



PREDICTING RAIN PATTERNS IN AUSTRALIA

JUSTIN GRISANTI

OBJECTIVES – 5 STEP MODEL



Obtain a Business Understanding



Understand the Data



Prepare Data for Modeling



Create Models



Generate Results

BUSINESS UNDERSTANDING

- Stakeholders:
 - Australian Bureau of Meteorology
 - Citizens of Australia
- Business Problem
 - Bureau of Meteorology historically underpredicts rainy days each year
- Goals/Value-add
 - Help the Bureau better inform the citizens of Australia of whether it will rain tomorrow
 - Provide a model to the Bureau of Meteorology that can be used in conjunction with their own findings to reduce the amount of underpredictions

DATA UNDERSTANDING

- Various Weather Data
 - Dates from 2008 - 2017
 - Location
 - Wind Patterns
 - Sunshine Data
 - Cloud Data
 - Humidity
 - Temperature
 - Rainfall



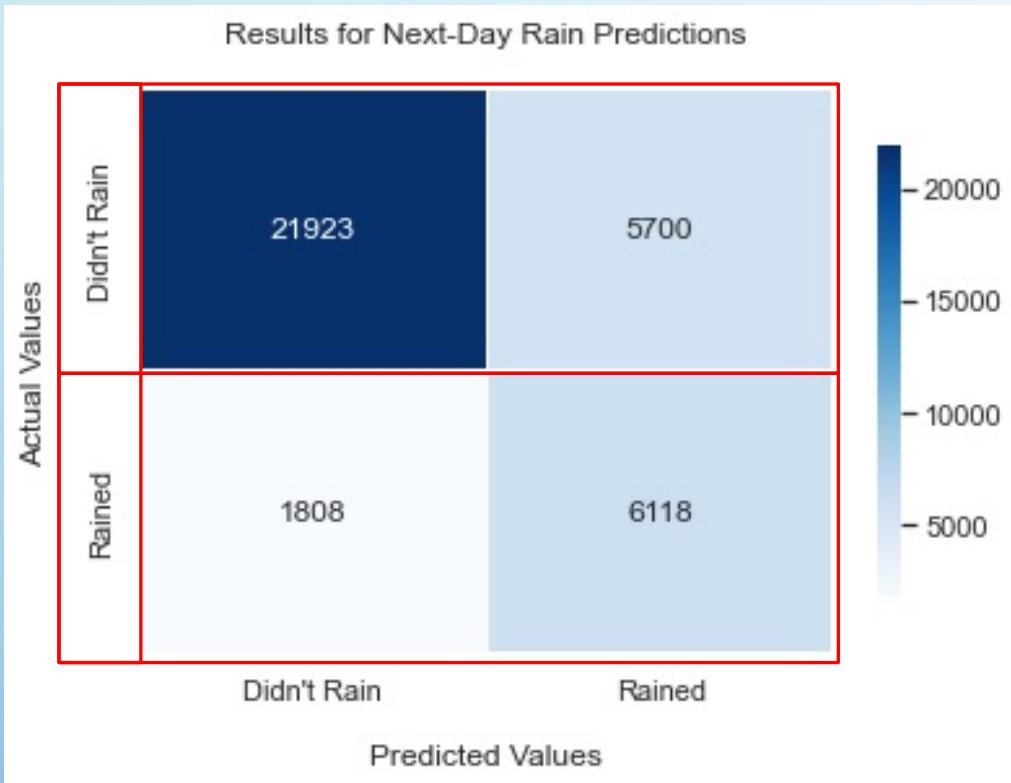
DATA PREPARATION

- Use data cleaning techniques to prepare data
 - Outliers
 - Nulls
 - Errors
- Better data equals a better model

MODELING

- Redefine the Problem
 - Bureau of Meteorology historically underpredicts rainy days each year
- Modeling Solution
 - Our model should minimize false negatives (Type II error)
 - False Negatives: Days predicted that it won't rain, and it actually ends up raining
 - Using Recall to minimize Type II error
 - Logistic Regression, k-Nearest Neighbors & Decision Trees

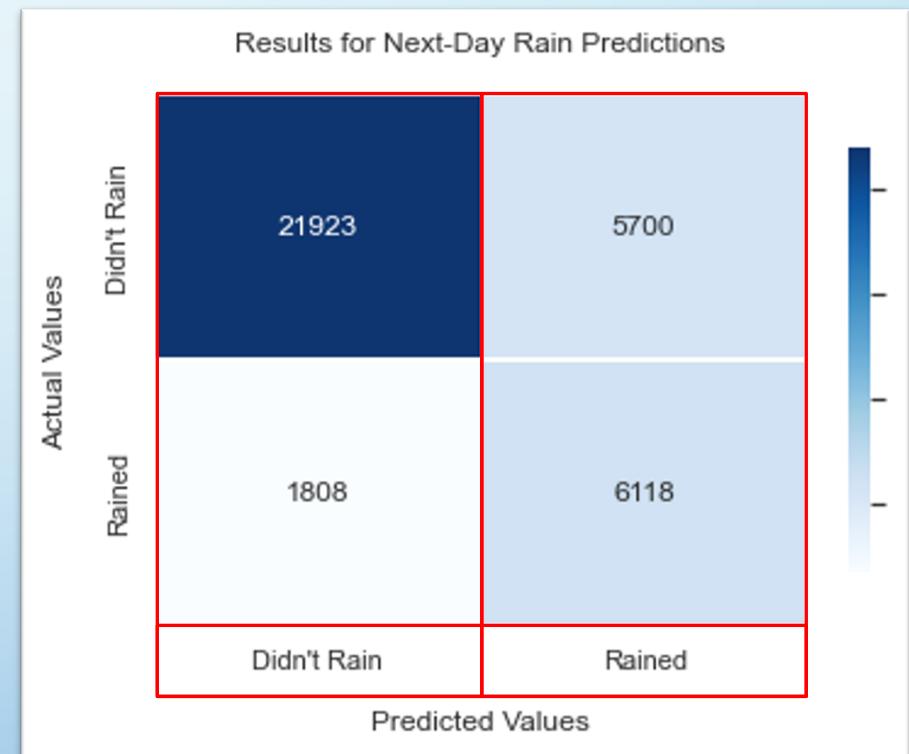
RESULTS



- Best Model: Logistic Regression with a weighted-average 79% recall score
- Actually Rained Next Day:
 - Correctly classified that it would rain 77% of the time
- Actually Did Not Rain The Next Day:
 - Correctly classified that it would not rain 79% of the time.
- Value: Model prevents False Negatives 80% of the time

NEXT STEPS:

- Find more data with different features (i.e. fog, high/low temperature, more cloud data)
- Try different model techniques
- Find a way to strengthen our Precision metric (reduce Type I error)



Thank You For Your Time!