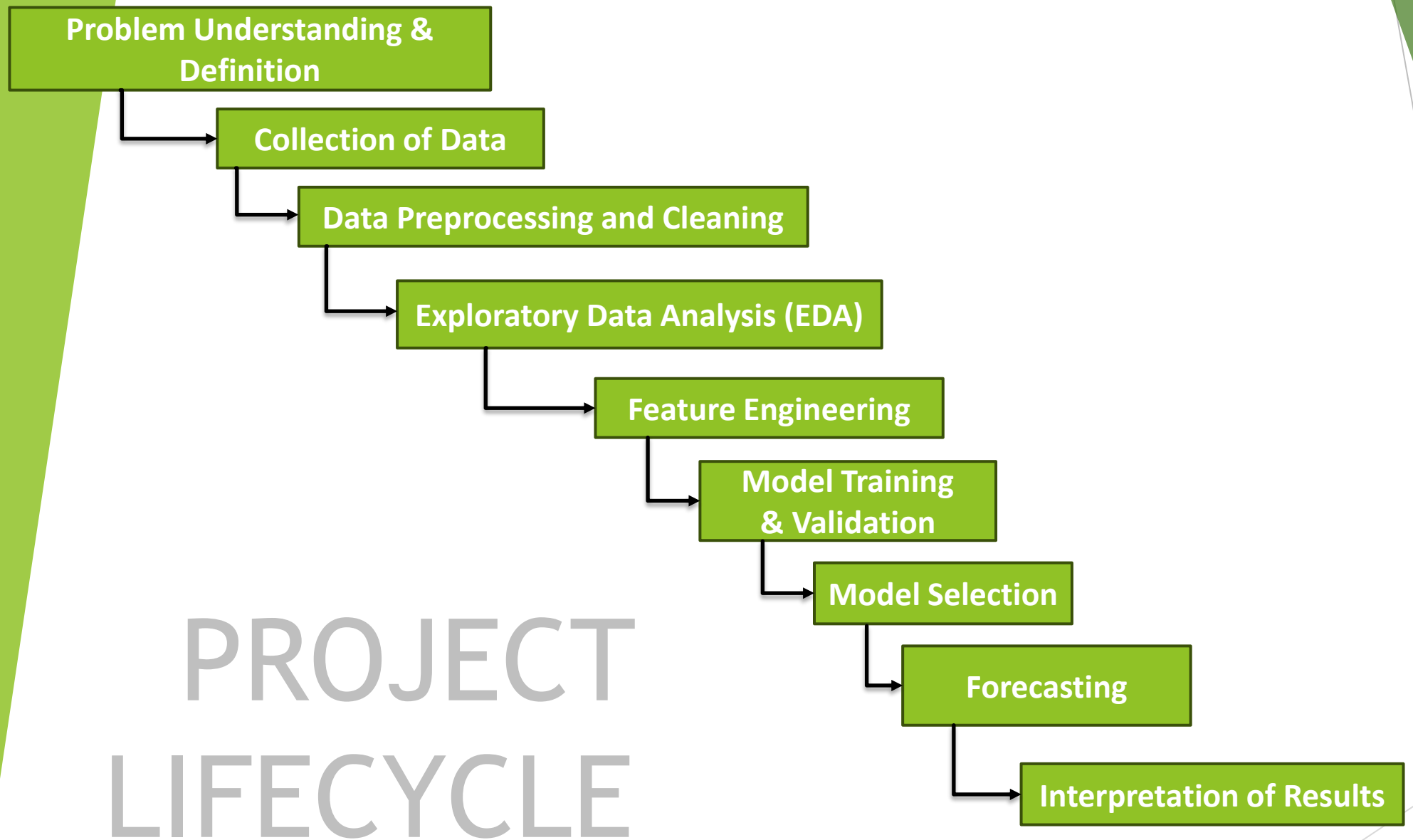
INFO: 523 Data Mining – Final Project

Urban Climate Patterns : Analysis of Urban Heat Islands(UHIs)

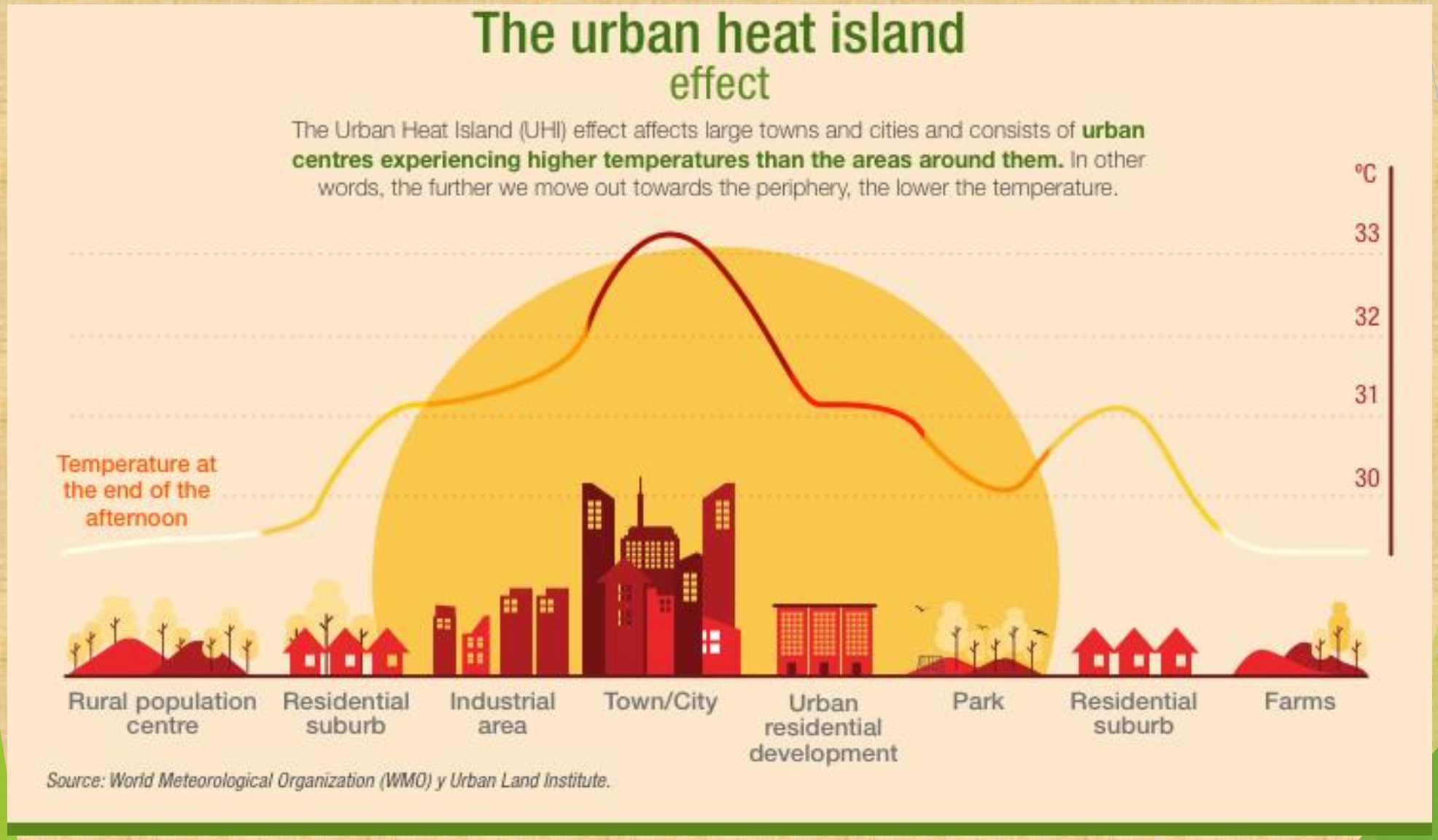
Project Members:

1. Ajinkya Deshmukh
2. Dhanyapriya Somasundaram
3. Kendall Beaver
4. Riyanshi Bohra
5. Udit Chaudhary

PROJECT LIFECYCLE



UHI Understanding & Definition



Collection of Data

- Details

1. Website
2. How many years of data
3. Cities we chose

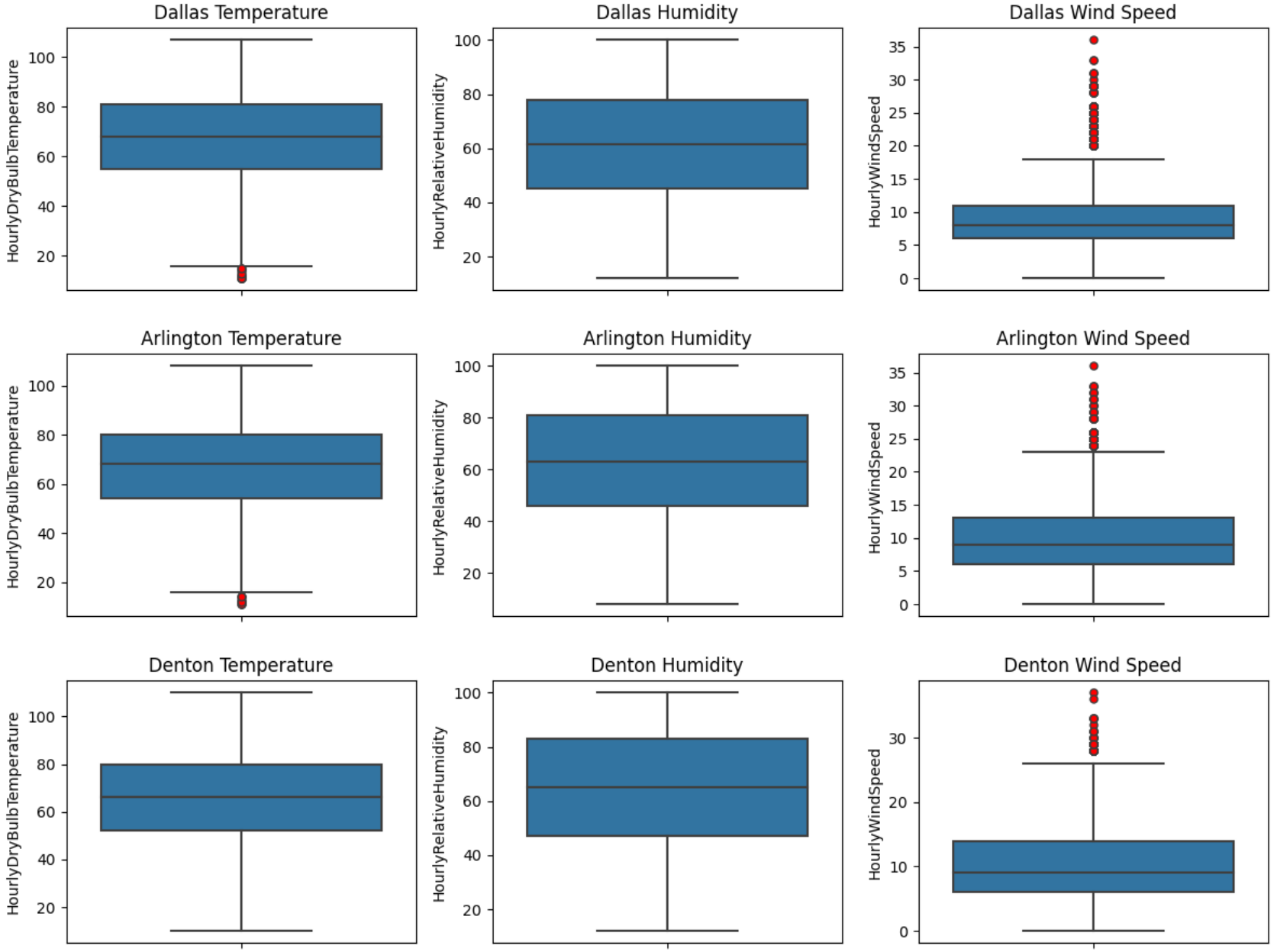
Data Preprocessing and Cleaning

- Rows, columns, Desired columns, NAN values, --KNN imputation(Numeric columns), No categorical features, removed the string and other characters , removal of duplicates, hourly data extraction

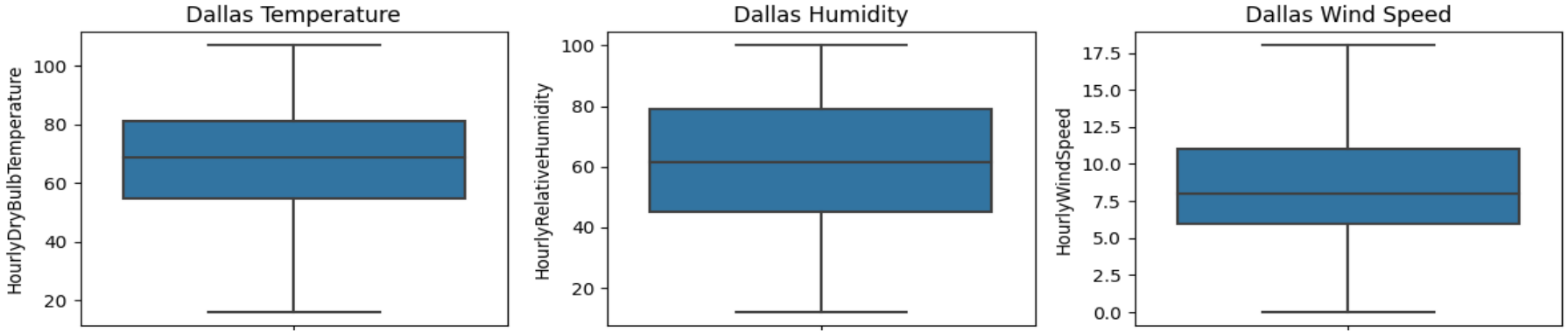
Exploratory Data Analysis (EDA)

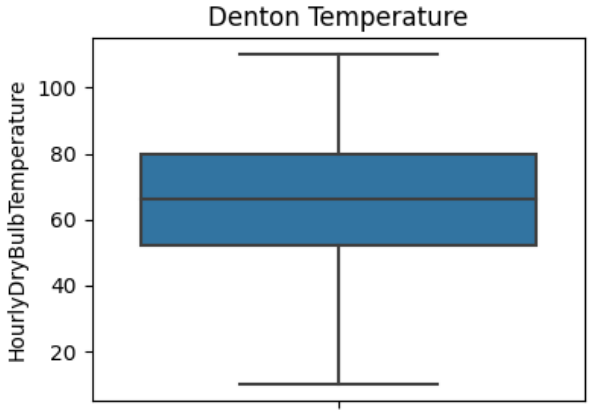
1. **Summary Statistics**
2. **Box Plots**
3. **Temperature plots**
4. **Histograms and Distributions**
5. **Correlation Analysis**
6. **Temporal Comparison**

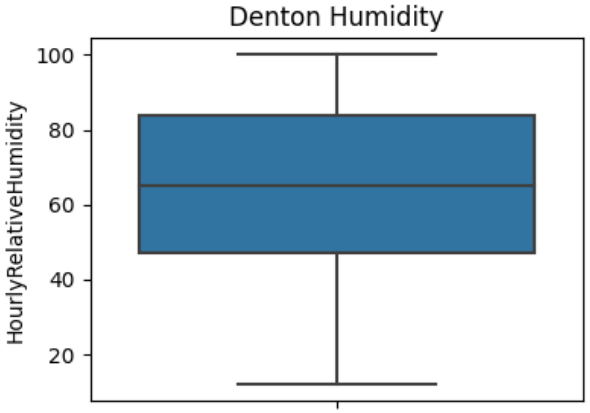
2. Box Plots: Outlier detection

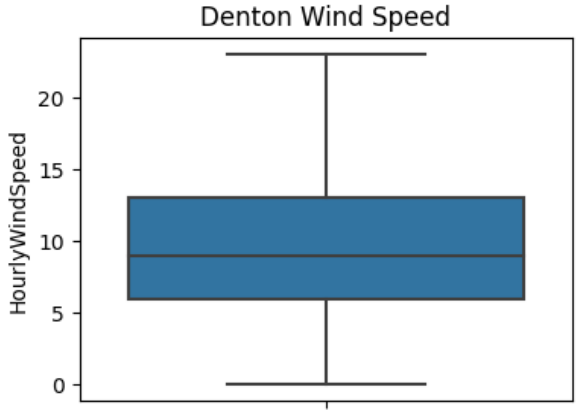


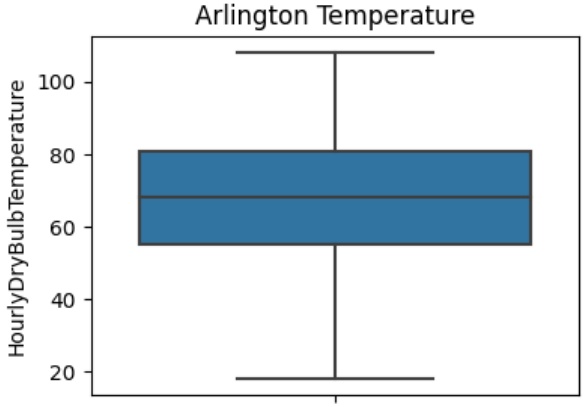
After cleaning…

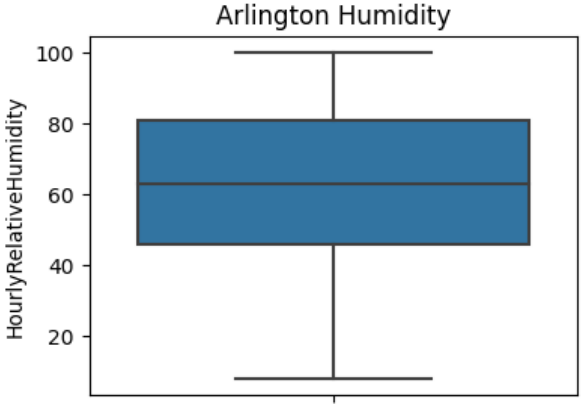


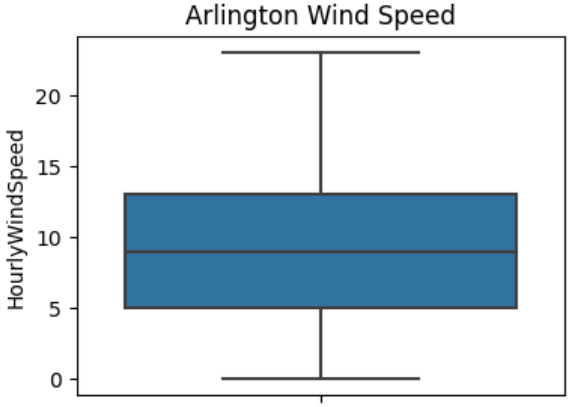




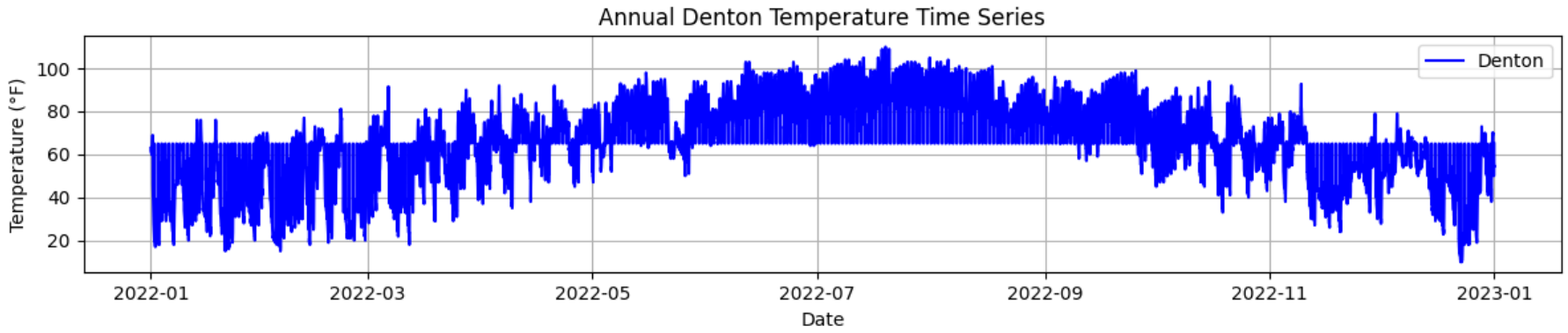


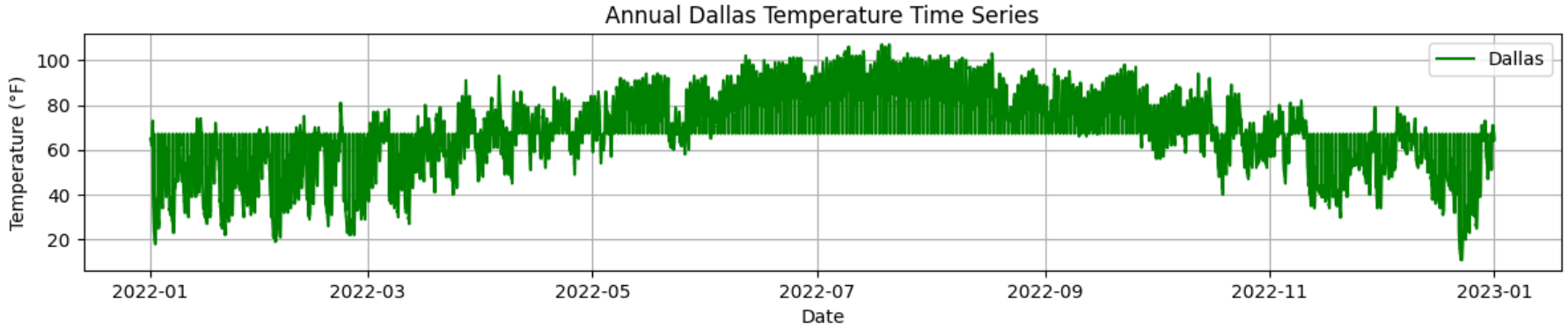


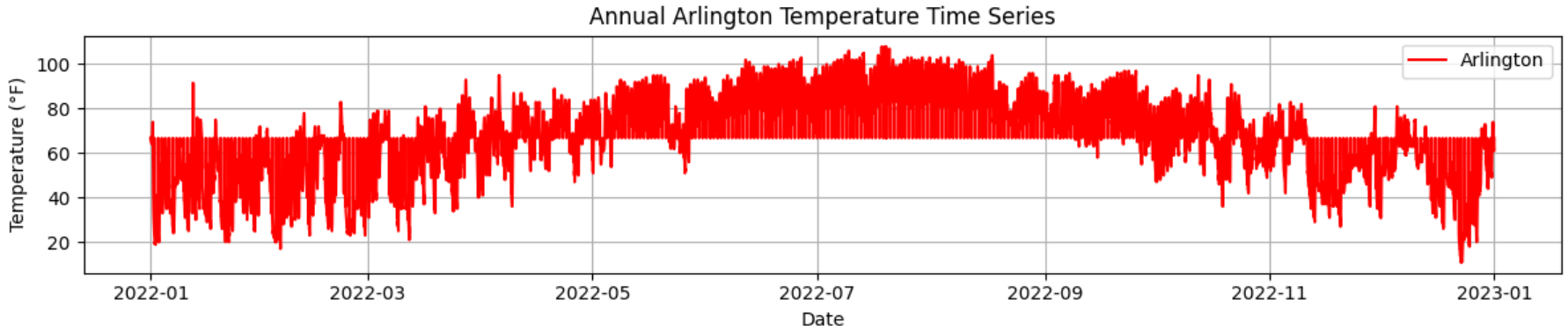




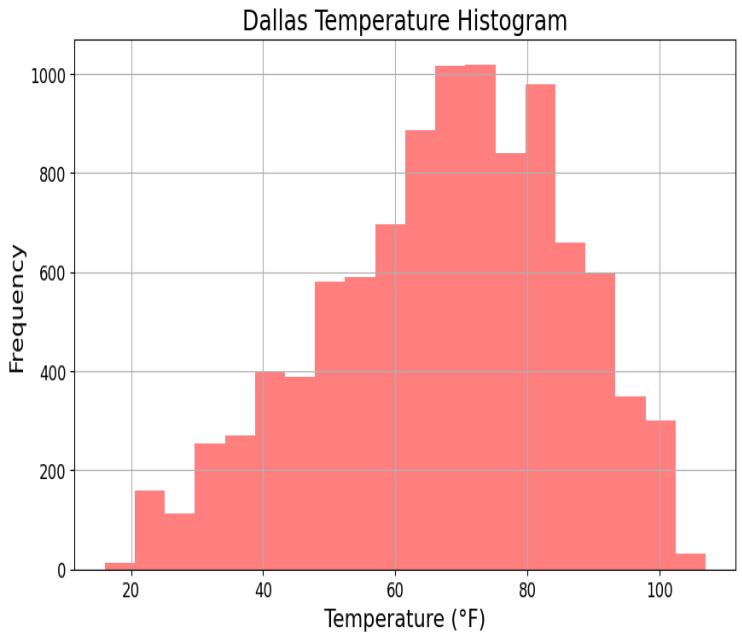
3. Annual variation of the temperature

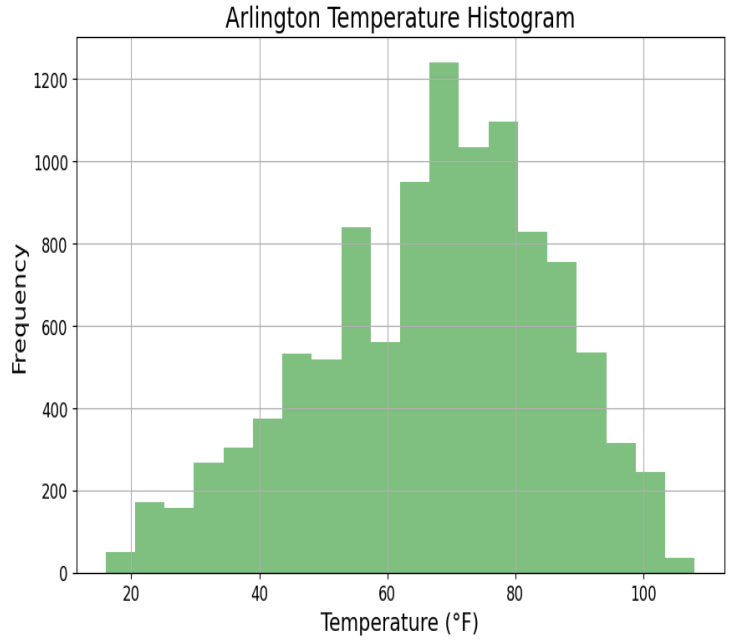


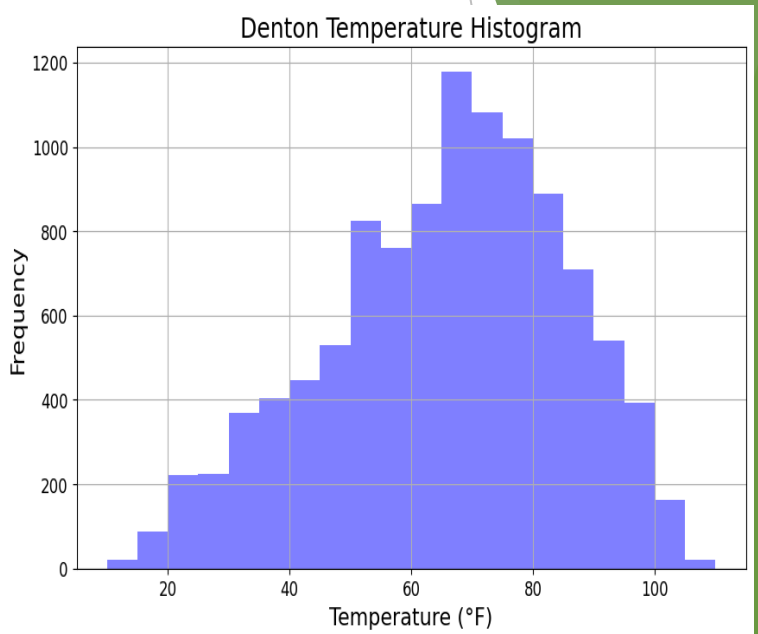




4. Histograms and Distributions:

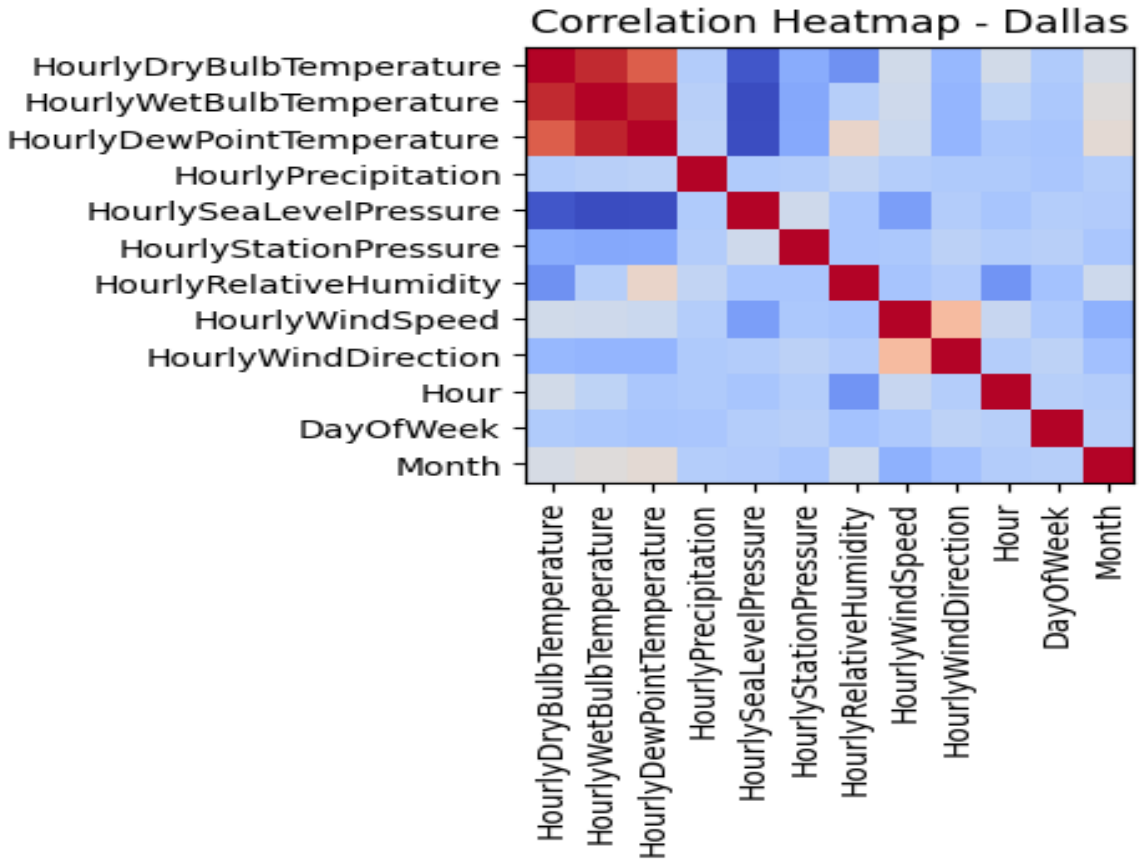




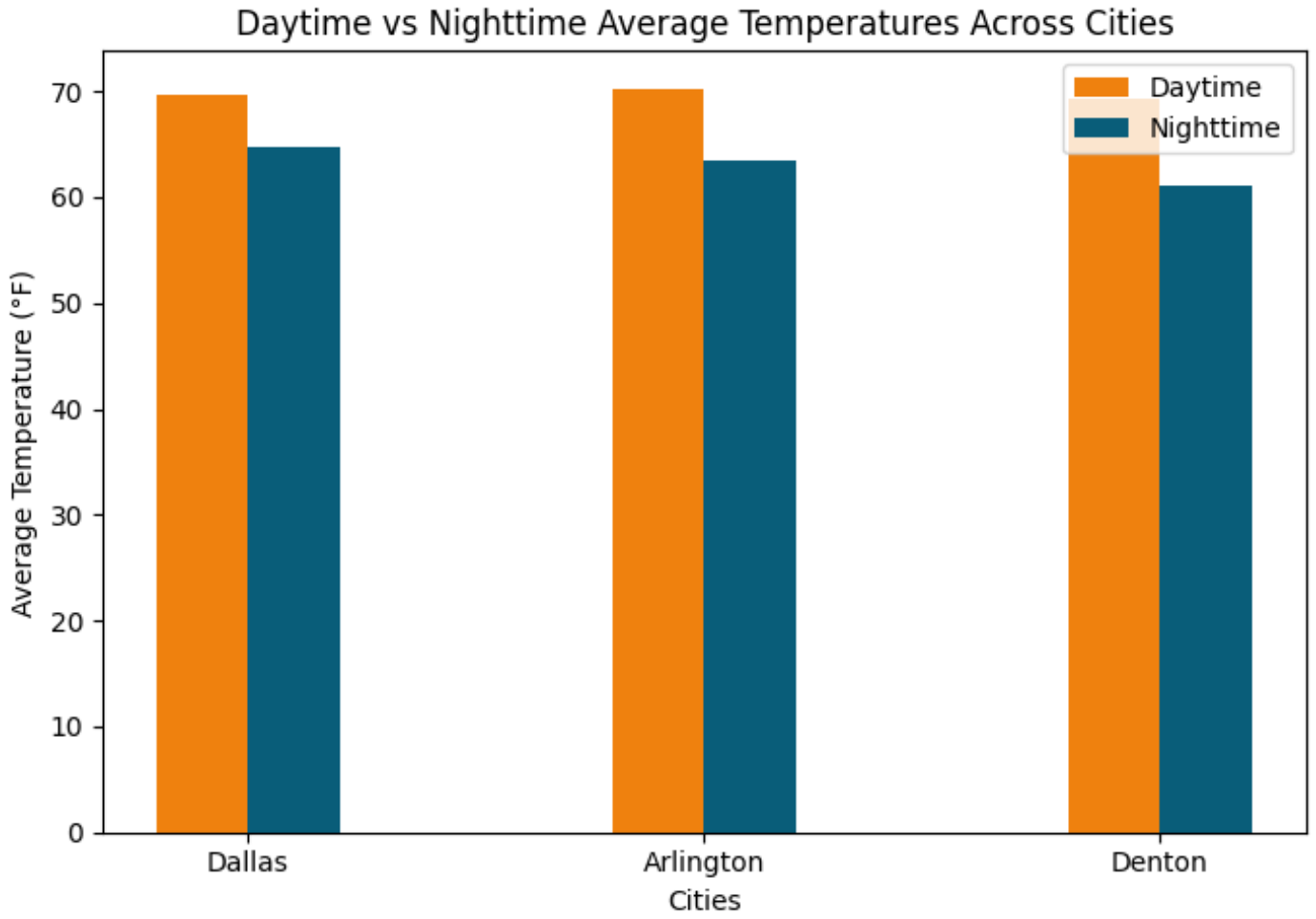


In all three cities, the **negative skew** suggests that there are **occasional periods of cooler** temperatures that pull the distribution's tail to the left. This might indicate that while the overall temperature range can be quite high, there are **fewer instances of extremely low** temperatures compared to the higher temperatures.

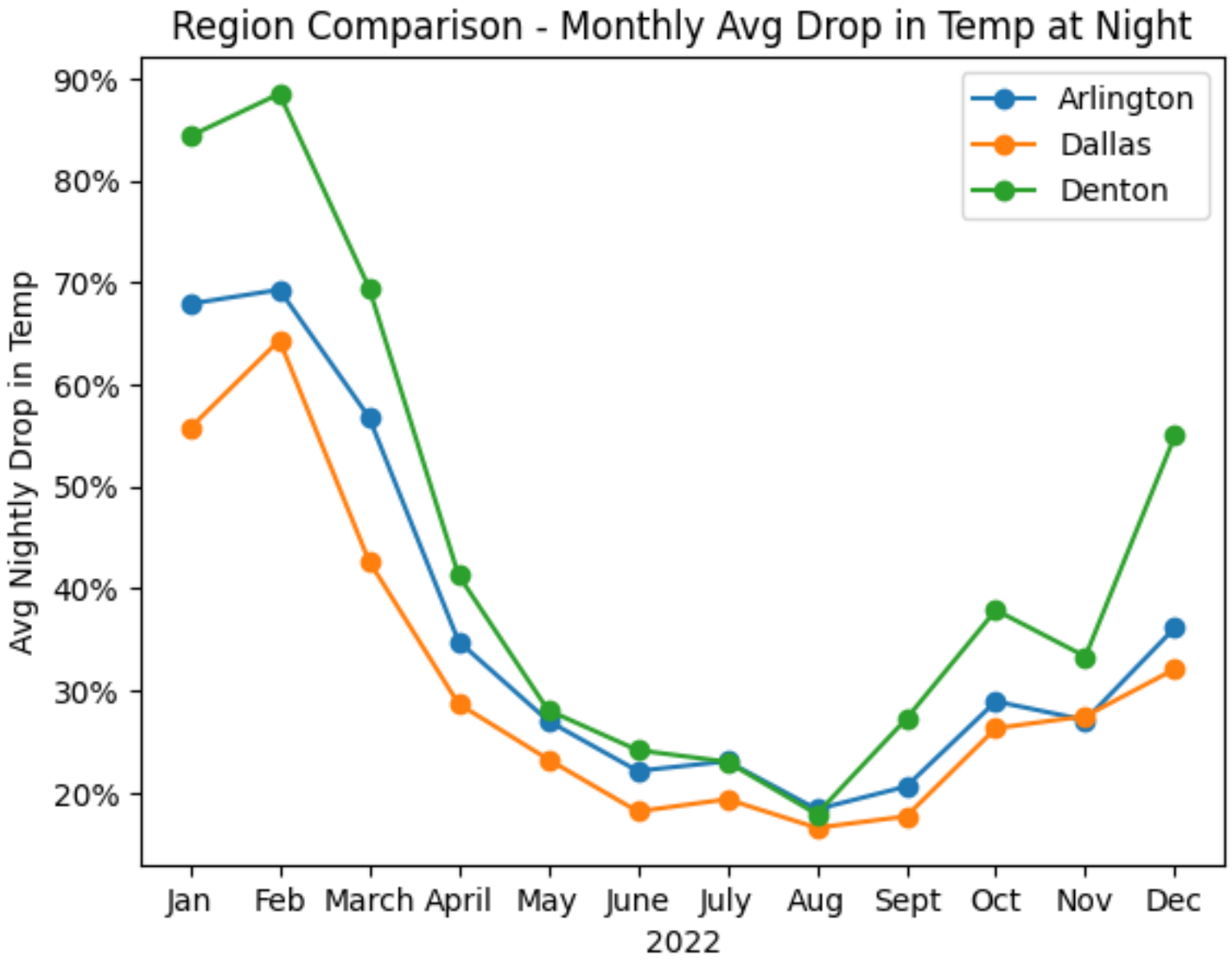
5. Correlation Analysis



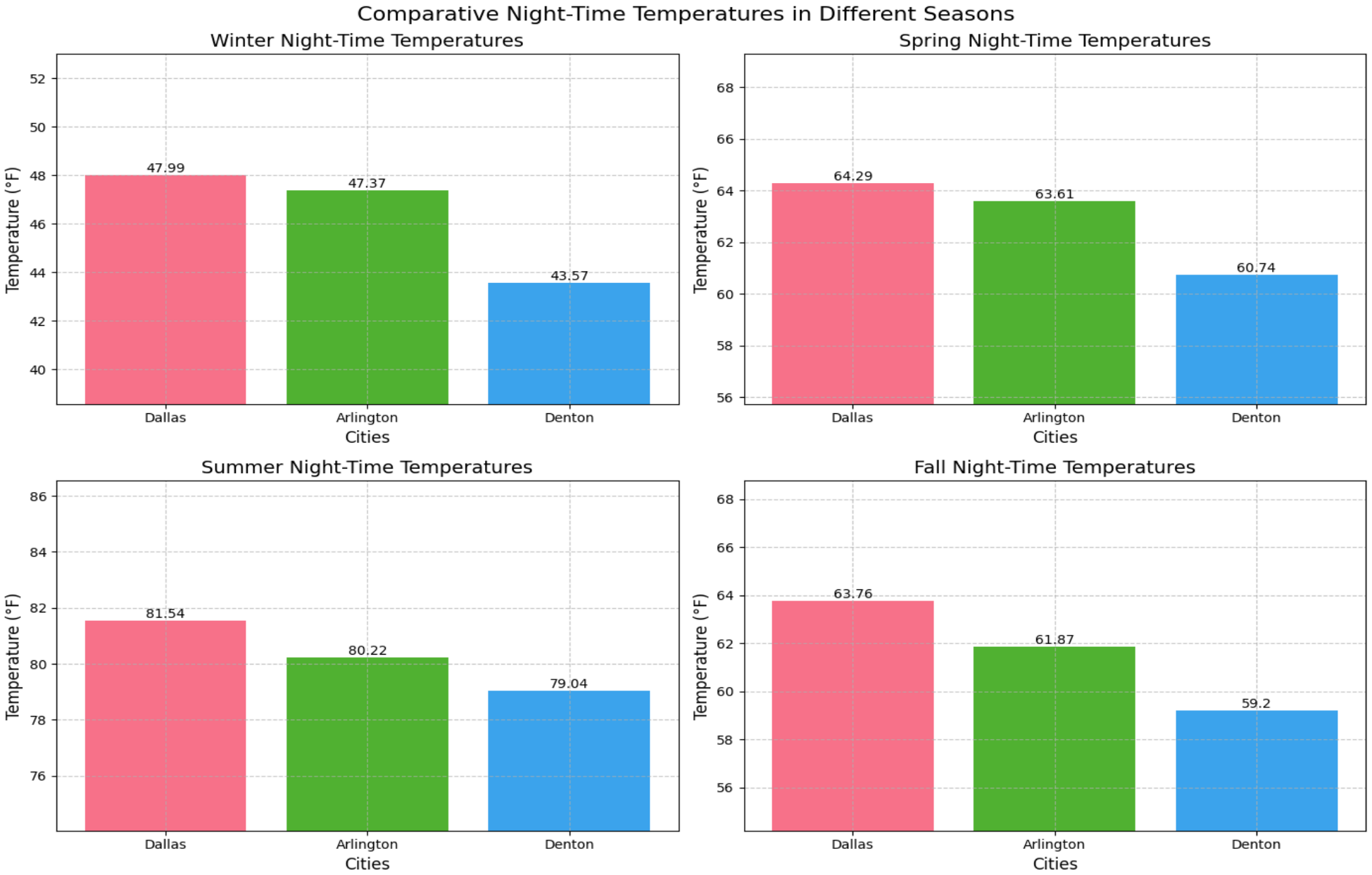
6. Temperature difference based on different times of the day

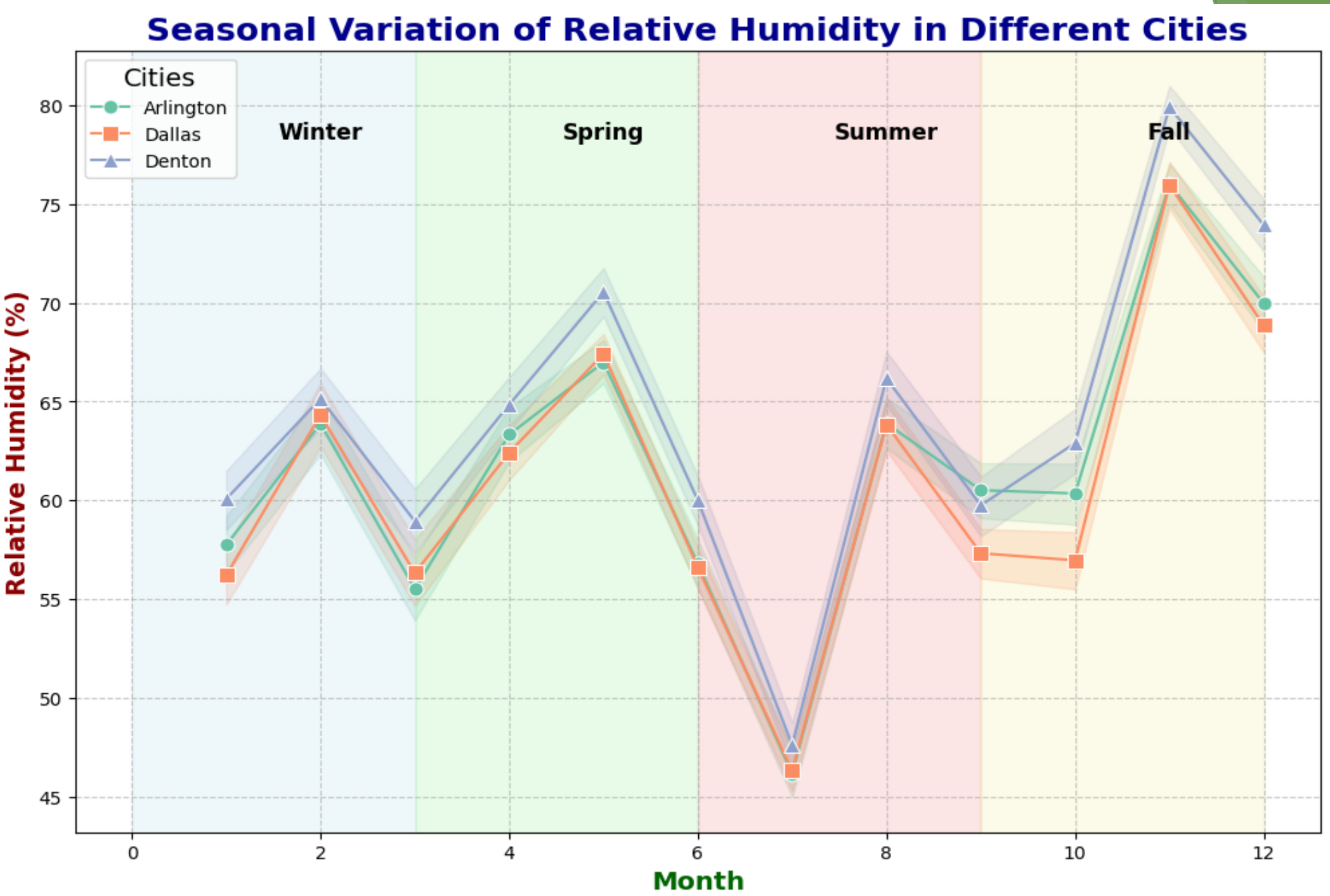


7. Region Comparison - Monthly Avg Drop in Temp at Night



8.Temporal Comparison




Seasonal Variation of Relative Humidity in Different Cities
Cities
Arlington
Dallas
Denton
Winter
Spring
Summer
Fall
Relative Humidity (%)
80
75
70
65
60
55
50
45
0
2
4
6
8
10
12
Month

Feature Engineering

RECAP: This project aims to study and analyze climatological data for **Dallas, Arlington, and Denton**, categorizing them based on an "Urban Heat Island" (UHI) Intensity scale. The goal is to understand the microclimatic effects of urbanization in different settings and **classify UHI intensity levels**. The project will focus on three key aspects:

1.Dallas (Significant City):

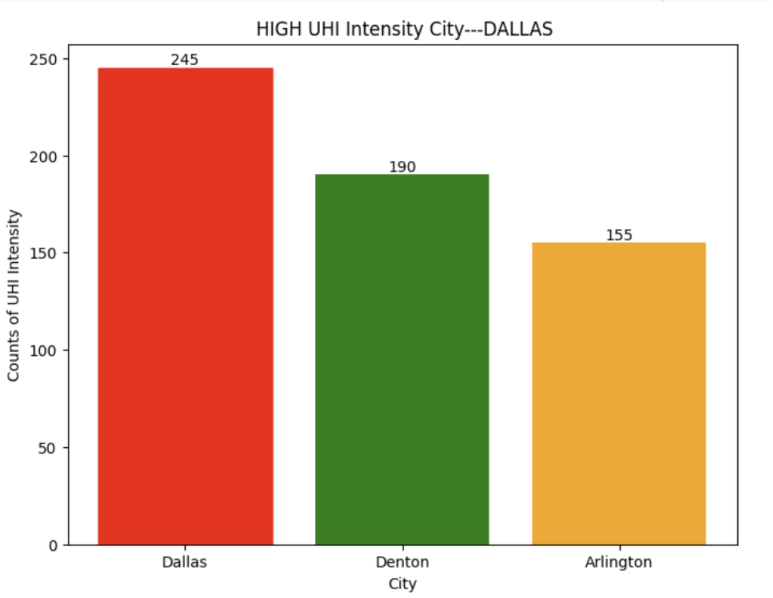
- Analyzing UHI in a major **metropolitan** area with **large** population density.
- Considering factors such as pollution, land use, and climate to determine UHI intensity.

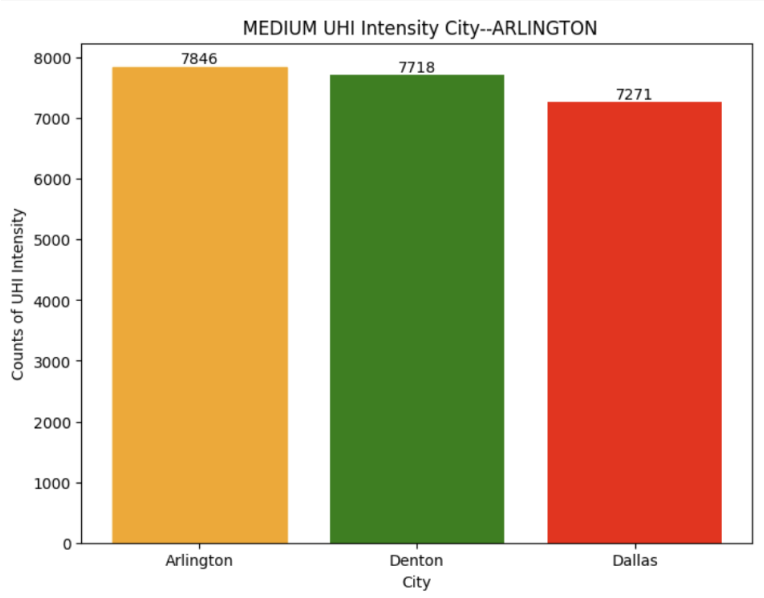
2. Arlington (Suburban Town):

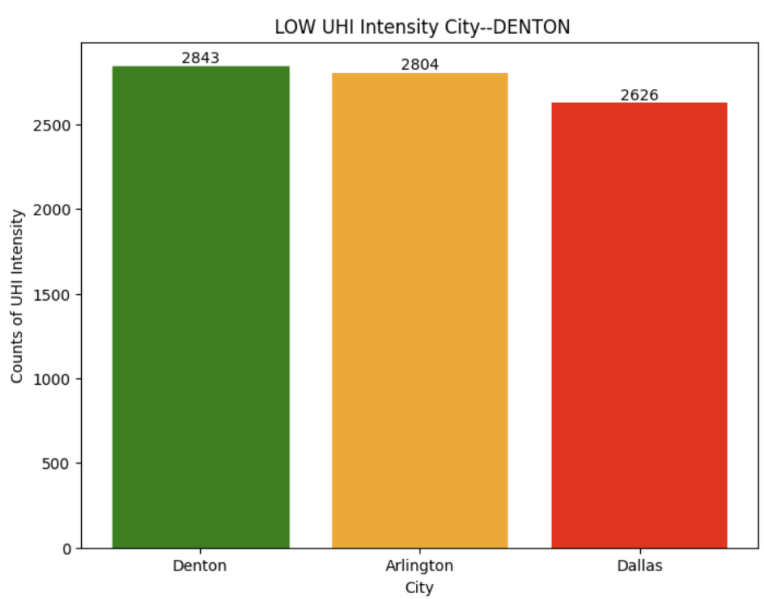
- Evaluating UHI in a **suburban** setting with **moderate** population density.

3. Denton (Rural City):

- Examining UHI in a **rural** city with **lower** population density.
- Considering factors like **reduced pollution** and different land use patterns.







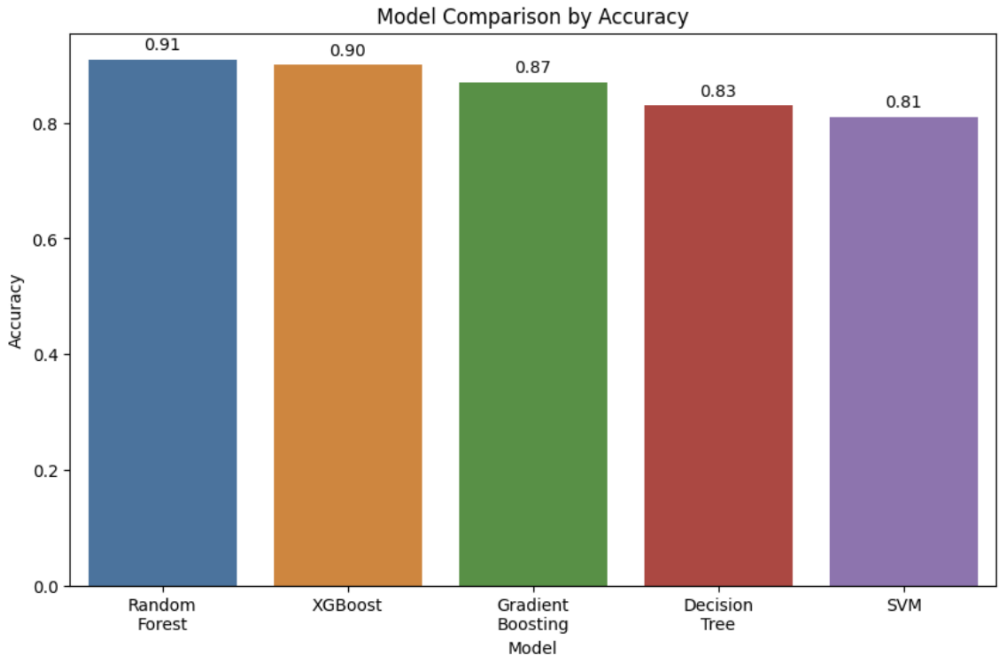
Model Selection

- Model 1: Decision Tree Classifier
- Model 2: XGBoost Classifier
- Model 3: Gradient Boost Classifier
- Model 4: SVM Classifier
- Model 5: Random Forest Classifier

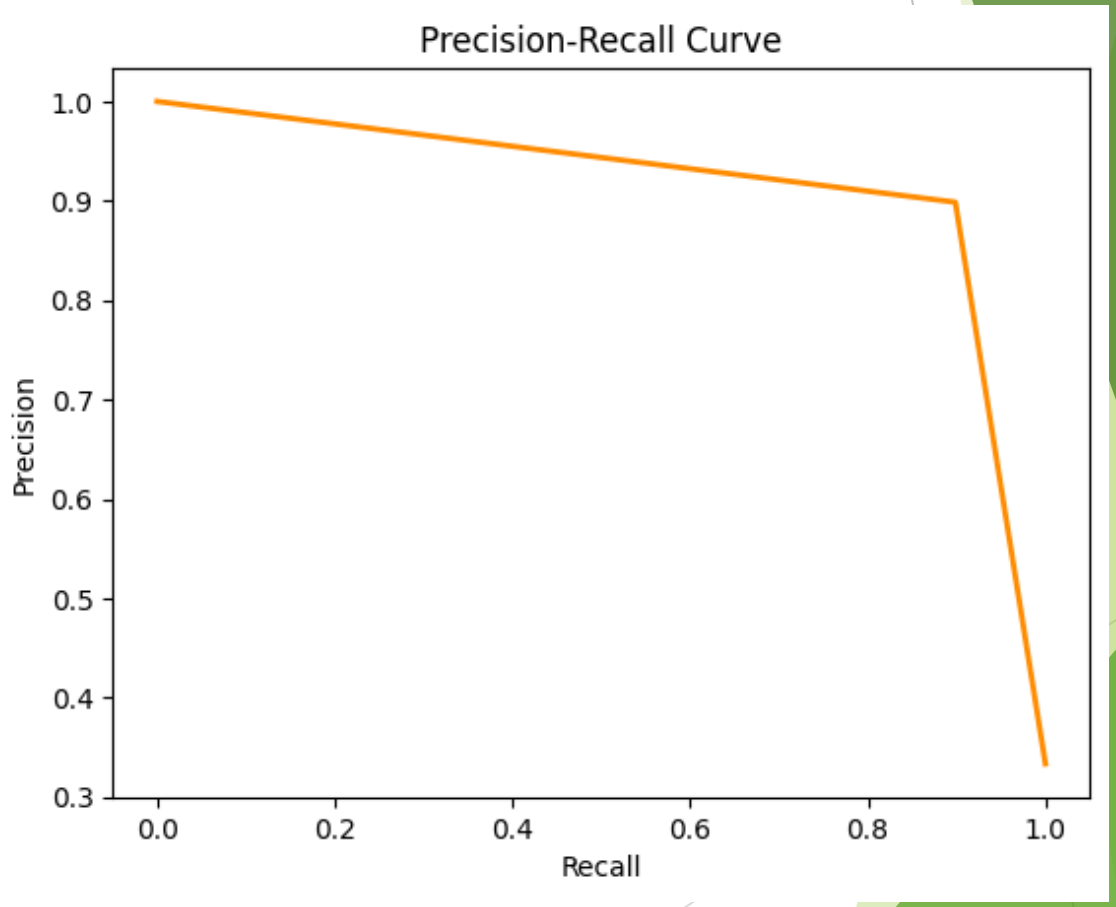
Model Training & Validation

- **Classification report:**

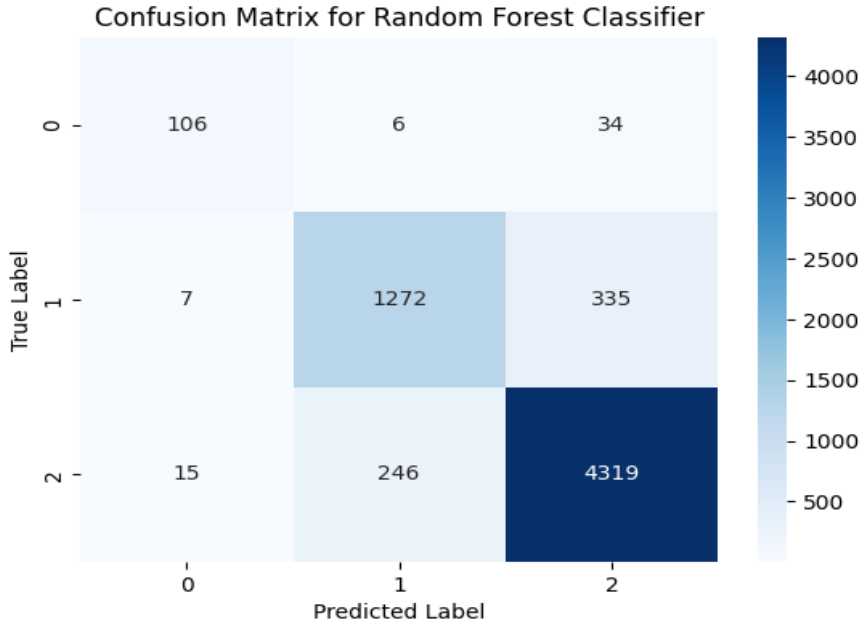
- **Choose Random forest**



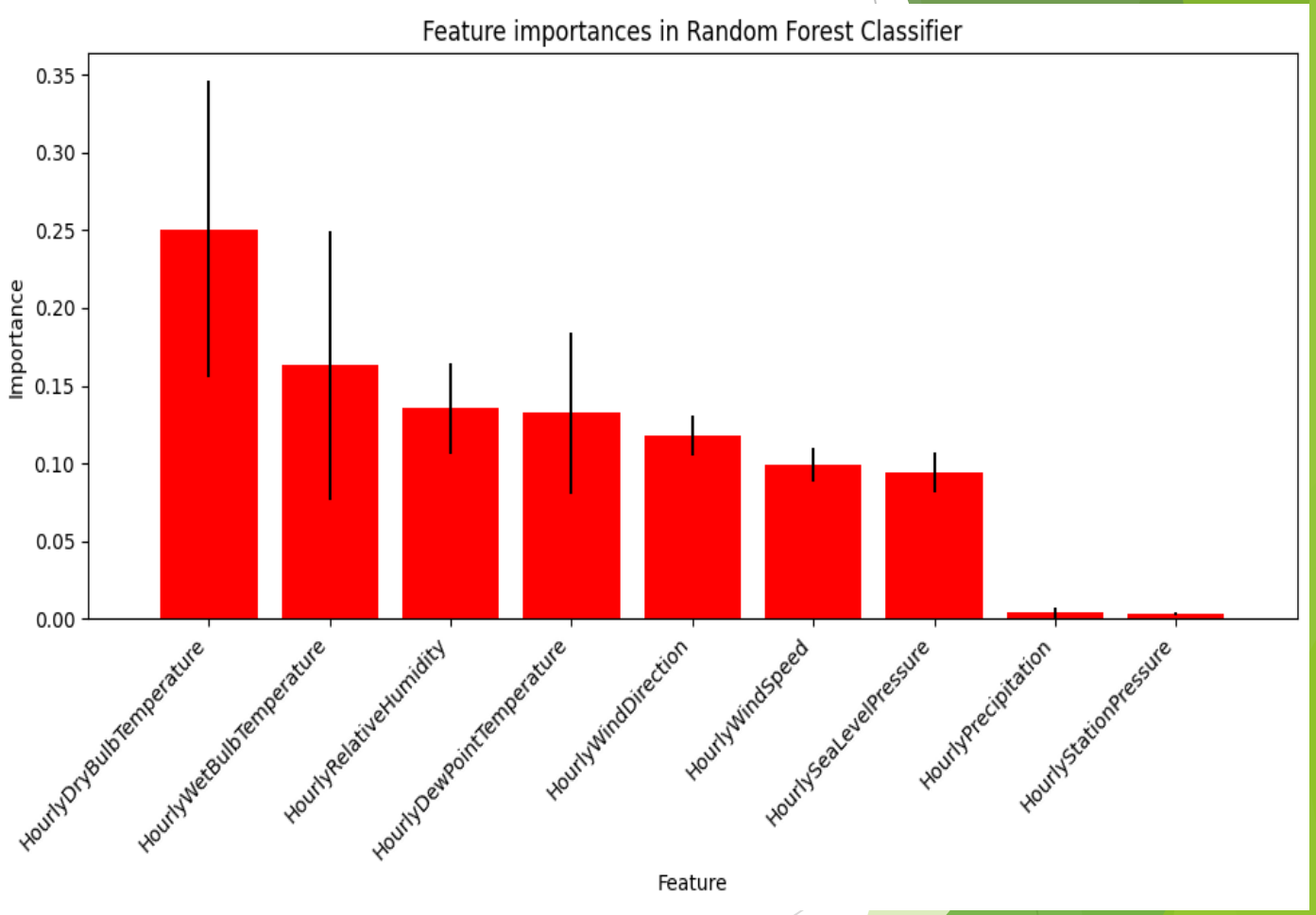
Precision-Recall Curve



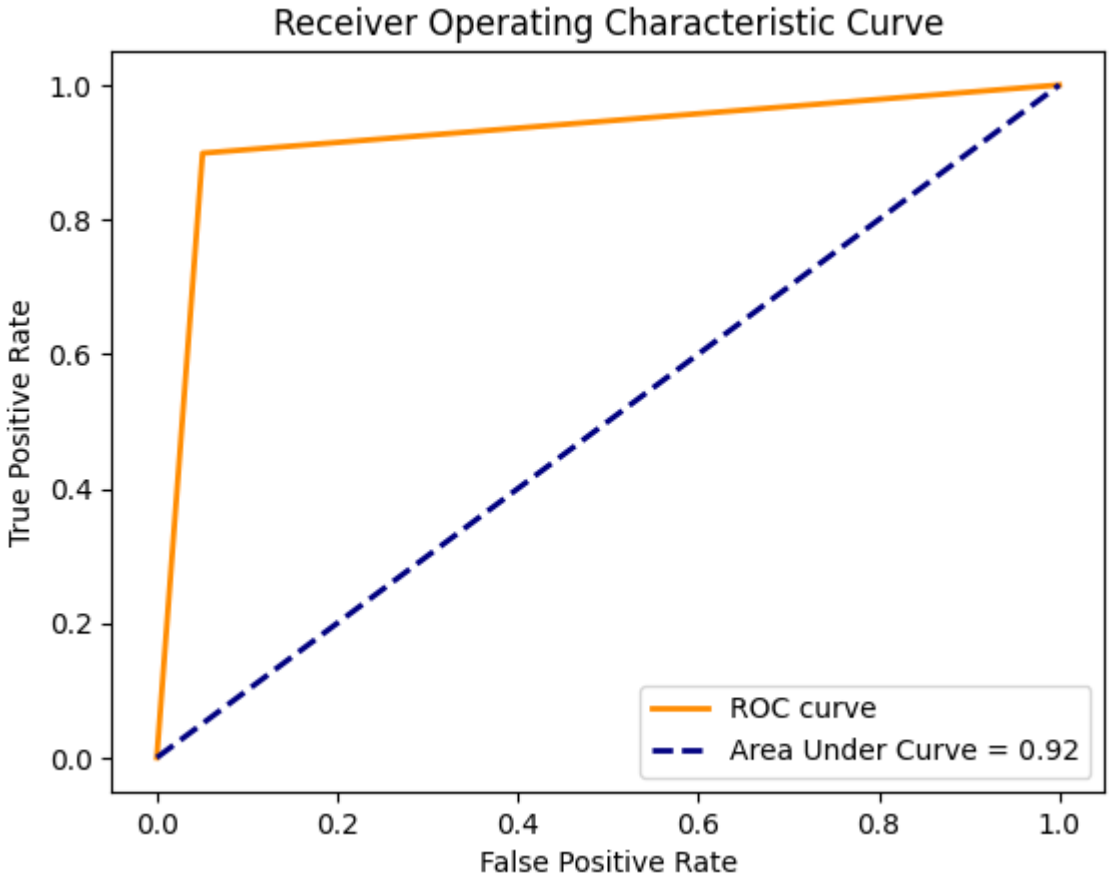
Confusion Matrix for Random Forest Classifier



Feature Importance from Random Forest



ROC and AUC curve



Forecasting

Interpretation of Results