PA2
CMPS 470

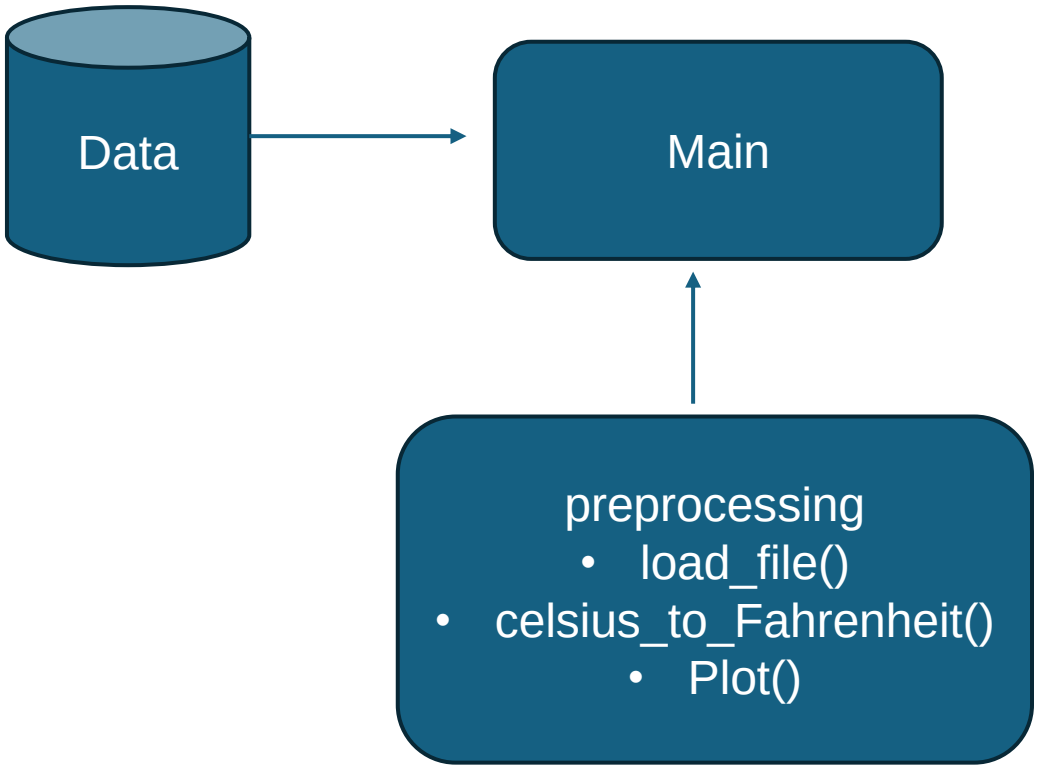
Byte Me: Sanele Harmon Taeden Kitchens

TEAM MEMBERS & ROLES

Sanele Harmon: Create the GitHub repository. Preprocess the temperature columns and plot

Taeden Kitchen: Preprocess the Rainfall column and plot. Visualize the distribution of each data column.

MODULE COMMUNICATION GRAPH



DESCRIPTION OF THE PROJECT

- Goal: Predicting whether there will be rainfall on a given day by means of ML algorithms ANN, SVM, DT, and K-NN.

Tasks:

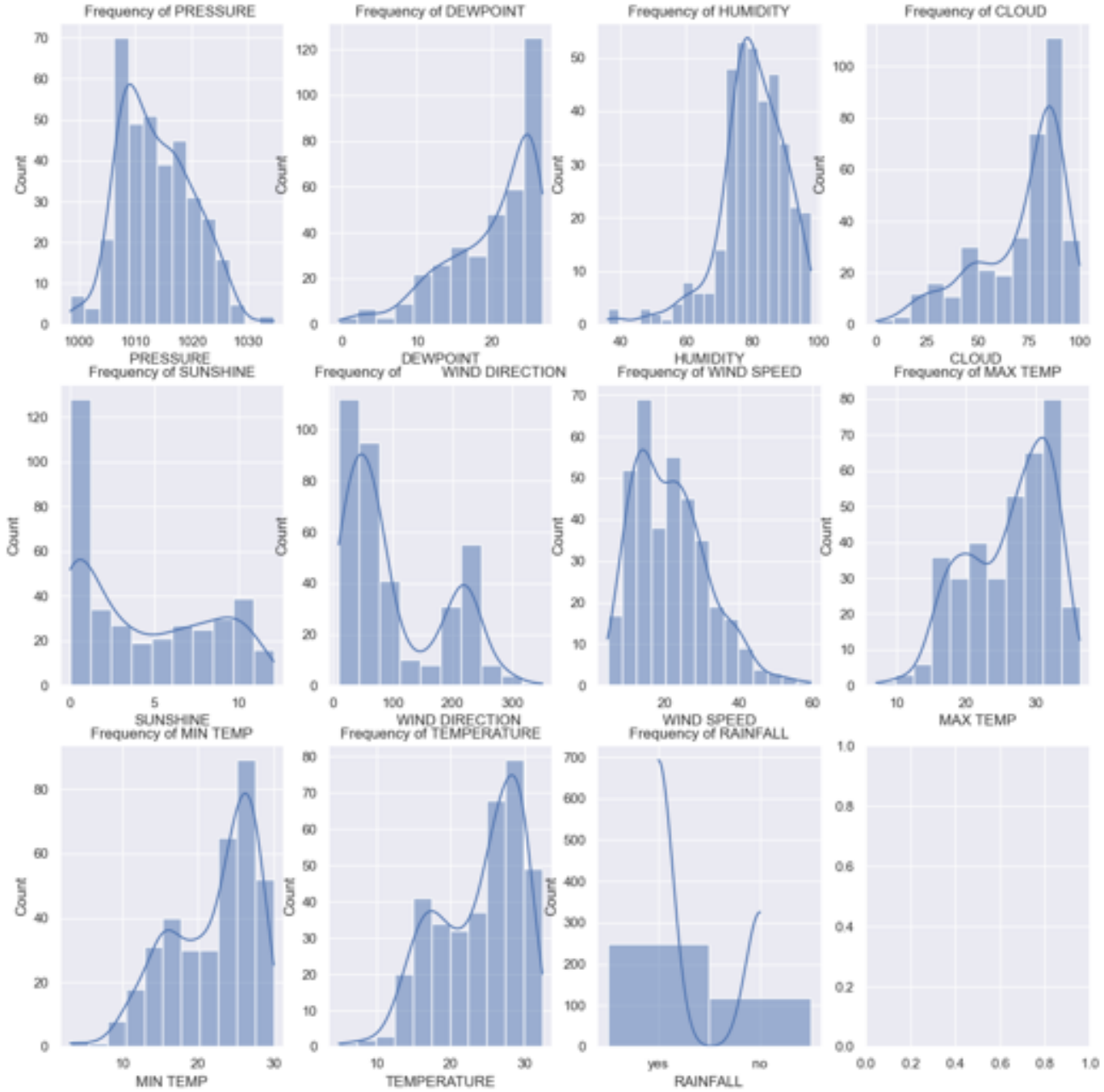
- Visualize data
- Preprocess data

DESCRIPTION OF THE RAW DATA

- Data Source: Seek Geek

ATTRIBUTE	TYPE
SAMPLE ID	INT
RAINFALL (TARGET)	STRING (‘yes’, ‘no’)
DAY	INT
PRESSURE	FLOAT
MAX TEMP	FLOAT
TEMPERATURE	FLOAT
MIN TEMP	FLOAT
DEW POINT	FLOAT
HUMIDITY	INT
CLOUD	INT
SUNSHINE	FLOAT
WIND DIRECTION	INT
WIND SPEED	FLOAT

DISTRIBUTION OF DATA

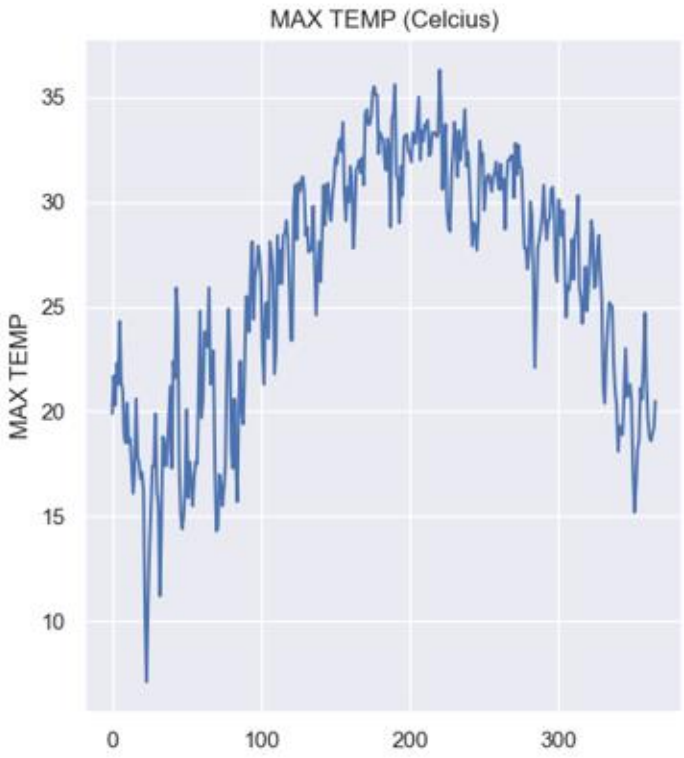


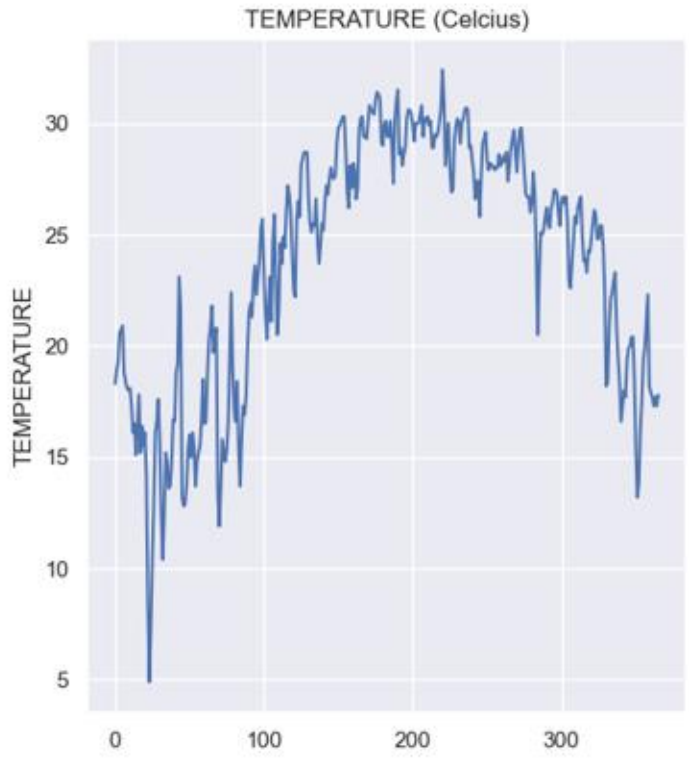
PREPROCESSING

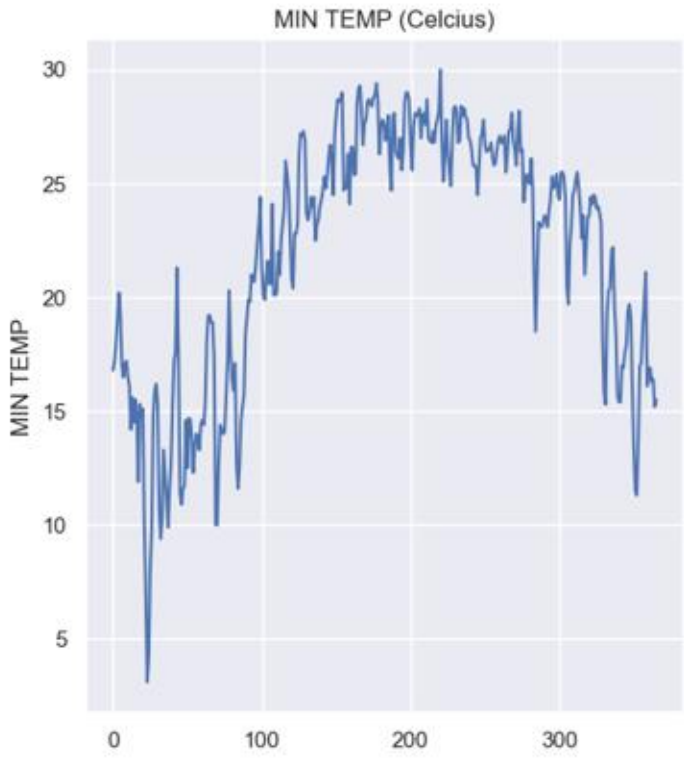
- For preprocessing, the values for the RAINFALL column were mapped from ‘yes’ and ‘no’ to 1 and 0, respectively
- The MAX TEMP, MIN TEMP, and TEMPERATURE columns were converted from Celsius to Fahrenheit

RAW DATA: MAX_TEMP, MIN_TEMP, TEMPERATURE

SAMPLE ID	RAINFALL	DAY	PRESSURE	MAX TEMP	TEMPERATURE	MIN TEMP
1	yes	1	1025.9	19.9	18.3	16.8
2	yes	2	1022	21.7	18.9	17.2
3	yes	3	1019.7	20.3	19.3	18
4	yes	4	1018.9	22.3	20.6	19.1
5	yes	5	1015.9	21.3	20.7	20.2
6	yes	6	1018.8	24.3	20.9	19.2
7	no	7	1021.8	21.4	18.8	17
8	no	8	1020.8	21	18.4	16.5
9	no	9	1020.6	18.9	18.1	17.1
10	yes	10	1017.5	18.5	18	17.2

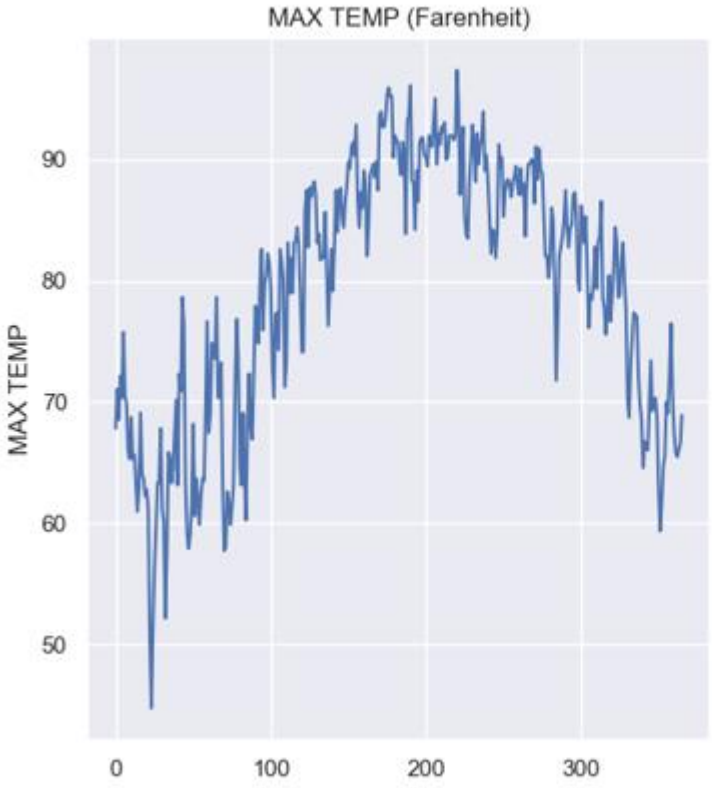


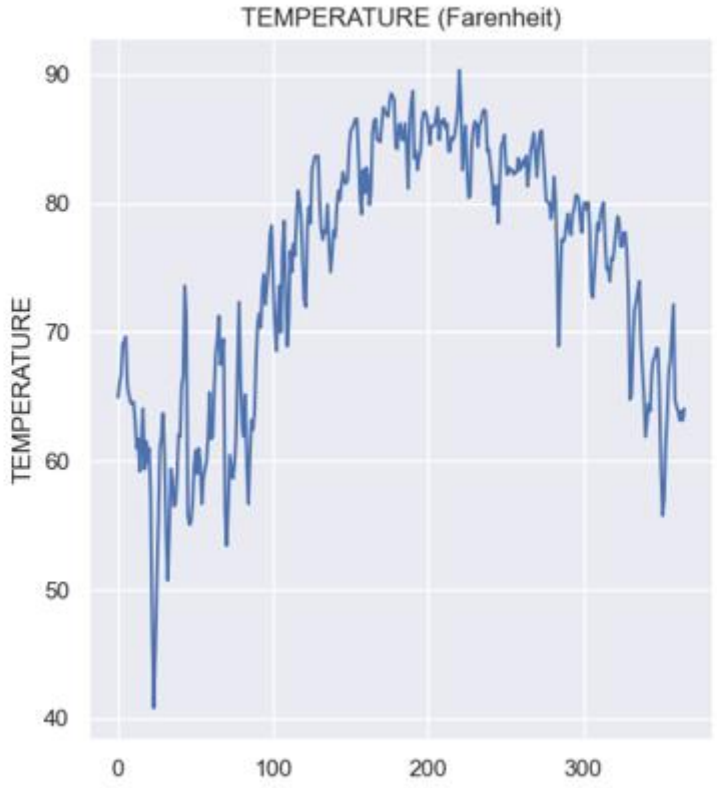


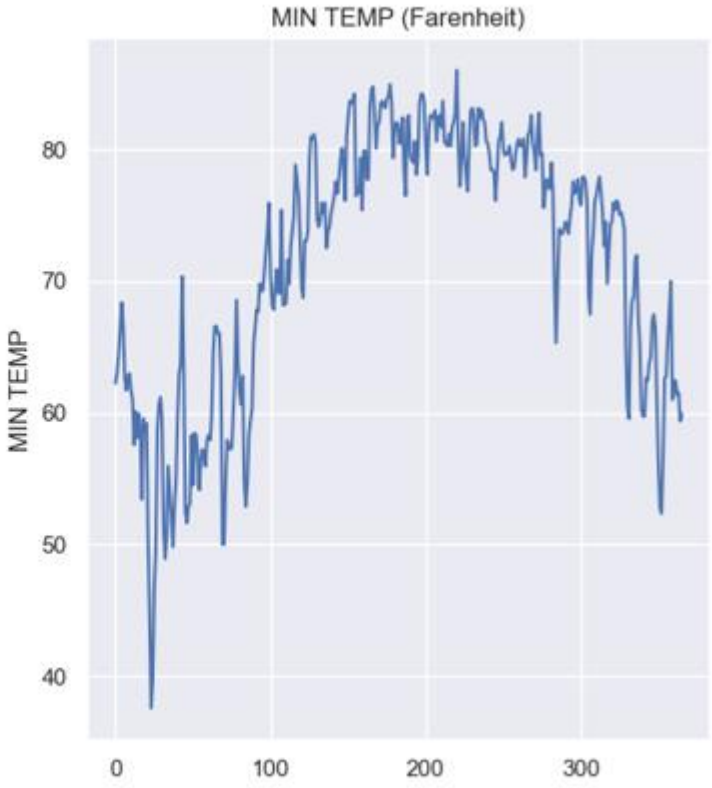


PREPROCESSED DATA: MAX_TEMP, MIN_TEMP, TEMPERATURE

SAMPLE ID	RAINFALL	DAY	PRESSURE	MAX TEMP	TEMPERATURE	MIN TEMP
1	yes	1	1025.9	67.82	64.94	62.24
2	yes	2	1022	71.06	66.02	62.96
3	yes	3	1019.7	68.54	66.74	64.4
4	yes	4	1018.9	72.14	69.08	66.38
5	yes	5	1015.9	70.34	69.26	68.36
6	yes	6	1018.8	75.74	69.62	66.56
7	no	7	1021.8	70.52	65.84	62.6
8	no	8	1020.8	69.8	65.12	61.7

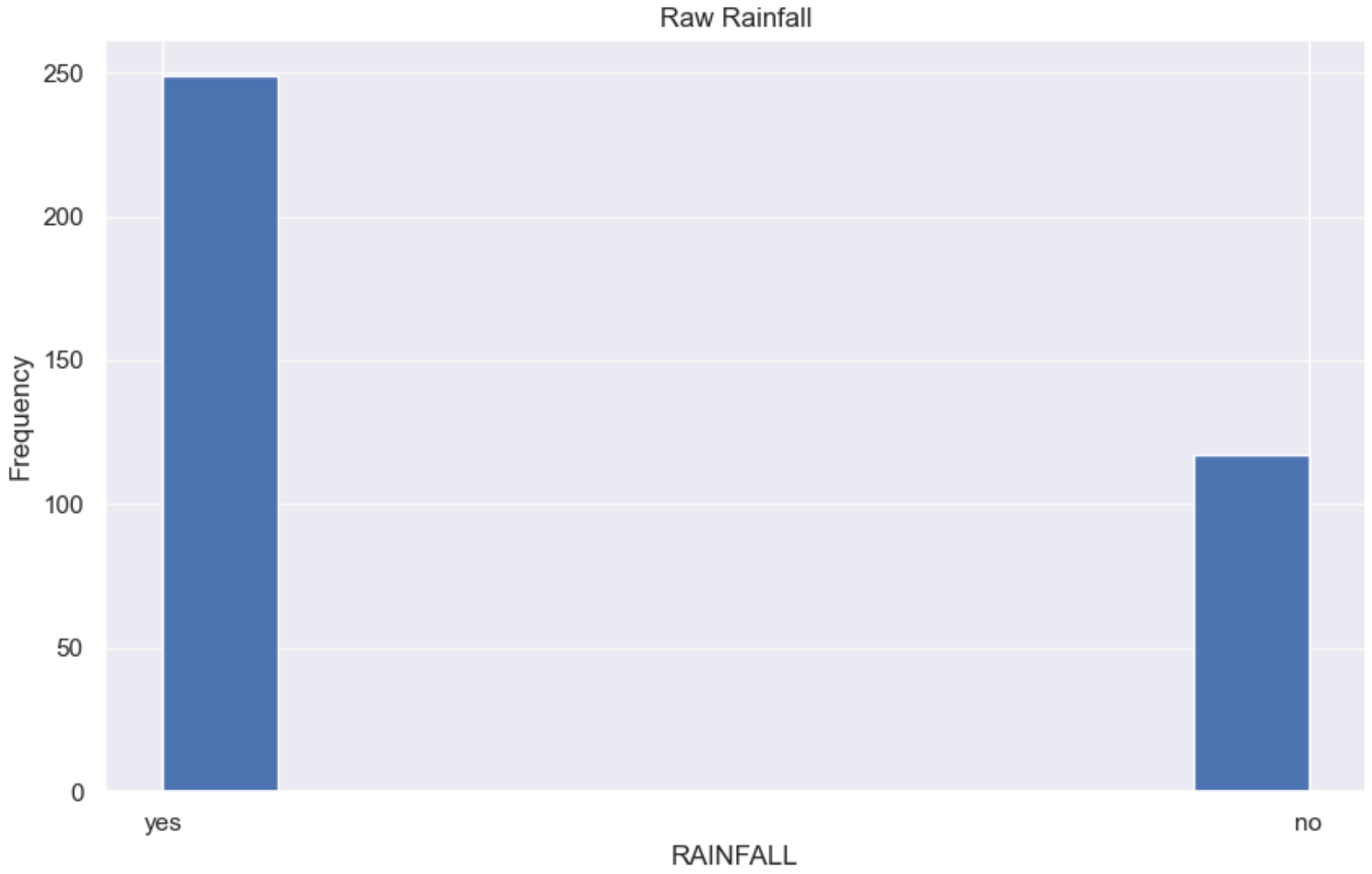






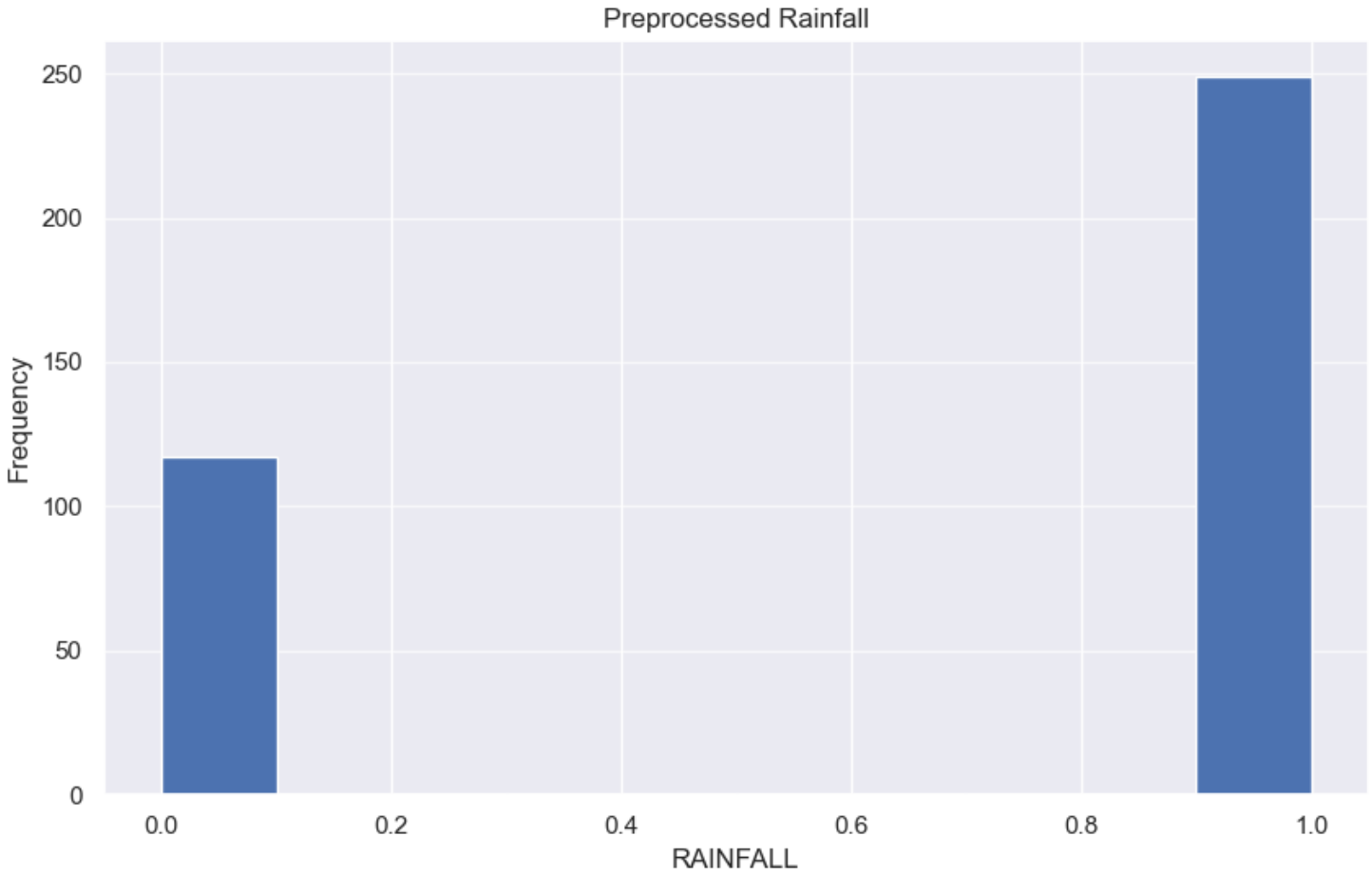
RAW DATA: Rainfall

SAMPLE ID	RAINFALL
1	yes
2	yes
3	yes
4	yes
5	yes
6	yes
7	no
8	no
9	no
10	yes
11	yes
12	no
13	no
14	yes
15	yes



PREPROCESSED DATA: Rainfall

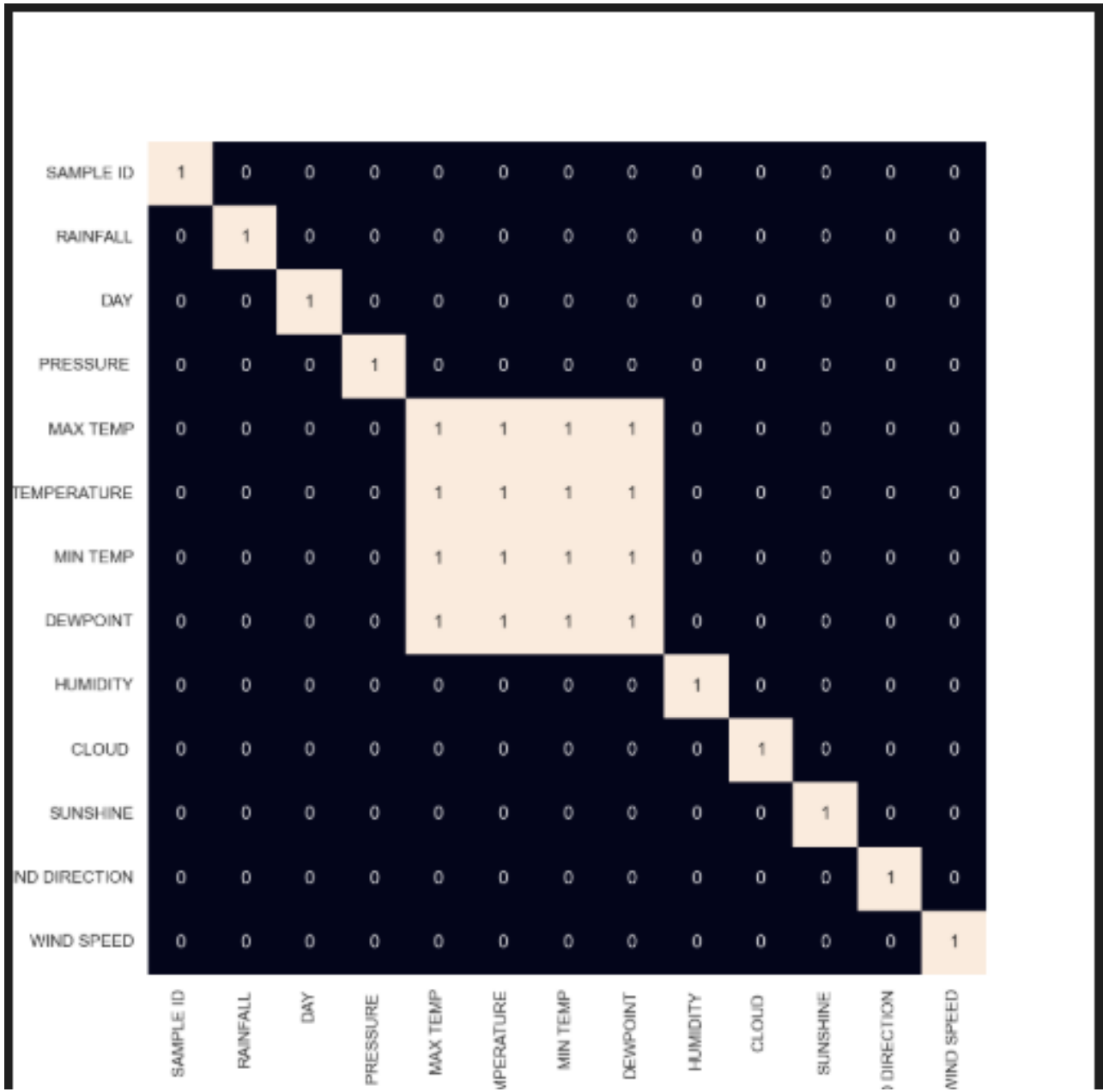
SAMPLE ID	RAINFALL
1	1
2	1
3	1
4	1
5	1
6	1
7	0
8	0
9	0
10	1
11	1
12	0
13	0
14	1
15	1



FEATURE EXTRACTION

- For feature extraction, we made a heat map for all the existing features, DAY, PRESSURE, MAX TEMP, TEMPERATURE, MIN_TEMP, DEWPOINT, HUMIDITY, CLOUD, SUNSHINE, WIND DIRECTION, WIND SPEED, to see which were highly correlated.
- Since TEMPERATURE, MIN TEMP, AND MAX_TEMP are highly correlated and provide redundant information, we dropped MIN_TEMP and MAX_TEMP.
- The DAY column provided no information for the ML algorithms, so that feature was also dropped.

Feature Heat Map:



Extracted Features

ATTRIBUTE	TYPE
SAMPLE ID	INT
RAINFALL (TARGET)	STRING (‘yes’, ‘no’)
PRESSURE	FLOAT
TEMPERATURE	FLOAT
DEW POINT	FLOAT
HUMIDITY	INT
CLOUD	INT
SUNSHINE	FLOAT
WIND DIRECTION	INT
WIND SPEED	FLOAT

Feature Samples

SAMPLE ID	RAINFALL	PRESSURE	TEMPERATURE	DEWPOINT	HUMIDITY	CLOUD	SUNSHINE	WIND DIRECTION	WIND SPEED
1	1	1025.9	64.94	13.1	72	49	9.3	80	26.3
2	1	1022	66.02	15.6	81	83	0.6	50	15.3
3	1	1019.7	66.74	18.4	95	91	0	40	14.2
4	1	1018.9	69.08	18.8	90	88	1	50	16.9
5	1	1015.9	69.26	19.9	95	81	0	40	13.7
6	1	1018.8	69.62	18	84	51	7.7	20	14.5
7	0	1021.8	65.84	15	79	56	3.4	30	21.5
8	0	1020.8	65.12	14.4	78	28	7.7	60	14.3
9	0	1020.6	64.58	14.3	78	79	3.3	70	39.3
10	1	1017.5	64.4	15.5	85	91	0	70	37.7
11	1	1016.5	64.58	16.4	90	90	2.1	40	23.3
12	0	1019.9	63.14	13.7	79	86	0.6	20	23.9
13	0	1020.8	60.98	12.1	77	34	9.1	30	24.4
14	1	1019.3	61.7	12.9	79	81	1.5	60	33.2
15	1	1015.4	59.18	14.6	97	97	0	50	37.5
16	1	1013.5	61.52	15.6	95	93	0	60	40
17	1	1011.5	64.04	16.1	90	79	1.6	20	23.4
18	0	1017.1	59.36	11.1	76	49	3.9	50	28.4
19	0	1020.1	61.52	12.5	78	84	1	60	38
20	1	1019.6	59.9	13.9	90	92	0	70	50.6

TASK COMPLETION REPORT

DATE	TASK NAME	STATUS	PERSON
2/26/25	CREATE GITHUB REPOSITORY	COMPLETE	SANELE HARMON
2/27/25	PLOT DISTRIBUTIONS OF THE DATA	COMPLETE	TAEDEN KITCHENS
2/29/25	PREPROCESS TEMPERATURE COLUMNS AND PLOT	COMPLETE	SANELE HARMON
2/29/25	PREPROCESS RAINFALL COLUMN AND PLOT	COMPLETE	TAEDEN KITCHENS
4/05/25	Discuss which features to drop	Complete	Sanele Harmon, Taeden Kitchens
4/07/25	Create Heat map	Complete	Taeden Kitchens
4/10/25	Add to PA2 report	Complete	Sanele Harmon