**Dataset Description:**

The dataset contains about 10 years of daily weather observations from many locations across Australia. The dataset has collected from Kaggle. Observations were drawn from numerous weather stations. The daily observations are available from <http://www.bom.gov.au/climate/data>

The dataset has a total 23 columns and there are 145460 observations. The dataset is mainly built with the intention of predicting whether based on the data available from 22 columns whether it will rain in the next day. RainTomorrow is the target variable to predict. This column is Yes if there was rain of 1 mm or more the next day or else No. Other attributes which the dataset contains are:

1. Date: The date in which the observation is recorded.
2. Location: the common name of the location of the weather station.
3. MinTemp: the minimum temperature in degree celsius.
4. MaxTemp: the maximum temperature in degree celsius.
5. Rainfall: the amount of rainfall recorded for the day in mm.
6. Evaporation: The so-called Class A pan evaporation(mm) in the 24 hours to 9 am.
7. Sunshine: The number of hours of bright sunshine in the day.
8. WindGustDir: the direction of the strongest wind gust in the 24 hours to midnight.
9. WindGustSpeed: the speed (km/h) of the strongest wind gust in the 24 hours to midnight.
10. WindDir9am: Direction of the wind at 9am.
11. WindDir3pm: Direction of the wind at 3pm.
12. WindSpeed9am: wind speed (km/hr) averaged over 10 minutes prior to 9am.
13. WindSpeed3pm: wind speed (km/hr) averaged over 10 minutes prior to 3am.
14. Humidity9am: humidity (present) at 9am.
15. Humidity3pm: humidity (present) at 3pm.
16. Pressure9am: Atmospheric pressure (hpa) reduced to mean sea level at 9am.
17. Pressure3pm: Atmospheric pressure (hpa) reduced to mean sea level at 3pm.
18. Cloud9am: Fraction of sky obscured by cloud at 9am.
19. Cloud3pm: Fraction of sky obscured by cloud at 3pm.
20. Temp9am: Temperature (degrees C) at 9am.
21. Temp3pm: Temperature (degrees C) at 3pm.
22. RainToday: Boolean: 1 if precipitation (mm) in the 24 hours to 9am exceed 1mm, otherwise 0.

**Initial plan for data exploration**

The first step would be to get some information about the dataset which will include how many missing values are there for each column. After that steps will be taken regarding the missing values i.e. whether to delete the attribute or the row or fill the missing values with some appropriate value. The next step would be to visualise the data and try to get some correlation among the various predictors and the target variable. Then Feature engineering and standardising the various columns of the data would be performed for the categorical variables. After which 3 hypothesis test will be performed.