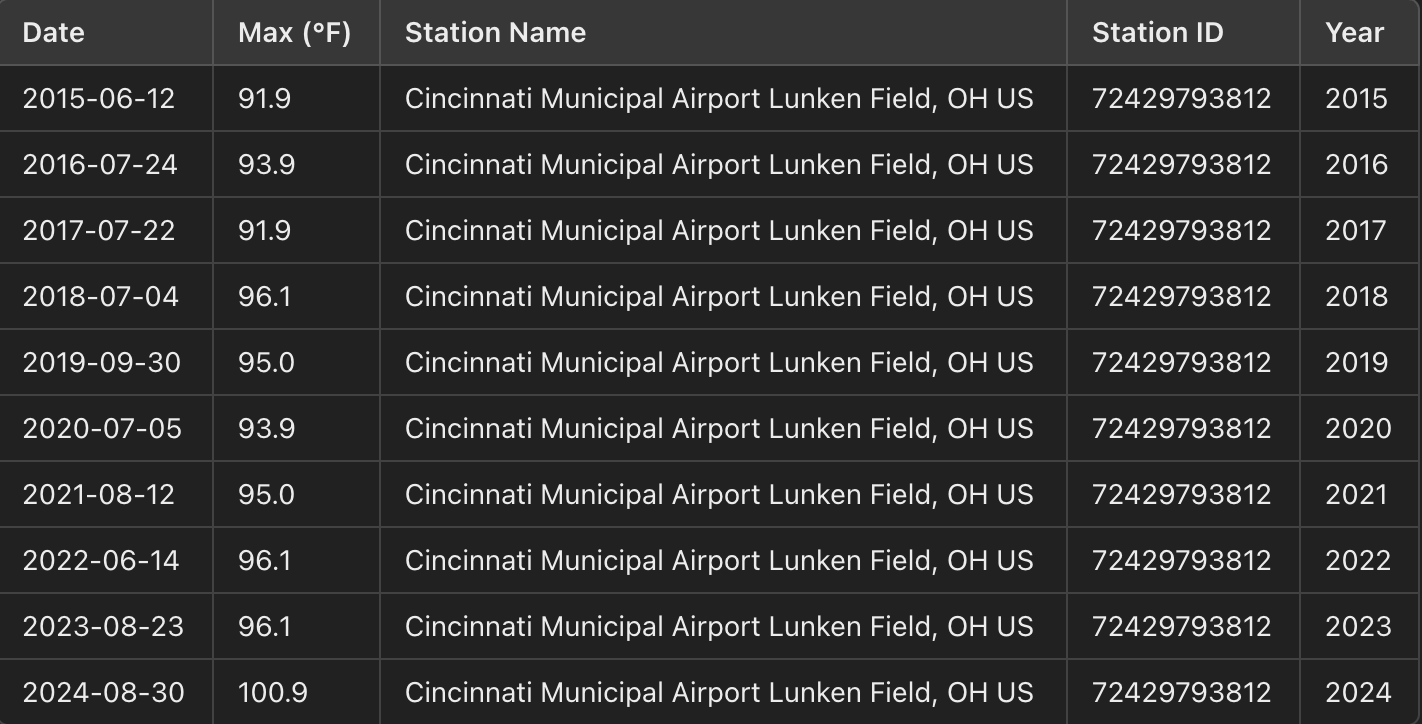
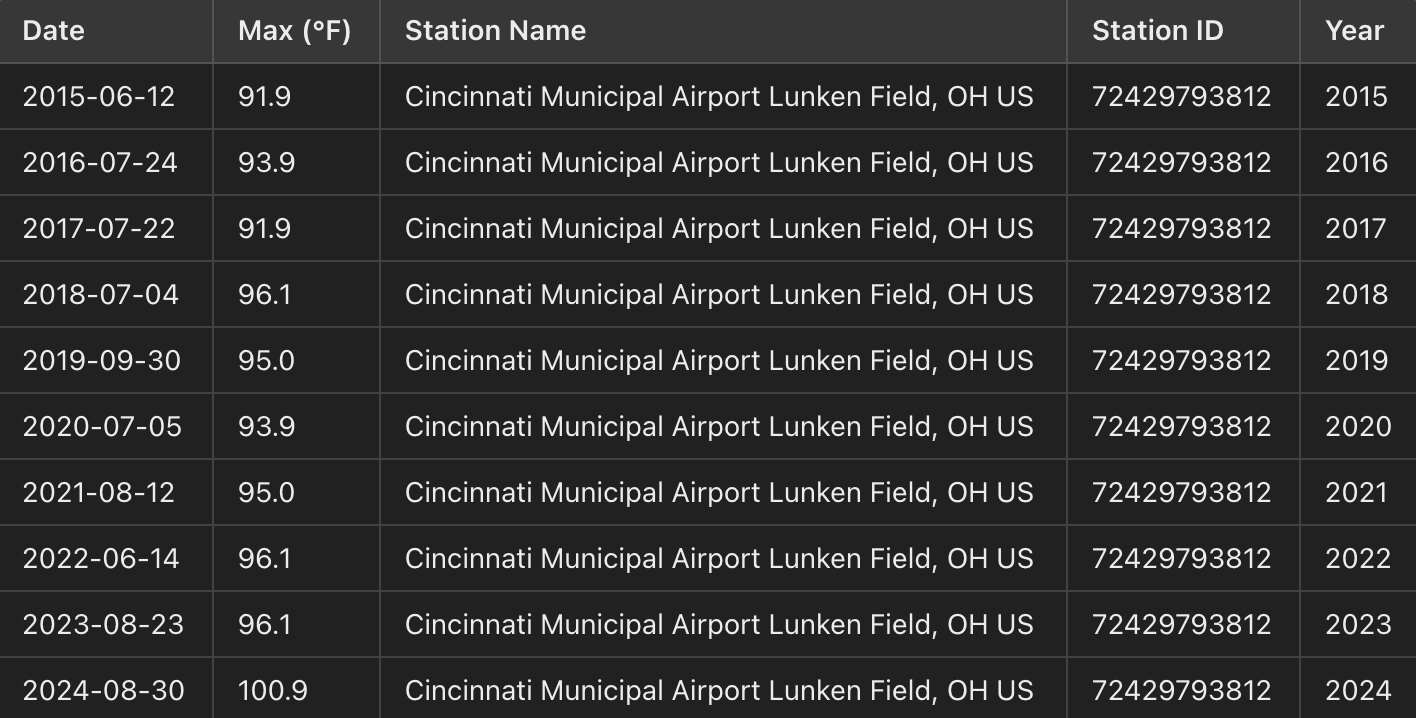
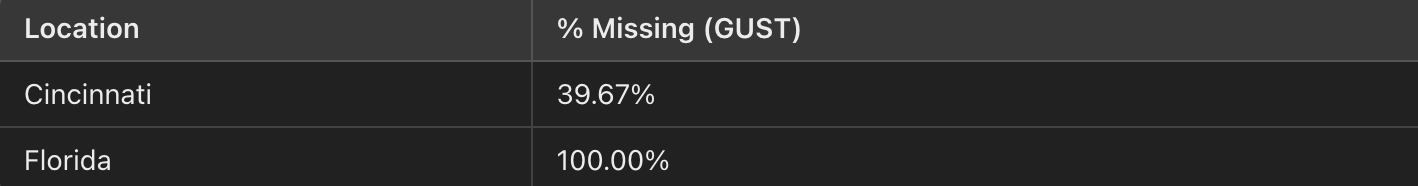
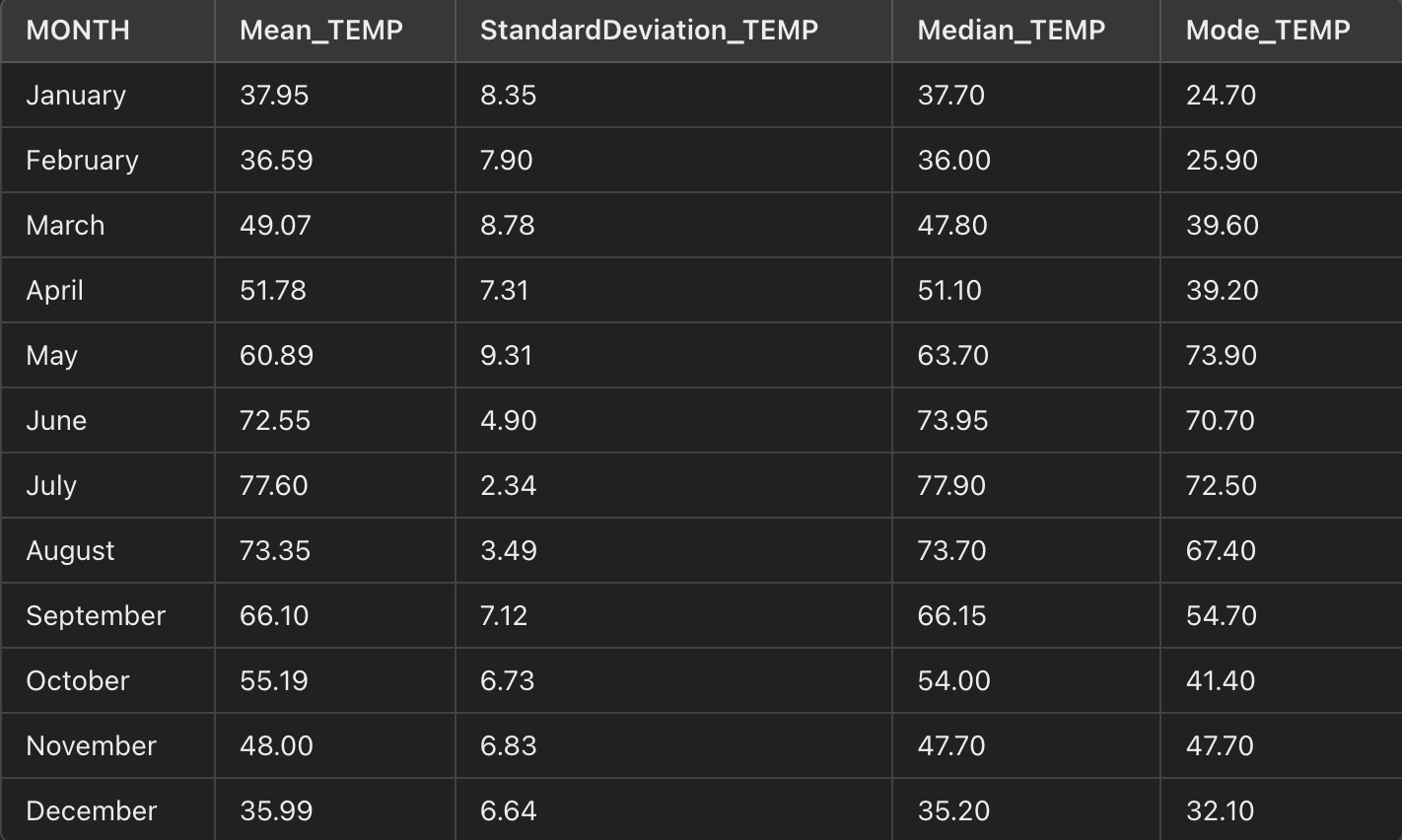
# **Cincinnati and Florida Temperature Analysis Report**

1. Hottest Days by Year (Cincinnati)
2. Hottest Days by Year (Florida)
3. Overall Hottest Day by Year (Cincinnati and Florida)

## Coldest Day Overall in March (2015-2024) across Cincinnati and Florida

## Year with Most Precipitation

1. Missing Values for Wind Gust in 2024
2. Monthly Temperature Statistics for Cincinnati in 2020
3. Top 10 Days with Lowest Wind Chill for Cincinnati in 2017

## Number of Days with Extreme Weather Conditions in Florida (2015–2024)

## 

## Model Evaluation Metrics

### **Linear Regression**

* **November Metrics**:
  + RMSE = 0.00
  + MAE = 0.00
  + R² = 1.00
* **December Metrics**:
  + RMSE = 0.00
  + MAE = 0.00
  + R² = 1.00

### **Random Forest**

* **November Metrics**:
  + RMSE = 8.29
  + MAE = 8.22
  + R² = -14.57
* **December Metrics**:
  + RMSE = 6.32
  + MAE = 6.30
  + R² = -38.94

**Areas of Improvements:**

To enhance the model, we could incorporate a longer span of historical data, ideally 10–20 years, instead of just two. This extended dataset would provide a clearer view of long-term trends and annual fluctuations, resulting in more reliable predictions. Additionally, implementing more advanced models—such as those that account for seasonal patterns or more complex trends like polynomial or time-series models—could improve accuracy by capturing temperature variations over time. Including other influential factors, such as climate events (e.g., El Niño or La Niña), greenhouse gas concentrations, or sea temperatures, would further enhance the model's accuracy by considering the underlying drivers of temperature changes. These improvements would help the model produce a more realistic forecast that aligns with real-world climate patterns affecting Cincinnati’s weather.

**Model Performance**

With only two years of data, the model may lack the robustness needed to capture broader patterns or seasonal fluctuations that occur over longer periods. Relying on a linear trend can lead to overly simplified predictions, as climate and weather patterns often exhibit nonlinear and cyclical variations. For instance, temperature changes may be influenced by unique yearly events, seasonal climate anomalies, or broader trends that evolve over decades, such as shifts due to climate change.

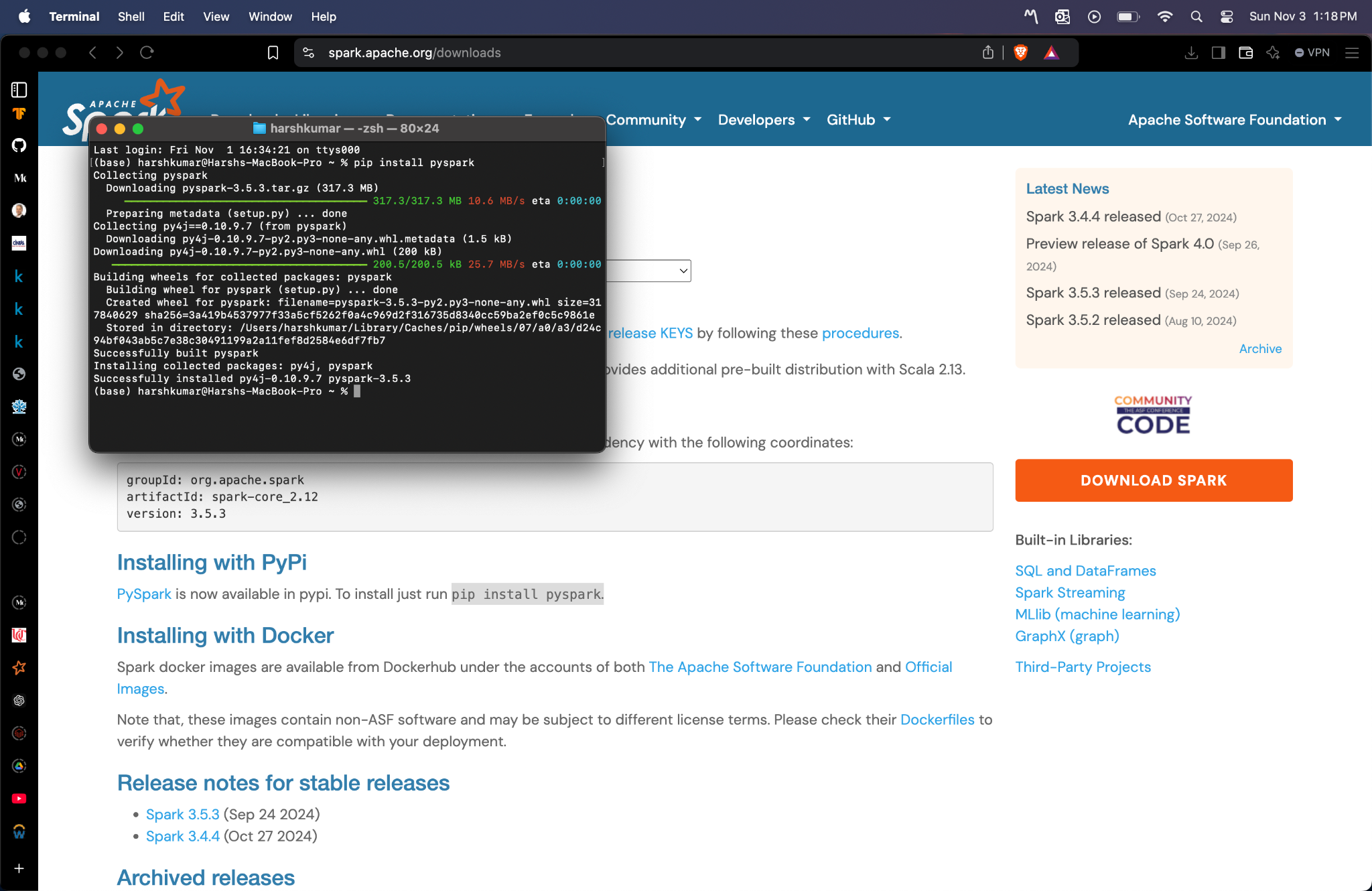
This model’s performance, while useful for short-term insights, could benefit from the inclusion of additional years of data, capturing a wider range of temperature trends and outliers. Expanding the dataset would allow the model to better reflect the natural variability in Cincinnati's weather, improving the accuracy and reliability of future predictions.

**Conclusion**

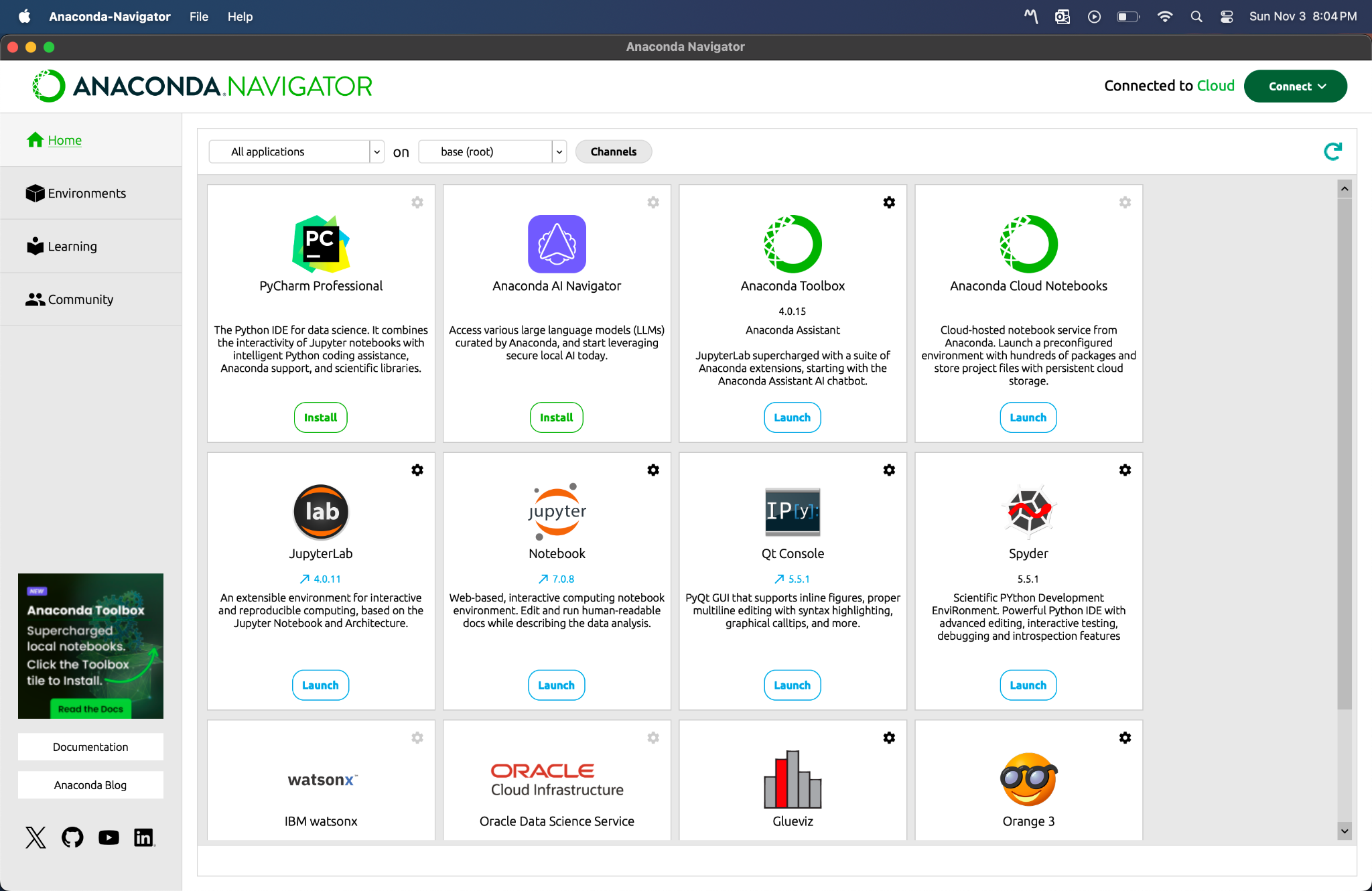
The linear regression model applied here offers a straightforward method for forecasting maximum temperatures in Cincinnati for November and December 2024, using historical data from 2022 and 2023. This basic approach provides a foundation for understanding potential trends but may be limited by a few factors.

**Screenshots:**

PySpark Installation:

****

Anaconda:

****

Jupyter Notebook:

