



# ClimateWins Weather Analysis

December 2024 | Ben Newmark

# Summary

- |   |                          |   |   |
|---|--------------------------|---|---|
| 1 | Cover Slide              | 6 | Supervised Learning and Algorithm Selection |
| 2 | Summary                  |   |   |
| 3 | Objective and Hypotheses | 7 | Summary and Next Steps                      |
| 4 | Data Sources and Biases  | 8 | Thanks and Context                          |
| 5 | Data Optimization        | 9 | Questions                                   |

# Objective and Hypotheses

**Objective:** Analyze weather patterns across different locations and determine favorable climates.

## Hypotheses:

Locations with similar climates to Madrid will have higher rates of pleasant weather.

Stations located in harsher climates will consistently report more bad weather, leading to data bias.

Locations with static weather patterns (e.g., always unpleasant) will negatively impact model predictions due to overfitting

# Data Sources and Potential Biases

**Data Source:** ClimateWins station reports.

## **Potential Biases:**

The majority of stations are in unfavorable climates, skewing the data toward bad weather.

Sonnblick always reports unpleasant weather, causing overfitting for that location.

Lack of transparency in how "pleasant weather" is defined introduces potential bias in labeling.



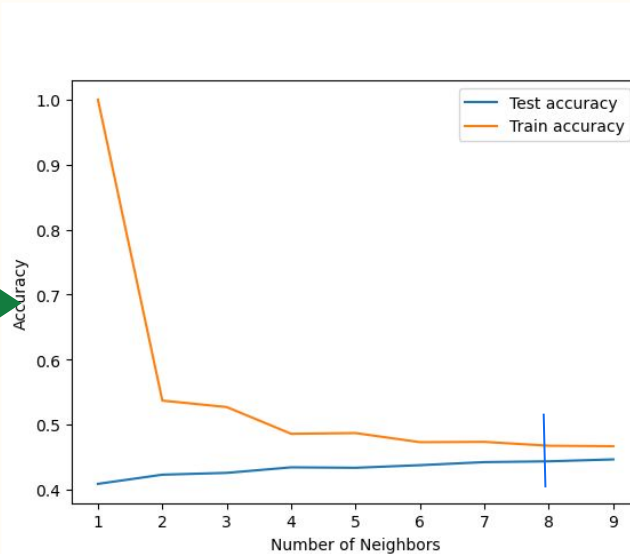
# Data Optimization

## Feature Scaling

Normalized data to ensure consistent input values for machine learning models.

## Finding Optimal K for KNN:

- Tested K values in the range of 1 to 10.
- Observed a plateau in accuracy improvement at K=8, indicating the optimal number of neighbors and avoiding overfitting



# Supervised Learning and Algorithm Selection

## Supervised Learning Methods Used:

### K-Nearest Neighbors (KNN):

- Best prediction accuracy: average 87.35% across all stations.
- Optimal K=8 ensures resource efficiency without compromising output quality.

### Decision Tree and ANN:

- Struggled with accuracy and overfitting, making them less effective.

## Why KNN?:

- Consistent accuracy across test locations.
- A simple and robust model for this dataset.

# Summary and Next Steps?



## Hypotheses Summary:

Similar climates predict pleasant weather trends.  
Harsher climates introduce bias.  
Static weather locations like Sonnblick need special handling.

## Chosen Methods:

KNN with  $K=8$  for optimal accuracy and efficiency.

## Next Steps:

Further analyze location-specific biases (e.g., Sonnblick and Madrid).  
Explore additional features or external datasets for better predictions.  
Conduct more cross-validation to ensure model robustness.

# Thank You for Your Attention!

Contact Information:

Name: Ben Newmark

Email: [Bbnew1287@gmail.com](mailto:Bbnew1287@gmail.com)





**Questions?**