922 and 12,403 individuals, respectively, within the same time frame.

We believe our predictions will promote policy makers to implement similar initiatives in hopes of reducing the homeless population. By providing these people with another opportunity for life, we can improve these ecosystems and promote development within our beloved cities.

Table of Contents

Q1. It Was the Best of Times

1.1 Defining the Problem

We are tasked to create a model that will predict the housing supply in Albuquerque, New Mexico and Seattle, Washington. We are then asked to utilize this model to examine and foresee the changes in housing supply in these two cities in the next 10, 20, and 50 years.

1.2 Local Assumptions

1-1. There are no new technological advancements or government policies that significantly impact any of our variables.
New technological advancements and new government policies are unpredictable and thus necessary to assume they won't occur in order to simplify our models.

1-2. Our independent variables can be expressed as a function in terms of time.
In order to predict the values of our features 10, 20, and 50 years into the future, we performed regression on the features with respect to time, allowing us to input those values at 10, 20, and 50 years into the future in our model to predict the number of housing units in those years.

1-3. The number of housing units can be used as a proxy for the housing supply.
We interpreted housing supply as the number of available housing units since a supply would be a physical stock of available housing in each city.

1-4. There is access to sufficient natural resources to construct new housing units and maintain existing ones.
In order to keep our model simpler, we assume that each city has the ability to construct and maintain housing units. In other words, a shortage of resources will not be accounted for when predicting the growth of the number of housing units.

1.3 Variables

Symbol	Definition	Unit
P_a	**Albuquerque population**: Number of people living in Albuquerque	people
P_s	**Seattle population**: Number of people living in Seattle	people

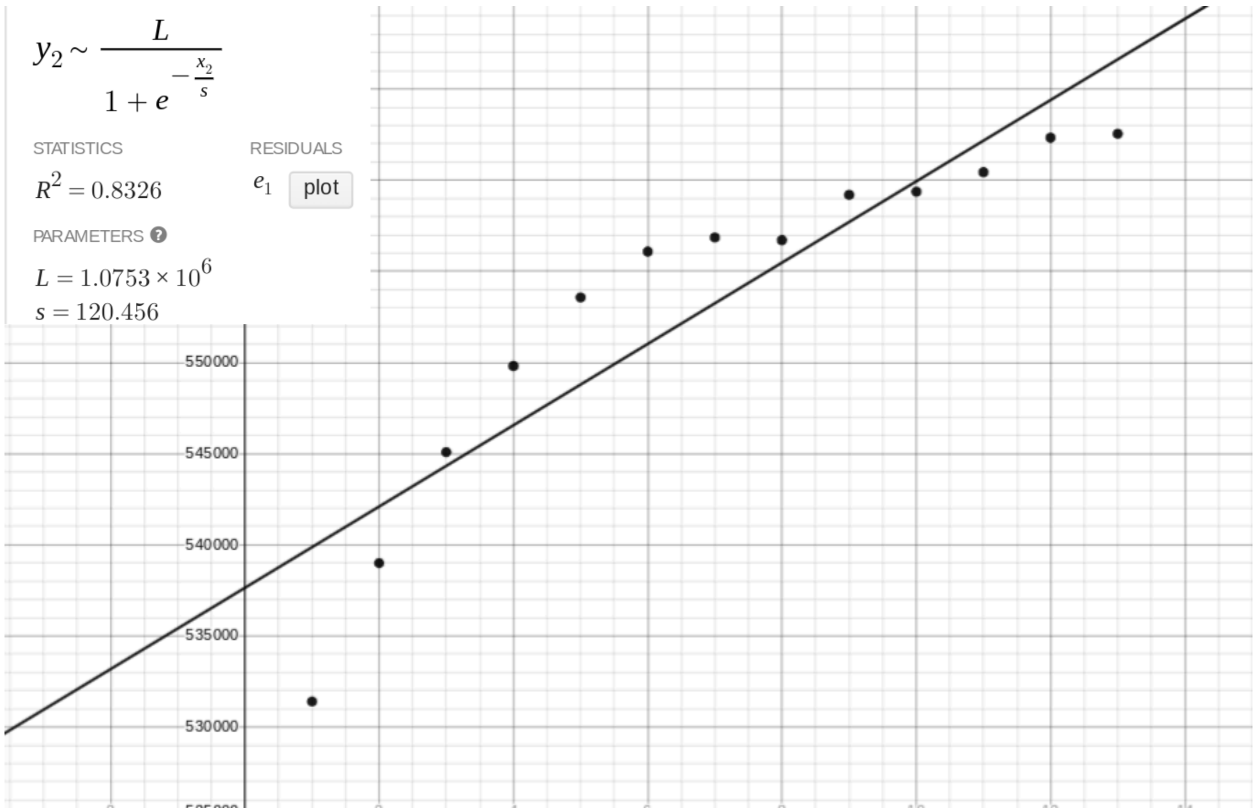
U_a	**Albuquerque unemployment rate**: Unemployment rate in Albuquerque	%
U_s	**Seattle unemployment rate:** Unemployment rate in Seattle	%
I_a	**Albuquerque median income**: Median income of people living in Albuquerque	$ (US Dollars)
I_s	**Seattle median income**: Median income of people living in Seattle	$ (US Dollars)
H_a	**Albuquerque housing units**: Number of housing units in Albuquerque	housing units
H_s	**Seattle housing units**: Number of housing units in Seattle	housing units

1.4 The Model

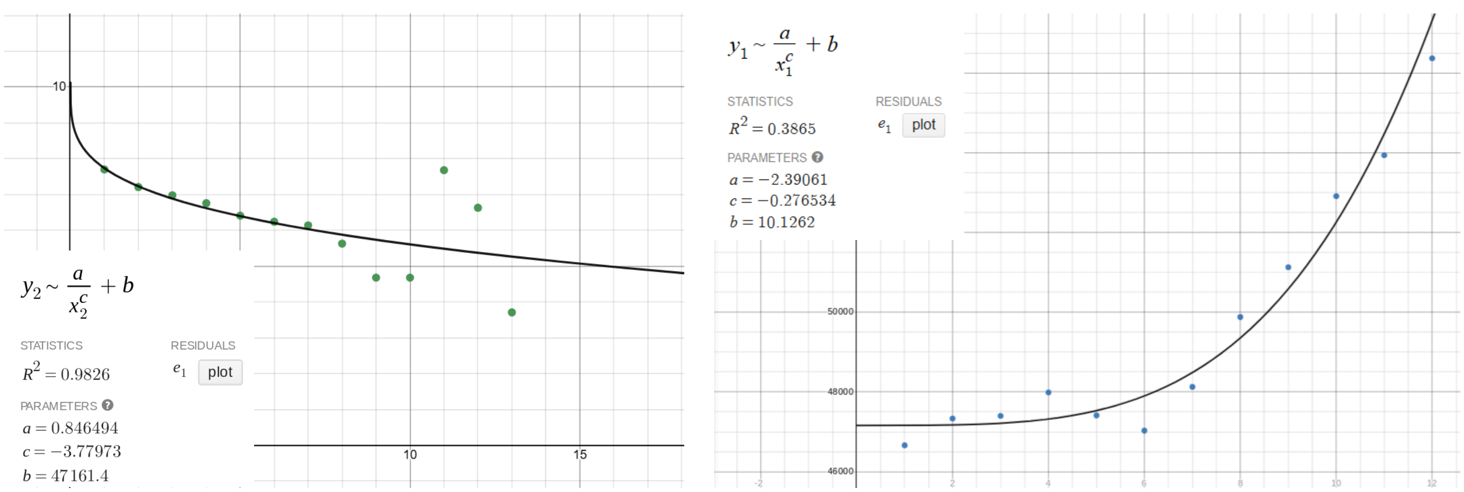
1.4.1 Factor Identification

The population (P) of each city is relevant to the number of housing units in these cities. With a higher population, the demand for housing will naturally increase, and play a role, and at least correlate with the number of housing units. Next, the unemployment rate (U) is relevant to the number of housing units, and one possible reason could be since investors will be unwilling to invest in a location where more people are unemployed, and thus unable to afford housing. Finally, the median income (I) of the working force is also correlated with the number of housing units due to a similar reasoning as why the unemployment rate would be correlated.

For each of these variables, we got historical data from the US census [8] we used either a polynomial or logistic regression to predict values for these variables (P, U, I) for the next 50 years. A logistic function was used for population related variables such as P because it considers the principle of exponential growth, but slows down as it reaches the carrying capacity.



Logistic Regression model of Albuquerque population



Polynomial regression models of Albuquerque unemployment rate and Albuquerque median income

1.4.2 Random Forest Regression

To predict housing unit supply decades into the future, we used the Scikit-learn Python library to train 2 random forest regression models on the features listed above (P, U, I), one model for Albuquerque, NM and one model for Seattle, WA. This supervised machine learning model utilizes an ensemble of decision trees to train alongside making an accurate and robust forecast for the housing supply 10, 20, and 50 years into the future. The random forest samples a random subset of the dataset and the features used to fit into each individual decision tree. This process, known as feature subspace selection, ensures variability between trees, thus reducing intercorrelation and allowing the model to determine optimal splits. With this, the decision trees are able to categorize inputs for each feature and produce a quantitative prediction for housing units. The final prediction is an average of all decision tree outputs. Our random forest model has 10 decision trees, which each are limited to a max depth of 5 nodes.

	housingUnits	population	unemploymentRate	medianIncome
0	234891.0	531403.0	7.70	46662.0
1	237735.0	539000.0	7.21	47333.0
2	239718.0	545083.0	6.98	47399.0
3	240277.0	549812.0	6.76	47989.0

First rows of training data for Albuquerque

	population	unemploymentRate	medianIncome
0	-2.392102	1.293183	-0.911149
1	-1.560875	0.873968	-0.755518
2	-0.895302	0.677194	-0.740209
3	-0.377878	0.488975	-0.603365

Normalized training data for Albuquerque

Both datasets were normalized using Scikit-learn's StandardScaler which finds the z-score for each data value, preventing data with larger scales from dominating the learning process. A 70/30 train test split was utilized to reserve a portion of the data to validate the housing unit predictions made by each model. The full code for data processing, model training, and model evaluations for both cities are shown in the appendix.

1.5 Results

The random forest regressor obtained a root mean squared error of 2,376 for Albuquerque and 3,123 for Seattle on its test data predictions of total housing units, which represents the average difference between predicted and actual values. These are reasonable values given the current degree of housing supply in each city.

Ultimately, the models predicted sizable increases in housing supply for each time period in both cities, as shown in the table below. These predictions are consistent with the expected results and can be attributed to population increases and economic growth. Due to the significant difference in years between training data and the final data used to produce results, confidence in each prediction decreases as years in the future increase. This is a result of potential overfitting of the regression model and unavoidable extrapolation.

City	Years in the Future	Total Housing Units
Alburquerque	10	287,075
	20	312,593
	50	382,767
Seattle	10	423,980
	20	512,389
	50	653,021

1.6 Strengths and Weaknesses

The random forest model allows us to make predictions by effectively combining numerous factors. The model doesn't require linear data, and is robust to skewed or non-linear data, which makes it an ideal choice for predicting based on our data. Additionally, random forest models are generally unaffected by multicollinearity [31], which is important in this case since our variables (such as unemployment rate and median income) are likely correlated with each other. Furthermore, the random forest model minimizes the chance of overfitting, which is an issue in many machine learning models.

However, the random first model is not easily interpretable due to its "black box" [8] nature, in which it employs many decision trees. In other words, the algorithm undergoes a complex decision process, making it difficult to pinpoint how exactly the model is built when it is training.

Q2. It Was the Worst of Times

2.1 Defining the Problem

The second question asks to predict the rise in homeless populations for increments of 10, 20, and 50 years in the cities Seattle and Albuquerque. By analyzing four factors we were able to determine the rate of homelessness.

2.2 Local Assumptions

2-1. Drug addicts and people with serious mental illnesses will not be able to purchase a home, but unafflicted homeless people with jobs may be able to purchase affordable housing.
The systemic nature of homelessness is compounded by drug addiction and psychiatric developments. [11] These factors also make it difficult to gain employment. The stigma surrounding drug addicts and mentally unwell makes it nearly impossible to find a job that can afford housing, and affordable housing is lacking in the first place. [35] Additionally, the initiatives and support systems made to benefit the homeless population are less effective for those with these afflictions. However, these established initiatives create the possibility of higher income employment opportunities, which may be taken advantage of by unafflicted homeless people. [15]

2-2. Everyone who did drugs before being homeless is homeless because of their drug use.
Alongside the expensive nature and psychological impacts of drug and alcohol misuse promotes the probability of homelessness. [34] Thus, drug use is a factor significant enough to cause homelessness.

2-3. People who are seriously mentally ill, but not drug addicts, that become homeless, are homeless because of mental illness.
Mental illness unveils psychological instability which affects their work performance. [35] Since drug addiction isn't a factor, it's reasonable to conclude the large impacts of serious mental illness will eventually spiral into homelessness.

2-4. There are few drug-addicted and mentally ill victims who are able to escape domestic abuse situations that they can be neglected; only unafflicted escapees need to be considered.
Abusive relationships are correlated with controlling partners which limit the victim's access to resources including drugs. Additionally, the motivation to escape abusive environments reveals resiliency reducing the likelihood of mental illness. [32]

2-5. Sheltered homeless people do not do drugs and have help in managing any mental illnesses.
Transitional shelters have the goal of providing homeless people with assistance to resist return to the streets. Thus, they will work to rehabilitate and provide aid to drug addicts and psychologically unstable. Shelters also have systems to screen residents for probationary products like drugs [29].

2-6. The distribution of drug addicted, mentally ill, and unafflicted migrant homeless people is the same as the distribution of Alburquerque or Seattle natives.
Since there is a lack of such specific data, we applied general percentages depending on each state to its respective city.

2-7. Homeless people can only die from drug-related causes.
There are too many other variables to consider for causes of death. Thus, to maintain the four factor model we've decided to implement this assumption as overdosing is the leading cause of death for the homeless population.

2-8. Escapees of domestic violence immediately become homeless.
Victims of abuse have often been isolated from their other support systems [30]. Thus, it's reasonable to conclude that they wouldn't have access to any previous resources they might have had prior to the relationship.

2-9. Someone below the poverty line who loses their job will immediately become homeless.
If someone is already below the poverty line then they are most likely to be renting a home. After losing all income, this can result in difficulty finding another job because of the competitive market. Additionally, living below the poverty line lowers the probability of having any savings to afford rent. Thus, it's reasonable to assume that they will be unable to afford housing and will be evicted [33].

2-10. The proportion of drug-addicted homeless people, mentally ill homeless people, unafflicted homeless people, domestic abuse victims, and poor people is independent of the total population.
Drug addiction rates, mental illness rates, domestic violence rates, and poverty rates are roughly the same location-wise over periods of time. Additionally, affordable housing is built at a rate roughly correlating to the amount of population in an area. [25] Without government intervention, which this model assumes there is none of, there will be no significant changes in the proportion of the homeless states of people.

2-11. People become homeless only as a result of drug use, mental illness, domestic abuse, or poverty.

Since our model only implements four factors, we picked the four largest factors of homelessness and considered the others negligible. Additionally, many of the other factors have interlaced or can be considered as a result of these four larger factors

2.3 Variables

Symbol	Definition	Unit
T	**Time**: years in the future	Years
D	**Dead:** total number of homeless people who have died from drug abuse	People
HDA	**Homeless drug addicts**: Number of homeless drug addicts	People
HMI	**Homeless mentally ill**: Number of homeless mentally ill that are not drug addicted	People
HU	**Homeless unafflicted:** Number of unafflicted homeless people	People
DV	**Domestic abuse**: Number of domestic abuse victims	People
P	**Poor**: amount of people earning below 15k or below the poverty line in Albuquerque or Seattle, respectively	People

2.4 The Model

We assume that the proportion of each variable is independent of the general population of each city. In Albuquerque (ABQ) 2023, the total number of homeless people was 2400. [4] The drug addiction rate among homeless people is 33%, while the mental illness rate is 50%, and 50.2% of mentally ill homeless people are drug addicts, so overall 24.9% of ABQ homeless are non-addicted mentally ill people. [14] Using this, the values of HDA, HMI, and HU for ABQ were calculated to be 800, 448, and 1152, respectively. [6][7] Repeated familial abuse happens at about a rate of 0.827% of women, and ABQ has 287391 women, meaning that there are an estimated 2378 women facing serial domestic abuse in Alburquerque. Albuquerque has an average unemployment rate of approximately 3.71% [5], which corresponds to 4.8% of residents earning less than 15 thousand dollars per year according to the model from Problem 1, which amounts to 27005 people. We consider D to be 0 at the start of 2023, because D is not part of the total homeless population and does not really matter. These numbers give us a starting vector for D, HDA, HMI, HU, DV, and P.

We then create a Markovian transition matrix based on the probability of each state to go to the next: 1125 homeless people were admitted into a two-week emergency shelter, while 292 were given transitional housing [4]. The probability of an occupant of two-week emergency shelter being safe from drugs and mental illnesses was considered to be 1 in 26, while for an occupant of transitional housing it was considered to be 1. This gives a 0.14 probability for any kind of homeless person to in HC at the end of the year. [12] Additionally, 51% of drug-addicted homeless people die of drug overdose. The average duration for which someone is homeless is 40 months, so the probability of a drug-addicted homeless person dying within a year is 51%*12/40. If a drug-addicted homeless person is not dead or unafflicted by the end of the year, they are still drug-addicted. 50.2% of mentally ill homeless people in ABQ are drug addicts, and multiplying this by 12/40 gives a 15.06 chance of a person in HMI to become HDA in a year. 33.3% of homeless people in general are drug addicts,this gives the probability of an unafflicted homeless person becoming homeless as 0.1. 48.8% of homeless people who are mentally ill are not drug addicts, meaning there is a 7% chance of an unafflicted homeless person becoming mentally ill. [16] [14] Finally, there is a 12/40=0.3 chance for a homeless person to purchase a house, in which case they will move to P. Otherwise, a member of HC will remain in HC. 15% of homeless people are escapees of domestic violence, and they are assumed to be in HC. Finally, statistically someone in P in Albuquerque has a 10.3% chance of having a substance use disorder, which means they have a 0.447% chance of becoming part of HDA, while their chance of becoming HMI was similarly calculated to be 0.1434%. [14] Finally, the amount of homeless people that are homeless due to losing their finances is 12.7%, corresponding with a 3.8% chance of someone in P becoming HU.

Besides change in status of ABQ natives over the course of a year, there are two more large sources of new homeless people-migration and children reaching adulthood. 16% of homeless people in ABQ are migrants, so in any given year there will be an amount equal to 6% of the already-present population arriving in ABQ, in the same distribution of drug users, mentally ill, and unafflicted. [26] Additionally, the amount of homeless children is 18% of the amount of other homeless people, and homeless children have a 70% addiction rate, so the addition of children and migrants can be split among the three categories. [9] Finally, 200-250 affordable housing units are built in ABQ per year, and there are plenty of unafflicted homeless people who can afford this, so 225 is subtracted from HU and added to P. [17] The multiplication of the current vector with the transition matrix, along with the addition of appropriate values, creates a state function that, when iterated 10, 20, or 50 times, can model the amount of homeless people in each category in 10, 20, or 50 years.

Roughly the same approach, with different numerical values, was taken for Seattle. The most significant difference is that Seattle has a much higher proportion of sheltering homeless people--53% of Seattle homeless are in long-term sheltering. [18] Seattle also has much more accessible affordable housing, with the municipality building 1000 units per year and offering housing at a price of 30% of any income to all homeless people and people below the poverty line. [19] [20] [27] [28]

Because the state function model was constructed without considering general increases in population over time, each set of results can be multiplied by the ratio of the projected population from Problem 1 to the population in either city in 2023 in order to predict the amount of people in each category in 10, 20, and 50 years.

2.5 Results

Results are printed as follows: T, D, HDA, HMI, HC, DV, P, total homeless population HDA + HMI + HC.

10	20	50
2385.84	6068.32	18705.8
2224.58	2558	2919.41
872.935	962.674	1097.32
2657.88	2874.4	3269.07
2333.86	2992.71	3661.03
24989.8	24639.1	25997.5
final count 8089.26	final count 9387.78	final count 10946.8

Albuquerque homeless population over time

The projected populations of ABQ at 10, 20, and 50 years in the future are 593,245; 615,217; 679,293; which compared to the current population of 562,599 mean the actual projected homeless populations will be 8,530; 9,735; and 13,214 people in 10, 20, and 50 years, respectively.

10 1267.74 1565.57 614.314 4175.9 6726.37 65412.8 final count 13082.2	20 2351.19 1427.22 558.846 3754.44 6353.64 59839.2 final count 12094.1	50 4938.48 959.068 359.236 2212.05 4387.88 46772.9 final count 7918.24

Seattle homeless population

The projected populations of Seattle at 10, 20, and 50 years in the future are 885,572; 971,190; and 1,102,518; respectively. Compared to the current population of 733,919; this means that the actual projected homeless populations will be 15,785; 16,004; and 11,895 in those years.

2.6 Strengths and Weaknesses

This Markov chain-like state function is a stochastic model that allows us to predict the number of homeless people in the future. The model we built accounts for the mentally ill and drug addict portions of homeless people, which make up a big percentage of homeless people in both cities. We were also able to account for domestic abuse victims and those in poverty, which are two of the other largest causes of homelessness. Additionally, the state-change form of the model allows us to conveniently output the long-run proportions of mentally ill people, drug addicts, those in poverty, and homeless people. [36] Drug-related deaths, which are the bulk of deaths in homeless populations, are also shown. [37]

However, the largest limitation of this model is that it does not account for every possible factor of homelessness. It only accounts for four: drug addiction, mental illness, domestic violence, and poverty. Other factors that cause homelessness are not accounted for, and could play a role in determining the number of homeless people 10, 20, and 50 years into the future. Finally, population growth is not accounted for directly in the function, and is rather multiplied on manually after the code is outputted. If population growth were worked directly into the function, it would be reflected more accurately.

Q3. Rising from This Abyss

3.1 Defining the Problem

The final questions tasks us to create a robust model that could aid a city, chosen to be Albuquerque, NM, in creating a long-term plan addressing homelessness. We are also tasked to test the stability of this model against unpredictable factors such as natural disasters, economic recessions, and increased migration.

3.2 Local Assumptions

3-1. A natural disaster will cause an increase in the homeless population equal to 30% of the poor population.
A research paper in the University of Chicago Press Journals found that a severe natural disaster resulted in a 3-5% decrease in home ownership, with more salient natural disasters decreasing home ownership even more [23]. For the purposes of this paper, we will assume that a natural disaster's effect is 5%, the upper limit of this range. Since around 15% of the population is classified as poor, we will implement this change as the homeless unafflicted population gets an increase of 30% of the poor population.

3-2. The government can only help around 20% of the homeless population during any given year.
According to Pareto's Principle, 80% of the outcomes result from 20% of the causes [24]. This is a common application in business and economics, but it can be applied to any field. Because this is such a well-known saying, we will assume that government policy is built around this principle, so they will only help around 20% of the people that need help every year.

3-3. A government policy to reduce mental illness will help 15% of people with mental illness every year.
According to psychiatry.org, 75% of people are helped by psychotherapy. [22]. By multiplying this value with the 20% from assumption 3-2, we get that 15% of people will be helped every year.

3-4. A government policy to help those with substance abuse problems will help 8% of people with substance abuse problems.
Less than 42% of people that enter for rehab for substance abuse complete the rehab process [21]. By multiplying this value with the 20% from assumption 3-2, we get that 8% of people will be helped every year.

3.3 Variables

Symbol	Definition	Unit
D	**Dead:** number of homeless people who have died from drug abuse	People
HDA	**Homeless drug addicts**: Number of homeless drug addicts	People
HMI	**Homeless mentally ill**: Number of homeless mentally ill	People
HU	**Homeless unafflicted:** Number of unafflicted homeless people	People
DV	**Domestic abuse**: Number of domestic abuse victims	People
P	**Poor**: amount of people earning below 15k or below the poverty line in Albuquerque or Seattle, respectively	People

3.4 The Model

To create a robust model that could aid a city, we choose to implement the Markov Chain model from part 2, but change some values around to simulate different government policies. We decided to examine 2 government policies:

1. the government helps people with substance abuse problems overcome their addiction
2. The government helps people with mental health issues get treatment

From our assumptions, 15% of people with mental health issues will recover every year, while 8% of people with substance abuse problems every year. We will implement this by decreasing the stationary transition matrix by these values, while increasing the transition to the poor state by the same value. We will do this once for each policy.

We also decided to examine 1 unforeseen circumstances:

1. Natural Disasters

In order to test the model against a destructive natural disaster, we increase the homeless unafflicted population by a value equal to 30% of the poor population after 25 years. We will run each situation with natural disasters, and once without.

3.5 Results

Our results from each of our simulations are shown below. The first value (50), represents how many years into the future we are looking. The next 6 values represent the dead, homeless addicts, homeless mentally ill, homeless unafflicted, domestic abuse, and poor population in that order. The final count represents the total homeless population.

	No Natural Disaster	Natural Disaster
No Government Policy	50 18705.8 2919.41 1097.32 3269.07 3661.03 25997.5 final count 10946.8	50 22078.2 3690.69 1390.94 4162.21 4702.83 31351.3 final count 13946.7
Government Mental Health Policy	50 16851.9 2561.13 656.852 3121.97 2942.7 27235.5 final count 9282.65	50 20048 3261.52 838.8 4011.26 3809.97 33237 final count 11921.6
Government Substance Abuse Policy	50 14062.2 2183.05 1125.13 3317.78 2994.71 29254.4 final count 9620.67	50 16832.9 2788.5 1443.88 4275.01 3895.82 35880.2 final count 12403.2

The projected population of Albuquerque is 679,293, which compared to the current population of 562,599, means that we would need to multiply all of our final counts by 1.207.

	No Natural Disaster	Natural Disaster
No Government policy	13,213.8	16,833.7
Government Mental Health Policy	11,204.2	14,389.4
Government Substance Abuse Policy	11,612.1	14,970.7

Based on our results, we recommend the government implement a policy to help homeless people with mental health issues since that was most effective in combating homelessness. A natural disaster did not seem to impact our result much since a higher homeless count without the natural disaster still correlated with a higher homeless count with the natural disaster.

3.6 Strengths and Weaknesses

An advantage of our model is that it is easy to manipulate. A change in government policy only results in a change to 2 values in our transition matrix, while an unexpected event just results in the addition of one if statement for when the event occurs and the change to be implemented because of it.

However, our model is limited in scope because it does not account for all possible variables. For instance, there are other external variables that cannot be explained. Additionally, the model doesn't consider people who fall through the policy. It's impossible for everyone who participates in the established initiatives to make a full recovery. We can't account for the number of people who have tried to participate and are unable to recuperate.

4. Conclusion

We provided forecasts for the number of housing units in Albuquerque, NM and Seattle, WA in the next 10, 20, and 50 years. By training a random forest regression model on population, median income, and unemployment rate, we were able to apply this model to predicted population, median income, and unemployment rate in both cities with regression functions in Desmos to forecast the number of housing units in each city achieving the following results:

City	Year in Future	Total Housing Units
Alburquerque	10	287,075

	20	312,593
	50	382,767
Seattle	10	423,980
	20	512,389
	50	653,021

Next, we used a state-change model of homeless people consisting of those who are homeless due to drug addiction, poverty, mental illness, and domestic violence. We obtained the following results:

City	**Year in Future**	**Total Homeless People**
Alburquerque	10	8,530
	20	9,735
	50	13,214
Seattle	10	15,785
	20	16,004
	50	11,895

We see that the decrease in the number of homeless people 50 years in the future for Seattle is consistent with the drastic increase in housing units predicted in question 1. This most likely reflects the success of Washington's homelessness programs--as of 2020, 53% of their homeless population is sheltered, and the government provides affordable housing at 30% of any wage that a homeless or impoverished person earns. We can also see individual statistics for homeless drug addicts, homeless mentally ill people, homeless unafflicted people, domestic abuse victims, poor people, and homeless drug deaths.

Finally, using our model from part 2, we implemented various government policies by changing the transition state matrix. We also simulated the result of a natural disaster by increasing the homeless population halfway through the simulation. We obtained the following results for the homeless population after 50 years:

	No Natural Disaster	Natural Disaster
No Government policy	13,213.8	16,833.7
Government Mental Health Policy	11,204.2	14,389.4
Government Substance Abuse Policy	11,612.1	14,970.7

5. References

[1] *Seattle Metro Area Population 1950-2020. (n.d.). Www.macrotrends.net. https://www.macrotrends.net/cities/23140/seattle/population*

[2] *Pickford, O. (2024, January 16). Seattle's Housing Construction Booms While Permitting Flashes Warning Signs - The Urbanist. Www.theurbanist.org. https://www.theurbanist.org/2024/01/16/seattles-housing-construction-booms-while-permitting-flashes-warning-signs/*

[3] *Zerger, S. (2002). SUBSTANCE ABUSE TREATMENT: WHAT WORKS FOR HOMELESS PEOPLE? A REVIEW OF THE LITERATURE PREPARED FOR TRANSLATING RESEARCH INTO PRACTICE SUBCOMMITTEE NATIONAL HCH COUNCIL & HCH CLINICIANS NETWORK RESEARCH COMMITTEE. https://www.currytbcenter.ucsf.edu/sites/default/files/product_tools/homelessnessandtbtoolkit/docs/background/Factsheet/Article_2002_Substance%20Abuse%20Treatment%20Lit%20Review.pdf*

[4] *2023 Point-In-Time Count Albuquerque and Balance of State Report Table of Contents*.(n.d.). https://www.nmceh.org/_files/ugd/ad7ad8_6d9bf66e3a5d407eaad310cc44ecaf82.pdf

[5] *Albuquerque, NM Unemployment Rate*. (n.d.). Ycharts.com. https://ycharts.com/indicators/albuquerque_nm_unemployment_rate

[6] *"Bleak Picture" for Mentally Ill: 80% Are Jobless - NAMI Illinois*. (2014, July 31). Namiillinois.org. https://namiillinois.org/bleak-picture-for-mentally-ill-80-are-jobless/

[7] *Employing and Managing People with Substance Use Addictions*. (n.d.). Www.shrm.org. https://www.shrm.org/topics-tools/tools/toolkits/employing-managing-people-substance-use-addictions

[8] Simon, S., Glaum, P., & Valdovinos, F. S. (2023). Interpreting random forest analysis of ecological models to move from prediction to explanation. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-30313-8

[9] Gomez, R., Thompson, S. J., & Barczyk, A. N. (2010). Factors Associated with Substance Use Among Homeless Young Adults. *Substance Abuse*, *31*(1), 24–34. https://doi.org/10.1080/08897070903442566

[10] Hahn, J. A., Kushel, M. B., Bangsberg, D. R., Riley, E., & Moss, A. R. (2006). Brief report: The aging of the homeless population: Fourteen-year trends in San Francisco. *Journal of General Internal Medicine*, *21*(7), 775–778. https://doi.org/10.1111/j.1525-1497.2006.00493.x

[11] ReferencesRobertson, I. E., Sagvaag, H., Selseng, L. B., & Nesvaag, S. (2021). The hunt for a job: narrating the process of gaining employment for people in recovery from lives dominated by drug use. Drugs: Education, Prevention and Policy, 1–10. https://doi.org/10.1080/09687637.2021.1973962

[12] Mosel, S. (2022, March 14). *Substance Abuse & Homelessness: Statistics & Rehab Treatment* (A. Sharp, Ed.). American Addiction Centers. https://americanaddictioncenters.org/rehab-guide/homeless

[13] ReferencesMcChesney, K. Y. (1990). Family Homelessness: A Systemic Problem. Journal of Social Issues, 46(4), 191–205. https://doi.org/10.1111/j.1540-4560.1990.tb01806.x

[14] National Institute on Drug Abuse. (2018, August 15). *Comorbidity: Substance Use and Other Mental Disorders*. National Institute on Drug Abuse. https://nida.nih.gov/research-topics/comorbidity/comorbidity-substance-use-other-mental-disorders-infographic

[15] ReferencesFosburg, L. B., & Dennis, D. L. (1999, August 1). Practical Lessons: The 1998 National Symposium on Homelessness Research (Arlington, Virginia, October 29-30, 1998). ERIC. https://eric.ed.gov/?id=ED443892

[16] Saldua, M. (2023). *Addressing Social Determinants of Health Among Individuals Experiencing Homelessness*. Samhsa.gov. https://www.samhsa.gov/blog/addressing-social-determinants-health-among-individuals-experiencing-homelessness

[17] *City Outlines Strategies to Increase Affordable Housing*. (n.d.). City of Albuquerque. Retrieved March 2, 2024, from https://www.cabq.gov/health-housing-homelessness/news/city-outlines-strategies-to-increase-affordable-housing-1

[18] *Seattle/King County Point-in-Time Count of Individuals Experiencing Homelessness 2020 Seattle/King County Point-in-Time Count of Individuals Experiencing Homelessness*. (2020). https://kcrha.org/wp-content/uploads/2020/06/Count-Us-In-2020-Final.pdf

[19] *Find Affordable Housing - Housing | seattle.gov*. (2018). Seattle.gov. https://www.seattle.gov/housing/renters/find-housing

[20] Bruce, M., Councilmember, H., Mosqueda, T., & Winkler-Chin, M. (2023). *2022 Annual Housing Investments Report Seattle Office of Housing*.

https://www.seattle.gov/documents/Departments/Housing/Reports/2022_AnnualInvestments_Final.pdf

[21] American Addiction Centers. (2023, July 12). *Drug Rehab Success Rates and Statistics* (W. Manwarren Generes, Ed.). American Addiction Centers. https://americanaddictioncenters.org/rehab-guide/success-rates-and-statistics

[22] American Psychiatric Association. (2019). *What is Psychotherapy?* Psychiatry.org; American Psychiatric Association. https://www.psychiatry.org/patients-families/psychotherapy

[23] Sheldon, T. L., & Zhan, C. (2019). The Impact of Natural Disasters on US Home Ownership. *Journal of the Association of Environmental and Resource Economists*, *6*(6), 1169–1203. https://doi.org/10.1086/705398

[24] Tardi, C. (2022, July 7). *How to Apply the 80-20 Rule*. Investopedia. https://www.investopedia.com/terms/1/80-20-rule.asp

[25] Mulder, C. (n.d.). *Sixty-ninth session of the UNECE Committee on Housing and Land Management Key note presentation The relationship between population and housing*. https://unece.org/fileadmin/DAM/hlm/archive/Key%20note%20population%20and%20housing.pdf

[26] Cardinale, J. (2023, October 6). *2,394 people experiencing homelessness in Albuquerque; 83% increase from 2022*. KOAT. https://www.koat.com/article/homeless-increase-in-albuquerque/45462181

[27] *How many homeless people in Seattle are from here?* (2018, October 26). The Seattle Times.

https://www.seattletimes.com/seattle-news/homeless/do-homeless-people-come-to-seattle-for-help/

[28] Brief, M. (n.d.). *Substance Use and Mental Disorders in the*. https://www.samhsa.gov/data/sites/default/files/NSDUHMetroBriefReports/NSDUHMetroBriefReports/NSDUH-Metro-Seattle.pdf

[29] Bardwell, G., Boyd, J., Kerr, T., & McNeil, R. (2018). Negotiating space & drug use in emergency shelters with peer witness injection programs within the context of an overdose crisis: A qualitative study. *Health & Place*, *53*, 86–93. https://doi.org/10.1016/j.healthplace.2018.07.011

[30] *Domestic Violence and Homelessness: Statistics (2016)*. (n.d.). Www.acf.hhs.gov. https://www.acf.hhs.gov/ofvps/fact-sheet/domestic-violence-and-homelessness-statistics-2016

[31] Liyanage, L. (2019, August 22). *Are Random Forests affected by multi-collinearity between features?* Research Gate. https://www.researchgate.net/post/Are-Random-Forests-affected-by-multi-collinearity-between-features

[32] Loxton, D., Schofield, M., & Hussain, R. (2006). Psychological Health in Midlife Among Women Who Have Ever Lived With a Violent Partner or Spouse. *Journal of Interpersonal Violence*, *21*(8), 1092–1107. https://doi.org/10.1177/0886260506290290

[33] *Poverty & Homelessness in the United States of America FacsNet*. (n.d.). FacsNet. https://facsnet.org/resources/poverty-homelessness-in-the-united-states/

[34] Thompson, R. G., Wall, M. M., Greenstein, E., Grant, B. F., & Hasin, D. S. (2013). Substance-Use Disorders and Poverty as Prospective Predictors of First-Time

Homelessness in the United States. *American Journal of Public Health*, *103*(S2), S282–S288. https://doi.org/10.2105/ajph.2013.301302

[35] Hennekam, S., Richard, S., & Grima, F. (2020). Coping with mental health conditions at work and its impact on self-perceived job performance. Emerald. https://www.emerald.com/insight/content/doi/10.1108/ER-05-2019-0211/full/html

[36] Arlington Life Shelter. (2023). *Arlington Life Shelter : How We Help : Resources : Causes of Homelessness*. Arlingtonlifeshelter.org. https://arlingtonlifeshelter.org/how-we-help/resources/causes-of-homelessness.html

[37] Leonard, C. (2023, November 27). *First-of-its-kind homelessness study reveals alarming statistic for California's unhoused*. San Francisco Chronicle. https://www.sfchronicle.com/california/article/homeless-death-18507431.php

6. Code Appendix

6.1 It Was the Best of Times

```
# Import relevant classes
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn import metrics
from sklearn.preprocessing import StandardScaler

# Import dataset
from google.colab import drive
drive.mount('/content/drive')

df = pd.read_csv('/content/drive/MyDrive/housingData.csv')
df.drop(['employed'], axis=1, inplace=True)

# Normalize dataset
scaler = StandardScaler()
scaler.fit(df.drop('housingUnits', axis=1))

scaled_values = scaler.transform(df.drop('housingUnits', axis=1))
df_feat = pd.DataFrame(scaled_values, columns=df.columns[1:])

# Split data into train and test
X = df_feat
y = df['housingUnits']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=101)

# Fit Random Forest Regression model to train data
rfc = RandomForestClassifier(n_estimators=10, max_depth=5)
rfc.fit(X_train, y_train)

# Use test data to make predictions
rfc_predict = rfc.predict(X_test)

# Measure accuracy using RMSE
metrics.mean_squared_error(y_test, rfc_predict, squared=False)
```

```python
# Import forecasted values for features in 10, 20, and 50 years
test = pd.read_csv('/content/drive/MyDrive/housingTestABQ.csv')
test.drop('employed', axis=1, inplace=True)

# Normalize forecasted dataset
scalerTest = StandardScaler()
scalerTest.fit(test)

scaled_valuesTest = scalerTest.transform(test)
df_test = pd.DataFrame(scaled_valuesTest, columns=test.columns)

# Use trained random forest regression model to make predictions on total housing units
predictions = rfc.predict(test)
```

6.2 It Was the Worst of Times

```cpp
#include <iostream>
#include <cstdio>
using namespace std;

int main() {
  double abq [6]= {0,800,448,1152,2378,27005}; //starting vector
  double transition [6][6]= {{1,0,0,0,0,0},{0.153,0.707,0,0,0.14,0},{0,0.1506,0.7094,0,0.14,0},{0,0.1,0.07,0.53,0,0.3},{0,0,0,0.15,0.85,0},
{0,0.00447,0.001434,0.038,0,0.956096}}; //transition  matrix
  double homeless [6]; //vector to represent state distribution
  for (int i=0;i<6;i++){
    homeless[i]=abq[i];
  }
  double homelessCount = 0;
  for(int i = 0; i<6; i++){
    homelessCount += homeless[i];
  }
  double placeholder[6];
  for(int i = 0; i<6; i++){
    cout << homeless[i] << "\n";
  }
```

```cpp
  for(int T = 0; T< 50; T++){
    for(int k = 0; k < 6; k++){
      placeholder[k] = 0; //clearing placeholder vector
    }
    for(int c = 0; c<6; c++){
      for(int r = 0; r<6; r++){
        placeholder[c]+=transition[r][c]*homeless[r]; //multiplication of transition matrix and homeless vector
      }
    }
      }
    }
    for(int i = 0; i<6; i++){
      homeless[i] = placeholder[i];
    }
    homeless[3] += 0.072* homeless[3] -225; //accounting for nonaddicted children, migrants, and homeless who gain hiouses
    homeless[5] += 225; //homeless who move into new affordabe houses
    homeless[1] += 0.096* homeless[1]; //acounting for addicted children and migrants
    homeless[2] += 0.054* homeless[2]; //accouunting for mentally ill migrants
    homelessCount = 0;
    for(int i = 1; i<5; i++){
      homelessCount += homeless[i]; //calculating # of homeless
    }
    cout << "\n" << T+1 << "\n";
    for(int i = 0; i<6; i++){
      cout << homeless[i] << "\n"; //printing results
    }
    cout << "final count "<< homelessCount << "\n";
  }
```

Code for stimulation of state function modeling Albuquerque’s homeless population

6.3 Rising from This Abyss

```cpp
#include <iostream>
#include <cstdio>
using namespace std;
bool mentaliIllnessPolicy=1;
bool drugAbusePolicy=0;
bool natDisaster=1;
//dead
//homeless addicts
//homeless mentally ill
//homeless unafflicted
//domestic abuse
//poor
int main() {
  double abq [6]= {0,800,448,1152,2378,27005};
  double transition [6][6]= {
  {1,0,0,0,0,0},
  {0.153,0.707,0,0,0.14,0},
  {0,0.1506,0.7094,0,0.14,0},
  {0,0.1,0.07,0.53,0,0.3},
  {0,0,0,0.15,0.85,0},
  {0,0.00447,0.001434,0.038,0,0.956096}};
  if (drugAbusePolicy){
    transition[1][1] = 0.627;
```

```
        transition[1][5] = 0.08;
    }
    if (mentaliIllnessPolicy){
        transition[2][2]=0.5594;
        transition[2][5]=0.15;
    }
    double homeless [6];
    for (int i=0;i<6;i++){
        homeless[i]=abq[i];
    }
    double homelessCount = 0;
    for(int i = 0; i<6; i++){
        homelessCount += homeless[i];
    }
    double placeholder[6];
    for(int i = 0; i<6; i++){
        cout << homeless[i] << "\n";
    }
    for(int T = 0; T< 50; T++){
        for(int k = 0; k < 6; k++){
            placeholder[k] = 0;
        }
        for(int c = 0; c<6; c++){
```

```
            for(int r = 0; r<6; r++){
                placeholder[c]+=transition[r][c]*homeless[r];
            }
        }
        for(int i = 0; i<6; i++){
            homeless[i] = placeholder[i];
        }
        homeless[3] += 0.072* homeless[3] -225;
        homeless[5] += 225;
        homeless[1] += 0.096* homeless[1];
        homeless[2] += 0.054* homeless[2];
        homelessCount = 0;
        for(int i = 1; i<5; i++){
            homelessCount += homeless[i];
        }
        cout << "\n" << T+1 << "\n";
        for(int i = 0; i<6; i++){
            cout << homeless[i] << "\n";
        }
        //nat disaster
        if (T==25&&natDisaster){
            homeless[3]+=0.3*homeless[5];
        }
        cout << "final count "<< homelessCount << "\n";
    }


}
```