# InsurTex

IST 718

Marley Akonnor, Alexander Klein, Natalia Uruska, Winston Vu
Alison Reikher

# Abstract

Over the last 21 years the climate has experienced many shifts that have radically changed the lives of Texans. Of the 7 cities examined, many experienced a 20-year spike in average temperatures in 2012. Though most cities began to exhibit observable cadences, there was no ubiquitous patterns that would declaratively suggest contemporary changes in temperature will persist going forward. Forecasts from the Texas Homeowners Insurance dataset showed a trendline of increasing premiums consistent with the rising premiums of the 21 years prior. Our predictions were in striking range of the actual values and demonstrated a low level of variance. There was a statistically significant change point in 2015.

Below are the recommended InsurTex premiums for the next 5 fiscal years based on Prophet forecasting. We are confident that these prices will grow our margins while providing affordable insurance to all our loyal customers in El Paso, Amarillo, Dallas, Houston, Austin, Brownsville, and Laredo.

| Fiscal Year | Premium Price |
|---|---|
| 2022-2023 | $1,406.10 |
| 2023-2024 | $1,446.09 ⇧ |
| 2024-2025 | $1,487.14 ⇧ |
| 2025-2026 | $1,516.30 ⇧ |
| 2026-2027 | $1,555.24 ⇧ |

# Specification:

Global winds, air masses, fronts, jet streams, ice ages, and global warming are all weather patterns that have had tremendous impacts on how humans interact with the earth. Contemporary science with all its advancement still relegates the discipline of meteorology to a paltry game of probability. Sun emojis with smiles and thunder emojis with frowns dilute our comprehension of nature's forces until we are unceremoniously reminded by a natural disaster. Hurricane Katrina still rings loudly in the minds of many. The 2005 hurricane was responsible for over 1,800 fatalities and caused approximately $108 billion in damage (National Weather Service, n.d.). Last February alone, 246 Texans over 77 counties died from an unprecedented winter storm (Diaz, 2022). The storm exposed vital infrastructure inadequacies such as energy, electrical, and water systems leaving millions without power to heat their homes and clean drinking water (Doss-Gollin, 2021). We endeavor to be a force of solace and assurance for the people of Texas. Our promise is, though the tides may ebb and flow, our protection will not. We are the property insurance company, InsurTex. We provide the largest insurance coverage system in the state of Texas, insuring millions of residents within 7 of Texas' metro areas: El Paso, Amarillo, Dallas, Houston, Austin, Brownsville and Laredo.

Our mission is to determine the best insurance premium costs for us as a business while ensuring affordability for our loyal customers. We are attempting to predict future weather trends and their impact on insurance premium costs over the next 5 years. This includes potential hurricanes and other natural disasters as well as major changes in temperature and rainfall. As the climate continually changes and weather patterns become increasingly unpredictable, we anticipate that premiums will rise marginally until the underlying infrastructure issues in Texas can and have been addressed.

We have collected the following data for research and analysis:

1) 2020 Insurance Council of Texas Property & Casualty Insurance Market Report
2) Assessment of Historic and Future Trends of Extreme Weather In Texas, 1900-2036
3) Corporate and Insurance Data 2008-2021
4) Monthly and Annual Totals – Precipitation Data for Texas
    a. El Paso Data + totals
    b. Amarillo Totals
    c. Dallas
    d. Houston
    e. Brownsville
    f. Laredo
5) FEMA National Risk Index
    a. El Paso
    b. Amarillo
    c. Dallas
    d. Houston
    e. Austin
    f. Brownsville
    g. Laredo
6) State Insurance Regulation: Key Facts and Market Trends
7) Insurance Information Institute
    a. Direct Premiums Written

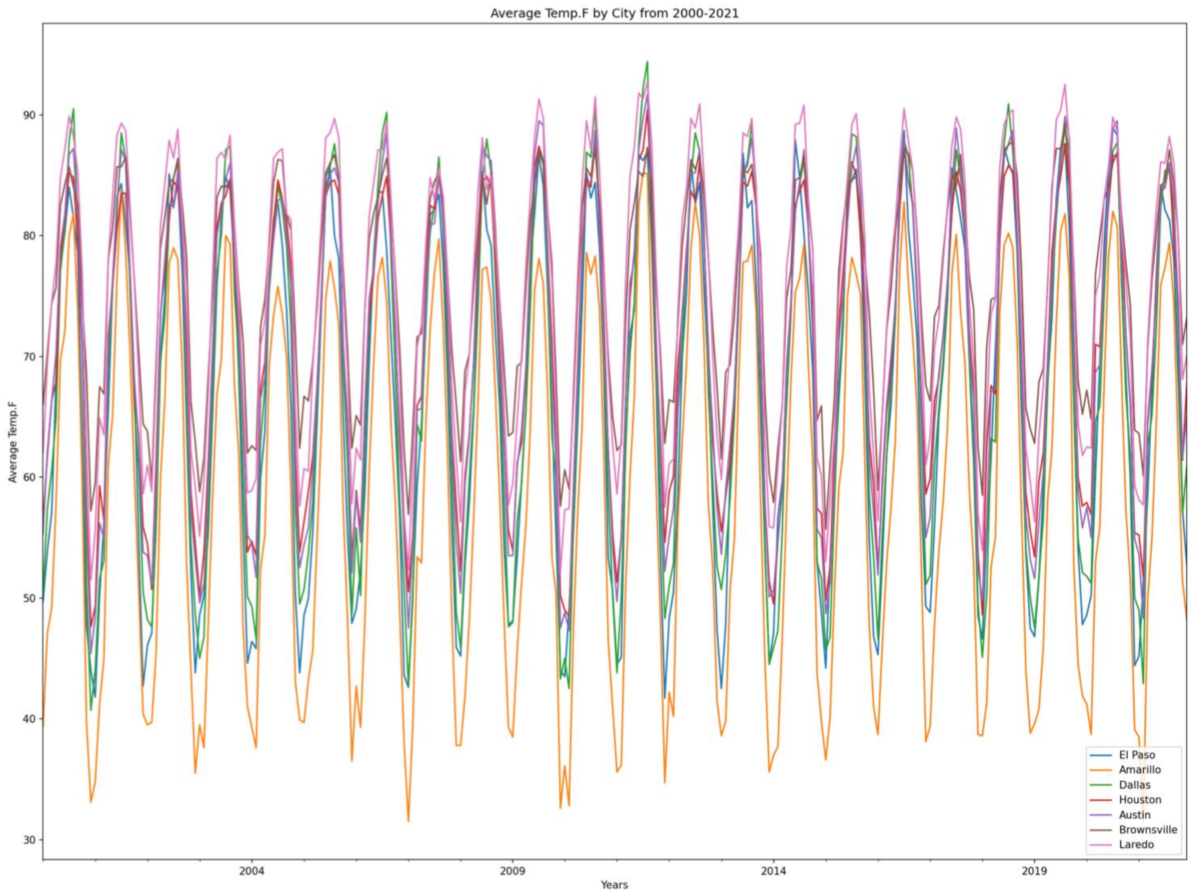
    b. Domestic Insurance Companies
    c. Property Casualty Insurance
8) Energy Star – Texas Weather

The data structures are as follows:

```
Date                      object
ElPasoAvgTemp            float64
ElPasoPrecipSum           object
ElPasoSnowSum             object
AmarilloAvgTemp          float64
AmarilloPrecipSum         object
AmarilloSnowSum           object
DallasAvgTemp            float64
DallasPrecipSum           object
DallasSnowSum             object
HoustonAvgTemp           float64
HoustonPrecipSum         float64
HoustonSnowSum            object
AustinAvgTemp            float64
AustinPrecipSum           object
AustinSnowSum             object
BrownsvilleAvgTemp       float64
BrownsvillePrecipSum      object
BrownsvilleSnowSum        object
LaredoAvgTemp            float64
LaredoPrecipSum           object
LaredoSnowSum             object
dtype: object
```

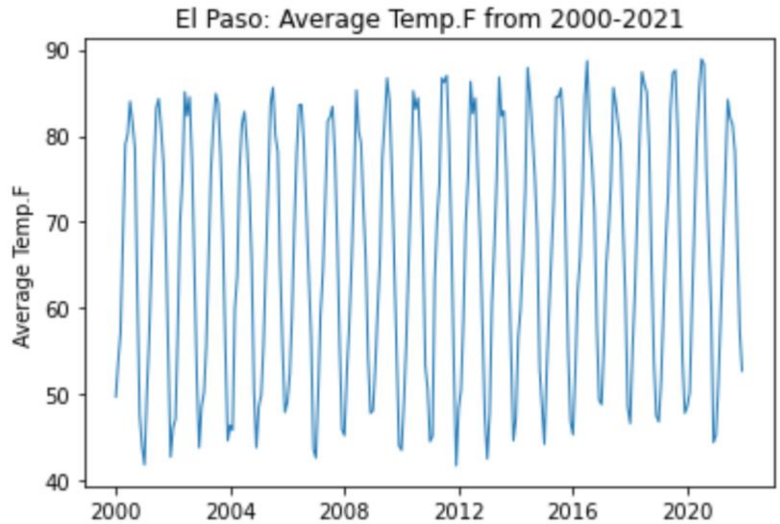
# Observation:

When observing the data, two trends became apparent. In nearly all 7 cities within our service area, the colder months have been getting slightly colder on average and the hotter months have been getting slightly warmer on average. Before examining individual differences, we began with an overlay of temperature fluctuations for all cities.

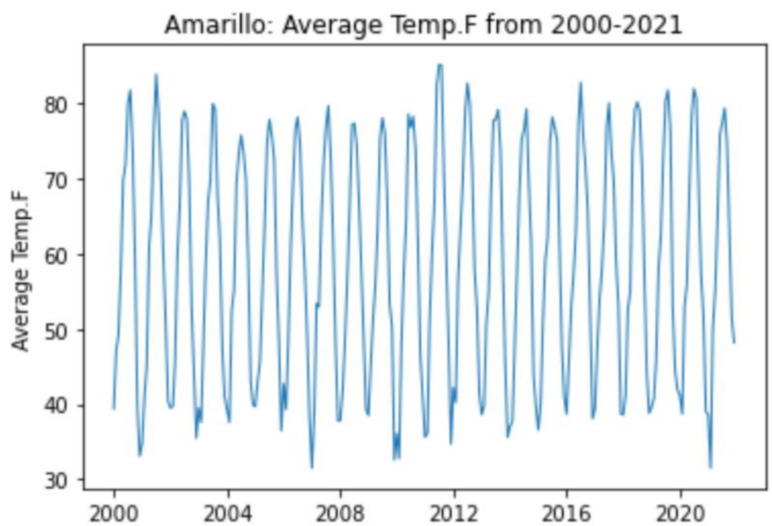



The above plot displays average temperature by city from the year 2000 to 2021. Though Dallas had the highest average temperature over the 21-year time horizon with a spike to ~92 degrees in 2012, Loredo has consistently been the hottest of the 7 cities. Conversely, Amarillo has consistently had the lowest average temperatures without competition dropping to temperatures as frigid as ~30 degrees in 2007 and achieving it once again in 2021.

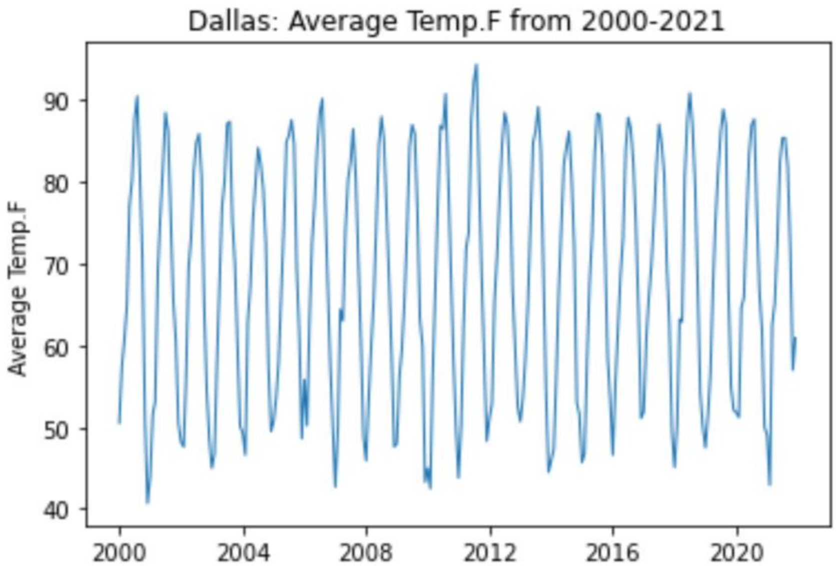
Transitioning to individual cities, we begin with El Paso, Texas. El Paso, interestingly, seemed to show more consistency at the higher end temperatures while showing more variance at the colder temperatures. The warmest point occurred in 2017 when the warmest average temperature was consistently around 90 degrees. Whereas, the coldest average temperature of the last 20 years occurred in 2012 at approximately 40 degrees. Regarding the most recent years, 2020 saw an average high temperature of ~90 degrees nearly matching 2017 as the hottest temperature of the last 21 years. 2021 dipped to an average low of 45 degrees.



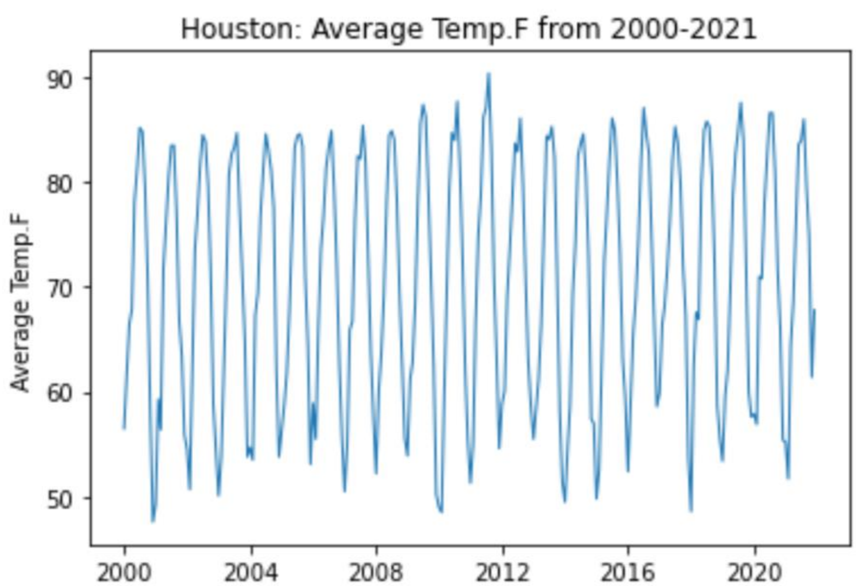

Amarillo displayed a rather unique cadence. There appeared to be a drastic spike of hot or cold averages temperatures and then a group of years in between with less variance. In 2002 and 2012 the average temperature spikes to around 85 degrees. The second warmest series was in 2000, 2013, and 2017 where the average highs were around 82 degrees. The cold spectrum matched that of the warmer temperatures. Most notably, 2007 and 2021 were around 30 degrees. The next closest grouping of years was 2001and 2009 which were around 35 degrees on average. If Amarillo continues this pattern, the next few winters shouldn't be as harsh.



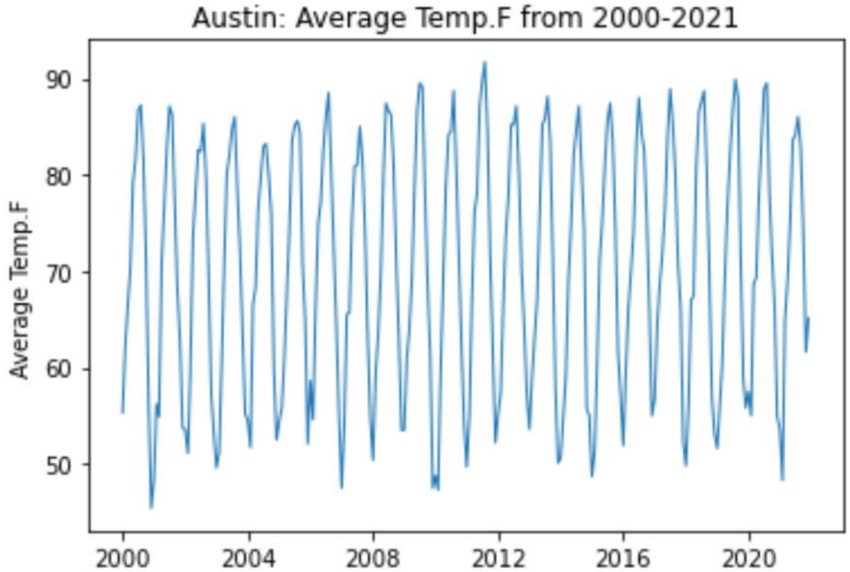

Dallas distinctly had its hottest average temperature in 2012 soaring to almost 95 degrees. The coldest average temperatures, conversely, were achieved at many different points in time. In 2001 the lowest average temperature was 40 degrees. In 2006, 2010, and most recently 2021, the lowest average temperature was in striking distance at ~37 degrees. A pattern of sorts still emerges on the cold end. There seems to be a cold spike and then 3 to 4 winters of rising temperature before another one. If this persist as true,  then the next few winters could be progressively warmer in Dallas.



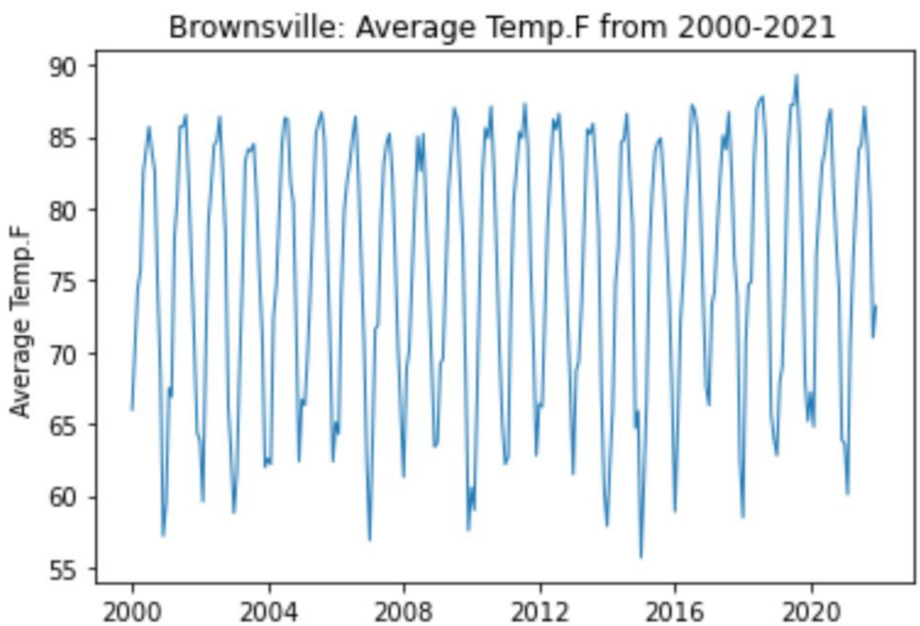

Houston, similarly, to Dallas and Amarillo, had its highest average temperature spike of the last 21 years in 2021 where it reached 90 degrees. Curiously, in Houston the average temperature high was almost level, hovering around 85 degrees from 2000 to 2010. In 2010 and 2011, there was a paired average temperature of ~87 degrees. After the heat spike of 2012 inconsistent high temperatures became apparent. On the cold end of the spectrum, there was a similar cadence to Dallas in that there was seemingly a cold spike and then several consecutive years of less frigid temperatures before another deep cold spike.



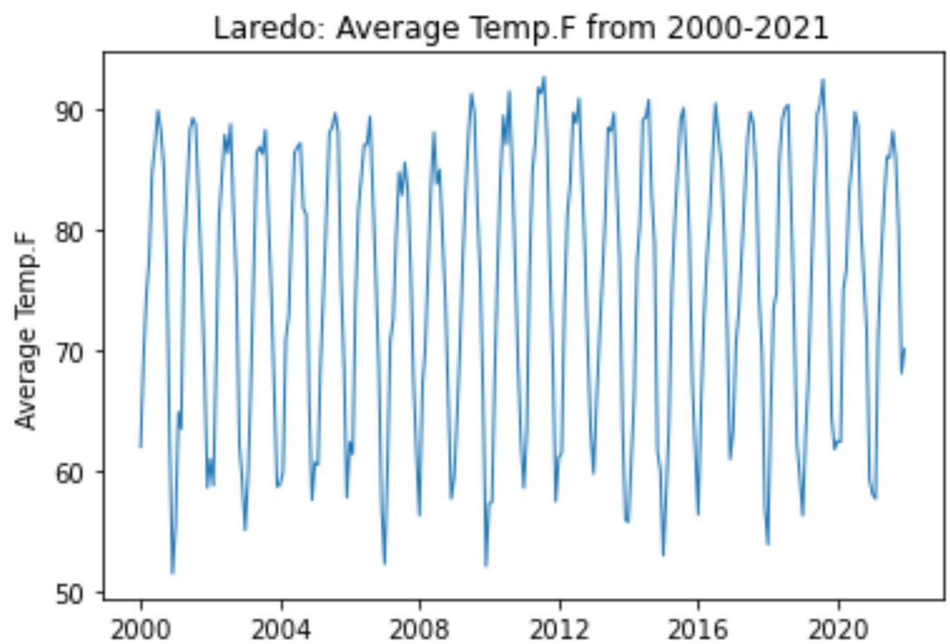

At this point it seems fair to assume there was a heat wave of sorts in 2012 because Austin did not diverge from the pattern either. Not only did it follow suit, 2012 was also the hottest average temperature of the last 21 years for Austin at ~92 degrees. Opposite from Houston, Austin had sporadic heat highs prior to 2012 but after, a lot more consistency became pervasive until 2021. On the opposite end of the spectrum, 2001was the coldest average temperature at a little above 40 degrees. The lowest average temperatures were warmer than other cities hovering around 50 degrees on average.




Brownsville joins El Paso as the second city to not have experienced the heat spike of 2012. Instead, the hottest average temperature of the last 21 years happened more recently. In 2020 the average temperature of Brownsville reached 90 degrees and has been cooler after. There was much less variance in the hottest average temperatures than in the coldest. The coldest average temperature in Brownville occurred in 2015 at 55 degrees. Thus far, Brownsville has held the warmest cold temperature of any city. 2021 was one of the warmer winters with an average low of about 60 degrees.




Finally, while Laredo experienced the infamous heat spike of 2012 as well, it matched that high in 2019 again reaching an average temperature of ~92 degrees. The years 2020 and 2021 have gotten cooler on average but are still in the mid-80's. The coldest point of the last 21 years has seemingly been reached on multiple occasions. In 2001, 2007, 2010, and again in 2015 the coldest average temperatures were in the very low 50s. This most recent winter was very warm comparatively, only reaching an average temperature of 60 degrees.




## Analysis:

The data being used in this project to forecast the weather is time series data with multiple variables. The model best suited for this type of data is the Skforecast model. The Skforecast model works with time series data that contains trends and seasonal components while allowing for exogenous variables to be included in the forecasting process. We will use the Skforecast model with the data we collected to forecast the future weather in all seven cities we provide insurance coverage to.

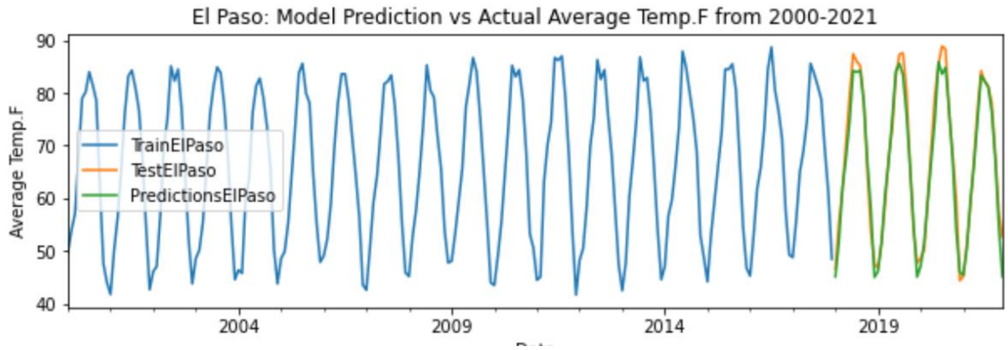
The Skforecast model uses Sklearn's RandomForestRegressor as a multi-step forecasting approach. Using the multi-step approach allows for exogenous variables to be included in the forecasting process while allowing more predictor customization along with the ability to get predictor importance. The RandomForestRegressor model utilizes ensemble learning by using multiple decision trees trained using sub-samples from the dataset and averages the results of each model to obtain more accurate classification results while minimizing potential overfitting.

The insurance premium data is univariate time series data. Facebook's Prophet model can be used to easily forecast future insurance premiums. That forecast can be used as a basis for our future insurance premiums. The future weather forecast combined with the extreme weather risk data will be used to adjust the future insurance premiums to the degree that is financially sustainable for all seven cities.

We began the analysis with the Skforecast model to set up the weather forecasts for all 7 cities. Test and training datasets using the last 48 months as the testing data were constructed. The trained model was then used to predict the next 48 months to allow us to observe forecast accuracy. The accuracy was demonstrated visually with 2 plots. The first graph, consisted of the training data, test data, and predictions. To validate the accuracy through quantification the MSE and RMSE values were calculated. Regarding the second plot, the model was trained using the entire dataset and was set to predict the next 60 months out.

Below are the results for all 7 cities:

## 1) El Paso City Weather Forecasting:

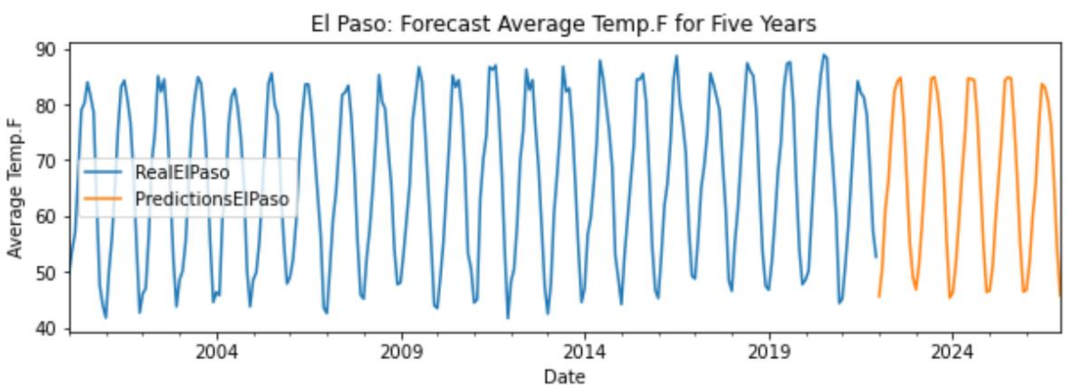



```
# Calculate the MSE (mean squared error)
MseEP = mean_squared_error(y_true = TestEP['ElPasoAvgTemp'], y_pred =␣
 ↪PredictionsEP)
print(f"El Paso MSE Value: {MseEP}")
```

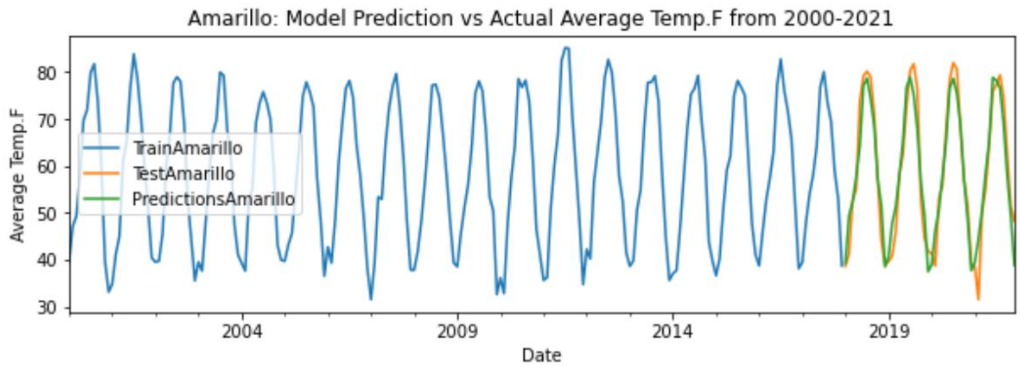
```
El Paso MSE Value: 6.927929874999948
```

```
# Calculate the RMSE (root mean squared error)
RmseEP = np.sqrt(MseEP)
print(f"El Paso RMSE Value: {RmseEP}")
```

```
El Paso RMSE Value: 2.6320960991194733
```



## 2) Amarillo City Weather Forecasting:

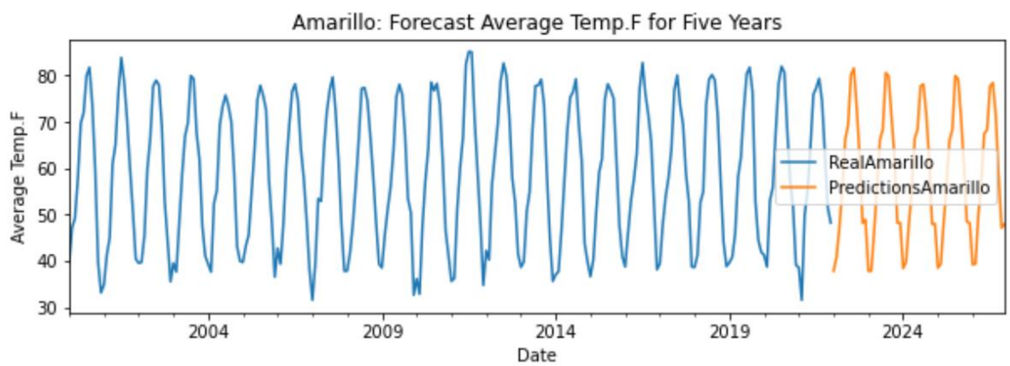



```
# Calculate the MSE (mean squared error)
MseAM = mean_squared_error(y_true = TestAM['AmarilloAvgTemp'], y_pred =␣
↪PredictionsAM)
print(f"Amarillo MSE Value: {MseAM}")
```

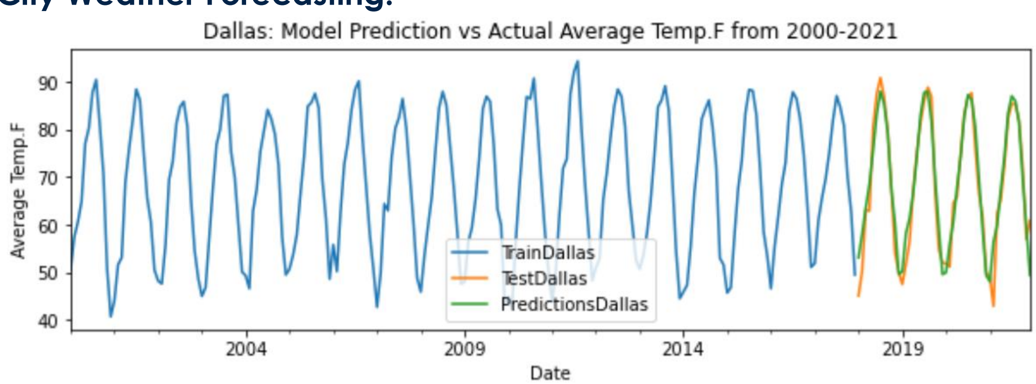
```
Amarillo MSE Value: 21.018619416666596
```

```
# Calculate the RMSE (root mean squared error)
RmseAM = np.sqrt(MseAM)
print(f"Amarillo RMSE Value: {RmseAM}")
```

```
Amarillo RMSE Value: 4.5846067897548854
```




## 3) Dallas City Weather Forecasting:

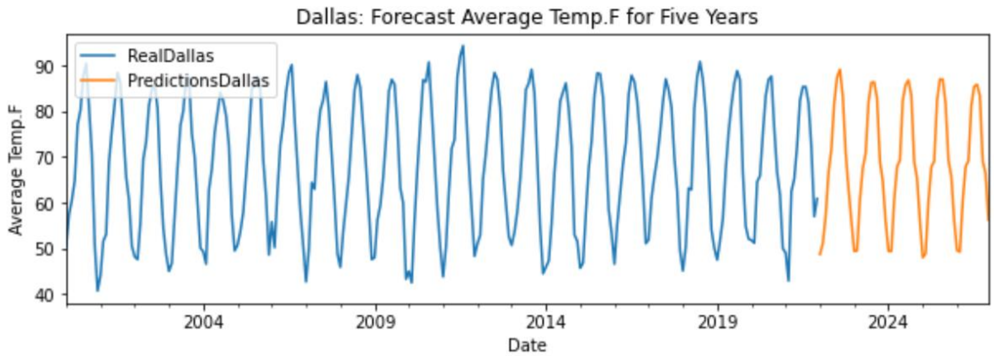


```python
# Calculate the MSE (mean squared error)
MseD = mean_squared_error(y_true = TestD['DallasAvgTemp'], y_pred =␣
 ↪PredictionsD)
print(f"Dallas MSE Value: {MseD}")
```

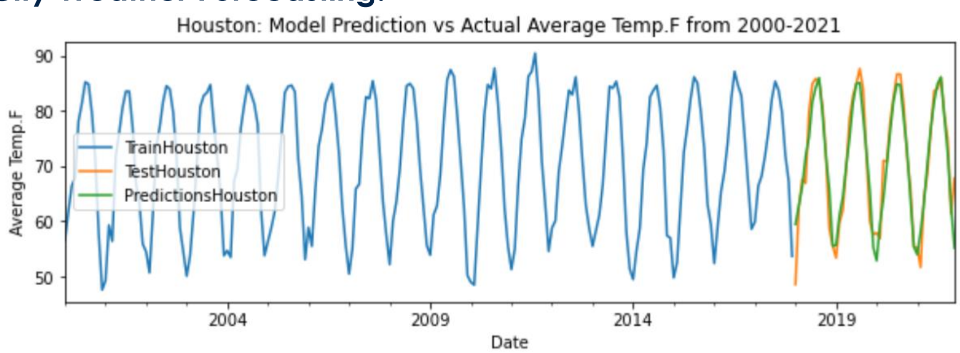
```
Dallas MSE Value: 17.26436185416662
```

```python
# Calculate the RMSE (root mean squared error)
RmseD = np.sqrt(MseD)
print(f"Dallas RMSE Value: {RmseD}")
```

```
Dallas RMSE Value: 4.155040535803065
```




## 4) Houston City Weather Forecasting:

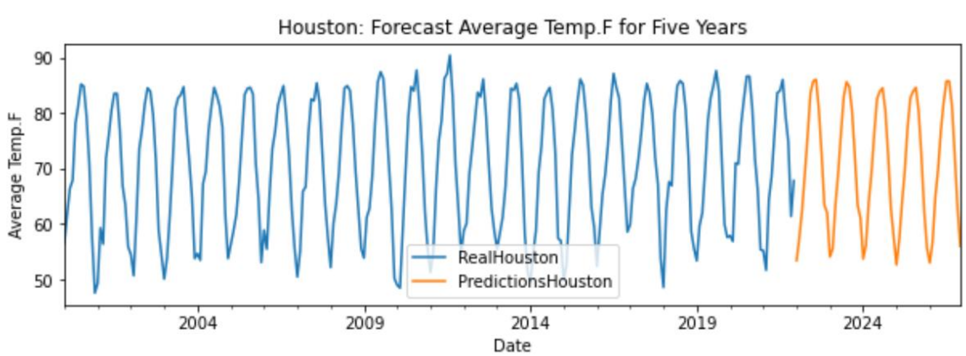



```python
# Calculate the MSE (mean squared error)
MseH = mean_squared_error(y_true = TestH['HoustonAvgTemp'], y_pred =␣
 ↪PredictionsH)
print(f"Houston MSE Value: {MseH}")
```

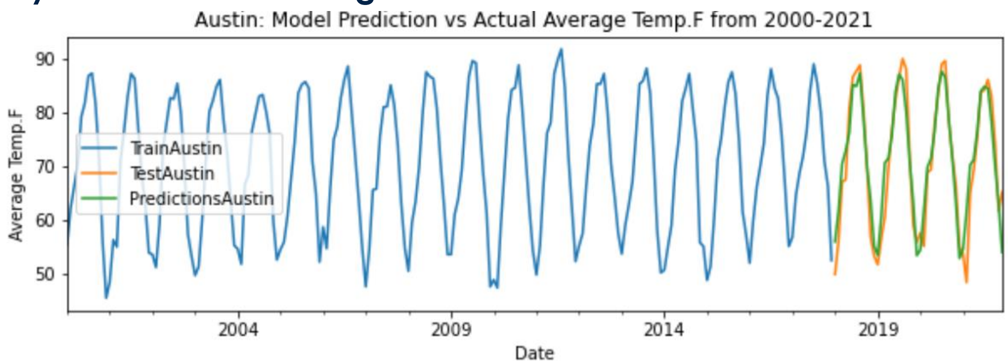
```
Houston MSE Value: 13.476819666666628
```

```python
# Calculate the RMSE (root mean squared error)
RmseH = np.sqrt(MseH)
print(f"Houston RMSE Value: {RmseH}")
```

```
Houston RMSE Value: 3.671078815098721
```



## 5) Austin City Weather Forecasting:

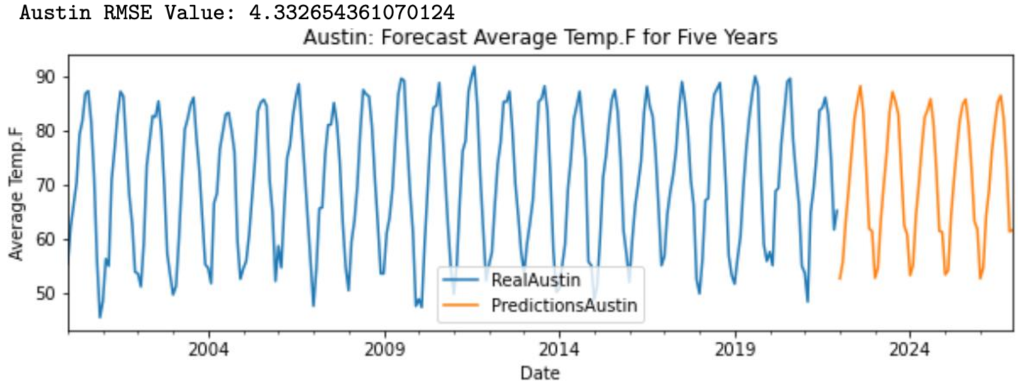



```python
# Calculate the MSE (mean squared error)
MseAU = mean_squared_error(y_true = TestAU['AustinAvgTemp'], y_pred =␣
 ↪PredictionsAU)
print(f"Austin MSE Value: {MseAU}")
```

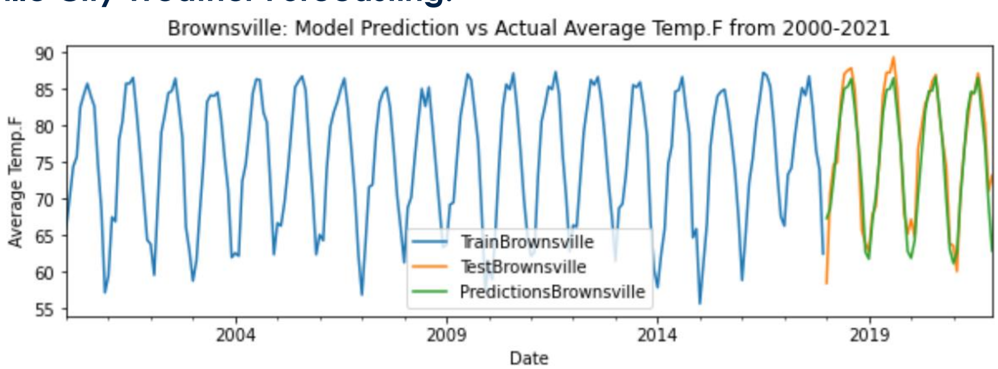
```
Austin MSE Value: 18.771893812499965
```

```python
# Calculate the RMSE (root mean squared error)
RmseAU = np.sqrt(MseAU)
print(f"Austin RMSE Value: {RmseAU}")
```

```
Austin RMSE Value: 4.332654361070124
```




## 6) Brownsville City Weather Forecasting:



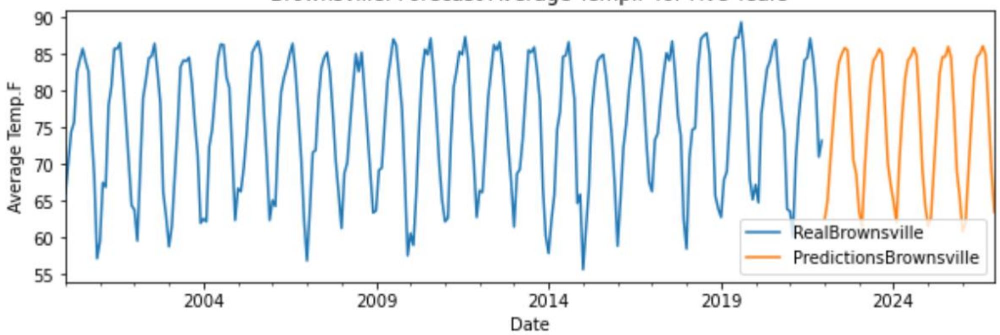

```python
# Calculate the MSE (mean squared error)
MseB = mean_squared_error(y_true = TestB['BrownsvilleAvgTemp'], y_pred =␣
 ↪PredictionsB)
print(f"Brownsville MSE Value: {MseB}")
```

```
Brownsville MSE Value: 9.249766083333373
```

```python
# Calculate the RMSE (root mean squared error)
RmseB = np.sqrt(MseB)
print(f"Brownsville RMSE Value: {RmseB}")
```

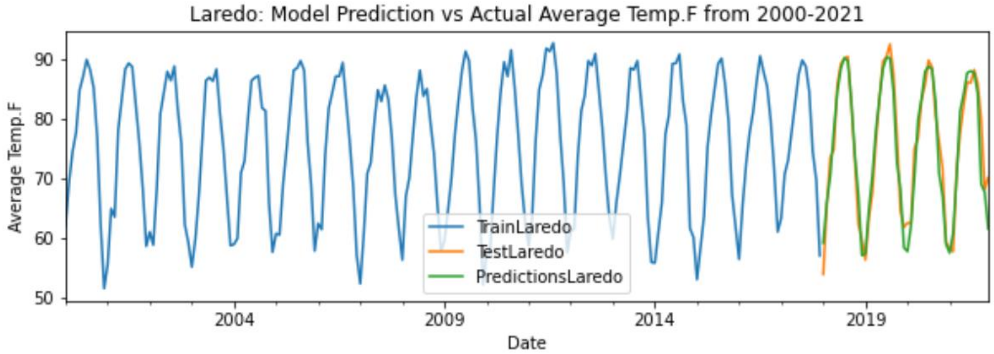
```
Brownsville RMSE Value: 3.04134280924288
```

Brownsville: Forecast Average Temp.F for Five Years




## 7) Laredo City Weather Forecasting:

Laredo: Model Prediction vs Actual Average Temp.F from 2000-2021



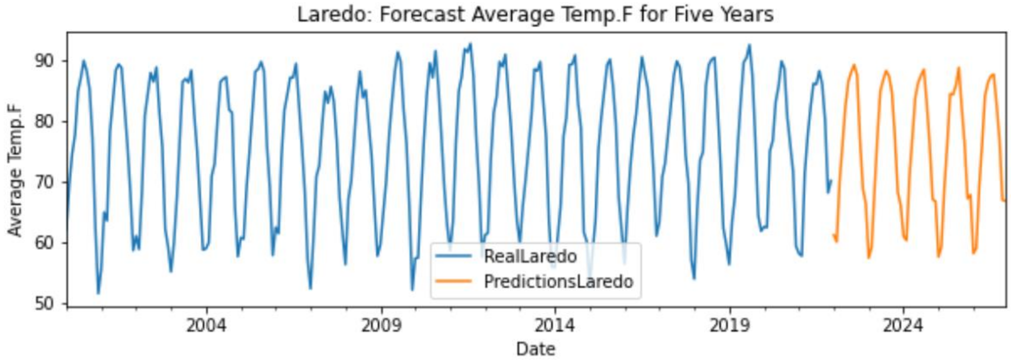

```python
# Calculate the MSE (mean squared error)
MseL = mean_squared_error(y_true = TestL['LaredoAvgTemp'], y_pred =␣
↪PredictionsL)
print(f"Laredo MSE Value: {MseL}")
```

```
Laredo MSE Value: 10.563244208333318
```

```python
# Calculate the RMSE (root mean squared error)
RmseL = np.sqrt(MseL)
print(f"Laredo RMSE Value: {RmseL}")
```

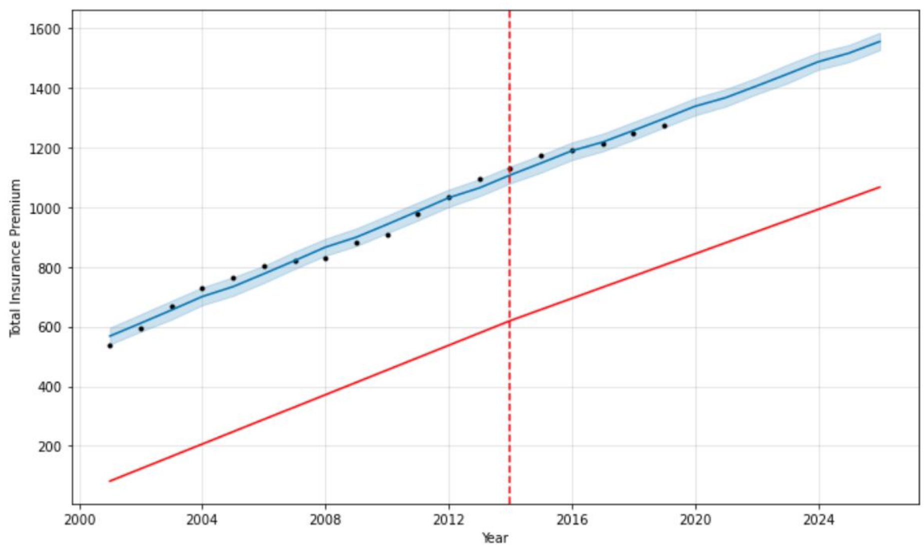
```
Laredo RMSE Value: 3.2501144915730764
```

Laredo: Forecast Average Temp.F for Five Years

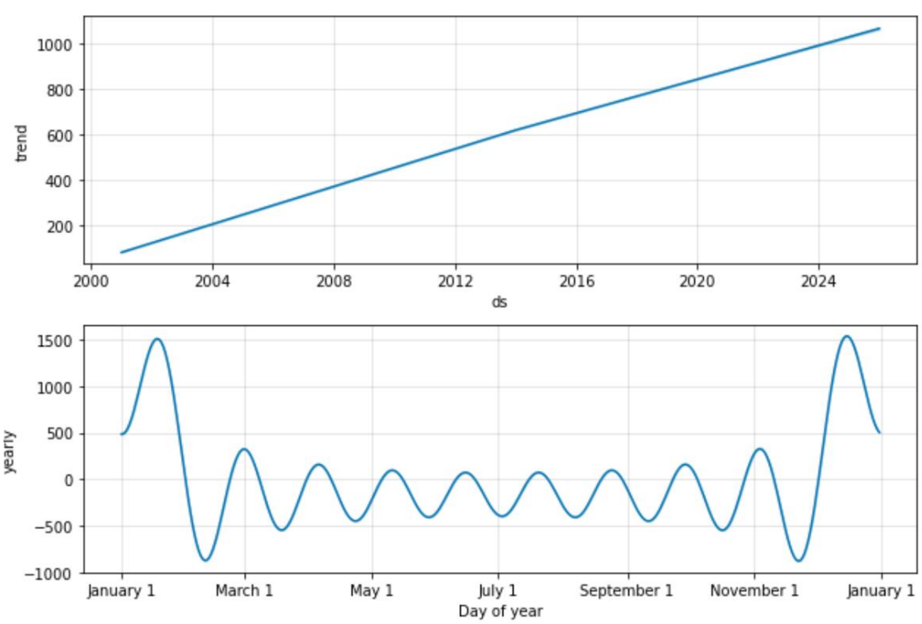


# Recommendation:

Our recommendations are based on the final step of our analysis. We utilized the Texas Homeowners Insurance dataset by manipulating it in preparation for use by the Prophet model. Below we see a projection of steadily increasing insurance rates stemming all the way back from 2000 and continuing into the foreseeable future. This parallels what the red trend line shows as well. There is a change point denoted by the perforated red line indicating a statistically significant shift occurring in 2014. The predicted values (black dots) are all very close to the actual data points (blue line) with limited variance.




We further isolated the trendline visually to emphasize the pattern of increasing rates over time. The third chart shows the variance over a fiscal year. The fluctuations remain consistent for 11 months out of the year but taking into consideration this is financial data, there is a drastic change in January when rates change for the new year.



Our final premium recommendations for InsurTex over the next 5 years are as follows:

| Fiscal Year | Premium Price |
|---|---|
| 2022-2023 | $1,406.10 |
| 2023-2024 | $1,446.09 |
| 2024-2025 | $1,487.14 |
| 2025-2026 | $1,516.30 |
| 2026-2027 | $1,555.24 |

# References:


1) Diaz, J. (2022, January 4). *Texas officials put the final death toll from last year's winter storm at 246*. NPR.Org. Retrieved August 24, 2022, from https://www.npr.org/2022/01/03/1069974416/texas-winter-storm-final-death-toll
2) Doss-Gollin, J. (2021, April 16). *How Unprecedented Was the February 2021 Texas Cold Snap?* State of the Planet. Retrieved August 24, 2022, from https://news.climate.columbia.edu/2021/03/16/unprecedented-texas-cold-snap/
3) National Weather Service. (n.d.). *Hurricane Katrina - August 2005*. Weather.Gov. Retrieved August 24, 2022, from https://www.weather.gov/mob/katrina