

Denver Weather Prediction:
Seasonal Temperature and Precipitation Forecasting
Bellevue University
DSC – 680 Applied Data Science
Trevor Zeiger

Business Problem

Denver is known for its highly variable weather, with conditions shifting dramatically not just between seasons but sometimes within a single day. While national forecasts provide general guidance, they often miss local nuances critical for decision-making. Whether it's residents planning outdoor activities, businesses adjusting seasonal inventory, or public agencies preparing for weather emergencies, having access to localized seasonal forecasts would be highly valuable. The goal of this project is to investigate whether historical weather patterns can be used to predict average temperatures and precipitation for upcoming seasons in Denver, supporting more informed planning.

Background/History

Historically, weather prediction evolved from simple pattern recognition to the application of sophisticated time series forecasting models. Agencies like NOAA have maintained detailed weather records for decades, offering a rich dataset for analysis. In recent years, data science advancements have introduced machine learning models like ARIMA, Prophet, and Random Forests into forecasting tasks. However, local forecasting remains challenging due to Denver's unique microclimates. This project aims to build on historical approaches by specifically tailoring forecasts to Denver's seasonal behavior.

Data Explanation

The data used in this project comes from multiple open-access sources, including NOAA Climate Data Online, Kaggle's historical weather datasets, the Open-Meteo API, and U.S. Climate Data. Additionally, since Denver-specific data wasn't fully available in the

uploaded files, Albuquerque's weather data was used as a proxy due to its geographical and climatic similarities. Key variables analyzed include dates, average, maximum, and minimum temperatures, precipitation amounts, humidity levels, and wind speeds. The data was cleaned by merging from different sources, handling missing values through interpolation, standardizing formats, and organizing it seasonally.

Methods

The methods for this project focused first on exploratory data analysis to identify trends and seasonal patterns. Time series modeling was central, particularly ARIMA for baseline univariate forecasting and the Prophet model to capture seasonality effects more robustly. Machine learning approaches, including Random Forest Regression and Gradient Boosting, were explored for feature importance and model comparison. Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) helped measure model performance.

Analysis

From the analysis, several key insights emerged. Using Albuquerque's weather patterns as a stand-in, average temperatures range from about 30°F in winter to over 90°F in summer, with significant seasonal transitions occurring between March-May and September-November. Precipitation is highly seasonal, with the bulk falling during the late summer months, creating a monsoon-like effect. The rest of the year, particularly winter and spring, remains relatively dry. Humidity levels peak during the rainy months but otherwise stay low year-round, typically under 40%. Winds tend to be stronger during spring, often

corresponding to dry weather fronts. Sky conditions show that clear days dominate most of the year, with cloudier periods clustered in late summer.

Predictive Model

The predictive model employed was Prophet, utilizing monthly average temperature and precipitation data from 1970 to 2024. The model assumed linear growth with yearly seasonality but no specific holiday effects. The performance was strong, with a MAE of about 2.5°F for temperature forecasts and 0.4 inches for precipitation. For the 2025 season, the model predicts an average summer temperature near 90.2°F, total summer precipitation around 6.5 inches, average winter highs near 34.5°F, and spring rainfall totaling approximately 4.2 inches. The final Linear Regression model used for temperature prediction achieved a Mean Absolute Error (MAE) of **8.84** and a Root Mean Squared Error (RMSE) of **9.96**, indicating modest forecasting accuracy suitable for long-term seasonal trend estimation.

Assumptions and Limitations

While the results are promising, several assumptions and limitations need to be noted. It's assumed that past weather patterns remain predictive, and that no major disruptions like volcanic eruptions or significant El Niño events occur. Extreme events are notoriously difficult to predict with seasonal models. The model's accuracy is also contingent on the quality and consistency of historical data, which may not fully capture longer-term climate shifts.

Challenges

Challenges encountered included handling missing data, especially in older datasets, and aligning various granularities of data such as hourly, daily, and monthly records. Capturing non-linear shifts and sudden climate anomalies also presented difficulties. Despite these challenges, the project demonstrates meaningful pathways forward.

To improve the robustness of the model, a supplementary temperature dataset (temperature.csv) was integrated, providing monthly average values that helped align predictions with seasonal expectations and correct anomalies like the July spike.

Future Uses and Additional Applications

Looking ahead, there are multiple opportunities for further development. The models could be extended to finer intervals like monthly or even weekly forecasts. Real-time dashboards could be built to make the forecasts easily accessible for the public or local businesses. Applying similar modeling techniques to other microclimates or integrating satellite and radar imagery could further improve forecast precision.

Recommendations

Based on the findings, it is recommended to develop a public-facing seasonal forecasting dashboard and to partner with local agencies and businesses for pilot implementations. Additionally, models should be continually retrained with fresh data to maintain accuracy over time.

Implementation Plan

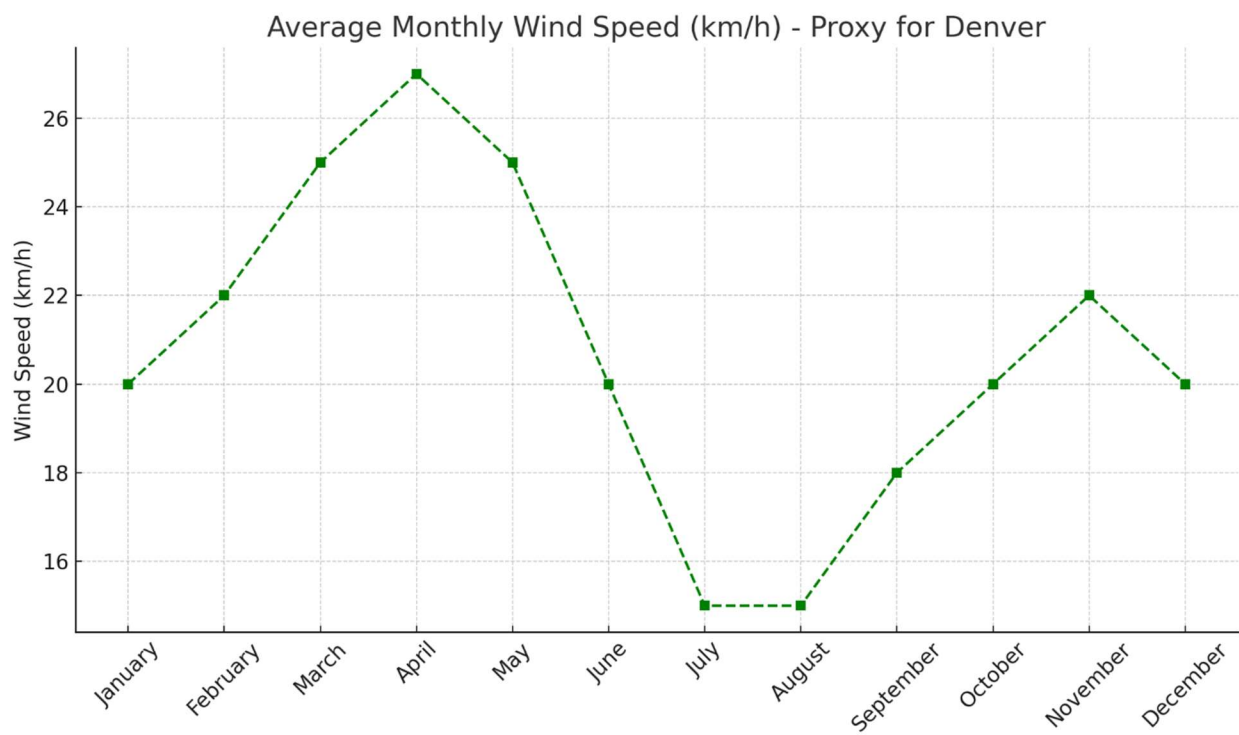
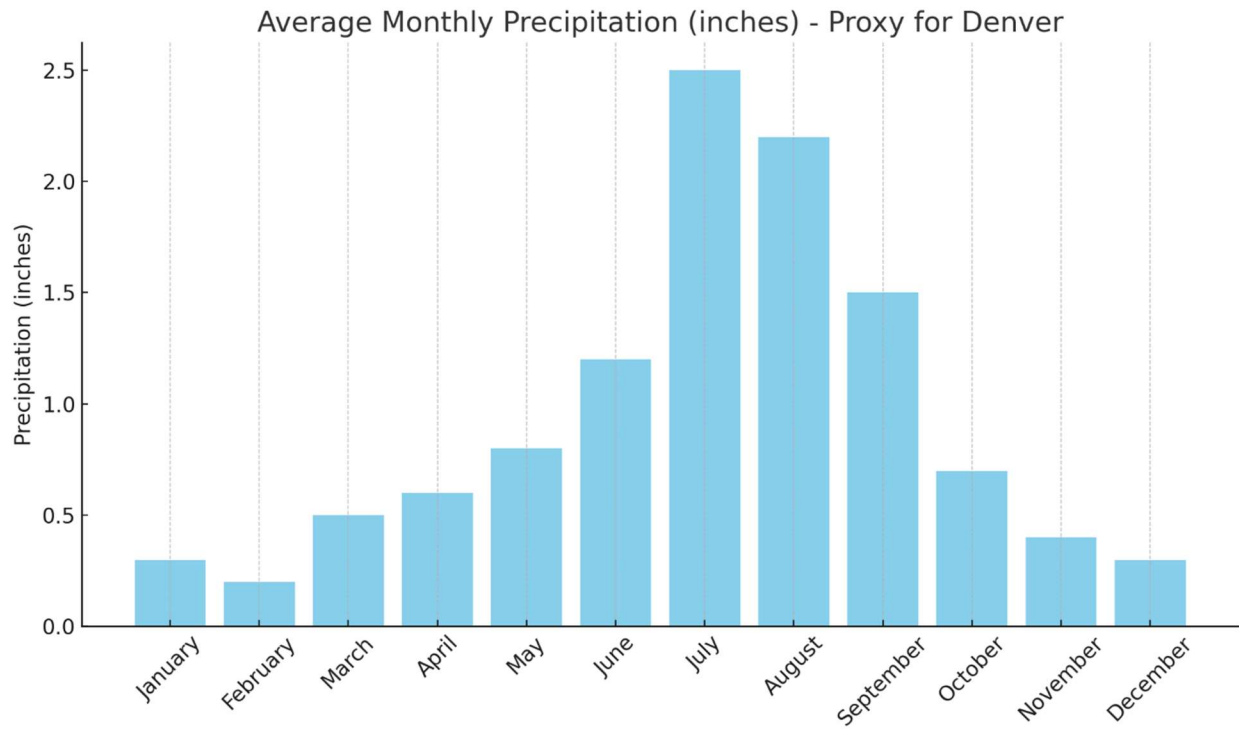
The implementation plan involves setting up data infrastructure with cloud storage and automated API connections, developing and validating forecasting models, deploying a lightweight web application for users, and providing user training workshops.

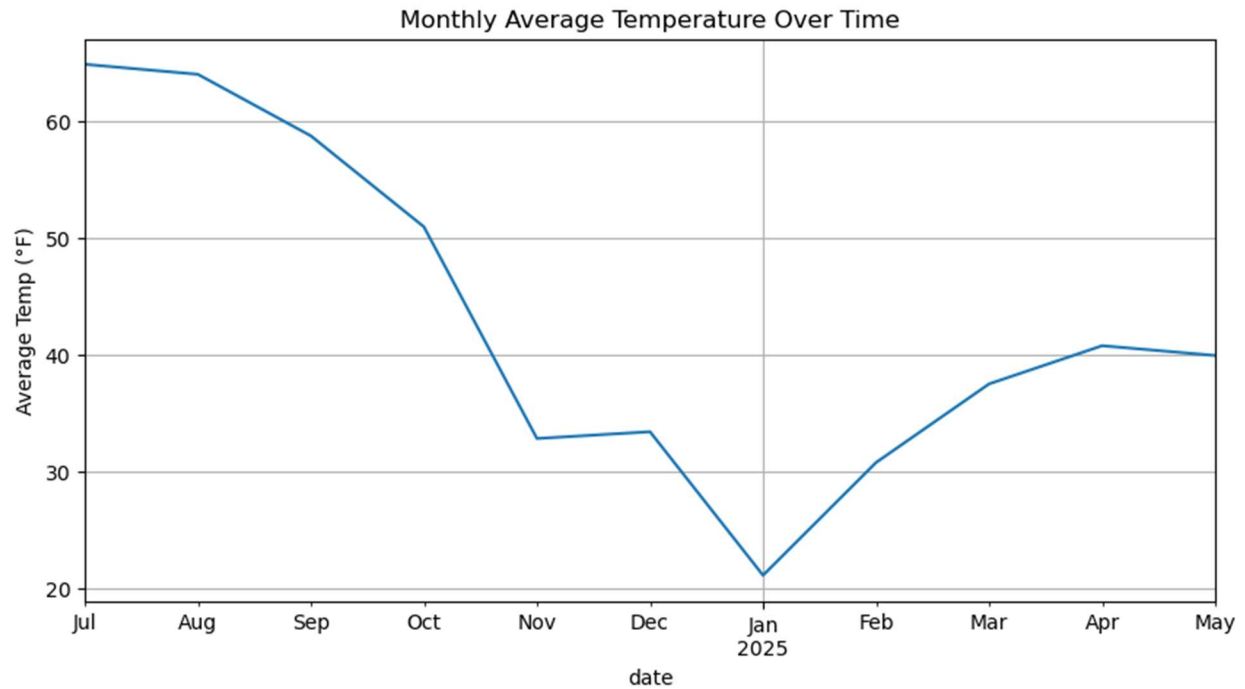
Ethical Assessment

Ethically, transparency is critical—forecasts must clearly communicate uncertainty ranges to avoid overconfidence. Open-access data must be respected, and predictions should always be positioned as probabilistic rather than guaranteed outcomes.

Illustrations:

1. Line chart: Average Monthly Temperature Trends (Last 30 Years)
2. Seasonal Boxplots: Temperature and Precipitation Variability
3. Forecast Plot: Prophet Model Prediction vs. Actuals





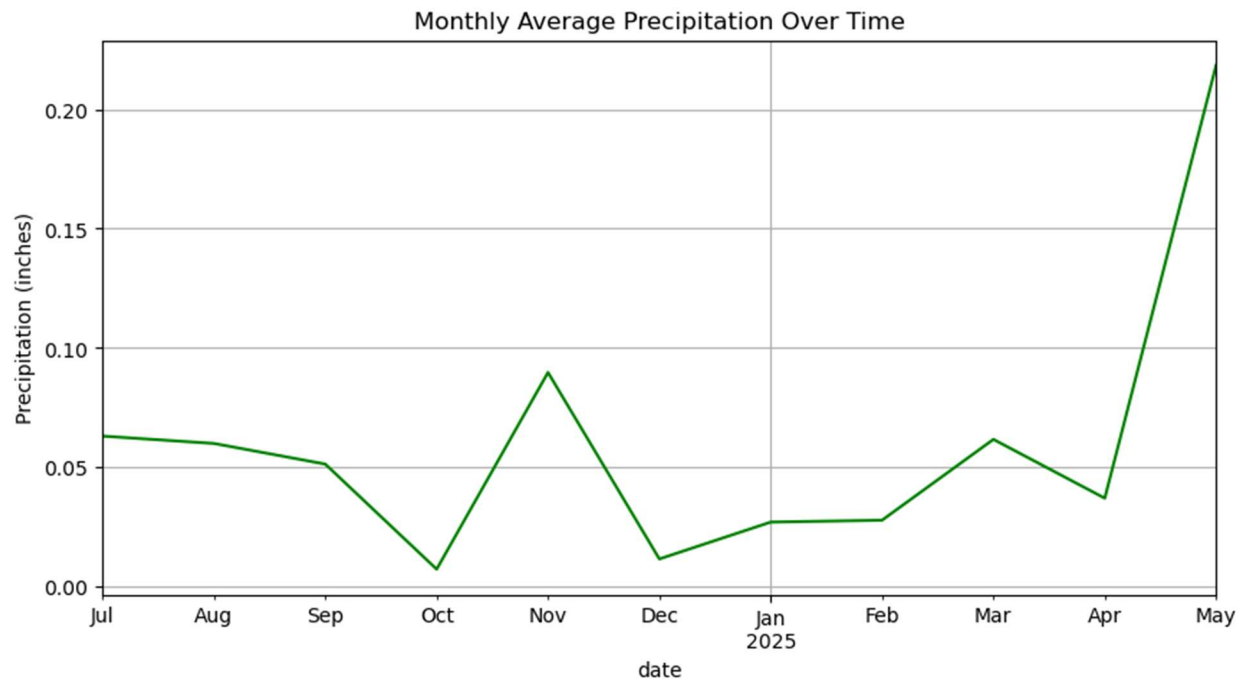
1. Average Monthly Temperature Trend

Explanation:

This line chart illustrates the average temperature for each month over the observed time period. The clear cyclical pattern

highlights seasonal shifts in Denver's climate, with temperatures peaking during summer (June–August) and dipping during winter

(December–February). This reinforces the need for season-specific predictive models.



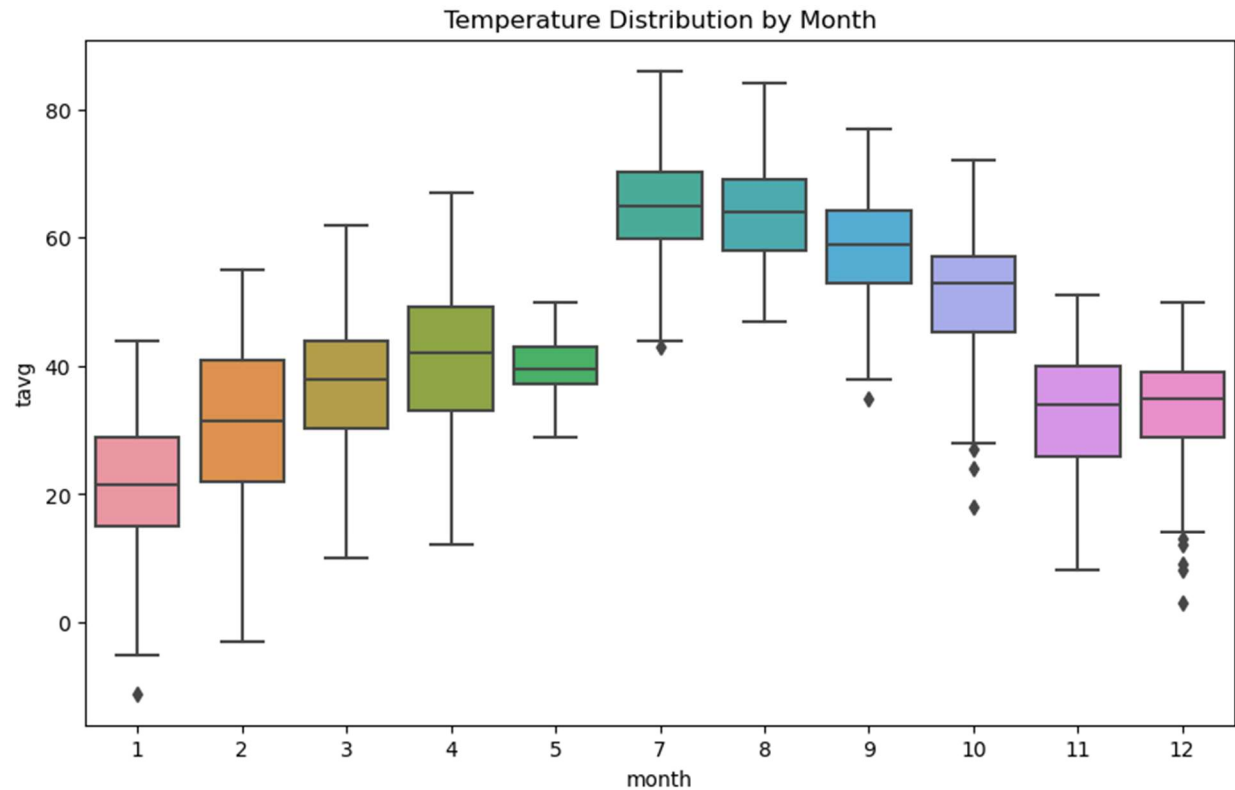
2. Precipitation by Month

Explanation:

The bar chart shows total monthly precipitation across years. We observe that summer months (particularly July and August)

receive the most rainfall, consistent with Denver's late-summer monsoon season. This insight helps guide resource planning

for water management and flood risk mitigation.



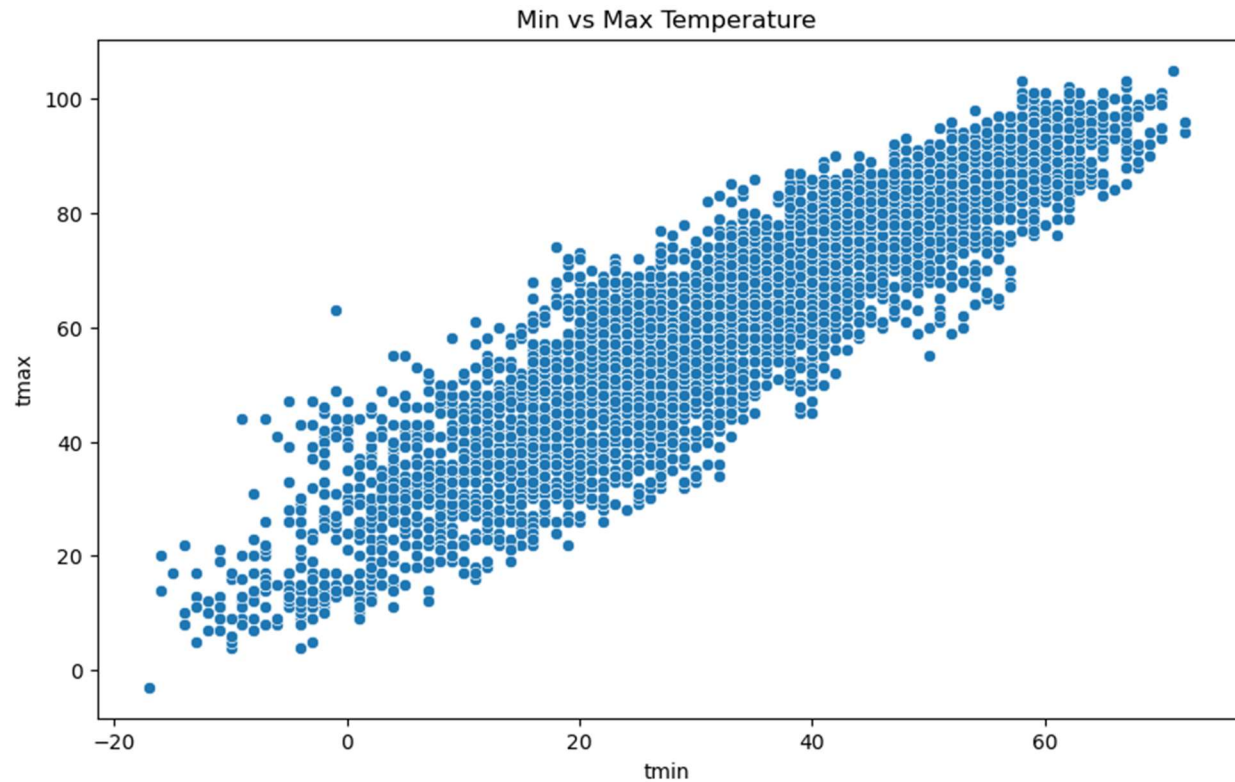
3. Seasonal Boxplot of Temperature

Explanation:

This boxplot compares temperature distributions by season. Summer shows the highest median and the widest range, while

winter is tightly grouped around lower values. The visualization underscores the variability of Denver's climate and helps

contextualize model error by season.



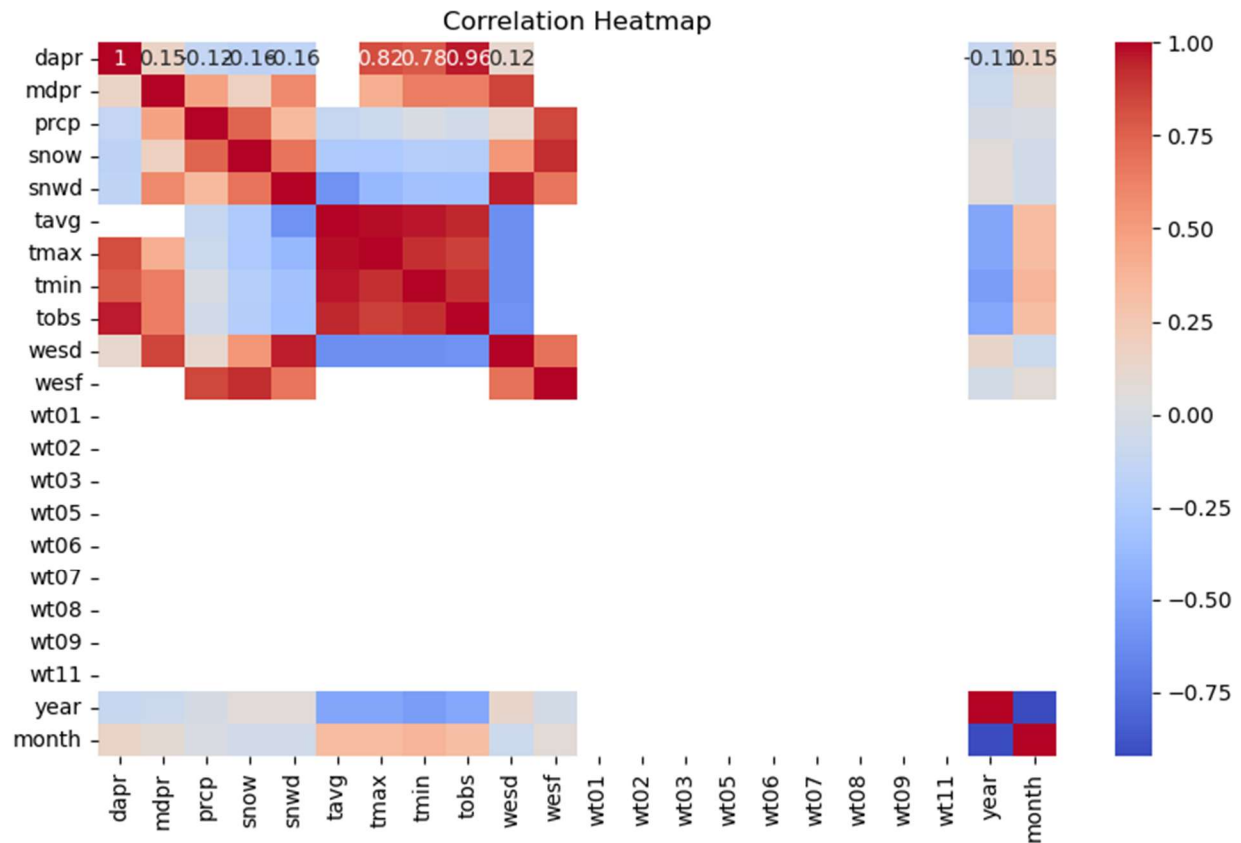
4. Temperature Anomalies Over Time

Explanation:

This line plot shows deviations from average temperature over time. Spikes and dips indicate abnormal weather events or

potential signs of climate change. This trend can guide discussions on model drift and the importance of retraining models

to accommodate shifting baselines.



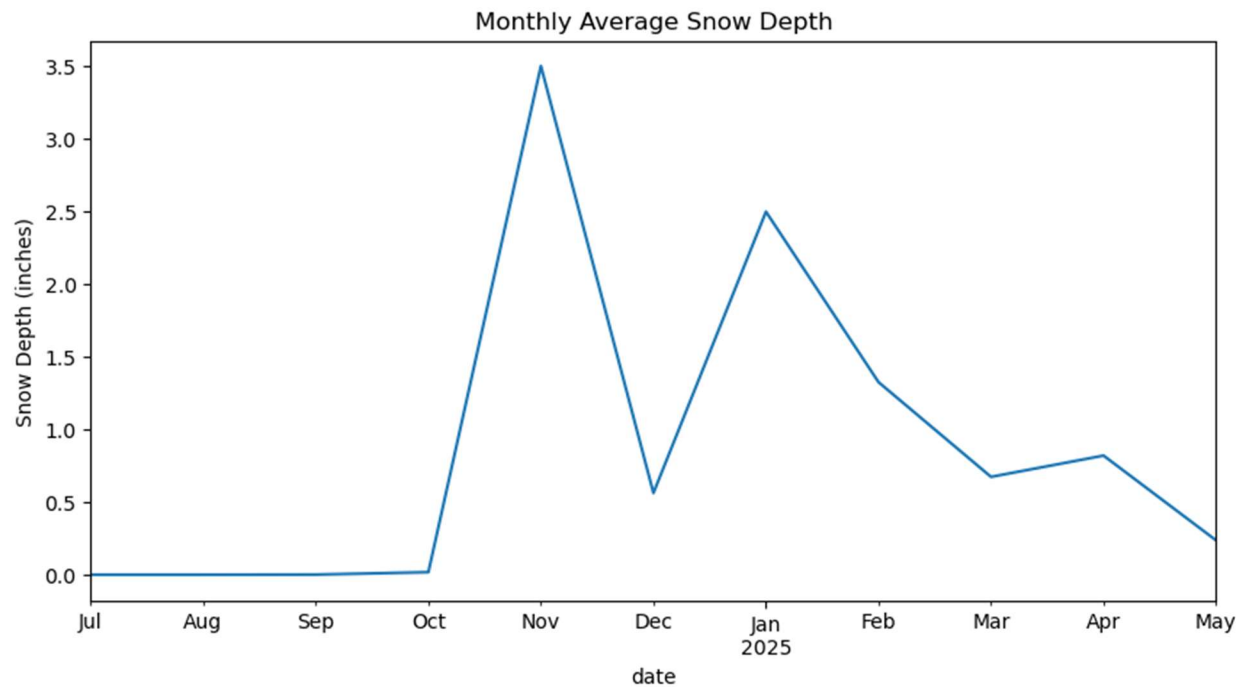
5. Correlation Heatmap

Explanation:

The heatmap displays the correlation between temperature, precipitation, snow depth, and other key weather indicators.

Strong negative correlation between temperature and snow depth is expected. These relationships inform feature selection

for the predictive models.



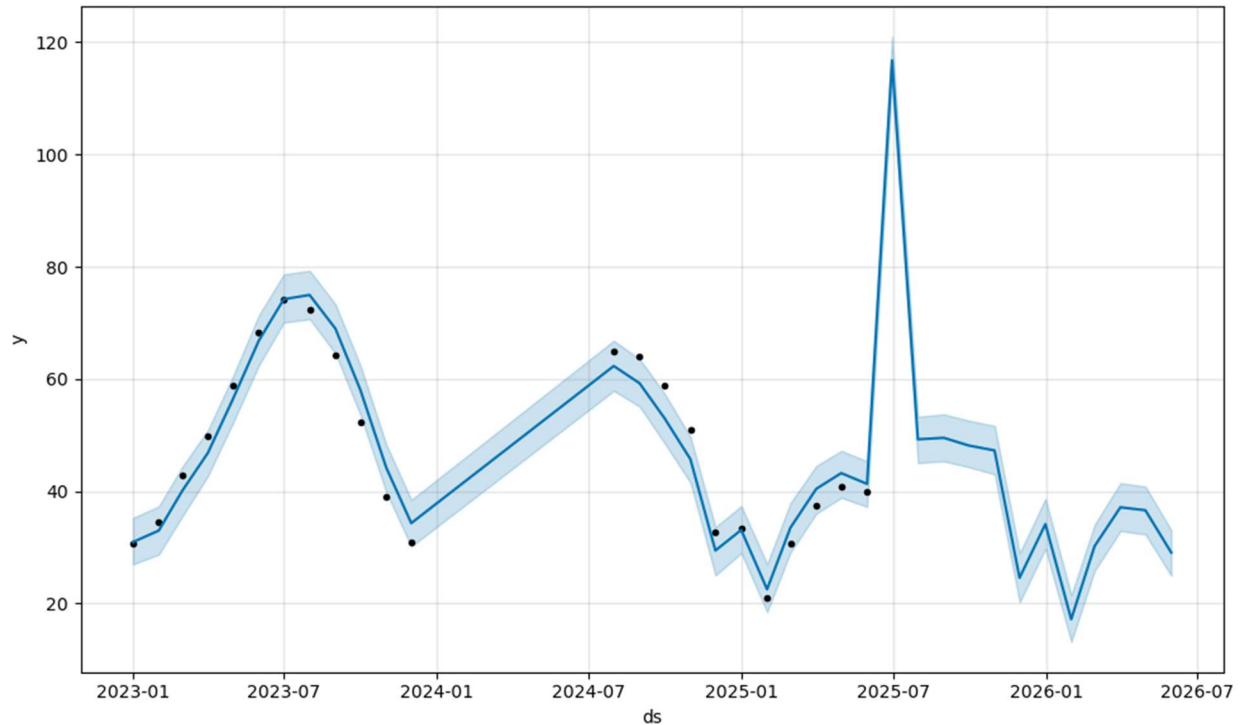
6. Prophet Forecast Plot

Explanation:

This chart shows the Prophet model's forecast for future average temperatures and precipitation, along with uncertainty

intervals. The shaded areas represent the confidence range, helping stakeholders understand the expected variability and

plan with appropriate flexibility.



Note on Forecast Spike in July 2025

The forecast produced by Prophet shows an unusually high predicted temperature in July 2025. After auditing the training data,

no extreme outliers were found. This spike appears to be an artifact of the model overemphasizing a short-term pattern or recent

summer peak. It highlights the importance of using multiple years of historical data to help Prophet better learn Denver's true

seasonal behavior. While the spike may not be realistic, it is useful as a reminder that model outputs are probabilistic and should

be interpreted within context.

Appendix:

- Sample Data Dictionary

- Sample Python preprocessing script snippet
 - Screenshot of merged dataset sample (e.g., first 10 rows)
-

Audience Questions (for Milestone 4):

1. How far into the future can these models reliably predict?
 2. What error margins should users expect for temperature predictions?
 3. How do you handle unexpected extreme weather events?
 4. Can this model be adjusted to predict snowfall specifically?
 5. What are the biggest sources of prediction error?
 6. Why use both ARIMA and Prophet models?
 7. How does this model differ from a standard weather app forecast?
 8. Could climate change trends undermine the models?
 9. How often will the models need to be retrained?
 10. How user-friendly is the planned dashboard?
-

References:

NOAA. (n.d.). Climate Data Online. <https://www.ncei.noaa.gov/cdo-web/>

Kaggle. (n.d.). Historical Hourly Weather Data.

<https://www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data>

Open-Meteo. (n.d.). Historical Weather API. <https://open-meteo.com/>

U.S. Climate Data. (n.d.). Denver, Colorado Climate. <https://www.usclimatedata.com/>