

Group Members: Anna Jacobson, Govind Kishan, Akil Rammohan, Alan Rocque, and Eileen McConville

Question: Can we predict whether or not it will rain on a given day based on the temperature, humidity, wind speed, cloud cover, and pressure?

Variables: temperature (Celsius), humidity (% or g/kg), wind speed (unknown units, maybe km/h), cloud cover (%), pressure (millibars), rain (categorical)

Methods:

- **Data Exploration:** The only necessary data engineering for this project would be to convert rain to a categorical variable by using one hot encoding
- **Feature Engineering:** We might create additional features that contain interactions between temperature, humidity, and other factors.
- **Modeling:** We will test several machine learning models such as logistic regression, decision trees and other methods. and choose the best model accordingly based on metrics such as accuracy, precision, and F1 score.

Code to read data:

We got this data set from Kaggle Hub. Here is the url:

<https://www.kaggle.com/datasets/zeeshier/weather-forecast-dataset>. This data measures Temperature, Humidity, Wind Speed, Cloud Cover, Pressure, Rain as its variables.

```
import pandas as pd

df = pd.read_csv('weather_forecast_data.csv')
df
```

	Temperature	Humidity	Wind_Speed	Cloud_Cover	Pressure	Rain
0	23.720338	89.592641	7.335604	50.501694	1032.378759	rain
1	27.879734	46.489704	5.952484	4.990053	992.614190	no rain
2	25.069084	83.072843	1.371992	14.855784	1007.231620	no rain
3	23.622080	74.367758	7.050551	67.255282	982.632013	rain
4	20.591370	96.858822	4.643921	47.676444	980.825142	no rain
...
2495	21.791602	45.270902	11.807192	55.044682	1017.686181	no rain
2496	27.558479	46.481744	10.884915	39.715133	1008.590961	no rain
2497	28.108274	43.817178	2.897128	75.842952	999.119187	no rain
2498	14.789275	57.908105	2.374717	2.378743	1046.501875	no rain
2499	26.554356	97.101517	18.563084	81.357508	1001.729176	no rain

2500 rows x 6 columns

Here is the Data Frame with One Hot Encoding:

```
m = df.join(pd.get_dummies(df.Rain, drop_first=True))
m
```

	Temperature	Humidity	Wind_Speed	Cloud_Cover	Pressure	Rain	rain
0	23.720338	89.592641	7.335604	50.501694	1032.378759	rain	1
1	27.879734	46.489704	5.952484	4.990053	992.614190	no rain	0
2	25.069084	83.072843	1.371992	14.855784	1007.231620	no rain	0
3	23.622080	74.367758	7.050551	67.255282	982.632013	rain	1
4	20.591370	96.858822	4.643921	47.676444	980.825142	no rain	0
...
2495	21.791602	45.270902	11.807192	55.044682	1017.686181	no rain	0
2496	27.558479	46.481744	10.884915	39.715133	1008.590961	no rain	0
2497	28.108274	43.817178	2.897128	75.842952	999.119187	no rain	0
2498	14.789275	57.908105	2.374717	2.378743	1046.501875	no rain	0
2499	26.554356	97.101517	18.563084	81.357508	1001.729176	no rain	0

2500 rows x 7 columns