Climate and US Migration

by

Sanjana Chintalapudi  
 Masters in Data Science  
 University of Colorado, Boulder  
 Colorado, USA  
 Sanjana.Chintalapudi@colorado.edu

Sai Rohitha Challa  
 Masters in Data Science  
 University of Colorado, Boulder  
 Colorado, USA  
Sai.Challa@colorado.edu

Minal Pawar  
 Masters in Data Science  
 University of Colorado, Boulder  
 Colorado, USA  
 Minal.Pawar@colorado.edu

ABSTRACT

Climate change has emerged as one of the most formidable challenges facing humanity, with its impacts resonating across the world. Annual global temperature has increased by about 1.8°F (1.0°C) according to a linear trend from 1901 to 2016. 2023 was the hottest year on record. With the rise of global temperatures and the increasing frequency of climate disasters, there is a profound consequence of shifting migration and settlement patterns. Individuals and communities are grappling with the short- and long-term ramifications of environmental changes and in turn rethinking where and how they live. We are going to dive deeper into understanding how changing weather patterns and natural disasters in the United States are impacting climate migration. We are going to look at both climate disasters that could force people to migrate and the slower-shifting weather patterns that are changing where people may want to live. By analyzing the different dimensions of migration in the United States and comparing them to national climate patterns, we will shed light on the impacts of climate change on US migration. Our goal is to comprehend the subtle effects of climate change that go beyond the obvious by looking at how shifting weather patterns affect migration to states.

1. Introduction

1.1 Climate change

Climate change has emerged as one of the most formidable challenges facing humanity, with its impacts resonating across the world. Annual global temperature has increased by about 1.8°F (1.0°C) according to a linear trend from 1901 to 2016. [1] 2023 was the hottest year on record. [2] According to the National Academy of Sciences, the earth will see a greater increase in temperature in the next 50 years than compared to the whole last 6,000 years combined. Extremely hot zones like the Sahara could cover a fifth of the land surface by 2070 [3]. By the next century, around 3 to 6 billion people could be trapped in places facing extreme heat and food scarcity. South Asia, which houses one-fourth of the global population, will be most affected soon given the current rates of global warming.  In places like India and China, even a few hours outside would lead to death, with the greenhouse gas emissions unabated. [4] As the earth heats up, it has led to volatile weather patterns and more frequent climate disasters, impacting people everywhere significantly. From 2000 to 2019, 7,348 major recorded disaster events claimed 1.23 million lives, affecting 4.2 billion people, resulting in approximately US$2.97 trillion in global economic losses. [5] That is more than half the global population that has been affected. In 2022 alone, there were 33 million natural disaster-related displacements. [6]

1.2 Climate Migration

With the rise of global temperatures and the increasing frequency of climate disasters, there is a profound consequence of shifting migration and settlement patterns. Individuals and communities are grappling with the short- and long-term ramifications of environmental changes and in turn rethinking where and how they live. This type of migration is called climate migration. Climate migration occurs when people leave their homes due to climate disasters, such as floods, droughts, and wildfires, as well as slower-moving climate challenges such as rising seas and increasing water scarcity. [9] People, regardless of immediate threat, are increasingly proactive in reevaluating their habitats. Across the United States, nearly 1 in 2 people will experience a decline in the quality of their environment, namely more heat and less water. [15] The term "billion-dollar disasters" has been coined to name the natural disasters that cause over a billion dollars in losses. These billion-dollar disasters cause about $60.5 billion in losses every year. [11] 14.5 million homes were impacted by natural disasters in 2021. That is about 1 in 10 homes in the US. [12] Droughts in the West are changing agricultural landscapes, forcing people to move due to their livelihoods.

2. Data Collection and Cleaning

The data for migration within states was collected from the US Census [website,](https://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-to-state-migration.html) and the weather-related data was collected from the National Oceanic and Atmospheric Administration ([NOAA](https://www.ncei.noaa.gov/)) website. The collected migration data includes the estimated number of people who migrated within the states in United States (internal migration) from 2005 to 2022 except for data of the year 2020, and for weather data, parameters like average temperature (F), average snow (in), average precipitation (in), maximum temperature (F) and minimum temperature (F) were gathered for 372 stations spread over all the states of US, except for Delaware, every month from 2005 to 2022. The list of stations was obtained from the [Global Summary of the Month](https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-month?pageNum=1&startDate=2005-01-01T23:59:59&endDate=2022-12-31T00:00:00) dataset of NOAA. API [link](https://www.ncei.noaa.gov/access/services/data/v1)

2.1 Weather data:

After requesting the data from NOAA, using an API, the data was cleaned, reduced, and transformed into seasonal and yearly data. First, the data types of the obtained data are converted to help with further operations. All the stations didn’t have the parameters requested, resulting in a lot of blanks. To handle the NaNs and blanks, statistics like average, maximum, and minimum values for the data were calculated by grouping for states and years. For data exploration, the calculated monthly statistics were further calculated for summer and winter months, where April to October are considered summer months and the rest of the months are considered as winter months. The data frame was sorted for ease of use and to help with merging the data frames.

A table with numbers and letters

Description automatically generated

Figure 1: Initial weather data obtained

A screenshot of a graph

Description automatically generated

**Figure 2: Weather Data frame after data-cleaning**

2.2 Migration data:

For migration data, the preliminary data cleaning was done using MS Excel, as the table formats in the files changed over the years. The given data is a cross table for all the states that include the in-state population, the population that moved from one state to another, and the margin of error of census data. The datasets on the website were given per year, so combined 17 years of data files into one large data frame. Since the data was given in segments, we transformed the dataset and calculated variables that can be leveraged in our analysis, Total\_Pop, Percentage, Top1, Top1 State.

Total\_Pop: We calculated the total population for each state per year, by adding the population currently living in that state and summing the populations that moved from other states to that current state.

Percentage: The percentage was calculated by dividing the people that moved to the state (Total Population - In-state population) by the total population.

Top 1: We found the top state that people moved from into the current state and the population moved by using the max function.

Top1\_State: The top1 population variable identified the top state.

The Excel sheets with these parameters data were read into separate data-frames. The data frames were concatenated and cleaned. Unnecessary columns were dropped, and the percentage column was transformed to show the values in percentages. After dropping the data for Delaware, the District of Columbia, and Puerto Rico (District of Columbia and Puerto Rico data was dropped as they weren’t states), the weather and migration data frames were merged on state and date columns.

A screenshot of a spreadsheet

Description automatically generated

Figure 3: Initial migration data obtained

A table with numbers and text

Description automatically generated

Figure 4: Migration Data-frame after data-cleaning

2.3 Merged data:

Both the Weather and the Migration data frames didn’t have any missing values, NaN values, or blanks after the data cleaning. There were data points with average snow as zero, this data is to be expected for some years and states.

We merged the weather and migration data frames so that each state and its corresponding year had both the migration data and the weather data. We did some additional cleanup by using the state’s full name instead of its abbreviation and changing the date column to a year column.

A screenshot of a graph

Description automatically generated

**Figure 5: Merged Data frame**

The merged data frame has 24 columns representing the data across movement between states and the state’s weather patterns.

We analyzed the distribution of the data for the merged data frame using qq plots. From the qq-plots of the merged distribution, we could see that in the weather-related columns, temperature appears to be relatively normally distributed, while the migration-related columns are non-normally distributed.

A screenshot of a computer code

Description automatically generated

**Figure 6: Merged Data frame Data types**

3. Visualizations

To explore the weather data and migration data for the United States, we used visualizations to uncover initial patterns.  Link of [website](https://sites.google.com/view/sanjminalchala/introduction/project-introduction?authuser=2).

First, we started by looking at the weather patterns across the years and the states.

A graph of the temperature of the sun

Description automatically generated

**Figure 7: Average Temperature in US (2005-2022)**

In the above plot, we can see that the average temperatures in the United States show a cyclical behavior with consecutive high and low dips in the temperature with the mean average temperature 55F and the range being 53F to 56.5F.

A graph of a number of precipitation

Description automatically generated

**Figure 8: Average Precipitation in US (2005-2022)**

The above plot shows the average precipitation (in) in the United States over the years, with the mean precipitation being 3 inches. Concerningly, we can see that after 2018, average precipitation is declining.

A graph with orange lines

Description automatically generated

**Figure 9: Average Snow in US (2005-2022)**

The above plot shows us the average snow(in) in the United States from 2005 to 2022. There was a sharp drop in the average snow in 2006 and 2012, after that, the average snow ranged between 2 to 3 inches a year.Conference Short Name:WOODSTOCK’18

A screenshot of a graph

Description automatically generated

**Figure 10: Correlation Heatmap for Temp, Rain, Snow**

The above heatmap shows the correlation between the temperatures, snow, and rain. We can see that the average snow has the highest correlation with average temperature and then with minimum temperature. Let's see how exactly the data is distributed with the help of scatter plots.

A graph of a graph showing the average temperature and snow

Description automatically generated

**Figure 11: Average Temperature and snow**

The above scatterplot shows that the yearly average snow and yearly average temperature tend to move together.

A graph of blue dots

Description automatically generated

**Figure 12: Minimum Temperature and Average Snow**

The above scatterplot shows that the yearly average snow and yearly minimum temperature tend to move together. We did not find any worthwhile trends when looking at temperature and precipitation.

A graph of snow fall

Description automatically generated

**Figure 13: Average Snowfall in Cold Months by States**

The above graph compares snowfall in the winter by state in 2005 and 2022. In most states, it appears that snowfall was more prevalent in 2005 compared to 2022. This aligns with the broader trends of climate change, including rising global temperatures resulting in lower snow patterns.

A graph of different colored lines

Description automatically generated

**Figure 14: Population Percentage Growth by States (2005-2023)**

To better understand the migration patterns for each of the states, we plotted migration trends over the years for each state. The line plot shows a significant common trend across all the states; between 2009 and 2010, every state’s population growth percentage dropped from over 5% to under 5%. One can speculate that the drop could be associated with the major event in the United States at that time, the Great Recession. To analyze the states’ migration data better, we split the data to look at 2005-2009 together and the data between 2010-2022 together.

A graph of different colored lines

Description automatically generated

**Figure 15: Population Percentage Decline by States (2005-2009)**

We first focused on exploring migration data from 2005 to 2009. We found that there were specific states that had decreasing population growth and significantly less population growth than most other states: less than 15%. California and Michigan had lower population growth than most states, less than 10% and declining.

A graph of different colored lines

Description automatically generated

**Figure 16: Population Percentage Growth by State (2010-2022)**

Between 2010 and 2022, most states did not have more than 10% population growth. There were specific states that were increasing greater than 3 percent and exponentially over most other states. In Vermont, there was a significant increase in population between 2017 and 2021. Keep in mind that there is no data for 2020, due to Covid so there is a gap for that year. Idaho had the most significant population growth percentage between 2010 and 2022 with over 5% in 2021.

A graph showing the growth of the average temperature

Description automatically generated

**Figure 17: Population Growth and Avg Temperature in Michigan (2005-2009)**

In Michigan, average temperature and population growth percentage had similar trend lines throughout 2005 and 2009. Between 2005 and 2006, population growth and average temperature increased. Between 2006 and 2009, the population percentage decreased and so did the average temperature. Both population growth and average temperature had similar slopes throughout the increase and the decrease.

A graph of the temperature

Description automatically generated

**Figure 18: Idaho’s Average Temperature (2010-2022)**

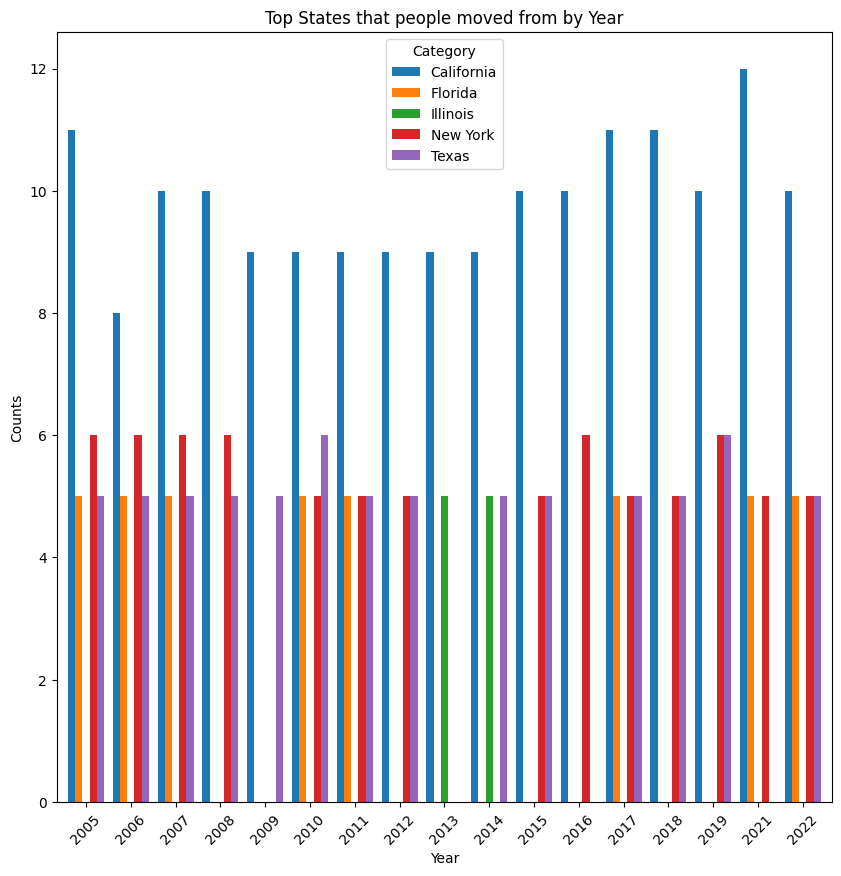
Idaho’s average temperature from 2010 to 2022 stayed consistent - around the 50s.

A graph showing the temperature of the year

Description automatically generated

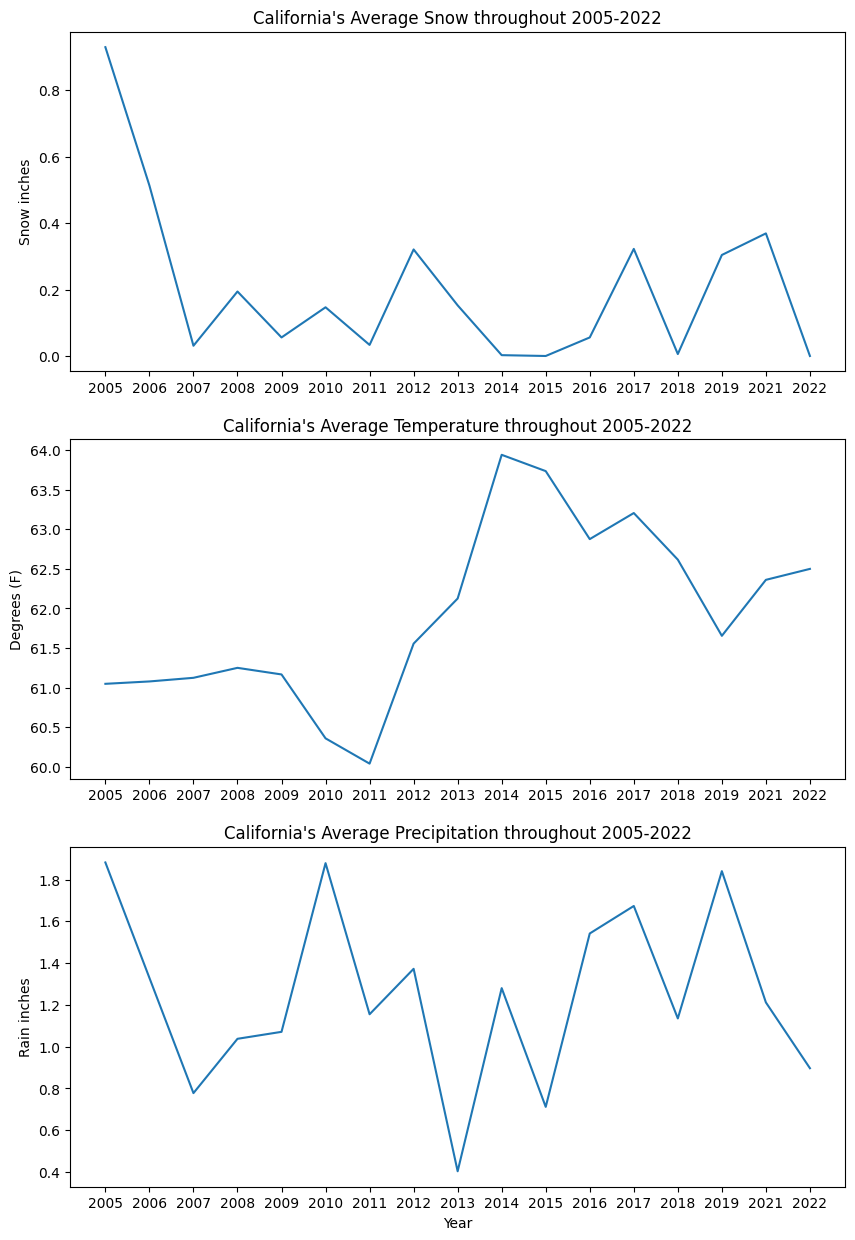
**Figure 19: Vermont’s Average Temperature (2010-2022)**

Vermont’s average temperature throughout 2010 to 2022 stayed consistent - around the high 40s, between 44- and 48-degrees Fahrenheit.



**Figure 20: Top states that people moved from by Year**

The above graph shows the top states that people move from to another state by year. The top states are California, Florida, Illinois, New York and Texas. California is the top state that people moved from throughout 2005 and 2022, surpassing other states throughout those years.



**Figure 21: California weather parameters (2005-2022)**

California’s average temperature increased by 3 points between 2011 and 2015. California’s snowfall dropped by a quarter since 2005. California’s rainfall is also volatile with it fluctuating between less than 1in to no more than 2 inches throughout the last 17 years.

4. Models Implemented

*Data Cleaning*: Our data is a combined dataset of weather and migration datasets. When we looked closely at the migration data, there was a sharp drop in the metrics calculated, which was caused due to the change in the data collection methods by the Census Department of the USA during 2009 and 2010. Therefore, the data before 2010 was removed from further inferences and analysis. We performed preliminary data analysis in R programming (the code and result snippets of Rare at the end of the report) to understand the data and evidence of multicollinearity was found due to the variables that were derived from other variables, such as temperatures that correlated to winter and summer. These were originally leveraged to see if there were any prominent trends between winter and summer weather patterns that needed to be taken into account for migration. However, during our analysis stage, we found that they weren’t significant. Therefore, these low-significant features were dropped.

For one feature in the dataset, named yearly\_average\_snow, NAs were replaced with 0s. For modeling purposes, the state names were transformed into numbers using scikit-library functions. The features or X variables of the models are the weather data, and the response or y variable of the model is the percentage of population migrated to a state in a given year.

Since the dataset comprises of time-series and numerical data, the models were chosen to accommodate for these factors, specifically that these factors would not affect the models’ performances. Originally, we attempted to split the data by years for our train and test data, with train being from 2010 to 2018 and test data being from 2019 to 2022. However, when tested against different machine learning models, the models performed poorly. We found that a random split worked better for most of the models as indicated below. We did leverage splitting the data by years for train and test in the ARIMA model.

We implemented 5 models: Random Forest, Support Vector Machine (SVM), Gradient Boosting Machines (GBM), Linear Regression, and Autoregressive integrated moving average (ARIMA).

To see the implementation of the models check the [link](https://colab.research.google.com/drive/1Dxm1Gdj45rOACEnNPpRSTrbFoXBYuw7x?usp=sharing).

A screenshot of a computer code

Description automatically generated

**Figure 22: Data before Transformation for models**

A computer code with black text

Description automatically generated 

X (features)

Y (response)

**Figure 23: Dataset after transforming for all the models’ implementation**

4.1 Random Forest Regression:

The Random Forest Regression was used as a model since it builds multiple decision trees during training and gives the average of their predictions. Since our dataset has multiple weather features that don’t exactly have a linear relationship with the percentage of migration, this model was suitable for predicting migration percentage. The X(feature) and y(response) were randomly split into 80% and 20%, named as train and test datasets. Hyper-parameters **max\_depth** and **max\_features** were used to create a more accurate model and different parameters were estimated. The hyper-parametermax\_depthcontrols the depth of the tree, with a deeper tree risking overfitting and a shallow tree risking underfitting. Max\_features defines the number of features to look at per decision tree in the random forest.

Mean Squared Error (MSE) and Mean Absolute Errors (MAE) were leveraged as metrics to test the accuracy and reliability of the model. The MSE indicates the average squared difference between the estimated values and the actual value, and its purpose is to test the quality of the model. Without hyperparameters, the MSE is 10.8%. With hyper-parameters of max\_depth = 20 and max\_features = “log2”, the Mean Squared Error (MSE) was 10.6%. Although the difference between having hyper-parameters and not is quite small, 0.002, it could be of some significance considering that the percentages of migration fall between 0-10%.

The Mean Absolute Error (MAE) of the model without hyperparameters is 0.234. With hyper-parameters, the MAE is 0.232 indicating that the model is about 0.23 percentage points away from the actual values.

MSE and the MAE are relatively small indicating that the Random Forest model is accurate and reliable for predicting migration patterns based on weather patterns.

4.2 Support Vector Machine (SVM):

The support vector regression model is used to predict continuous outcomes by fitting as many data points as possible within a specified error margin while reducing the model complexity to prevent overfitting. Similar to Random Forest the X(feature) and y(response) were randomly split into 80% and 20% train and test data. Without hyper-parameters, the MSE for the SVM was 18% and the MAE is 0.32.

The hyper-parameters used for SVM were C, gamma, and epsilon. The C parameter controls the complexity and accuracy of the model; gamma defines how much influence individual data points have on the prediction model; epsilon controls how sensitive the model is to the “noise” in the dataset. The hyperparameters chosen that allowed for a decrease in MSE and MAE were C = 5, gamma = 0.1, and epsilon = 0.2.

With hyper-parameters, the MSE is 16% and the MAE is 0.30. The MSE difference of 0.03 amounts to an 11.11% error reduction which is a significant reduction for migration patterns as the migration percentages are small. With hyper-parameters, the model was more accurate and reliable, albeit not as accurate as random forest.

4.3 Gradient Boosting Machines (GBM):

In Gradient Boosting Machine (GBM) model, the "gradient" helps the model get better by fixing its errors step by step. Boosting means each new iteration tries to learn from the previous ones to perform better. GBM builds trees, with each new tree fixing the mistakes of the ones before it.

Our dataset was split into two parts: one for training the model and the other for testing it. We set 80% of our data for training (X\_train, y\_train) and kept 20% for testing (X\_test, y\_test).

Mean Squared Error (MSE) was used to measure how close the model’s predictions were to the actual values. The MSE of our model was approximately 14.94%. This number indicates that on average, the GBM model's predictions are about 0.1494 away from the real values. This was not the best MSE of the models that have been implemented so far. To make the model better, we can adjust its settings, understand which factors are most important, or test it with different data to see if it works well in general.

4.4 Multiple Linear Regression Model (MLR):

The Multiple linear regression model is one of the standard statistical models to evaluate the relation between the feature and response variables. Its assumptions make it easier to interpret and infer about the data. The regression models assume homoscedasticity, normality and linearity from the data for the model to be unbiased and as efficient as possible. When checked for the assumptions using residual vs fitted plot and Q-Q plot below, we find that the assumptions are satisfied but not in its entirety. The X(feature) and y(response) were randomly split into 80% and 20%, named as train and test datasets. The model is trained on the train dataset and tested for accuracy and errors on test data. After fitting the linear regression model, the MSE (Mean-squared-error) metric was used to check the models’ performance.

A collage of graphs and charts

Description automatically generated

**Figure 24: Residual Plots**

Below we find the fitted MLR model summary, and we can see that the p-values for many variables are very low. When the p-values are lower than 0.05 (default significance level), that means that those feature variables are significant in the model and that those features contribute to explaining the patterns in the response variable.

A screenshot of a computer

Description automatically generated

**Figure 25: Summary of Preliminary Linear regression**

The **MSE** turned out to be more than **30%** on any run, which indicates the model is performing poorly. There are several factors for the result we obtained, like the small size of the dataset, the data being time-series data while the MLR assumes the data points to be independent, the features not capturing the total changes in the response variable, and so on. We next perform a model that is used mainly for time-series data.

4.5 Autoregressive Integrated Moving Average Model (ARIMA):

The ARIMA model is often used to forecast time-series data. The model takes the response variable as the input, accounts for the seasonality of the variable, identifies the trend of the data, and forecasts the next value for the future time-point. Given the methodology, we fitted an ARIMA model for each of the 50 states in the United States and calculated the MSE for each state separately. The response variable for each state consists of 11 data points which were divided into train data (2010-2018) and test data (2019-2022) of 80% and 20% split.

The **mean MSE** for all the states is **13%** with a standard deviation of MSE being 22%**,** the median of the MSEs calculated for all the states is 4% and the model for Virginia state leads to an MSE of 0.4%, the best-performing model out of all states. The median gives us a better idea of the model statistics, as it is robust to outliers, and since there is a huge difference between the mean and median we can say that there is a presence of outliers, meaning some of the ARIMA models performed very poorly.

As we know lower MSE means that the predicted values are close to the actual test data values and that the model is performing well, looking at the obtained MSEs above, we can say that ARIMA models are one of the best-performing models for time-series data, but with more data-points, this would lead to better results overall and not just for one or two states.

5. Research Questions and Objectives

In this section, we answer the research questions that were formed based on our interest in the effect of climate change on internal migration in the United States. As our dataset consists only of population migration data in the United States overall and weather data, the questions related to age, demographics, household, and socioeconomic status were unanswered. These are the factors we believe will help us build a more efficient and unbiased model in the future.

**What states are going to be more populated in the next five years?**

From the data we have, we can predict that California, Texas, Florida, New York, and Illinois, might be the most populated states in the next 5 years. The states have Interestingly, these states happen to be the same 5 states from which people migrate the most as we saw in one of our visualizations.

**What regions are going to be more affected by weather patterns and natural disasters and what is the population there going to be?**

California, and Arizona had record high temperatures during the summers and during the year overall for the past decade and are expected to continue to have hot summers in the future. While Alaska, Minnesota, and Montana are going to be on the other end of the climate spectrum with the coldest weather conditions as they endured the coldest winters with the highest records of snowfall over the past decade. The wildfires in California are claiming more and more forest areas and people over the years. Using one of the models implemented, the percentage growth was forecasted for the next 3 years. The expected percentage growth of population due to internal migration would be 4.75% on average for the next 3 years in Alaska, 3.72% in Arizona, 1.26% in California, 1.93% in Minnesota, and 4.06% in Montana.

**How are regions with seasonal changes affecting migration patterns versus regions that are mostly consistent throughout the year?**

California, Michigan, and New York are the states with consistent weather throughout the year and the population growth for these rose consistently as well. Whereas regions such as Wyoming, Alaska, and North Dakota, which have varying weather depending on the seasons, had the opposite trend in seeing lows and highs in their population growths over the past 5-6 years.

**How do changing weather patterns influence migration patterns in the United States?**

When we look at our data and the models’ results, we can say that the regions with consistent weather seem to have a lot of internal migration, like the places California, Texas, and Florida. While places like Colorado, and Minnesota with drastic weather seem to have relatively lower internal migration. Therefore, we can say that the weather patterns are one of the contributing factors to migration numbers.

6. Key Results

|  |  |
| --- | --- |
| Model | MSE |
| Random Forest | 10.80% |
| Random Forest with Hyper-parameters | 10.60% |
| Support Vector Machine (SVM) | 18.00% |
| SVM with Hyper-parameters | 16.00% |
| Gradient Boosting Machine (GBM) | 14.94% |
| Linear Regression | 30.00% |
| ARIMA | 13.00% |

The table above lists the models implemented (Random Forest, Support Vector Machine- SVM, Gradient Boosting Machine – GBM, Linear Regression – MLR, ARIMA) and their respective Mean squared Error (MSE). Some models were tuned using hyper-parameters. The lower the MSE is the closer the predicted values are to the actual values of test data. Out of all the models, the Random Forest model with hyper-parameters had the least MSE of 10.6%. The models in general do not account for the data being time-series data and the fact that not all significant features were present in the model, the MSE turned out to be higher. For further decreasing the MSE, increasing the dataset size in factors and sample size might be helpful.

7. Conclusion

Our research explores the impacts climate change has on US internal migration, specifically focusing on the changing weather patterns over the last 20 years. The data analysis highlighted that fluctuations in average state temperatures correlate with population change. Notably, California experienced a decline in population growth rates as well as led the chart of states from which people are relocating; this coincides with the state's patterns of erratic temperatures and precipitation levels.

Among the five predictive models implemented to predict population growth from weather data such as temperature, precipitation, and snowfall, the Random Forest Regression and the ARIMA (Autoregressive Integrated Moving Average) models were the best in accuracy and reliability.

The insights from the research illustrate the influence of weather patterns on US state migration. Research like this can be used for urban planning, state policies, and city/state infrastructure planning. Policymakers can leverage these insights to guide infrastructure development, resource allocation, and community support initiatives, ultimately shaping a sustainable and adaptable future for the nation's evolving demographic landscape.

To expand on this research, it would be interesting to include natural disaster impacts in the analysis. This would enrich our understanding of migration patterns, offering a nuanced view of the connection between climate disasters and choice of living. Moreover, it would be worth also considering the economic consequences of climate change, such as its effect on insurance premiums, providing a holistic picture of the various factors impacting migration as it relates to climate change.

As climate change becomes the of our time, this study takes a step forward toward understanding and addressing the complex relationship between environmental shifts and migratory patterns within the United States.

8. Data Sources:

[1]<https://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-to-state-migration.html>

[2] <https://www.ncei.noaa.gov/>

9. R-code:

weather <- read.csv('data/merged\_dataframe.csv') #reading file

weather <- weather [!weather$year %in% c(2005, 2006, 2007, 2008, 2009), ] #dropping those year rows

rownames(weather) <- NULL #reindexing

weather$state <- factor(weather$state) #converting the state to numerical

weather$state <- as.numeric(weather$state)

colSums(is.na(weather)) #checking nas

weather$yearly\_avg\_snow[is.na(weather$yearly\_avg\_snow)] <- 0 #replacing nas with 0s

weather <- weather[, !(names(weather) %in% c("top1.state"))] #dropping top1 state as it is a string variable

weatherlm <- lm(percentage ~ ., data=weather)

summary(weatherlm)

options(repr.plot.width = 9, repr.plot.height = 9)

par(mfrow = c(2,2))

plot(weatherlm) #plots of the fitted regression model

#(plot in the model section)

10. References:

[1] USGCRP, 2018: Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II [Reidmiller, D.R., C.W. Avery, D.R. Easterling, K.E. Kunkel, K.L.M. Lewis, T.K. Maycock, and B.C. Stewart (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, 1515 pp. doi: 10.7930/NCA4.2018.

[2] Rebecca Lindsey, Luann Dahlman, 2024. Climate Change: Global Temperature. [Link](https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature#:~:text=2023%20was%20the%20warmest%20year,1850%20by%20a%20wide%20margin.)

[3] Abrahm Lustgarten, 2020. Where will everyone go? [Link](https://features.propublica.org/climate-migration/model-how-climate-refugees-move-across-continents/)

[4] Abrahm Lustgarten, 2023. Climate Crisis Is on Track to Push One-Third of Humanity Out of its most livable environment. [Link](https://www.propublica.org/article/climate-crisis-niche-migration-environment-population)

[5] The human cost of disasters: an overview of the last 20 years (2000-2019). [Link](https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019)

[6] Lawrence Huang, 2023. Climate Migration 101: An explainer. [Link](https://www.migrationpolicy.org/article/climate-migration-101-explainer#origins)

[7] Rigaud, Kanta Kumari; de Sherbinin, Alex; Jones, Bryan; Bergmann,Jonas; Clement, Viviane; Ober, Kayly; Schewe, Jacob; Adamo, Susana; McCusker, Brent; Heuser, Silke; Midgley, Amelia. 2018. Groundswell: Preparing for Internal Climate Migration.  WorldBank, Washington, DC. http://hdl.handle.net/10986/29461 License: [CC BY 3.0 IGO](http://creativecommons.org/licenses/by/3.0/igo).

[8] Climate and weather-related disasters surge five-fold over 50 years, but early warnings save lives – WMO report, 2021. [Link](https://news.un.org/en/story/2021/09/1098662)

[9] Mia Prange, 2022. Climate Change is Fueling Migration. DO Climate Migrants Have Legal Protections? [Link](https://www.cfr.org/in-brief/climate-change-fueling-migration-do-climate-migrants-have-legal-protections#:~:text=Climate%20migration%20occurs%20when%20people,seas%20and%20intensifying%20water%20stress)

[10] McLeman, R. Thresholds in climate migration. Popul Environ 39, 319–338 (2018). https://doi.org/10.1007/s11111-017-0290-2

[11] NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters (2024). <https://www.ncei.noaa.gov/access/billions/>,DOI: [10.25921/stkw-7w73](https://www.doi.org/10.25921/stkw-7w73)

[12] Samantha Allen, 2024. 30% of Americans cite climate change as a motivator to move in 2024. [Link](https://www.forbes.com/home-improvement/features/americans-moving-climate-change/)

[13] Pat Howard, 2023. Home insurance prices up 21% as homeowners are left to deal with climate change, turbulent market. [Link](https://www.policygenius.com/homeowners-insurance/home-insurance-pricing-report-2023/)

[14] Anna Bahney, 2023. Climate change is impacting the home insurance industry and damaging the housing market. [Link](https://www.cnn.com/2023/11/07/homes/homeowners-insurance-climate-real-estate/index.html)

[15] Abrahm Lustgarten, 2020. Climate Change will force a New American Migration. [Link](https://www.propublica.org/article/climate-change-will-force-a-new-american-migration)