WEATHER FORECAST FOR OUTDOOR ACTIVITIES

By Huynh Anh Thu.

OVERVIEW

1 PROBLEM DEFINITION OF PROJECT

2 REQUIREMENT SPECIFICATION

3 EXPLORATORY DATA ANALYSIS

4 MACHINE LEARNING MODEL

5 CONCLUSION

1. PROBLEM DEFINITION OF PROJECT

Introduction

As a business specializing in providing products for picnics, BBQs and camping, we fully understand the importance of creating memorable experiences. We offer everything our customers need to make their outdoor gatherings perfect.

We are committed to delivering a complete and comfortable experience for our customers, no matter the weather conditions.

1. PROBLEM DEFINITION OF PROJECT

Problems

One of the factors that determines the success of a picnic is the weather.

Goals

- Analyzing weather patterns to provide customers with the best recommendations on timing and the necessary preparations for their outdoor activities.
- Providing products that meet all needs and weather conditions.
- So that we can have a weather-appropriate business strategy.

2. REQUIREMENT SPECIFICATION

Specific Requirements

Requirement 1:

How do various weather patterns (e.g., clear skies, overcast, rainy) affect the overall success and enjoyment of a picnic?

Requirement 2:

How can predictive models be built and evaluated to recommend ideal picnic days based on weather conditions?

Requirement 3:

How to use weather data analysis to make recommendations that deliver the best customer experience and create appropriate business strategies?

3. EXPLORATORY DATA ANALYSIS

Initial Data

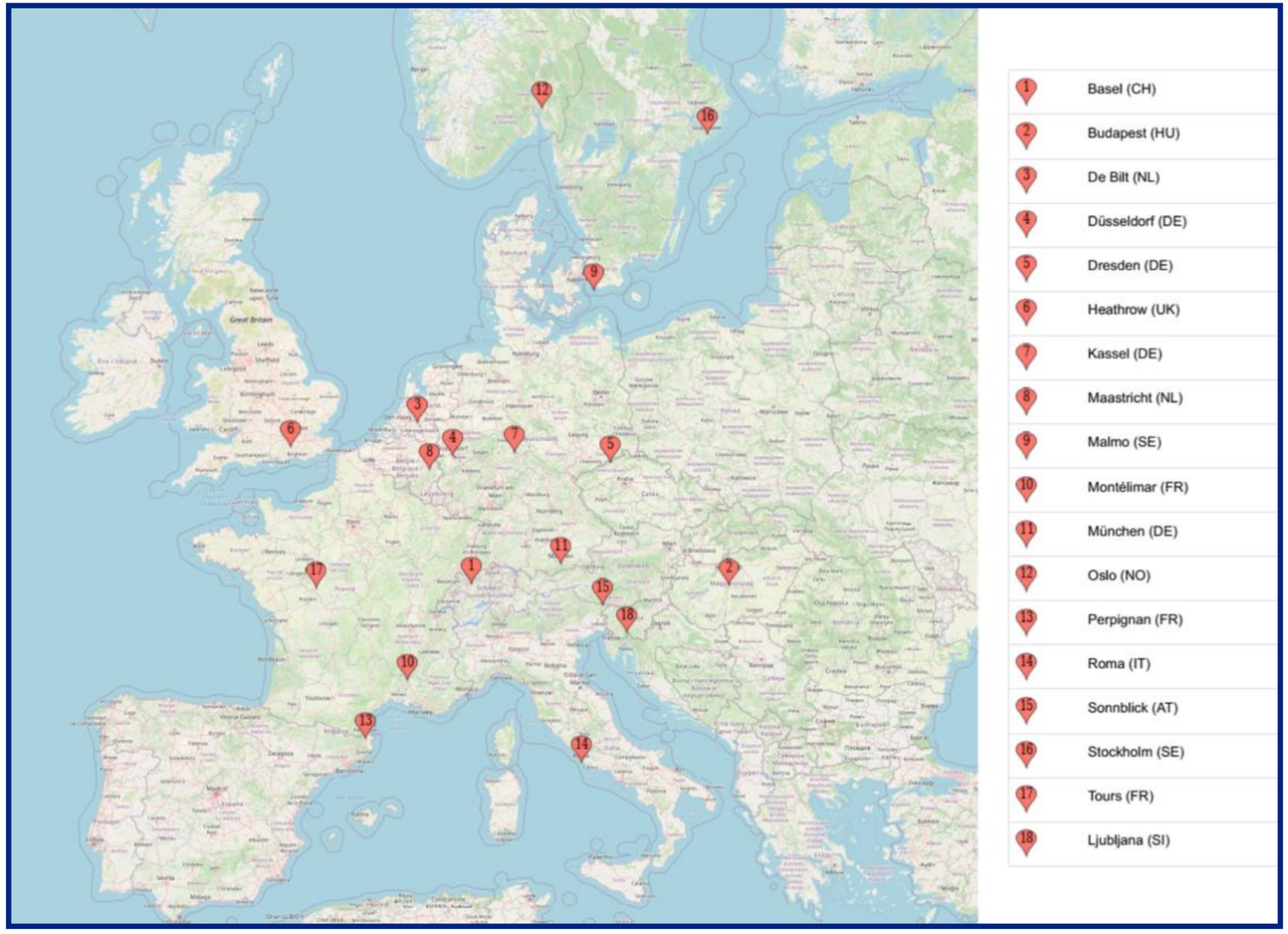


Figure 2. Map of 18 observed cities (Source: ECA&D)

- The "weather prediction dataset" is intuitively accessible weather observations from 18 locations in Europe. While all selected locations provide data for the variables ‘mean temperature’, ‘max temperature’, and ‘min temperature’, we also included data for the variables 'cloud_cover', 'wind_speed', 'wind_gust', 'humidity', 'pressure', 'global_radiation', 'precipitation', 'sunshine' wherever those were available.
- Location: Europe.
- Number of observations: 3654.
- Time: from 2000 to 2010.
- Source: Klein Tank, A.M.G. and Coauthors, 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. of Climatol., 22, 1441-1453.

3. EXPLORATORY DATA ANALYSIS

Initial Data

Proceed to collect data by code because the data in the table is separate and arranged by city. If left like that, it will be impossible to review all 18 cities.

Table 1. A portion of the dataset

DATE	MONTH	BASEL_cloud_cover	BASEL_humidity	BASEL_pressure	BASEL_global_radiation	BASEL_precipitation	BASEL_sunshine	BASEL_temp_mean	BASEL_temp_min	BASEL_temp_max
20000101	1	8	0.89	1.0286	0.2	0.03	0.0	2.9	1.6	3.9
20000102	1	8	0.87	1.0318	0.25	0.0	0.0	3.6	2.7	4.8
20000103	1	5	0.81	1.0314	0.5	0.0	3.7	2.2	0.1	4.8
20000104	1	7	0.79	1.0262	0.63	0.35	6.9	3.9	0.5	7.5
20000105	1	5	0.9	1.0246	0.51	0.07	3.7	6.0	3.8	8.6

3. EXPLORATORY DATA ANALYSIS

Data Processing Steps

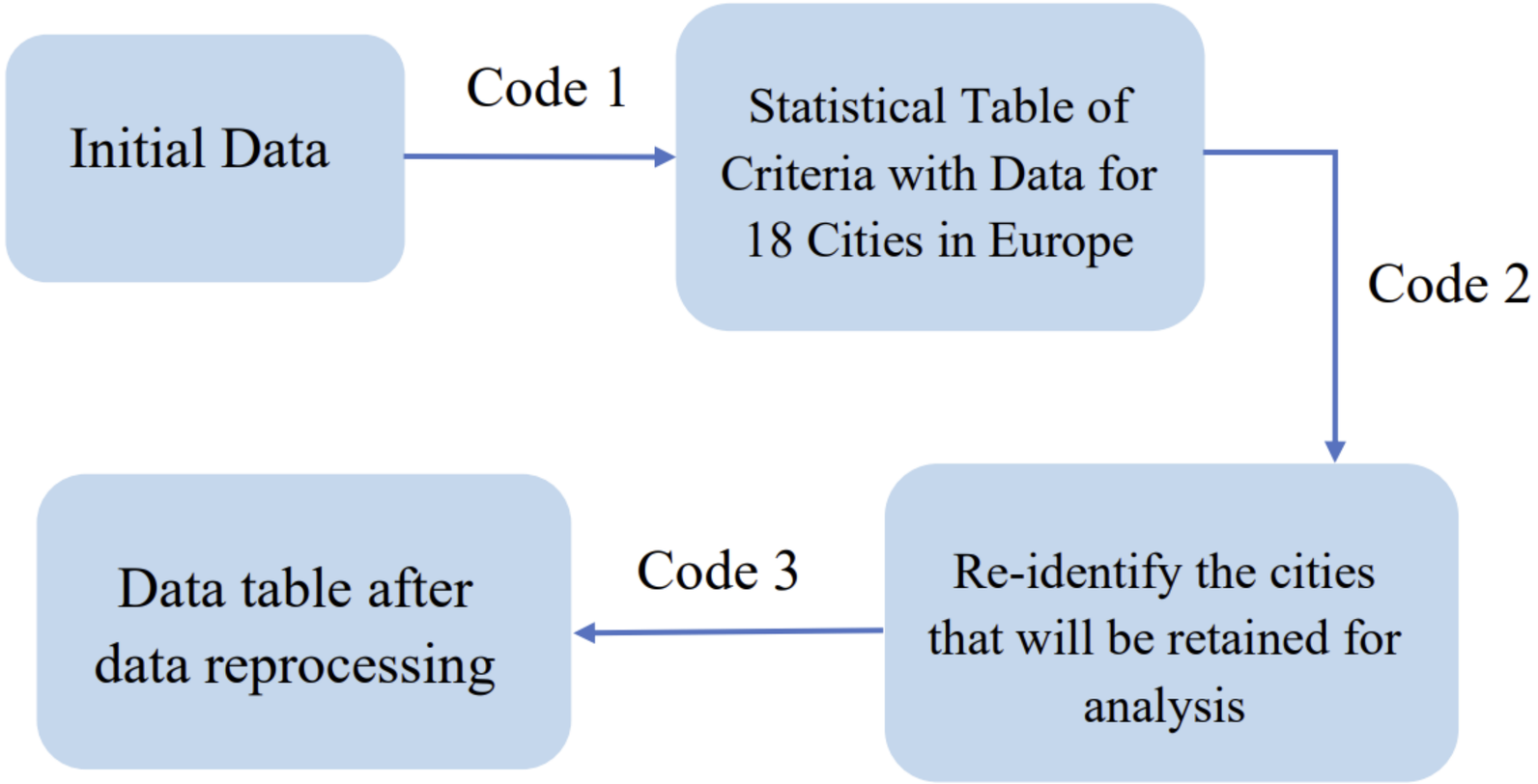


Figure 1. Data processing steps

3. EXPLORATORY DATA ANALYSIS

Statistical Table of Criteria with Data

- Statistical Tables: Created for each city to aid decision-making.
- 18 Cities Analysis: Only 2 common variables: Temp Max and Temp Mean.
- Optimization: Consider eliminating cities with fewer criteria to focus on more comprehensive data.

	cloud_cover	wind_speed	wind_gust	humidity	pressure	global_radiation	precipitation	sunshine	temp_mean	temp_min	temp_max
BASEL	x			x	x	x	x	x	x	x	x
BUDAPEST	x			x	x	x	x	x	x		x
DE_BILT	x	x	x	x	x	x	x	x	x	x	x
DRESDEN	x	x	x	x		x	x	x	x	x	x
DUSSELDORF	x	x	x	x	x	x	x	x	x	x	x
HEATHROW	x			x	x	x	x	x	x	x	x
KASSEL		x	x	x	x	x	x	x	x	x	x
LJUBLJANA	x	x		x	x	x	x	x	x	x	x
MAASTRICHT	x	x	x	x	x	x	x	x	x	x	x
MALMO		x					x		x	x	x
MONTELIMAR		x		x	x	x	x		x	x	x
MUENCHEN	x	x	x	x	x	x	x	x	x	x	x
OSLO	x	x	x	x	x	x	x	x	x	x	x
PERPIGNAN		x		x	x	x	x		x	x	x
ROMA	x			x	x			x	x	x	x
SONNBLICK	x			x		x	x	x	x	x	x
STOCKHOLM	x				x		x	x	x	x	x
TOURS		x		x	x	x	x		x	x	x

3. EXPLORATORY DATA ANALYSIS

Re-identify The Cities

These cities do not provide the amount of data needed for a comprehensive analysis if the goal is to include at least six common criteria across cities.

Budapest and Sonnblick can still be retained because adding or removing these 2 cities still meets the 6 criteria, enough to meet the requirements for effective data analysis.

- Budapest (8 criteria)
- Roma (6 criteria)
- Malmo (5 criteria)
- Stockholm (7 criteria)
- Montelimar (8 criteria)
- Sonnblick (8 criteria)
- Perpignan (8 criteria)
- Tours (8 criteria)

3. EXPLORATORY DATA ANALYSIS

Re-identify The Cities

These are the remaining cities in the data used for analysis.

- Basel
- Budapest
- De Bilt
- Dresden
- Dusseldorf
- Heathrow
- Kassel
- Ljubljana
- Maastricht
- Muenchen
- Oslo
- Stockholm

3. EXPLORATORY DATA ANALYSIS

Data Table After Data Reprocessing

Use R to remake the new data with the Go Picnic column incorporated

Table 3. Data table after data reprocessing

DATE	MONTH	SEASON	humidity	global_radiation	precipitation	sunshine	temp_mean	temp_max	GO_PICNIC
20000101	1	Winter	0.89	0.2	0.03	0	2.9	3.9	NO
20000102	1	Winter	0.87	0.25	0	0	3.6	4.8	NO
20000103	1	Winter	0.81	0.5	0	3.7	2.2	4.8	NO
20000104	1	Winter	0.79	0.63	0.35	6.9	3.9	7.5	NO
20000105	1	Winter	0.9	0.51	0.07	3.7	6	8.6	NO
20000106	1	Winter	0.85	0.56	0	5.7	4.2	6.9	NO
20000107	1	Winter	0.84	0.2	0	0	4.7	6.2	NO

3. EXPLORATORY DATA ANALYSIS

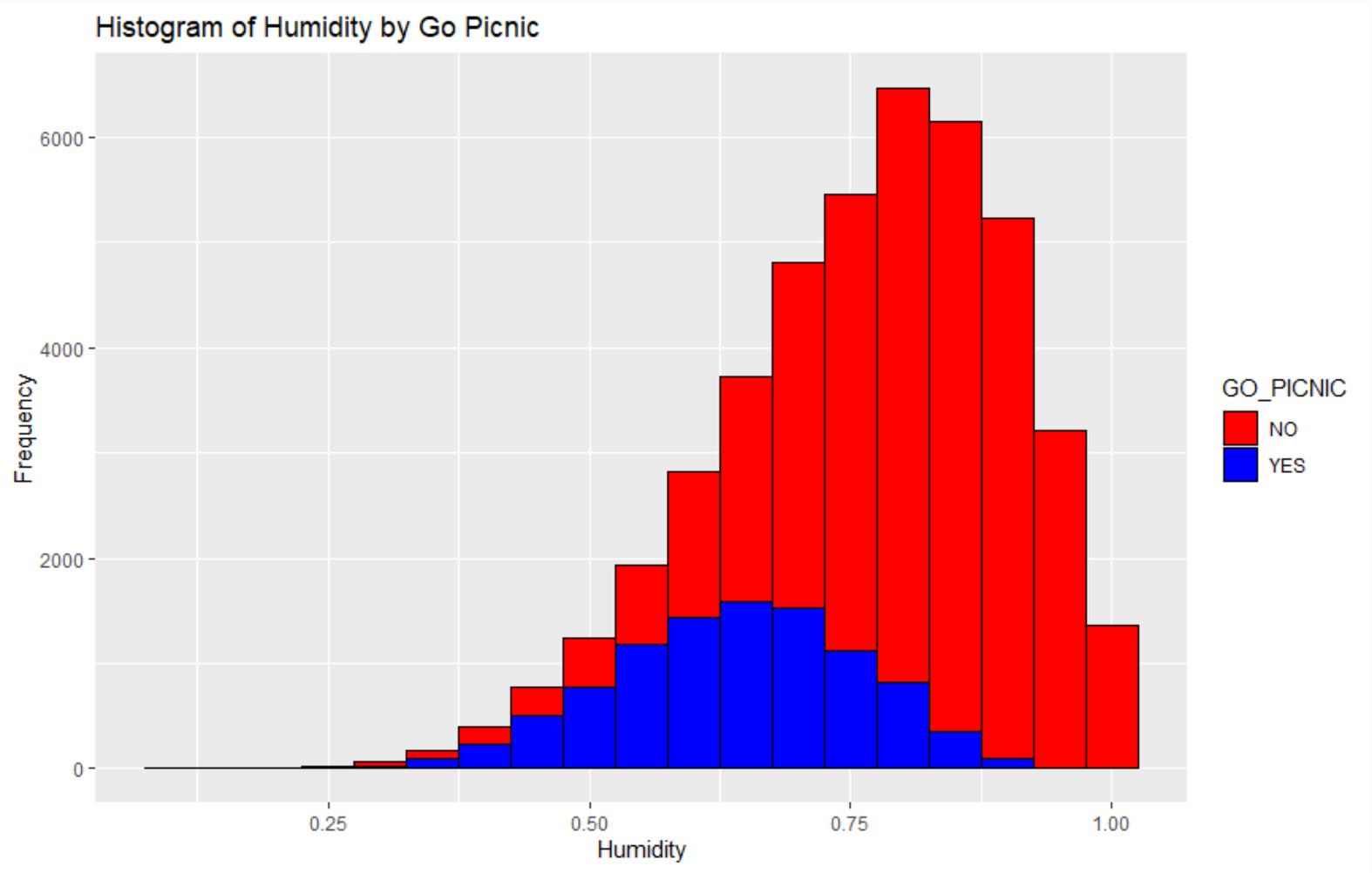
Structure Data

Table 4. Structure data

No.	Field Name	Data Types	Level of measurements	Description
1	DATE	Character	Obligatory	Date, month, and year where the observation took place.
2	MONTH	Character	Obligatory	The month where the observation took place.
3	SEASON	Character	Obligatory	The season where the observation took place.
4	humidity	Numeric	Obligatory	Humidity (unit: %)
5	global_radiation	Numeric	Obligatory	Global radiation (unit: 100 W/m2)
6	precipitation	Numeric	Obligatory	Precipitation amount (unit: 10 mm)
7	sunshine	Numeric	Obligatory	Sunshine in 1 Hours
8	temp_mean	Numeric	Obligatory	Mean temperature (unit: Celsius)
9	temp_max	Numeric	Obligatory	Max temperature (unit: Celsius)
10	GO_PICNIC	Character	Obligatory	Decide if the weather is suitable for a picnic (Yes/No)

3. EXPLORATORY DATA ANALYSIS

Visualization & Basics Evaluation



Humidity:

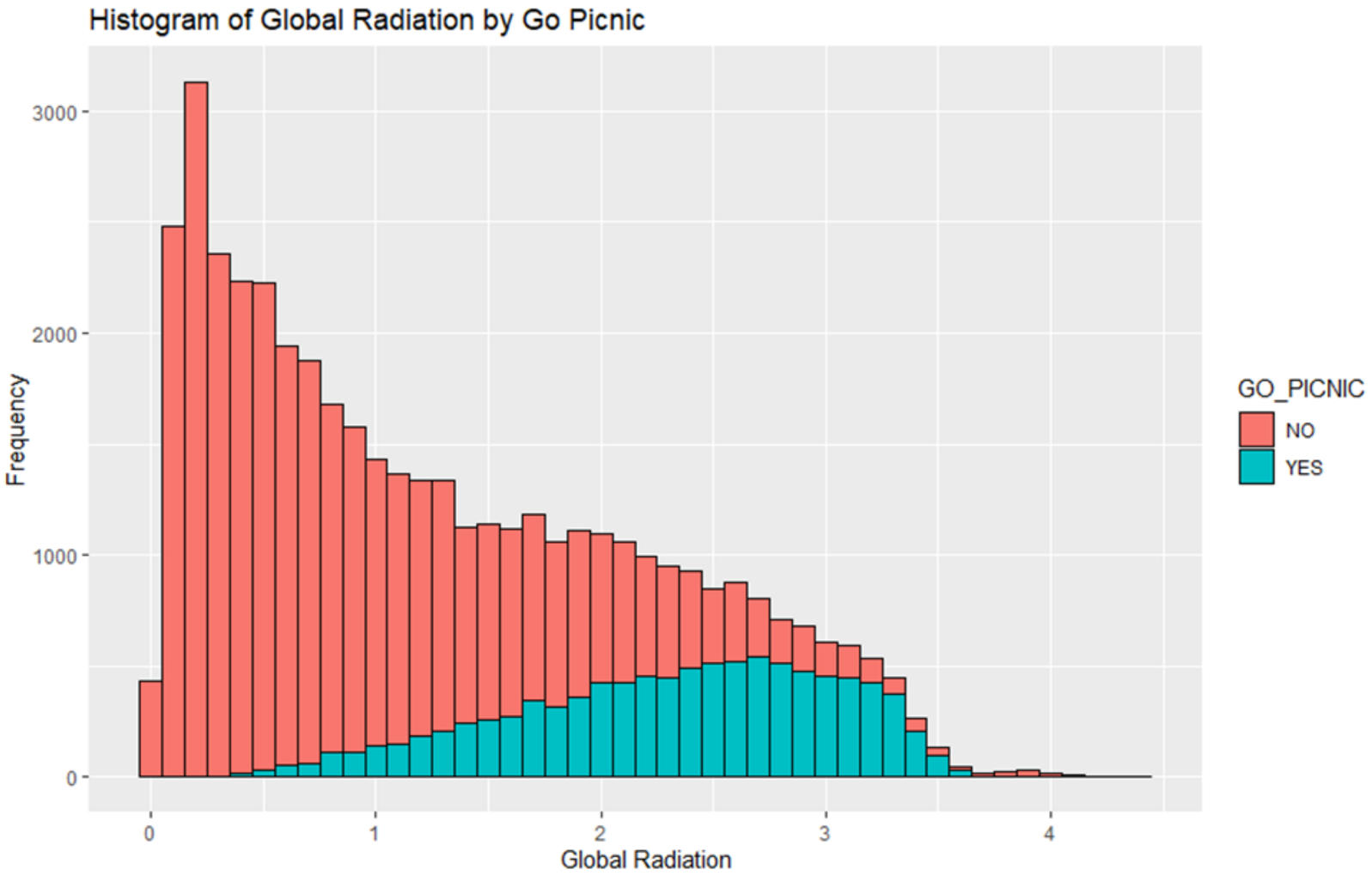
Mainly focus

0.75 - 0.8

However, the number of NO is higher than the number of YES, suggesting that when humidity is high people are less likely to picnic.

3. EXPLORATORY DATA ANALYSIS

Visualization & Basics Evaluation



Global Radiation:

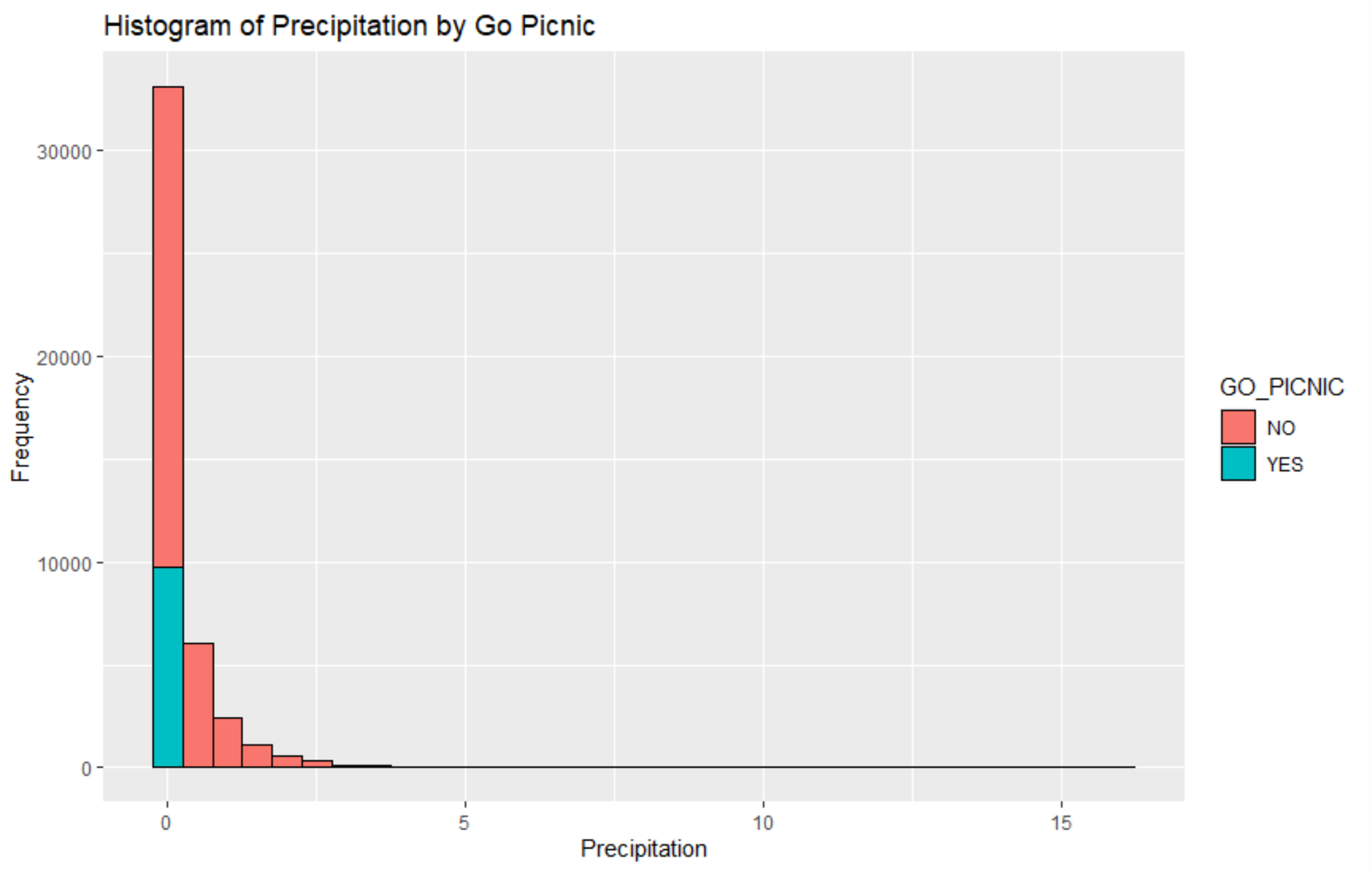
Most concentrated left deviation

0.3

Mostly NO at 0 - 1 and decreasing at 2 - 3 as the number of YES increases shows that people often go picnicking when the radiation levels are at medium level.

3. EXPLORATORY DATA ANALYSIS

Visualization & Basics Evaluation



Precipitation:

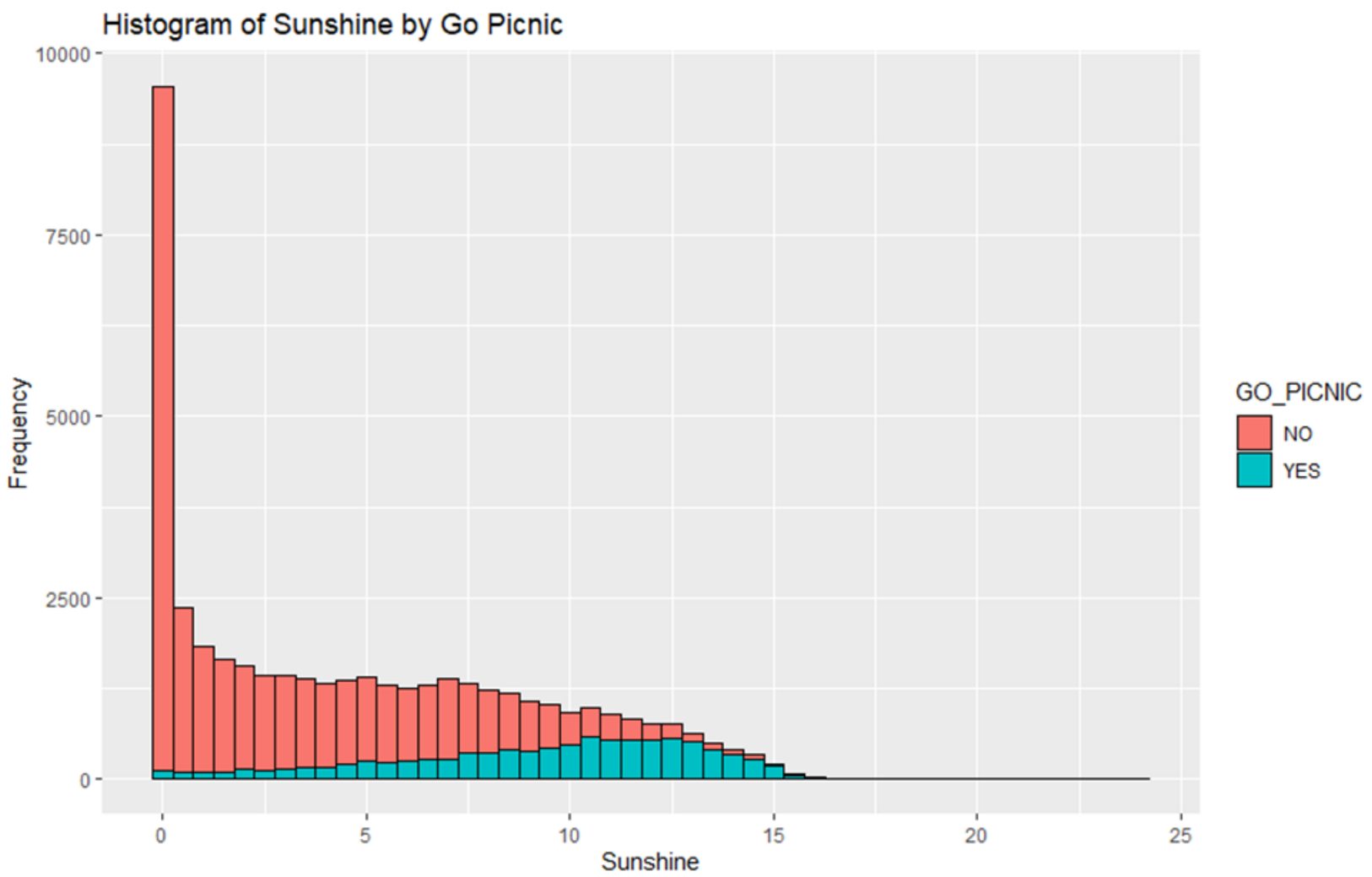
Most concentrated left deviation

0

There is a contradiction here that most of them are NO but there is still a part that says YES. From there, it is assumed that Precipitation does not affect people's decision to go or not too much.

3. EXPLORATORY DATA ANALYSIS

Visualization & Basics Evaluation



Sunshine:

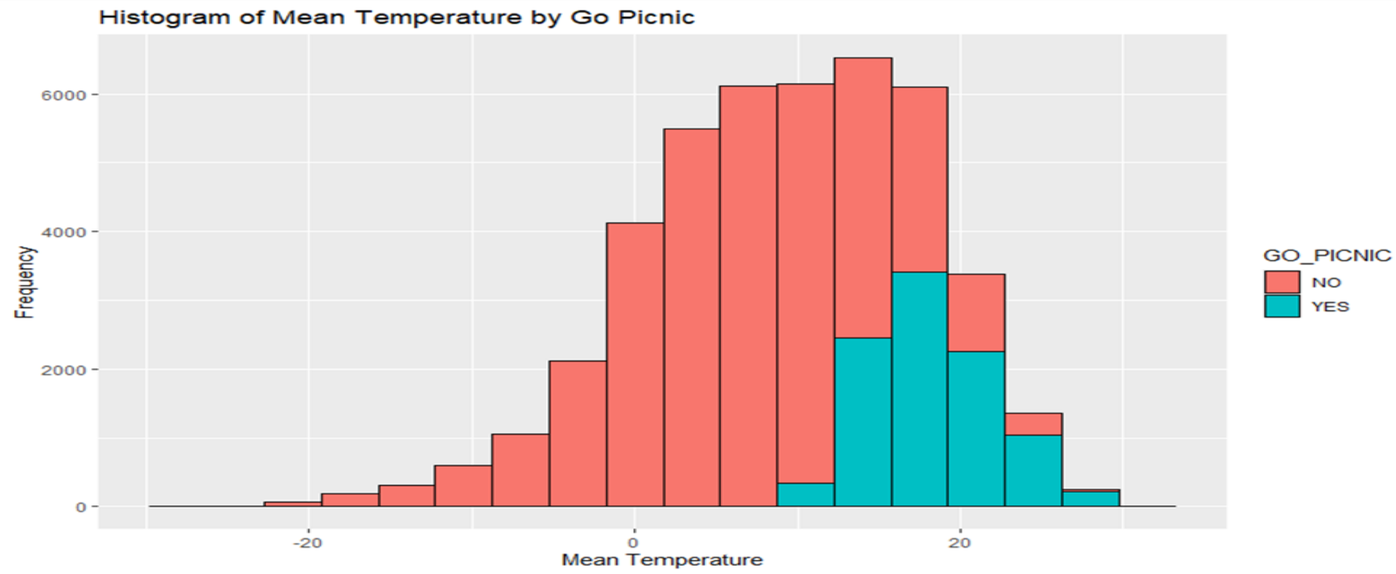
Most concentrated left deviation

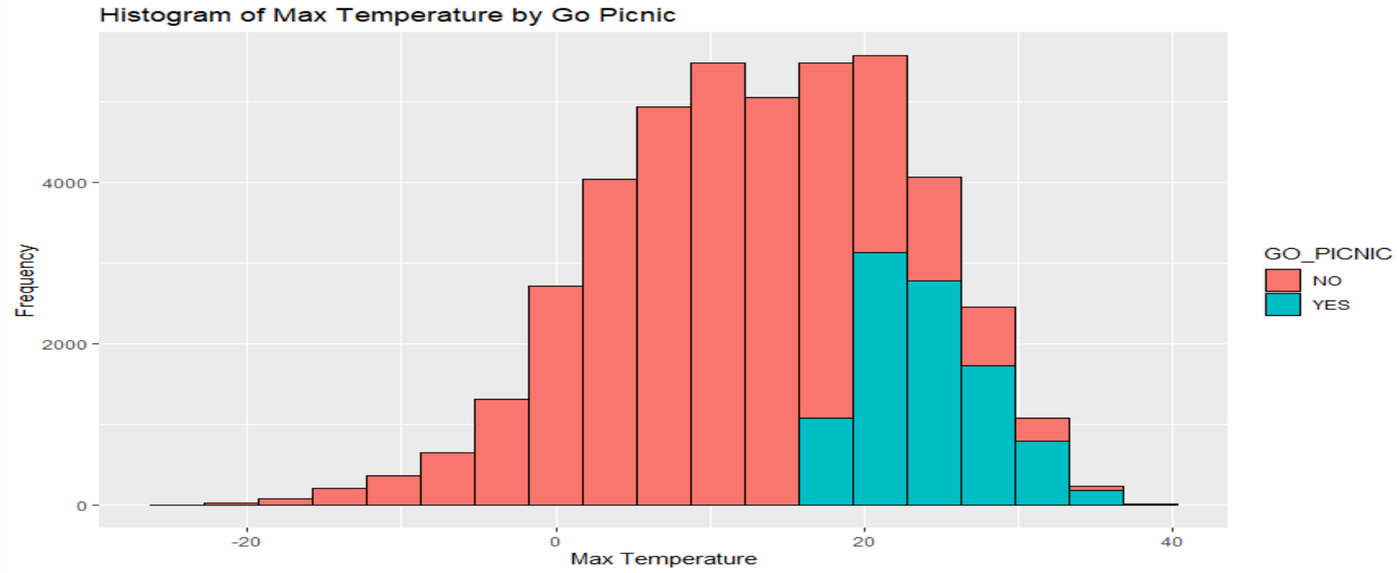
0

Most of the data shows that the sunshine level in 1 hour is 0.
Most of the data shows that with 0 sunshine, most people don't go, but increases gradually as the sunlight level increases to 10 - 15.

3. EXPLORATORY DATA ANALYSIS

Visualization & Basics Evaluation





Mean & Max Temperature:

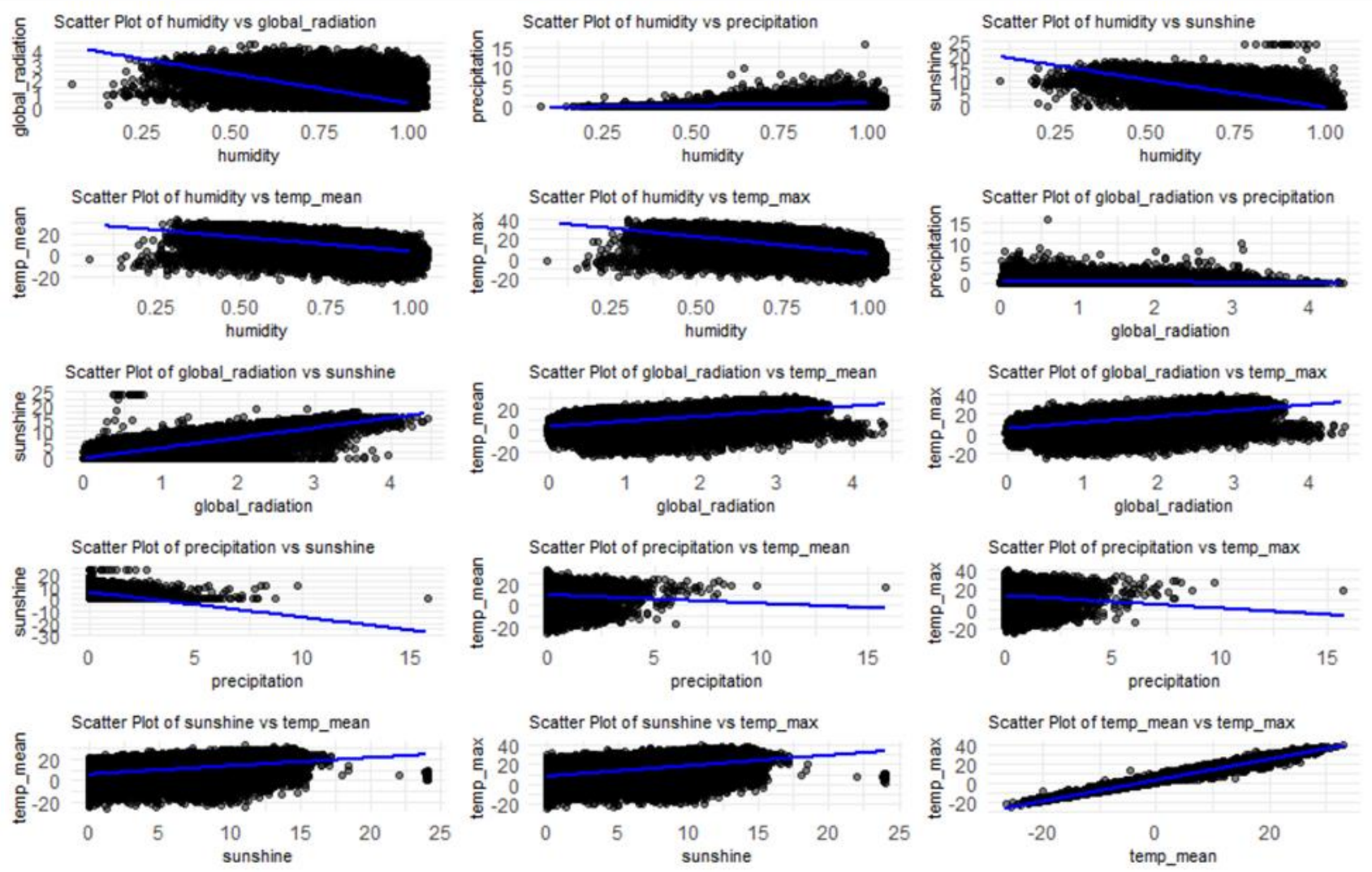
Evenly distributed, most concentrated at temperature

5 - 20°C

NO numbers still predominate at low temperatures, but YES numbers increase as the weather begins to warm up.

3. EXPLORATORY DATA ANALYSIS

Visualization & Basics Evaluation



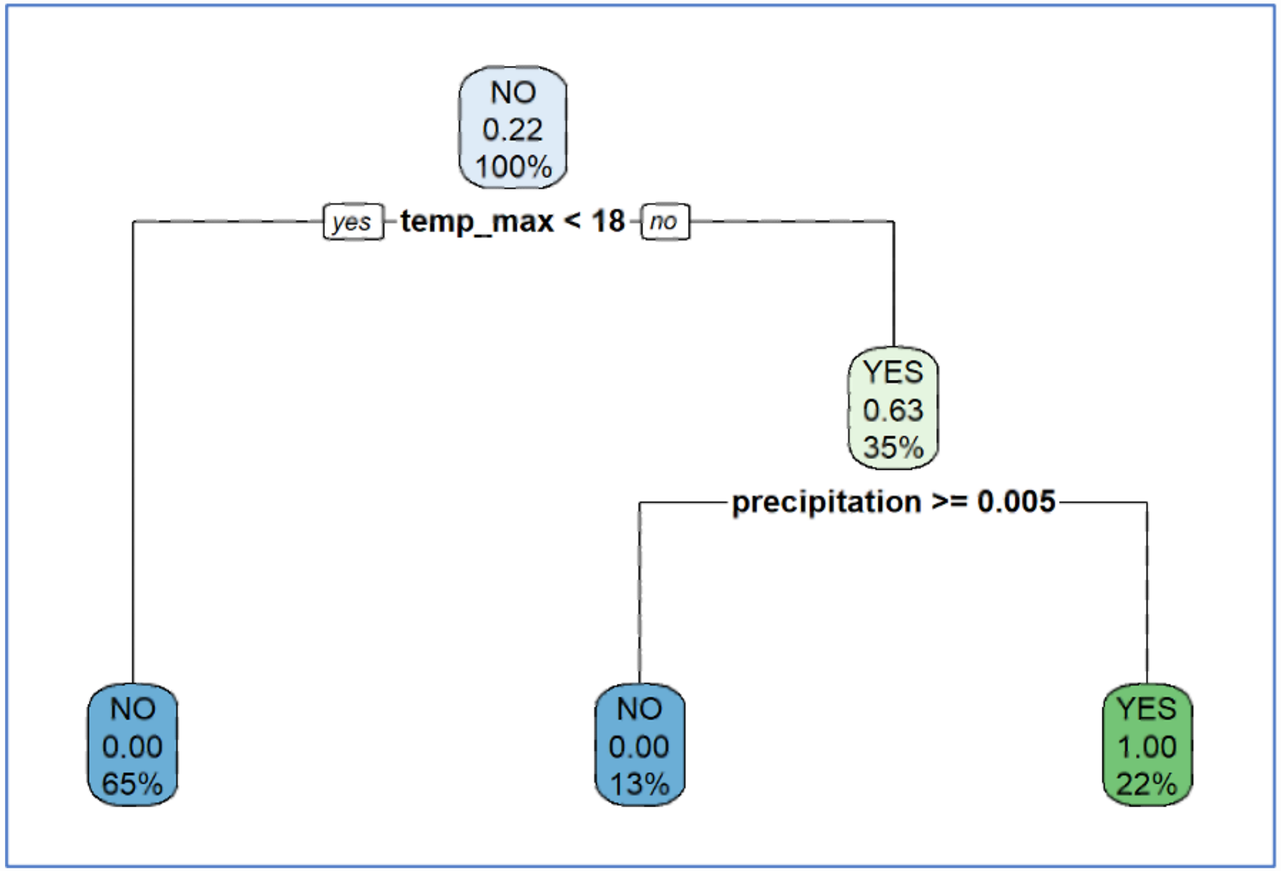
Relationship between variables:

Scatter plot of temp_mean and temp_max shows a tight clustering of points along the regression line

Removing one of the two variables is very difficult, so we decided to model both cases: removing temp_mean and removing temp_max.

4. MACHINE LEARNING MODEL

Case 1: Building a Decision Tree model with columns in Main_Data



However, accuracy = 1

↓

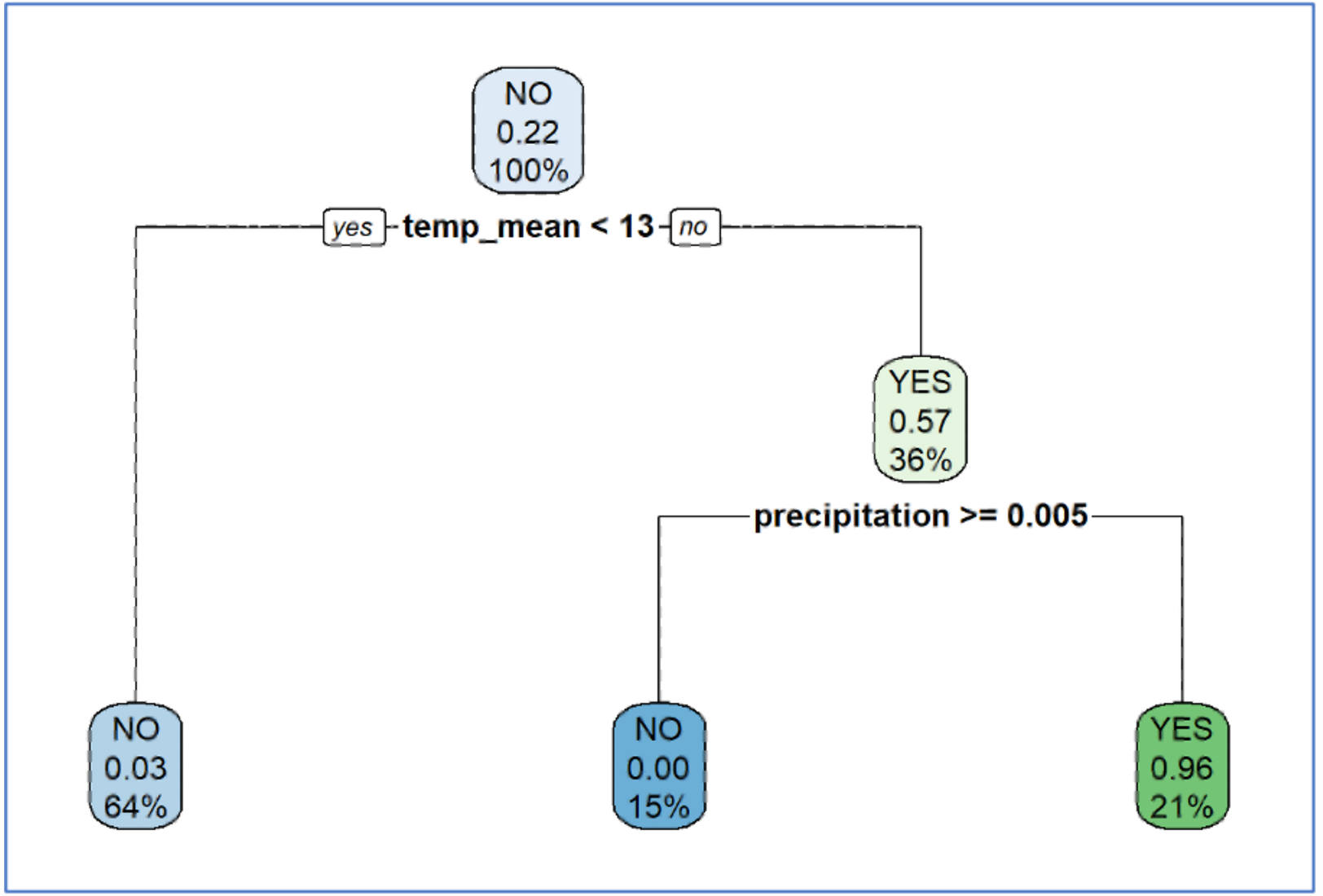
This is absolute accuracy

↓

The input data may show some anomalies (errors in data merging) ?

4. MACHINE LEARNING MODEL

Case 2: Building a Decision Tree model without the temp_max column

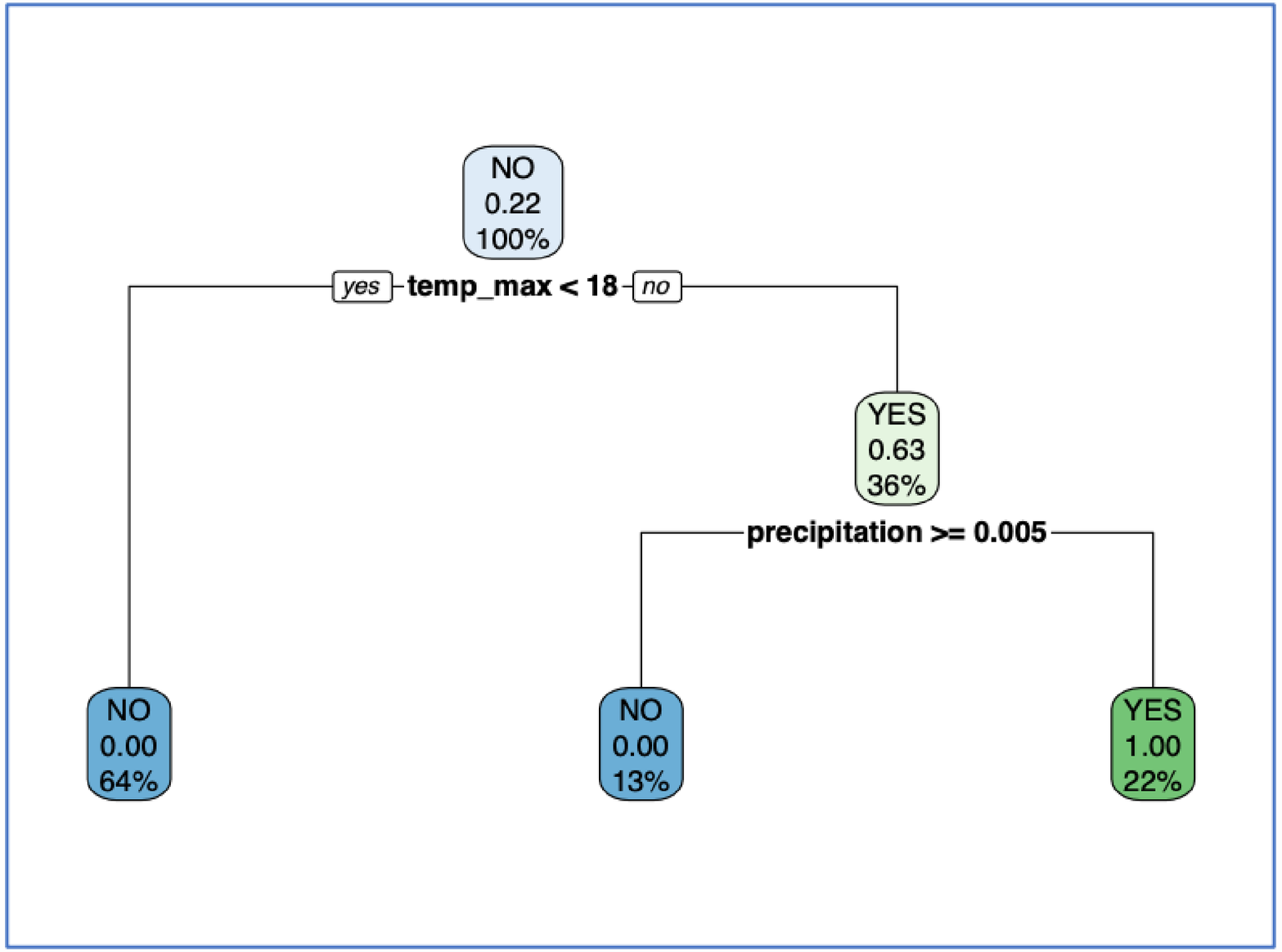


Accaracy =

0.9726

4. MACHINE LEARNING MODEL

Case 3: Building a Decision Tree model without the temp_mean column



Accaracy =

1

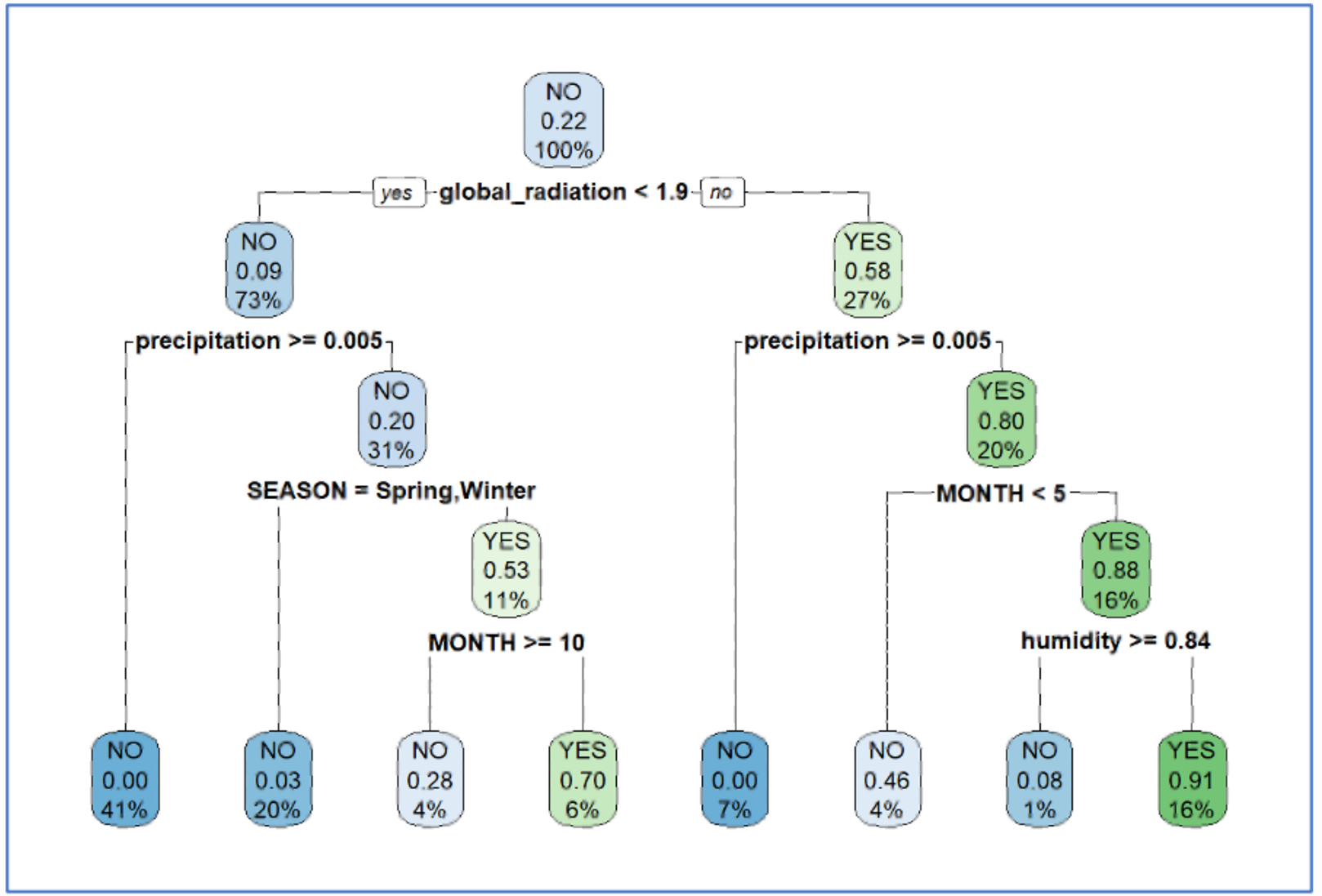
4. MACHINE LEARNING MODEL

The analysis process is divided into 6 cases as follows:

	Main_Data										
	DATE	MONTH	SEASON	humidity	global_radiation	precipitation	sunshine	temp_mean	temp_max	CITY	GO_PICNIC
Case 1 - data1											
Case 2 - data2											
Case 3 - data3											
	Keep the variable (temp_mean)										
Case 4 - data4											
Case 5 - data5											
Case 6 - data6											

4. MACHINE LEARNING MODEL

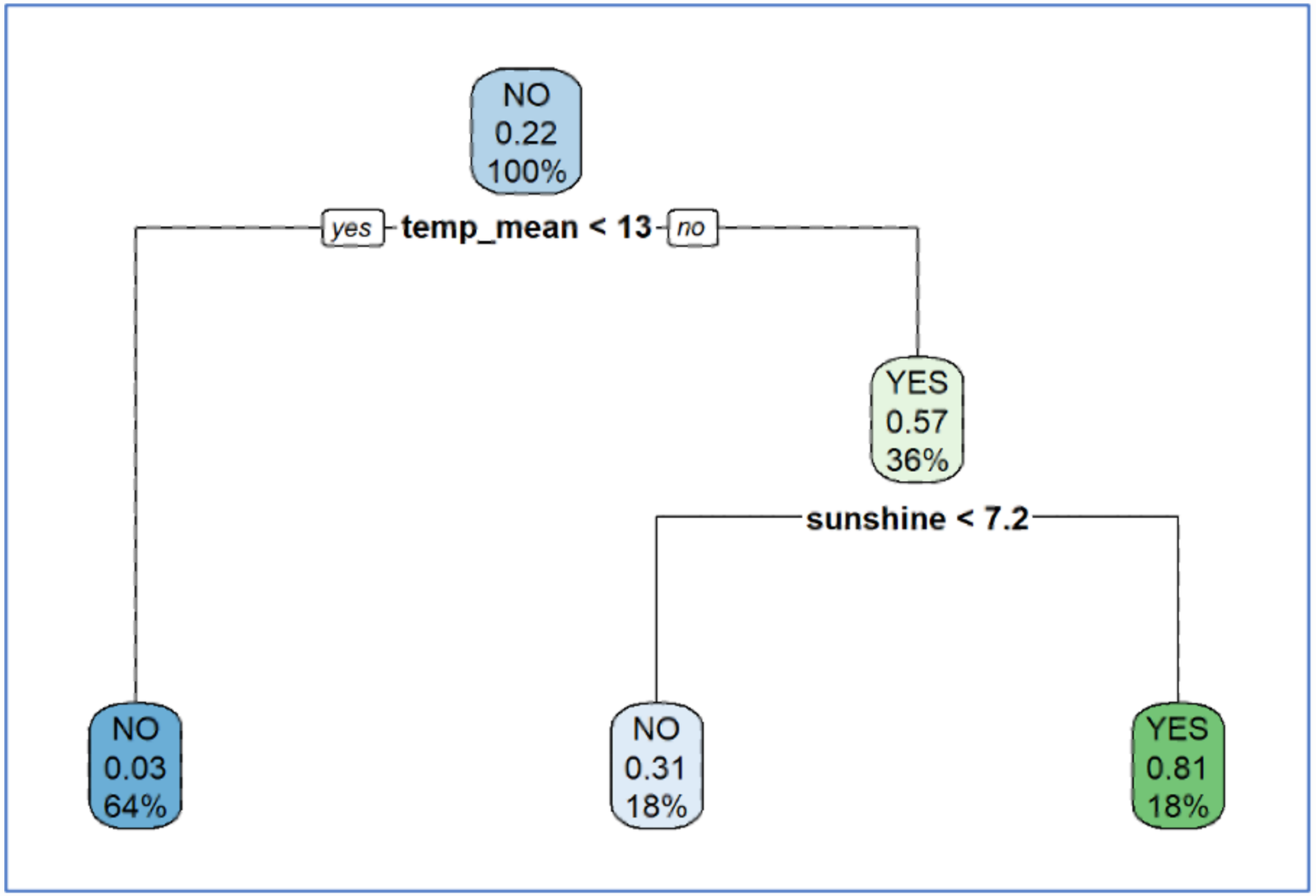
Case 4: Building the Decision Tree model without the temp_mean and temp_max column



Accaracy = 0.9292

4. MACHINE LEARNING MODEL

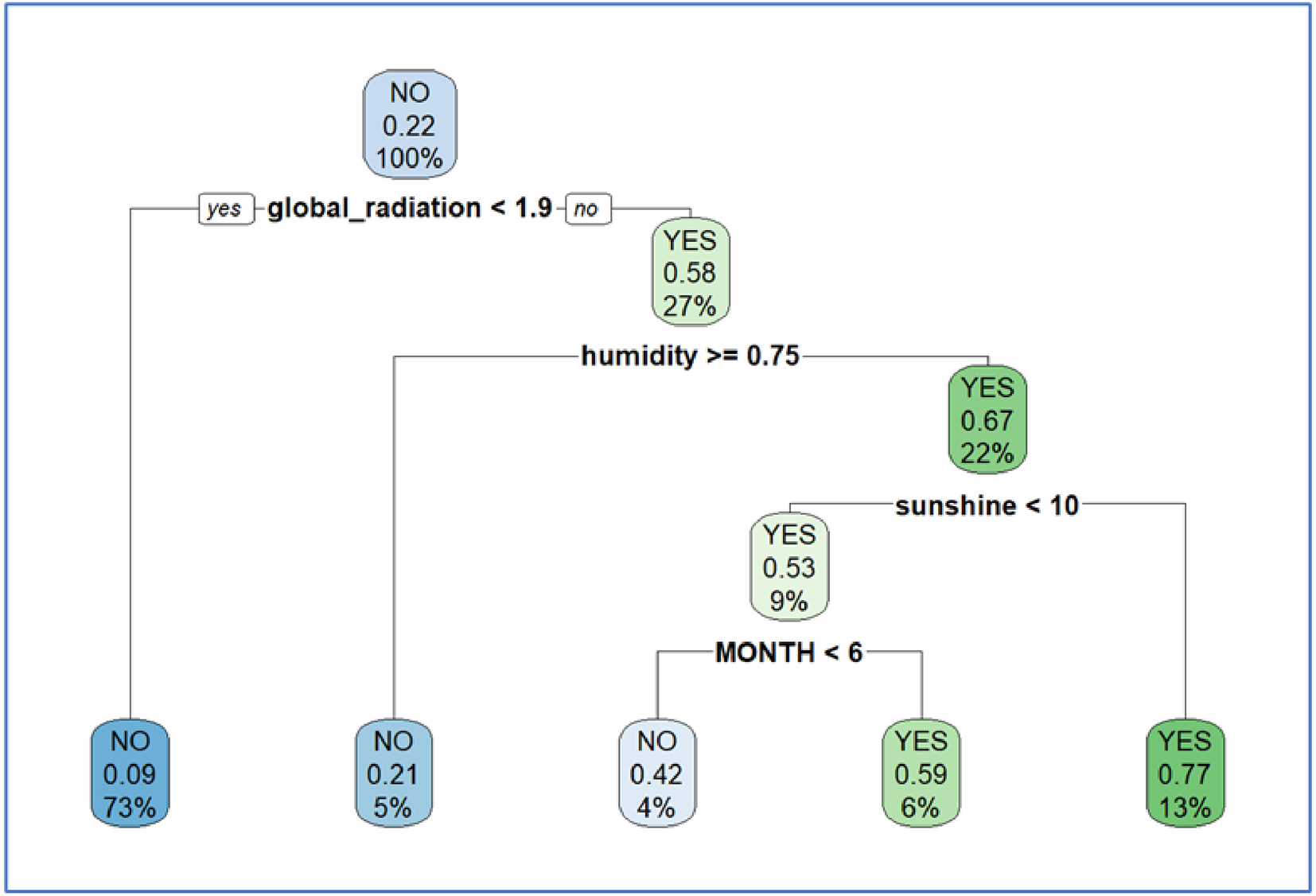
Case 5: Building the Decision Tree model without the precipitation column and keep temp_mean column



Accaracy = 0.8927

4. MACHINE LEARNING MODEL

Case 6: Continue building the Decision Tree model without precipitation and temp_mean columns



Accaracy = 0.8592

CONCLUSION

Average temperature and precipitation are two key factors in predicting the likelihood of going for a picnic.

Removing either factor significantly decreases the model's accuracy, indicating that they not only have independent value but also interact with each other in influencing outdoor activity decisions.

Ideal Weather

SEASON	AVERAGE TEMPERTURE	PRECIPITATION	HUMIDITY	GLOBAL_RADIATION	SUNSHINE
IN SUMMER AND AUTUMN	>=13	< 0.005	< 0.84	> 1.9	>= 10

6. CONCLUSION

BUSINESS STRATEGY

Weather Adaptation Strategies

Develop the ability to combine opportunity and risk

Marketing communications

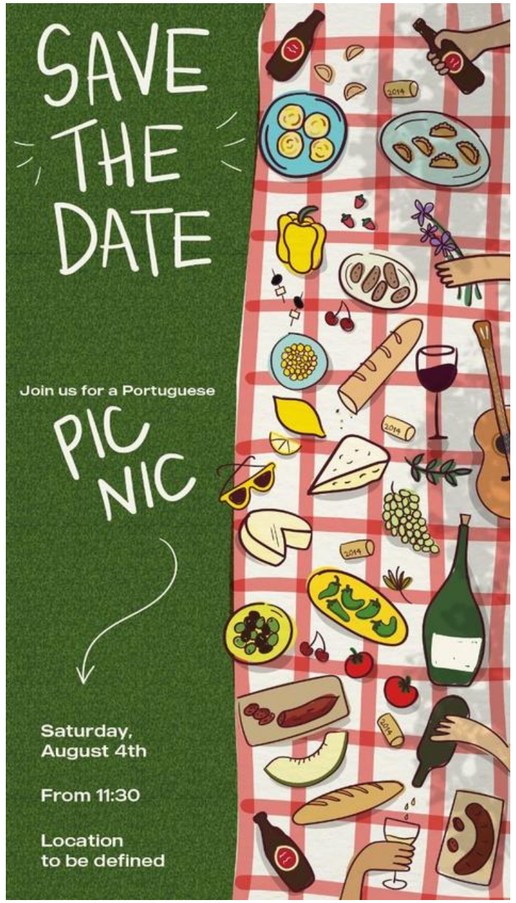
Combining sales and distribution channels

6. CONCLUSION

Develop the ability to combine opportunity and risk

01 Leveraging weather opportunities to increase revenue

Update a diverse range of products to meet weather conditions



Enhance advertising campaigns and promotions

Recommend picnic products people should prepare.

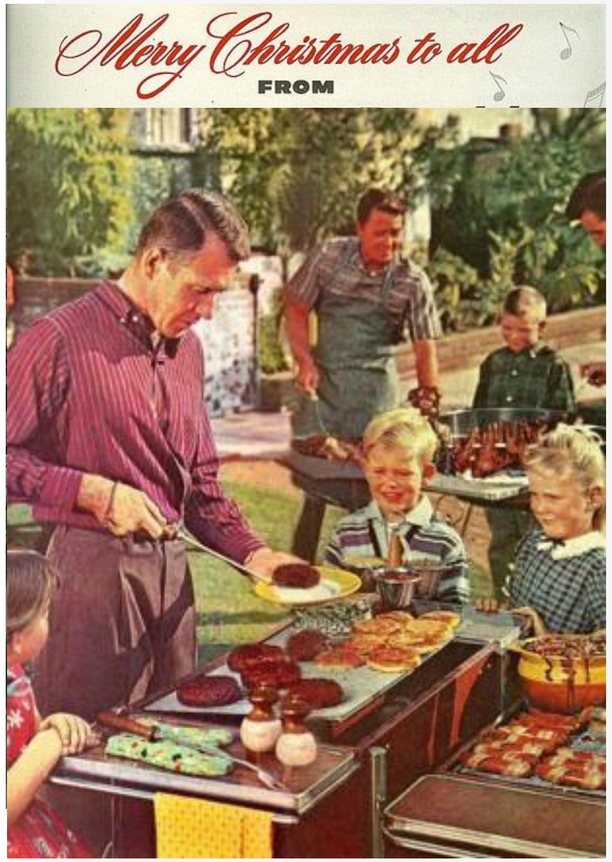
6. CONCLUSION

Develop the ability to combine opportunity and risk

02 Managing business risks when the weather is unfavorable

Limit off-season product imports and offer discounts on remaining inventory

Enhance advertising campaigns before unfavorable weather periods



Take advantage of holidays to create marketing campaign

6. CONCLUSION

Combining sales and distribution channels

Enhance online sales channels.

Expand distribution with local partners such as stores, tourist areas, or event-organizing companies.

Marketing communications

Develop digital advertising and marketing platform.


THANK YOU
FOR LISTENING


F&M