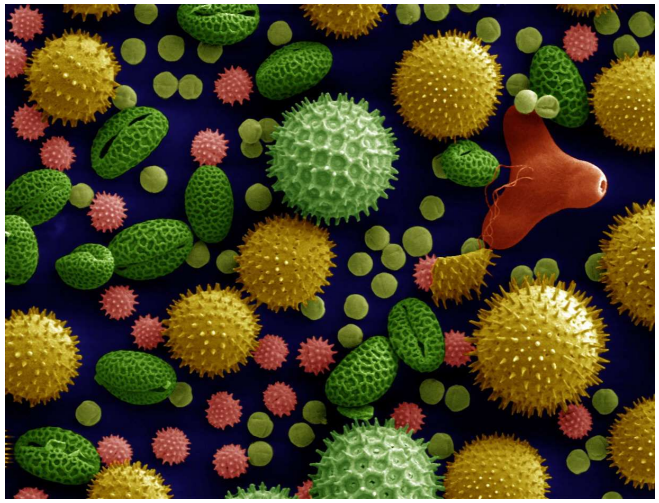
Houston,
We Have an *Allergy* Problem



Predicting Pollen Counts in Texas

May, 2022

Nicholas Ross O'Keefe Kennedy

Flat Iron Data Science Graduate
www.github.com/nikennedy
nrokennedy@gmail.com

Agenda:

1. Introduction & Business Understanding
2. Business Problem
3. Data Understanding
4. Initial Analysis
5. Modeling: *Classification & Time-Series*
6. Results
7. Species-Specific Diagnostic Tool Proposal
8. Further Considerations
9. Conclusion
10. References & Appendix

Introduction:

Allergies & Public Health After Covid-19

1. Heightened Public Awareness of Allergies
2. Seasonal Strain on Healthcare Providers

Allergies:
A Business Problem

Preparing Medical Providers & the General Public
for Allergy Season Through Public Health Initiatives

The Data: *Pollen*

- All Pollen Data From Houston Health Dept.
- Monthly Tallies of Species-Specific Pollen Counts per m^3
- Business Days from January 2013 to April 2022

The Data: *Climate*

- Daily Climate Data from

- 2 Testing Centers in Houston, TX, 1 from Shreveport, LA

- Average Daily Temp (F),
  Average Wind Speed (mph),
  Precipitation (inches)

Pollen Counts & Allergies

TREE POLLEN:

- 90-1499 /m^3: *Heavy*
- 1500+ /m^3: *Extremely Heavy*

GRASS POLLEN:

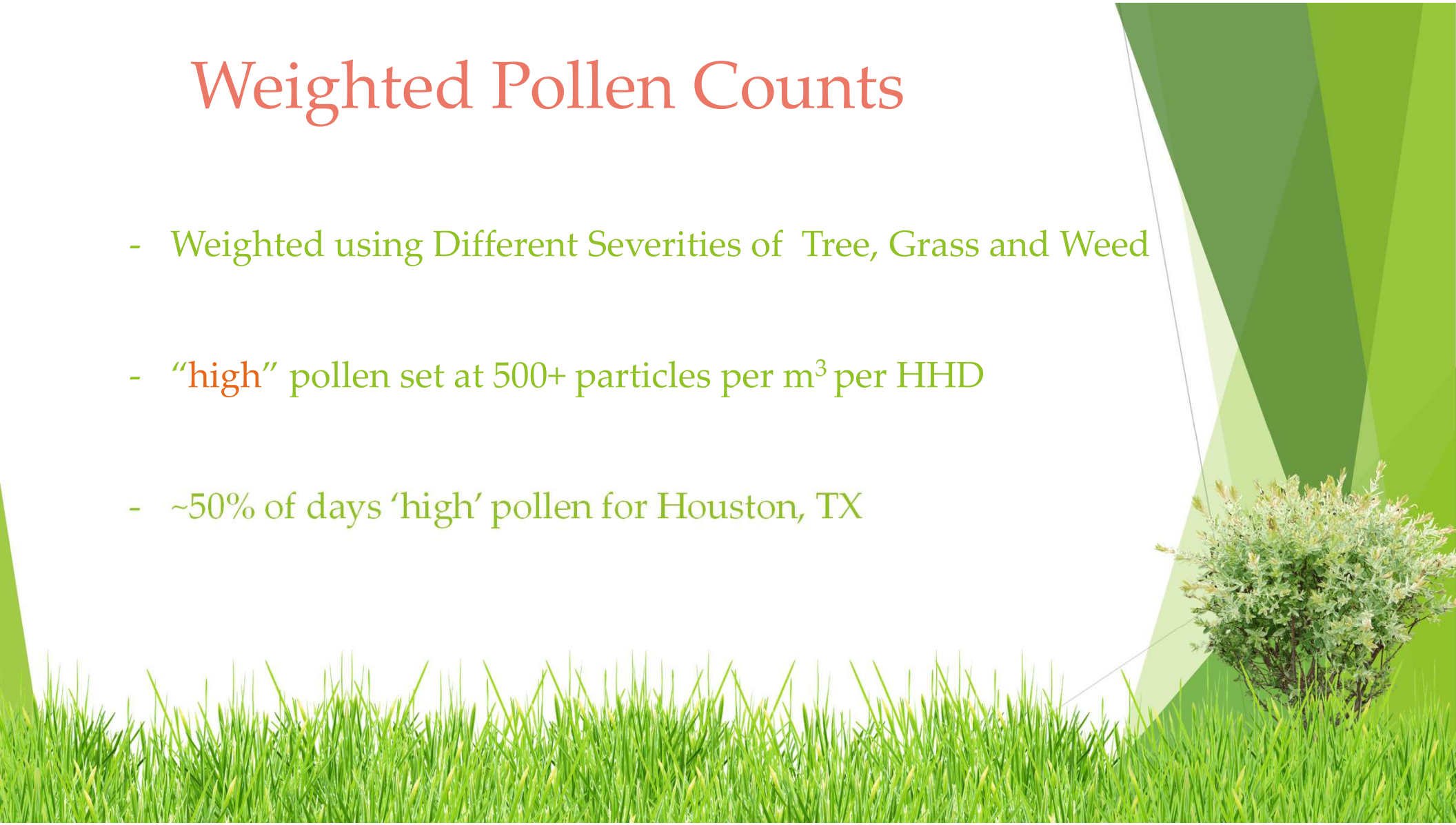
- 20-199/m^3: *Heavy*
- 200+ /m^3: *Extremely Heavy*

WEED POLLEN:

- 50-499 /m^3 : *Heavy*
- 500+ /m^3: *Extremely Heavy*

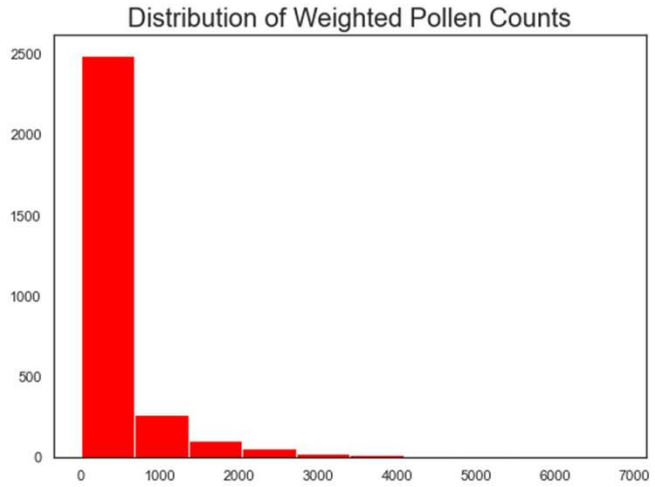
Weighted Pollen Counts

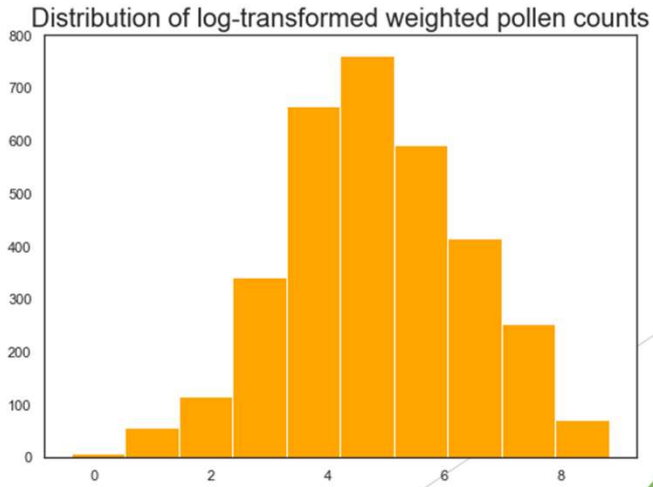
- Weighted using Different Severities of Tree, Grass and Weed
- "high" pollen set at 500+ particles per m^3 per HHD
- ~50% of days 'high' pollen for Houston, TX



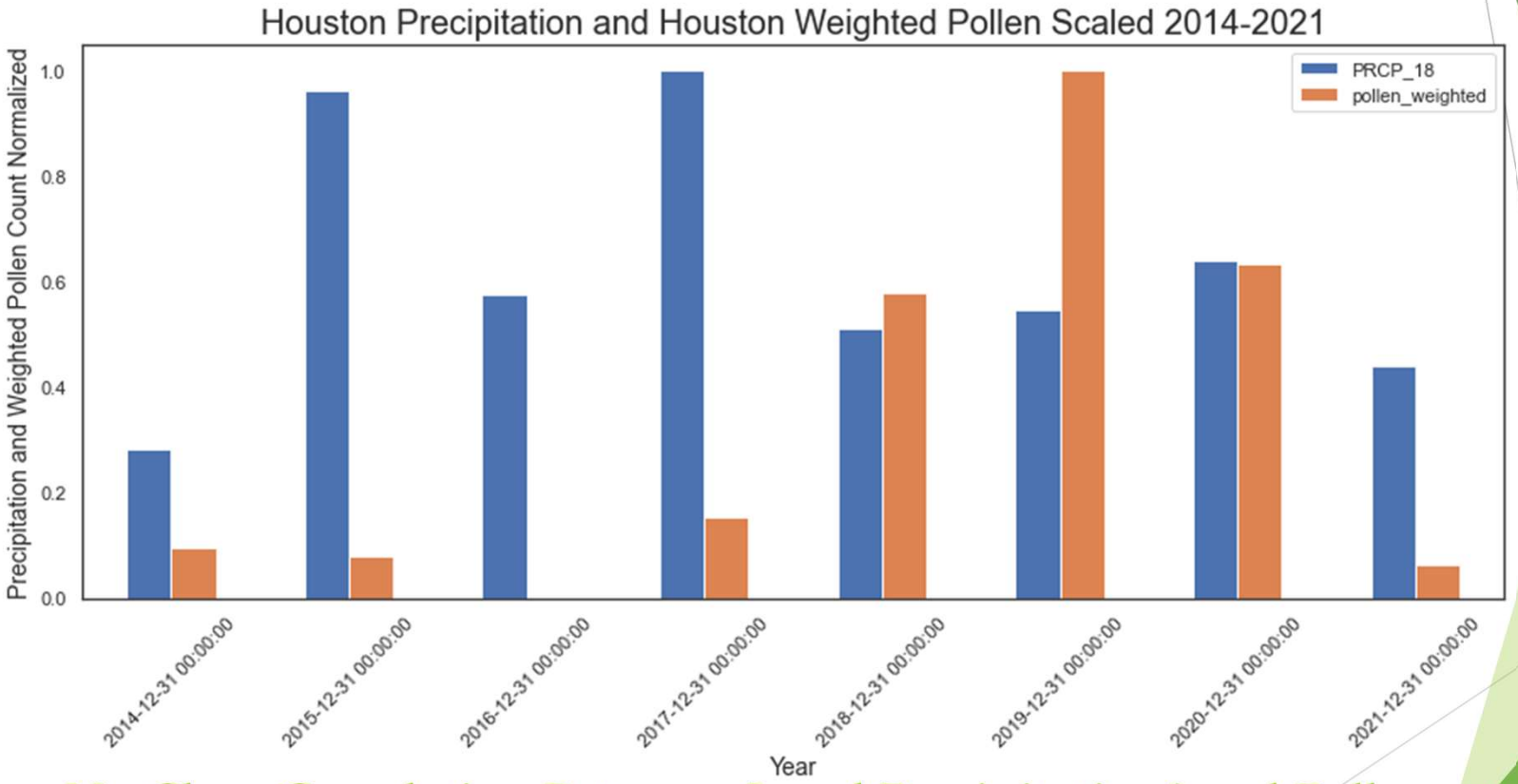
Exploratory Findings

- Many Days had pollen counts of zero especially in summertime
- Max Pollen Count of over 9,000 per m^3 occurred in March, 2019
- Pollen's Logarithmic Distribution





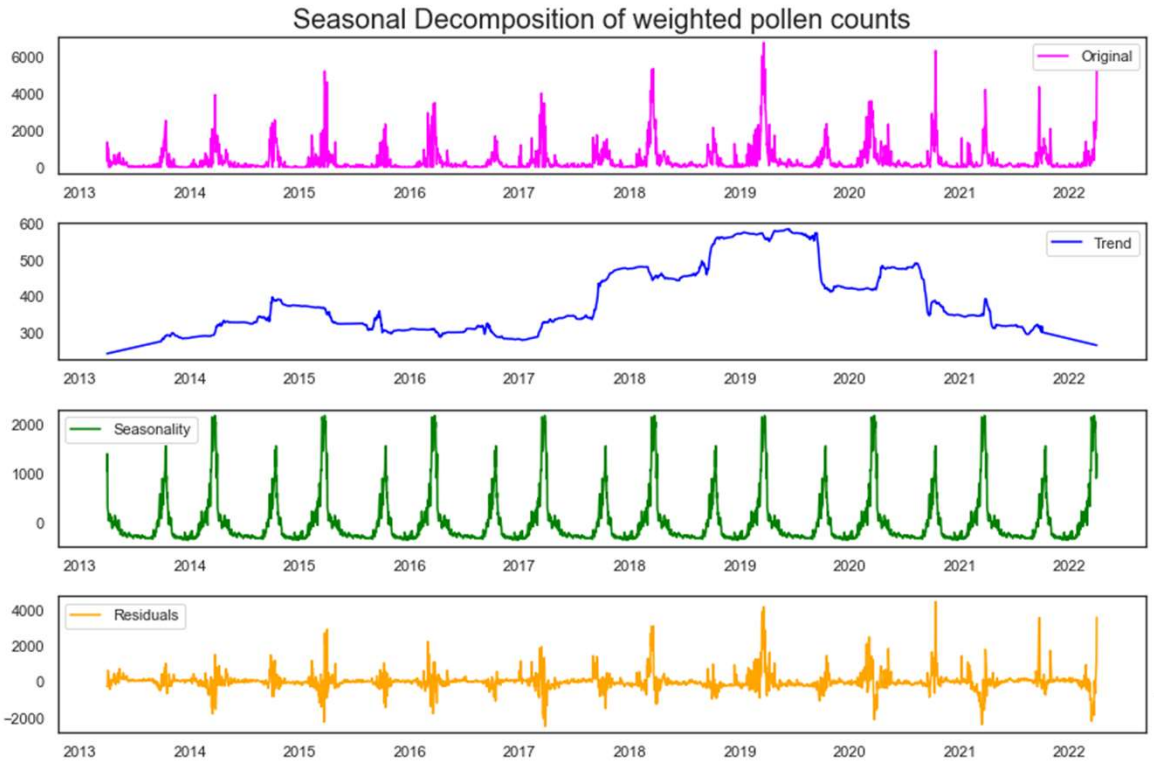
Exploratory Findings



No Clear Correlation Between Local Precipitation* and Pollen Counts for Current or Prior Year

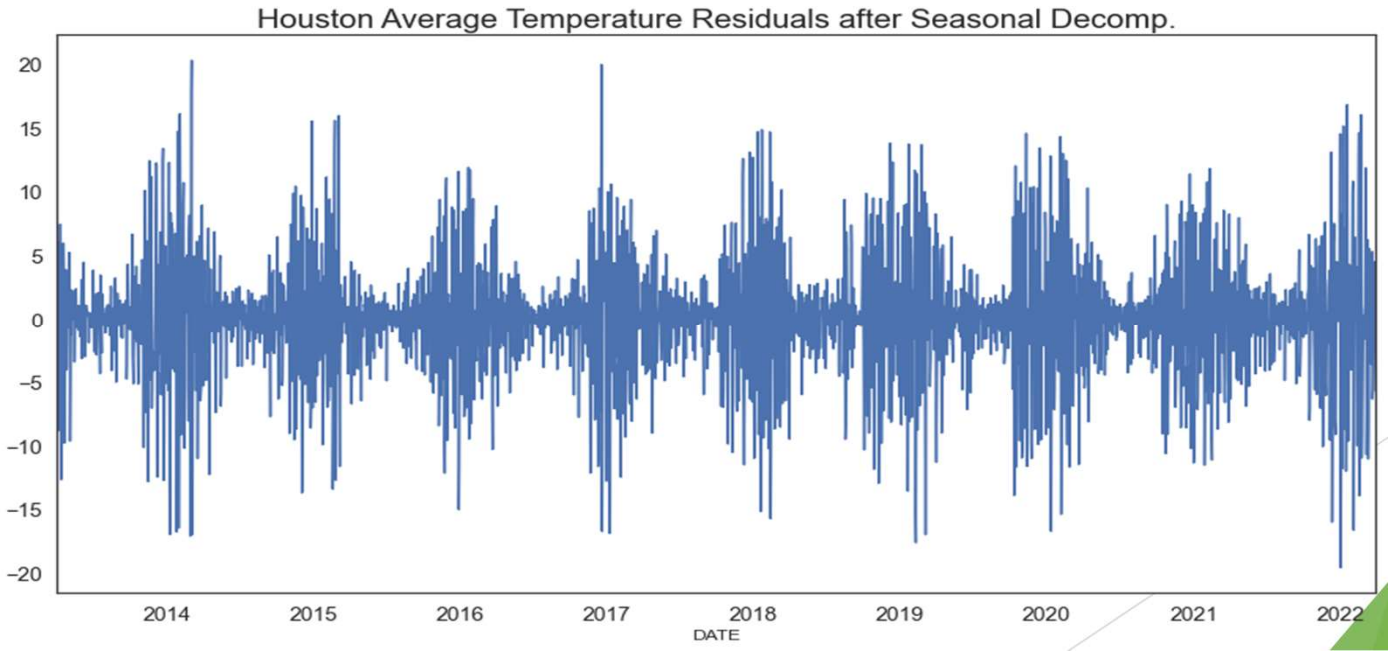
* Also True for Shreveport, & other variables (see Appendix)

Classification Modeling:
The Seasonality Problem

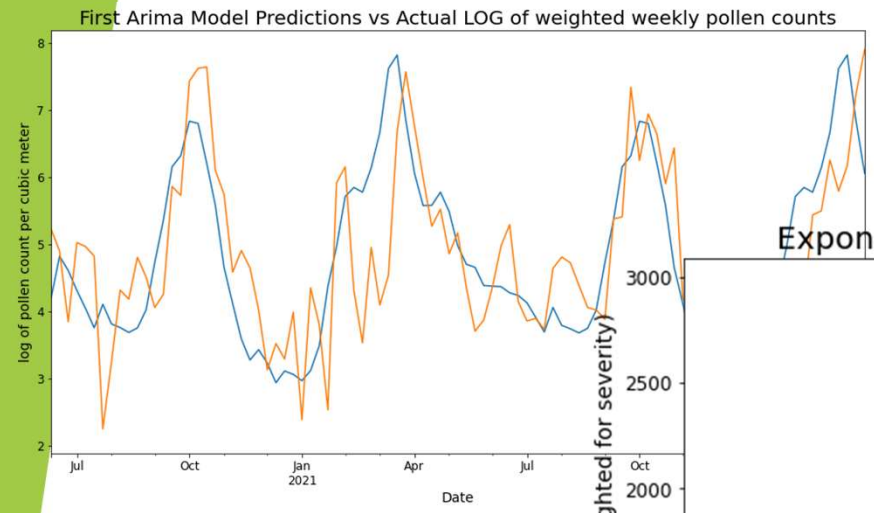


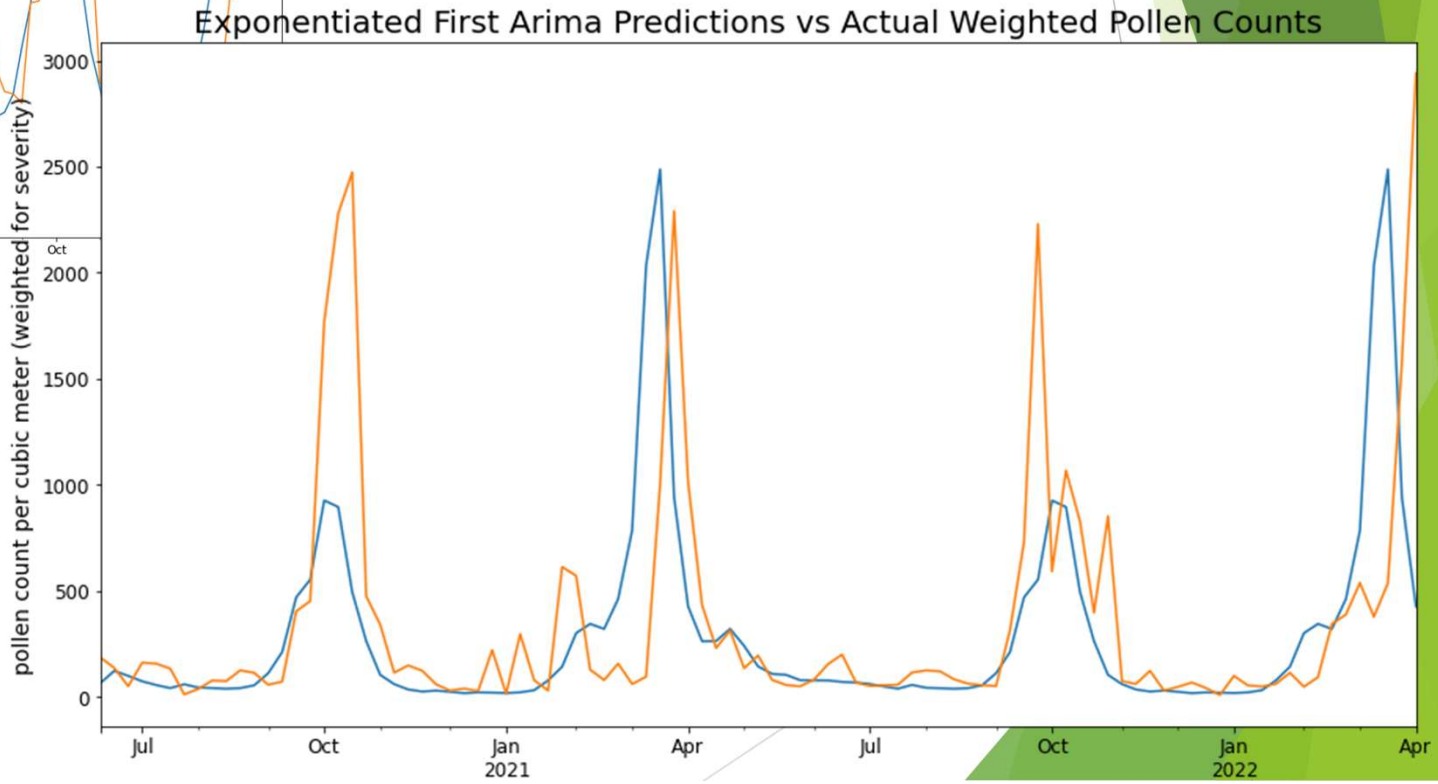
Classification Modeling

- Decision Tree Model: Local Temp most important feature
- Random Forest Classifier Achieved 60% Accuracy
- Seasonality after decomposition?



Time-Series Modeling





Results

Classification:

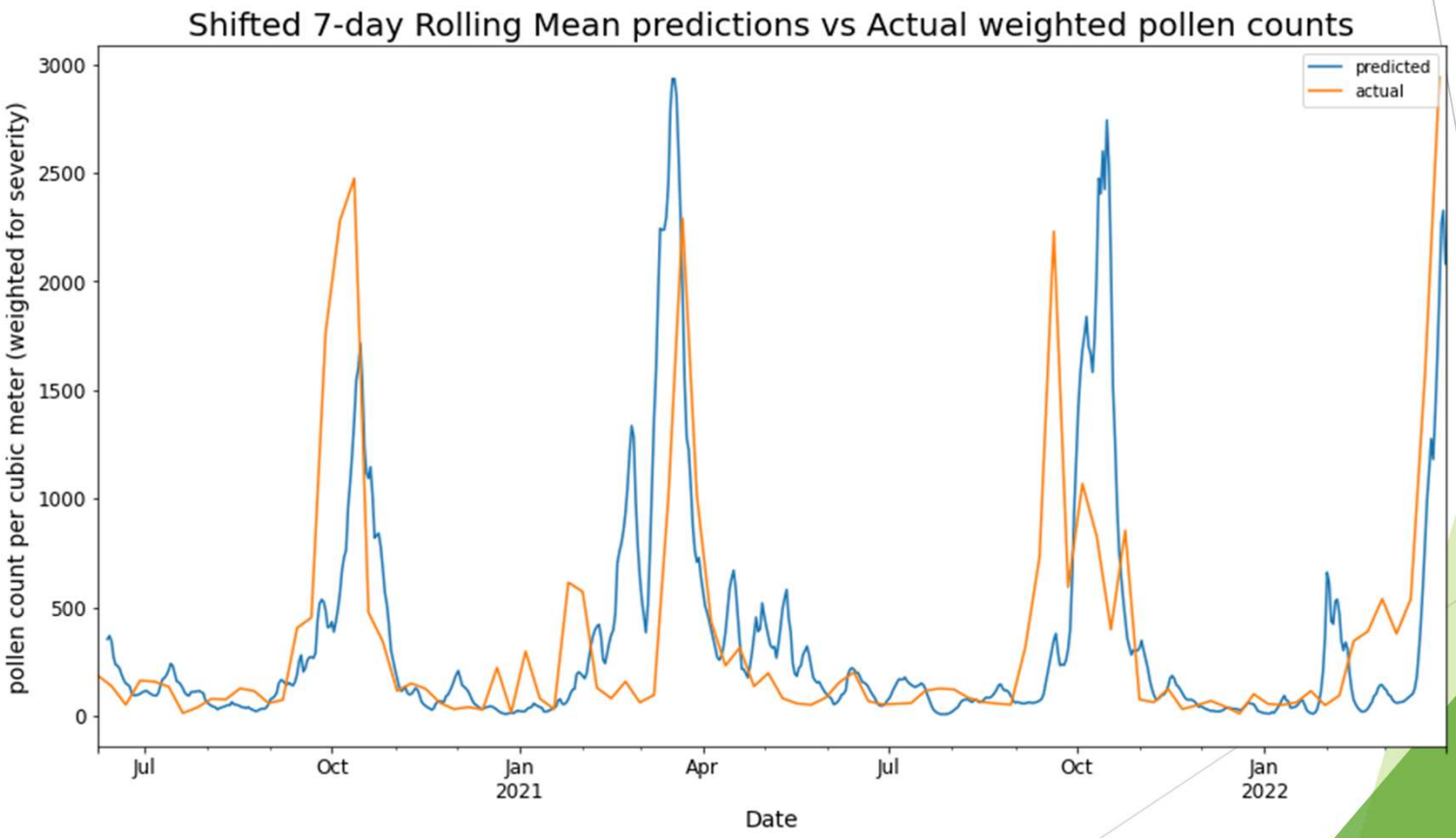
- Best Model was Random Forest 60% Accuracy, 59% Precision

Time Series:

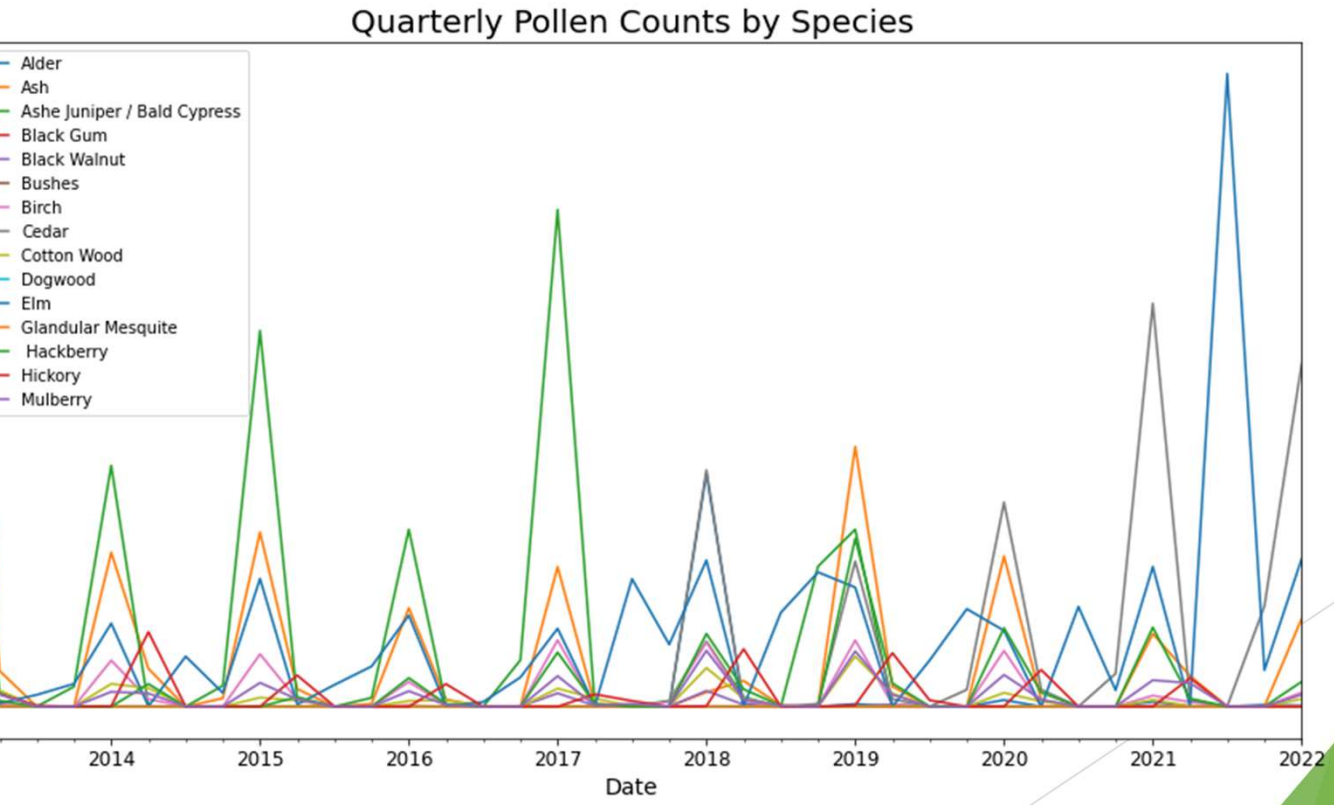
- Best Predictor was Prior Year's Pollen Counts
- Off by ~ 425 particles per m^3 on average

Results:

Best Predictions from Last Year

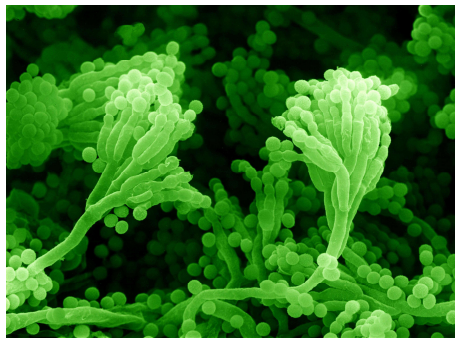


Proposal for Next Steps:
Species-Specific Diagnostic Tool



Further Considerations:

1. Mold Spores



2. More Local and Statewide Data

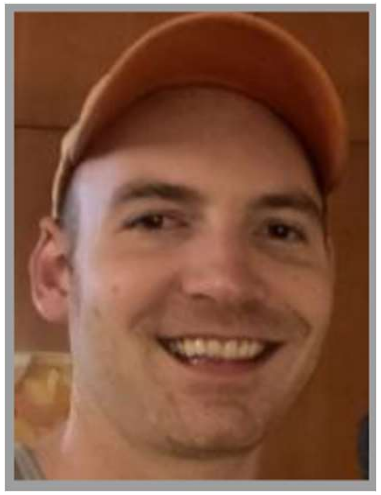
Conclusion

Seasonality is King

- The local seasonality of pollen production is the best predictor of pollen counts.

Thanks!

Questions and Comments



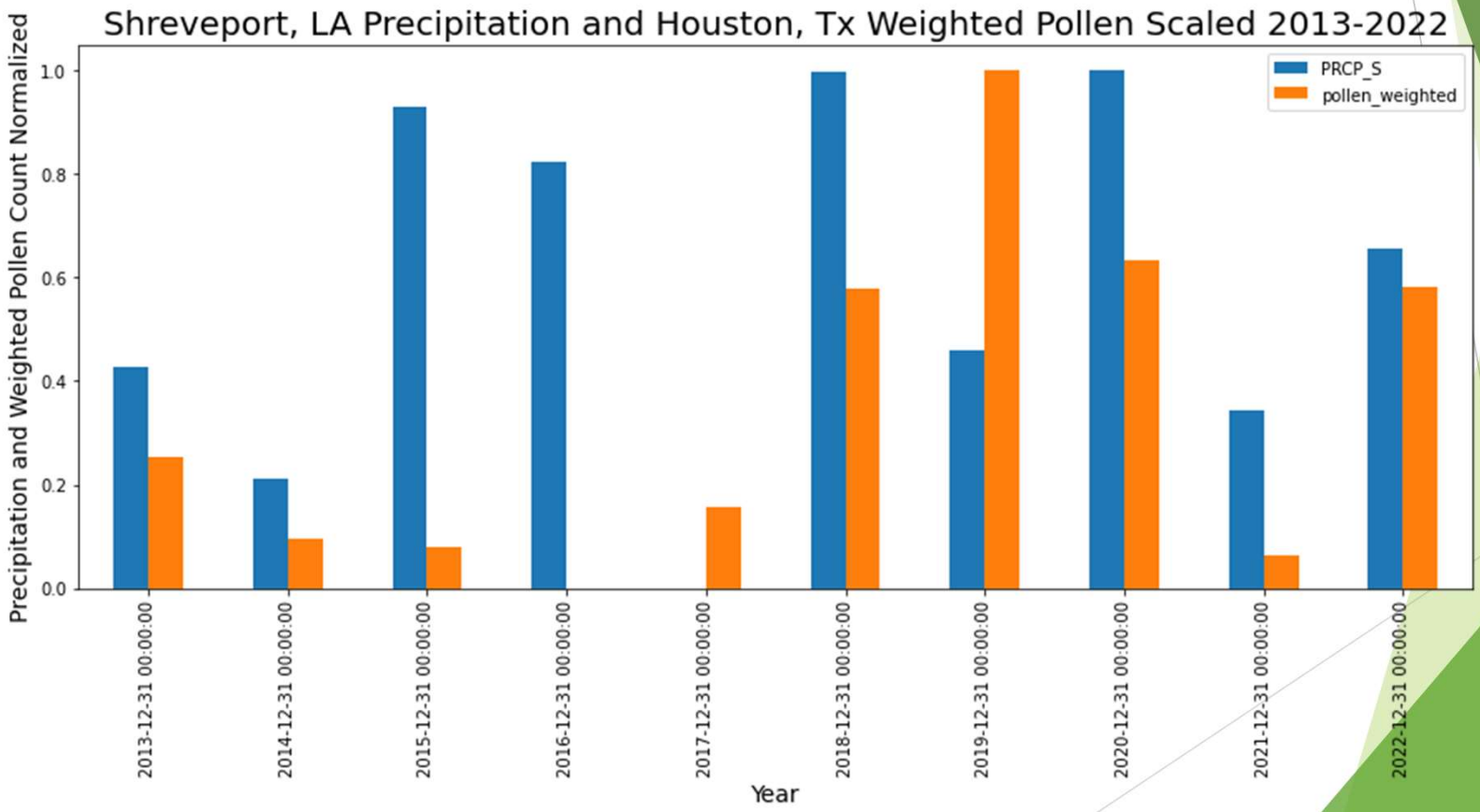
Flat Iron Data Science Graduate
www.github.com/nikennedy
nrokennedy@gmail.com

References

1. Mold Spores

1. More Local and Statewide Data

Appendix


Shreveport, LA Precipitation and Houston, Tx Weighted Pollen Scaled 2013-2022
Precipitation and Weighted Pollen Count Normalized
PRCP_S
pollen_weighted
1.0
0.8
0.6
0.4
0.2
0.0
2013-12-31 00:00:00
2014-12-31 00:00:00
2015-12-31 00:00:00
2016-12-31 00:00:00
2017-12-31 00:00:00
2018-12-31 00:00:00
2019-12-31 00:00:00
2020-12-31 00:00:00
2021-12-31 00:00:00
2022-12-31 00:00:00
Year