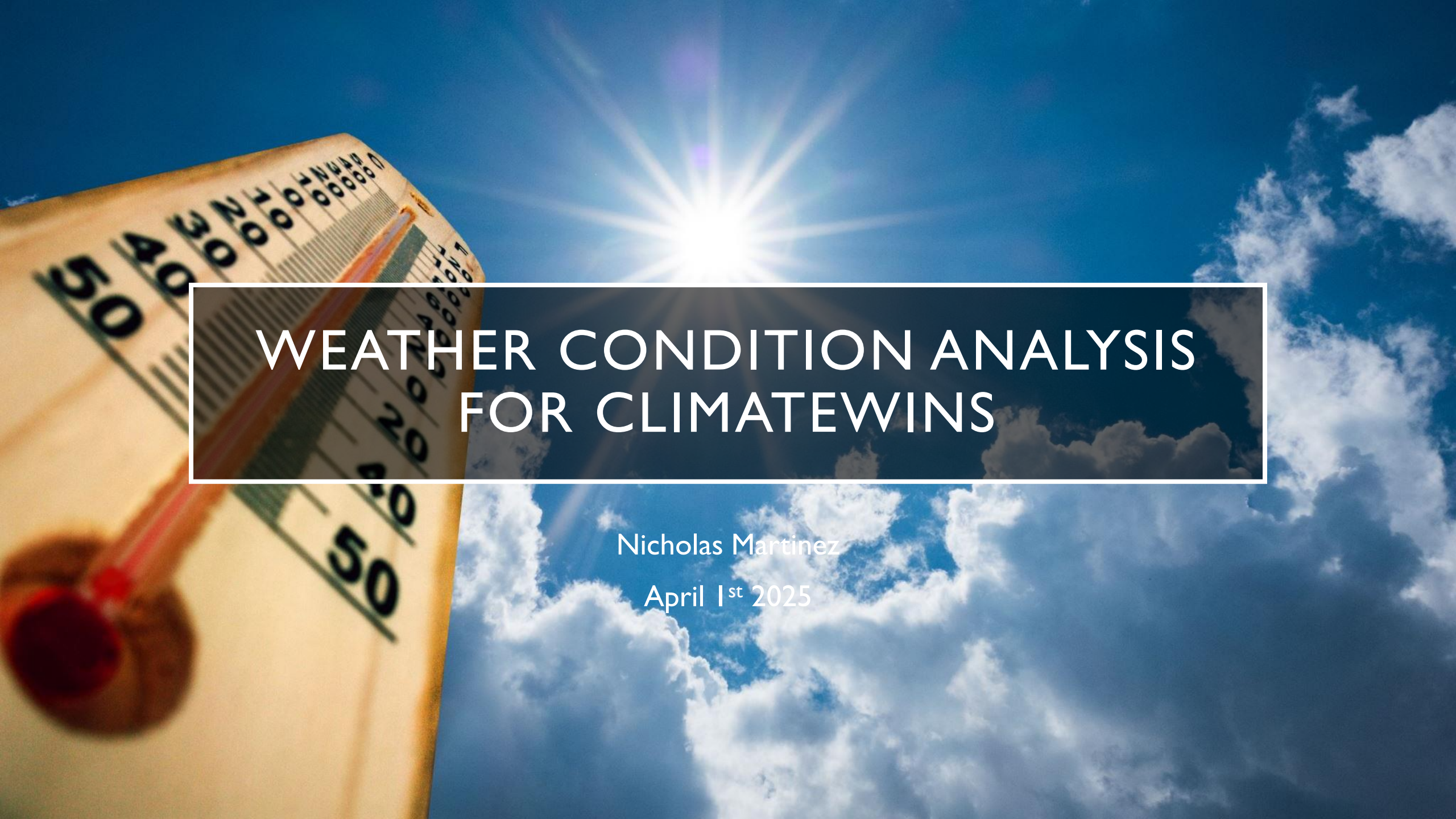


# WEATHER CONDITION ANALYSIS FOR CLIMATEWINS

Nicholas Martinez

April 1st 2025

# Introduction

Climate change is occurring at a rapid pace around the world. Consequently, increased extreme weather events are occurring. To better predict these complex events machine learning is being leveraged to sort the large amounts of data inherent to climate analysis.

# Hypotheses

Rising temperatures will correspond with increased extreme weather events.

Machine learning will be able to accurately predict extreme weather events.

Through machine learning the factors that most contribute to extreme weather events can found.

# THE DATA

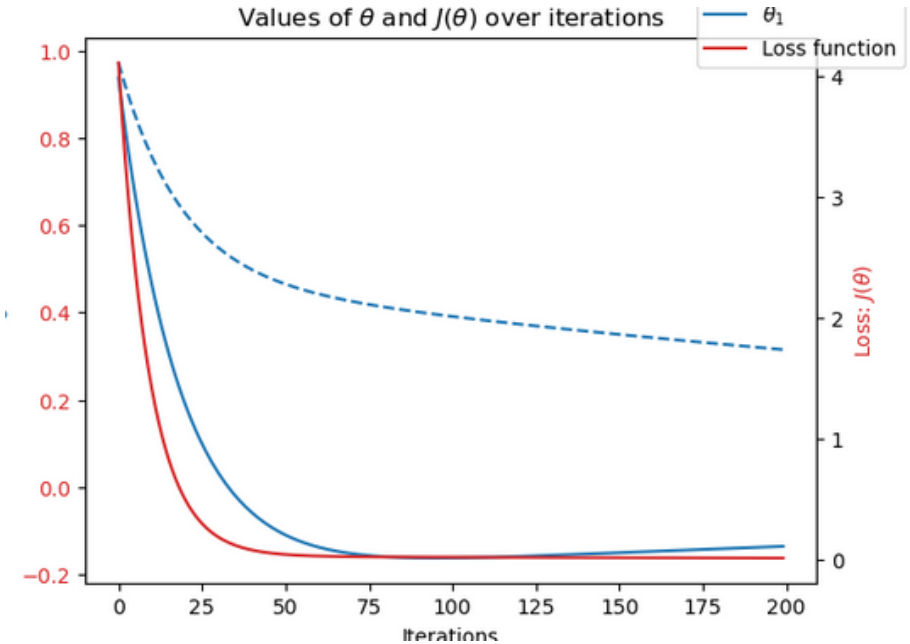
- Collected from 18 weather stations from across Europe
- Dates of collection range from the 1800's to 2022
- Collected by the European Climate Assessment & Data Set project
- Data is cleaned and has no missing or null values
- Download here

# DATA BIAS

- Collection Bias: There is data only from 18 stations, which may not represent to totality of Europe.
- Measurement Bias:  Older data points, before 1970, may have issues with measurement methods and instrument accuracy
- Temporal Bias: Recent extreme climate change spans around 20 years making the bulk of the older data nonrepresentative
- Sample Bias: The climate is a global phenomena trying to predict it through just Europe, a relatively smaller land mass, is prone to some error.

# DATA OPTIMIZATION

- The data was optimized through the Gradient Descent method.
- Method details
  - Represents the convergence of the method through the loss function
  - Parameters are adjusted to reduce loss until a stagnation occurs



# SUPERVISED MACHINE LEARNING

- All supervised methods were tested by seeing how well they predicted if the climate data reflected a pleasant day
- Compared to a data set collected by residents of the locations with the weather stations
- Pleasant day data exists for only 15 of the original 18 weather stations

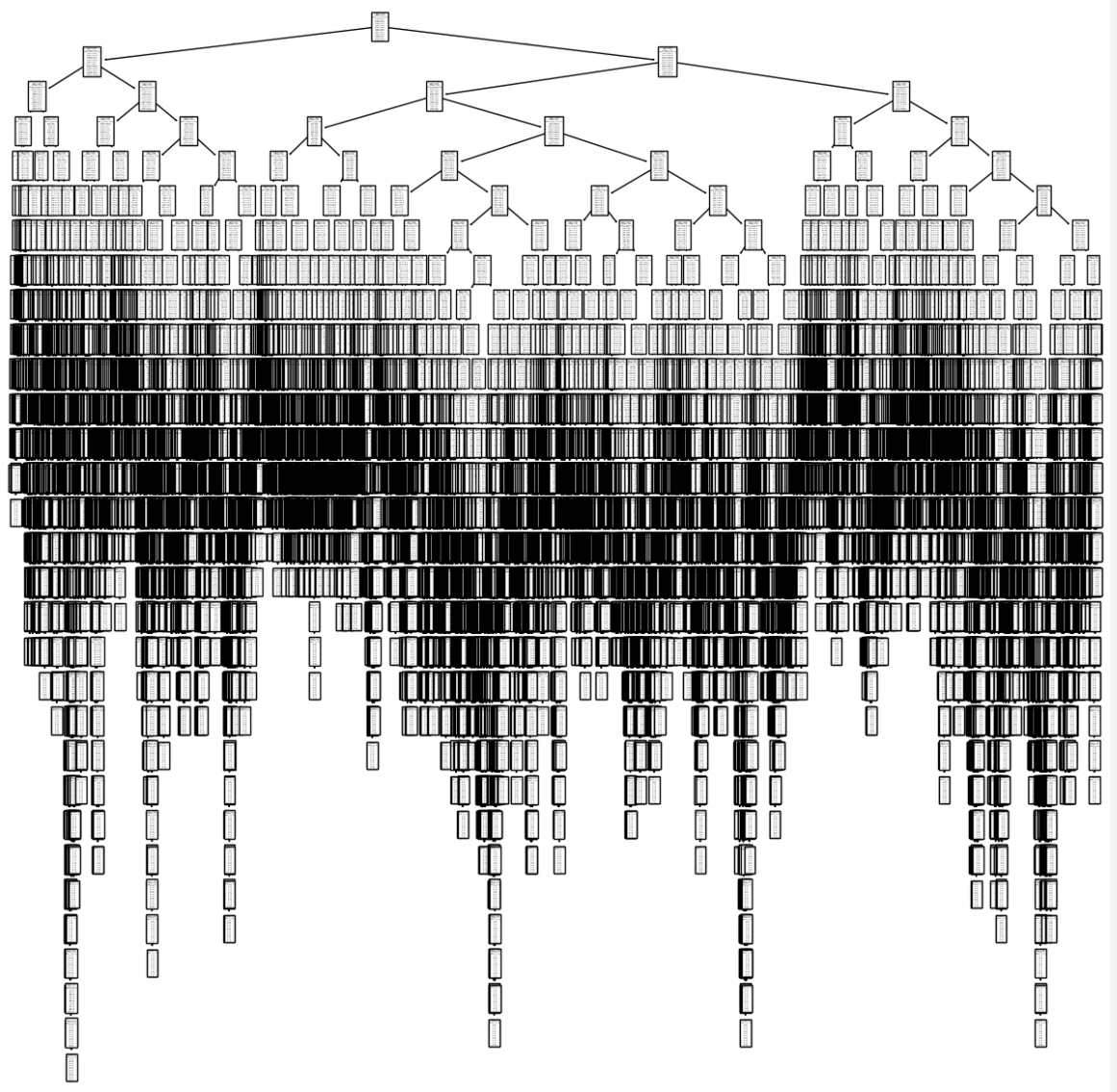
# KNN (K-NEAREST NEIGHBOR)

- The k-nearest neighbors (KNN) algorithm makes predictions by measuring the distance between new data points and nearby data points, or neighbors, then groups the new data with nearby data points based on the number of neighbors in each group.
- With a k-range value of 3 an average of 88% accuracy was obtained
- Model needs the removal of all stations that have one type of weather, in this case Sonnblick

| Weather Station | Accurate Predictions | | False Positives | False negatives | Accuracy rate |
|---|---|---|---|---|---|
| Basel | 3917 | 961 | 421 | 439 | 85% |
| Belgrade | 3252 | 1544 | 524 | 418 | 84% |
| Budapest | 3424 | 1462 | 476 | 376 | 85% |
| Debilt | 4320 | 723 | 317 | 378 | 88% |
| Dusseldorf | 4164 | 810 | 343 | 421 | 87% |
| Heathrow | 4138 | 744 | 432 | 424 | 85% |
| Kassel | 4563 | 614 | 252 | 309 | 90% |
| Ljubljana | 3740 | 1180 | 455 | 363 | 86% |
| Maastricht | 4253 | 824 | 309 | 352 | 88% |
| Madrid | 2750 | 2261 | 418 | 309 | 87% |
| Munchenb | 4237 | 792 | 309 | 400 | 88% |
| Oslo | 4637 | 512 | 242 | 347 | 90% |
| Sonnblick | 5738 | 0 | 0 | 0 | 100% |
| Stockholm | 4483 | 607 | 283 | 365 | 89% |
| Valentia | 5404 | 74 | 50 | 202 | 96% |
| | | | | Average | 88% |

# DECISION TREE

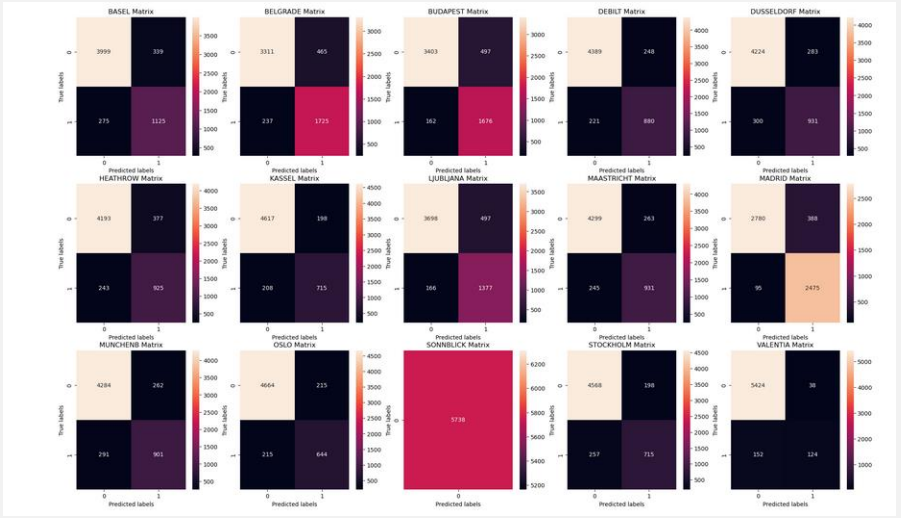
- The decision tree method starts with a root of all the data, then using various binary splits to create branches and leaves to arrive at a conclusion
- Like how a human mind works splits a larger decision into its relevant components
- Heavy overfitting and low accuracy occurred with a training accuracy of 46% and test accuracy of 45.7%



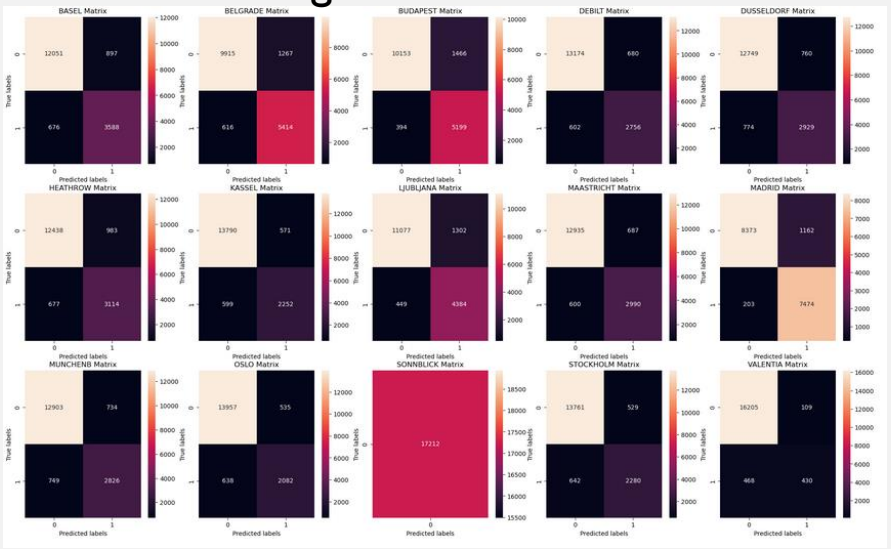
# ANN (ARTIFICIAL NEURAL NETWORK)

- ANN seeks to reproduce the way the human brain processes information. The method consists of up to three hidden layers with a set of nodes for each layer. The method can often handle more complex data were is unpredictable to the point that the same inputs can result in different outcomes.
- Number of hidden layers and nodes were tested for greatest accuracy, best model was a dual layer with 50 nodes each
- The best achieved accuracy for the model was 52% for training data and a 50% for the test data

Test confusion matrix




Training confusion matrix



# SUMMARIZATION

- KNN is the current best model when compared to the simplistic binary of good or bad weather, but ANN maybe to better choice for more complicated predictions
- Rising temperatures will correspond with increased extreme weather events.
- Machine learning will be able to accurately predict extreme weather events.
- Through machine learning the factors that most contribute to extreme weather events can found.

# NEXT STEPS AND FUTURE ANALYSIS

Start the process of more complex machine learning with unsupervised machine learning

Add in more relevant data such as cloud cover and radiation

Continue model refinement

Predict the weather

# THANK YOU

- For questions or discussion please email me at mnicholas63@yahoo.com
- See my GitHub page for the detailed analysis
- Please see this link for recorded presentation