Table of Contents

Go to:

Rain in Australia

Introduction

The future weather is an important information for lots of areas of life e.g. agriculture or leisure activities

The aim of the project is to predict if it is raining tomorrow based on today weather.

Dataset

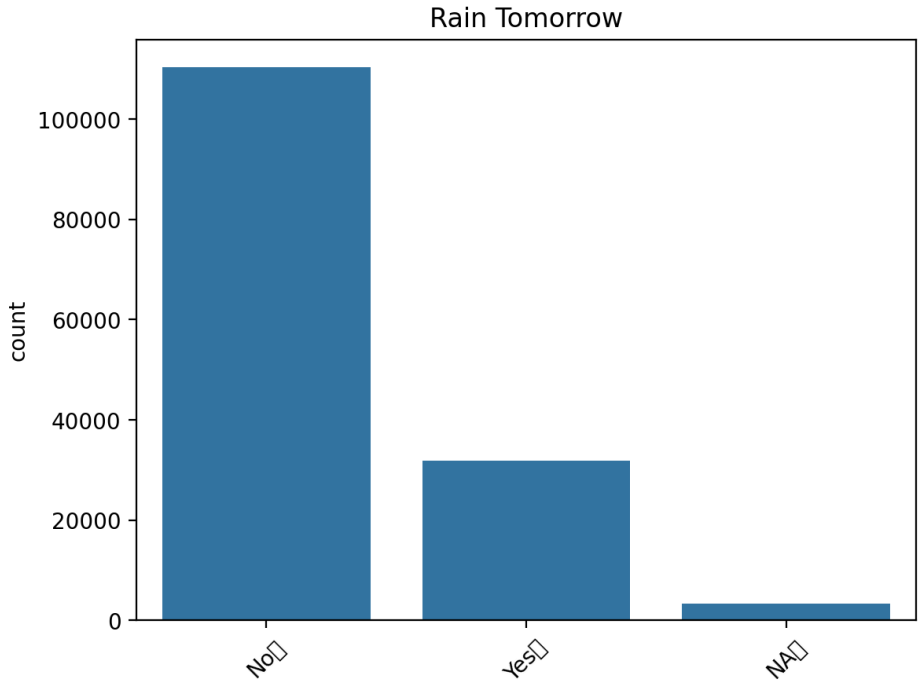
The given Dataset from Australia has `23` different columns and contains `145460` data entrys. The target column is RainTomorrow, which is a Boolean.

	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGu
0	2008-12-01	Albury	13.4	22.9	0.6	0	0	W	
1	2008-12-02	Albury	7.4	25.1	0	0	0	WNW	
2	2008-12-03	Albury	12.9	25.7	0	0	0	WSW	
3	2008-12-04	Albury	9.2	28	0	0	0	NE	
4	2008-12-05	Albury	17.5	32.3	1	0	0	W	
5	2008-12-06	Albury	14.6	29.7	0.2	0	0	WNW	
6	2008-12-07	Albury	14.3	25	0	0	0	W	
7	2008-12-08	Albury	7.7	26.7	0	0	0	W	
8	2008-12-09	Albury	9.7	31.9	0	0	0	NNW	
9	2008-12-10	Albury	13.1	30.1	1.4	0	0	W	

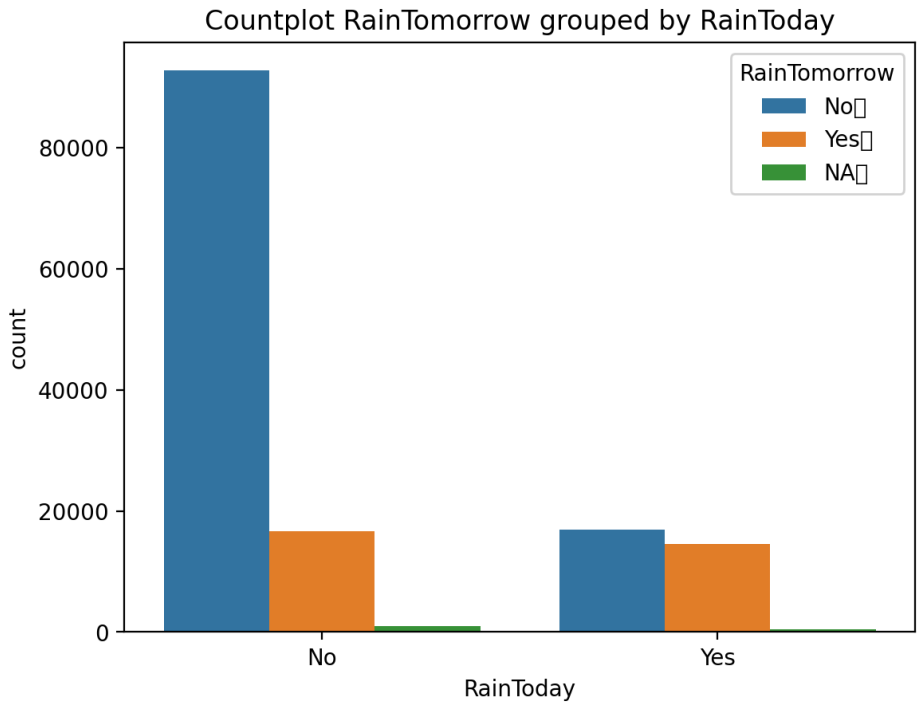
The Data for today contains information about the date, the city, temperature, humidity, pressure, wind, clouds, sunshine and rain. Most variables are numeric. Categorical values are the location and wind related values (e.g. wind direction). Rains today is a boolean. The dataset contains measurements from 49 citys/places.

First observations

The target is unevenly distibuted (fewer rainy days).



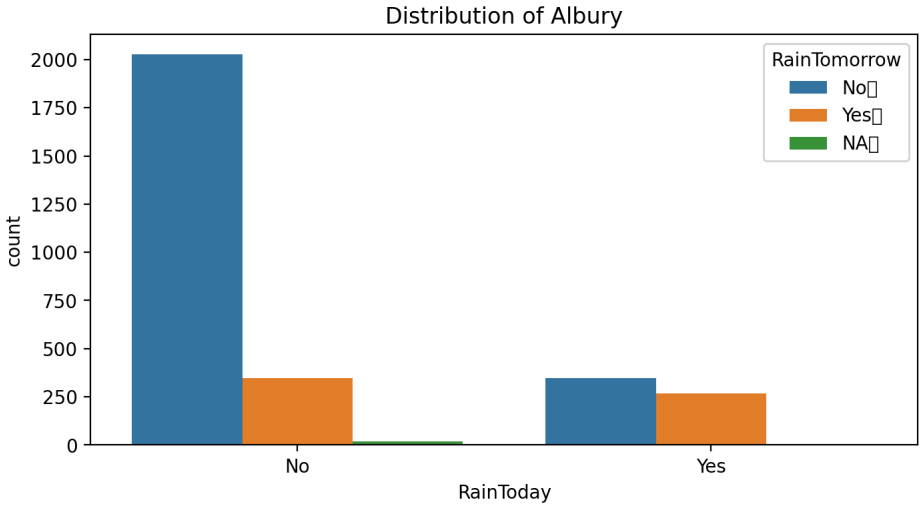
If it rains today, there is 50% chance that it also rains tomorrow. If it does not rain today, it will most likely also not rain tomorrow.



Show missing values

Select a Location:

Albury


Distribution of Albury
RainTomorrow
No
Yes
NA
count
2000
1750
1500
1250
1000
750
500
250
0
No
Yes
RainToday