Rain Prediction

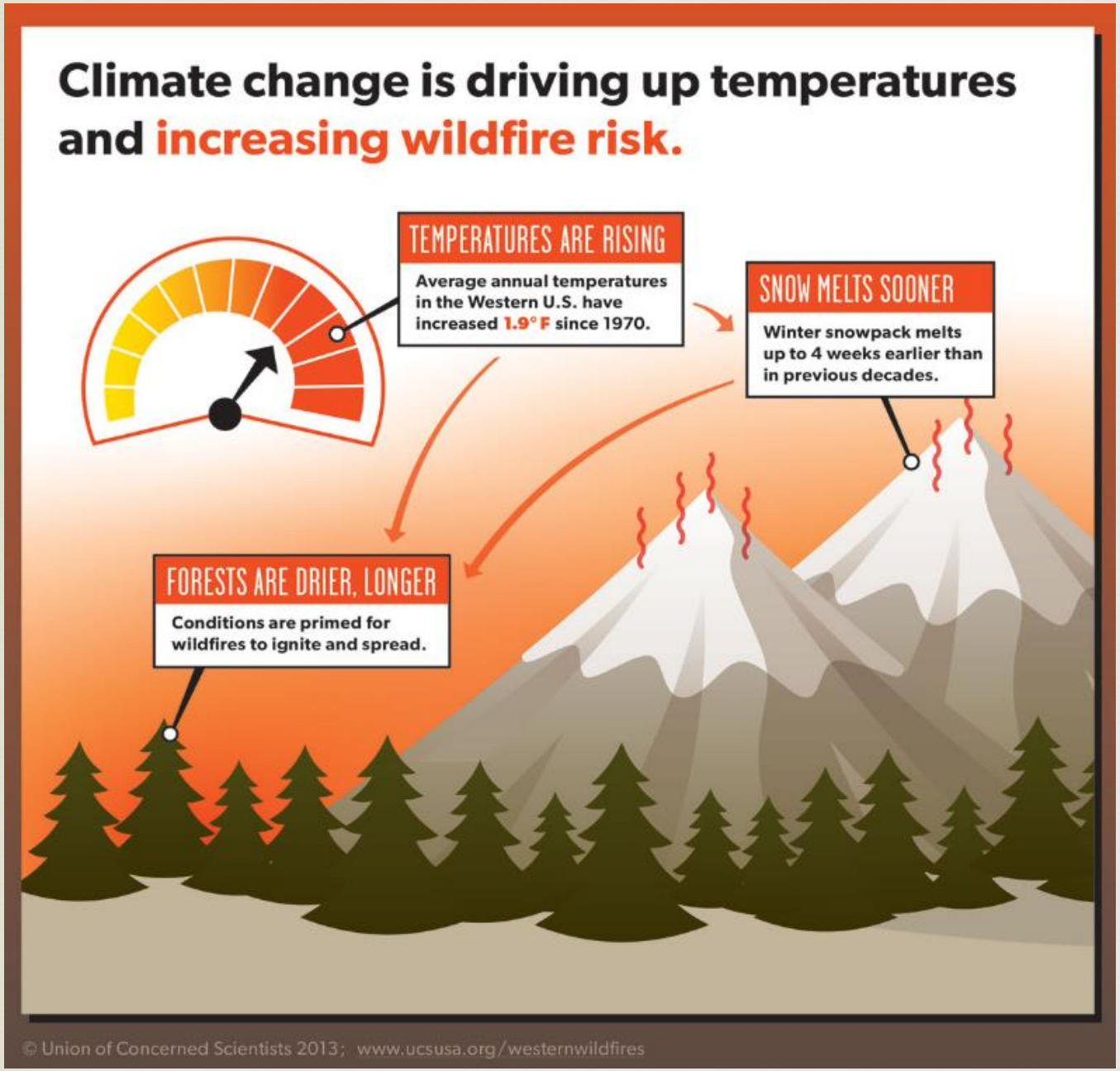
Classification
Project Presentation

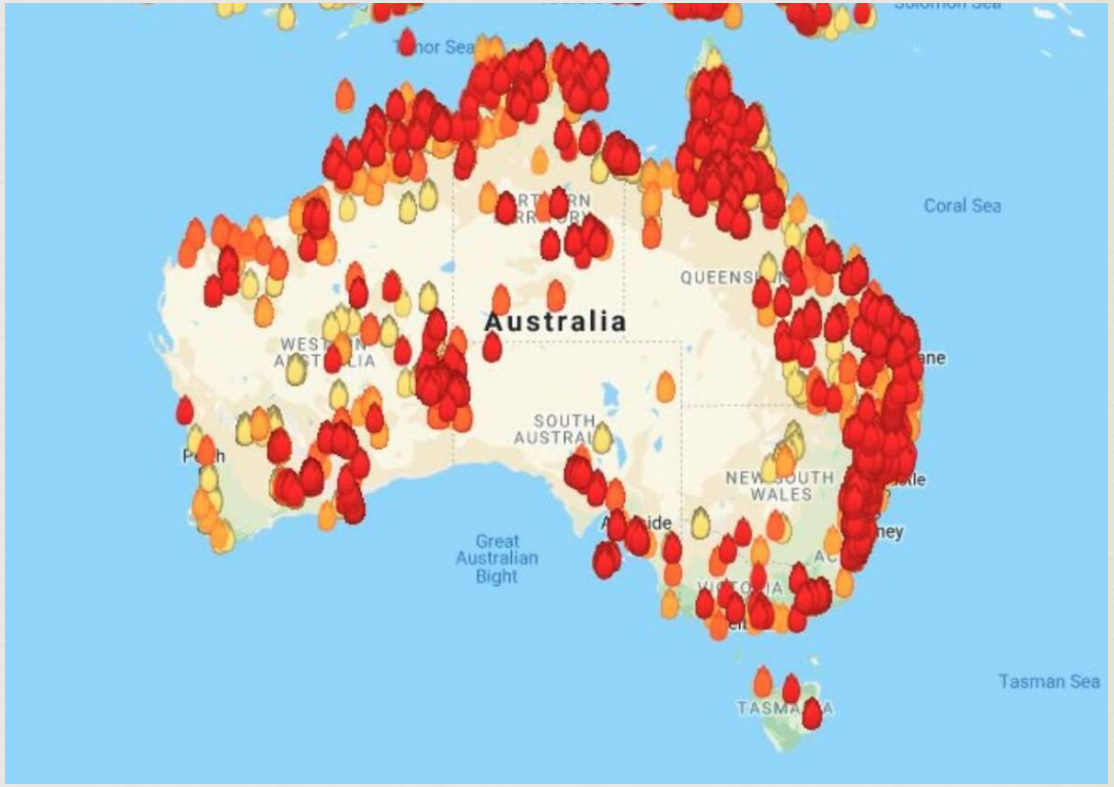
Meshal Alamr **Norah Alkhalifah**

31 October 2021

Backstory

Backstory


Climate change is driving up temperatures and increasing wildfire risk.
TEMPERATURES ARE RISING
Average annual temperatures in the Western U.S. have increased 1.9° F since 1970.
SNOW MELTS SOONER
Winter snowpack melts up to 4 weeks earlier than in previous decades.
FORESTS ARE DRIER, LONGER
Conditions are primed for wildfires to ignite and spread.
© Union of Concerned Scientists 2013; www.ucsusa.org/westernwildfires


Australia
Coral Sea
Great Australian Bight
Tasman Sea

Goal

- Create a model that will predict will it rain tomorrow or not
- Help firefighters fight forest fires

Dataset

Dataset

145,460 Rows

- Date
- Location
- Temperature
- Rain

23 Columns

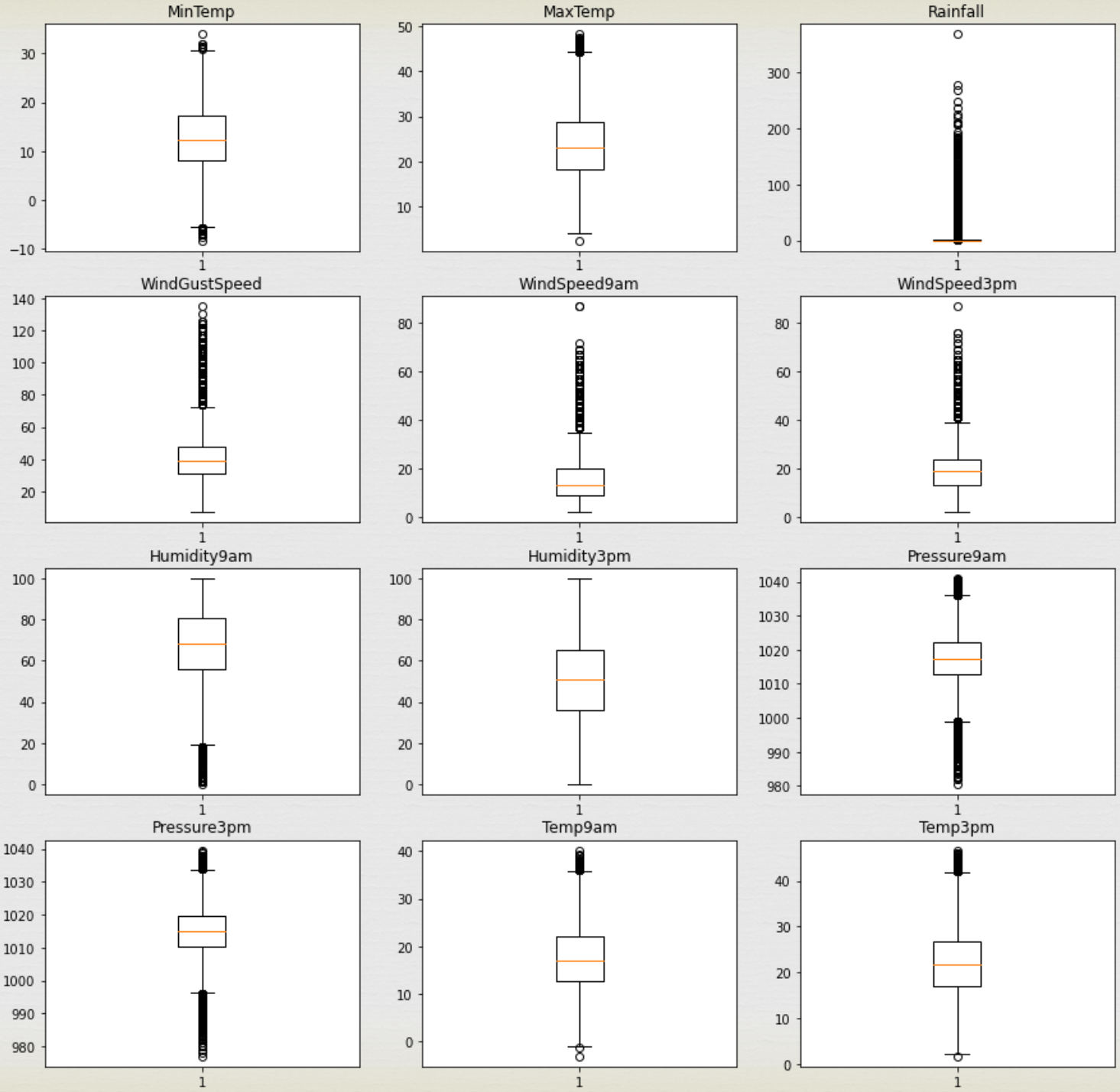
- Humidity
- Wind
- Cloud
- Pressure

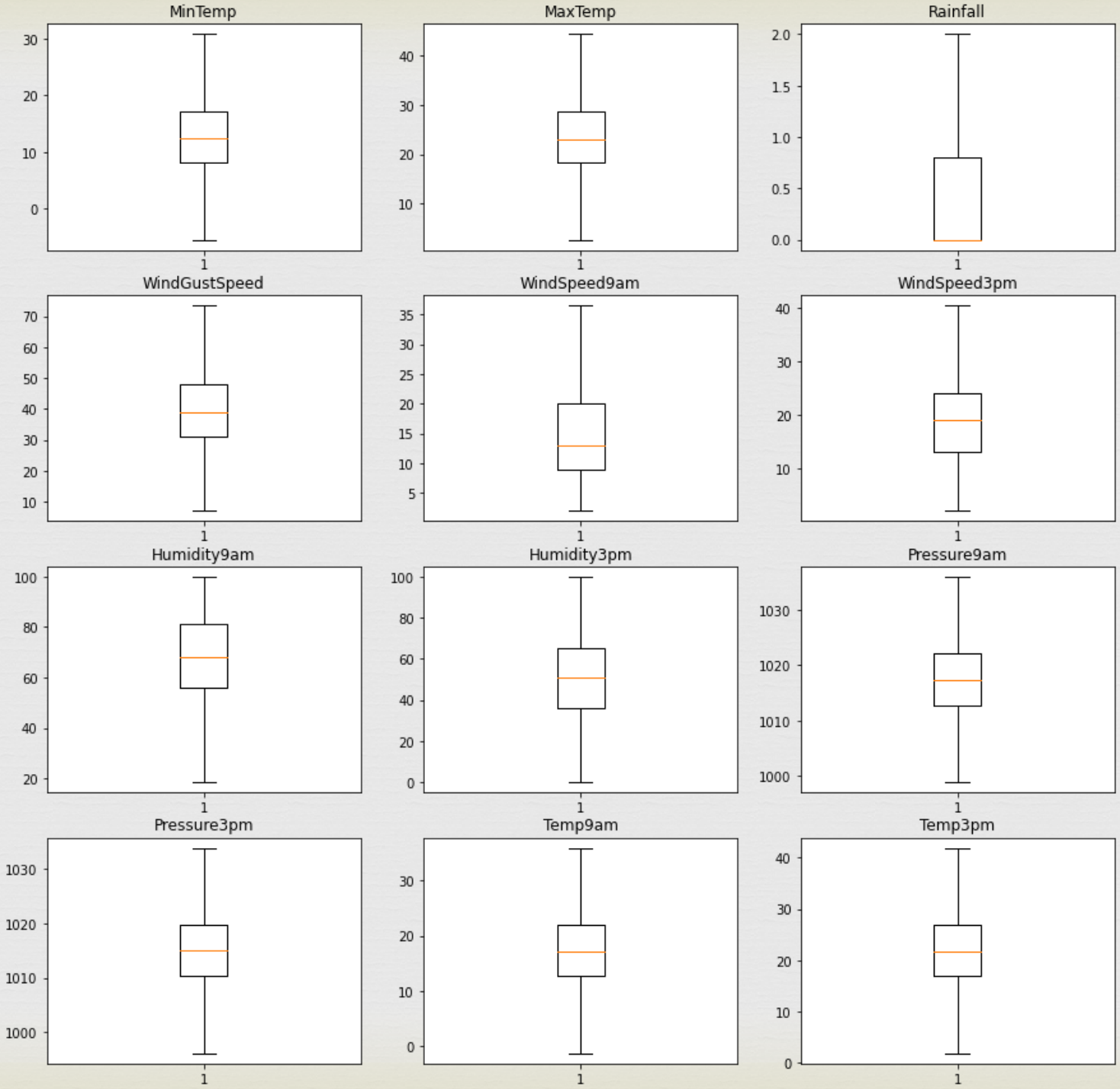
EDA

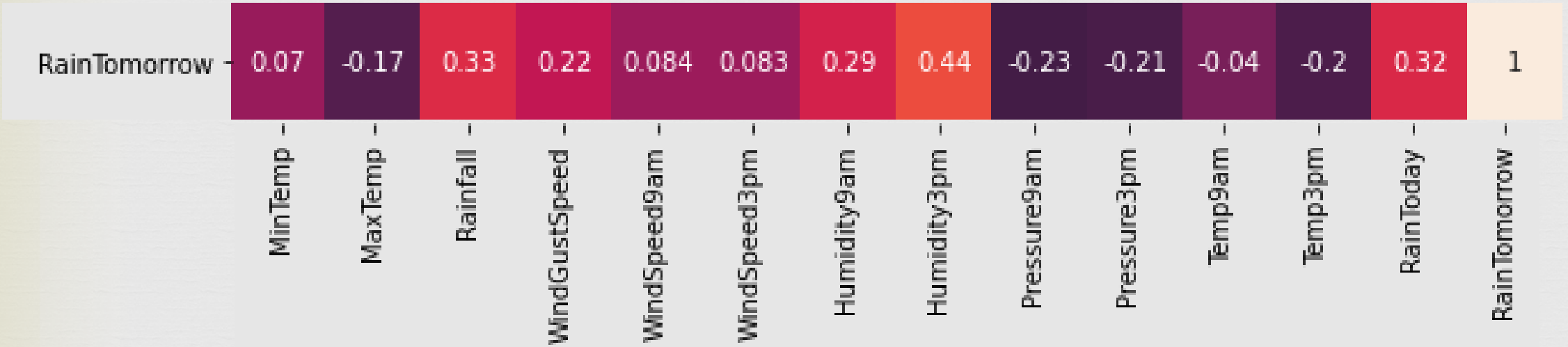
Null Values

Sunshine	69835
Evaporation	62790
Cloud3pm	59358
Cloud9am	55888
Pressure9am	15065
Pressure3pm	15028
WindDir9am	10566
WindGustDir	10326
WindGustSpeed	10263
Humidity3pm	4507
WindDir3pm	4228
Temp3pm	3609
RainTomorrow	3267
Rainfall	3261
RainToday	3261
WindSpeed3pm	3062
Humidity9am	2654
Temp9am	1767
WindSpeed9am	1767
MinTemp	1485
MaxTemp	1261
Location	0
Date	0

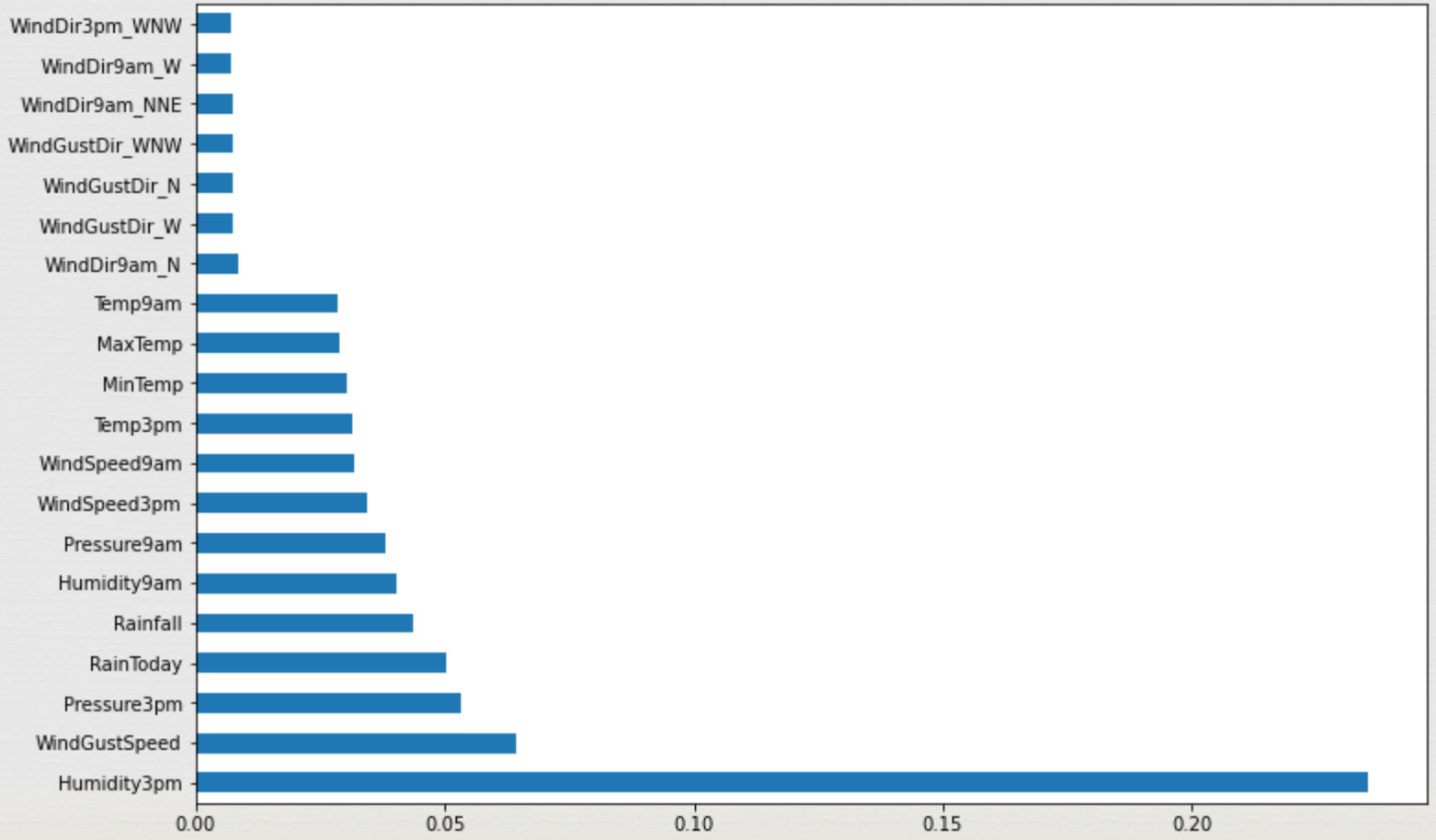
- Dropping all null:
  - Lose 89,040 rows ~ 60%
- Drop top 4 columns then drop nulls:
  - Lose 32,535 rows ~ 20%


MinTemp
MaxTemp
Rainfall
WindGustSpeed
WindSpeed9am
WindSpeed3pm
Humidity9am
Humidity3pm
Pressure9am
Pressure3pm
Temp9am
Temp3pm

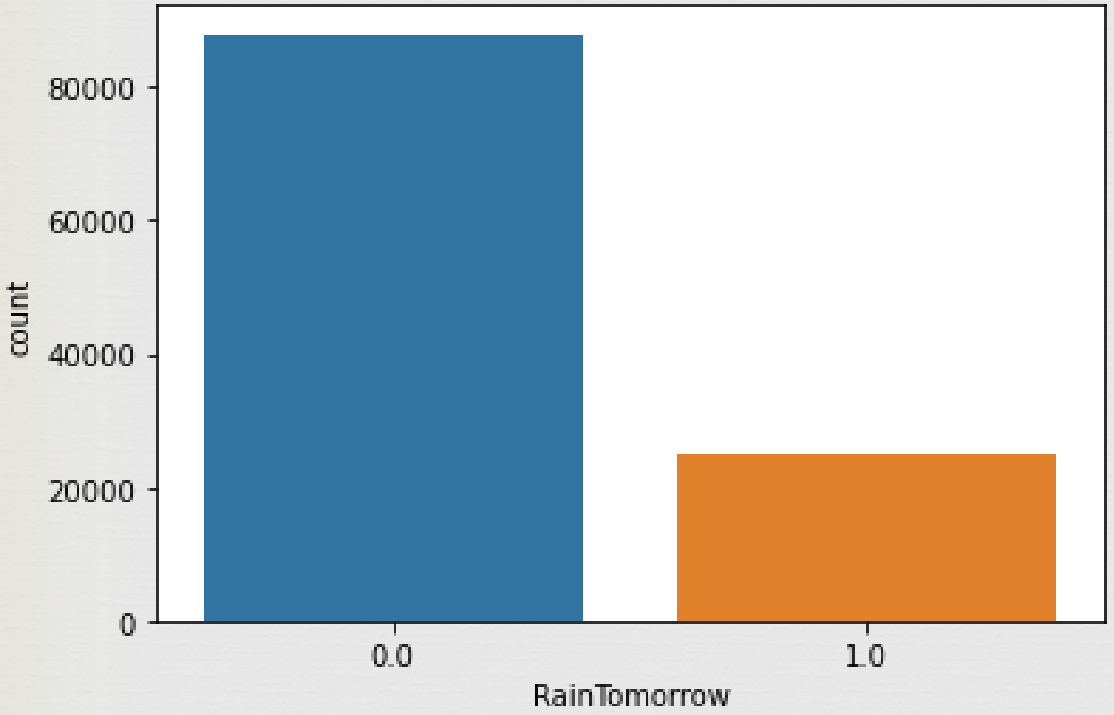

MinTemp
MaxTemp
Rainfall
WindGustSpeed
WindSpeed9am
WindSpeed3pm
Humidity9am
Humidity3pm
Pressure9am
Pressure3pm
Temp9am
Temp3pm


RainTomorrow
0.07
-0.17
0.33
0.22
0.084
0.083
0.29
0.44
-0.23
-0.21
-0.04
-0.2
0.32
1
MinTemp
MaxTemp
Rainfall
WindGustSpeed
WindSpeed9am
WindSpeed3pm
Humidity9am
Humidity3pm
Pressure9am
Pressure3pm
Temp9am
Temp3pm
RainToday
RainTomorrow

Feature Importance

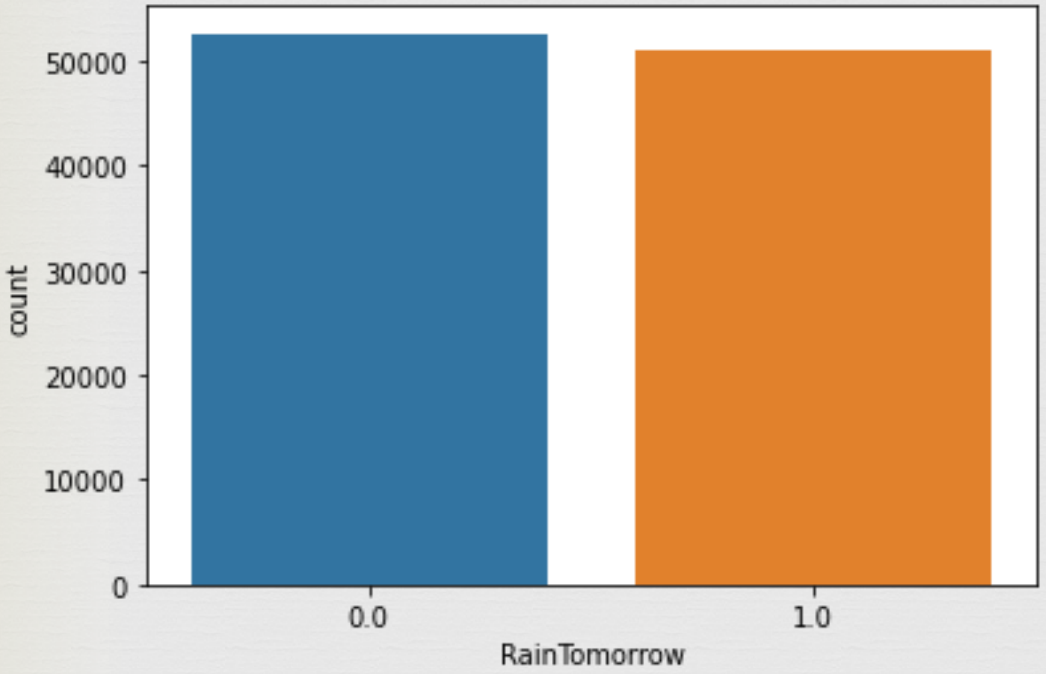


Imbalance

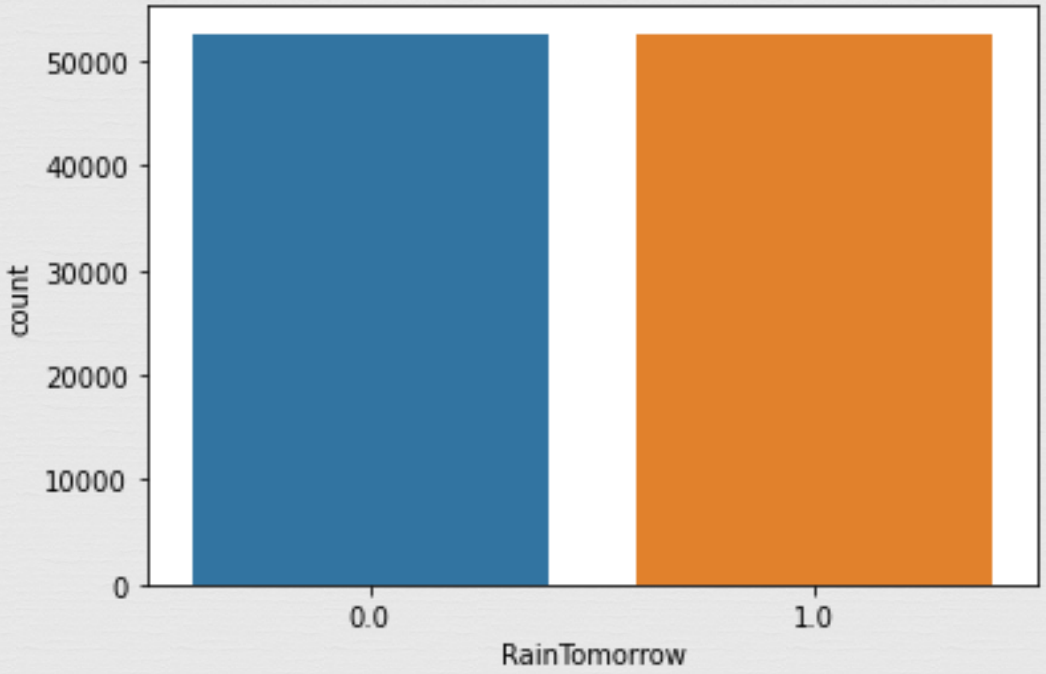


- SMOTE
- ADASYN
- Random Oversampling

Imbalance



ADASYN



SMOTE / Random Oversampling

Results and Experiments

Experiments

- Models:
  - Random Forest
  - Logistic Regression
  - XGBoost
  - CatBoost
- Sampling:
  - Imbalanced
  - SMOTE
  - ADASYN
  - Random Oversampling

16 Experiments

Results

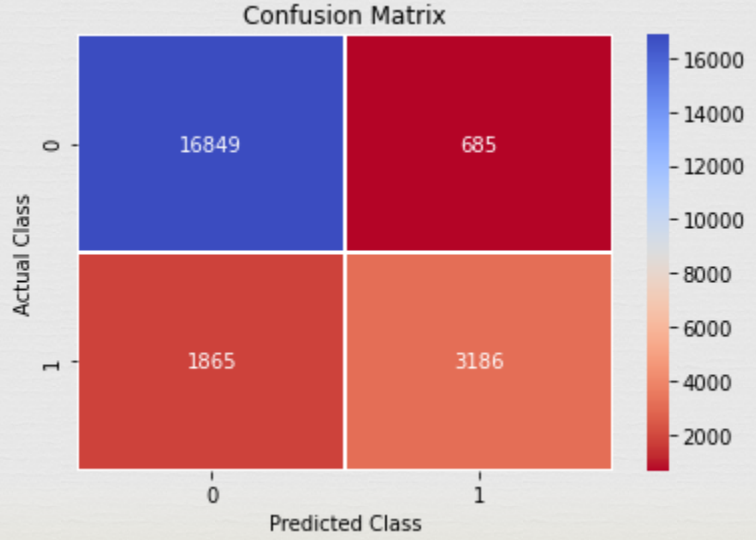
ADASYN

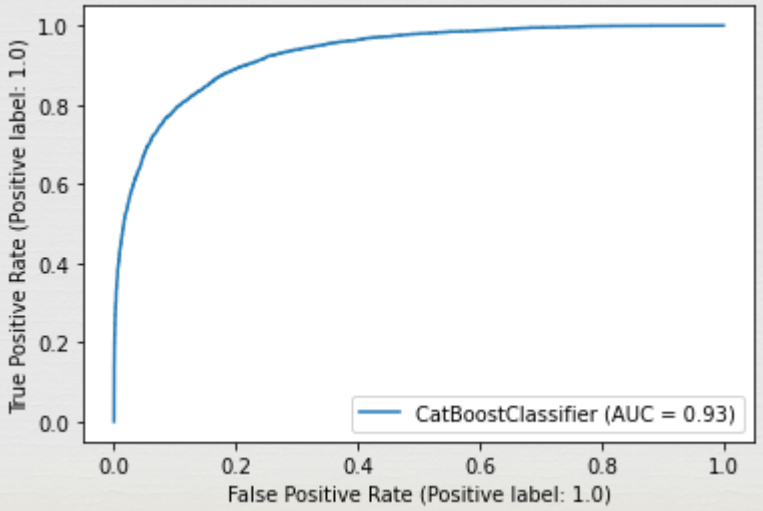
Model / Metric	Precision	Recall	F1	Accuracy
Random Forest	0.63	0.67	0.65	0.84
Logistic Regression	0.52	0.77	0.62	0.79
XGBoost	0.72	0.55	0.62	0.85
CatBoost	0.73	0.55	0.63	0.86

Final Model

Final Model - ADASYN:

Model / Metric	Precision	Recall	F1	Accuracy
CatBoost	0.82	0.63	0.71	0.89





Other Experiments

- Label-Encoding VS One-Hot-Encoding
- No-Scaling VS Min-Max Scaling
- Feature Engineering (+5 features)

12 Experiments

Recommendations

- Consider other scaling methods
- Consider Filling Nulls:
  - Another dataset / Scraped dataset
  - Random Sample Imputation
  - Feature Engineering

App Demo