Plant Growth Data Classification

image.png

Plant Growth Data Classification

Problem Statement

Agricultural productivity relies on understanding and managing soil conditions to match the nutrient requirements of specific crops. With varying needs for nitrogen (N), phosphorus (P), and potassium (K), as well as sensitivity to soil temperature and pH levels, it is critical for farmers to determine which crops are most suitable for their soil conditions to optimize yield.

This dataset includes key soil properties—nutrient levels (N, P, K), average soil temperature (°F), and pH (acidity or basicity)—and maps them to specific crop types (e.g., rice, maize, various legumes, fruits, and cash crops like cotton and coffee). Using this data, we aim to develop a predictive model that suggests the most suitable crop types based on given soil characteristics.

Objective

To build a machine learning model that recommends the best crop(s) based on soil nutrient levels (N, P, K), temperature, and pH, thereby supporting informed decision-making for farmers and improving agricultural efficiency.

Goals

1. **Analyze Soil-Condition Requirements for Different Crops**: Explore the soil nutrient composition, temperature, and pH ranges ideal for each crop in the dataset.
2. **Develop Predictive Model**: Train a classification model that suggests optimal crop choices based on the input conditions of N, P, K levels, temperature, and pH.
3. **Evaluate Model Accuracy**: Assess the model's accuracy in correctly recommending crop types by validating it against test data. Provide Insights for Crop Rotation or Soil Amendment: Use the model to suggest soil amendments or crop rotation strategies to achieve better suitability for desired crops.

Importing Libraries and Loading dataset

```
import os
os.getcwd()
```

```
'C:\\Users\\Lenovo\\Desktop\\HealthCare\\Agreeculture Domain'
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
sns.set()
%matplotlib inline

from sklearn import model_selection
from sklearn import preprocessing
from sklearn.pipeline import make_pipeline
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.neighbors import KNeighborsClassifier

from sklearn.metrics import confusion_matrix, classification_report,
accuracy_score

import warnings
warnings.filterwarnings("ignore")

df = pd.read_csv("/content/Plan_Growth_recommendation.csv")
df.head()
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 2200,\n  \"fields\":
[\n    {\n      \"column\": \"N\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 36,\n        \"min\": 0,\n
\"max\": 140,\n        \"num_unique_values\": 137,\n
\"samples\": [\n          106,\n          101,\n          88\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"P\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 32,\n        \"min\": 5,\n        \"max\": 145,\n
\"num_unique_values\": 117,\n        \"samples\": [\n          69,\n
37,\n          11\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"K\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 50,\n        \"min\": 5,\n        \"max\": 205,\n
\"num_unique_values\": 73,\n        \"samples\": [\n          42,\n
12,\n          15\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"temperature\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 5.063748599958843,\n        \"min\":
8.825674745,\n        \"max\": 43.67549305,\n
\"num_unique_values\": 2200,\n        \"samples\": [\n
29.49401389,\n          26.1793464,\n          43.36051537\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"humidity\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 22.263811589761115,\n        \"min\":
14.25803981,\n        \"max\": 99.98187601,\n
\"num_unique_values\": 2200,\n        \"samples\": [\n
94.72981338,\n          86.52258079,\n          93.35191636\
```

```
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"ph\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 0.7739376880298721,\n        \"min\": 3.504752314,\n
\"max\": 9.93509073,\n        \"num_unique_values\": 2200,\n
\"samples\": [\n          6.185053234,\n          6.25933595,\n
6.941496806\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"rainfall\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 54.95838852487811,\n        \"min\":
20.21126747,\n        \"max\": 298.5601175,\n
\"num_unique_values\": 2200,\n        \"samples\": [\n
26.30820876,\n          49.43050977,\n          114.778071\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"label\",\n      \"properties\": {\n        \"dtype\": \"category\",\
n        \"num_unique_values\": 22,\n        \"samples\": [\n
\"rice\",\n          \"watermelon\",\n          \"lentil\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```


```
print("Shape of the dataframe :", df.shape)
df.isna().sum()
```

```
Shape of the dataframe : (2200, 8)
```

```
N              0
P              0
K              0
temperature    0
humidity       0
ph             0
rainfall       0
label          0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2200 entries, 0 to 2199
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   N            2200 non-null   int64
 1   P            2200 non-null   int64
 2   K            2200 non-null   int64
 3   temperature  2200 non-null   float64
 4   humidity     2200 non-null   float64
 5   ph           2200 non-null   float64
```

```
 6   rainfall     2200 non-null   float64
 7   label        2200 non-null   object
dtypes: float64(4), int64(3), object(1)
memory usage: 137.6+ KB
```

```
df.describe()
```

```
                 N            P            K  temperature     humidity
\
count  2200.000000  2200.000000  2200.000000  2200.000000  2200.000000

mean     50.551818    53.362727    48.149091    25.616244    71.481779

std      36.917334    32.985883    50.647931     5.063749    22.263812

min       0.000000     5.000000     5.000000     8.825675    14.258040

25%      21.000000    28.000000    20.000000    22.769375    60.261953

50%      37.000000    51.000000    32.000000    25.598693    80.473146

75%      84.250000    68.000000    49.000000    28.561654    89.948771

max     140.000000   145.000000   205.000000    43.675493    99.981876


                ph     rainfall
count  2200.000000  2200.000000
mean      6.469480   103.463655
std       0.773938    54.958389
min       3.504752    20.211267
25%       5.971693    64.551686
50%       6.425045    94.867624
75%       6.923643   124.267508
max       9.935091   298.560117
```

```
df.dtypes
```

```
N                int64
P                int64
K                int64
temperature    float64
humidity       float64
ph             float64
rainfall       float64
label           object
dtype: object
```

```
df['label'].value_counts()
```

```
label
rice            100
maize           100
chickpea        100
kidneybeans     100
pigeonpeas      100
mothbeans       100
mungbean        100
blackgram       100
lentil          100
pomegranate     100
banana          100
mango           100
grapes          100
watermelon      100
muskmelon       100
apple           100
orange          100
papaya          100
coconut         100
cotton          100
jute            100
coffee          100
Name: count, dtype: int64
```

EDA - Exploratory Data Analysis

sweetviz

```
#!pip install sweetviz

import sweetviz as sv
report  = sv.analyze(df)
report.show_html('sweetviz_report.html')
```


```
{"model_id":"34cf0c2ff5e84e52956fa3dc0a0dec8a","version_major":2,"vers
ion_minor":0}
```


```
Report sweetviz_report.html was generated! NOTEBOOK/COLAB USERS: the
web browser MAY not pop up, regardless, the report IS saved in your
notebook/colab files.
```

autoviz

```
#!pip install autoviz==0.0.6
```

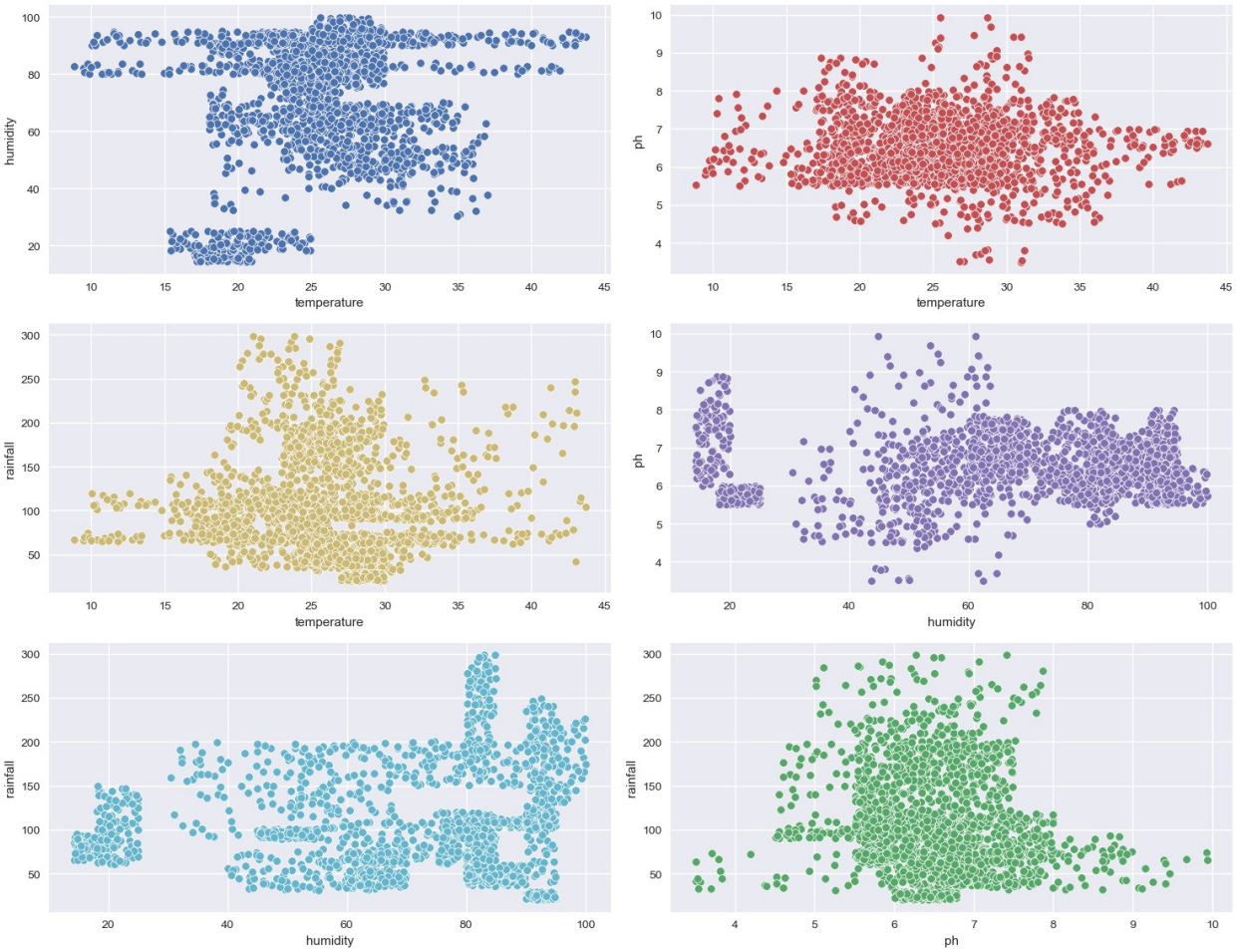
```
from autoviz.AutoViz_Class import AutoViz_Class
autoviz = AutoViz_Class()
autoviz.AutoViz(filename = "", dfte=df)
```

```
Shape of your Data Set: (2200, 8)
Classifying variables in data set...
    8 Predictors classified...
        This does not include the Target column(s)
    1 variables removed since they were ID or low-information
variables
Number of All Scatter Plots = 10
Could not draw Violin Plot
Could not draw Bar Plots
Time to run AutoViz (in seconds) = 3.910
```

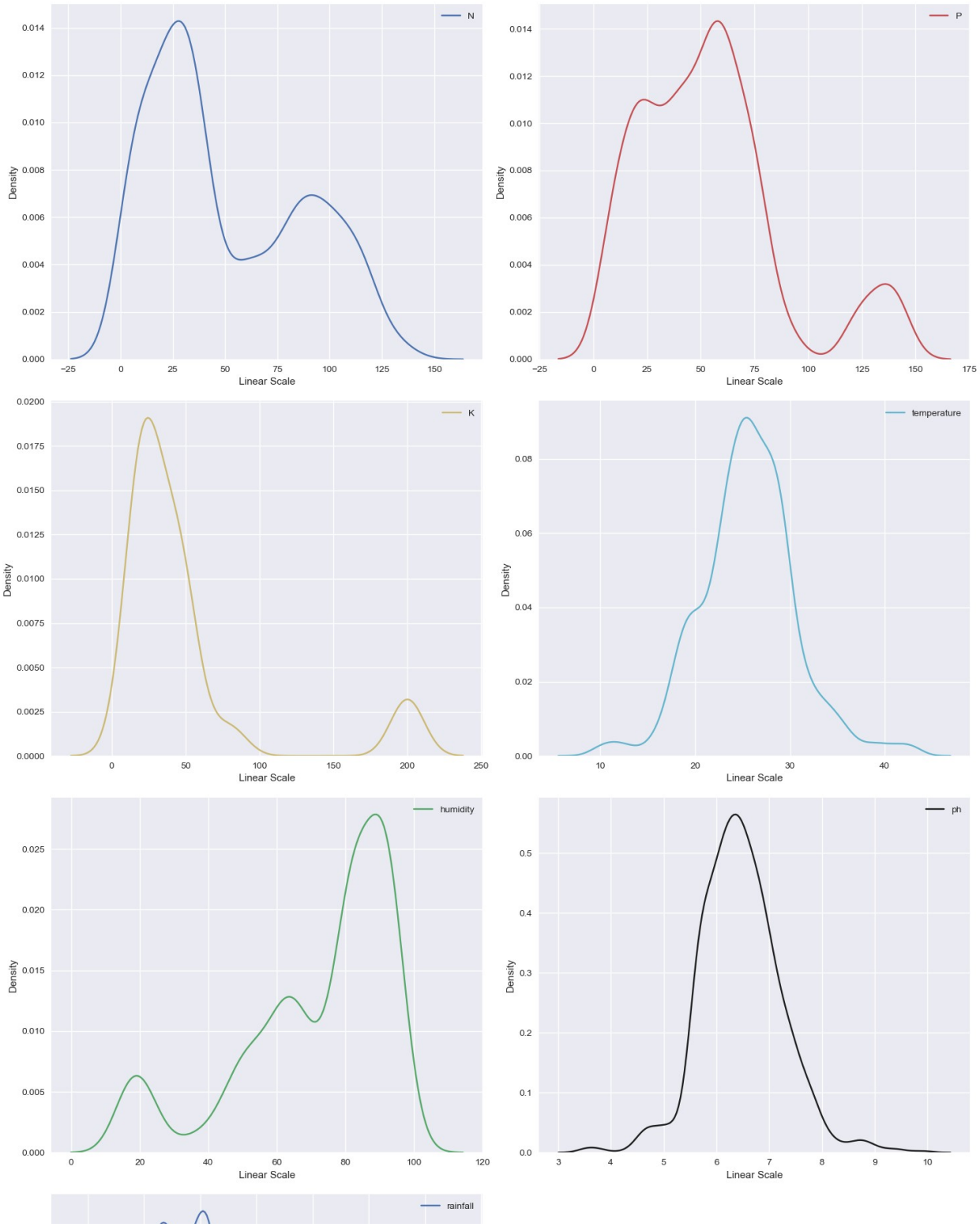
	N	P	K	temperature	humidity	ph	rainfall
0	90	42	43	20.879744	82.002744	6.502985	202.935536
1	85	58	41	21.770462	80.319644	7.038096	226.655537
2	60	55	44	23.004459	82.320763	7.840207	263.964248
3	74	35	40	26.491096	80.158363	6.980401	242.864034
4	78	42	42	20.130175	81.604873	7.628473	262.717340
...	...	..	..	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507
2196	99	15	27	27.417112	56.636362	6.086922	127.924610
2197	118	33	30	24.131797	67.225123	6.362608	173.322839
2198	117	32	34	26.272418	52.127394	6.758793	127.175293
2199	104	18	30	23.603016	60.396475	6.779833	140.937041

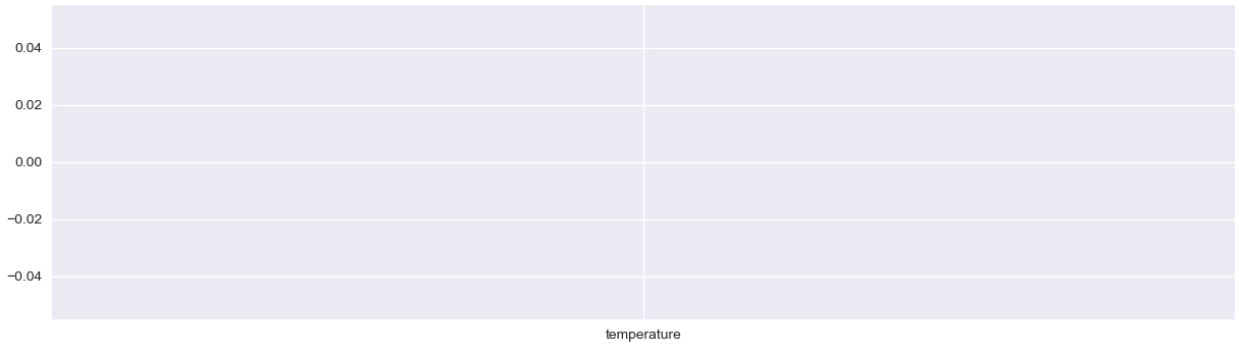
```
[2200 rows x 7 columns]
```

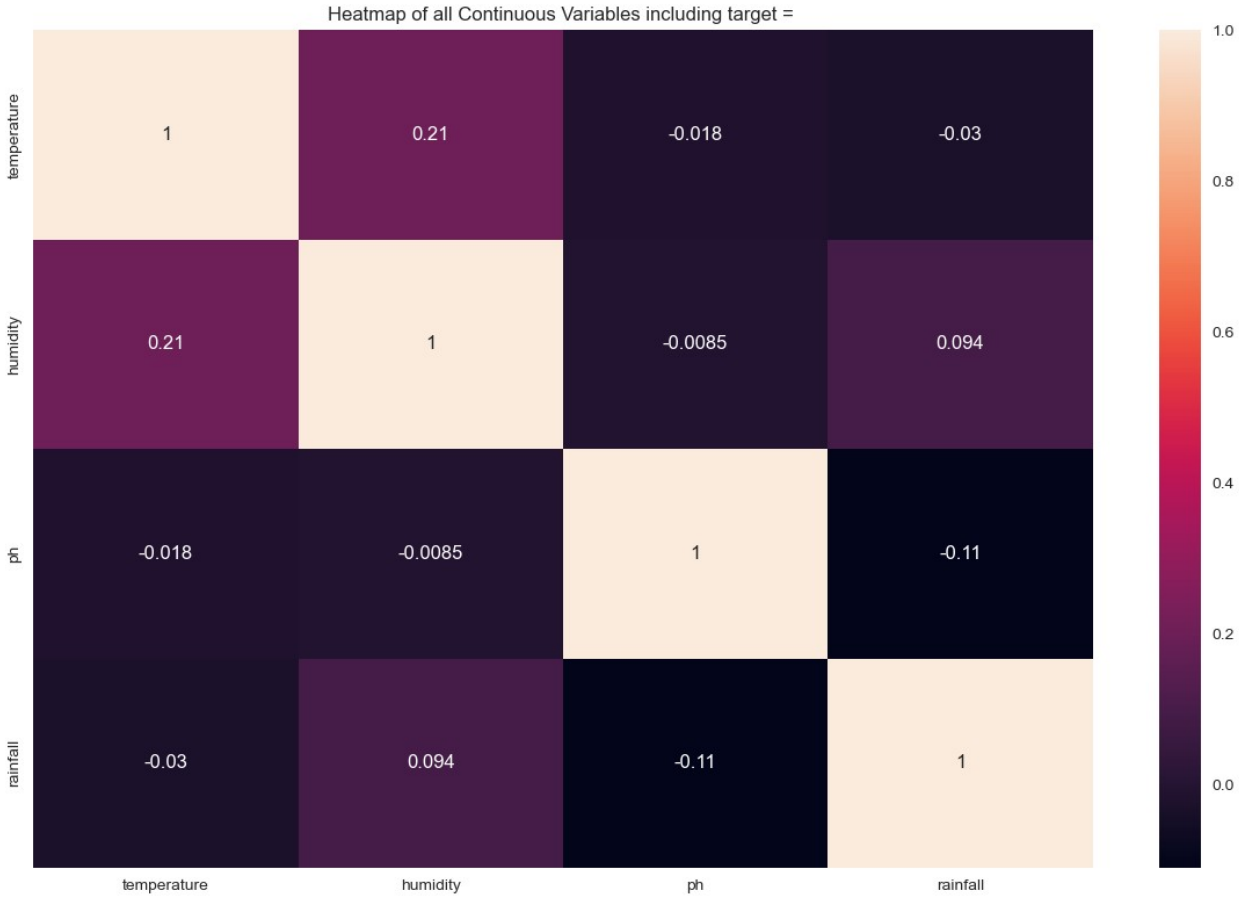
Pair-wise Scatter Plot of all Continuous Variables

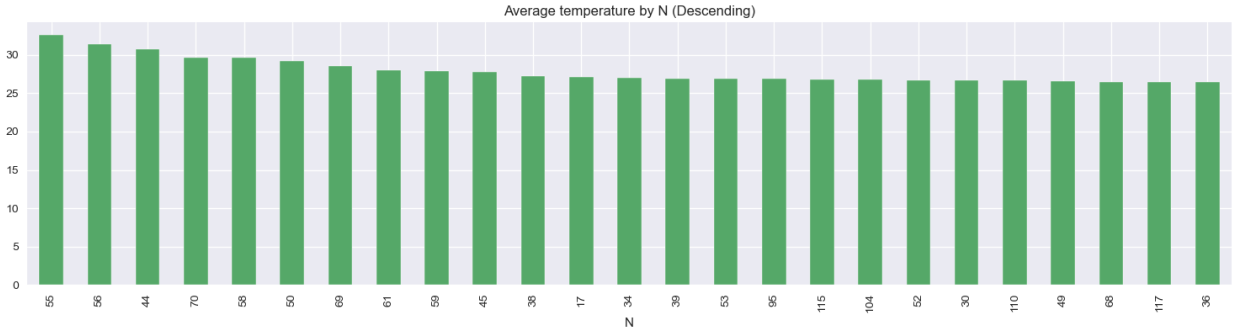


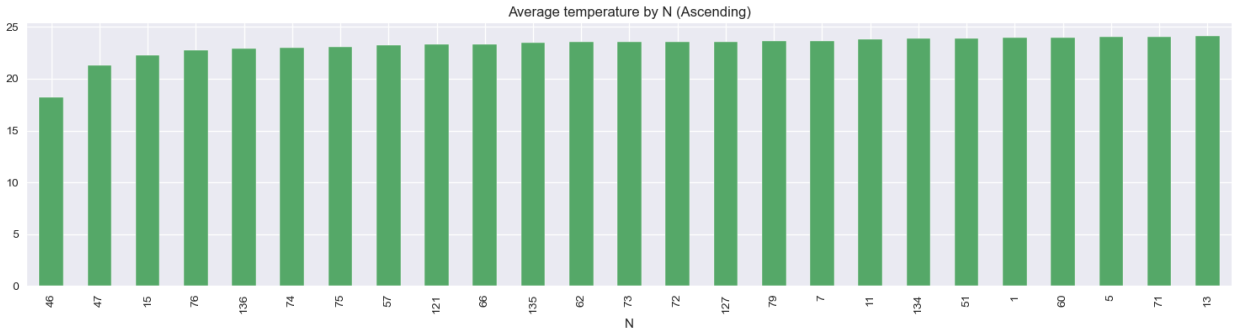
Histograms (KDE plots) of all Continuous Variables

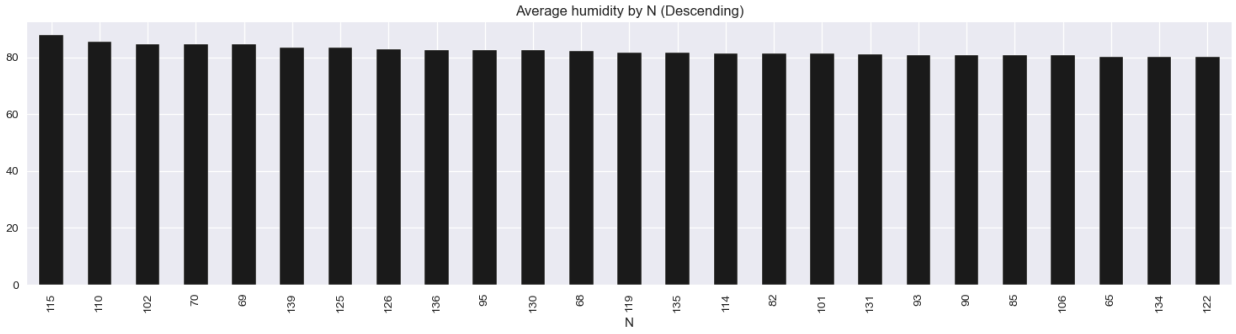


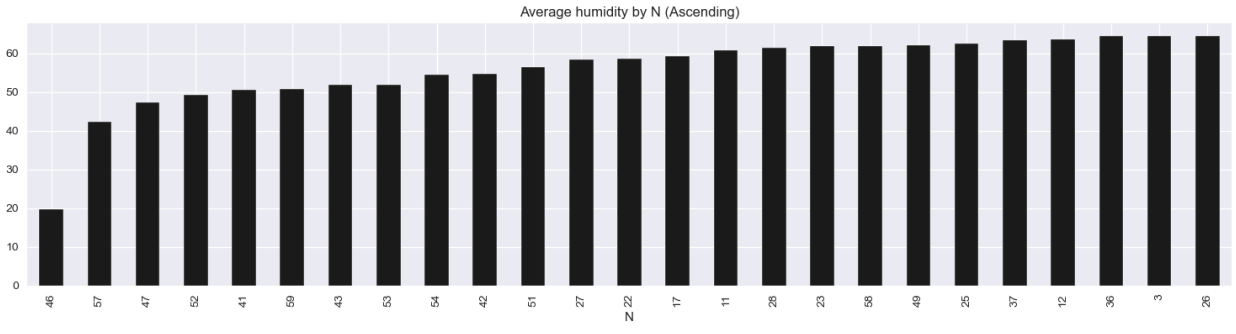

0.04
0.02
0.00
−0.02
−0.04
temperature

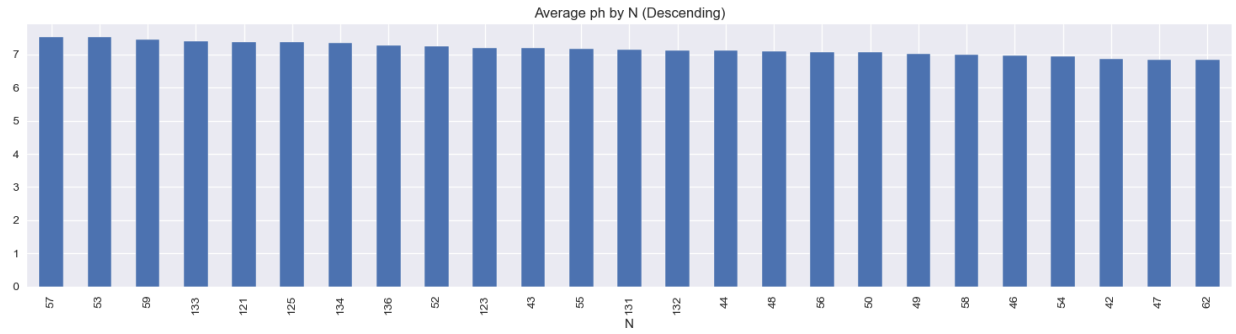

Heatmap of all Continuous Variables including target =
temperature
humidity
ph
rainfall
1
0.21
-0.018
-0.03
0.21
1
-0.0085
0.094
-0.018
-0.0085
1
-0.11
-0.03
0.094
-0.11
1
temperature
humidity
ph
rainfall
1.0
0.8
0.6
0.4
0.2
0.0

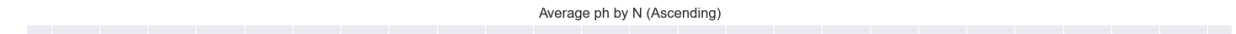

Average temperature by N (Descending)
0
5
10
15
20
25
30
55 56 44 70 58 50 69 61 59 45 38 17 34 39 53 95 115 104 52 30 110 49 68 117 36
N

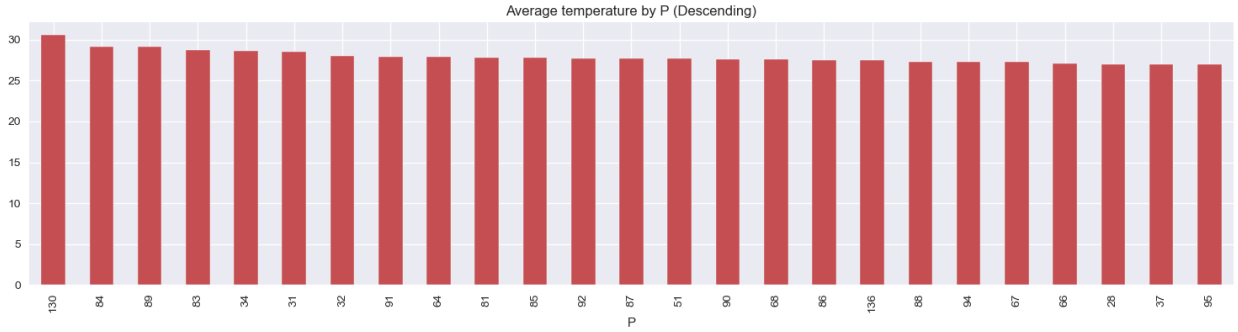

Average temperature by N (Ascending)
0
5
10
15
20
25
46 47 15 76 136 74 75 57 121 66 135 62 73 72 127 79 7 11 134 51 1 60 5 71 13
N

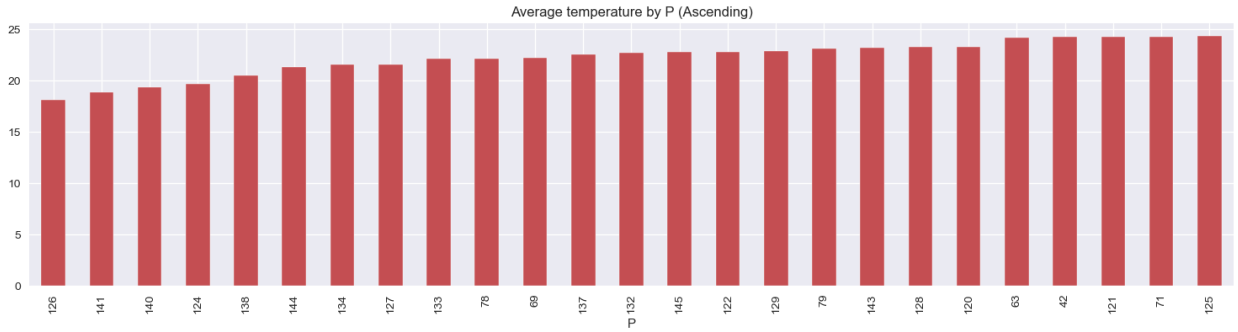

Average humidity by N (Descending)
0
20
40
60
80
115 110 102 70 69 139 125 126 136 95 130 68 119 135 114 82 101 131 93 90 85 106 65 134 122
N

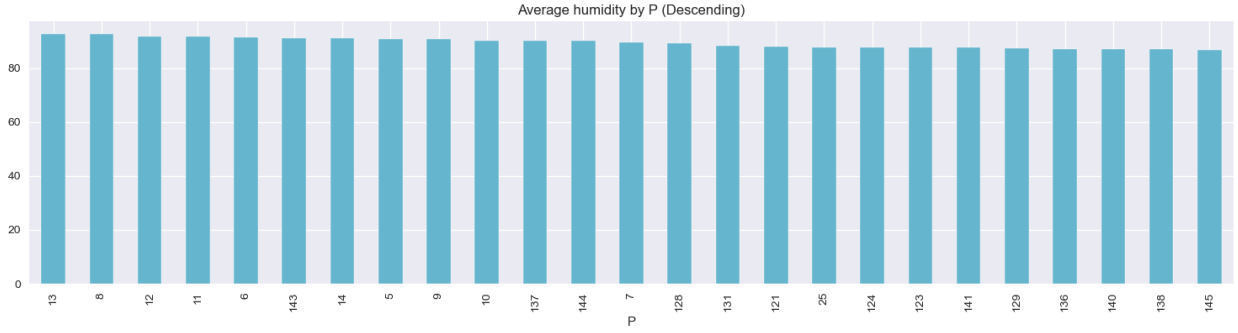

Average humidity by N (Ascending)
0
10
20
30
40
50
60
46 57 47 52 41 59 43 53 54 42 51 27 22 17 11 28 23 58 49 25 37 12 36 3 26
N

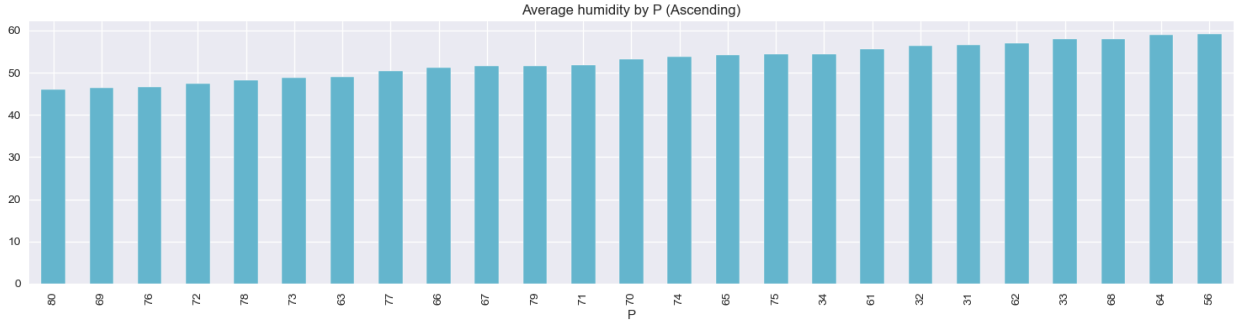

Average ph by N (Descending)
0
1
2
3
4
5
6
7
57 53 59 133 121 125 134 136 52 123 43 55 131 132 44 48 56 50 49 58 46 54 42 47 62
N

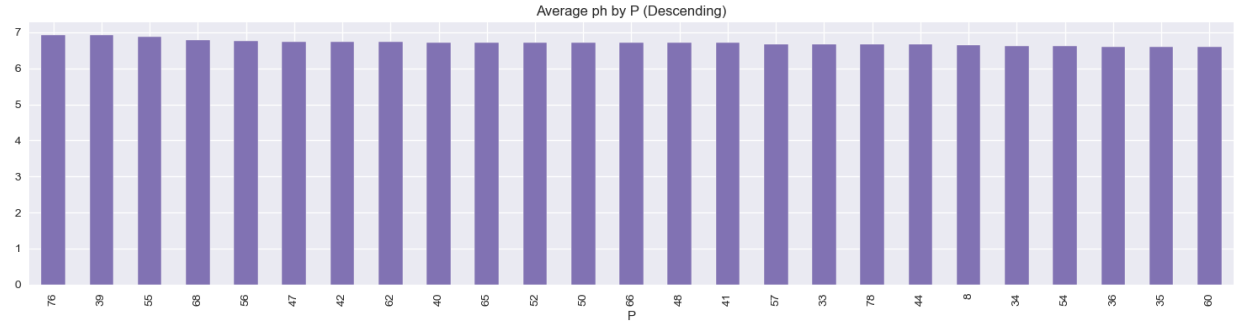

Average ph by N (Ascending)

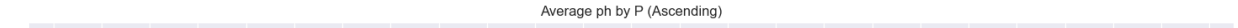

Average temperature by P (Descending)
30
25
20
15
10
5
0
130 84 89 83 34 31 32 91 64 81 85 92 87 51 90 68 86 136 88 94 67 66 28 37 95
P


Average temperature by P (Ascending)
25
20
15
10
5
0
126 141 140 124 138 144 134 127 133 78 69 137 132 145 122 129 79 143 128 120 63 42 121 71 125
P


Average humidity by P (Descending)
80
60
40
20
0
13 8 12 11 6 143 14 5 9 10 137 144 7 128 131 121 25 124 123 141 129 136 140 138 145
P


Average humidity by P (Ascending)
60
50
40
30
20
10
0
80 69 76 72 78 73 63 77 66 67 79 71 70 74 65 75 34 61 32 31 62 33 68 64 56
P


Average ph by P (Descending)
7
6
5
4
3
2
1
0
76 39 55 68 56 47 42 62 40 65 52 50 66 48 41 57 33 78 44 8 34 54 36 35 60
P


Average ph by P (Ascending)

```
<Figure size 1500x4800 with 0 Axes>
```

```
!pip install --upgrade autoviz
```

```
Collecting autoviz
  Downloading autoviz-0.1.905-py3-none-any.whl.metadata (14 kB)
Requirement already satisfied: xlrd in /usr/local/lib/python3.10/dist-
packages (from autoviz) (2.0.1)
Requirement already satisfied: wordcloud in
/usr/local/lib/python3.10/dist-packages (from autoviz) (1.9.4)
Collecting emoji (from autoviz)
  Downloading emoji-2.14.0-py3-none-any.whl.metadata (5.7 kB)
Collecting pyamg (from autoviz)
  Downloading pyamg-5.2.1-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (8.1 kB)
Requirement already satisfied: scikit-learn in
/usr/local/lib/python3.10/dist-packages (from autoviz) (1.5.2)
Requirement already satisfied: statsmodels in
/usr/local/lib/python3.10/dist-packages (from autoviz) (0.14.4)
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-
packages (from autoviz) (3.9.1)
Requirement already satisfied: textblob in
/usr/local/lib/python3.10/dist-packages (from autoviz) (0.17.1)
Collecting xgboost<1.7,>=0.82 (from autoviz)
  Downloading xgboost-1.6.2-py3-none-manylinux2014_x86_64.whl.metadata
(1.8 kB)
Requirement already satisfied: fsspec>=0.8.3 in
/usr/local/lib/python3.10/dist-packages (from autoviz) (2024.10.0)
Requirement already satisfied: typing-extensions>=4.1.1 in
/usr/local/lib/python3.10/dist-packages (from autoviz) (4.12.2)
Collecting pandas-dq>=1.29 (from autoviz)
  Downloading pandas_dq-1.29-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: numpy>=1.24.0 in
/usr/local/lib/python3.10/dist-packages (from autoviz) (1.26.4)
Collecting hvplot>=0.9.2 (from autoviz)
  Downloading hvplot-0.11.1-py3-none-any.whl.metadata (15 kB)
Requirement already satisfied: holoviews>=1.16.0 in
/usr/local/lib/python3.10/dist-packages (from autoviz) (1.20.0)
Requirement already satisfied: panel>=1.4.0 in
/usr/local/lib/python3.10/dist-packages (from autoviz) (1.5.4)
Requirement already satisfied: pandas>=2.0 in
/usr/local/lib/python3.10/dist-packages (from autoviz) (2.2.2)
Requirement already satisfied: matplotlib>3.7.4 in
/usr/local/lib/python3.10/dist-packages (from autoviz) (3.8.0)
Requirement already satisfied: seaborn>0.12.2 in
/usr/local/lib/python3.10/dist-packages (from autoviz) (0.13.2)
Requirement already satisfied: bokeh>=3.1 in
/usr/local/lib/python3.10/dist-packages (from holoviews>=1.16.0-
>autoviz) (3.6.1)
Requirement already satisfied: colorcet in
```

```
/usr/local/lib/python3.10/dist-packages (from holoviews>=1.16.0-
>autoviz) (3.1.0)
Requirement already satisfied: packaging in
/usr/local/lib/python3.10/dist-packages (from holoviews>=1.16.0-
>autoviz) (24.2)
Requirement already satisfied: param<3.0,>=2.0 in
/usr/local/lib/python3.10/dist-packages (from holoviews>=1.16.0-
>autoviz) (2.1.1)
Requirement already satisfied: pyviz-comms>=2.1 in
/usr/local/lib/python3.10/dist-packages (from holoviews>=1.16.0-
>autoviz) (3.0.3)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>3.7.4-
>autoviz) (1.3.1)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>3.7.4-
>autoviz) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>3.7.4-
>autoviz) (4.54.1)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>3.7.4-
>autoviz) (1.4.7)
Requirement already satisfied: pillow>=6.2.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>3.7.4-
>autoviz) (11.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>3.7.4-
>autoviz) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.10/dist-packages (from matplotlib>3.7.4-
>autoviz) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas>=2.0->autoviz)
(2024.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.10/dist-packages (from pandas>=2.0->autoviz)
(2024.2)
Requirement already satisfied: bleach in
/usr/local/lib/python3.10/dist-packages (from panel>=1.4.0->autoviz)
(6.2.0)
Requirement already satisfied: linkify-it-py in
/usr/local/lib/python3.10/dist-packages (from panel>=1.4.0->autoviz)
(2.0.3)
Requirement already satisfied: markdown in
/usr/local/lib/python3.10/dist-packages (from panel>=1.4.0->autoviz)
(3.7)
Requirement already satisfied: markdown-it-py in
/usr/local/lib/python3.10/dist-packages (from panel>=1.4.0->autoviz)
```

```
(3.0.0)
Requirement already satisfied: mdit-py-plugins in
/usr/local/lib/python3.10/dist-packages (from panel>=1.4.0->autoviz)
(0.4.2)
Requirement already satisfied: requests in
/usr/local/lib/python3.10/dist-packages (from panel>=1.4.0->autoviz)
(2.32.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-
packages (from panel>=1.4.0->autoviz) (4.66.6)
Requirement already satisfied: scipy>=1.6.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn->autoviz)
(1.13.1)
Requirement already satisfied: joblib>=1.2.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn->autoviz)
(1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/usr/local/lib/python3.10/dist-packages (from scikit-learn->autoviz)
(3.5.0)
Requirement already satisfied: click in
/usr/local/lib/python3.10/dist-packages (from nltk->autoviz) (8.1.7)
Requirement already satisfied: regex>=2021.8.3 in
/usr/local/lib/python3.10/dist-packages (from nltk->autoviz)
(2024.9.11)
Requirement already satisfied: patsy>=0.5.6 in
/usr/local/lib/python3.10/dist-packages (from statsmodels->autoviz)
(1.0.1)
Requirement already satisfied: Jinja2>=2.9 in
/usr/local/lib/python3.10/dist-packages (from bokeh>=3.1-
>holoviews>=1.16.0->autoviz) (3.1.4)
Requirement already satisfied: PyYAML>=3.10 in
/usr/local/lib/python3.10/dist-packages (from bokeh>=3.1-
>holoviews>=1.16.0->autoviz) (6.0.2)
Requirement already satisfied: tornado>=6.2 in
/usr/local/lib/python3.10/dist-packages (from bokeh>=3.1-
>holoviews>=1.16.0->autoviz) (6.3.3)
Requirement already satisfied: xyzservices>=2021.09.1 in
/usr/local/lib/python3.10/dist-packages (from bokeh>=3.1-
>holoviews>=1.16.0->autoviz) (2024.9.0)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7-
>matplotlib>3.7.4->autoviz) (1.16.0)
Requirement already satisfied: webencodings in
/usr/local/lib/python3.10/dist-packages (from bleach->panel>=1.4.0-
>autoviz) (0.5.1)
Requirement already satisfied: uc-micro-py in
/usr/local/lib/python3.10/dist-packages (from linkify-it-py-
>panel>=1.4.0->autoviz) (1.0.3)
Requirement already satisfied: mdurl~=0.1 in
/usr/local/lib/python3.10/dist-packages (from markdown-it-py-
```

```
>panel>=1.4.0->autoviz) (0.1.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->panel>=1.4.0-
>autoviz) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.10/dist-packages (from requests->panel>=1.4.0-
>autoviz) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->panel>=1.4.0-
>autoviz) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->panel>=1.4.0-
>autoviz) (2024.8.30)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from Jinja2>=2.9->bokeh>=3.1-
>holoviews>=1.16.0->autoviz) (3.0.2)
Downloading autoviz-0.1.905-py3-none-any.whl (67 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 67.5/67.5 kB 4.7 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 161.2/161.2 kB 8.9 MB/s eta
0:00:00
anylinux2014_x86_64.whl (255.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 255.9/255.9 MB 5.6 MB/s eta
0:00:00
oji-2.14.0-py3-none-any.whl (586 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 586.9/586.9 kB 43.7 MB/s eta
0:00:00
g-5.2.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
(2.0 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.0/2.0 MB 87.5 MB/s eta
0:00:00
oji, xgboost, pyamg, pandas-dq, hvplot, autoviz
  Attempting uninstall: xgboost
    Found existing installation: xgboost 2.1.2
    Uninstalling xgboost-2.1.2:
      Successfully uninstalled xgboost-2.1.2
Successfully installed autoviz-0.1.905 emoji-2.14.0 hvplot-0.11.1
pandas-dq-1.29 pyamg-5.2.1 xgboost-1.6.2
```

```
import autoviz
print(autoviz.__version__)
```

```
Imported v0.1.905. Please call AutoViz in this sequence:
    AV = AutoViz_Class()
    %matplotlib inline
    dfte = AV.AutoViz(filename, sep=',', depVar='', dfte=None,
header=0, verbose=1, lowess=False,

chart_format='svg',max_rows_analyzed=150000,max_cols_analyzed=30,
```

```
save_plot_dir=None)
0.1.905

from autoviz.AutoViz_Class import AutoViz_Class
autoviz = AutoViz_Class()
autoviz.AutoViz(filename = "", dfte=df)

Shape of your Data Set loaded: (2200, 8)
#####################################################################
##################
######################### C L A S S I F Y I N G  V A R I A B L E S
####################
#####################################################################
##################
Classifying variables in data set...
    Number of Numeric Columns =  4
    Number of Integer-Categorical Columns =  3
    Number of String-Categorical Columns =  1
    Number of Factor-Categorical Columns =  0
    Number of String-Boolean Columns =  0
    Number of Numeric-Boolean Columns =  0
    Number of Discrete String Columns =  0
    Number of NLP String Columns =  0
    Number of Date Time Columns =  0
    Number of ID Columns =  0
    Number of Columns to Delete =  0
    8 Predictors classified...
        No variables removed since no ID or low-information variables
found in data set
To fix these data quality issues in the dataset, import FixDQ from
autoviz...
    All variables classified into correct types.

<pandas.io.formats.style.Styler at 0x7bc9efe9f700>

Number of All Scatter Plots = 10
All Plots done
Time to run AutoViz = 4 seconds

 ###################### AUTO VISUALIZATION Completed
########################

{"summary":"{\n  \"name\": \"autoviz\",\n  \"rows\": 2200,\n
\"fields\": [\n    {\n      \"column\": \"N\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 36,\n
\"min\": 0,\n        \"max\": 140,\n        \"num_unique_values\":
137,\n        \"samples\": [\n          106,\n          101,\n
88\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"P\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
```

```
\"std\": 32,\n        \"min\": 5,\n        \"max\": 145,\n
\"num_unique_values\": 117,\n        \"samples\": [\n          69,\n
37,\n          11\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"K\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 50,\n        \"min\": 5,\n        \"max\": 205,\n
\"num_unique_values\": 73,\n        \"samples\": [\n          42,\n
12,\n          15\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"temperature\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 5.063748599958843,\n        \"min\":
8.825674745,\n        \"max\": 43.67549305,\n
\"num_unique_values\": 2200,\n        \"samples\": [\n
29.49401389,\n          26.1793464,\n          43.36051537\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"humidity\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 22.263811589761115,\n        \"min\":
14.25803981,\n        \"max\": 99.98187601,\n
\"num_unique_values\": 2200,\n        \"samples\": [\n
94.72981338,\n          86.52258079,\n          93.35191636\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"ph\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 0.7739376880298721,\n        \"min\": 3.504752314,\n
\"max\": 9.93509073,\n        \"num_unique_values\": 2200,\n
\"samples\": [\n          6.185053234,\n          6.25933595,\n
6.941496806\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"rainfall\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 54.95838852487811,\n        \"min\":
20.21126747,\n        \"max\": 298.5601175,\n
\"num_unique_values\": 2200,\n        \"samples\": [\n
26.30820876,\n          49.43050977,\n          114.778071\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"label\",\n      \"properties\": {\n        \"dtype\": \"category\",\
n        \"num_unique_values\": 22,\n        \"samples\": [\n
\"rice\",\n          \"watermelon\",\n          \"lentil\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\n}","type":"dataframe"}
```


```
<google.colab._quickchart_helpers.SectionTitle at 0x7bc9e5a029b0>
```

```
from matplotlib import pyplot as plt
_df_0['index'].plot(kind='hist', bins=20, title='index')
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_1['N'].plot(kind='hist', bins=20, title='N')
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_2['P'].plot(kind='hist', bins=20, title='P')
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_3['K'].plot(kind='hist', bins=20, title='K')
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
<google.colab._quickchart_helpers.SectionTitle at 0x7bca05822260>
```

```
from matplotlib import pyplot as plt
_df_4.plot(kind='scatter', x='index', y='N', s=32, alpha=.8)
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_5.plot(kind='scatter', x='N', y='P', s=32, alpha=.8)
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_6.plot(kind='scatter', x='P', y='K', s=32, alpha=.8)
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_7.plot(kind='scatter', x='K', y='temperature', s=32, alpha=.8)
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
<google.colab._quickchart_helpers.SectionTitle at 0x7bc9e5980160>
```

```
from matplotlib import pyplot as plt
import seaborn as sns
def _plot_series(series, series_name, series_index=0):
  palette = list(sns.palettes.mpl_palette('Dark2'))
  xs = series['index']
  ys = series['N']

  plt.plot(xs, ys, label=series_name, color=palette[series_index %
len(palette)])

fig, ax = plt.subplots(figsize=(10, 5.2), layout='constrained')
df_sorted = _df_8.sort_values('index', ascending=True)
_plot_series(df_sorted, '')
sns.despine(fig=fig, ax=ax)
plt.xlabel('index')
_ = plt.ylabel('N')
```

```
from matplotlib import pyplot as plt
import seaborn as sns
def _plot_series(series, series_name, series_index=0):
```

```
  palette = list(sns.palettes.mpl_palette('Dark2'))
  xs = series['index']
  ys = series['P']

  plt.plot(xs, ys, label=series_name, color=palette[series_index %
len(palette)])

fig, ax = plt.subplots(figsize=(10, 5.2), layout='constrained')
df_sorted = _df_9.sort_values('index', ascending=True)
_plot_series(df_sorted, '')
sns.despine(fig=fig, ax=ax)
plt.xlabel('index')
_ = plt.ylabel('P')
```

```
from matplotlib import pyplot as plt
import seaborn as sns
def _plot_series(series, series_name, series_index=0):
  palette = list(sns.palettes.mpl_palette('Dark2'))
  xs = series['index']
  ys = series['K']

  plt.plot(xs, ys, label=series_name, color=palette[series_index %
len(palette)])

fig, ax = plt.subplots(figsize=(10, 5.2), layout='constrained')
df_sorted = _df_10.sort_values('index', ascending=True)
_plot_series(df_sorted, '')
sns.despine(fig=fig, ax=ax)
plt.xlabel('index')
_ = plt.ylabel('K')
```

```
from matplotlib import pyplot as plt
import seaborn as sns
def _plot_series(series, series_name, series_index=0):
  palette = list(sns.palettes.mpl_palette('Dark2'))
  xs = series['index']
  ys = series['temperature']

  plt.plot(xs, ys, label=series_name, color=palette[series_index %
len(palette)])

fig, ax = plt.subplots(figsize=(10, 5.2), layout='constrained')
df_sorted = _df_11.sort_values('index', ascending=True)
_plot_series(df_sorted, '')
sns.despine(fig=fig, ax=ax)
plt.xlabel('index')
_ = plt.ylabel('temperature')
```

```
<google.colab._quickchart_helpers.SectionTitle at 0x7bc9e5983a00>
```

```
from matplotlib import pyplot as plt
_df_12['index'].plot(kind='line', figsize=(8, 4), title='index')
plt.gca().spines[['top', 'right']].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_13['N'].plot(kind='line', figsize=(8, 4), title='N')
plt.gca().spines[['top', 'right']].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_14['P'].plot(kind='line', figsize=(8, 4), title='P')
plt.gca().spines[['top', 'right']].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_15['K'].plot(kind='line', figsize=(8, 4), title='K')
plt.gca().spines[['top', 'right']].set_visible(False)
```

Amazing automation for Model Building

PYCARET PACKAGE

```
!pip install pycaret
```

```
Collecting pycaret
  Downloading pycaret-3.3.2-py3-none-any.whl.metadata (17 kB)
Requirement already satisfied: ipython>=5.5.0 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (7.34.0)
Requirement already satisfied: ipywidgets>=7.6.5 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (7.7.1)
Requirement already satisfied: tqdm>=4.62.0 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (4.66.6)
Requirement already satisfied: numpy<1.27,>=1.21 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (1.26.4)
Collecting pandas<2.2.0 (from pycaret)
  Downloading pandas-2.1.4-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (18 kB)
Requirement already satisfied: jinja2>=3 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (3.1.4)
Collecting scipy<=1.11.4,>=1.6.1 (from pycaret)
  Downloading scipy-1.11.4-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (60 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 60.4/60.4 kB 3.9 MB/s eta
0:00:00
 pycaret)
  Downloading joblib-1.3.2-py3-none-any.whl.metadata (5.4 kB)
Requirement already satisfied: scikit-learn>1.4.0 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (1.5.2)
```

```
Collecting pyod>=1.1.3 (from pycaret)
  Downloading pyod-2.0.2.tar.gz (165 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 165.8/165.8 kB 12.2 MB/s eta
0:00:00
etadata (setup.py) ... ent already satisfied: imbalanced-learn>=0.12.0
in /usr/local/lib/python3.10/dist-packages (from pycaret) (0.12.4)
Collecting category-encoders>=2.4.0 (from pycaret)
  Downloading category_encoders-2.6.4-py2.py3-none-any.whl.metadata
(8.0 kB)
Requirement already satisfied: lightgbm>=3.0.0 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (4.5.0)
Requirement already satisfied: numba>=0.55.0 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (0.60.0)
Requirement already satisfied: requests>=2.27.1 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (2.32.3)
Requirement already satisfied: psutil>=5.9.0 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (5.9.5)
Requirement already satisfied: markupsafe>=2.0.1 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (3.0.2)
Requirement already satisfied: importlib-metadata>=4.12.0 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (8.5.0)
Requirement already satisfied: nbformat>=4.2.0 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (5.10.4)
Requirement already satisfied: cloudpickle in
/usr/local/lib/python3.10/dist-packages (from pycaret) (3.1.0)
Collecting deprecation>=2.1.0 (from pycaret)
  Downloading deprecation-2.1.0-py2.py3-none-any.whl.metadata (4.6 kB)
Collecting xxhash (from pycaret)
  Downloading xxhash-3.5.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting matplotlib<3.8.0 (from pycaret)
  Downloading matplotlib-3.7.5-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.7 kB)
Collecting scikit-plot>=0.3.7 (from pycaret)
  Downloading scikit_plot-0.3.7-py3-none-any.whl.metadata (7.1 kB)
Requirement already satisfied: yellowbrick>=1.4 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (1.5)
Requirement already satisfied: plotly>=5.14.0 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (5.24.1)
Collecting kaleido>=0.2.1 (from pycaret)
  Downloading kaleido-0.4.1-py3-none-any.whl.metadata (3.9 kB)
Collecting schemdraw==0.15 (from pycaret)
  Downloading schemdraw-0.15-py3-none-any.whl.metadata (2.2 kB)
Collecting plotly-resampler>=0.8.3.1 (from pycaret)
  Downloading plotly_resampler-0.10.0-py3-none-any.whl.metadata (13
kB)
Requirement already satisfied: statsmodels>=0.12.1 in
/usr/local/lib/python3.10/dist-packages (from pycaret) (0.14.4)
Collecting sktime==0.26.0 (from pycaret)
```

```
  Downloading sktime-0.26.0-py3-none-any.whl.metadata (29 kB)
Collecting tbats>=1.1.3 (from pycaret)
  Downloading tbats-1.1.3-py3-none-any.whl.metadata (3.8 kB)
Collecting pmdarima>=2.0.4 (from pycaret)
  Downloading pmdarima-2.0.4-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_28_x86_64.whl.m
etadata (7.8 kB)
Collecting wurlitzer (from pycaret)
  Downloading wurlitzer-3.1.1-py3-none-any.whl.metadata (2.5 kB)
Requirement already satisfied: packaging in
/usr/local/lib/python3.10/dist-packages (from sktime==0.26.0->pycaret)
(24.2)
Collecting scikit-base<0.8.0 (from sktime==0.26.0->pycaret)
  Downloading scikit_base-0.7.8-py3-none-any.whl.metadata (8.8 kB)
Collecting scikit-learn>1.4.0 (from pycaret)
  Downloading scikit_learn-1.4.2-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (11 kB)
Requirement already satisfied: patsy>=0.5.1 in
/usr/local/lib/python3.10/dist-packages (from category-
encoders>=2.4.0->pycaret) (1.0.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from imbalanced-
learn>=0.12.0->pycaret) (3.5.0)
Requirement already satisfied: zipp>=3.20 in
/usr/local/lib/python3.10/dist-packages (from importlib-
metadata>=4.12.0->pycaret) (3.21.0)
Requirement already satisfied: setuptools>=18.5 in
/usr/local/lib/python3.10/dist-packages (from ipython>=5.5.0->pycaret)
(75.1.0)
Collecting jedi>=0.16 (from ipython>=5.5.0->pycaret)
  Downloading jedi-0.19.2-py2.py3-none-any.whl.metadata (22 kB)
Requirement already satisfied: decorator in
/usr/local/lib/python3.10/dist-packages (from ipython>=5.5.0->pycaret)
(4.4.2)
Requirement already satisfied: pickleshare in
/usr/local/lib/python3.10/dist-packages (from ipython>=5.5.0->pycaret)
(0.7.5)
Requirement already satisfied: traitlets>=4.2 in
/usr/local/lib/python3.10/dist-packages (from ipython>=5.5.0->pycaret)
(5.7.1)
Requirement already satisfied: prompt-toolkit!=3.0.0,!
=3.0.1,<3.1.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from
ipython>=5.5.0->pycaret) (3.0.48)
Requirement already satisfied: pygments in
/usr/local/lib/python3.10/dist-packages (from ipython>=5.5.0->pycaret)
(2.18.0)
Requirement already satisfied: backcall in
/usr/local/lib/python3.10/dist-packages (from ipython>=5.5.0->pycaret)
(0.2.0)
```

```
Requirement already satisfied: matplotlib-inline in
/usr/local/lib/python3.10/dist-packages (from ipython>=5.5.0->pycaret)
(0.1.7)
Requirement already satisfied: pexpect>4.3 in
/usr/local/lib/python3.10/dist-packages (from ipython>=5.5.0->pycaret)
(4.9.0)
Requirement already satisfied: ipykernel>=4.5.1 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.6.5-
>pycaret) (5.5.6)
Requirement already satisfied: ipython-genutils~=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.6.5-
>pycaret) (0.2.0)
Requirement already satisfied: widgetsnbextension~=3.6.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.6.5-
>pycaret) (3.6.10)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in
/usr/local/lib/python3.10/dist-packages (from ipywidgets>=7.6.5-
>pycaret) (3.0.13)
Collecting choreographer>=0.99.6 (from kaleido>=0.2.1->pycaret)
  Downloading choreographer-0.99.6-py3-none-any.whl.metadata (5.7 kB)
Requirement already satisfied: async-timeout in
/usr/local/lib/python3.10/dist-packages (from kaleido>=0.2.1->pycaret)
(4.0.3)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib<3.8.0-
>pycaret) (1.3.1)
Requirement already satisfied: cycler>=0.10 in
/usr/local/lib/python3.10/dist-packages (from matplotlib<3.8.0-
>pycaret) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib<3.8.0-
>pycaret) (4.54.1)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib<3.8.0-
>pycaret) (1.4.7)
Requirement already satisfied: pillow>=6.2.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib<3.8.0-
>pycaret) (11.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib<3.8.0-
>pycaret) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.10/dist-packages (from matplotlib<3.8.0-
>pycaret) (2.8.2)
Requirement already satisfied: fastjsonschema>=2.15 in
/usr/local/lib/python3.10/dist-packages (from nbformat>=4.2.0-
>pycaret) (2.20.0)
Requirement already satisfied: jsonschema>=2.6 in
/usr/local/lib/python3.10/dist-packages (from nbformat>=4.2.0-
```

```
>pycaret) (4.23.0)
Requirement already satisfied: jupyter-core!=5.0.*,>=4.12 in
/usr/local/lib/python3.10/dist-packages (from nbformat>=4.2.0-
>pycaret) (5.7.2)
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in
/usr/local/lib/python3.10/dist-packages (from numba>=0.55.0->pycaret)
(0.43.0)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas<2.2.0->pycaret)
(2024.2)
Requirement already satisfied: tzdata>=2022.1 in
/usr/local/lib/python3.10/dist-packages (from pandas<2.2.0->pycaret)
(2024.2)
Requirement already satisfied: tenacity>=6.2.0 in
/usr/local/lib/python3.10/dist-packages (from plotly>=5.14.0->pycaret)
(9.0.0)
Collecting dash>=2.9.0 (from plotly-resampler>=0.8.3.1->pycaret)
  Downloading dash-2.18.2-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: orjson<4.0.0,>=3.8.0 in
/usr/local/lib/python3.10/dist-packages (from plotly-
resampler>=0.8.3.1->pycaret) (3.10.11)
Collecting tsdownsample>=0.1.3 (from plotly-resampler>=0.8.3.1-
>pycaret)
  Downloading tsdownsample-0.1.3-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (7.9 kB)
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in
/usr/local/lib/python3.10/dist-packages (from pmdarima>=2.0.4-
>pycaret) (3.0.11)
Requirement already satisfied: urllib3 in
/usr/local/lib/python3.10/dist-packages (from pmdarima>=2.0.4-
>pycaret) (2.2.3)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.27.1-
>pycaret) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.27.1-
>pycaret) (3.10)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.27.1-
>pycaret) (2024.8.30)
Requirement already satisfied: Flask<3.1,>=1.0.4 in
/usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-
resampler>=0.8.3.1->pycaret) (3.0.3)
Collecting Werkzeug<3.1 (from dash>=2.9.0->plotly-resampler>=0.8.3.1-
>pycaret)
  Downloading werkzeug-3.0.6-py3-none-any.whl.metadata (3.7 kB)
Collecting dash-html-components==2.0.0 (from dash>=2.9.0->plotly-
resampler>=0.8.3.1->pycaret)
  Downloading dash_html_components-2.0.0-py3-none-any.whl.metadata
```

```
(3.8 kB)
Collecting dash-core-components==2.0.0 (from dash>=2.9.0->plotly-
resampler>=0.8.3.1->pycaret)
  Downloading dash_core_components-2.0.0-py3-none-any.whl.metadata
(2.9 kB)
Collecting dash-table==5.0.0 (from dash>=2.9.0->plotly-
resampler>=0.8.3.1->pycaret)
  Downloading dash_table-5.0.0-py3-none-any.whl.metadata (2.4 kB)
Requirement already satisfied: typing-extensions>=4.1.1 in
/usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-
resampler>=0.8.3.1->pycaret) (4.12.2)
Collecting retrying (from dash>=2.9.0->plotly-resampler>=0.8.3.1-
>pycaret)
  Downloading retrying-1.3.4-py3-none-any.whl.metadata (6.9 kB)
Requirement already satisfied: nest-asyncio in
/usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-
resampler>=0.8.3.1->pycaret) (1.6.0)
Requirement already satisfied: jupyter-client in
/usr/local/lib/python3.10/dist-packages (from ipykernel>=4.5.1-
>ipywidgets>=7.6.5->pycaret) (6.1.12)
Requirement already satisfied: tornado>=4.2 in
/usr/local/lib/python3.10/dist-packages (from ipykernel>=4.5.1-
>ipywidgets>=7.6.5->pycaret) (6.3.3)
Requirement already satisfied: parso<0.9.0,>=0.8.4 in
/usr/local/lib/python3.10/dist-packages (from jedi>=0.16-
>ipython>=5.5.0->pycaret) (0.8.4)
Requirement already satisfied: attrs>=22.2.0 in
/usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6-
>nbformat>=4.2.0->pycaret) (24.2.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
/usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6-
>nbformat>=4.2.0->pycaret) (2024.10.1)
Requirement already satisfied: referencing>=0.28.4 in
/usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6-
>nbformat>=4.2.0->pycaret) (0.35.1)
Requirement already satisfied: rpds-py>=0.7.1 in
/usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6-
>nbformat>=4.2.0->pycaret) (0.21.0)
Requirement already satisfied: platformdirs>=2.5 in
/usr/local/lib/python3.10/dist-packages (from jupyter-core!
=5.0.*,>=4.12->nbformat>=4.2.0->pycaret) (4.3.6)
Requirement already satisfied: ptyprocess>=0.5 in
/usr/local/lib/python3.10/dist-packages (from pexpect>4.3-
>ipython>=5.5.0->pycaret) (0.7.0)
Requirement already satisfied: wcwidth in
/usr/local/lib/python3.10/dist-packages (from prompt-toolkit!=3.0.0,!
=3.0.1,<3.1.0,>=2.0.0->ipython>=5.5.0->pycaret) (0.2.13)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7-
```

```
>matplotlib<3.8.0->pycaret) (1.16.0)
Requirement already satisfied: notebook>=4.4.1 in
/usr/local/lib/python3.10/dist-packages (from
widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (6.5.5)
Requirement already satisfied: itsdangerous>=2.1.2 in
/usr/local/lib/python3.10/dist-packages (from Flask<3.1,>=1.0.4-
>dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (2.2.0)
Requirement already satisfied: click>=8.1.3 in
/usr/local/lib/python3.10/dist-packages (from Flask<3.1,>=1.0.4-
>dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (8.1.7)
Requirement already satisfied: blinker>=1.6.2 in
/usr/local/lib/python3.10/dist-packages (from Flask<3.1,>=1.0.4-
>dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (1.9.0)
Requirement already satisfied: pyzmq<25,>=17 in
/usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1-
>widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (24.0.1)
Requirement already satisfied: argon2-cffi in
/usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1-
>widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (23.1.0)
Requirement already satisfied: nbconvert>=5 in
/usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1-
>widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (7.16.4)
Requirement already satisfied: Send2Trash>=1.8.0 in
/usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1-
>widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (1.8.3)
Requirement already satisfied: terminado>=0.8.3 in
/usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1-
>widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (0.18.1)
Requirement already satisfied: prometheus-client in
/usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1-
>widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (0.21.0)
Requirement already satisfied: nbclassic>=0.4.7 in
/usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1-
>widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (1.1.0)
Requirement already satisfied: notebook-shim>=0.2.3 in
/usr/local/lib/python3.10/dist-packages (from nbclassic>=0.4.7-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (0.2.4)
Requirement already satisfied: beautifulsoup4 in
/usr/local/lib/python3.10/dist-packages (from nbconvert>=5-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (4.12.3)
Requirement already satisfied: bleach!=5.0.0 in
/usr/local/lib/python3.10/dist-packages (from nbconvert>=5-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (6.2.0)
Requirement already satisfied: defusedxml in
/usr/local/lib/python3.10/dist-packages (from nbconvert>=5-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
```

```
>pycaret) (0.7.1)
Requirement already satisfied: jupyterlab-pygments in
/usr/local/lib/python3.10/dist-packages (from nbconvert>=5-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (0.3.0)
Requirement already satisfied: mistune<4,>=2.0.3 in
/usr/local/lib/python3.10/dist-packages (from nbconvert>=5-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (3.0.2)
Requirement already satisfied: nbclient>=0.5.0 in
/usr/local/lib/python3.10/dist-packages (from nbconvert>=5-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (0.10.0)
Requirement already satisfied: pandocfilters>=1.4.1 in
/usr/local/lib/python3.10/dist-packages (from nbconvert>=5-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (1.5.1)
Requirement already satisfied: tinycss2 in
/usr/local/lib/python3.10/dist-packages (from nbconvert>=5-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (1.4.0)
Requirement already satisfied: argon2-cffi-bindings in
/usr/local/lib/python3.10/dist-packages (from argon2-cffi-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (21.2.0)
Requirement already satisfied: webencodings in
/usr/local/lib/python3.10/dist-packages (from bleach!=5.0.0-
>nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0-
>ipywidgets>=7.6.5->pycaret) (0.5.1)
Requirement already satisfied: jupyter-server<3,>=1.8 in
/usr/local/lib/python3.10/dist-packages (from notebook-shim>=0.2.3-
>nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbextension~=3.6.0-
>ipywidgets>=7.6.5->pycaret) (1.24.0)
Requirement already satisfied: cffi>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from argon2-cffi-bindings-
>argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.6.0-
>ipywidgets>=7.6.5->pycaret) (1.17.1)
Requirement already satisfied: soupsieve>1.2 in
/usr/local/lib/python3.10/dist-packages (from beautifulsoup4-
>nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0-
>ipywidgets>=7.6.5->pycaret) (2.6)
Requirement already satisfied: pycparser in
/usr/local/lib/python3.10/dist-packages (from cffi>=1.0.1->argon2-
cffi-bindings->argon2-cffi->notebook>=4.4.1-
>widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (2.22)
Requirement already satisfied: anyio<4,>=3.1.0 in
/usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8-
>notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1-
>widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (3.7.1)
```

```
Requirement already satisfied: websocket-client in
/usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8-
>notebook-shim>=0.2.3->nbclassic>=0.4.7->notebook>=4.4.1-
>widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycaret) (1.8.0)
Requirement already satisfied: sniffio>=1.1 in
/usr/local/lib/python3.10/dist-packages (from anyio<4,>=3.1.0-
>jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbclassic>=0.4.7-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (1.3.1)
Requirement already satisfied: exceptiongroup in
/usr/local/lib/python3.10/dist-packages (from anyio<4,>=3.1.0-
>jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbclassic>=0.4.7-
>notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>=7.6.5-
>pycaret) (1.2.2)
Downloading pycaret-3.3.2-py3-none-any.whl (486 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 486.1/486.1 kB 24.5 MB/s eta
0:00:00
draw-0.15-py3-none-any.whl (106 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 106.8/106.8 kB 11.3 MB/s eta
0:00:00
e-0.26.0-py3-none-any.whl (21.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 21.8/21.8 MB 94.6 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 82.0/82.0 kB 7.6 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 302.2/302.2 kB 28.7 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 3.8/3.8 MB 106.5 MB/s eta
0:00:00
atplotlib-3.7.5-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (11.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 11.6/11.6 MB 116.5 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (12.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 12.3/12.3 MB 114.7 MB/s eta
0:00:00
pler-0.10.0-py3-none-any.whl (80 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 80.7/80.7 kB 8.2 MB/s eta
0:00:00
darima-2.0.4-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_28_x86_64.whl
(2.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.1/2.1 MB 86.2 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (12.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 12.1/12.1 MB 106.3 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (36.4 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 36.4/36.4 MB 14.0 MB/s eta
```

```
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 44.0/44.0 kB 3.6 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 194.1/194.1 kB 20.0 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 7.8/7.8 MB 110.1 MB/s eta
0:00:00
ponents-2.0.0-py3-none-any.whl (3.8 kB)
Downloading dash_html_components-2.0.0-py3-none-any.whl (4.1 kB)
Downloading dash_table-5.0.0-py3-none-any.whl (3.9 kB)
Downloading jedi-0.19.2-py2.py3-none-any.whl (1.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.6/1.6 MB 77.9 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 130.1/130.1 kB 13.9 MB/s eta
0:00:00
ple-0.1.3-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
(2.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.1/2.1 MB 87.2 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 228.0/228.0 kB 23.7 MB/s eta
0:00:00
e=pyod-2.0.2-py3-none-any.whl size=198469
sha256=0ecfba96bf280f601c2b58bc6107ce82e0eabee9d7850d64f829ff53a8601cb
3
  Stored in directory:
/root/.cache/pip/wheels/77/c2/20/34d1f15b41b701ba69f42a32304825810d680
754d509f91391
Successfully built pyod
Installing collected packages: dash-table, dash-html-components, dash-
core-components, xxhash, wurlitzer, Werkzeug, tsdownsample, scipy,
scikit-base, schemdraw, retrying, joblib, jedi, deprecation,
choreographer, scikit-learn, pandas, matplotlib, kaleido, sktime,
scikit-plot, pyod, dash, pmdarima, plotly-resampler, category-
encoders, tbats, pycaret
  Attempting uninstall: Werkzeug
    Found existing installation: Werkzeug 3.1.3
    Uninstalling Werkzeug-3.1.3:
      Successfully uninstalled Werkzeug-3.1.3
  Attempting uninstall: scipy
    Found existing installation: scipy 1.13.1
    Uninstalling scipy-1.13.1:
      Successfully uninstalled scipy-1.13.1
  Attempting uninstall: joblib
    Found existing installation: joblib 1.4.2
    Uninstalling joblib-1.4.2:
      Successfully uninstalled joblib-1.4.2
  Attempting uninstall: scikit-learn
    Found existing installation: scikit-learn 1.5.2
```

```
    Uninstalling scikit-learn-1.5.2:
      Successfully uninstalled scikit-learn-1.5.2
  Attempting uninstall: pandas
    Found existing installation: pandas 2.2.2
    Uninstalling pandas-2.2.2:
      Successfully uninstalled pandas-2.2.2
  Attempting uninstall: matplotlib
    Found existing installation: matplotlib 3.8.0
    Uninstalling matplotlib-3.8.0:
      Successfully uninstalled matplotlib-3.8.0
ERROR: pip's dependency resolver does not currently take into account
all the packages that are installed. This behaviour is the source of
the following dependency conflicts.
google-colab 1.0.0 requires pandas==2.2.2, but you have pandas 2.1.4
which is incompatible.
mizani 0.13.0 requires pandas>=2.2.0, but you have pandas 2.1.4 which
is incompatible.
plotnine 0.14.1 requires matplotlib>=3.8.0, but you have matplotlib
3.7.5 which is incompatible.
plotnine 0.14.1 requires pandas>=2.2.0, but you have pandas 2.1.4
which is incompatible.
Successfully installed Werkzeug-3.0.6 category-encoders-2.6.4
choreographer-0.99.6 dash-2.18.2 dash-core-components-2.0.0 dash-html-
components-2.0.0 dash-table-5.0.0 deprecation-2.1.0 jedi-0.19.2
joblib-1.3.2 kaleido-0.4.1 matplotlib-3.7.5 pandas-2.1.4 plotly-
resampler-0.10.0 pmdarima-2.0.4 pycaret-3.3.2 pyod-2.0.2 retrying-
1.3.4 schemdraw-0.15 scikit-base-0.7.8 scikit-learn-1.4.2 scikit-plot-
0.3.7 scipy-1.11.4 sktime-0.26.0 tbats-1.1.3 tsdownsample-0.1.3
wurlitzer-3.1.1 xxhash-3.5.0
```


```
{"id":"08646319f5904c9f8a0e8b4517bd06fa","pip_warning":{"packages":
["joblib","matplotlib","mpl_toolkits","sklearn"]}}
```


```
df.head()
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 2200,\n  \"fields\":
[\n    {\n      \"column\": \"N\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 36,\n        \"min\": 0,\n
\"max\": 140,\n        \"num_unique_values\": 137,\n
\"samples\": [\n          106,\n          101,\n          88\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"P\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 32,\n        \"min\": 5,\n        \"max\": 145,\n
\"num_unique_values\": 117,\n        \"samples\": [\n          69,\n
37,\n          11\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"K\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 50,\n        \"min\": 5,\n        \"max\": 205,\n
\"num_unique_values\": 73,\n        \"samples\": [\n          42,\n
```

```
12,\n          15\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"temperature\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 5.063748599958843,\n        \"min\":
8.825674745,\n        \"max\": 43.67549305,\n
\"num_unique_values\": 2200,\n        \"samples\": [\n
29.49401389,\n          26.1793464,\n          43.36051537\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"humidity\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 22.263811589761115,\n        \"min\":
14.25803981,\n        \"max\": 99.98187601,\n
\"num_unique_values\": 2200,\n        \"samples\": [\n
94.72981338,\n          86.52258079,\n          93.35191636\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"ph\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 0.7739376880298721,\n        \"min\": 3.504752314,\n
\"max\": 9.93509073,\n        \"num_unique_values\": 2200,\n
\"samples\": [\n          6.185053234,\n          6.25933595,\n
6.941496806\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"rainfall\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 54.95838852487811,\n        \"min\":
20.21126747,\n        \"max\": 298.5601175,\n
\"num_unique_values\": 2200,\n        \"samples\": [\n
26.30820876,\n          49.43050977,\n          114.778071\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"label\",\n      \"properties\": {\n        \"dtype\": \"category\",\
n        \"num_unique_values\": 22,\n        \"samples\": [\n
\"rice\",\n          \"watermelon\",\n          \"lentil\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```


```
data = df.sample(frac=0.8, random_state=42).reset_index(drop=True) # Training data - 80%
data_unseen = df.drop(data.index).reset_index(drop=True) # Test data - 20%

print(data.shape, data_unseen.shape)
```

```
(1760, 8) (440, 8)
```

```
from pycaret.classification import *

classification_model = setup(data = data, target = 'label', session_id = 123)
```

```
<pandas.io.formats.style.Styler at 0x7e6fd9305750>
```

```
compare_models()
```

```
<IPython.core.display.HTML object>
```

```
<pandas.io.formats.style.Styler at 0x7e70737ba6b0>
```

```
{"model_id":"df788f1ced6844398ed33cfc5425b1bb","version_major":2,"vers
ion_minor":0}
```

```
<IPython.core.display.HTML object>
```

```
GaussianNB(priors=None, var_smoothing=1e-09)
```

```
rf = create_model('rf')
```

```
<IPython.core.display.HTML object>
```

```
<pandas.io.formats.style.Styler at 0x7e702e902530>
```

```
{"model_id":"7433c47bcb174e459d30df727f9218fd","version_major":2,"vers
ion_minor":0}
```

```
<IPython.core.display.HTML object>
```

```
# HyperParameters tuning
tuned_rf = tune_model(rf)
```

```
<IPython.core.display.HTML object>
```

```
<pandas.io.formats.style.Styler at 0x7e7039e1d5a0>
```

```
{"model_id":"407d15fd2db74f728ff7e7f19ad1535b","version_major":2,"vers
ion_minor":0}
```

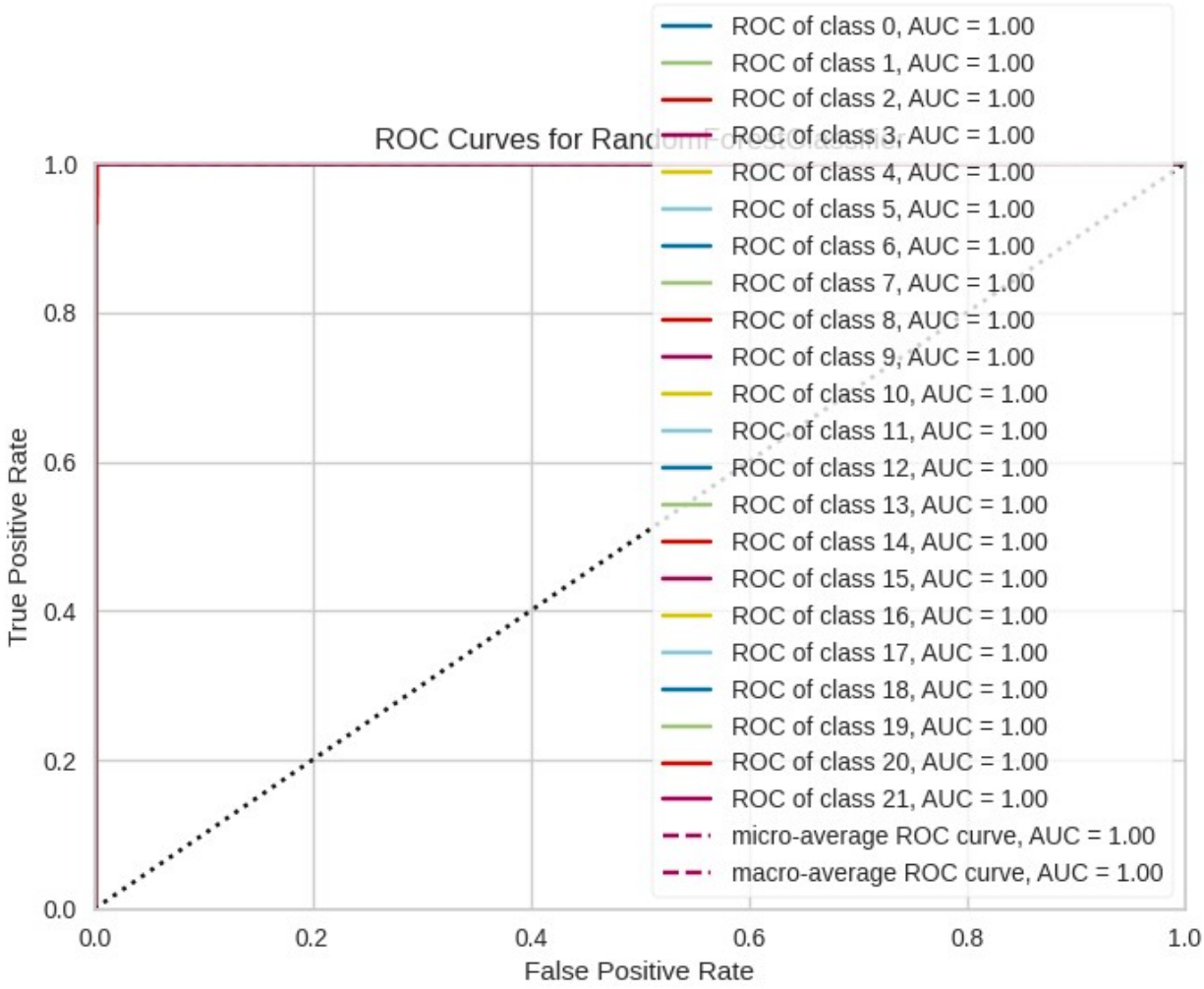
```
Fitting 10 folds for each of 10 candidates, totalling 100 fits
```

```
<IPython.core.display.HTML object>
```

```
Original model was better than the tuned model, hence it will be
returned. NOTE: The display metrics are for the tuned model (not the
original one).
```

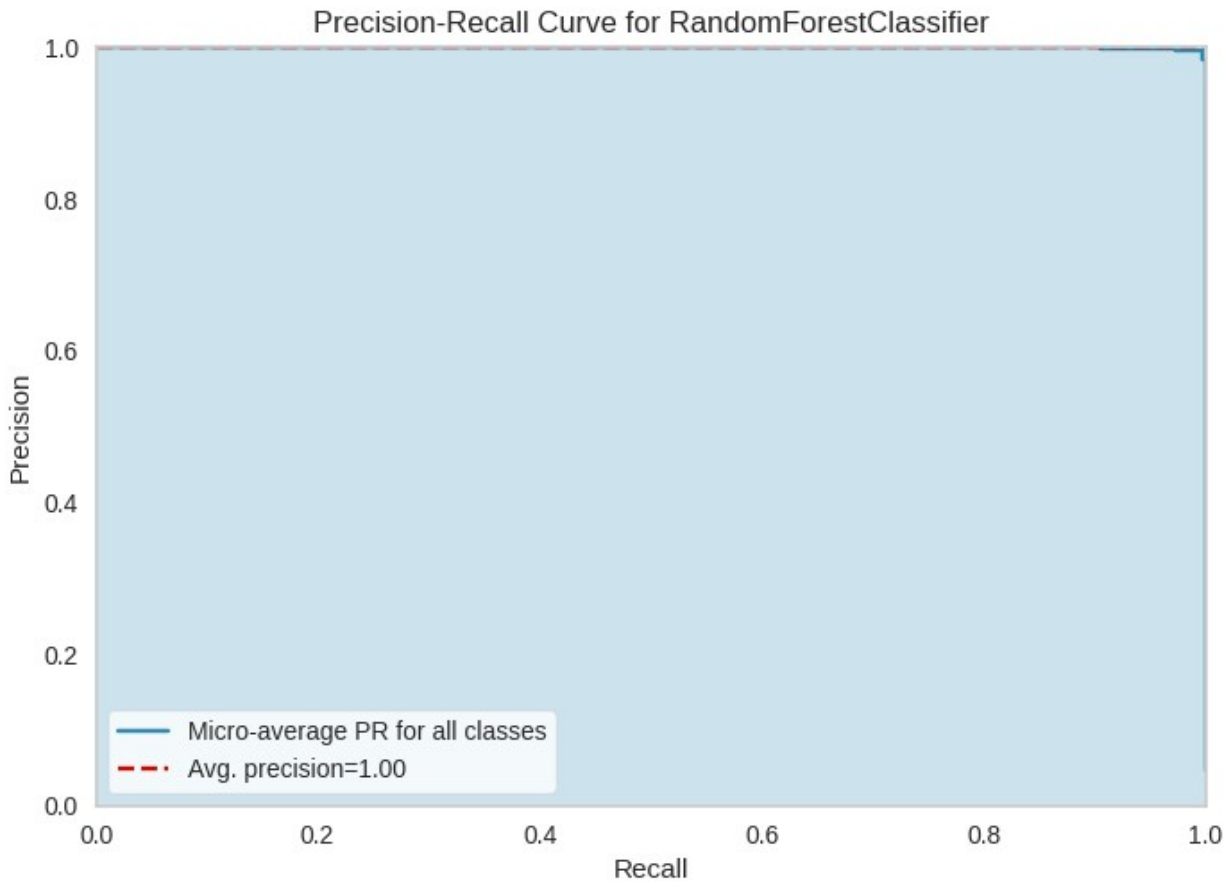
```
plot_model(tuned_rf, plot='auc')
```

```
<IPython.core.display.HTML object>
```



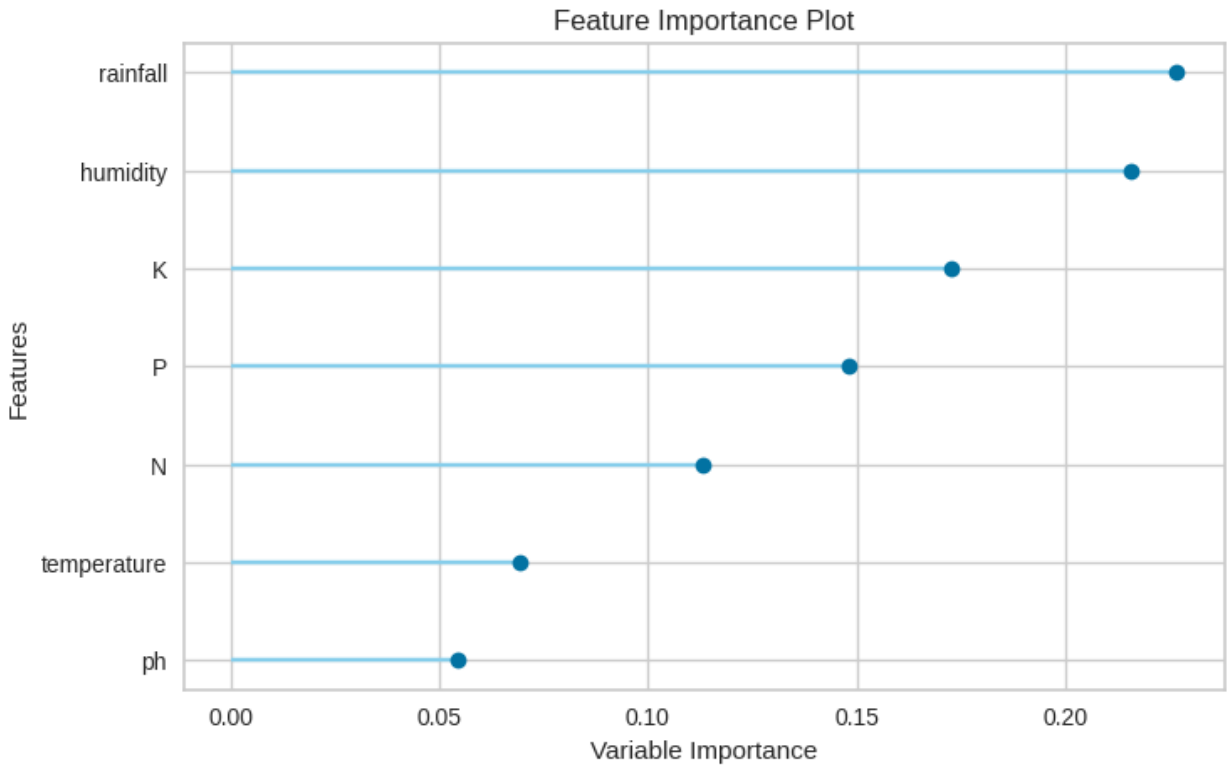

```
plot_model(tuned_rf, plot='pr')
```

```
<IPython.core.display.HTML object>
```



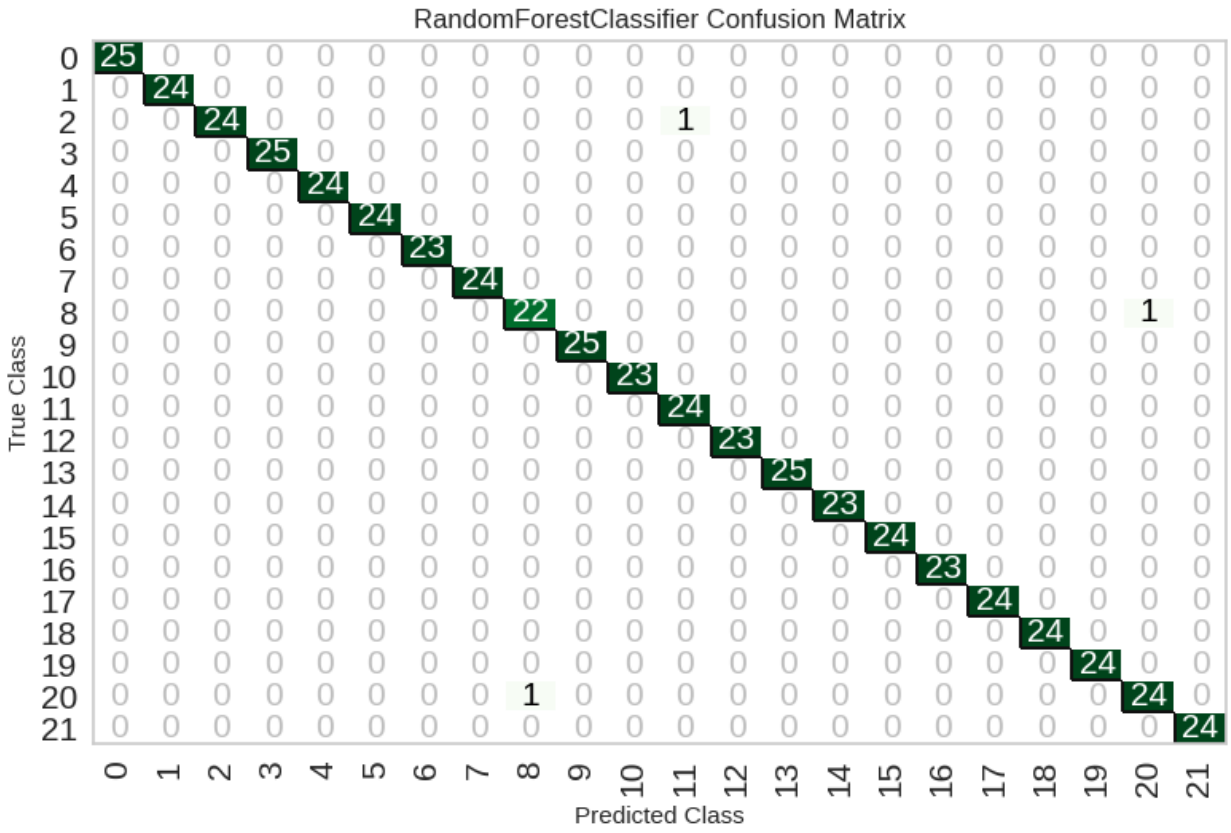

```
plot_model(tuned_rf, plot='feature')
```

```
<IPython.core.display.HTML object>
```




```
plot_model(tuned_rf, plot='confusion_matrix')
```

```
<IPython.core.display.HTML object>
```




```
evaluate_model(tuned_rf)
```


```
{"model_id":"f6b31157efe54c208a3d41286ec832ed","version_major":2,"vers
ion_minor":0}
```


```
predict_model(tuned_rf)
```

```
<pandas.io.formats.style.Styler at 0x7e6fd9bf0820>
```


```
{"summary":"{\n  \"name\": \"predict_model(tuned_rf)\",\n  \"rows\":
528,\n  \"fields\": [\n    {\n      \"column\": \"N\",\n
\"properties\": {\n        \"dtype\": \"int32\",\n
\"num_unique_values\": 116,\n        \"samples\": [\n          86,\n
12,\n          22\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"P\",\n      \"properties\": {\n        \"dtype\": \"int32\",\n
\"num_unique_values\": 106,\n        \"samples\": [\n          133,\n
50,\n          61\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"K\",\n      \"properties\": {\n        \"dtype\": \"int32\",\n
\"num_unique_values\": 73,\n        \"samples\": [\n          24,\n
9,\n          33\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"temperature\",\n      \"properties\": {\n        \"dtype\":
\"float32\",\n        \"num_unique_values\": 528,\n
```

```
\"samples\": [\n          29.607187271118164,\n
30.554725646972656,\n          35.538448333740234\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"humidity\",\n      \"properties\":
{\n        \"dtype\": \"float32\",\n        \"num_unique_values\":
528,\n        \"samples\": [\n          93.15642547607422,\n
90.90343475341797,\n          52.94641876220703\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"ph\",\n      \"properties\": {\n
\"dtype\": \"float32\",\n        \"num_unique_values\": 528,\n
\"samples\": [\n          6.573980331420898,\n
7.1892595291137695,\n          4.934964656829834\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"rainfall\",\n      \"properties\":
{\n        \"dtype\": \"float32\",\n        \"num_unique_values\":
528,\n        \"samples\": [\n          62.68710708618164,\n
106.07119750976562,\n          91.54560089111328\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"label\",\n      \"properties\": {\
n        \"dtype\": \"category\",\n        \"num_unique_values\": 22,\
n        \"samples\": [\n          \"maize\",\n          \"apple\",\n
\"rice\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"prediction_label\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 22,\n
\"samples\": [\n          \"maize\",\n          \"apple\",\n
\"rice\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"prediction_score\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0.0760333877941529,\n        \"min\":
0.49,\n        \"max\": 1.0,\n        \"num_unique_values\": 36,\n
\"samples\": [\n          0.73,\n          0.88,\n          0.78\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\n}","type":"dataframe"}
```


```
WARNING:root:Quickchart encountered unexpected dtypes in columns:
"(['label'],)"
```

```
<google.colab._quickchart_helpers.SectionTitle at 0x7e702e305630>
```

```
from matplotlib import pyplot as plt
_df_0['N'].plot(kind='hist', bins=20, title='N')
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_1['P'].plot(kind='hist', bins=20, title='P')
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_2['K'].plot(kind='hist', bins=20, title='K')
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_3['temperature'].plot(kind='hist', bins=20, title='temperature')
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
<google.colab._quickchart_helpers.SectionTitle at 0x7e701a1d1630>
```

```
from matplotlib import pyplot as plt
_df_4.plot(kind='scatter', x='N', y='P', s=32, alpha=.8)
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_5.plot(kind='scatter', x='P', y='K', s=32, alpha=.8)
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_6.plot(kind='scatter', x='K', y='temperature', s=32, alpha=.8)
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_7.plot(kind='scatter', x='temperature', y='humidity', s=32,
alpha=.8)
plt.gca().spines[['top', 'right',]].set_visible(False)
```

```
<google.colab._quickchart_helpers.SectionTitle at 0x7e702e304370>
```

```
from matplotlib import pyplot as plt
_df_8['N'].plot(kind='line', figsize=(8, 4), title='N')
plt.gca().spines[['top', 'right']].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_9['P'].plot(kind='line', figsize=(8, 4), title='P')
plt.gca().spines[['top', 'right']].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_10['K'].plot(kind='line', figsize=(8, 4), title='K')
plt.gca().spines[['top', 'right']].set_visible(False)
```

```
from matplotlib import pyplot as plt
_df_11['temperature'].plot(kind='line', figsize=(8, 4),
title='temperature')
plt.gca().spines[['top', 'right']].set_visible(False)
```

```
unseen_predict = predict_model(tuned_rf, data=data_unseen)
unseen_predict.head()
```

```
<pandas.io.formats.style.Styler at 0x7e703a33d9f0>
```

```
{"summary":"{\n  \"name\": \"unseen_predict\",\n  \"rows\": 440,\n
\"fields\": [\n    {\n      \"column\": \"N\",\n      \"properties\":
{\n        \"dtype\": \"int32\",\n        \"num_unique_values\": 130,\
n        \"samples\": [\n          6,\n          9,\n          53\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"P\",\n      \"properties\": {\n
\"dtype\": \"int8\",\n        \"num_unique_values\": 65,\n
\"samples\": [\n          56,\n          54,\n          55\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"K\",\n      \"properties\": {\n        \"dtype\": \"int8\",\n
\"num_unique_values\": 39,\n        \"samples\": [\n          36,\n
37,\n          50\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"temperature\",\n      \"properties\": {\n        \"dtype\":
\"float32\",\n        \"num_unique_values\": 440,\n
\"samples\": [\n          24.394594192504883,\n
26.312705993652344,\n          23.078954696655273\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"humidity\",\n      \"properties\":
{\n        \"dtype\": \"float32\",\n        \"num_unique_values\":
440,\n        \"samples\": [\n          89.89106750488281,\n
98.6204833984375,\n          63.65861511230469\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"ph\",\n      \"properties\": {\n
\"dtype\": \"float32\",\n        \"num_unique_values\": 440,\n
\"samples\": [\n          6.551130294799805,\n
5.804965019226074,\n          7.1848015785217285\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"rainfall\",\n      \"properties\":
{\n        \"dtype\": \"float32\",\n        \"num_unique_values\":
440,\n        \"samples\": [\n          197.12200927734375,\n
208.11813354492188,\n          129.8765411376953\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"label\",\n      \"properties\": {\
n        \"dtype\": \"category\",\n        \"num_unique_values\": 5,\n
\"samples\": [\n          \"coconut\",\n          \"coffee\",\n
\"cotton\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"prediction_label\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 6,\n        \"samples\":
[\n          \"papaya\",\n          \"coconut\",\n
\"coffee\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"prediction_score\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0.07710880131083693,\n        \"min\":
0.51,\n        \"max\": 1.0,\n        \"num_unique_values\": 37,\n
\"samples\": [\n          0.92,\n          0.64,\n          0.97\n
```

```
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\n}","type":"dataframe","variable_name":"unseen_predict"}
```

```
save_model(tuned_rf, 'tuned_rf_model')
```

```
Transformation Pipeline and Model Successfully Saved
```

```
(Pipeline(memory=Memory(location=None),
          steps=[('label_encoding',
                  TransformerWrapperWithInverse(exclude=None,
include=None,

transformer=LabelEncoder())),
                 ('numerical_imputer',
                  TransformerWrapper(exclude=None,
                                     include=['N', 'P', 'K',
'temperature',
                                              'humidity', 'ph',
'rainfall'],

transformer=SimpleImputer(add_indicator=False,

copy=True,

fill_value=None,

keep_empty_fea...
                  RandomForestClassifier(bootstrap=True,
ccp_alpha=0.0,
                                         class_weight=None,
criterion='gini',
                                         max_depth=None,
max_features='sqrt',
                                         max_leaf_nodes=None,
max_samples=None,
                                         min_impurity_decrease=0.0,
                                         min_samples_leaf=1,
min_samples_split=2,
                                         min_weight_fraction_leaf=0.0,
                                         monotonic_cst=None,
n_estimators=100,
                                         n_jobs=-1, oob_score=False,
                                         random_state=123, verbose=0,
                                         warm_start=False))],
          verbose=False),
 'tuned_rf_model.pkl')
```