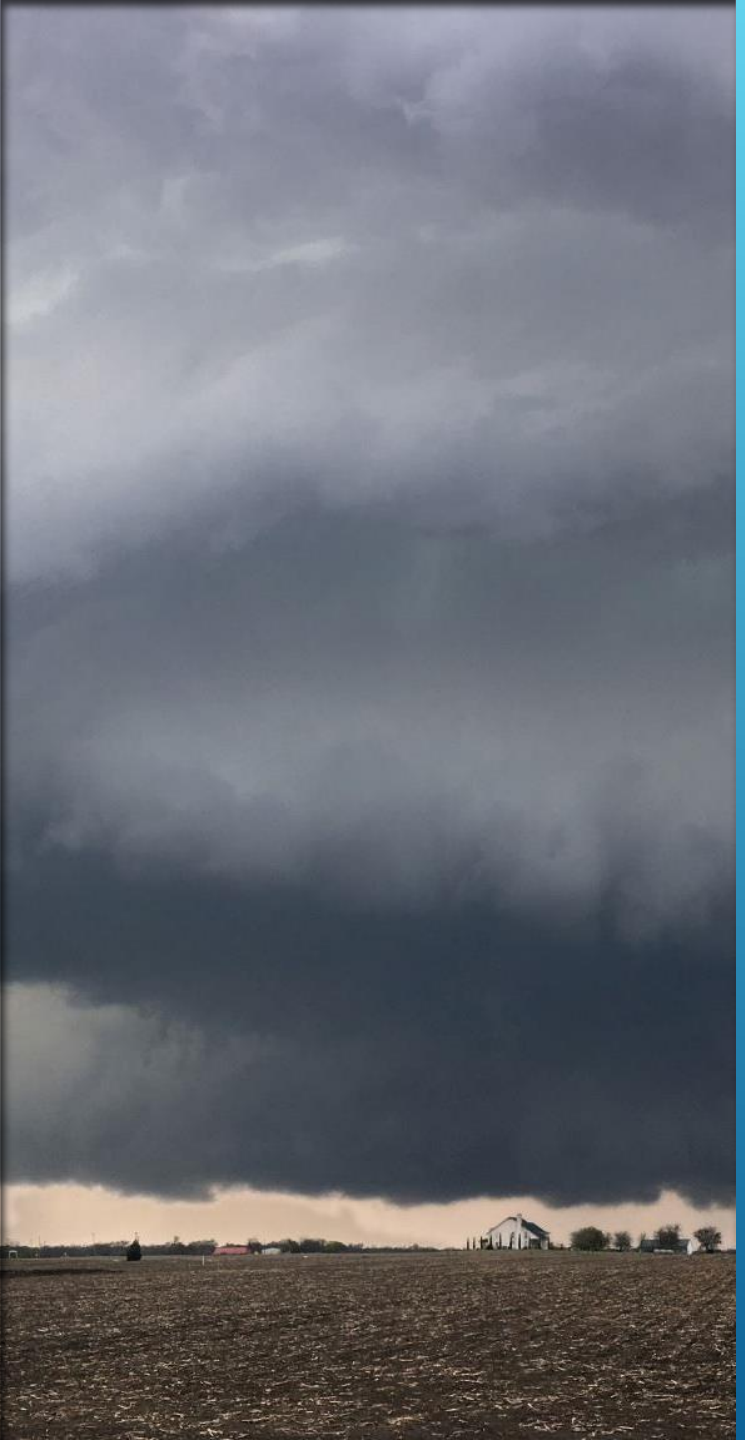
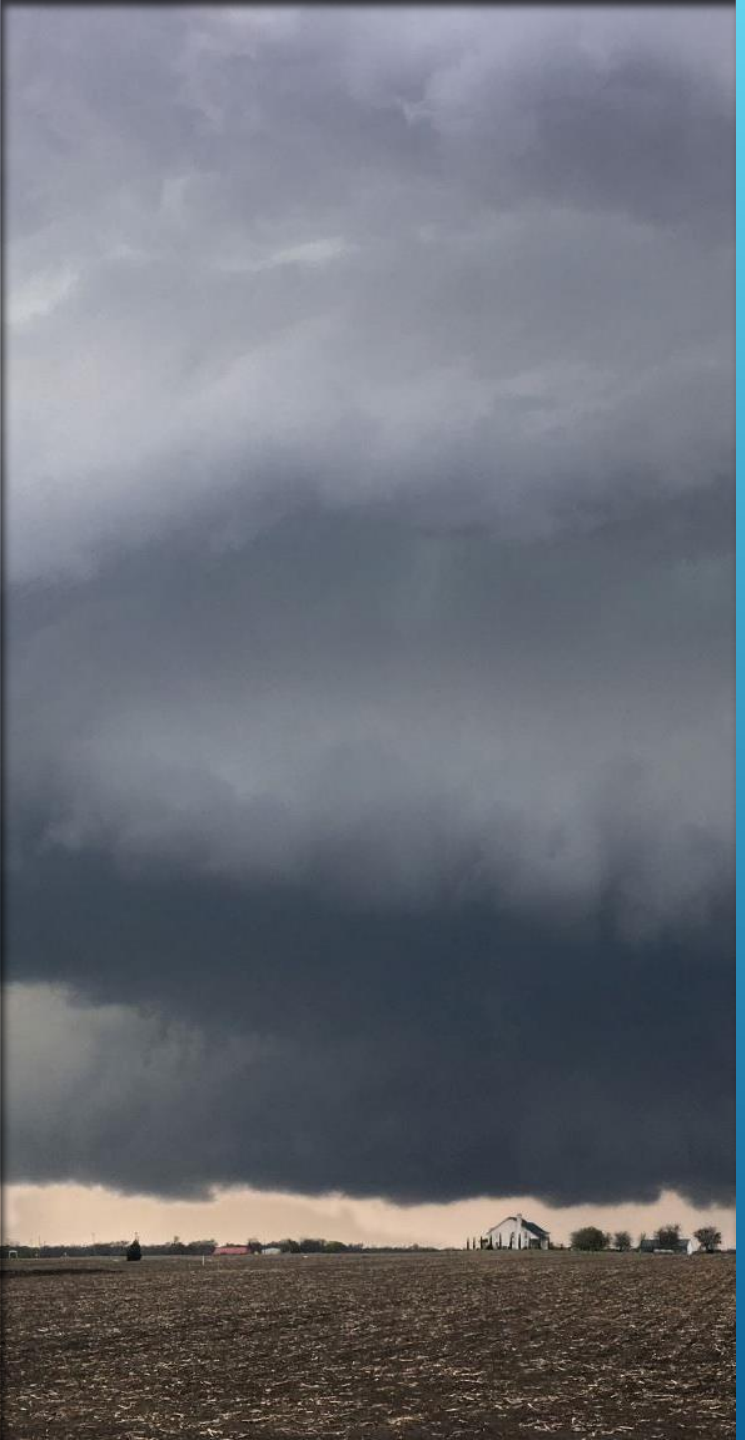
PRECIPITATION IN LONDON

Author: Mathew Thomas



PURPOSE

- Precipitation - any liquid or frozen water that forms in the atmosphere and falls back to the earth
- Using ML models, to try and predict this
- Provide crucial information to a wide variety of groups and entities: Agriculture sector, Meteorologists, Energy companies, Urban planners

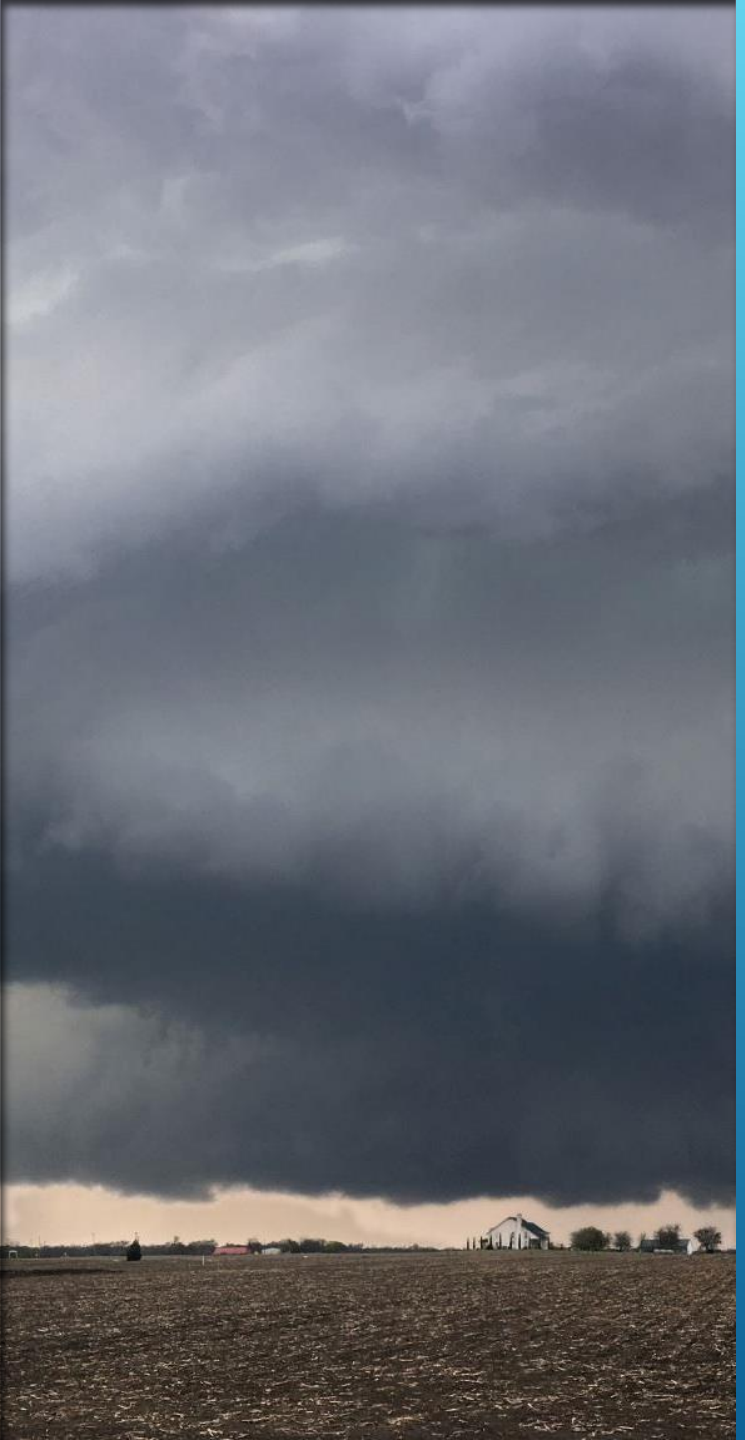


DATASET

- Aggregate of datasets compiled from ECA & D (European Climate Assessment & Dataset)
- Most measurements taken from station near Heathrow Airport
- Jan 1st 1979 to Dec 31st 2022
- File size 1.2 MB

Data dictionary:

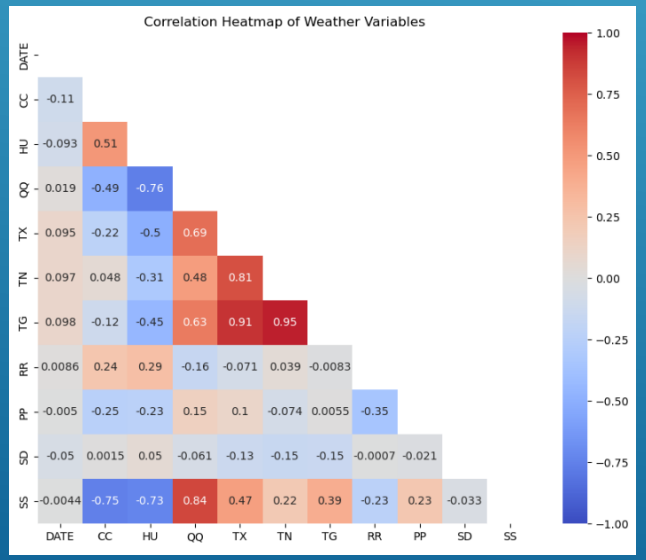
- `DATE` : recorded date of measurement
- `CC` : Cloud Cover, measurement in oktas
- `HU` : Humidity, measurement in %
- `QQ` : Global Radiation, irradiance measurement in Watt per square meter (W/m2)
- `TX` : Temperature Maximum, maximum temperature recorded in degrees Celsius (°C)
- `TG` : Temperature Mean, mean temperature in degrees Celsius (°C)
- `TN` : Temperature Minimum, minimum temperature recorded in degrees Celsius (°C)
- `RR` : Precipitation, precipitation measurement in millimeters (mm)
- `PP` : Pressure, pressure measurement in Pascals (hPa)
- `SD` : Snow Depth, depth measurement in centimeters (cm)
- `SS` : Sunshine, measurement in hours (hrs)

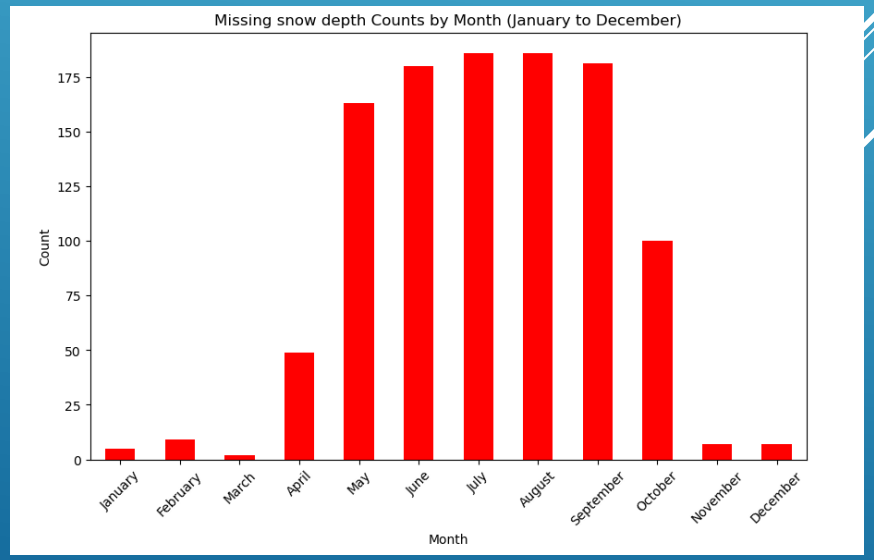


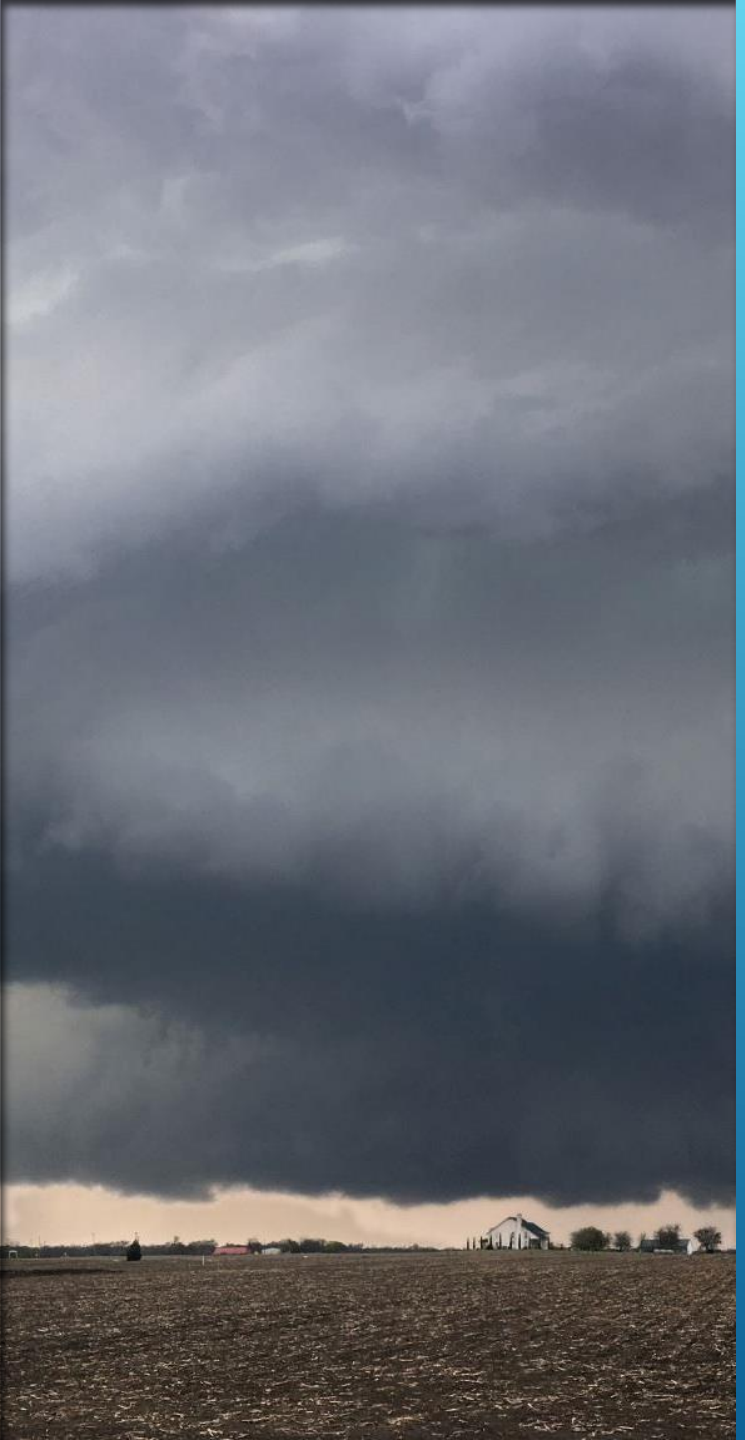
EDA

- Cleaned the dataset off any null values using statistical measures of mean & median
- Some values SD used 0
- Low correlation to RR (Precipitation)

DATE	0
CC	18
HU	57
QQ	25
TX	0
TN	0
TG	0
RR	0
PP	0
SD	1075
SS	0

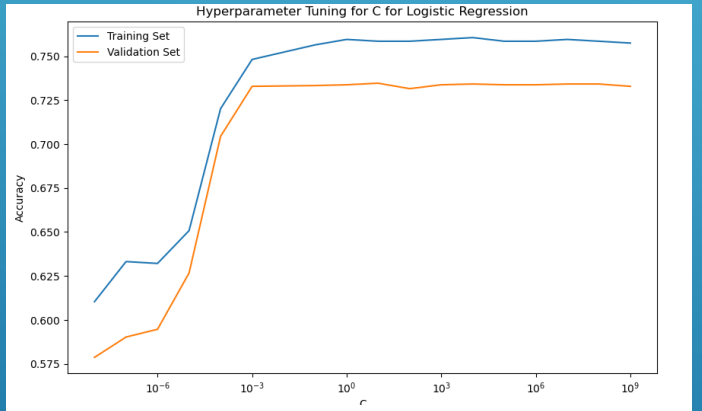






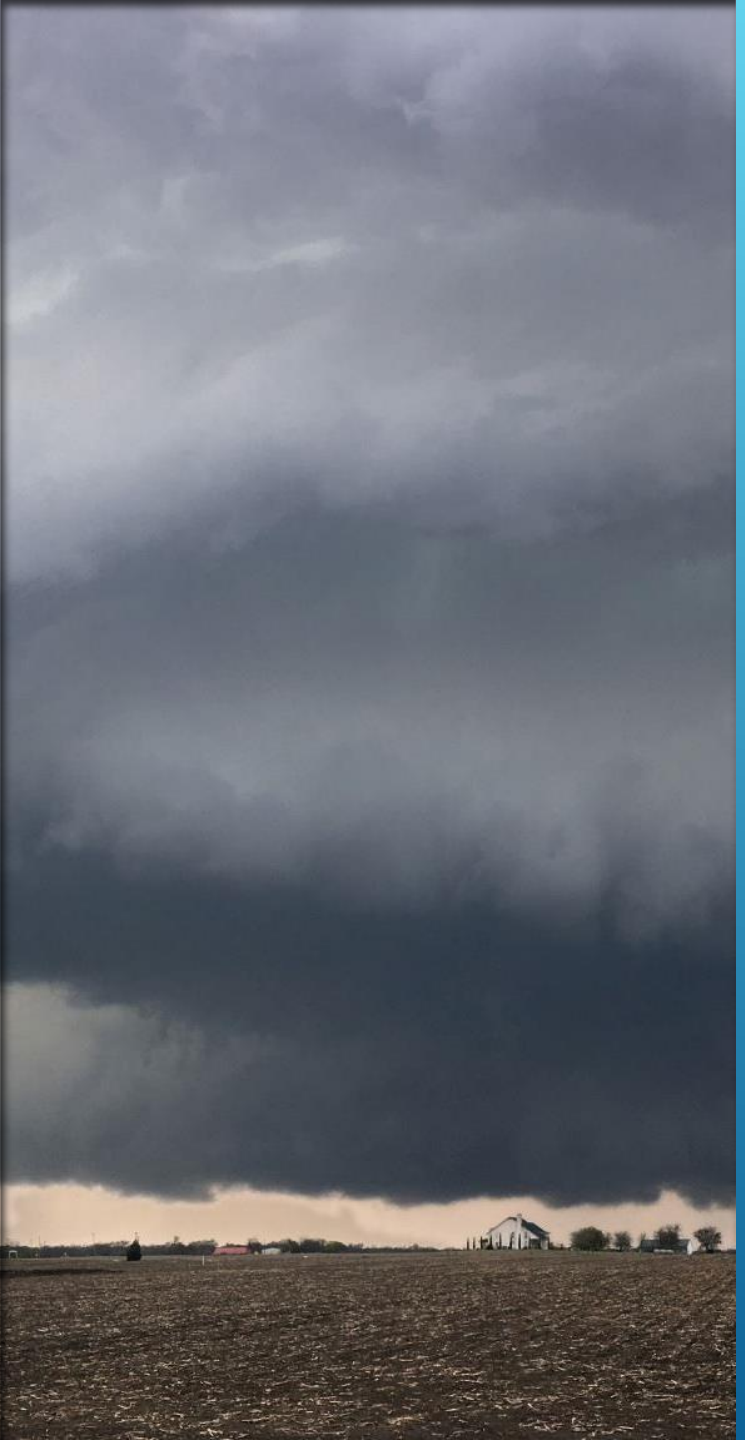
MODEL 1

- Approach it in two steps:1. Classification, 2.Regression
- Far too many multicollinearity issues



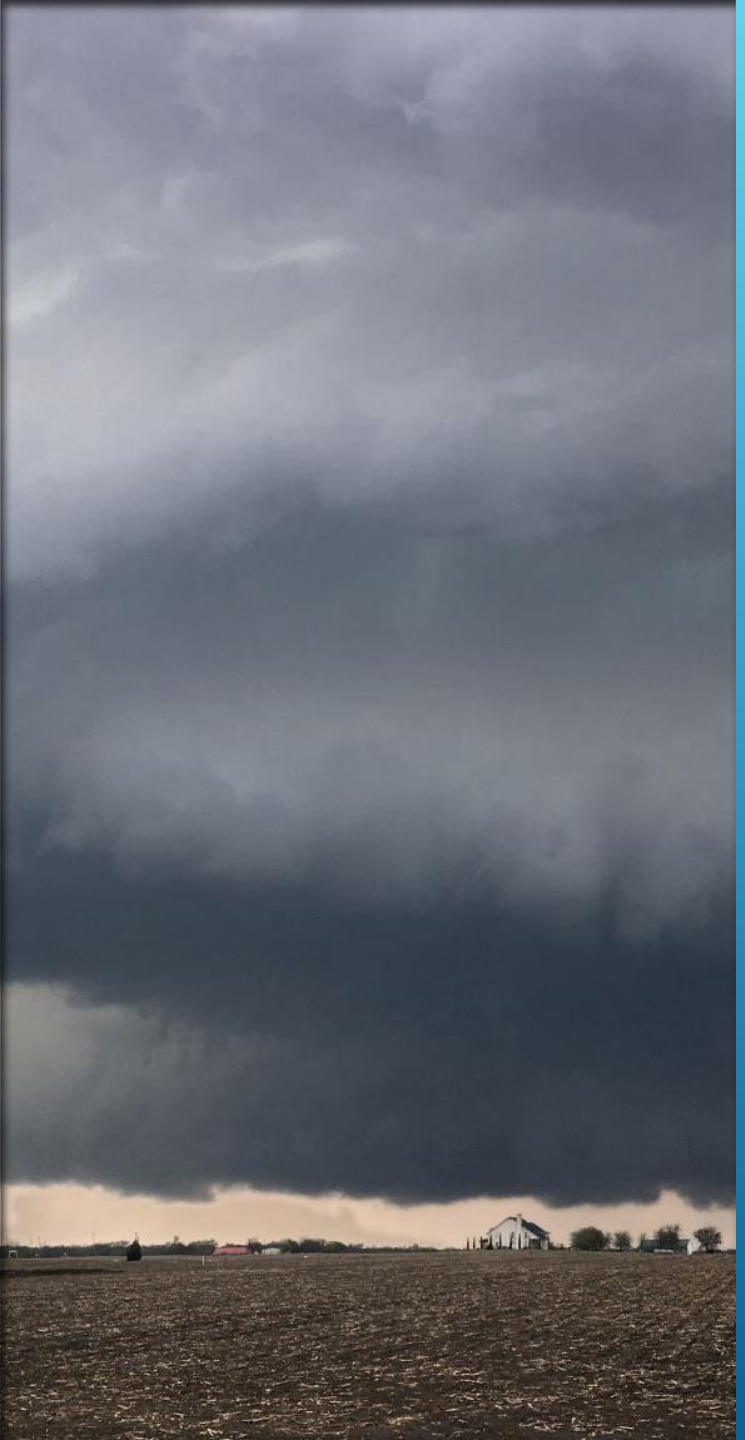
	precision	recall	f1-score	support
0	0.77	0.71	0.74	1193
1	0.70	0.76	0.73	1057
accuracy			0.73	2250
macro avg	0.73	0.73	0.73	2250
weighted avg	0.74	0.73	0.73	2250

CC	3.073403
HU	2.718286
QQ	7.487177
TX	7.151541
TG	29.429480
TN	17.225219
PP	1.140106
SD	1.027989
SS	7.812878
Year	1.079712
Month	1.275171
Day	1.001400



MODEL 2 & 3

- Model 2 – Decision Tree Regressor
- Overfitting unseen data too much, unable to generalize
- Model 3 – Random Forest Generator
- Running a range of hyperparameters



NEXT STEPS

- Deep Learning
- RNN – LSTM or GRU (faster but not as powerful)
- Predict precipitation anywhere in the world