Q1: Predicting Housing Supply

1.1 Defining the problem
In this problem, we are tasked with developing a model that predicts change in the housing supply in Seattle, Washington and Albuquerque, New Mexico in the next 10, 20, and 50 years.

1.2 Assumptions
1-1 *Housing supply is directly correlated to demand.*

Justification: This serves as the foundational premise for the examination of Q1, and is an assumption commonly accepted in economics and urban planning, as housing markets typically respond to changes in demand through adjustments in supply.

1-2 *All housing is built by private companies.*

Justification: Public housing accounts for less than 1% of housing in the US, a negligible amount when considering the larger trend. Therefore, housing companies seek profit and the general market follows a supply-and-demand relationship.

1-3 No significant, unexpected situations that would drastically alter the parameters.

Justification: We are unable to predict these situations reliably.

Assumptions in the evaluation of specific factors are outlined as they are introduced.

1.3 Population Projection Model
Housing supply is directly correlated to demand. As such, we begin our examination by attempting to model the demand for housing in Albuquerque, New Mexico, and Seattle, Washington, by projecting the changes in population size in these regions for the next 10, 20 and 50 years.

We identified three key components to the population change of a city: fertility rates, mortality rate (estimated with life expectancy), and immigration.
We model the population change in the recursive relation:

$$P_{t+1} = P_t + B_t - D_t + I_t$$

Where P_{t+1} represents the city's population at year t+1, P_t is the city's population at year t, and B_t, D_t and I_t are the city's births, deaths, and net immigration respectively during year t. The components are calculated as follows:

1.3.1 Number of Births (B_t)
B_t is calculated by multiplying the number of women of reproductive age with the annualized TFR (Total Fertility Rate, average number of children a woman would have over her reproductive lifespan (15-49 years)) for a city:

$B_t = \ Women\ of\ Reproductive\ Age_t \times \ Annualized\ TFR_t$

The number of women of reproductive age is given by a sum of women aged 15-49, which is assumed to be half of total population:

$$Women\ of\ Reproductive\ Age_t = \sum_{age=15}^{49} (Age\ Distribution_{age,t} \cdot P_t) \cdot 0.5\ .$$

The recursive structure of our population calculation also allows for a step-wise adjustment of population structure in the two cities. While the data provided us with the age distribution in 2022, our model adjusts for the change each year:

$Age\ Distribution_{<5,\ t+1}$	$\frac{B_t}{P_t} \times 100$
$Age\ Distribution_{age,\ t+1}$	$Age\ Distribution_{previous\ age,\ t+1} \cdot (1 - Mortality\ Fraction_{previous\ age,t})$
$Age\ Distribution_{85>,\ t+1}$	$Age\ Distribution_{85>,\ t+1} \cdot (1 - Mortality\ Fraction_{85>,t})$

The annualized TFR is given by $\frac{Predicted\ Fertility_{,t} \cdot Fertility\ Adjustment\ Factor}{35}$, dividing the total fertility by the approximate number of reproductive years (age 15-49). $Predicted\ Fertility_t$ is calculated based on a linear regression of the historical fertility rates of Albuquerque and Seattle, respectively. The projected fertility rates are then applied to an age-distributed model of the cities' specific populations to forecast future births. The value is adjusted by a $Fertility\ Adjustment\ Facto$r, important in the adaptability of our model to external factors and will be further discussed in Q3.

Finally, incorporating these into the formula for B_t, we have

$$B_t = \left(\sum_{age=15}^{49} (Age\ Distribution_{age,t} \cdot P_t) \cdot 0.5\right) \left(\frac{Predicted\ Fertility_{,t} \cdot Fertility\ Adjustment\ Factor}{35}\right)$$

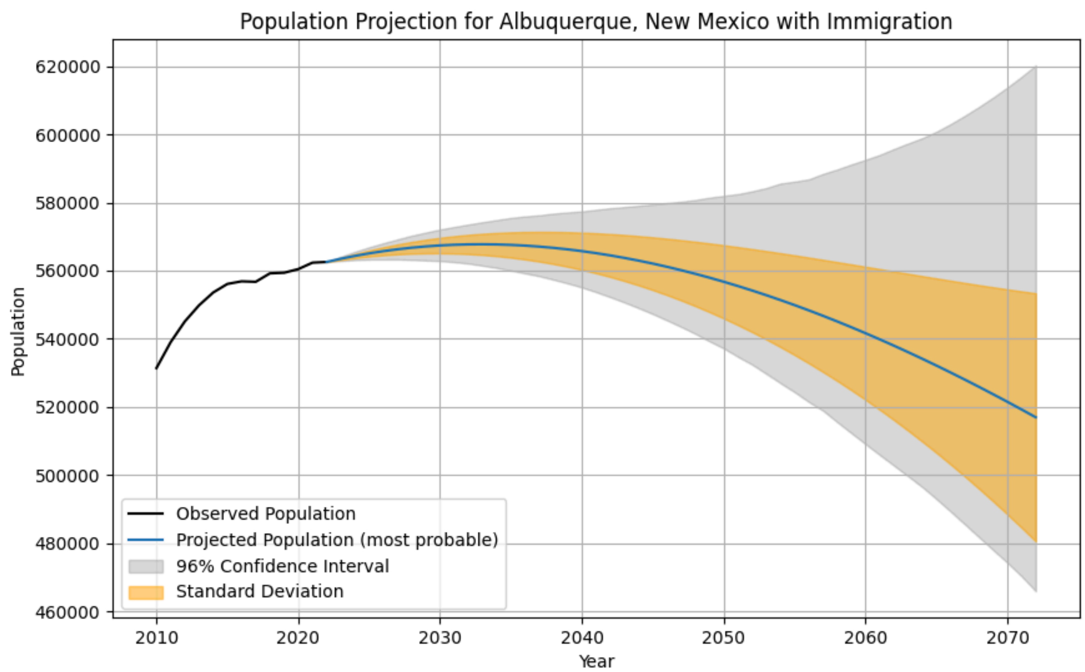
1.3.2 Number of Deaths (D_t)

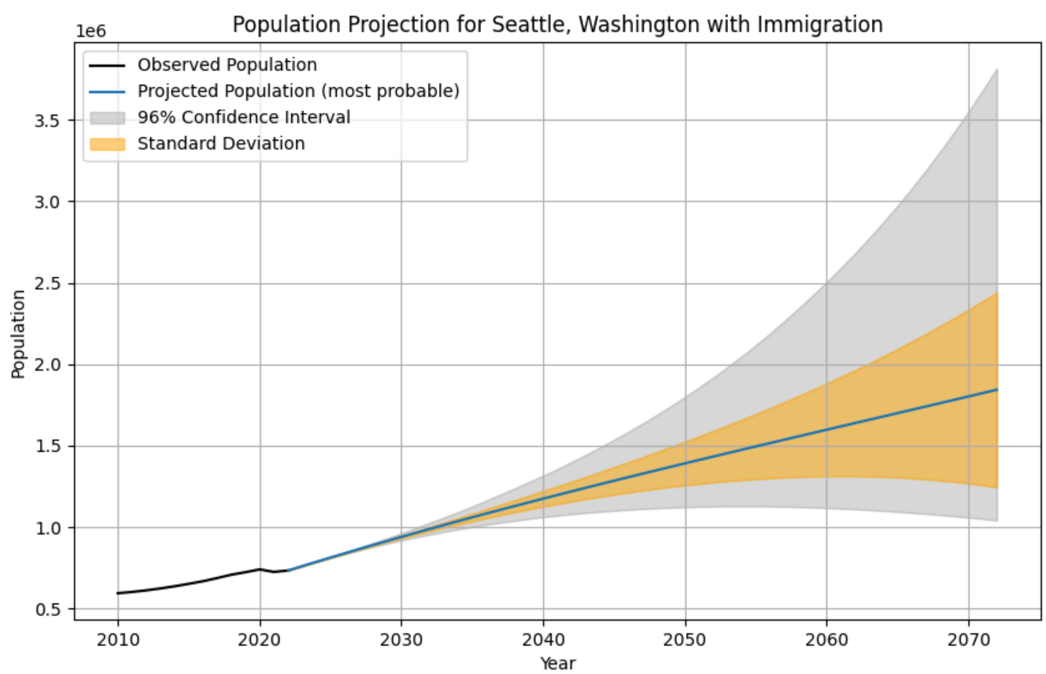
In our research, we found the global increase in life expectancy to be roughly linear. Assuming a similar slope in the two cities, the projections $Life\ Expectancy_t$ for each year are calculated by adding the cumulative increase in life expectancy to the 2022 value – 75.5 years and 79.73 years for Albuquerque and Seattle respectively. Then, D_t is calculated by the formula $D_t = \frac{P_t}{Life\ Expectancy_t}$.

1.3.3 Net Immigration (I_t)

Immigration rates are known to be highly volatile and unpredictable, historically unstable, and dependent on a wide variety of factors that could induce extreme future changes in either direction. Many previous studies assumed constant yearly immigrant populations. For the next 50 years, we used a simple exponential regression model based on data from 2000-2022.

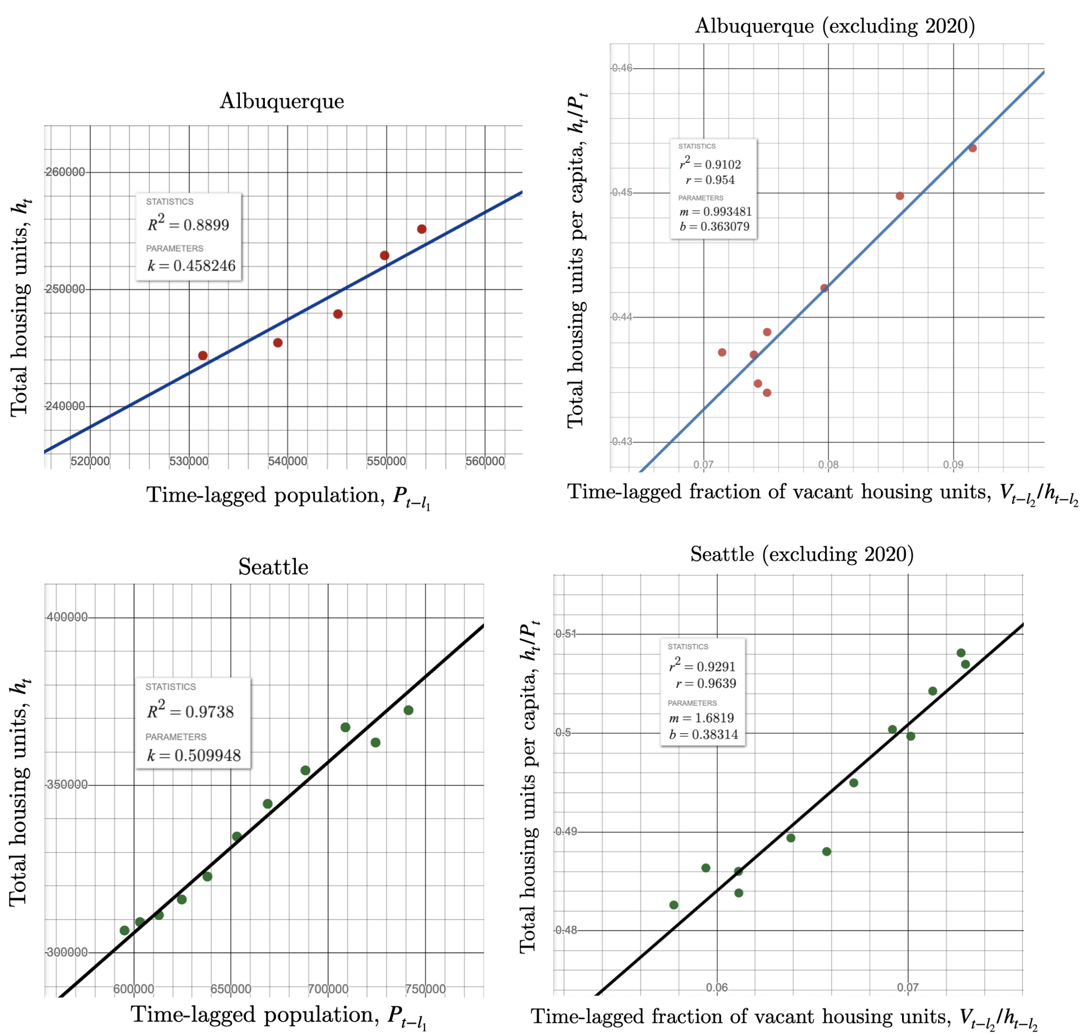
Combining all factors, our projection for the population of Albuquerque and Seattle are mapped in the following figures:





To derive confidence intervals for our predictions of demand, we made assumptions about the normal distribution of variance in two key parameters: fertility rate and the exponential increase factor of immigration. While subjective, our choice of standard deviation reflects a preference for conservative estimates to ensure a stable future outlook. We opted for a standard deviation of 0.05 (5%) for the fertility rate, indicating our expectation of minimal societal or policy shifts in the United States. Though fertility rates may fluctuate, we anticipate no extreme deviations absent significant interventions or demographic changes. Similarly, the standard deviation of 0.01 for the exponent of immigration increase, centered around a mean of 1.019 and 1.013, suggests a moderate prediction of change without accounting for extreme events impacting immigration patterns.

Using a Monte Carlo simulation, we generated 1000 projection curves sampling from the normal distributions and subsequently calculated the standard deviation of our population projection for each year. This enables us to delineate confidence intervals as highlighted in orange in Fig 1.



We then model the total number of housing units as being directly proportional to the population, accounting for a fit-determined time lag, which also seems to agree with existing data,[5] as shown above. For instance, following an increase in population, a time lag will occur – corresponding to the time taken for developers to recognize growing demand, obtain construction permits, and the process of new housing units being developed – before the housing supply increases and adjusts to the new housing demand.

Similarly, we model the total number of housing units per capita as a linear function of the fraction of vacant housing units (excluding the vacant housing fraction from 2020 due to the effects of the COVID-19 pandemic on housing occupancy rates) with a fit-determined time lag. Once again, it takes time for housing developers to adjust to new vacancy rates and adjust their speed of housing development.

For each plot, we performed an optimization analysis over a range of time lags to determine the time lag that best fit existing data and assumed this time lag (representing the "reaction speed" of the housing market suppliers) would remain constant (time taken for obtaining construction permits and building housing units is approximately constant). While these assumptions of constant parameters are indeed weaknesses and potential oversimplifications of our model, within the input datasets we were provided, they seem justified by the goodness-of-fits we observe (see the above R^2 and r^2 values).

From our optimization calculation code (see Appendix), we obtain the time lags given below:

For Albuquerque, we obtained time lags of l_1=6 and l_2=5 years.
For Seattle, we obtained time lags of l_1=2 and l_2=0 years.

From the above fits, we derive linear equations and rearrange for the number of vacant houses, V, in terms of the fit constants k, b, and m, as well as the previously calculated population P. Subscripts indicate the time in years.

$$\frac{h_t}{P_t} = m\frac{V_{t-l_2}}{h_{t-l_2}} + b$$

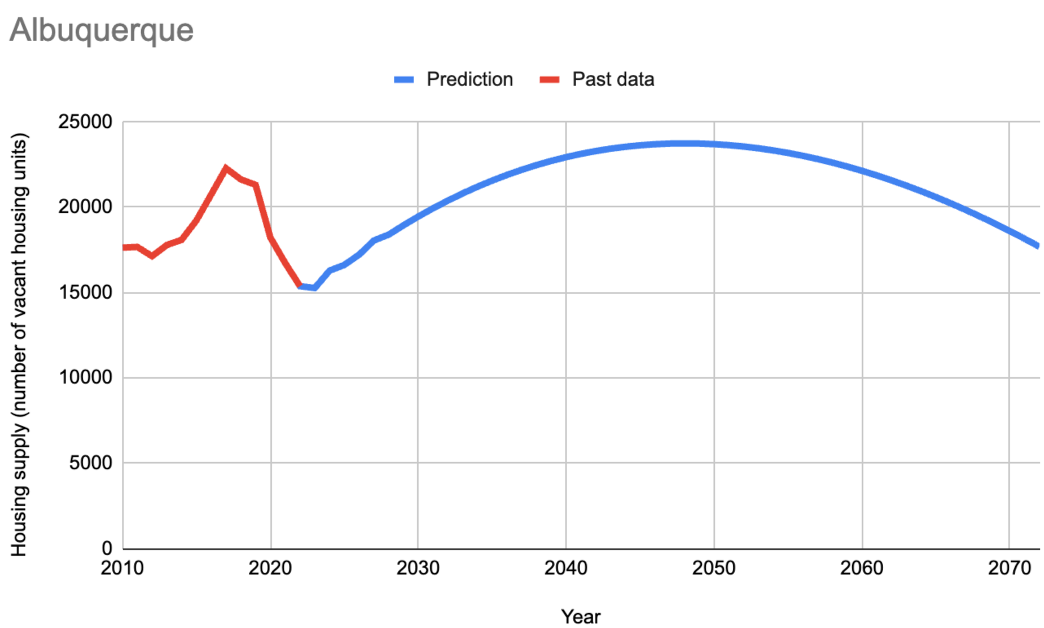
$$h_t = kP_{t-l_1}$$

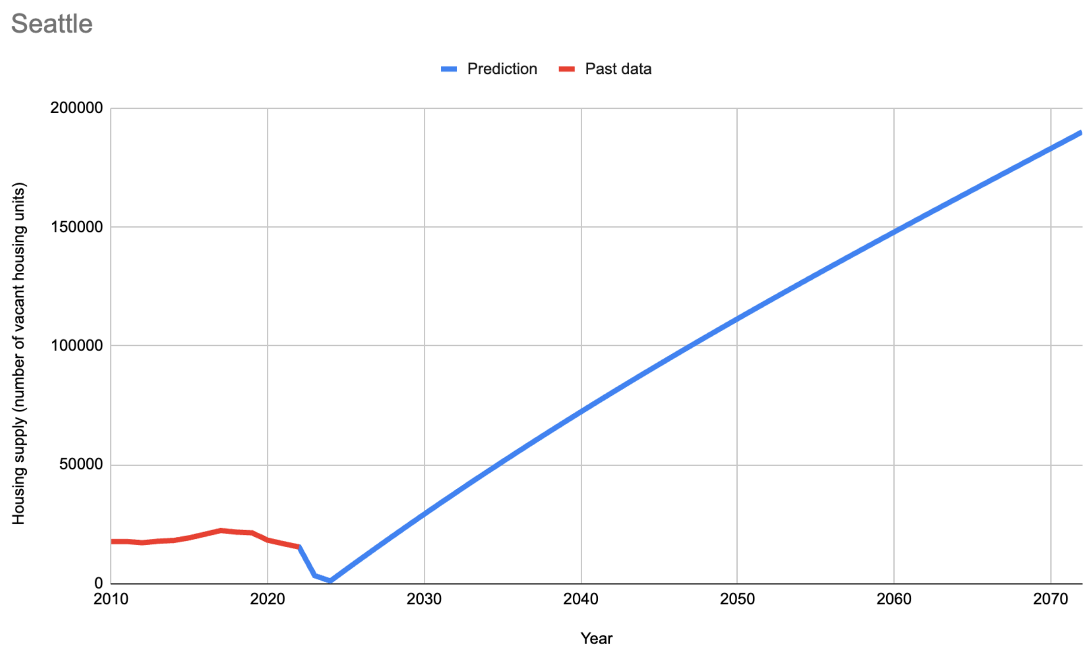
$$h_{t-l_2} = kP_{t-l_1-l_2}$$

$$V_{t-l_2} = \frac{kP_{t-l_1-l_2}}{m}\left(\frac{kP_{t-l_1}}{P_t} - b\right)$$

$$V_t - V_{t-1} = \frac{kP_{t-l_1}}{m}\left(\frac{kP_{t-l_1+l_2}}{P_{t+l_2}} - b\right) - \frac{kP_{t-l_1-1}}{m}\left(\frac{kP_{t-l_1+l_2-1}}{P_{t+l_2-1}} - b\right)$$

Applying this recursive relation to the given data for the number of vacant housing units, we obtain the following plots for Albuquerque and Seattle:





Using the partial derivative method of computing an error propagation formula, we further calculated standard deviations, and in turn the corresponding 95% confidence intervals for our 10, 20, and 50-year predictions.

For Albuquerque, they are: (2032, 0 to 114390), (2042, 0 to 116292), and (2072, 0 to 97684).

For Seattle, they are: (2032, 0 to 167751), (2042, 0 to 241403), and (2072, 110103 to 350757).

Q2: Modeling Homeless Population

2.1 Defining the problem

We are tasked with modeling the homeless population in Seattle and Albuquerque in 10, 20, and 50 years, respectively. Here, we define homelessness in the same way as under the education subtitle of the McKinney-Vento Homeless Assistance Act– individuals who lack a ***fixed, regular, and adequate nighttime residence***. This specifically includes children and youth whose living situations fall into one of the categories below.

1. Temporarily Staying with Other People
2. Staying in Emergency or Transitional Shelters
3. Staying in Motels, Campgrounds, Cars, Parks, Abandoned Buildings, Bus or Train Stations, or any Public or Private Place not Designed for Humans to Live in