Project Report

Project Title: Data-Driven Weather Prediction: Leveraging Data Mining for Enhanced Predictive Analytics

Group Number: 02

Names of Group Members:

1. Yash Jayesh Bhadreshwara: bhadreshwara.y@northeastern.edu
2. Siddhant Sunil Chavan: chavan.sid@northeastern.edu
3. Akshun Singh: singh.akshu@northeastern.edu
4. Bhakti Paithankar: paithankar.b@northeastern.edu

Contribution Percentage:

Yash Jayesh Bhadreshwara: 25%

Siddhant Sunil Chavan: 25%

Akshun Singh: 25%

Bhakti Paithankar: 25%

Submission Date: April 19th, 2024.

1. Introduction

This project focuses on the application of data mining techniques to predict weather patterns in Boston, a city known for its rich and varied climate. Utilizing a dataset that encompasses a decade of daily weather data from the Meteostat API, Through the meticulous analysis of extensive historical weather data, this endeavor seeks to uncover nuanced insights into the factors driving Boston's climate variability. By harnessing the power of advanced data mining techniques, we aim to develop predictive models capable of anticipating shifts in temperature, precipitation, wind patterns, and other meteorological phenomena with heightened accuracy. Moreover, by considering socio-economic dependencies alongside meteorological factors, this project endeavors to provide actionable insights that can inform decision-making processes across various sectors. From urban planning and emergency management to transportation and agriculture, the implications of precise weather predictions extend far and wide, shaping the resilience and sustainability of Boston's diverse communities.The city's experience with extreme weather events and the availability of extensive historical weather data provides a robust foundation for developing predictive models. This project not only addresses scientific curiosity but also serves a practical need, aiming to mitigate the impacts of adverse weather on Boston’s population and infrastructure.

2. Problem Statement and Definition

The primary challenge this project seeks to address is the development of an accurate and reliable predictive model for Boston's weather conditions, leveraging historical data covering significant meteorological parameters from 2013 to 2023. This challenge is critical due to the direct impact of weather on public safety, transportation, energy consumption, and overall economic activities in the region. The project will explore various data mining techniques to forecast daily weather metrics such as temperature, precipitation, and wind patterns. Improving prediction accuracy is not merely a technical goal but a necessity to enhance urban planning, emergency preparedness, and resource management in the face of climate variability and change. By focusing on Boston, this study also contributes to understanding the broader implications of climate trends and the increasing frequency of extreme weather events, providing insights that are vital for future climate resilience strategies.

3. <u>Data Exploration</u>

We obtained the Boston weather dataset spanning from 2013 to 2023 from Kaggle. This dataset comprises the following columns:

- Date: Represented in string format
- Average Temperature (tavg): Recorded in Celsius, as a float
- Minimum Temperature (tmin): Recorded in Celsius, as a float
- Maximum Temperature (tmax): Recorded in Celsius, as a float
- Precipitation (prcp): Daily total measured in millimeters, as a float
- Wind Direction (wdir): Average direction in degrees, as a float
- Wind Speed (wspd): Average speed in kilometers per hour, as a float
- Air Pressure (pres): Average sea-level pressure in hectopascals, as a float

4. <u>Data Preprocessing and Handling Missing Values</u>

Our initial exploration revealed several important insights into the Boston weather dataset. Firstly, an examination of the first 10 rows provided a glimpse into its structure and contents. Secondly, detailed assessments using df.info() and df.describe() unveiled crucial information regarding data types, missing values, and statistical summaries.

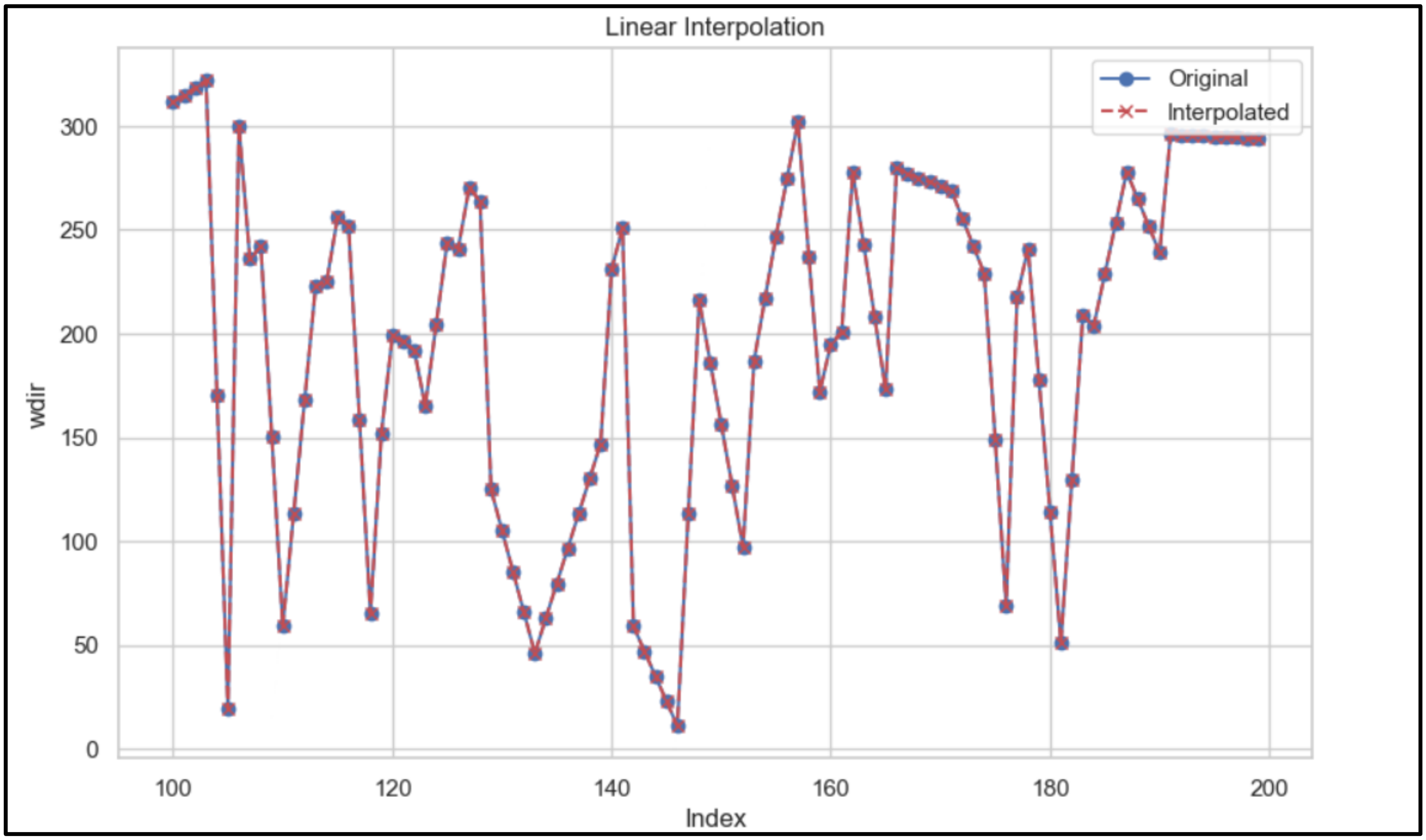
One notable observation was the presence of missing values in the 'tavg' column, prompting the decision to impute them by aligning with corresponding values from other columns. This meticulous approach ensured uniformity throughout the dataset. Moreover, to address missing values in the 'wdir' column, linear interpolation was employed, enhancing dataset completeness. Additionally, transforming the 'time' column into datetime format facilitated temporal analysis, with supplementary columns for year and month extraction further enhancing granularity. Lastly, preprocessing techniques like standardization and normalization were applied to standardize data across features, preparing it for subsequent analytical and modeling tasks. These comprehensive data processing steps underscore our commitment to ensuring the dataset's readiness for insightful exploration and meaningful interpretation of Boston's weather patterns.

Handling Missing Values

```
df['tavg'] = df.apply(lambda row: (row['tmin'] + row['tmax']) / 2 if pd.isnull(row['tavg']) and not pd.isnull(row['tmin']) and not pd.isnull(row['tmax']) else row['tavg'], axis=1)
```

```
df.interpolate(method='linear', inplace=True)
```

```
start_index = 100
end_index = 200

df_interpolated = df.interpolate(method='linear')

plt.figure(figsize=(10, 6))

plt.plot(df.index[start_index:end_index], df['wdir'][start_index:end_index], marker='o', linestyle='-', color='b', label='Original')

plt.plot(df_interpolated.index[start_index:end_index], df_interpolated['wdir'][start_index:end_index], marker='x', linestyle='--', color='r', label='Interpolated')

plt.xlabel('Index')
plt.ylabel('wdir')
plt.title('Linear Interpolation')
plt.legend()
plt.grid(True)
plt.show()
```

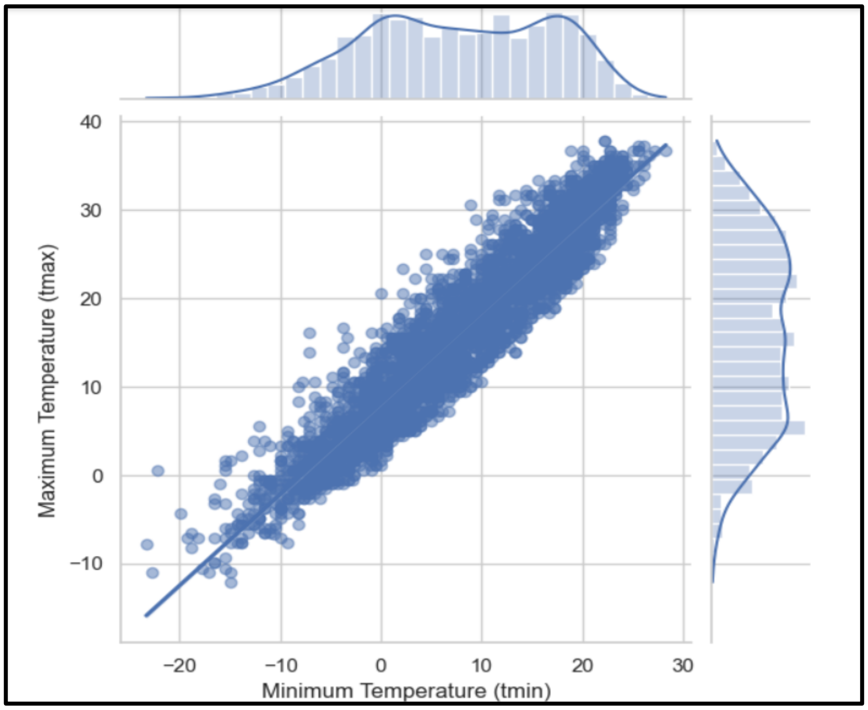


In the following image, we fixed the missing values, and every column has equal amount of data.

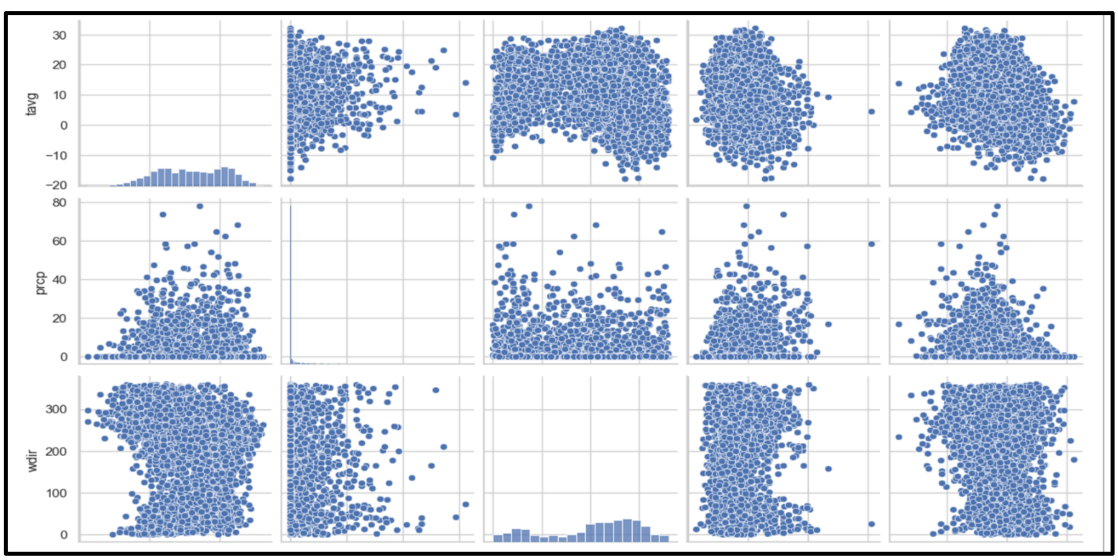
```
[ ]  df.info()

     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 3653 entries, 0 to 3652
     Data columns (total 11 columns):
      #   Column       Non-Null Count  Dtype
     ---  ------       --------------  -----
      0   time         3653 non-null   datetime64[ns]
      1   tavg         3653 non-null   float64
      2   tmin         3653 non-null   float64
      3   tmax         3653 non-null   float64
      4   prcp         3653 non-null   float64
      5   wdir         3653 non-null   float64
      6   wspd         3653 non-null   float64
      7   pres         3653 non-null   float64
      8   year         3653 non-null   int64
      9   month_index  3653 non-null   int64
      10  month        3653 non-null   object
     dtypes: datetime64[ns](1), float64(7), int64(2), object(1)
     memory usage: 314.1+ KB
```

5. Exploratory Data Analysis:

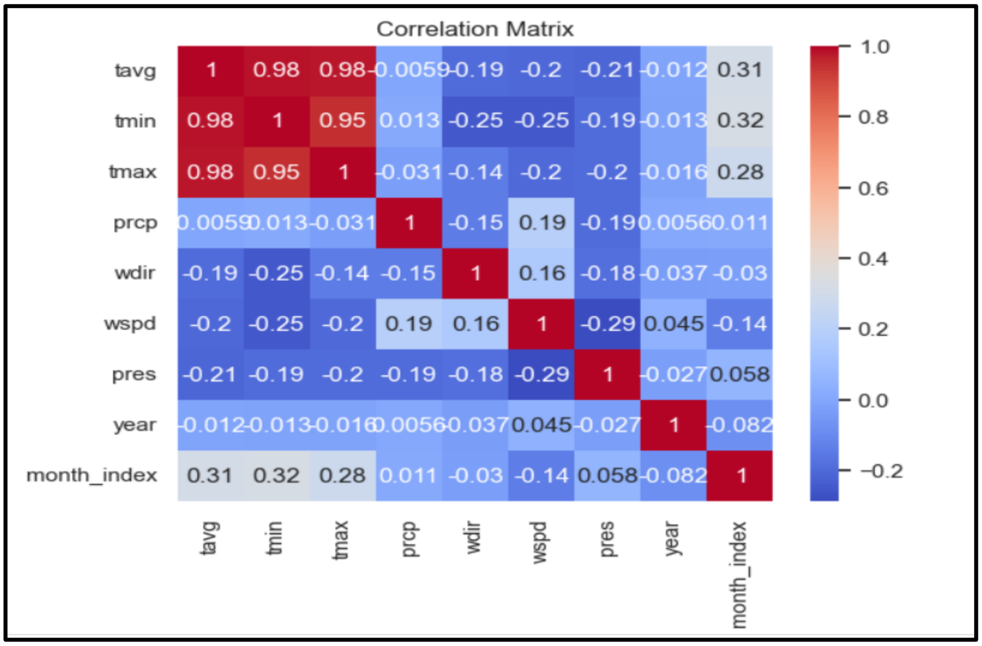
In the exploratory data analysis (EDA) of the Boston weather dataset, a comprehensive approach was adopted to understand the relationships between various meteorological parameters. A scatter plot with regression line analysis was conducted specifically to explore the connection between the minimum and maximum temperatures ('tmin' vs. 'tmax'). This analysis revealed insights into how changes in minimum temperatures might influence maximum temperatures, aiding in understanding temperature dynamics and weather patterns.



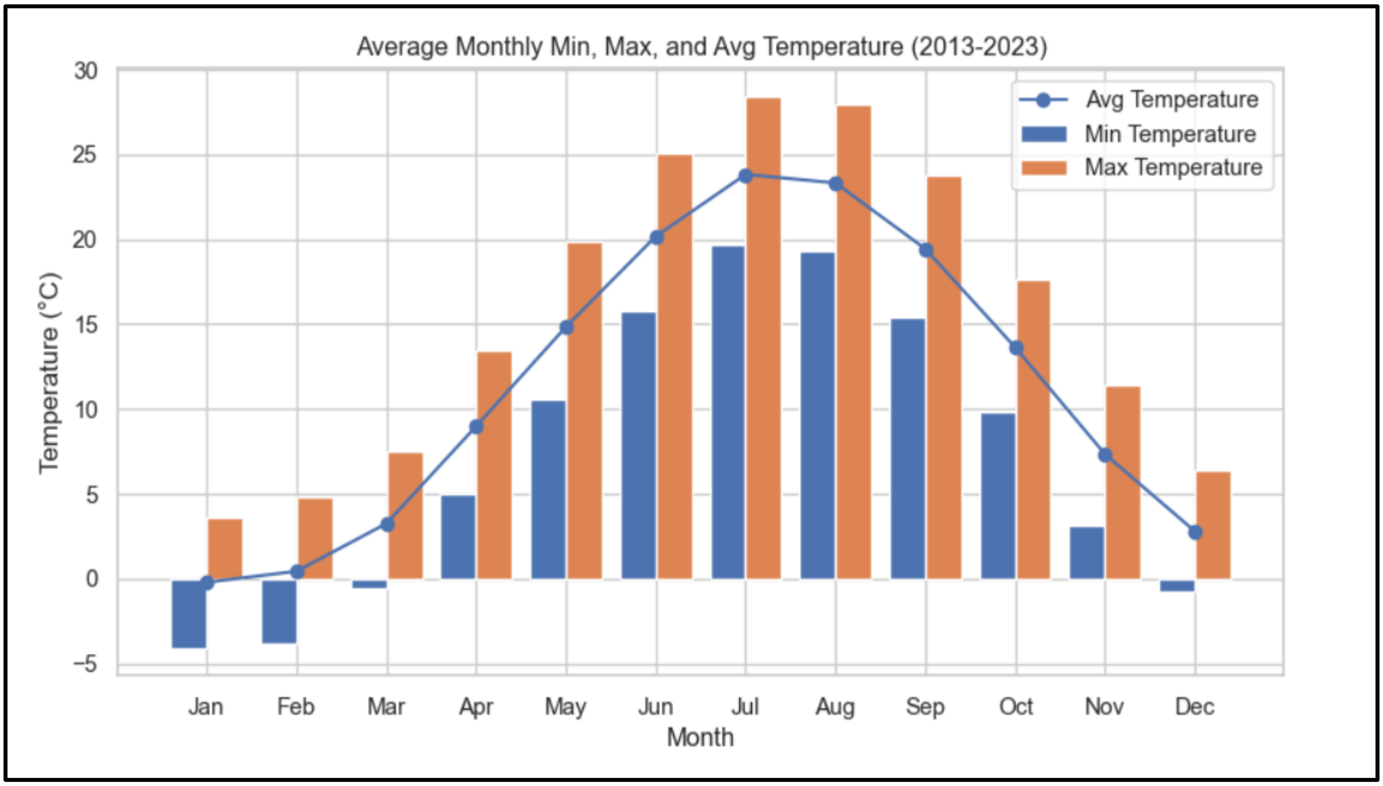
Moreover, to delve deeper into the relationships among different meteorological variables, a pairplot using Seaborn (sns.pairplot()) was employed. This visualization technique facilitated the simultaneous examination of multiple pairwise relationships by plotting scatterplots for each pair of variables and histograms for individual variables along the diagonal. The variables included in the pairplot were 'tavg' (average temperature), 'prcp' (precipitation), 'wdir' (wind direction), 'wspd' (wind speed), and 'pres' (pressure). Additionally, a correlation matrix was generated to quantify the strength and direction of linear relationships between these variables. The correlation matrix provided correlation coefficients for each pair of variables, ranging from -1 to 1, where values closer to 1 indicated a strong positive correlation, values closer to -1 indicated a strong negative correlation, and values close to 0 indicated little to no correlation.



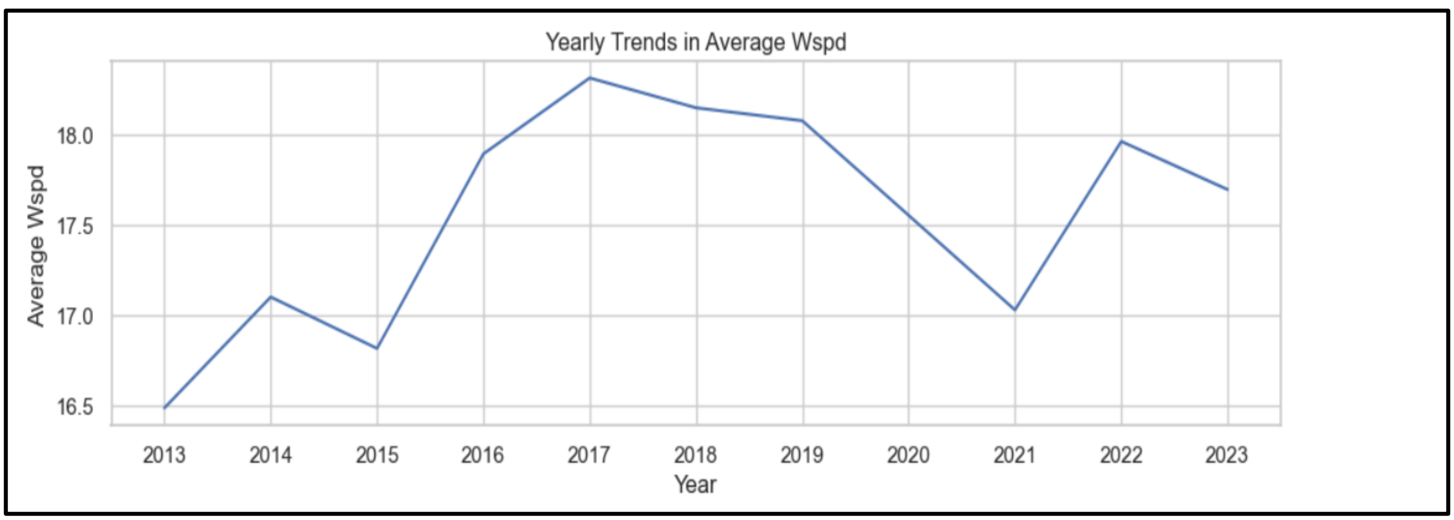
By analyzing the correlation matrix, we gained valuable insights into the interdependencies among meteorological parameters. For instance, we could identify which variables exhibited strong correlations, such as temperature and pressure, or wind speed and wind direction. These insights are crucial for understanding the complex interactions within the weather system and for informing further analysis and modeling efforts. Overall, the combined use of scatter plots, regression lines, pairplots, and correlation matrices provided a comprehensive framework for exploring the Boston weather dataset, enabling us to uncover patterns, relationships, and trends essential for understanding and predicting weather phenomena.



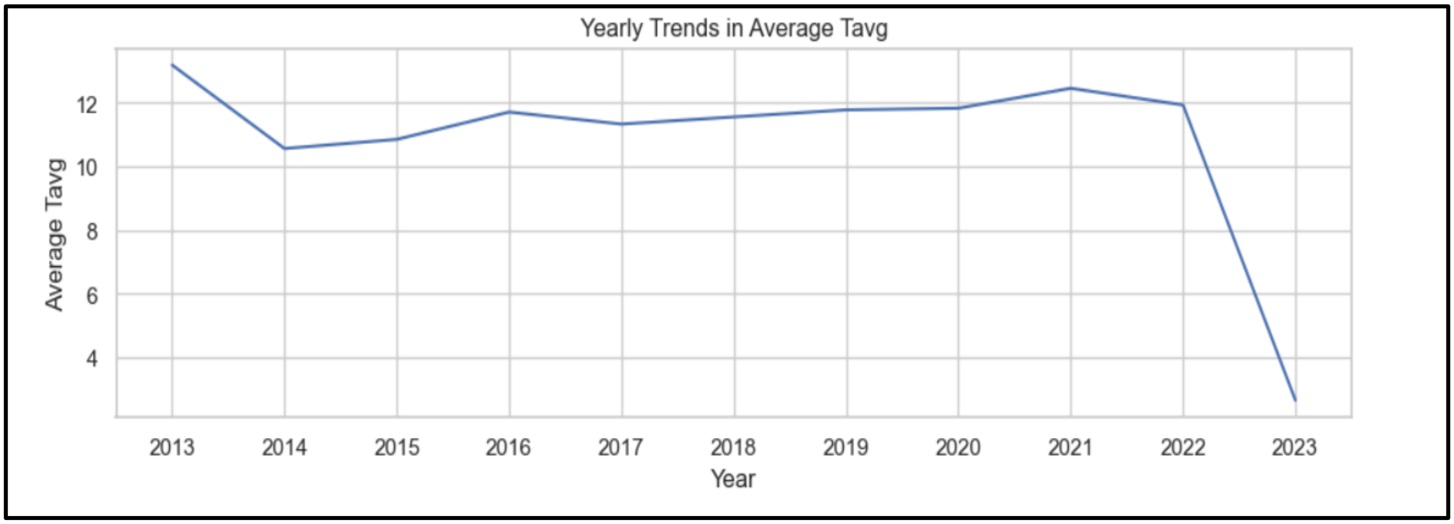
6. Data Visualization



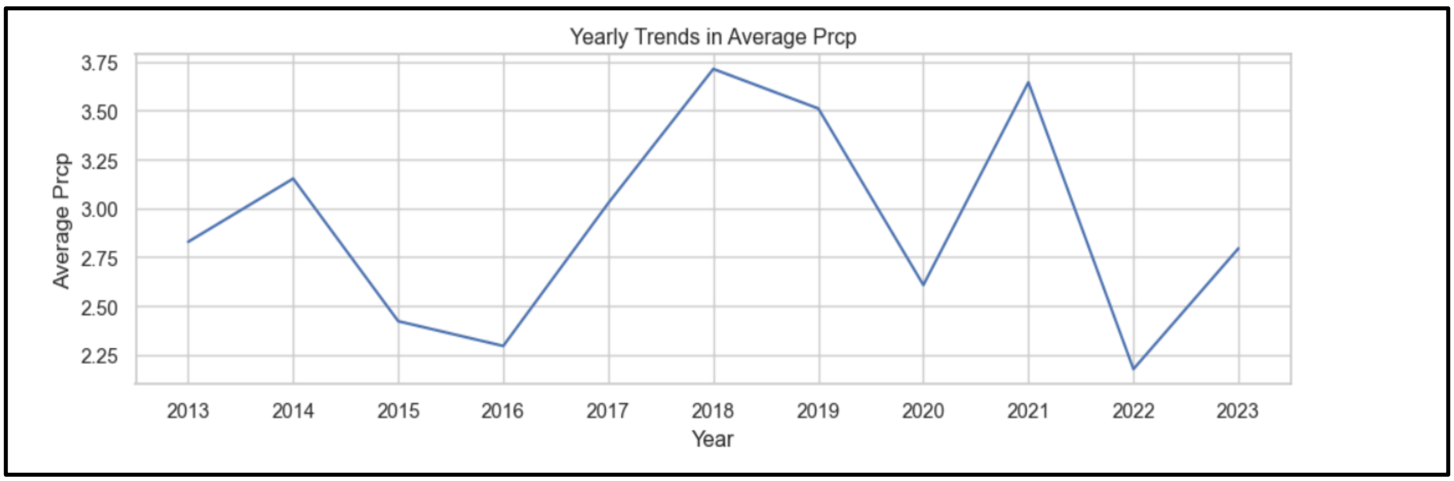
The bar chart depicting temperature trends in Boston over the span of 2013 to 2023 offers a compelling narrative of the city's climatic oscillations. It showcases a cyclical pattern of temperature variation, with winter months, especially December through February, consistently registering sub-zero averages, while summer months, notably May through September, experience peaks surpassing 30 degrees Celsius. The chart's overarching trend line illustrates a gradual ascent from winter's icy grip to summer's warmth, followed by a gradual descent as the year draws to a close. This visualization serves as a valuable tool for stakeholders, providing insights crucial for urban planning, resource allocation, and climate adaptation strategies tailored to Boston's distinct seasonal rhythms.



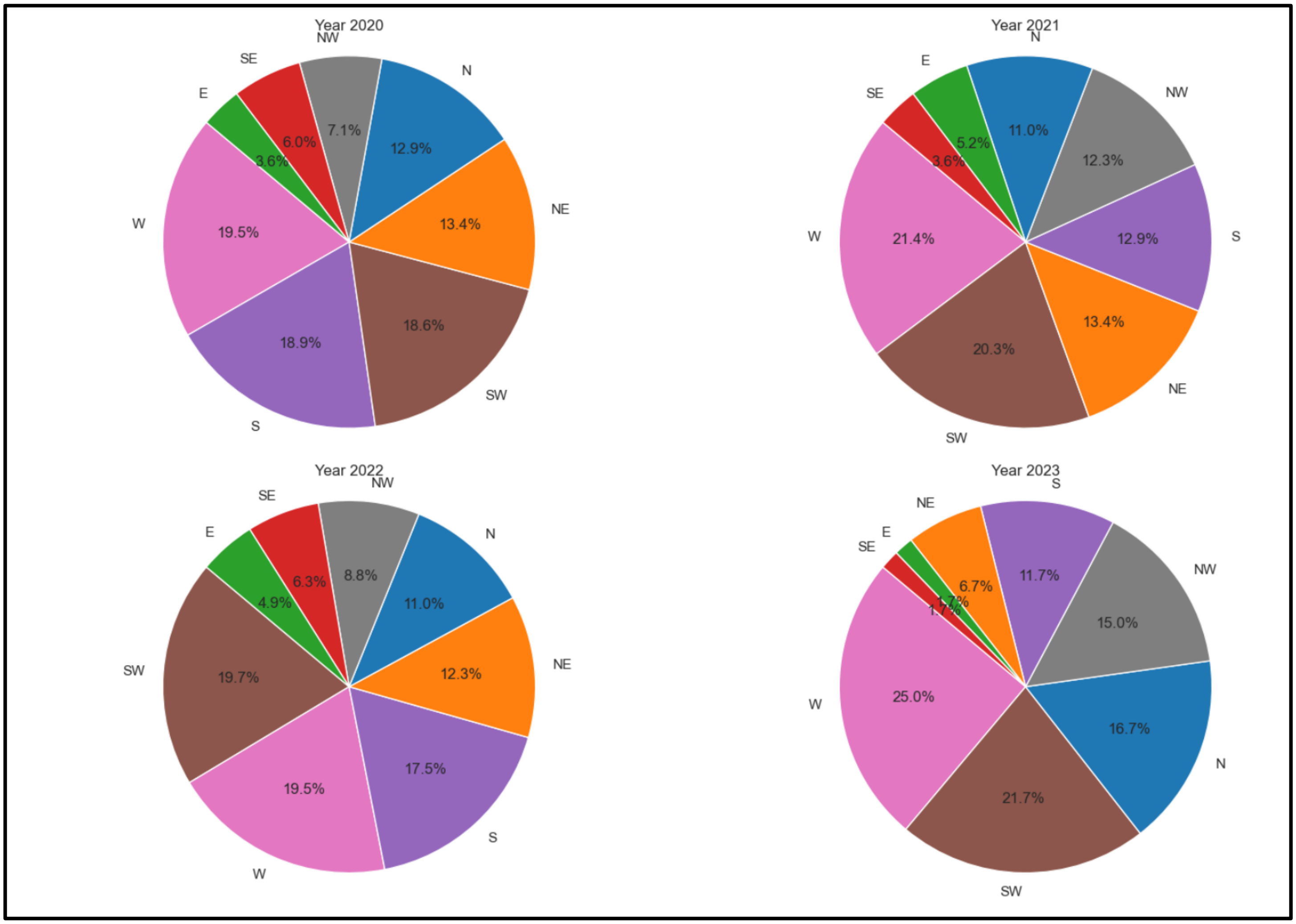
The first line plot detailing the yearly trend in average wind speed in Boston from 2013 to 2023 offers a nuanced perspective on atmospheric dynamics. It commences at 16.5 mph in 2013, witnessing a slight uptick to 17 mph in 2014 before experiencing a discernible downturn. Notably, in 2017, there is a remarkable surge, with wind speeds peaking at approximately 20 mph, indicative of heightened atmospheric activity. Subsequent to this peak, another downturn is evident, particularly noticeable in 2021. Post-2021, the plot stabilizes around 17.5 mph, implying a return to more typical wind conditions, albeit slightly elevated compared to earlier years.



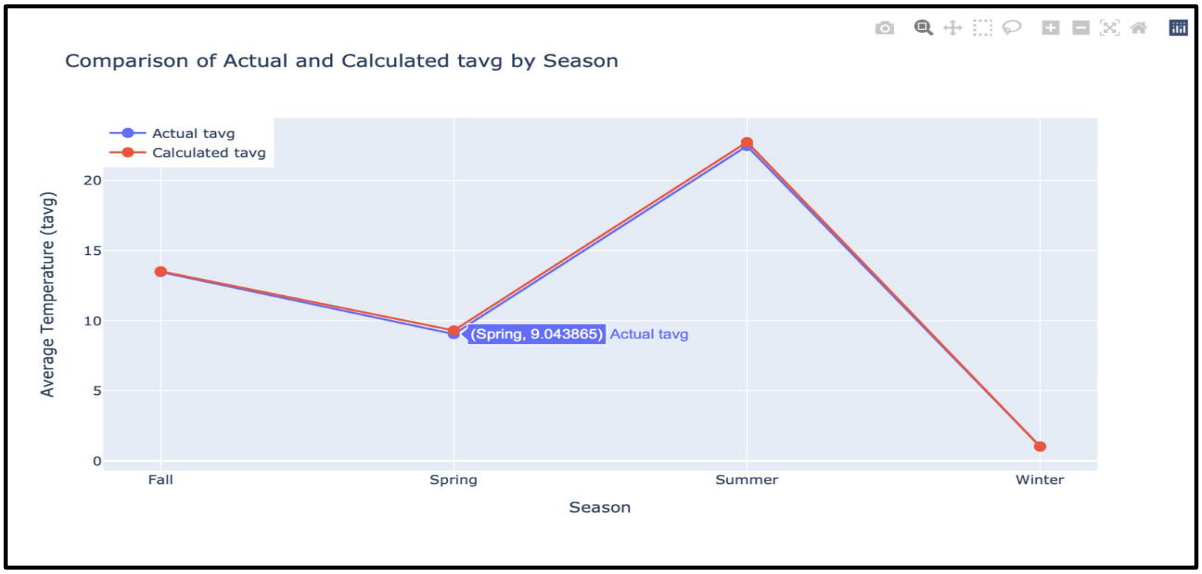
Moving on to the second plot, which traces the yearly trends in average temperature for Boston over the same period, starting at 12 degrees Celsius in 2013. The temperature remains relatively stable until 2021, suggesting a period of climatic equilibrium. However, an abrupt deviation occurs in 2023, with temperatures plummeting to 4 degrees Celsius. This sharp decline signifies either an anomalous weather event or a notable shift in larger climatic patterns, warranting further investigation into its underlying causes.



Lastly, the third plot illustrates the yearly trend in average precipitation levels, revealing the intricate variability in Boston's rainfall patterns. Commencing at 2.75 inches in 2013, a gradual decline ensues until 2016, indicating a diminishing trend in overall precipitation during this period. However, a significant departure from this trend occurs in 2018, characterized by a sudden upsurge in precipitation levels, soaring to 4 inches. This spike suggests the occurrence of a substantial rainfall event or a notable alteration in weather patterns. Following 2018, precipitation levels exhibit a rollercoaster-like fluctuation, oscillating between periods of heightened and diminished precipitation until 2023. These detailed insights provide a comprehensive understanding of the multifaceted dynamics observed in each plot, underscoring the intricate interplay of meteorological factors shaping Boston's weather landscape over the specified timeframe.

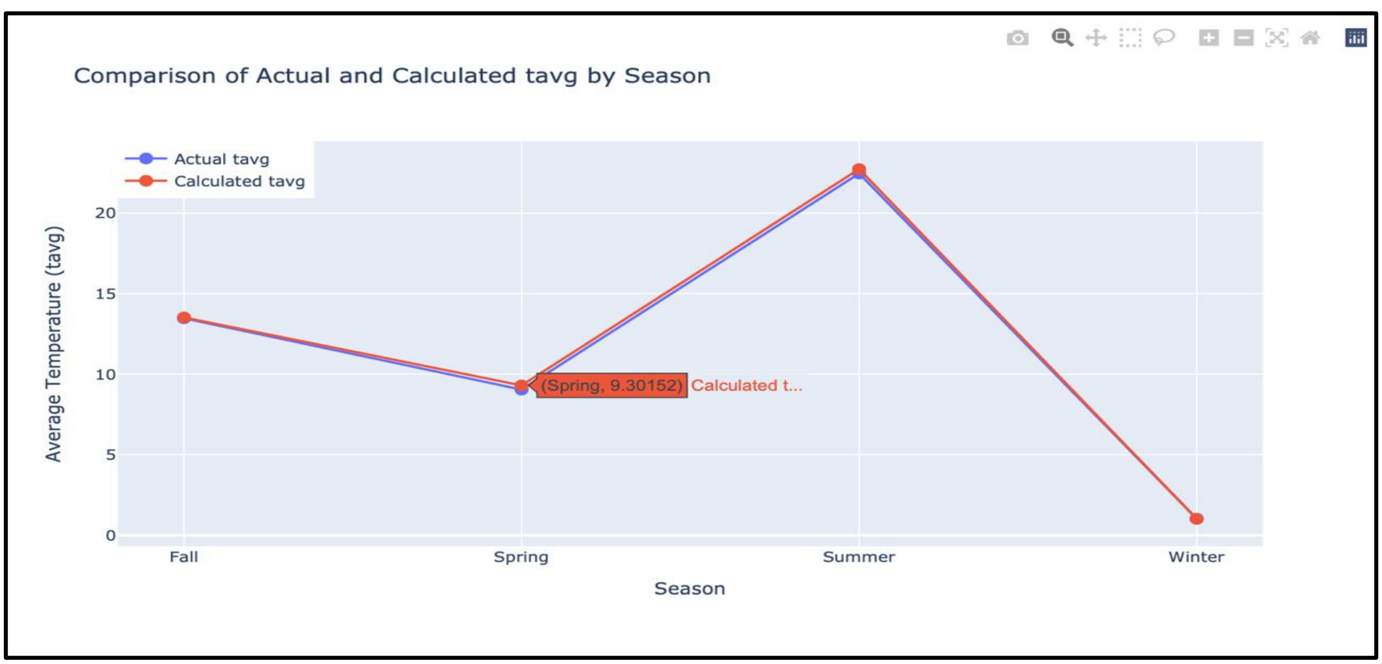


The third visualization presents pie charts representing the distribution of wind directions over the past four years. Each chart reveals notable shifts in wind patterns across different compass directions. In the southern and western directions, there are observable variations, albeit with distinct characteristics. The southern and western wind directions exhibit relatively stable distributions across the four years, with minimal fluctuations. However, in the northern direction, there is a significant increase in wind frequency, rising from 11% to 17% over the specified period. This pronounced surge in northern wind occurrences highlights a noteworthy shift in atmospheric dynamics, which could have implications for local weather patterns and climate conditions. Overall, the pie charts provide a clear visual representation of the evolving wind patterns over the past four years, underscoring the importance of monitoring and understanding these changes for various meteorological and environmental considerations.



The fourth visualization illustrates a comparison between the actual average temperature (tavg) and the calculated average temperature across different seasons: Spring, Summer, Fall, and Winter. Each data point represents the mean temperature for the respective season. Notably, while the actual average temperature across all seasons is recorded as 9.04, the calculated average temperature is slightly higher at 9.3. The graph reveals interesting insights into the accuracy of the calculated average temperature in relation to the actual values across different seasons. In general, the calculated temperature closely aligns with the actual temperature trends, exhibiting similar fluctuations and patterns. However, there are slight discrepancies between the two, particularly evident in certain seasons.

For instance, during the Fall season, the calculated average temperature tends to slightly exceed the actual temperature, indicating a potential overestimation. Conversely, in Spring and Winter, the calculated average temperature closely mirrors the actual temperature, suggesting a higher level of accuracy in these seasons. Overall, this visualization provides valuable insights into the effectiveness of the calculated average temperature in approximating the actual temperature trends across different seasons.



7. Model Exploration and Selection

Model Considerations – We proceeded to train various regression models to predict the target variable, average air temperature (tomavg). Our analysis included testing logistic regression, random forests, and gradient boosting regressors. Among these, the linear regression model emerged as the most suitable choice for our project. We chose linear regression because of its simplicity, interpretability, and efficiency in handling numerical data. Despite its straightforward nature, the linear regression model proved to be highly effective in capturing the underlying relationships between the predictor variables and the target variable. Moreover, since our dataset primarily consisted of numerical features, linear regression was well-suited to the task at hand. Additionally, linear regression offered valuable insights into the linear dependencies between the features and the target variable, allowing us to interpret the coefficients and understand the magnitude and direction of their impact on the average air temperature.

Linear Regression Model

```
[ ] from sklearn.model_selection import train_test_split
    from sklearn.linear_model import LinearRegression
    from sklearn.metrics import mean_squared_error
    from sklearn.metrics import r2_score, mean_absolute_error

    model = LinearRegression()
    model.fit(X_train, y_train)

    y_pred = model.predict(X_test)

    r_squared = r2_score(y_test, y_pred)
    mae = mean_absolute_error(y_test, y_pred)
    mse = mean_squared_error(y_test, y_pred)

    print("Mean Squared Error:", mse)
    print("R-squared:", r_squared)
    print("Mean Absolute Error:", mae)

    Mean Squared Error: 5.816303261712841
    R-squared: 0.9337660301972853
    Mean Absolute Error: 1.878176480026237
```

XG boost regressor Model

```
import xgboost as xgb

xgb_model = xgb.XGBRegressor()
xgb_model.fit(X_train, y_train)

y_pred_xgb = xgb_model.predict(X_test)

mse_xgb = mean_squared_error(y_test, y_pred_xgb)
r2_xgb = r2_score(y_test, y_pred_xgb)
mae_xgb = mean_absolute_error(y_test, y_pred_xgb)

print("Mean Squared Error:", mse_xgb)
print("R-squared:", r2_xgb)
print("Mean Absolute Error:", mae_xgb)

Mean Squared Error: 6.316726513304519
R-squared: 0.9280673901774837
Mean Absolute Error: 1.9526714188608605
```

Random Forest Regressor Model

```python
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error

rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

y_pred_rf = rf_model.predict(X_test)

mse_rf = mean_squared_error(y_test, y_pred_rf)
r2_rf = r2_score(y_test, y_pred_rf)
mae_rf = mean_absolute_error(y_test, y_pred_rf)

print("Mean Squared Error:", mse_rf)
print("R-squared:", r2_rf)
print("Mean Absolute Error:", mae_rf)
```

```
Mean Squared Error: 5.829089544459645
R-squared: 0.9336204246077495
Mean Absolute Error: 1.8896785225718193
```

Model Implementation - We implemented the models using Python's scikit-learn library. Key parameters tuned included the number of estimators for the Random Forest Regressor, set to 100 for optimal performance, and we utilized the default parameters for the XGBoost Regressor to capture inherent complexities within the data.

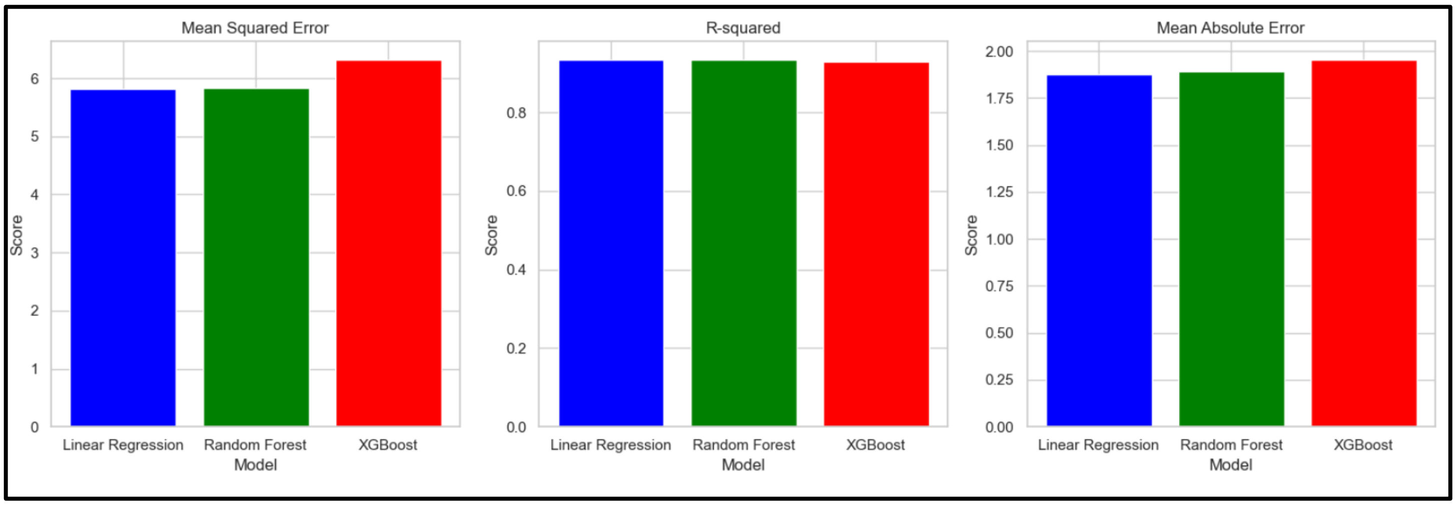
8. Model Performance Evaluation

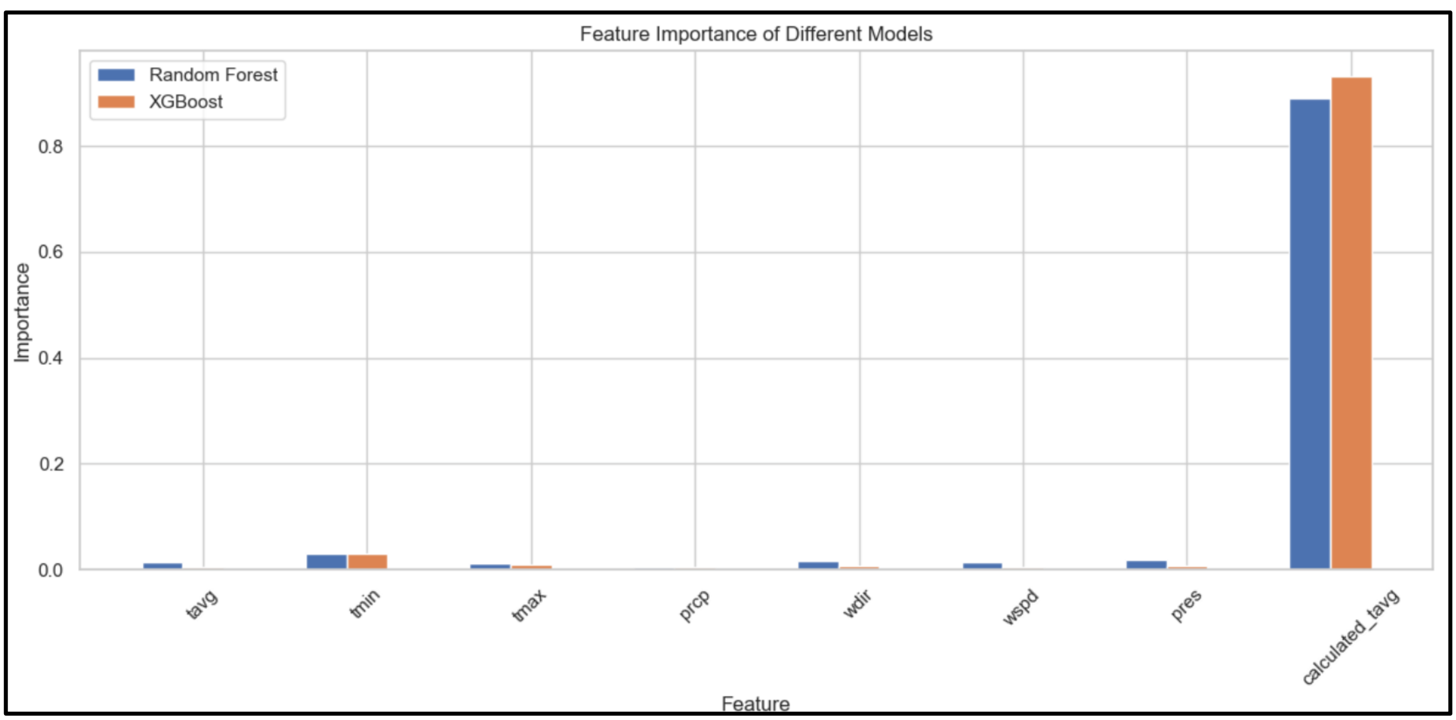
In this evaluation, we began by preparing our dataset, implementing various regression models, and assessing their performance metrics. The initial step involved creating a training dataset by shifting the 'tomavg' column and filling any missing values. Subsequently, we split the dataset into training and testing sets using a test size of 20%.

Three regression models were then trained and evaluated:

1. Linear Regression
2. Random Forest Regressor
3. XGBoost Regressor

For each model, we computed key metrics including Mean Squared Error (MSE), R-squared, and Mean Absolute Error (MAE) to quantify their predictive accuracy. Visualizations such as bar plots were utilized to compare the performance of these models across different metrics. Through visualizations like scatter plots and histograms, we gained insights into the accuracy and distribution of predicted values compared to actual values.





After the initial model testing, the results showed moderate performance

Linear Regression Model

Mean Squared Error: 5.816303261712841

R-squared: 0.9337660301972853

Mean Absolute Error: 1.878176480026237

Random Forest Regressor Model

Mean Squared Error: 5.829089544459645

R-squared: 0.9336204246077495

Mean Absolute Error: 1.8896785225718193

XG boost regressor Model

Mean Squared Error: 6.316726513304519

R-squared: 0.9280673901774837

Mean Absolute Error: 1.9526714188608605

To enhance the model's accuracy, additional predictors were incorporated. This involved implementing functions to calculate percentage differences and rolling averages for various weather attributes such as maximum temperature (tmax), minimum temperature (tmin), average temperature (tavg), and calculated average temperature (calculated_tavg) over different time horizons. Specifically, rolling averages were computed for horizons of 3 and 14 days, providing a more comprehensive dataset for improved predictive modeling.

```
 14  rolling_3_tmax                  3639 non-null   float64
 15  rolling_3_tmax_pct              3639 non-null   float64
 16  rolling_3_tmin                  3639 non-null   float64
 17  rolling_3_tmin_pct              3639 non-null   float64
 18  rolling_3_tavg                  3639 non-null   float64
 19  rolling_3_tavg_pct              3639 non-null   float64
 20  rolling_3_calculated_tavg       3639 non-null   float64
 21  rolling_3_calculated_tavg_pct   3639 non-null   float64
 22  rolling_14_tmax                 3639 non-null   float64
 23  rolling_14_tmax_pct             3639 non-null   float64
 24  rolling_14_tmin                 3639 non-null   float64
 25  rolling_14_tmin_pct             3639 non-null   float64
 26  rolling_14_tavg                 3639 non-null   float64
 27  rolling_14_tavg_pct             3639 non-null   float64
 28  rolling_14_calculated_tavg      3639 non-null   float64
 29  rolling_14_calculated_tavg_pct  3639 non-null   float64
dtypes: datetime64[ns](1), float64(25), int64(2), object(2)
memory usage: 853.0+ KB
```

After we included more predictors, we went back to train the dataset again, and this time, we saw better accuracy.

Linear Regression Model

Mean Squared Error: 5.427388862589086

R-squared: 0.9402155084302812

Mean Absolute Error: 1.813181159218727

Random Forest Regressor Model

Mean Squared Error: 5.894400571428572

R-squared: 0.9350712192929159

Mean Absolute Error: 1.8438763736263737

XG boost regressor Model

Mean Squared Error: 6.201930171038774

R-squared: 0.9316836785765936

Mean Absolute Error: 1.8879793859291618

In the end, we normalized the data, leading to the highest accuracy achieved by the model. The Linear Regression model exhibited the highest R-squared value of 94.02%, indicating its superior predictive capability. Further analysis involved visualizing the distribution of absolute differences between predicted and actual values, providing insights into model accuracy.

```
          time      tavg      tmin      tmax      prcp      wdir      wspd  \
14 2013-03-15  0.335329  0.333333  0.342685  0.000000  0.769444  0.374787
15 2013-03-16  0.383234  0.430233  0.330661  0.000000  0.843056  0.183986
16 2013-03-17  0.359281  0.375969  0.308617  0.000000  0.916667  0.270869
17 2013-03-18  0.309381  0.333333  0.254509  0.038462  0.552778  0.240204
18 2013-03-19  0.363273  0.408915  0.286573  0.439744  0.188889  0.563884

        pres  year  month_index  ... rolling_3_calculated_tavg  \
14  0.413675  2013            3  ...                  0.342561
15  0.429060  2013            3  ...                  0.299500
16  0.569231  2013            3  ...                  0.301423
17  0.764103  2013            3  ...                  0.284506
18  0.473504  2013            3  ...                  0.271819

   rolling_3_calculated_tavg_pct  rolling_14_tmax  rolling_14_tmax_pct  \
14                      0.320373         0.314410         4.989754e-14
15                      0.343173         0.312240         4.988045e-14
16                      0.318266         0.307178         4.984859e-14
17                      0.335530         0.299223         4.975928e-14
18                      0.321033         0.292171         4.982032e-14

    rolling_14_tmin  rolling_14_tmin_pct  rolling_14_tavg  \
14         0.359076         1.526711e-13         0.337580
15         0.354697         1.451904e-13         0.334298
16         0.345740         1.464346e-13         0.329666
17         0.334793         1.465754e-13         0.321753
18         0.330414         1.449095e-13         0.316927

    rolling_14_tavg_pct  rolling_14_calculated_tavg  \
14             0.294003                    0.341099
15             0.296147                    0.337826
16             0.295067                    0.330798
17             0.292303                    0.321267
18             0.295361                    0.315394

    rolling_14_calculated_tavg_pct
14                        0.657417
15                        0.657784
16                        0.657449
17                        0.656904
18                        0.657505

[5 rows x 30 columns]
```

Finally, we summarized the accuracy of each model using a heatmap, highlighting the MSE, R-squared, and MAE scores for comparison. The Linear Regression model emerged as the best performer, achieving an accuracy of 94.02%. Overall, this comprehensive evaluation process enabled us to identify the most effective regression model for our dataset and assess its predictive capabilities accurately.

Metrics	Mean Squared Error	R-squared	Mean Absolute Error
Linear Regression 1	5.816303	0.933766	1.878176
Linear Regression 2	5.427389	0.940216	1.813181
Linear Regression 3	5.427620	0.940213	1.813175
Random Forest Regressor 1	5.829090	0.933620	1.889679
Random Forest Regressor 2	5.894401	0.935071	1.843876
Random Forest Regressor 3	5.769103	0.936451	1.824979
XG Boost Regressor 1	6.316727	0.928067	1.952671
XG Boost Regressor 2	6.201930	0.931684	1.887979
XG Boost Regressor 3	6.259601	0.931048	1.904149

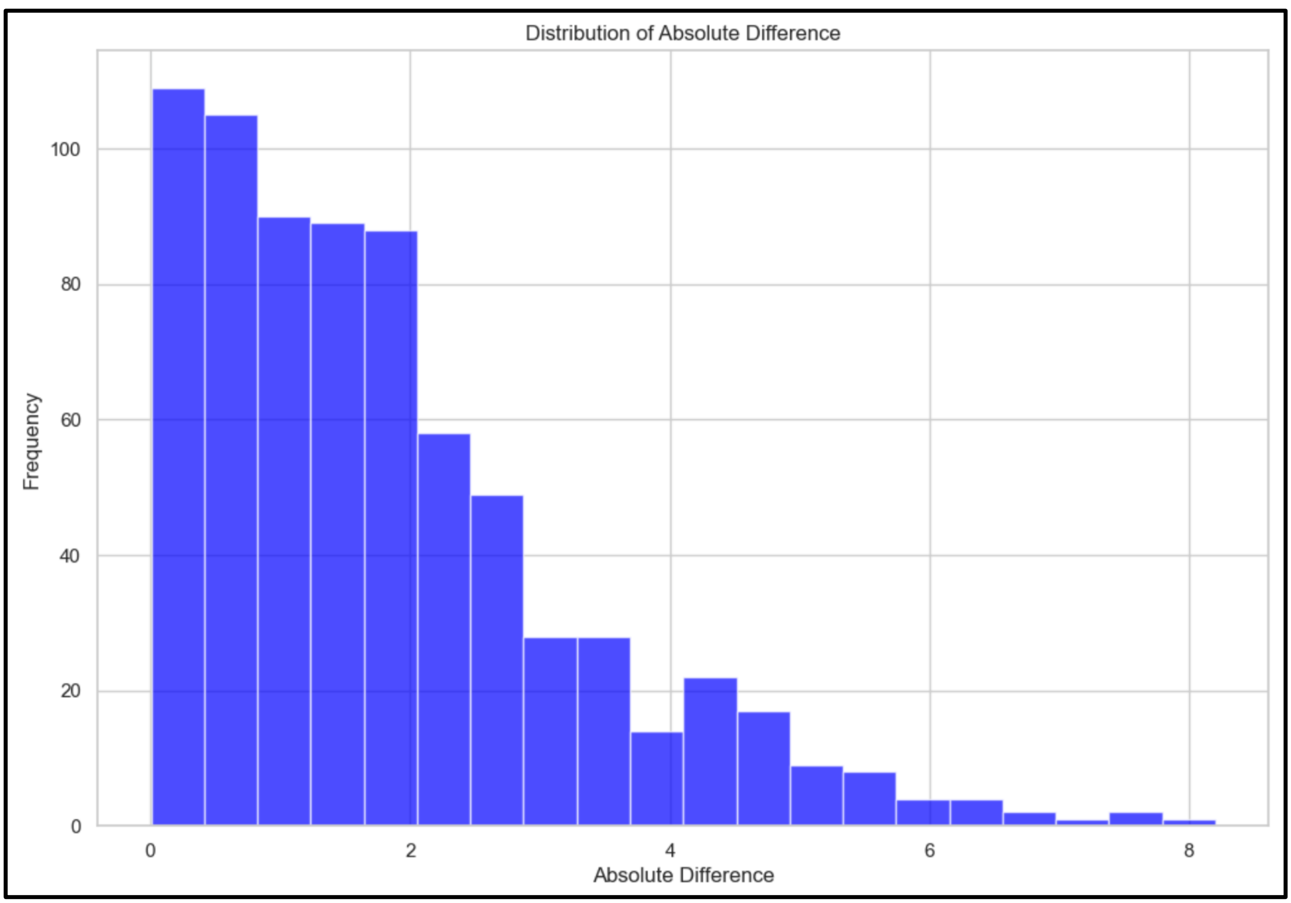
Linear Regression was the best model for our dataset

9. Performance Diagnostic and Visualization:

Diagnostic- After enriching our dataset with additional predictors and performing further preprocessing, including rolling averages and normalization, we retrained and evaluated our models. The linear regression model consistently demonstrated superior performance across multiple evaluation metrics. In our analysis, we employed a distribution of absolute differences to gauge the accuracy of the linear regression model in predicting Boston's average air temperature. The resulting R-squared value of 0.9402 indicated that approximately 94.02% of the variability in the target variable could be accounted for by the linear relationship with the predictor variables. This high level of explanatory power underscores the effectiveness of the model in capturing the underlying patterns within the dataset. Consequently, we quantified the accuracy of the model at 94.02%, affirming its robust performance in predicting average air temperatures in Boston. This finding highlights the model's reliability and its potential utility in informing decision-making processes and enhancing meteorological forecasts.

```
[ ]  r_squared = 0.9401977896073723
     accuracy_r2 = r_squared * 100
     print("Accuracy of the model is", accuracy_r2, "%")

     Accuracy of the model is 94.01977896073723 %
```

Visualization- The histogram illustrates the distribution of absolute differences between predicted and actual values. It provides insights into the accuracy of the predictions by showing how often the absolute differences fall within specific intervals. The frequency of occurrence is depicted on the y-axis, while the absolute difference intervals are represented on the x-axis.



10. Conclusion:

Summary: Our project successfully applied data mining techniques to forecast weather patterns in Boston, utilizing a decade of daily weather data. Through meticulous preprocessing, feature engineering, and model evaluation, we aimed to improve the accuracy of weather predictions. The findings revealed the superior performance of the linear regression model, with an impressive R-squared value of 0.94, indicating its ability to explain 94% of the variability in the average air temperature. Visualizations, such as bar charts and heatmaps, provided comprehensive insights into the comparative performance of different models across evaluation metrics.

Impact: The results of our project not only contribute to the advancement of meteorological forecasting techniques but also hold significant implications for various sectors and stakeholders. Improved weather predictions enable urban planners to better anticipate and mitigate the impacts of extreme weather events on city infrastructure, such as flooding and heatwaves. Emergency management agencies can utilize more accurate forecasts to enhance preparedness and response strategies, potentially saving lives and minimizing property damage during severe weather incidents. Moreover, industries heavily reliant on weather conditions, such as agriculture and renewable energy, stand to benefit from more precise forecasts. Farmers can make informed decisions about crop planting, irrigation, and pest management, optimizing yields and reducing resource wastage. Similarly, energy companies can better anticipate fluctuations in renewable energy generation, optimizing grid stability and reducing reliance on fossil fuels during peak demand periods.

Beyond immediate applications, the insights gained from our project contribute to a deeper understanding of Boston's weather dynamics and their interconnectedness with broader climate patterns. This knowledge can inform long-term planning efforts aimed at building resilience to climate change and its associated impacts. By fostering a proactive approach to climate adaptation and mitigation, our project plays a crucial role in safeguarding the well-being and sustainability of Boston's communities for generations to come.

Future Work: Moving forward, future research could explore more sophisticated ensemble techniques or delve deeper into feature engineering methods to further refine predictive models. Additionally, investigating the integration of real-time data sources and incorporating insights from climate change research could provide a more holistic understanding of Boston's weather dynamics. By continually refining and expanding upon our methodologies, we can strive towards even greater accuracy and reliability in weather forecasting, ultimately contributing to the resilience and well-being of Boston's inhabitants and infrastructure.

11. References:

Dataset : https://www.kaggle.com/datasets/swaroopmeher/boston-weather-2013-2023/data