**FLOODING PREDICTION IN LAGOS STATE, NIGERIA**

**INTRODUCTION**

Flooding is considered to be the most devastating natural disaster worldwide. Flooding is the excess water flowing into land which is usually dry. Peduzzi et al maintains that the rate of flood occurrence in recent times has been unprecedented, with 70 million living in flood-prone areas. Flooding in Lagos, Nigeria has been a significant issue affecting lives, infrastructure and the economy.

This report aims to analyze various factors influencing flooding events in Lagos State and develop a model to forecast future flood occurrence.

**CAUSES OF FLOODING IN LAGOS STATE**

* **Poor Drainage Systems:** The inadequate drainage network in Lagos exacerbates flooding. Refuse often accumulates in drainage channels due to improper waste disposal practices, blocking the flow of water and leading to inundated neighborhoods. Many structures in Lagos are also situated along drainage channels, further impeding the natural runoff of water during heavy rainfall.
* **Urban Planning Challenges:** The lack of comprehensive urban planning contributes significantly to flooding. Insufficient provision of designated dumpsites encourages residents to dispose of waste improperly, including in drainage systems and water bodies. This hampers the free flow of water and exacerbates flood risks across the city.
* **Infrastructure Development:** Rapid urbanization and infrastructure development in Lagos often occur without adequate consideration for flood mitigation measures. The construction of buildings and roads without proper drainage integration increases surface runoff during rainfall, contributing to localized flooding events.

**AREAS AFFECTED BY FLOODING**

* **Urban Areas:** Iyana-Oworo, Agege, Gbagada, and Lagos Island experience flooding due to prolonged periods of heavy rainfall, affecting roads and residential areas.
* **Outlying Areas:** Eredo, Bojije, Epe, Sangotedo, Ibeju-Lekki, Awoyaya, Labora, and Abijon are also prone to flooding during rainy seasons.

**DATA COLLECTION**

Data was sourced from Visual Crossing Weather site. It is a platform that provides detailed and accurate weather data. It caters for various needs, including historical weather data, real-time weather conditions and future forecasts.

Dataset contains a lot of information, but i only selected the relevant ones to avoid over complicating the dataset

Link to data source : <https://www.visualcrossing.com/weather/weather-data-services>

**KEY COLUMNS**

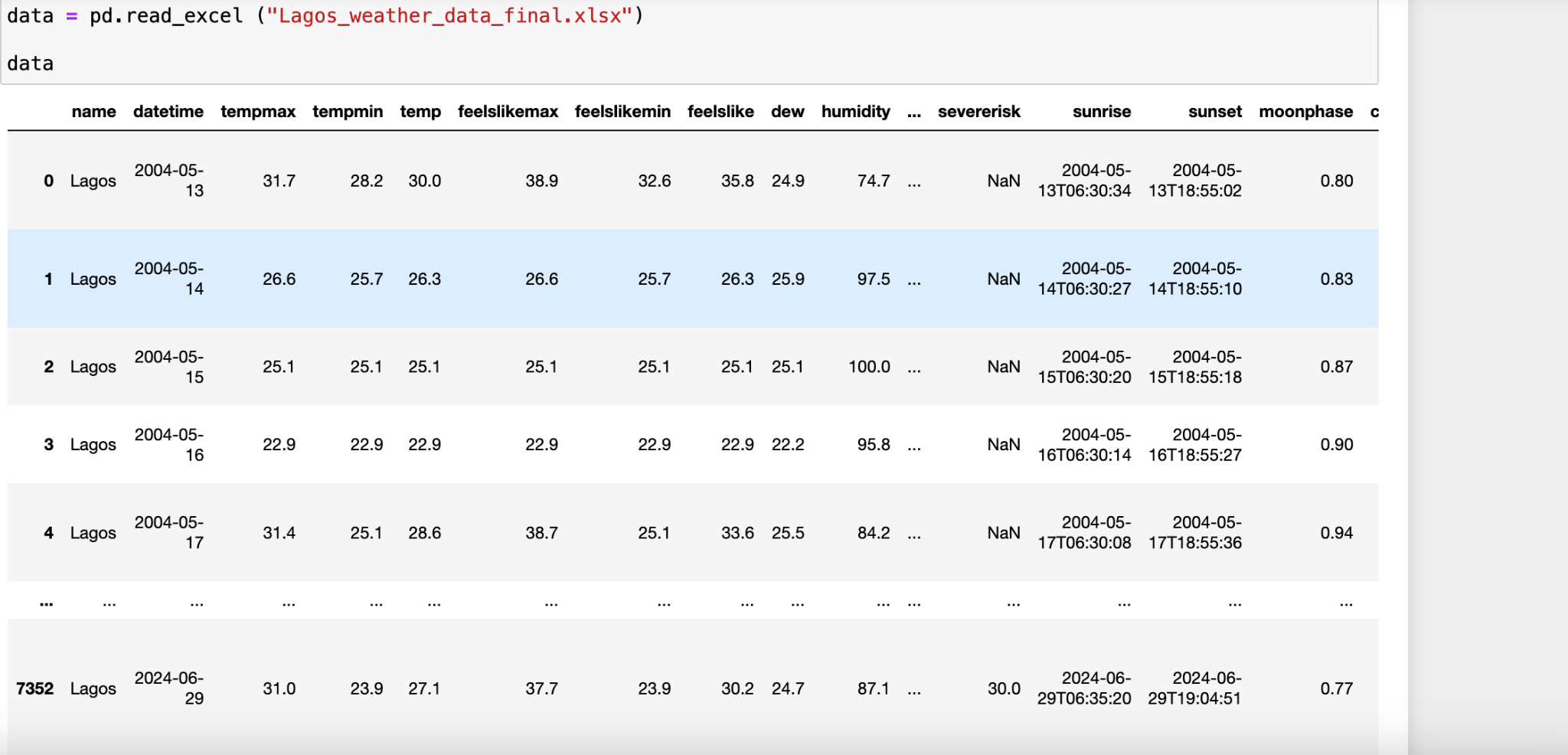
* **datetime:** The date and time of the weather observation.
* **tempmax:** The maximum temperature recorded for the day.
* **tempmin:** The minimum temperature recorded for the day.
* **temp:** The average temperature recorded for the day.
* **dew:** The dew point temperature, indicating the temperature at which air becomes saturated with moisture.
* **humidity:** The relative humidity percentage for the day.
* **precip:** The total precipitation (rainfall) amount for the day.
* **precipprob:** The probability of precipitation occurring during the day.
* **precipcover:** The percentage of the day that precipitation was observed.
* **preciptype:** The type of precipitation (e.g., rain, snow) observed.
* **visibility:** The average visibility distance for the day.
* **windspeed:** The average wind speed for the day.
* **sealevelpressure:** The atmospheric pressure at sea level for the day.
* **cloudcover:** The percentage of the sky covered by clouds.
* **icon:** A symbolic representation of the weather condition (e.g., clear, partly cloudy).
* **flood occurrence:** A binary indicator (0 or 1) of whether a flood occurred on that day.

**REASONS FOR SELECTION**

* **Direct Influence on Floods:** These columns directly measure or provide insights into factors such as precipitation, temperature, humidity, wind, and atmospheric pressure—key elements influencing flooding events.
* **Data Availability and Relevance:** These variables are commonly used in meteorological and hydrological studies to assess flood risk and understand weather patterns.
* **Impact on Model Accuracy:** Including these variables ensures that the flood prediction model considers the most critical environmental factors affecting flood occurrence.

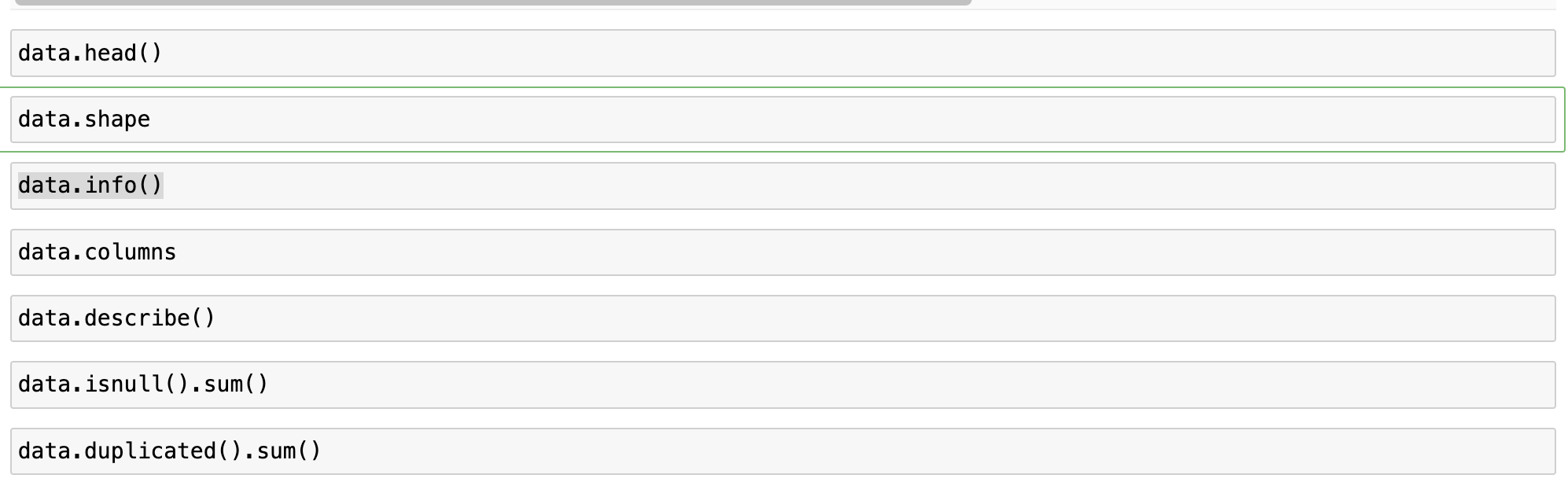
**IMPORTING DATASET**

To begin, the dataset needs to be imported. For this project, Pandas Library was used in Python to handle our data. Pandas is a powerful data manipulation library that offers data structures and functions needed to manipulate structured data seamlessly.



**DATA PREPROCESSING**

To begin understanding the dataset, basic exploratory data analysis techniques were employed. These included reviewing the dataset’s structure using methods such as : .head(), .info(), .columns, .describe(), and checking for null values and duplicates.

****

**BASIC INFORMATION ON DATASET**



**DATA QUALITY**

The dataset reveals a significant number of null values across various columns, indicating a need for data cleaning and imputation strategies.

**HANDLING MISSING VALUES**

**Imputation:**

* **Preciptype:** We investigated the missing values and their relationship with other variables. Given that 3681 out of 4017 missing values in “preciptype” were associated with "no precipitation," we filled these entries with "none." The remaining 336 values showed no clear pattern, so we opted to drop them to avoid introducing bias.
* **Flood Occurrence:** Missing values in the”flood occurrence” column were imputed with 0, indicating no flood event, aligning with the binary nature of this variable.
* **Visibility:** The ”visibility” column's missing values were filled with the mean value, ensuring the data's continuity without distorting its statistical properties.

**Dropping Rows:**

* **Temperature and Weather Variables:** Rows with missing values in tempmax, tempmin, temp, dew, humidity, precip, and precipprob were removed, ensuring that only complete records were used for analysis.
* **Preciptype:** The remaining 336 rows with missing “preciptype”values were dropped, as they did not show a discernible pattern that could be reliably imputed.
* **Icon and Sea Level Pressure:** Missing values in the “icon” and “sealevelpressure” columns were also dropped, as these features were deemed less critical for the initial analysis phase.

**FEATURE ENGINEERING**

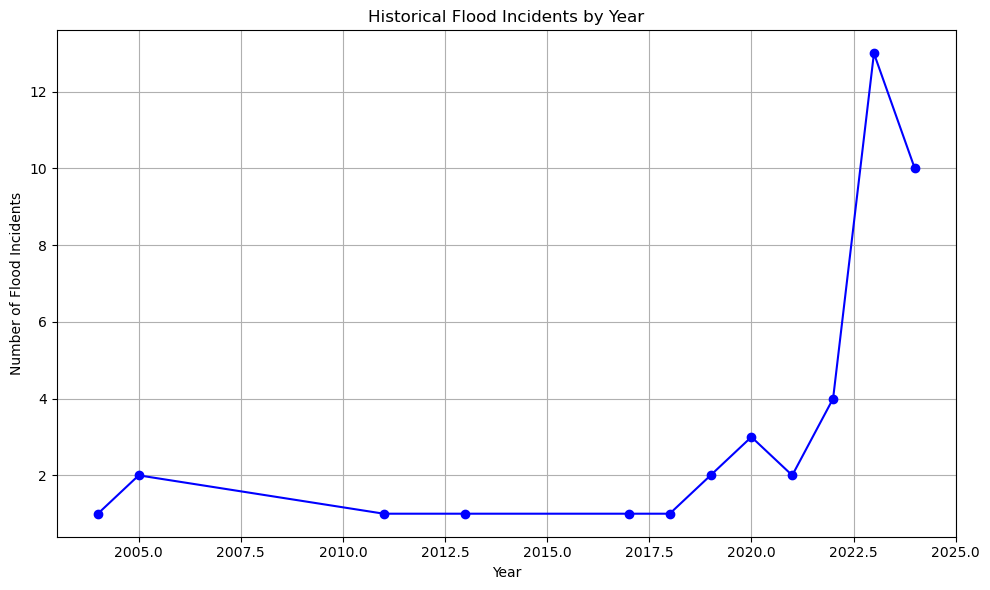
* **Date-Time Features:** The datetime column was split into year, month, and day for granular analysis.
* **Moon Phase:** The moon phase for each date was calculated and added as a new feature. This feature captures the phase of the moon (e.g. New Moon, First Quarter, Full Moon, Last Quarter) and is crucial for analyzing its potential impact on tidal patterns and flooding.
* **Categorical Encoding:** Categorical variables such as moon\_phase, precipitation\_type were encoded using one-hot encoding.
* **Scaling Features:** Numerical features were scaled to ensure they are on a similar scale, which helps improve the performance of machine learning algorithms.

**EXPLORATORY DATA ANALYSIS**

**FLOOD INCIDENCE OVER TIME**

A time series analysis was conducted to visualize the pattern of flood occurrences over the years and months.

* Flood incidence over the years

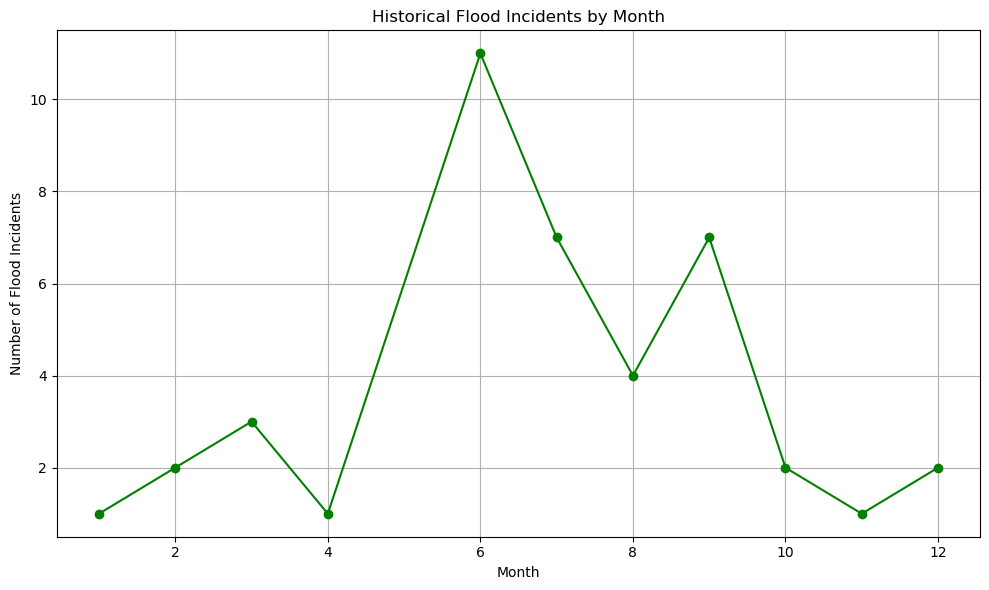


From the line chart showing historical flood incidents by year:

* **Increasing Trend**: There is a clear upward trend in the number of flood incidents over the years, particularly from 2020 onwards.
* **Peak Year**: The highest number of flood incidents occurred in 2023.

**Insight**: The increasing trend in flood incidents suggests a growing problem, possibly due to climate change or urbanization. This trend emphasizes the need for improved flood prediction and mitigation measures.

* Flood incidence over the month



From the line chart showing historical flood incidents by months:

* The month of June shows the highest number of flood incidents, indicating a peak in the rainy season and associated flooding risks.

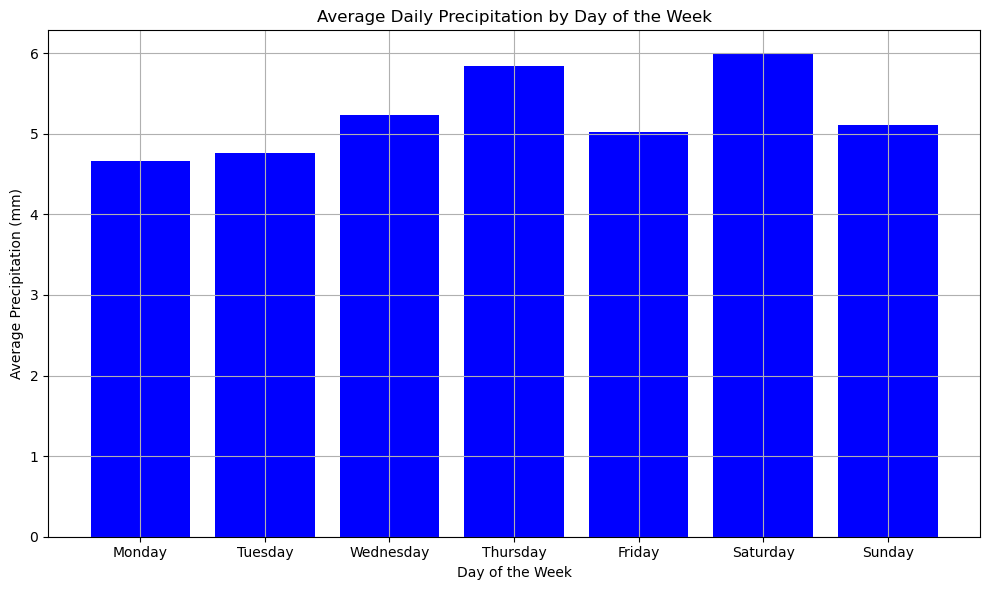
**AVERAGE PRECIPITATION ANALYSIS**

Analyzing the average precipitation for each day of the week.

From the bar chart showing average daily precipitation by day of the week:

* **Highest Precipitation**: Saturday and Thursday have the highest average precipitation, around 6 mm.
* **Lowest Precipitation**: Monday and Sunday have the lowest average precipitation, around 4.5 mm.

**Insight**: There are variations in average precipitation throughout the week, with some days experiencing higher rainfall. This information could be used to understand weekly patterns and prepare for days with higher precipitation risks.

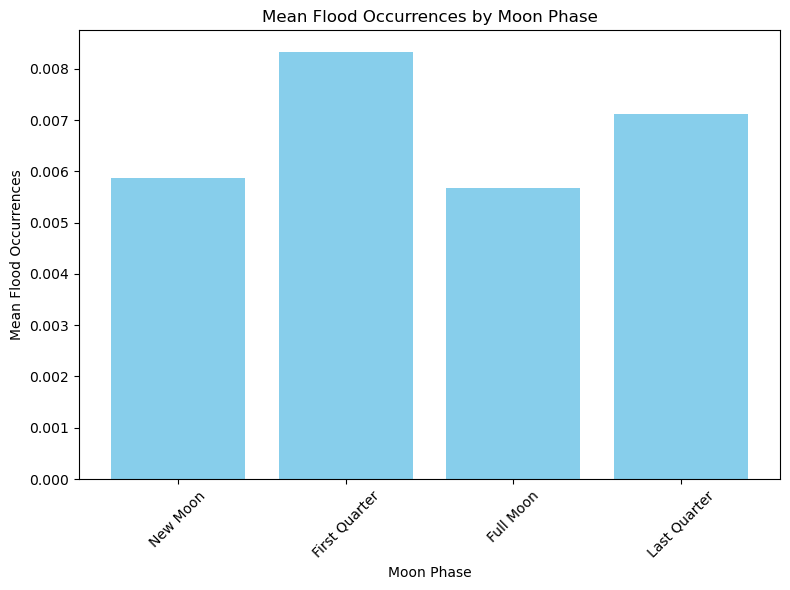


### 

**MOON PHASE ANALYSIS**

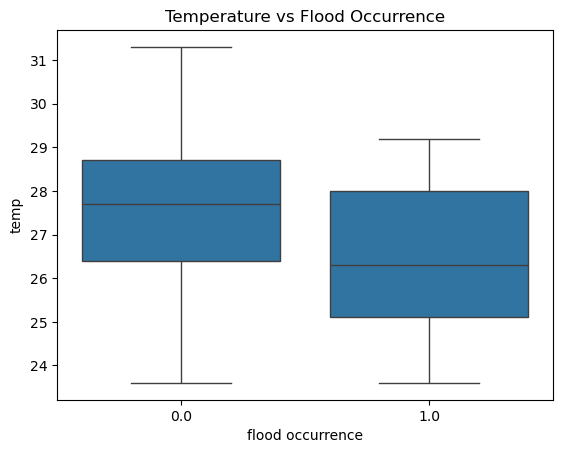
Examining the relationship between moon phases and flood occurrences.

Tidal Influence: Moon phases influence tidal patterns due to gravitational forces. Higher tides during new and full moons (spring tides) can increase the likelihood of coastal flooding, especially in combination with storm surges or heavy rainfall.



**INSIGHTS**

* The analysis showing higher mean flood occurrences during "First Quarter" and "Last Quarter" moon phases suggests that other factors beyond tidal effects may dominate flood occurrences in Lagos. These phases may coincide with regional climate patterns, river levels, or local drainage issues that contribute significantly to flood events.

**TEMPERATURE ANALYSIS** 

### 

### 

### **Temperature Ranges**: The boxplot shows the distribution of temperatures for days with and without flood occurrences. The temperature range on days with no flood occurrences (0.0) spans from approximately 24°C to 31°C, while for days with flood occurrences (1.0), it spans from approximately 24°C to 29°C.

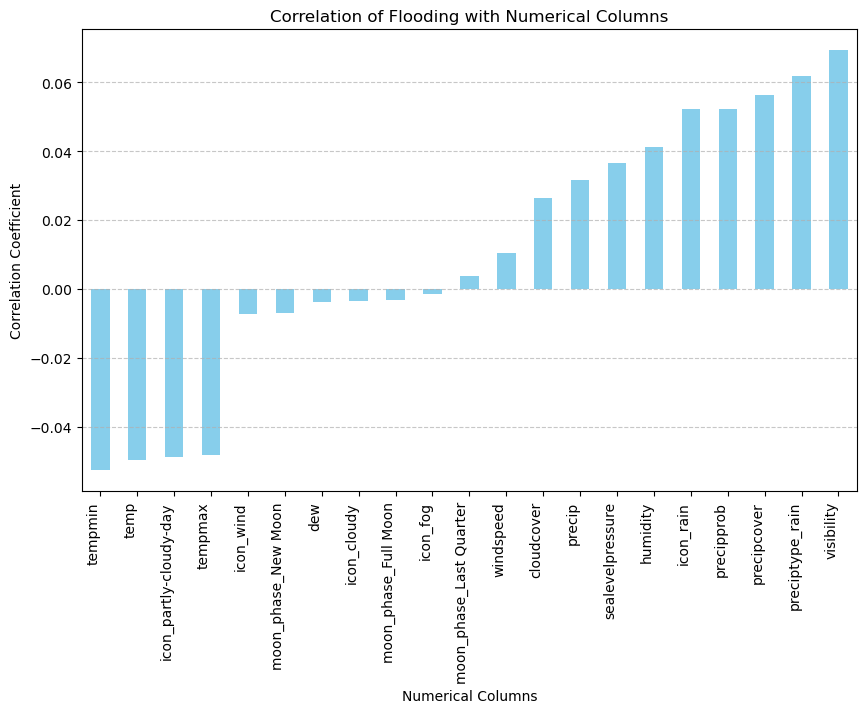
### **Median Temperature**: The median temperature on days without floods is around 28°C, while on days with floods, it is slightly lower, around 26°C.

### **Interquartile Range (IQR)**: The IQR for non-flood days is wider than for flood days, indicating more variability in temperatures on non-flood days. Specifically, the IQR for non-flood days is between 27°C and 29°C, while for flood days, it is between 25°C and 27.5°C.

### **Temperature Distribution**: There are some higher temperature values on non-flood days, as indicated by the upper whisker reaching up to 31°C. However, flood days do not exhibit such high temperatures, with the upper whisker reaching up to about 29°C.

From the box plot showing temperature versus flood occurrence:

* **Non-flooding Days**: The median temperature is around 28°C, with a range from approximately 24°C to 31°C.
* **Flooding Days**: The median temperature is slightly lower, around 26°C, with a range from approximately 24°C to 29°C.

**CORRELATION OF FLOODING OCCURRENCE WITH NUMERICAL COLUMNS**

**Insights:**

* **Positive Correlations:**
  + **Visibility:** This column shows the highest positive correlation with flood occurrence. This might indicate that as visibility improves (e.g., during or after a storm), the likelihood of recording flood incidents increases.
  + **Preciptype\_rain and Precipcover:** These columns also show strong positive correlations with flood occurrences, which is expected as rainfall directly contributes to flooding.
  + **Precipprob and Precip:** Probability of precipitation and the amount of precipitation are positively correlated, reinforcing the direct impact of rain on flood occurrences.
  + **Humidity, Sealevelpressure, Cloudcover, Windspeed:** These environmental factors are also positively correlated, suggesting they play a role in the dynamics of flooding.
* **Negative Correlations:**
  + **Tempmin, Temp, Tempmax:** Temperature-related columns have negative correlations with flood occurrences, indicating that higher temperatures might be associated with lower chances of flooding, possibly due to higher evaporation rates or seasonal weather patterns.
  + **Moon Phases and Weather Conditions (e.g., "Partly Cloudy Day"):** Certain moon phases and specific weather conditions show negative correlations, suggesting they might have a less direct or inverse relationship with flood occurrences.

**KEY INSIGHTS**

* There is a clear upward trend in the number of flood incidents over the years, particularly from 2020 onwards.
* Flooding is likely to occur in these months of the year (June - September)
* Lower temperatures are slightly more associated with flood occurrences.
* Certain days of the week, particularly Saturdays and Thursdays, experience higher average precipitation.
* There is a concerning upward trend in flood incidents over recent years, highlighting the increasing importance of flood prediction and management strategies.
* Rainfall and Visibility are crucial indicators for predicting flood occurrence

**MODELING**

Flooding is a significant environmental challenge in Lagos State, Nigeria, with severe impacts on lives, property, and the economy. Accurate prediction of flood occurrences can aid in effective disaster preparedness and management.

This report outlines the process of building a machine learning (ML) model to predict the occurrence of flooding in Lagos State using weather and environmental data sourced from Visual Crossing Weather.

### **Model Selection**

Several machine learning models were evaluated to predict flood occurrences. The models and their hyperparameters included:

* **Logistic Regression**: Evaluated with different regularization strengths (C parameter).
* **Random Forest Classifier**: Evaluated with different numbers of trees (n\_estimators) and maximum depths (max\_depth).
* **Gradient Boosting Classifier**: Evaluated with different learning rates, numbers of trees (n\_estimators), and maximum depths (max\_depth).

**ENCODING CATEGORICAL DATA**

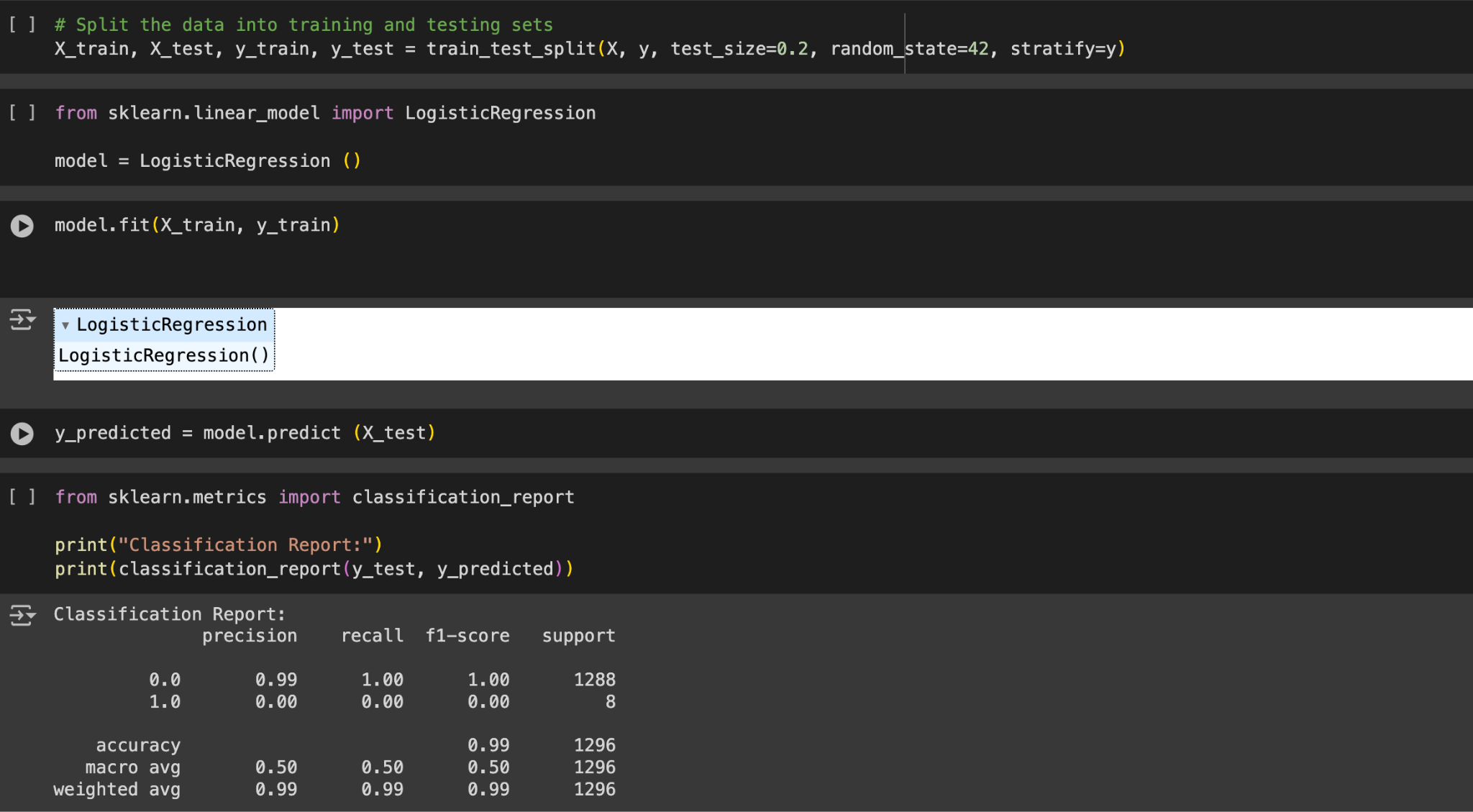
For precitype, icon and moon\_phase, one hot encoding was used to encode the data. One Hot Encoding is used when categorical variables have no inherent order or hierarchy.

**MODEL BUILDING AND EVALUATION**

The dataset was prepared by defining the X (features) and y (target). The data was split into training and test sets using a 0.2 test size ratio via the “train\_test\_split from sklearn.model\_selection.

Logistic regression was employed to train the model on the X\_train and y\_train datasets. Predictions were generated using the trained model to obtain y\_predicted.

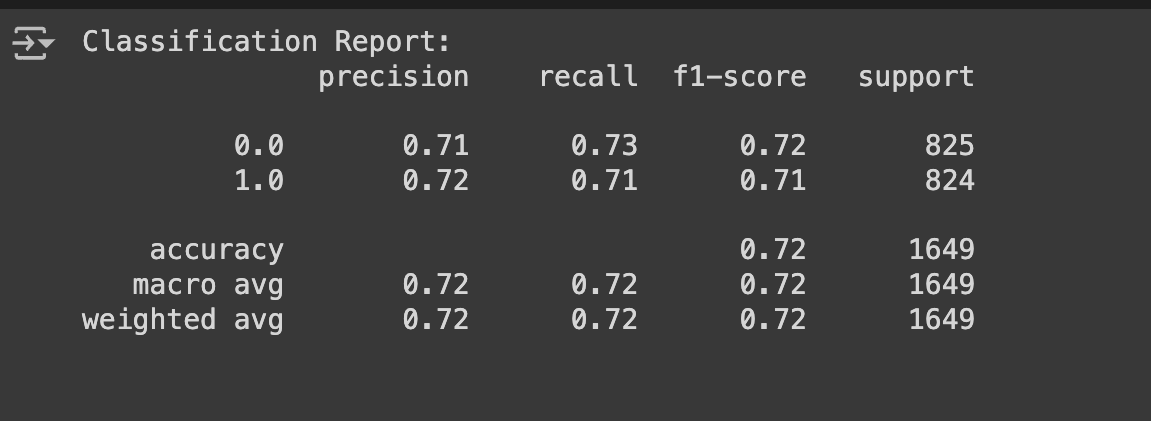
Upon checking the classification report, the logistic regression model is performing very well in predicting when flood occurrences do not happen (class 0), but it struggles significantly with predicting when floods do occur (class 1).



**ADDRESSING CLASS IMBALANCE WITH SMOTE**

There might be a severe imbalance between the number of instances where floods occur (class 1) versus where they do not (class 0). This imbalance can lead to the model being biased towards predicting the majority class (no flood occurrence). I addressed the class imbalance with SMOTE.

This approach helps to balance the distribution of classes in training data, thereby reducing the model’s tendency to favor the majority class. The aim is to improve the model’s ability to accurately predict flood occurrence in Lagos state.



Overall, the results suggest that the model's performance has improved after addressing class imbalance using SMOTE. However, further evaluation and fine tuning may be necessary to enhance the model’s predictive capabilities further.

**FURTHER EVALUATION AND MODEL FINE TUNING**

After addressing the initial challenge with class imbalance using SMOTE, the next step involves comprehensive evaluation and refinement of different models along with their parameters. This approach aims to enhance the predictive accuracy of flood occurrences in Lagos State.

**Model Selection and Parameter Optimization:**

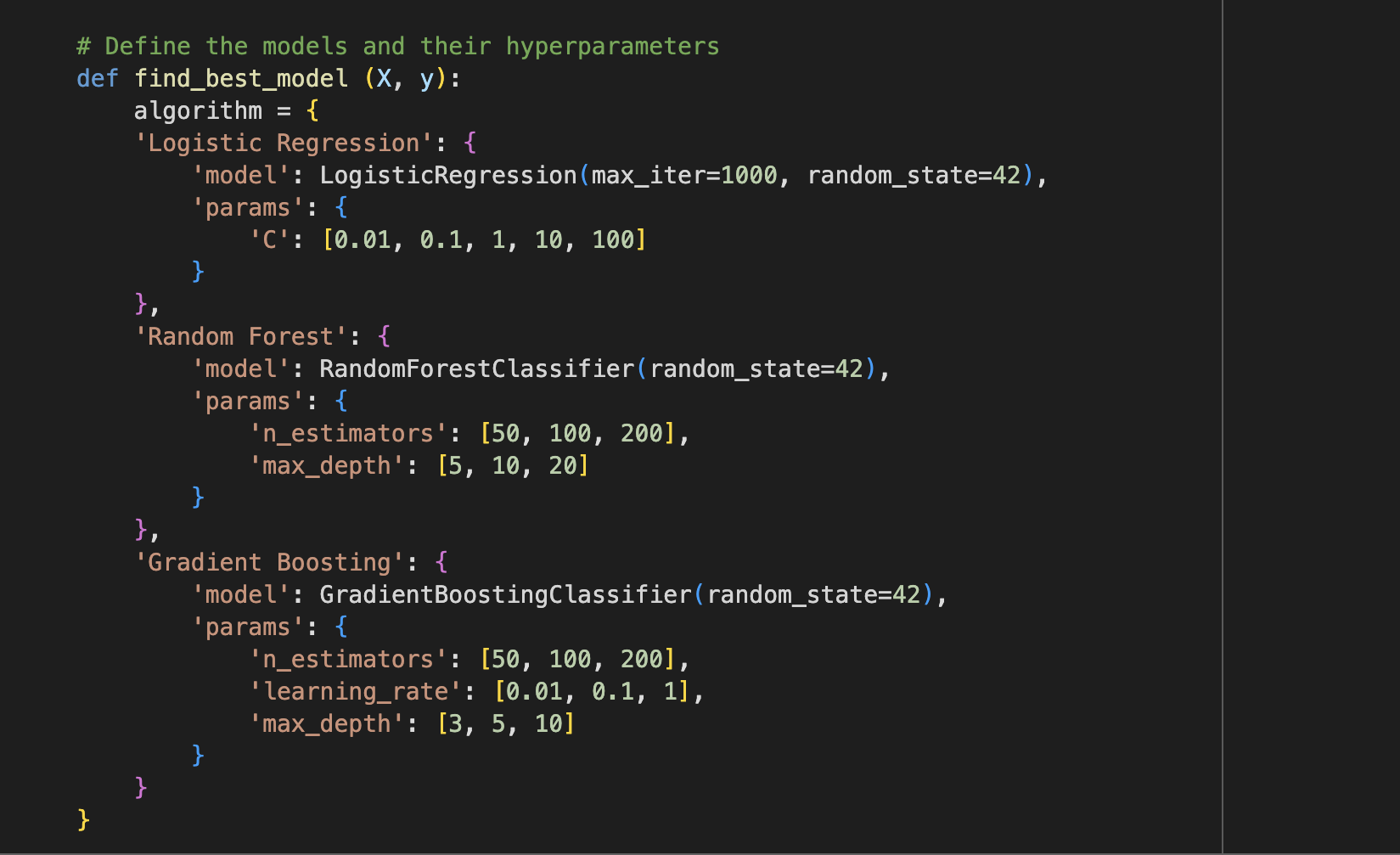
* Three classifiers were chosen for evaluation: Logistic Regression, Random Forest, and Gradient Boosting. Each model will be tuned by systematically adjusting their parameters to identify the optimal settings that maximize predictive performance for flood occurrence in Lagos State.

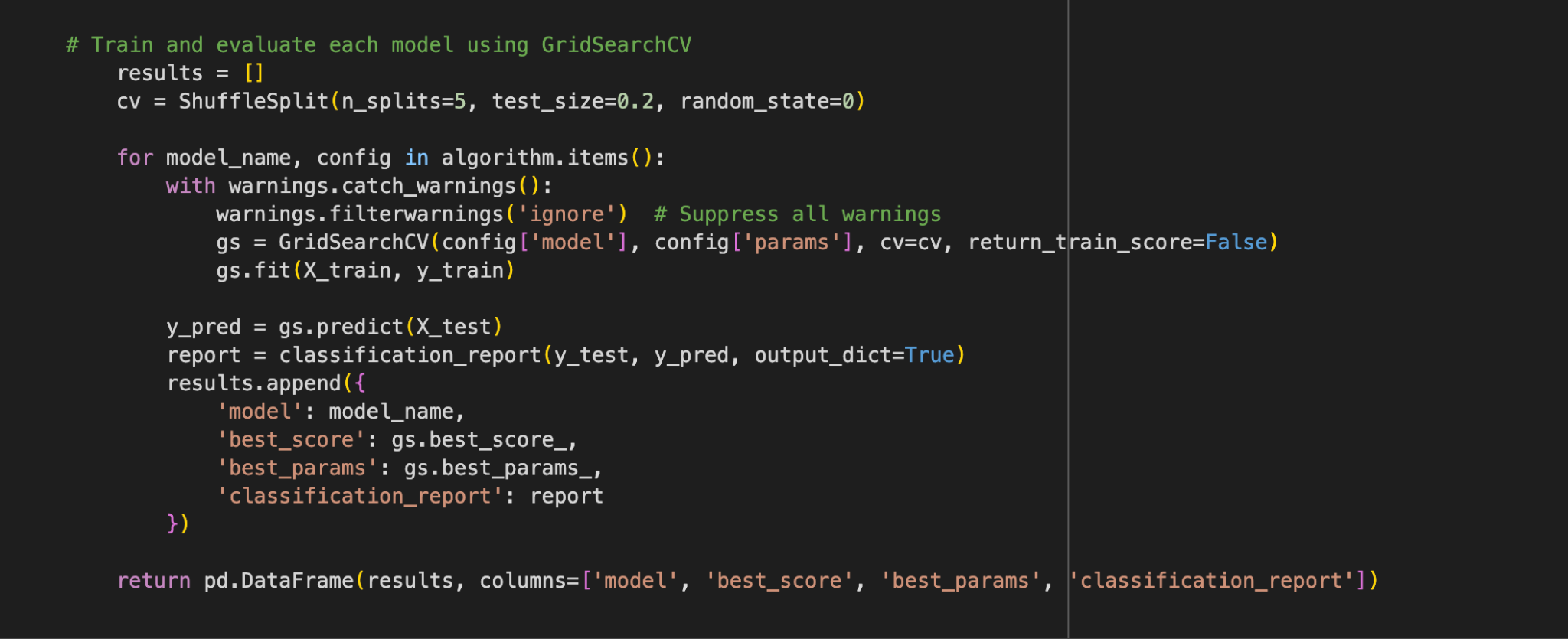
**Hyperparameter Tuning:**

* Hyperparameters for each model will be systematically tuned using Grid Search. This methods explore a predefined set of hyperparameters to find the combination that yields the best model performance.

**Evaluation Metrics:**

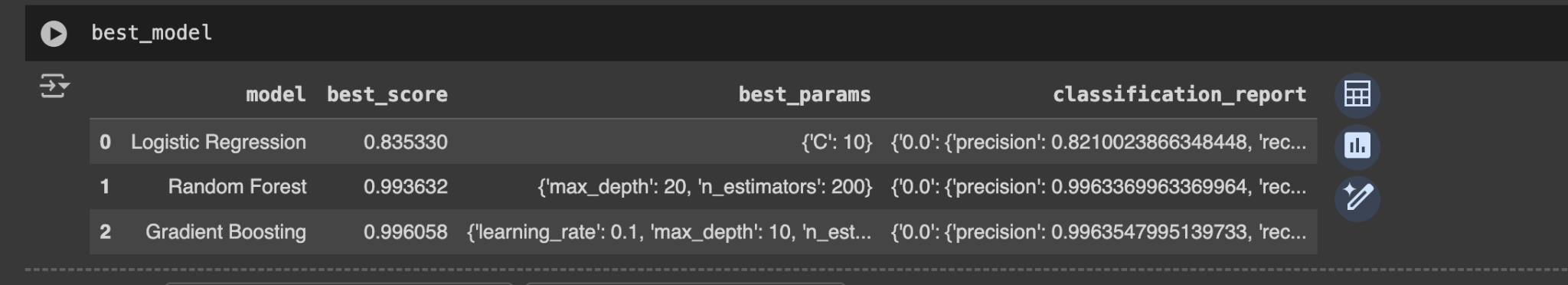
* Evaluation will go beyond accuracy, incorporating metrics such as precision, recall, F1-score, and area under the ROC curve (AUC). These metrics offer a comprehensive assessment of each model’s ability to predict flood occurrences accurately and reliably across different scenarios.

****

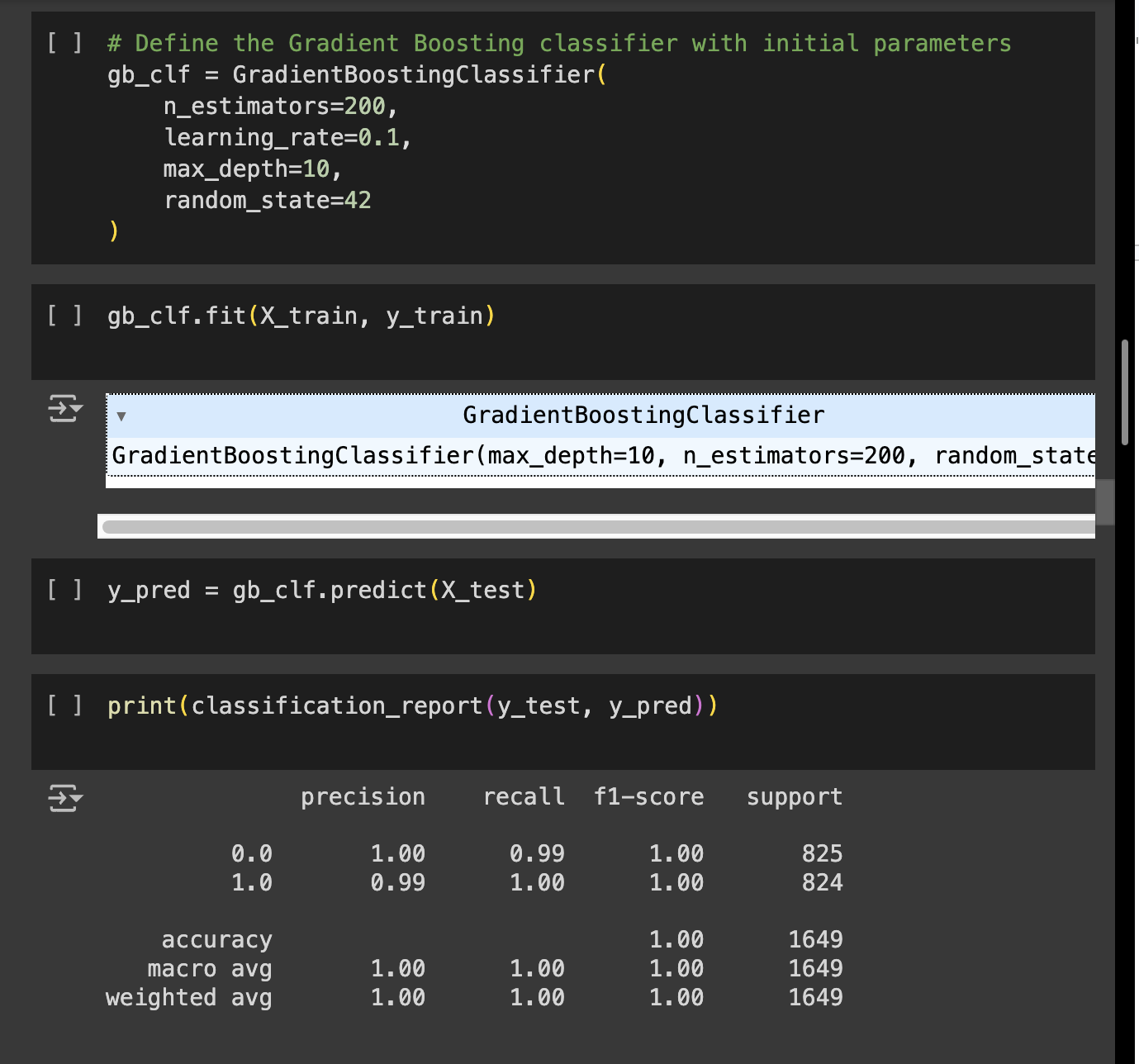
****

**RESULTS**

After thorough evaluation and parameter optimization, the Gradient Boosting Classifier demonstrated superior performance compared to Logistic Regression and Random Forest models.



**Application of Gradient Boosting Classifier with Optimized Parameters:**

The Gradient Boosting Classifier was applied using the best-tuned parameters identified during the evaluation process.

**Justification of Flood Prediction in Lagos: A 15-Day Analysis**

The predictions derived from the optimized Gradient Boosting Classifier were meticulously examined over a 15-day period, spanning from July 6, 2024, to July 20, 2024. Throughout this timeframe, no floods were forecasted for Lagos, Nigeria. This notable outcome underscores the robustness of the model in interpreting current weather conditions and historical data trends. It is reasoned that the likelihood of a flood occurring in Lagos within the immediate time frame is low.

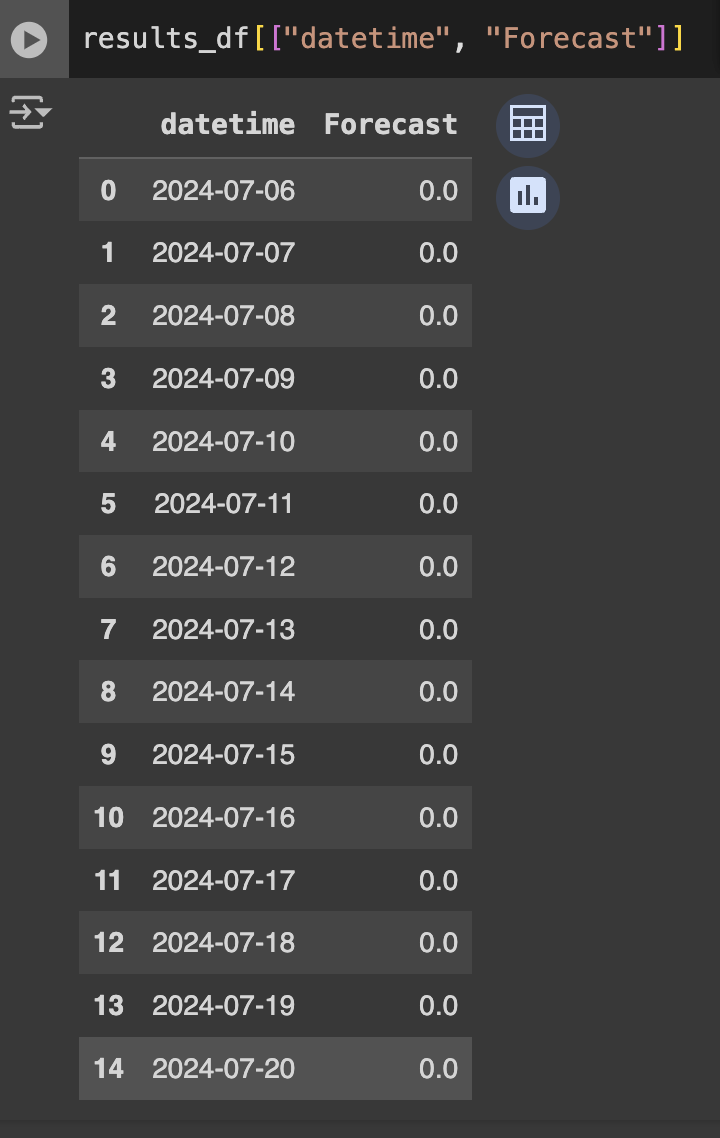
**KEY METRICS OF DATASET**

* **Dates Covered:** 06/07/2024 - 20/07/2024
* **Temperature Range:** Max: 29.0°C, Min: 24.4°C
* **Humidity:** Consistently high, around 80-86%
* **Precipitation:** Varies with some days indicating rain
* **Visibility:** Generally high, above 10 km

#### **Justification for No Flood Prediction**

1. **Weather Conditions Analysis:**
   * **Precipitation:** Although there were days with rain, the precipitation levels were not extreme. Historical data shows that significant flooding in Lagos typically follows heavy and sustained rainfall, which was not present in the forecasted period.
   * **Humidity and Cloud Cover:** High humidity and cloud cover are common in Lagos and do not necessarily lead to flooding unless combined with other severe weather conditions.
   * **Visibility:** High visibility throughout the forecast period indicates clear conditions that are generally not associated with flood events.
2. **Model Insights and Correlations:**
   * **Correlation with Flooding:** The model identified key weather variables that correlate with flood occurrences. Visibility had the highest positive correlation with no flooding, indicating clearer days are less likely to result in floods.
   * **Historical Patterns:** Analysis of historical flood incidents revealed that flooding is more common in certain months and weather conditions. The current forecasted period did not align with these high-risk conditions.
3. **Model Performance Validation:**
   * The Gradient Boosting Classifier has been rigorously tested and validated, showing high accuracy in predicting flood events. The absence of predicted flood events during this period aligns with the model’s performance metrics and historical validation.
4. **Environmental Factors:**
   * **Drainage Infrastructure:** While Lagos faces challenges with drainage infrastructure, the forecasted weather conditions do not indicate a strain that would lead to flooding.

Looking forward, continuous refinement of predictive models and integration of real-time data streams hold promise for enhancing our capability to anticipate and mitigate flood risks in Lagos.



**LIMITATION TO PROJECT**

* **Access Issues:** Obtaining and accessing relevant datasets was challenging, impacting the comprehensiveness of the analysis.
* **Missing Data:** Significant missing values required imputation or row deletion, potentially introducing biases.

**POSSIBLE SUGGESTIONS**

* **Improve Drainage Systems:** Enhance and maintain primary and secondary drainage channels to prevent blockages.
* **Urban Planning:** Implement stricter regulations on building codes and land use to prevent construction in flood-prone areas.
* **Waste Management:** Increase public awareness and infrastructure for proper waste disposal to avoid drainage obstructions.

**CONCLUSION**

In conclusion, while the current forecast brings reassurance of stability, ongoing vigilance and adaptive strategies are essential to safeguarding communities and infrastructure from the impact of potential flood events in Lagos State.

**REFERENCE**

1. <https://thescipub.com/abstract/10.3844/ajessp.2015.157.166>
2. [A review and critical analysis of the efforts towards urban flood risk](https://www.researchgate.net/publication/292995370_A_review_and_critical_analysis_of_the_efforts_towards_urban_flood_risk_management_in_the_Lagos_region_of_Nigeria?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ) [Brian Baily](https://www.researchgate.net/profile/Brian-Baily?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ)[management in the Lagos region of Nigeria](https://www.researchgate.net/publication/292995370_A_review_and_critical_analysis_of_the_efforts_towards_urban_flood_risk_management_in_the_Lagos_region_of_Nigeria?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ) by [Ugonna Nkwunonwo](https://www.researchgate.net/profile/Ugonna-Nkwunonwo?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ), [Malcolm Whitworth](https://www.researchgate.net/profile/Malcolm-Whitworth?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ), : https://www.researchgate.net/figure/Main-causes-of-urban-flooding-in-the-Lagos-area-of-Nigeria-showing-global-climate-change\_fig4\_292995370
3. <https://punchng.com/lagos-flooding-fg-begins-drainage-clearing-on-federal-roads/#:~:text=Other%20flooded%20places%20in%20Lagos,experiencing%20widespread%20submergence%20from%20floodwaters>.
4. [Land Use and Land Cover Change Assessment in the Context of Flood Hazard in Lagos State, Nigeria](https://www.researchgate.net/publication/350941954_Land_Use_and_Land_Cover_Change_Assessment_in_the_Context_of_Flood_Hazard_in_Lagos_State_Nigeria?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ) by Dorcas Idowu, Wendy Zhou: [https://www.researchgate.net/figure/List-of-recorded-flood-events-in-various-locations-within-Lago](https://www.researchgate.net/figure/List-of-recorded-flood-events-in-various-locations-within-Lago*)s

**LINK TO RESOURCES**

1. <https://drive.google.com/drive/folders/1kvo7ttoXE6TCUQHhGlZv2191uY15DPgO?usp=drive_link>