{

"cells": [

{

"cell\_type": "markdown",

"metadata": {

"id": "YF4zpAwQZ434"

},

"source": [

"\n",

"\*\*1. Importing Necessary Libraries\*\*"

]

},

{

"cell\_type": "code",

"execution\_count": 28,

"metadata": {

"id": "9IErcoiwSEic"

},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"Requirement already satisfied: openpyxl in c:\\users\\vicky reddy\\anaconda3\\lib\\site-packages (3.0.9)\n",

"Requirement already satisfied: et-xmlfile in c:\\users\\vicky reddy\\anaconda3\\lib\\site-packages (from openpyxl) (1.1.0)\n"

]

}

],

"source": [

"import pandas as pd\n",

"import numpy as np\n",

"import matplotlib as mlt\n",

"import sklearn\n",

"import scipy\n",

"import seaborn as sb\n",

"import missingno as msno\n",

"import warnings\n",

"warnings.filterwarnings('ignore')\n",

"!pip3 install openpyxl"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "uSBDwc5eaQmJ"

},

"source": [

"\*\*2. Importing the Dataset\*\*"

]

},

{

"cell\_type": "code",

"execution\_count": 29,

"metadata": {

"id": "8Y-31FFCSIoU"

},

"outputs": [],

"source": [

"data=pd.read\_csv(\"weather.csv\")"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "FPuqDPh9aYzx"

},

"source": [

"\*\*3. Analyse the data\*\*"

]

},

{

"cell\_type": "code",

"execution\_count": 30,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 364

},

"id": "QvMdJuGMZlGL",

"outputId": "a3316a5b-21f3-416a-c857-825b9e7381ec"

},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>MinTemp</th>\n",

" <th>MaxTemp</th>\n",

" <th>Rainfall</th>\n",

" <th>Evaporation</th>\n",

" <th>Sunshine</th>\n",

" <th>WindGustSpeed</th>\n",

" <th>WindSpeed9am</th>\n",

" <th>WindSpeed3pm</th>\n",

" <th>Humidity9am</th>\n",

" <th>Humidity3pm</th>\n",

" <th>Pressure9am</th>\n",

" <th>Pressure3pm</th>\n",

" <th>Cloud9am</th>\n",

" <th>Cloud3pm</th>\n",

" <th>Temp9am</th>\n",

" <th>Temp3pm</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>count</th>\n",

" <td>143975.000000</td>\n",

" <td>144199.000000</td>\n",

" <td>142199.000000</td>\n",

" <td>82670.000000</td>\n",

" <td>75625.000000</td>\n",

" <td>135197.000000</td>\n",

" <td>143693.000000</td>\n",

" <td>142398.000000</td>\n",

" <td>142806.000000</td>\n",

" <td>140953.000000</td>\n",

" <td>130395.00000</td>\n",

" <td>130432.000000</td>\n",

" <td>89572.000000</td>\n",

" <td>86102.000000</td>\n",

" <td>143693.000000</td>\n",

" <td>141851.00000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>mean</th>\n",

" <td>12.194034</td>\n",

" <td>23.221348</td>\n",

" <td>2.360918</td>\n",

" <td>5.468232</td>\n",

" <td>7.611178</td>\n",

" <td>40.035230</td>\n",

" <td>14.043426</td>\n",

" <td>18.662657</td>\n",

" <td>68.880831</td>\n",

" <td>51.539116</td>\n",

" <td>1017.64994</td>\n",

" <td>1015.255889</td>\n",

" <td>4.447461</td>\n",

" <td>4.509930</td>\n",

" <td>16.990631</td>\n",

" <td>21.68339</td>\n",

" </tr>\n",

" <tr>\n",

" <th>std</th>\n",

" <td>6.398495</td>\n",

" <td>7.119049</td>\n",

" <td>8.478060</td>\n",

" <td>4.193704</td>\n",

" <td>3.785483</td>\n",

" <td>13.607062</td>\n",

" <td>8.915375</td>\n",

" <td>8.809800</td>\n",

" <td>19.029164</td>\n",

" <td>20.795902</td>\n",

" <td>7.10653</td>\n",

" <td>7.037414</td>\n",

" <td>2.887159</td>\n",

" <td>2.720357</td>\n",

" <td>6.488753</td>\n",

" <td>6.93665</td>\n",

" </tr>\n",

" <tr>\n",

" <th>min</th>\n",

" <td>-8.500000</td>\n",

" <td>-4.800000</td>\n",

" <td>0.000000</td>\n",

" <td>0.000000</td>\n",

" <td>0.000000</td>\n",

" <td>6.000000</td>\n",

" <td>0.000000</td>\n",

" <td>0.000000</td>\n",

" <td>0.000000</td>\n",

" <td>0.000000</td>\n",

" <td>980.50000</td>\n",

" <td>977.100000</td>\n",

" <td>0.000000</td>\n",

" <td>0.000000</td>\n",

" <td>-7.200000</td>\n",

" <td>-5.40000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>25%</th>\n",

" <td>7.600000</td>\n",

" <td>17.900000</td>\n",

" <td>0.000000</td>\n",

" <td>2.600000</td>\n",

" <td>4.800000</td>\n",

" <td>31.000000</td>\n",

" <td>7.000000</td>\n",

" <td>13.000000</td>\n",

" <td>57.000000</td>\n",

" <td>37.000000</td>\n",

" <td>1012.90000</td>\n",

" <td>1010.400000</td>\n",

" <td>1.000000</td>\n",

" <td>2.000000</td>\n",

" <td>12.300000</td>\n",

" <td>16.60000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>50%</th>\n",

" <td>12.000000</td>\n",

" <td>22.600000</td>\n",

" <td>0.000000</td>\n",

" <td>4.800000</td>\n",

" <td>8.400000</td>\n",

" <td>39.000000</td>\n",

" <td>13.000000</td>\n",

" <td>19.000000</td>\n",

" <td>70.000000</td>\n",

" <td>52.000000</td>\n",

" <td>1017.60000</td>\n",

" <td>1015.200000</td>\n",

" <td>5.000000</td>\n",

" <td>5.000000</td>\n",

" <td>16.700000</td>\n",

" <td>21.10000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>75%</th>\n",

" <td>16.900000</td>\n",

" <td>28.200000</td>\n",

" <td>0.800000</td>\n",

" <td>7.400000</td>\n",

" <td>10.600000</td>\n",

" <td>48.000000</td>\n",

" <td>19.000000</td>\n",

" <td>24.000000</td>\n",

" <td>83.000000</td>\n",

" <td>66.000000</td>\n",

" <td>1022.40000</td>\n",

" <td>1020.000000</td>\n",

" <td>7.000000</td>\n",

" <td>7.000000</td>\n",

" <td>21.600000</td>\n",

" <td>26.40000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>max</th>\n",

" <td>33.900000</td>\n",

" <td>48.100000</td>\n",

" <td>371.000000</td>\n",

" <td>145.000000</td>\n",

" <td>14.500000</td>\n",

" <td>135.000000</td>\n",

" <td>130.000000</td>\n",

" <td>87.000000</td>\n",

" <td>100.000000</td>\n",

" <td>100.000000</td>\n",

" <td>1041.00000</td>\n",

" <td>1039.600000</td>\n",

" <td>9.000000</td>\n",

" <td>9.000000</td>\n",

" <td>40.200000</td>\n",

" <td>46.70000</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" MinTemp MaxTemp Rainfall Evaporation \\\n",

"count 143975.000000 144199.000000 142199.000000 82670.000000 \n",

"mean 12.194034 23.221348 2.360918 5.468232 \n",

"std 6.398495 7.119049 8.478060 4.193704 \n",

"min -8.500000 -4.800000 0.000000 0.000000 \n",

"25% 7.600000 17.900000 0.000000 2.600000 \n",

"50% 12.000000 22.600000 0.000000 4.800000 \n",

"75% 16.900000 28.200000 0.800000 7.400000 \n",

"max 33.900000 48.100000 371.000000 145.000000 \n",

"\n",

" Sunshine WindGustSpeed WindSpeed9am WindSpeed3pm \\\n",

"count 75625.000000 135197.000000 143693.000000 142398.000000 \n",

"mean 7.611178 40.035230 14.043426 18.662657 \n",

"std 3.785483 13.607062 8.915375 8.809800 \n",

"min 0.000000 6.000000 0.000000 0.000000 \n",

"25% 4.800000 31.000000 7.000000 13.000000 \n",

"50% 8.400000 39.000000 13.000000 19.000000 \n",

"75% 10.600000 48.000000 19.000000 24.000000 \n",

"max 14.500000 135.000000 130.000000 87.000000 \n",

"\n",

" Humidity9am Humidity3pm Pressure9am Pressure3pm \\\n",

"count 142806.000000 140953.000000 130395.00000 130432.000000 \n",

"mean 68.880831 51.539116 1017.64994 1015.255889 \n",

"std 19.029164 20.795902 7.10653 7.037414 \n",

"min 0.000000 0.000000 980.50000 977.100000 \n",

"25% 57.000000 37.000000 1012.90000 1010.400000 \n",

"50% 70.000000 52.000000 1017.60000 1015.200000 \n",

"75% 83.000000 66.000000 1022.40000 1020.000000 \n",

"max 100.000000 100.000000 1041.00000 1039.600000 \n",

"\n",

" Cloud9am Cloud3pm Temp9am Temp3pm \n",

"count 89572.000000 86102.000000 143693.000000 141851.00000 \n",

"mean 4.447461 4.509930 16.990631 21.68339 \n",

"std 2.887159 2.720357 6.488753 6.93665 \n",

"min 0.000000 0.000000 -7.200000 -5.40000 \n",

"25% 1.000000 2.000000 12.300000 16.60000 \n",

"50% 5.000000 5.000000 16.700000 21.10000 \n",

"75% 7.000000 7.000000 21.600000 26.40000 \n",

"max 9.000000 9.000000 40.200000 46.70000 "

]

},

"execution\_count": 30,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"data.describe()"

]

},

{

"cell\_type": "code",

"execution\_count": 31,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "Zawkx9NhZykm",

"outputId": "e97dbe7e-5e88-4ce8-bdf3-d51c93f0943a"

},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"<class 'pandas.core.frame.DataFrame'>\n",

"RangeIndex: 145460 entries, 0 to 145459\n",

"Data columns (total 23 columns):\n",

" # Column Non-Null Count Dtype \n",

"--- ------ -------------- ----- \n",

" 0 Date 145460 non-null object \n",

" 1 Location 145460 non-null object \n",

" 2 MinTemp 143975 non-null float64\n",

" 3 MaxTemp 144199 non-null float64\n",

" 4 Rainfall 142199 non-null float64\n",

" 5 Evaporation 82670 non-null float64\n",

" 6 Sunshine 75625 non-null float64\n",

" 7 WindGustDir 135134 non-null object \n",

" 8 WindGustSpeed 135197 non-null float64\n",

" 9 WindDir9am 134894 non-null object \n",

" 10 WindDir3pm 141232 non-null object \n",

" 11 WindSpeed9am 143693 non-null float64\n",

" 12 WindSpeed3pm 142398 non-null float64\n",

" 13 Humidity9am 142806 non-null float64\n",

" 14 Humidity3pm 140953 non-null float64\n",

" 15 Pressure9am 130395 non-null float64\n",

" 16 Pressure3pm 130432 non-null float64\n",

" 17 Cloud9am 89572 non-null float64\n",

" 18 Cloud3pm 86102 non-null float64\n",

" 19 Temp9am 143693 non-null float64\n",

" 20 Temp3pm 141851 non-null float64\n",

" 21 RainToday 142199 non-null object \n",

" 22 RainTomorrow 142193 non-null object \n",

"dtypes: float64(16), object(7)\n",

"memory usage: 25.5+ MB\n"

]

}

],

"source": [

"data.info()"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "casodfQFagxa"

},

"source": [

"\*\*4. Handling Missing Values\*\*"

]

},

{

"cell\_type": "code",

"execution\_count": 32,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "q520qItFaftk",

"outputId": "59017e11-8faa-42ae-8794-954652473dec"

},

"outputs": [

{

"data": {

"text/plain": [

"Date 0\n",

"Location 0\n",

"MinTemp 1485\n",

"MaxTemp 1261\n",

"Rainfall 3261\n",

"Evaporation 62790\n",

"Sunshine 69835\n",

"WindGustDir 10326\n",

"WindGustSpeed 10263\n",

"WindDir9am 10566\n",

"WindDir3pm 4228\n",

"WindSpeed9am 1767\n",

"WindSpeed3pm 3062\n",

"Humidity9am 2654\n",

"Humidity3pm 4507\n",

"Pressure9am 15065\n",

"Pressure3pm 15028\n",

"Cloud9am 55888\n",

"Cloud3pm 59358\n",

"Temp9am 1767\n",

"Temp3pm 3609\n",

"RainToday 3261\n",

"RainTomorrow 3267\n",

"dtype: int64"

]

},

"execution\_count": 32,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"data.isnull().sum()"

]

},

{

"cell\_type": "code",

"execution\_count": 33,

"metadata": {

"id": "M20Lu7jpbVty"

},

"outputs": [],

"source": [

"import missingno as msno"

]

},

{

"cell\_type": "code",

"execution\_count": 34,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 421

},

"id": "wnCHAMUPbaZQ",

"outputId": "ca75d023-7707-42eb-cb80-75e13be6adfb"

},

"outputs": [

{

"data": {

"text/plain": [

"<AxesSubplot:>"

]

},

"execution\_count": 34,

"metadata": {},

"output\_type": "execute\_result"

},

{

"data": {

"image/png": "\n",

"text/plain": [

"<Figure size 1800x720 with 2 Axes>"

]

},

"metadata": {

"needs\_background": "light"

},

"output\_type": "display\_data"

}

],

"source": [

"msno.matrix(data,color=(0.55,0.255,0.255),fontsize=16)"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "0yXYHW-tboZA"

},

"source": [

"\n",

"\n",

"> Imputing Data\n",

"\n"

]

},

{

"cell\_type": "code",

"execution\_count": 35,

"metadata": {

"id": "3azEm4-eTsu9"

},

"outputs": [],

"source": [

"data\_c=data[[\"RainToday\",\"WindGustDir\",\"WindDir9am\",\"WindDir3pm\"]]"

]

},

{

"cell\_type": "code",

"execution\_count": 36,

"metadata": {

"id": "VVs0FP55Txlz"

},

"outputs": [],

"source": [

"data.drop(columns=[\"Evaporation\",\"Sunshine\",\"Cloud9am\",\"Cloud3pm\"],axis=1,inplace=True)\n",

"data.drop(columns=[\"RainToday\",\"WindGustDir\",\"WindDir9am\",\"WindDir3pm\"],axis=1,inplace=True)"

]

},

{

"cell\_type": "code",

"execution\_count": 37,

"metadata": {

"id": "s6V9x87fTzcB"

},

"outputs": [],

"source": [

"data['MinTemp'].fillna(data['MinTemp'].mean(),inplace=True)\n",

"data['MaxTemp'].fillna(data['MaxTemp'].mean(),inplace=True)\n",

"data['Rainfall'].fillna(data['Rainfall'].mean(),inplace=True)\n",

"data['WindGustSpeed'].fillna(data['WindGustSpeed'].mean(),inplace=True)\n",

"data['WindSpeed9am'].fillna(data['WindSpeed9am'].mean(),inplace=True)\n",

"data['WindSpeed3pm'].fillna(data['WindSpeed3pm'].mean(),inplace=True)\n",

"data['Humidity3pm'].fillna(data['Humidity3pm'].mean(),inplace=True)\n",

"data['Humidity9am'].fillna(data['Humidity9am'].mean(),inplace=True)\n",

"data['Temp9am'].fillna(data['Temp9am'].mean(),inplace=True)\n",

"data['Temp3pm'].fillna(data['Temp3pm'].mean(),inplace=True)"

]

},

{

"cell\_type": "code",

"execution\_count": 38,

"metadata": {

"id": "n14FJpFiT1Ve"

},

"outputs": [],

"source": [

"c\_names=data\_c.columns"

]

},

{

"cell\_type": "code",

"execution\_count": 39,

"metadata": {

"id": "PXpcAmxQT43j"

},

"outputs": [],

"source": [

"from sklearn.impute import SimpleImputer"

]

},

{

"cell\_type": "code",

"execution\_count": 40,

"metadata": {

"id": "fJ\_RgqwOT6sL"

},

"outputs": [],

"source": [

"imp\_mode=SimpleImputer(missing\_values=np.nan,strategy=\"most\_frequent\")"

]

},

{

"cell\_type": "code",

"execution\_count": 41,

"metadata": {

"id": "KH0KO2piGFuD"

},

"outputs": [],

"source": [

"data\_c=imp\_mode.fit\_transform(data\_c)"

]

},

{

"cell\_type": "code",

"execution\_count": 42,

"metadata": {

"id": "tfZbioMCUQcu"

},

"outputs": [],

"source": [

"data\_c=pd.DataFrame(data\_c,columns=c\_names)"

]

},

{

"cell\_type": "code",

"execution\_count": 43,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 206

},

"id": "hx7rz8xpGuDv",

"outputId": "792c5995-6adf-4b6e-e6b1-2171d2aaf634"

},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>RainToday</th>\n",

" <th>WindGustDir</th>\n",

" <th>WindDir9am</th>\n",

" <th>WindDir3pm</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>145455</th>\n",

" <td>No</td>\n",

" <td>E</td>\n",

" <td>SE</td>\n",

" <td>ENE</td>\n",

" </tr>\n",

" <tr>\n",

" <th>145456</th>\n",

" <td>No</td>\n",

" <td>NNW</td>\n",

" <td>SE</td>\n",

" <td>N</td>\n",

" </tr>\n",

" <tr>\n",

" <th>145457</th>\n",

" <td>No</td>\n",

" <td>N</td>\n",

" <td>SE</td>\n",

" <td>WNW</td>\n",

" </tr>\n",

" <tr>\n",

" <th>145458</th>\n",

" <td>No</td>\n",

" <td>SE</td>\n",

" <td>SSE</td>\n",

" <td>N</td>\n",

" </tr>\n",

" <tr>\n",

" <th>145459</th>\n",

" <td>No</td>\n",

" <td>W</td>\n",

" <td>ESE</td>\n",

" <td>ESE</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" RainToday WindGustDir WindDir9am WindDir3pm\n",

"145455 No E SE ENE\n",

"145456 No NNW SE N\n",

"145457 No N SE WNW\n",

"145458 No SE SSE N\n",

"145459 No W ESE ESE"

]

},

"execution\_count": 43,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"data\_c.tail()"

]

},

{

"cell\_type": "code",

"execution\_count": 44,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 357

},

"id": "5byvcS40G6u2",

"outputId": "6b3a1a48-cf50-4ae9-97ae-e49039739c44"

},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>Date</th>\n",

" <th>Location</th>\n",

" <th>MinTemp</th>\n",

" <th>MaxTemp</th>\n",

" <th>Rainfall</th>\n",

" <th>WindGustSpeed</th>\n",

" <th>WindSpeed9am</th>\n",

" <th>WindSpeed3pm</th>\n",

" <th>Humidity9am</th>\n",

" <th>Humidity3pm</th>\n",

" <th>Pressure9am</th>\n",

" <th>Pressure3pm</th>\n",

" <th>Temp9am</th>\n",

" <th>Temp3pm</th>\n",

" <th>RainTomorrow</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>2008-12-01</td>\n",

" <td>Albury</td>\n",

" <td>13.4</td>\n",

" <td>22.9</td>\n",

" <td>0.6</td>\n",

" <td>44.0</td>\n",

" <td>20.0</td>\n",

" <td>24.0</td>\n",

" <td>71.0</td>\n",

" <td>22.0</td>\n",

" <td>1007.7</td>\n",

" <td>1007.1</td>\n",

" <td>16.9</td>\n",

" <td>21.8</td>\n",

" <td>No</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>2008-12-02</td>\n",

" <td>Albury</td>\n",

" <td>7.4</td>\n",

" <td>25.1</td>\n",

" <td>0.0</td>\n",

" <td>44.0</td>\n",

" <td>4.0</td>\n",

" <td>22.0</td>\n",

" <td>44.0</td>\n",

" <td>25.0</td>\n",

" <td>1010.6</td>\n",

" <td>1007.8</td>\n",

" <td>17.2</td>\n",

" <td>24.3</td>\n",

" <td>No</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>2008-12-03</td>\n",

" <td>Albury</td>\n",

" <td>12.9</td>\n",

" <td>25.7</td>\n",

" <td>0.0</td>\n",

" <td>46.0</td>\n",

" <td>19.0</td>\n",

" <td>26.0</td>\n",

" <td>38.0</td>\n",

" <td>30.0</td>\n",

" <td>1007.6</td>\n",

" <td>1008.7</td>\n",

" <td>21.0</td>\n",

" <td>23.2</td>\n",

" <td>No</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>2008-12-04</td>\n",

" <td>Albury</td>\n",

" <td>9.2</td>\n",

" <td>28.0</td>\n",

" <td>0.0</td>\n",

" <td>24.0</td>\n",

" <td>11.0</td>\n",

" <td>9.0</td>\n",

" <td>45.0</td>\n",

" <td>16.0</td>\n",

" <td>1017.6</td>\n",

" <td>1012.8</td>\n",

" <td>18.1</td>\n",

" <td>26.5</td>\n",

" <td>No</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>2008-12-05</td>\n",

" <td>Albury</td>\n",

" <td>17.5</td>\n",

" <td>32.3</td>\n",

" <td>1.0</td>\n",

" <td>41.0</td>\n",

" <td>7.0</td>\n",

" <td>20.0</td>\n",

" <td>82.0</td>\n",

" <td>33.0</td>\n",

" <td>1010.8</td>\n",

" <td>1006.0</td>\n",

" <td>17.8</td>\n",

" <td>29.7</td>\n",

" <td>No</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" Date Location MinTemp MaxTemp Rainfall WindGustSpeed \\\n",

"0 2008-12-01 Albury 13.4 22.9 0.6 44.0 \n",

"1 2008-12-02 Albury 7.4 25.1 0.0 44.0 \n",

"2 2008-12-03 Albury 12.9 25.7 0.0 46.0 \n",

"3 2008-12-04 Albury 9.2 28.0 0.0 24.0 \n",

"4 2008-12-05 Albury 17.5 32.3 1.0 41.0 \n",

"\n",

" WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am \\\n",

"0 20.0 24.0 71.0 22.0 1007.7 \n",

"1 4.0 22.0 44.0 25.0 1010.6 \n",

"2 19.0 26.0 38.0 30.0 1007.6 \n",

"3 11.0 9.0 45.0 16.0 1017.6 \n",

"4 7.0 20.0 82.0 33.0 1010.8 \n",

"\n",

" Pressure3pm Temp9am Temp3pm RainTomorrow \n",

"0 1007.1 16.9 21.8 No \n",

"1 1007.8 17.2 24.3 No \n",

"2 1008.7 21.0 23.2 No \n",

"3 1012.8 18.1 26.5 No \n",

"4 1006.0 17.8 29.7 No "

]

},

"execution\_count": 44,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"data.head()"

]

},

{

"cell\_type": "code",

"execution\_count": 45,

"metadata": {

"id": "pHjibky7UaZ6"

},

"outputs": [],

"source": [

"data=pd.concat([data,data\_c],axis=1)"

]

},

{

"cell\_type": "code",

"execution\_count": 46,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 357

},

"id": "OKway2emG53W",

"outputId": "3423eef6-934c-45aa-cedf-18d194cc0097"

},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>Date</th>\n",

" <th>Location</th>\n",

" <th>MinTemp</th>\n",

" <th>MaxTemp</th>\n",

" <th>Rainfall</th>\n",

" <th>WindGustSpeed</th>\n",

" <th>WindSpeed9am</th>\n",

" <th>WindSpeed3pm</th>\n",

" <th>Humidity9am</th>\n",

" <th>Humidity3pm</th>\n",

" <th>Pressure9am</th>\n",

" <th>Pressure3pm</th>\n",

" <th>Temp9am</th>\n",

" <th>Temp3pm</th>\n",

" <th>RainTomorrow</th>\n",

" <th>RainToday</th>\n",

" <th>WindGustDir</th>\n",

" <th>WindDir9am</th>\n",

" <th>WindDir3pm</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>2008-12-01</td>\n",

" <td>Albury</td>\n",

" <td>13.4</td>\n",

" <td>22.9</td>\n",

" <td>0.6</td>\n",

" <td>44.0</td>\n",

" <td>20.0</td>\n",

" <td>24.0</td>\n",

" <td>71.0</td>\n",

" <td>22.0</td>\n",

" <td>1007.7</td>\n",

" <td>1007.1</td>\n",

" <td>16.9</td>\n",

" <td>21.8</td>\n",

" <td>No</td>\n",

" <td>No</td>\n",

" <td>W</td>\n",

" <td>W</td>\n",

" <td>WNW</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>2008-12-02</td>\n",

" <td>Albury</td>\n",

" <td>7.4</td>\n",

" <td>25.1</td>\n",

" <td>0.0</td>\n",

" <td>44.0</td>\n",

" <td>4.0</td>\n",

" <td>22.0</td>\n",

" <td>44.0</td>\n",

" <td>25.0</td>\n",

" <td>1010.6</td>\n",

" <td>1007.8</td>\n",

" <td>17.2</td>\n",

" <td>24.3</td>\n",

" <td>No</td>\n",

" <td>No</td>\n",

" <td>WNW</td>\n",

" <td>NNW</td>\n",

" <td>WSW</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>2008-12-03</td>\n",

" <td>Albury</td>\n",

" <td>12.9</td>\n",

" <td>25.7</td>\n",

" <td>0.0</td>\n",

" <td>46.0</td>\n",

" <td>19.0</td>\n",

" <td>26.0</td>\n",

" <td>38.0</td>\n",

" <td>30.0</td>\n",

" <td>1007.6</td>\n",

" <td>1008.7</td>\n",

" <td>21.0</td>\n",

" <td>23.2</td>\n",

" <td>No</td>\n",

" <td>No</td>\n",

" <td>WSW</td>\n",

" <td>W</td>\n",

" <td>WSW</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>2008-12-04</td>\n",

" <td>Albury</td>\n",

" <td>9.2</td>\n",

" <td>28.0</td>\n",

" <td>0.0</td>\n",

" <td>24.0</td>\n",

" <td>11.0</td>\n",

" <td>9.0</td>\n",

" <td>45.0</td>\n",

" <td>16.0</td>\n",

" <td>1017.6</td>\n",

" <td>1012.8</td>\n",

" <td>18.1</td>\n",

" <td>26.5</td>\n",

" <td>No</td>\n",

" <td>No</td>\n",

" <td>NE</td>\n",

" <td>SE</td>\n",

" <td>E</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>2008-12-05</td>\n",

" <td>Albury</td>\n",

" <td>17.5</td>\n",

" <td>32.3</td>\n",

" <td>1.0</td>\n",

" <td>41.0</td>\n",

" <td>7.0</td>\n",

" <td>20.0</td>\n",

" <td>82.0</td>\n",

" <td>33.0</td>\n",

" <td>1010.8</td>\n",

" <td>1006.0</td>\n",

" <td>17.8</td>\n",

" <td>29.7</td>\n",

" <td>No</td>\n",

" <td>No</td>\n",

" <td>W</td>\n",

" <td>ENE</td>\n",

" <td>NW</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" Date Location MinTemp MaxTemp Rainfall WindGustSpeed \\\n",

"0 2008-12-01 Albury 13.4 22.9 0.6 44.0 \n",

"1 2008-12-02 Albury 7.4 25.1 0.0 44.0 \n",

"2 2008-12-03 Albury 12.9 25.7 0.0 46.0 \n",

"3 2008-12-04 Albury 9.2 28.0 0.0 24.0 \n",

"4 2008-12-05 Albury 17.5 32.3 1.0 41.0 \n",

"\n",

" WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am \\\n",

"0 20.0 24.0 71.0 22.0 1007.7 \n",

"1 4.0 22.0 44.0 25.0 1010.6 \n",

"2 19.0 26.0 38.0 30.0 1007.6 \n",

"3 11.0 9.0 45.0 16.0 1017.6 \n",

"4 7.0 20.0 82.0 33.0 1010.8 \n",

"\n",

" Pressure3pm Temp9am Temp3pm RainTomorrow RainToday WindGustDir \\\n",

"0 1007.1 16.9 21.8 No No W \n",

"1 1007.8 17.2 24.3 No No WNW \n",

"2 1008.7 21.0 23.2 No No WSW \n",

"3 1012.8 18.1 26.5 No No NE \n",

"4 1006.0 17.8 29.7 No No W \n",

"\n",

" WindDir9am WindDir3pm \n",

"0 W WNW \n",

"1 NNW WSW \n",

"2 W WSW \n",

"3 SE E \n",

"4 ENE NW "

]

},

"execution\_count": 46,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"data.head()"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "yZroCR\_nb5Ys"

},

"source": [

"\*\*5. Data Visualization\*\*"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "XmN85uNScCrB"

},

"source": [

"\n",

"\n",

"> Data Correlation\n",

"\n"

]

},

{

"cell\_type": "code",

"execution\_count": 47,

"metadata": {

"id": "Y3-jYKy9UggM"

},

"outputs": [],

"source": [

"corr=data.corr()"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "ssSgCsWZcIWl"

},

"source": [

"\n",

"\n",

"> Heat-Map\n",

"\n"

]

},

{

"cell\_type": "code",

"execution\_count": 48,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 358

},

"id": "Oc6-QvSNUlZ7",

"outputId": "f4973c4e-e82c-4f99-f58d-3f1ba35ff598"

},

"outputs": [

{

"data": {

"text/plain": [

"<AxesSubplot:>"

]

},

"execution\_count": 48,

"metadata": {},

"output\_type": "execute\_result"

},

{

"data": {

"image/png": "\n",

"text/plain": [

"<Figure size 432x288 with 2 Axes>"

]

},

"metadata": {

"needs\_background": "light"

},

"output\_type": "display\_data"

}

],

"source": [

"sb.heatmap(data=corr,xticklabels=corr.columns.values,yticklabels=corr.columns.values)"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "zzFHLB5pcLcS"

},

"source": [

"\n",

"\n",

"> Pair Plot\n",

"\n"

]

},

{

"cell\_type": "code",

"execution\_count": 49,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 519

},

"id": "cQKagJZoU8lk",

"outputId": "d5d47455-16fc-4ecc-dcb5-a9662989230d"

},

"outputs": [

{

"data": {

"text/plain": [

"<seaborn.axisgrid.JointGrid at 0x2d5ab394ee0>"

]

},

"execution\_count": 49,

"metadata": {},

"output\_type": "execute\_result"

},

{

"data": {

"image/png": "\n",

"text/plain": [

"<Figure size 432x432 with 3 Axes>"

]

},

"metadata": {

"needs\_background": "light"

},

"output\_type": "display\_data"

}

],

"source": [

"sb.jointplot(data[\"MinTemp\"],data['Rainfall'])"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "H\_4v5OJycQTH"

},

"source": [

"\n",

"\n",

"> Box Plot\n",

"\n"

]

},

{

"cell\_type": "code",

"execution\_count": 50,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 343

},

"id": "eFz3FGe9VPBj",

"outputId": "c488dc0f-ec99-4f88-c9f6-0539385c85fd"

},

"outputs": [

{

"data": {

"text/plain": [

"<AxesSubplot:>"

]

},

"execution\_count": 50,

"metadata": {},

"output\_type": "execute\_result"

},

{

"data": {

"image/png": "\n",

"text/plain": [

"<Figure size 432x288 with 1 Axes>"

]

},

"metadata": {

"needs\_background": "light"

},

"output\_type": "display\_data"

}

],

"source": [

"data.boxplot()"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "kQMT3f1ecXPw"

},

"source": [

"\n",

"\n",

"> Hist Plot\n",

"\n"

]

},

{

"cell\_type": "code",

"execution\_count": 51,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 298

},

"id": "vwpvl2-VXsMI",

"outputId": "b6aea678-509c-4194-fa52-bd0e2c539cc0"

},

"outputs": [

{

"data": {

"text/plain": [

"<AxesSubplot:xlabel='RainTomorrow', ylabel='Count'>"

]

},

"execution\_count": 51,

"metadata": {},

"output\_type": "execute\_result"

},

{

"data": {

"image/png": "\n",

"text/plain": [

"<Figure size 432x288 with 1 Axes>"

]

},

"metadata": {

"needs\_background": "light"

},

"output\_type": "display\_data"

}

],

"source": [

"sb.histplot(data['RainTomorrow'])"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "ppKSsV4xchkD"

},

"source": [

"\n",

"\n",

"> Scatter Plot\n",

"\n"

]

},

{

"cell\_type": "code",

"execution\_count": 52,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 357

},

"id": "vjH83mWAX24Q",

"outputId": "06f43615-c600-4403-966b-dd02746923cb"

},

"outputs": [

{

"data": {

"text/plain": [

"<AxesSubplot:xlabel='MaxTemp', ylabel='Rainfall'>"

]

},

"execution\_count": 52,

"metadata": {},

"output\_type": "execute\_result"

},

{

"data": {

"image/png": "\n",

"text/plain": [

"<Figure size 432x288 with 1 Axes>"

]

},

"metadata": {

"needs\_background": "light"

},

"output\_type": "display\_data"

}

],

"source": [

"sb.scatterplot(data['MaxTemp'],data['Rainfall'])"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "cBTfCR2Tcs5Z"

},

"source": [

"\n",

"\n",

"> Distribution Plot\n",

"\n"

]

},

{

"cell\_type": "code",

"execution\_count": 53,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 388

},

"id": "nDJl5aQnX68X",

"outputId": "63ad448f-f736-4b65-bf02-ab239c75b06a"

},

"outputs": [

{

"data": {

"text/plain": [

"<seaborn.axisgrid.FacetGrid at 0x2d5ba9119d0>"

]

},

"execution\_count": 53,

"metadata": {},

"output\_type": "execute\_result"

},

{

"data": {

"image/png": "\n",

"text/plain": [

"<Figure size 360x360 with 1 Axes>"

]

},

"metadata": {

"needs\_background": "light"

},

"output\_type": "display\_data"

}

],

"source": [

"sb.displot(data['MinTemp'])"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "5oSLBnync5W8"

},

"source": [

"\*\*6. Splitting The Dateset Into Dependent And Independent Variable\*\*"

]

},

{

"cell\_type": "code",

"execution\_count": 54,

"metadata": {

"id": "LKBxqSX2YcEb"

},

"outputs": [],

"source": [

"from sklearn.preprocessing import StandardScaler"

]

},

{

"cell\_type": "code",

"execution\_count": 55,

"metadata": {

"id": "Tg7ryshQJJ65"

},

"outputs": [],

"source": [

"data = data[data['RainTomorrow'].notnull()]"

]

},

{

"cell\_type": "code",

"execution\_count": 56,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "tkoc-QUyuG36",

"outputId": "283ec314-fe3f-4e30-b2c2-e8ea41be2b56"

},

"outputs": [],

"source": [

"data['Pressure9am'].fillna(data['Pressure9am'].mean(),inplace=True)\n",

"data['Pressure3pm'].fillna(data['Pressure3pm'].mean(),inplace=True)"

]

},

{

"cell\_type": "code",

"execution\_count": 57,

"metadata": {

"id": "1wbvuzYkYjSz"

},

"outputs": [],

"source": [

"y=data['RainTomorrow']\n",

"x=data.drop('RainTomorrow',axis=1)"

]

},

{

"cell\_type": "code",

"execution\_count": 58,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "biHAAPY6IGh4",

"outputId": "f5cf0f8c-bc7e-457d-9d86-8e93171997c6"

},

"outputs": [

{

"data": {

"text/plain": [

"{'No', 'Yes'}"

]

},

"execution\_count": 58,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"set(y)"

]

},

{

"cell\_type": "code",

"execution\_count": 59,

"metadata": {

"id": "haPqgb7jZGvX"

},

"outputs": [],

"source": [

"x=x.drop('Date',axis=1)"

]

},

{

"cell\_type": "code",

"execution\_count": 60,

"metadata": {

"id": "A44pQlh1YnAa"

},

"outputs": [],

"source": [

"names=x.columns"

]

},

{

"cell\_type": "code",

"execution\_count": 61,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "Y5JId4zNYrDp",

"outputId": "52a20b62-62b7-4de7-9f99-d664837ac80e"

},

"outputs": [

{

"data": {

"text/plain": [

"Index(['Location', 'MinTemp', 'MaxTemp', 'Rainfall', 'WindGustSpeed',\n",

" 'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm',\n",

" 'Pressure9am', 'Pressure3pm', 'Temp9am', 'Temp3pm', 'RainToday',\n",

" 'WindGustDir', 'WindDir9am', 'WindDir3pm'],\n",

" dtype='object')"

]

},

"execution\_count": 61,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"names"

]

},

{

"cell\_type": "code",

"execution\_count": 62,

"metadata": {

"id": "Aprl2NKLYsae"

},

"outputs": [],

"source": [

"\n",

"sc=StandardScaler()"

]

},

{

"cell\_type": "code",

"execution\_count": 63,

"metadata": {

"id": "rfWxdrD2tagY"

},

"outputs": [],

"source": [

"from sklearn.preprocessing import LabelEncoder, MinMaxScaler"

]

},

{

"cell\_type": "code",

"execution\_count": 64,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "E-GZ6YQvKVWL",

"outputId": "1f9f6723-aa0f-40d2-eb66-ff99e78aa359"

},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"142193 142193\n"

]

}

],

"source": [

"print(len(x),len(y))"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "zVQjYLiYgI0H"

},

"source": [

"\*\*7. Label Encoding\*\*"

]

},

{

"cell\_type": "code",

"execution\_count": 65,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 270

},

"id": "yXPvaP\_jY3Xb",

"outputId": "b72b98c1-47ee-4d5f-ca6b-a2430e291484"

},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>Location</th>\n",

" <th>MinTemp</th>\n",

" <th>MaxTemp</th>\n",

" <th>Rainfall</th>\n",

" <th>WindGustSpeed</th>\n",

" <th>WindSpeed9am</th>\n",

" <th>WindSpeed3pm</th>\n",

" <th>Humidity9am</th>\n",

" <th>Humidity3pm</th>\n",

" <th>Pressure9am</th>\n",

" <th>Pressure3pm</th>\n",

" <th>Temp9am</th>\n",

" <th>Temp3pm</th>\n",

" <th>RainToday</th>\n",

" <th>WindGustDir</th>\n",

" <th>WindDir9am</th>\n",

" <th>WindDir3pm</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>2</td>\n",

" <td>13.4</td>\n",

" <td>22.9</td>\n",

" <td>0.6</td>\n",

" <td>44.0</td>\n",

" <td>20.0</td>\n",

" <td>24.0</td>\n",

" <td>71.0</td>\n",

" <td>22.0</td>\n",

" <td>1007.7</td>\n",

" <td>1007.1</td>\n",

" <td>16.9</td>\n",

" <td>21.8</td>\n",

" <td>0</td>\n",

" <td>13</td>\n",

" <td>13</td>\n",

" <td>14</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>2</td>\n",

" <td>7.4</td>\n",

" <td>25.1</td>\n",

" <td>0.0</td>\n",

" <td>44.0</td>\n",

" <td>4.0</td>\n",

" <td>22.0</td>\n",

" <td>44.0</td>\n",

" <td>25.0</td>\n",

" <td>1010.6</td>\n",

" <td>1007.8</td>\n",

" <td>17.2</td>\n",

" <td>24.3</td>\n",

" <td>0</td>\n",

" <td>14</td>\n",

" <td>6</td>\n",

" <td>15</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>2</td>\n",

" <td>12.9</td>\n",

" <td>25.7</td>\n",

" <td>0.0</td>\n",

" <td>46.0</td>\n",

" <td>19.0</td>\n",

" <td>26.0</td>\n",

" <td>38.0</td>\n",

" <td>30.0</td>\n",

" <td>1007.6</td>\n",

" <td>1008.7</td>\n",

" <td>21.0</td>\n",

" <td>23.2</td>\n",

" <td>0</td>\n",

" <td>15</td>\n",

" <td>13</td>\n",

" <td>15</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>2</td>\n",

" <td>9.2</td>\n",

" <td>28.0</td>\n",

" <td>0.0</td>\n",

" <td>24.0</td>\n",

" <td>11.0</td>\n",

" <td>9.0</td>\n",

" <td>45.0</td>\n",

" <td>16.0</td>\n",

" <td>1017.6</td>\n",

" <td>1012.8</td>\n",

" <td>18.1</td>\n",

" <td>26.5</td>\n",

" <td>0</td>\n",

" <td>4</td>\n",

" <td>9</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>2</td>\n",

" <td>17.5</td>\n",

" <td>32.3</td>\n",

" <td>1.0</td>\n",

" <td>41.0</td>\n",

" <td>7.0</td>\n",

" <td>20.0</td>\n",

" <td>82.0</td>\n",

" <td>33.0</td>\n",

" <td>1010.8</td>\n",

" <td>1006.0</td>\n",

" <td>17.8</td>\n",

" <td>29.7</td>\n",

" <td>0</td>\n",

" <td>13</td>\n",

" <td>1</td>\n",

" <td>7</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>"

],

"text/plain": [

" Location MinTemp MaxTemp Rainfall WindGustSpeed WindSpeed9am \\\n",

"0 2 13.4 22.9 0.6 44.0 20.0 \n",

"1 2 7.4 25.1 0.0 44.0 4.0 \n",

"2 2 12.9 25.7 0.0 46.0 19.0 \n",

"3 2 9.2 28.0 0.0 24.0 11.0 \n",

"4 2 17.5 32.3 1.0 41.0 7.0 \n",

"\n",

" WindSpeed3pm Humidity9am Humidity3pm Pressure9am Pressure3pm Temp9am \\\n",

"0 24.0 71.0 22.0 1007.7 1007.1 16.9 \n",

"1 22.0 44.0 25.0 1010.6 1007.8 17.2 \n",

"2 26.0 38.0 30.0 1007.6 1008.7 21.0 \n",

"3 9.0 45.0 16.0 1017.6 1012.8 18.1 \n",

"4 20.0 82.0 33.0 1010.8 1006.0 17.8 \n",

"\n",

" Temp3pm RainToday WindGustDir WindDir9am WindDir3pm \n",

"0 21.8 0 13 13 14 \n",

"1 24.3 0 14 6 15 \n",

"2 23.2 0 15 13 15 \n",

"3 26.5 0 4 9 0 \n",

"4 29.7 0 13 1 7 "

]

},

"execution\_count": 65,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"## RainToday\tWindGustDir\tWindDir9am\tWindDir3pm\n",

"\n",

"LE = LabelEncoder()\n",

"x['Location'] = LE.fit\_transform(x['Location'])\n",

"x.head()\n",

"\n",

"LE = LabelEncoder()\n",

"x['RainToday'] = LE.fit\_transform(x['RainToday'])\n",

"x.head()\n",

"\n",

"LE = LabelEncoder()\n",

"x['WindGustDir'] = LE.fit\_transform(x['WindGustDir'])\n",

"x.head()\n",

"\n",

"LE = LabelEncoder()\n",

"x['WindDir9am'] = LE.fit\_transform(x['WindDir9am'])\n",

"x.head()\n",

"\n",

"LE = LabelEncoder()\n",

"x['WindDir3pm'] = LE.fit\_transform(x['WindDir3pm'])\n",

"x.head()\n"

]

},

{

"cell\_type": "code",

"execution\_count": 66,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "yF\_W\_zo6HgJO",

"outputId": "6777d39a-56b7-458b-fabb-8c9d7796892a"

},

"outputs": [],

"source": [

"LE = LabelEncoder()\n",

"y=pd.DataFrame(y)\n",

"y = LE.fit\_transform(y)\n"

]

},

{

"cell\_type": "code",

"execution\_count": 67,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "oNsfvKze\_JvL",

"outputId": "74eacc4a-cb6e-4ffd-b29a-4736f90f08cb"

},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"142193 142193\n"

]

}

],

"source": [

"print(len(x),len(y))"

]

},

{

"cell\_type": "code",

"execution\_count": 68,

"metadata": {

"id": "QcWmILDICDxI"

},

"outputs": [],

"source": [

"sc=StandardScaler()\n"

]

},

{

"cell\_type": "code",

"execution\_count": 69,

"metadata": {

"id": "5Hd66Qt7CGfp"

},

"outputs": [],

"source": [

"x=sc.fit\_transform(x)"

]

},

{

"cell\_type": "code",

"execution\_count": 70,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "jFOx-KnACIkQ",

"outputId": "417d7b0f-bf34-47d1-c543-93957fefdc63"

},

"outputs": [

{

"data": {

"text/plain": [

"array([[-1.5270045 , 0.1899491 , -0.04596252, -0.2077696 , 0.30539521,\n",

" 0.67761657, 0.61479645, 0.11386682, -1.43600466, -1.47545613,\n",

" -1.22096552, -0.01352387, 0.01642307, -0.53296232, 1.05255576,\n",

" 1.32893289, 1.36627749],\n",

" [-1.5270045 , -0.74917952, 0.26348131, -0.27900154, 0.30539521,\n",

" -1.13007826, 0.38547865, -1.31228915, -1.28989124, -1.04558606,\n",

" -1.11620276, 0.03282863, 0.38028454, -0.53296232, 1.26558231,\n",

" -0.2215706 , 1.58623629],\n",

" [-1.5270045 , 0.11168839, 0.34787508, -0.27900154, 0.45762138,\n",

" 0.56463565, 0.84411424, -1.6292127 , -1.04636888, -1.49027923,\n",

" -0.98150779, 0.61996026, 0.22018549, -0.53296232, 1.47860885,\n",

" 1.32893289, 1.58623629],\n",

" [-1.5270045 , -0.46744093, 0.67138453, -0.27900154, -1.21686656,\n",

" -0.33921177, -1.10508701, -1.25946856, -1.7282315 , -0.00796867,\n",

" -0.36789739, 0.17188612, 0.70048263, -0.53296232, -0.86468316,\n",

" 0.4429309 , -1.71314577],\n",

" [-1.5270045 , 0.831687 , 1.27620655, -0.16028164, 0.07705594,\n",

" -0.79113548, 0.15616086, 0.69489333, -0.90025546, -1.01593985,\n",

" -1.3855927 , 0.12553362, 1.16622532, -0.53296232, 1.05255576,\n",

" -1.32907309, -0.17343414]])"

]

},

"execution\_count": 70,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"x[:5]"

]

},

{

"cell\_type": "code",

"execution\_count": 71,

"metadata": {

"id": "QKsQijoFCJt1"

},

"outputs": [],

"source": [

"x=pd.DataFrame(x,columns=names)"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "Dx\_Gp4vIgYha"

},

"source": [

"\*\*8. Splitting The Data Into Train And Test\*\*"

]

},

{

"cell\_type": "code",

"execution\_count": 72,

"metadata": {

"id": "6UCjUvN\_CYJ4"

},

"outputs": [],

"source": [

"from sklearn import model\_selection"

]

},

{

"cell\_type": "code",

"execution\_count": 73,

"metadata": {

"id": "olNItHq7Cf8n"

},

"outputs": [],

"source": [

"x\_train,x\_test,y\_train,y\_test=model\_selection.train\_test\_split(x,y,test\_size=0.2,random\_state=0)"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "0Adz9U77gn5s"

},

"source": [

"\*\*9. Training And Testing The Model\*\*"

]

},

{

"cell\_type": "code",

"execution\_count": 74,

"metadata": {

"id": "wamKRG4-C-yj"

},

"outputs": [],

"source": [

"from sklearn.ensemble import RandomForestClassifier \n",

"from sklearn.ensemble import GradientBoostingClassifier"

]

},

{

"cell\_type": "code",

"execution\_count": 75,

"metadata": {

"id": "SGT03rNPD2mF"

},

"outputs": [],

"source": [

"RFC=RandomForestClassifier()"

]

},

{

"cell\_type": "code",

"execution\_count": 76,

"metadata": {

"id": "WPhjVhZgMo50"

},

"outputs": [],

"source": [

"GBC=GradientBoostingClassifier()"

]

},

{

"cell\_type": "code",

"execution\_count": 77,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "uqM5VpewEXJ0",

"outputId": "f2f53c3a-dbd1-421a-e142-caf24071b5d8"

},

"outputs": [

{

"data": {

"text/plain": [

"False"

]

},

"execution\_count": 77,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"np.any(np.isnan(x))"

]

},

{

"cell\_type": "code",

"execution\_count": 78,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "2HHaHHSzMtPy",

"outputId": "cc902c6d-7f8c-46ae-abf0-8dd63c74f387"

},

"outputs": [

{

"data": {

"text/plain": [

"GradientBoostingClassifier()"

]

},

"execution\_count": 78,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"GBC.fit(x\_train,y\_train)"

]

},

{

"cell\_type": "code",

"execution\_count": 79,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "NCOvHh3aDHcX",

"outputId": "793b3168-194b-430d-8494-8abc01bd304c"

},

"outputs": [

{

"data": {

"text/plain": [

"RandomForestClassifier()"

]

},

"execution\_count": 79,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"RFC.fit(x\_train,y\_train)"

]

},

{

"cell\_type": "code",

"execution\_count": 80,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "OYTcMJ77tQWH",

"outputId": "b2c3602f-7959-4c0b-ba63-3595cea6b319"

},

"outputs": [

{

"data": {

"text/plain": [

"Date False\n",

"Location False\n",

"MinTemp False\n",

"MaxTemp False\n",

"Rainfall False\n",

"WindGustSpeed False\n",

"WindSpeed9am False\n",

"WindSpeed3pm False\n",

"Humidity9am False\n",

"Humidity3pm False\n",

"Pressure9am False\n",

"Pressure3pm False\n",

"Temp9am False\n",

"Temp3pm False\n",

"RainTomorrow False\n",

"RainToday False\n",

"WindGustDir False\n",

"WindDir9am False\n",

"WindDir3pm False\n",

"dtype: bool"

]

},

"execution\_count": 80,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"data.isnull().any()"

]

},

{

"cell\_type": "code",

"execution\_count": 86,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "76AjF0O\_toWD",

"outputId": "aae3cd67-a139-4356-eb9f-3de032e4aee0"

},

"outputs": [

{

"data": {

"text/plain": [

"Location False\n",

"MinTemp False\n",

"MaxTemp False\n",

"Rainfall False\n",

"WindGustSpeed False\n",

"WindSpeed9am False\n",

"WindSpeed3pm False\n",

"Humidity9am False\n",

"Humidity3pm False\n",

"Pressure9am False\n",

"Pressure3pm False\n",

"Temp9am False\n",

"Temp3pm False\n",

"RainToday False\n",

"WindGustDir False\n",

"WindDir9am False\n",

"WindDir3pm False\n",

"dtype: bool"

]

},

"execution\_count": 86,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"x.isnull().any()"

]

},

{

"cell\_type": "code",

"execution\_count": 87,

"metadata": {

"id": "abyHf3lduBEw"

},

"outputs": [],

"source": [

"p1=RFC.predict(x\_train)"

]

},

{

"cell\_type": "code",

"execution\_count": 88,

"metadata": {

"id": "sRVbTXIUzkVK"

},

"outputs": [],

"source": [

"p2=RFC.predict(x\_test)"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "jAYeCaZtgwkm"

},

"source": [

"\*\*10. Model Evaluation\*\*"

]

},

{

"cell\_type": "code",

"execution\_count": 89,

"metadata": {

"id": "\_n1qAJrfwQGA"

},

"outputs": [],

"source": [

"import sklearn.metrics as metrics"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "pSf7E3L4g2nF"

},

"source": [

"\n",

"\n",

"> Accuracy\_score\n",

"\n"

]

},

{

"cell\_type": "code",

"execution\_count": 90,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "WMZ3zjOLu7lz",

"outputId": "e1e49d42-82b3-4b7a-c086-8127fed3da7b"

},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"0.9999472546020359\n"

]

}

],

"source": [

"print(metrics.accuracy\_score(y\_train,p1))"

]

},

{

"cell\_type": "code",

"execution\_count": 91,

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "CwDYIUM5zxLV",

"outputId": "24855768-645f-48d9-aa70-a74ac0ade232"

},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"0.8567460177924681\n"

]

}

],

"source": [

"print(metrics.accuracy\_score(y\_test,p2))"

]

},

{

"cell\_type": "markdown",

"metadata": {

"id": "h2BUtjkcg8cg"

},

"source": [

"\*\*11. Save The Model\*\*"

]

},

{

"cell\_type": "code",

"execution\_count": 94,

"metadata": {

"id": "MKtSLd5dwGRi"

},

"outputs": [],

"source": [

"import pickle"

]

},

{

"cell\_type": "code",

"execution\_count": 95,

"metadata": {

"id": "Ejzj2MENy0cC"

},

"outputs": [],

"source": [

"pickle.dump(RFC,open('rainfall.pkl','wb'))\n",

"pickle.dump(LE,open('encoder.pkl','wb'))\n",

"pickle.dump(imp\_mode,open('imputer.pkl','wb'))\n",

"pickle.dump(sc,open('scale.pkl','wb'))"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {},

"outputs": [],

"source": []

}

],

"metadata": {

"colab": {

"provenance": []

},

"kernelspec": {

"display\_name": "Python 3 (ipykernel)",

"language": "python",

"name": "python3"

},

"language\_info": {

"codemirror\_mode": {

"name": "ipython",

"version": 3

},

"file\_extension": ".py",

"mimetype": "text/x-python",

"name": "python",

"nbconvert\_exporter": "python",

"pygments\_lexer": "ipython3",

"version": "3.9.12"

}

},

"nbformat": 4,

"nbformat\_minor": 1

}