

# Using Moon Phases and Weather Data to Predict Temperature

CS506 Fall 2024

By: Esraa Sabr, Julissa Mijares, Lily Beaudion

Presentation Link: <https://youtu.be/vcG--q5QIQw>

## Introduction

This project seeks to examine the relationship between moon phases and fluctuations in weather patterns. By leveraging historical weather data collected from various cities— Boston, MA; Chicago, IL; Los Angeles, CA— the study aims to determine if moon phases can serve as predictors for temperature. The primary objective is to develop an accurate predictive model that integrates moon phases with other meteorological variables such as precipitation and wind speed.

## Data Cleaning and Processing

The data cleaning process involved several key steps to prepare the datasets for analysis. Weather data was sourced from the Open-Meteo API, covering the period from October 29, 2019, to October 29, 2024. This dataset includes daily minimum and maximum temperatures (in °F), total precipitation (in mm), and maximum wind speed (in mph) for three cities: Los Angeles, Chicago, and Boston. Simultaneously, moon phase data was obtained from the USNO Astronomical Application API, spanning the same timeframe and detailing the different moon phases (New Moon, First Quarter, Full Moon, Last Quarter), along with their corresponding dates and times in Universal Time (UT). To create a comprehensive dataset, we combined the moon phase data with the weather data for each city based on matching dates. Additionally, we computed and added two new columns—range and mean temperatures (in °F)—to enhance the dataset for further analysis.

## Analyzing Various Cities - Boston, Chicago, Los Angeles

For each city, we used the moon phases, precipitation, and wind speed data to predict mean temperature. We tested three different models: linear regression, decision tree regression, and random forest regression.

We used the linear regression model as a baseline to get a basic understanding of the relationship between the features and the mean temperature. However, we found that the linear regression model was too simple to showcase the relationship and showed an overly flat trend.

The next model we used was the decision tree regression model. It was an improvement compared to the linear regression model, but had a lot of variability and did not show any significant insights.

The final model used was the random forest model. This was a good medium between the linear regression and decision tree models, as it improved the prediction accuracy. The estimated fit line showed a more positive trend closer to the perfect fit line. There is still high variability, yet was an improvement from the previous model.

## Final Model (at Midpoint)

When testing the potential of moon phases as predictors of weather across Los Angeles, Boston, and Chicago, Decision Tree and Random Forest models showed some promising results, though improvements are needed for accuracy. Across all three cities, Decision Tree and Random Forest models outperformed Linear Regression, especially in capturing more complex, nonlinear relationships that may exist between moon phases and weather patterns. Despite their limitations, these models suggest there is potential for further exploration especially with ensemble approaches.

In Los Angeles, Random Forest demonstrated the most promise, achieving a test R-squared of 0.185 and a lower test mean squared error (MSE) of 63.76 compared to other models. This slight increase in predictive power over Linear Regression (test R-squared of 0.072) suggests that Random Forest could be more effective at capturing subtle variations potentially associated with moon phases. Similarly, Decision Tree achieved a high training R-squared of 0.588 in Los Angeles, although its test performance (-0.018) indicates some degree of overfitting. These mixed results imply that moon phases may still hold some predictive relevance but require more refined modeling to harness effectively.

In Boston and Chicago, both Decision Tree and Random Forest models showed high training R-squared values but faced challenges in generalization, particularly in Boston, where the Decision Tree model achieved a training R-squared of 0.683 but negative test R-squared (-0.769). In Chicago, Random Forest once again outperformed Linear Regression, though with only marginal gains, achieving a training R-squared of 0.599, suggesting it captures slightly more variance related to weather patterns than linear methods.

To improve both accuracy and the importance of the moon phase feature in predictions, further steps could include testing advanced models like gradient boosting, combining models into ensemble approaches, or adding more context-relevant features to enrich the moon phase data. By combining models or leveraging more sophisticated machine learning techniques, we hope to refine these preliminary results and explore the potential hidden relationships between moon phases and weather outcomes across diverse climates.

## Future Steps

As we reach the midpoint of our project, we have identified several key steps to enhance our model before the final due date. First, we plan to explore ensemble methods and other modeling techniques to improve both the accuracy of our predictions and the feature importance of the moon phases. Additionally, we hope to improve our dataset by adding or creating new features, seeking out supplementary weather information that could provide valuable insights without detracting from the focus on moon phases. Finally, we intend to conduct hyperparameter tuning to refine our model further, testing various hyperparameters through methods such as Grid Search or manual adjustments. By pursuing these strategies, we hope to develop a more robust and reliable predictive model.

Chart 1:

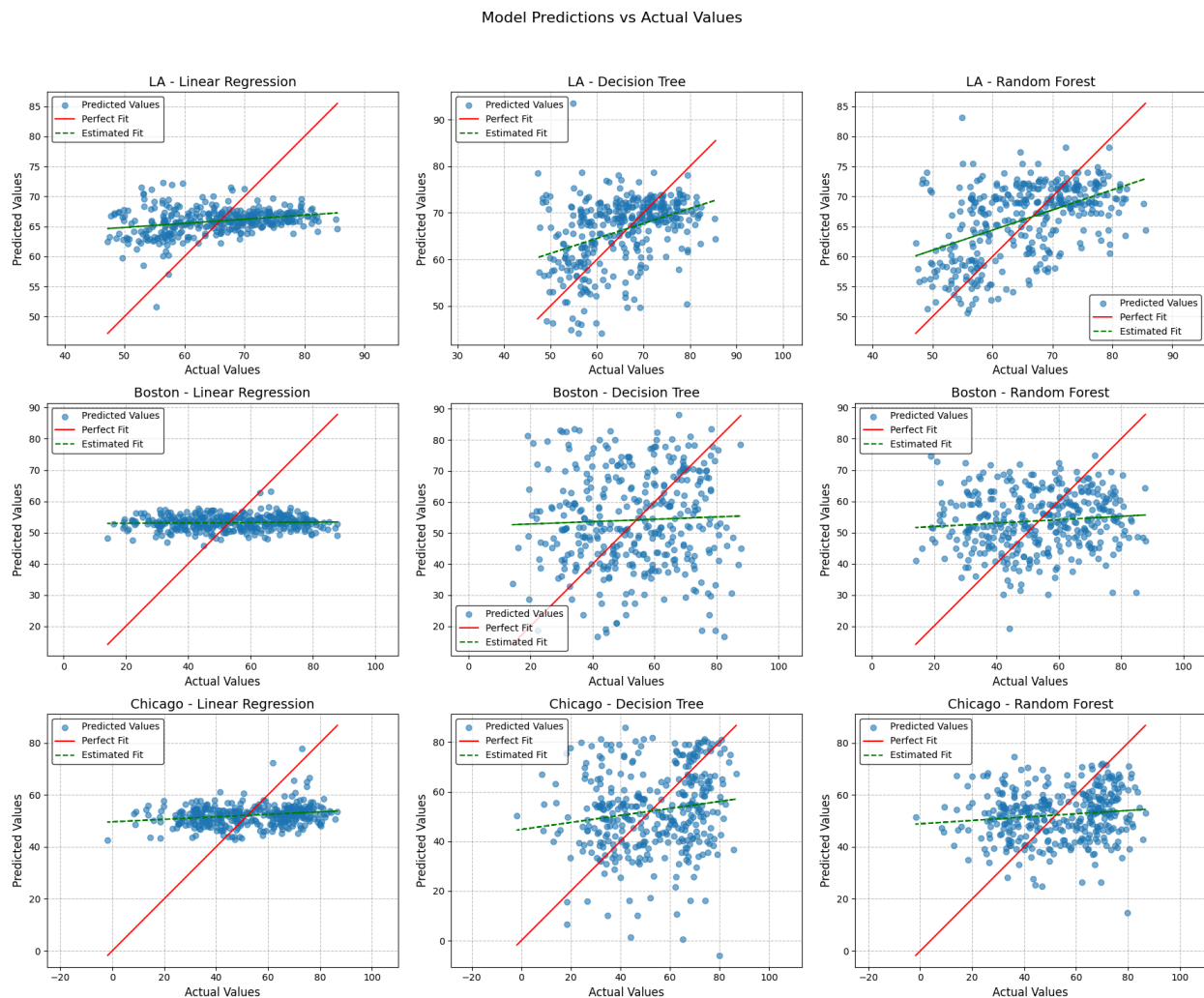


Chart 2:

