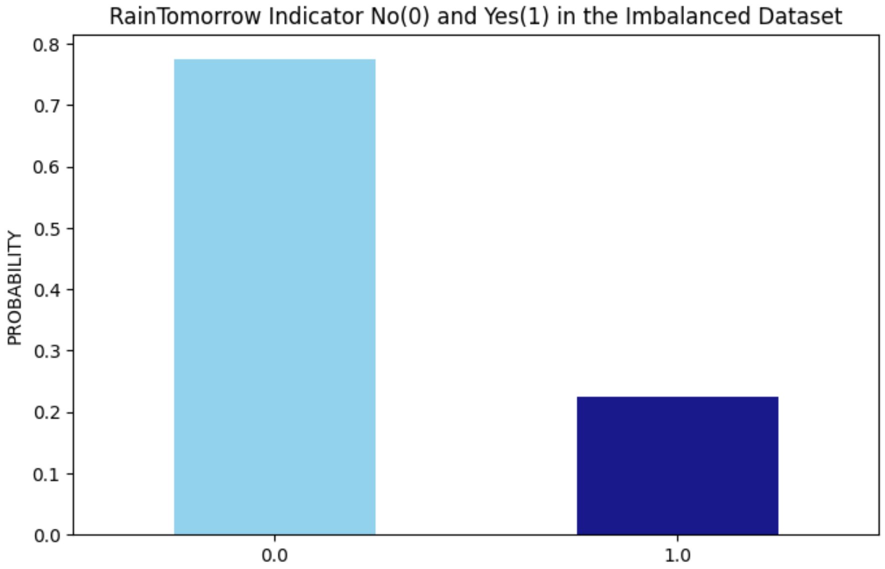
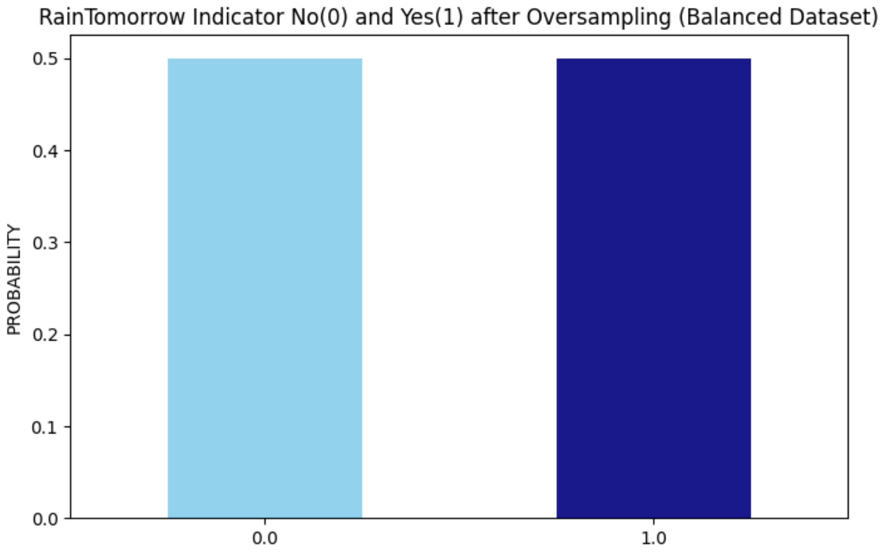
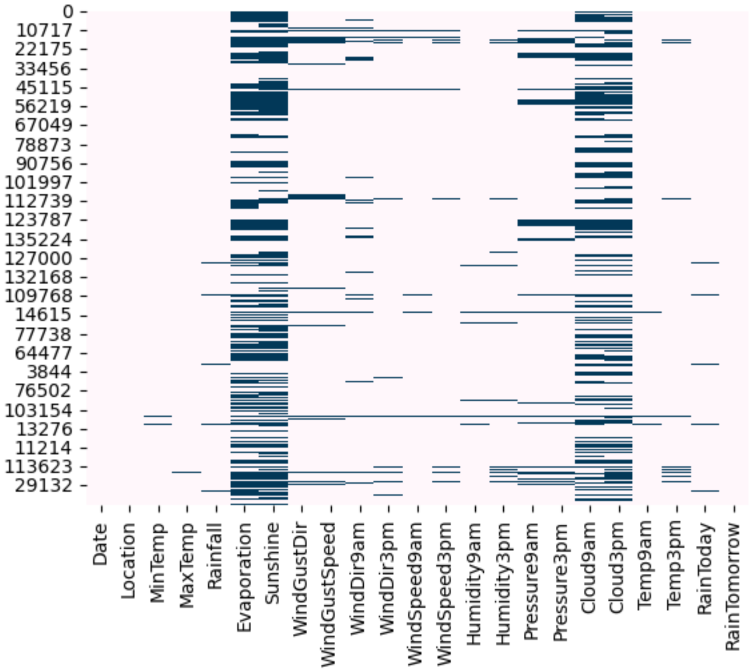
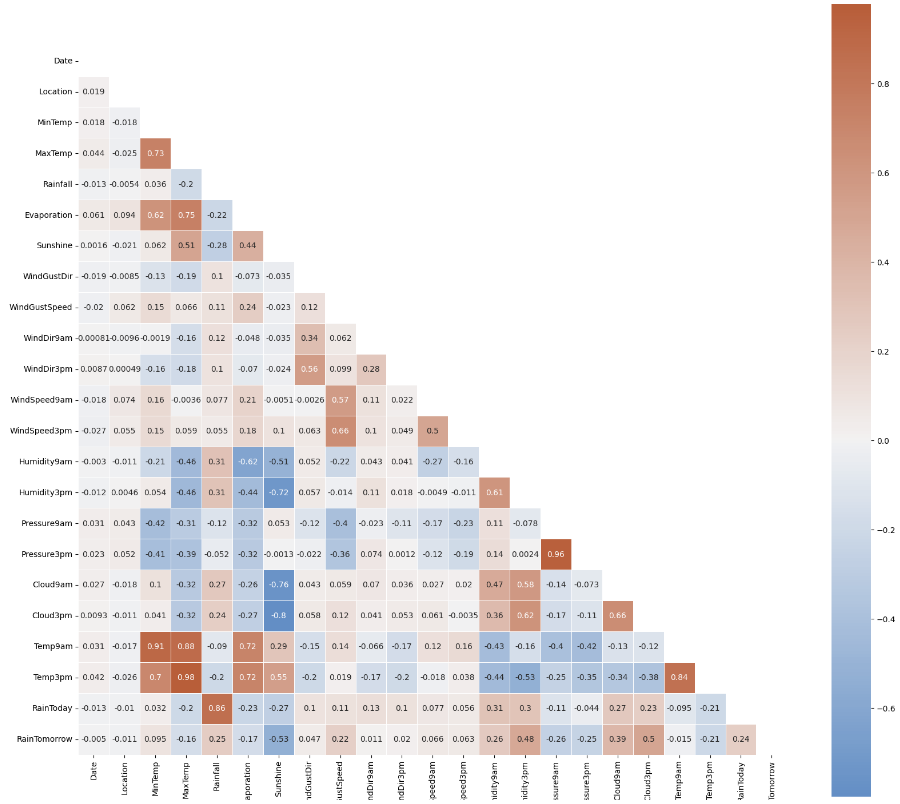
```
import pandas as pd
import numpy as np
full_data = pd.read_csv('/content/weather.csv')
full_data.head()
full_data.shape
full_data.info()
full_data['RainToday'].replace({'No': 0, 'Yes': 1},inplace = True)
full_data['RainTomorrow'].replace({'No': 0, 'Yes': 1},inplace = True)
import matplotlib.pyplot as plt
fig = plt.figure(figsize = (8,5))
full_data.RainTomorrow.value_counts(normalize = True).plot(kind='bar', color= ['skyblue','navy'], alpha = 0.9, rot=0)
plt.title('RainTomorrow Indicator No(0) and Yes(1) in the Imbalanced Dataset')
plt.ylabel('PROBABILITY')
plt.show()
from sklearn.utils import resample
no = full_data[full_data.RainTomorrow == 0]
yes = full_data[full_data.RainTomorrow == 1]
yes_oversampled = resample(yes, replace=True, n_samples=len(no), random_state=123)
oversampled = pd.concat([no, yes_oversampled])
fig = plt.figure(figsize = (8,5))
oversampled.RainTomorrow.value_counts(normalize = True).plot(kind='bar', color= ['skyblue','navy'], alpha = 0.9, rot=0)
plt.title('RainTomorrow Indicator No(0) and Yes(1) after Oversampling (Balanced Dataset)')
plt.ylabel('PROBABILITY')
plt.show()

import seaborn as sns
sns.heatmap(oversampled.isnull(), cbar=False, cmap='PuBu')

total = oversampled.isnull().sum().sort_values(ascending=False)
percent = (oversampled.isnull().sum()/oversampled.isnull().count()).sort_values(ascending=False)
missing = pd.concat([total, percent], axis=1, keys=['Total', 'Percent'])
missing.head(4)

oversampled.select_dtypes(include=['object']).columns
oversampled['Date'] = oversampled['Date'].fillna(oversampled['Date'].mode()[0])
oversampled['Location'] = oversampled['Location'].fillna(oversampled['Location'].mode()[0])
oversampled['WindGustDir'] = oversampled['WindGustDir'].fillna(oversampled['WindGustDir'].mode()[0])
oversampled['WindDir9am'] = oversampled['WindDir9am'].fillna(oversampled['WindDir9am'].mode()[0])
oversampled['WindDir3pm'] = oversampled['WindDir3pm'].fillna(oversampled['WindDir3pm'].mode()[0])
from sklearn.preprocessing import LabelEncoder
lencoders = {}
for col in oversampled.select_dtypes(include=['object']).columns:
    lencoders[col] = LabelEncoder()
    oversampled[col] = lencoders[col].fit_transform(oversampled[col])

import warnings
warnings.filterwarnings("ignore")

from sklearn.experimental import enable_iterative_imputer
from sklearn.impute import IterativeImputer
MiceImputed = oversampled.copy(deep=True)
mice_imputer = IterativeImputer()
MiceImputed.iloc[:, :] = mice_imputer.fit_transform(oversampled)
MiceImputed.head()
MiceImputed.isna()
Q1 = MiceImputed.quantile(0.25)
Q3 = MiceImputed.quantile(0.75)
IQR = Q3 - Q1
print(IQR)
MiceImputed = MiceImputed[~((MiceImputed < (Q1 - 1.5 * IQR)) |(MiceImputed > (Q3 + 1.5 * IQR))).any(axis=1)]
MiceImputed.shape
corr = MiceImputed.corr()
mask = np.triu(np.ones_like(corr, dtype=np.bool_))
f, ax = plt.subplots(figsize=(20, 20))
cmap = sns.diverging_palette(250, 25, as_cmap=True)
sns.heatmap(corr, mask=mask, cmap=cmap, vmax=None, center=0,square=True, annot=True, linewidths=.5, cbar_kws={"shrink": .9})
sns.pairplot( data=MiceImputed, vars=('MaxTemp','MinTemp','Pressure9am','Pressure3pm', 'Temp9am', 'Temp3pm', 'Evaporation'), hue='RainTomorrow
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 145460 entries, 0 to 145459
Data columns (total 23 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   Date           145460 non-null  object 
 1   Location       145460 non-null  object 
 2   MinTemp        143975 non-null  float64
 3   MaxTemp        144199 non-null  float64
 4   Rainfall       142199 non-null  float64
 5   Evaporation    82670 non-null   float64
 6   Sunshine       75625 non-null   float64
 7   WindGustDir    135134 non-null  object 
 8   WindGustSpeed  135197 non-null  float64
 9   WindDir9am     134894 non-null  object 
 10  WindDir3pm     141232 non-null  object 
 11  WindSpeed9am   143693 non-null  float64
 12  WindSpeed3pm   142398 non-null  float64
 13  Humidity9am    142806 non-null  float64
 14  Humidity3pm    140953 non-null  float64
 15  Pressure9am    130395 non-null  float64
 16  Pressure3pm    130432 non-null  float64
 17  Cloud9am       89572 non-null   float64
 18  Cloud3pm       86102 non-null   float64
 19  Temp9am        143693 non-null  float64
 20  Temp3pm        141851 non-null  float64
 21  RainToday      142199 non-null  object 
 22  RainTomorrow   142193 non-null  object 
dtypes: float64(16), object(7)
memory usage: 25.5+ MB
```



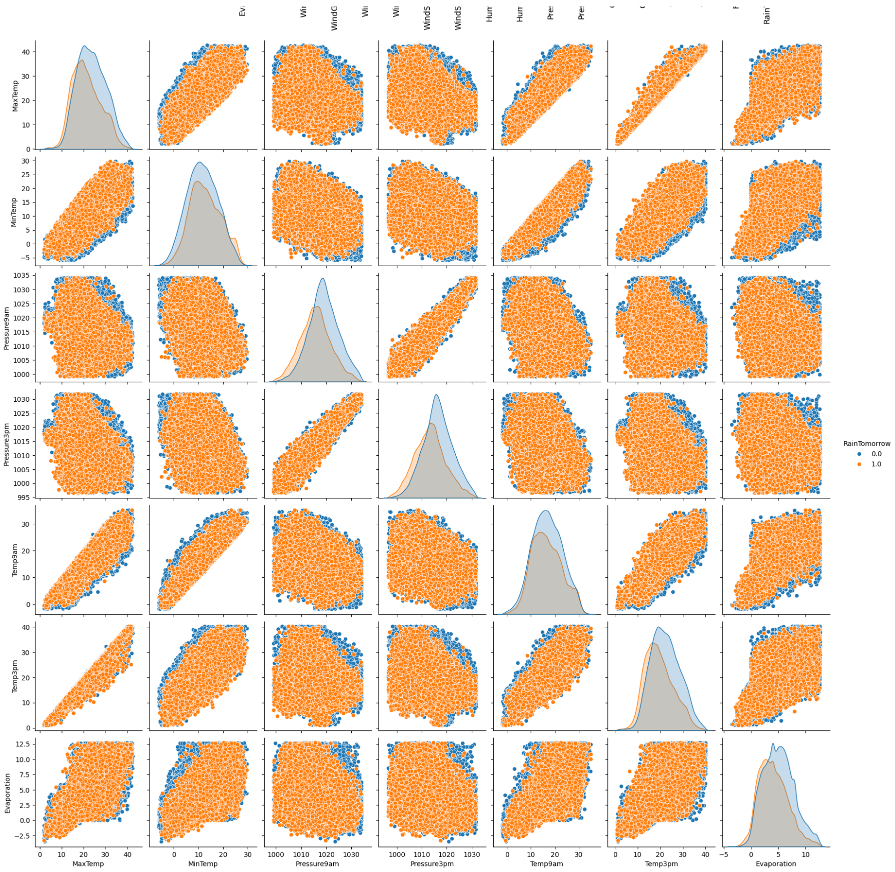


```
Date            1535.000000
Location          25.000000
MinTemp            9.300000
MaxTemp           10.200000
Rainfall           2.400000
Rainfall         2.400000
Evaporation      4.120044
Sunshine         5.979485
WindGustDir      9.000000
WindGustSpeed   19.000000
WindDir9am       8.000000
WindDir3pm       8.000000
WindSpeed9am    13.000000
WindSpeed3pm    11.000000
Humidity9am     26.000000
Humidity3pm     30.000000
Pressure9am      8.800000
Pressure3pm      8.800000
Cloud9am         4.000000
Cloud3pm         3.684676
Temp9am          9.300000
Temp3pm          9.800000
RainToday        1.000000
RainTomorrow     1.000000
dtype: float64
<seaborn.axisgrid.PairGrid at 0x78088ce9e230>
```





	Date	Location	MinTemp	MaxTemp	Rainfall	aporation	Sunshine	ndGustDir	iustSpeed	ndDir9am	ndDir3pm	peed9am	peed3pm	nidity9am	nidity3pm	ssure9am	ssure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	ainToday	Tomorrow
Date																							
Location	0.019																						
MinTemp	0.018	-0.018																					
MaxTemp	0.044	-0.025	0.73																				
Rainfall	-0.013	-0.0054	0.036	-0.2																			
Evaporation	0.061	0.094	0.62	0.75	-0.22																		
Sunshine	0.0016	-0.021	0.062	0.51	-0.28	0.44																	
WindGustDir	-0.019	-0.0085	-0.13	-0.19	0.1	-0.073	-0.035																
WindGustSpeed	-0.02	0.062	0.15	0.066	0.11	0.24	-0.023	0.12															
WindDir9am	-0.00081	-0.0096	-0.0019	-0.16	0.12	-0.048	-0.035	0.34	0.062														
WindDir3pm	0.0087	0.00049	-0.16	-0.18	0.1	-0.07	-0.024	0.56	0.099	0.28													
WindSpeed9am	-0.018	0.074	0.16	-0.0036	0.077	0.21	-0.0051	-0.0026	0.57	0.11	0.022												
WindSpeed3pm	-0.027	0.055	0.15	0.059	0.055	0.18	0.1	0.063	0.66	0.1	0.049	0.5											
Humidity9am	-0.003	-0.011	-0.21	-0.46	0.31	-0.62	-0.51	0.052	-0.22	0.043	0.041	-0.27	-0.16										
Humidity3pm	-0.012	0.0046	0.054	-0.46	0.31	-0.44	-0.72	0.057	-0.014	0.11	0.018	-0.0049	-0.011	0.61									
Pressure9am	0.031	0.043	-0.42	-0.31	-0.12	-0.32	0.053	-0.12	-0.4	-0.023	-0.11	-0.17	-0.23	0.11	-0.078								
Pressure3pm	0.023	0.052	-0.41	-0.39	-0.052	-0.32	-0.0013	-0.022	-0.36	0.074	0.0012	-0.12	-0.19	0.14	0.0024	0.96							
Cloud9am	0.027	-0.018	0.1	-0.32	0.27	-0.26	-0.76	0.043	0.059	0.07	0.036	0.027	0.02	0.47	0.58	-0.14	-0.073						
Cloud3pm	0.0093	-0.011	0.041	-0.32	0.24	-0.27	-0.8	0.058	0.12	0.041	0.053	0.061	-0.0035	0.36	0.62	-0.17	-0.11	0.66					
Temp9am	0.031	-0.017	0.91	0.88	-0.09	0.72	0.29	-0.15	0.14	-0.066	-0.17	0.12	0.16	-0.43	-0.16	-0.4	-0.42	-0.13	-0.12				
Temp3pm	0.042	-0.026	0.7	0.98	-0.2	0.72	0.55	-0.2	0.019	-0.17	-0.2	-0.018	0.038	-0.44	-0.53	-0.25	-0.35	-0.34	-0.38	0.84			
RainToday	-0.013	-0.01	0.032	-0.2	0.86	-0.23	-0.27	0.1	0.11	0.13	0.1	0.077	0.056	0.31	0.3	-0.11	-0.044	0.27	0.23	-0.095	-0.21		
RainTomorrow	-0.005	-0.011	0.095	-0.16	0.25	-0.17	-0.53	0.047	0.22	0.011	0.02	0.066	0.063	0.26	0.48	-0.26	-0.25	0.39	0.5	-0.015	-0.21	0.24	


Ev
Wi
WindG
Wi
Wi
WindS
WindS
Hum
Hum
Pres
Pres
Rain
MaxTemp
MinTemp
Pressure9am
Pressure3pm
Temp9am
Temp3pm
Evaporation
0 10 20 30 40
0 10 20 30
1000 1010 1020 1030
1000 1010 1020 1030
0 10 20 30
0 10 20 30 40
−5 0 5 10
RainTomorrow
0.0
1.0