

**UC San Diego**

**JACOBS SCHOOL OF ENGINEERING**  
Electrical and Computer Engineering

---

# **Decoding Pandemic Patterns: Integrating Wastewater Viral Load, Weather, and COVID-19 Epidemiology**

ECE 143 - Programming for Data Analytics

Final Project

---

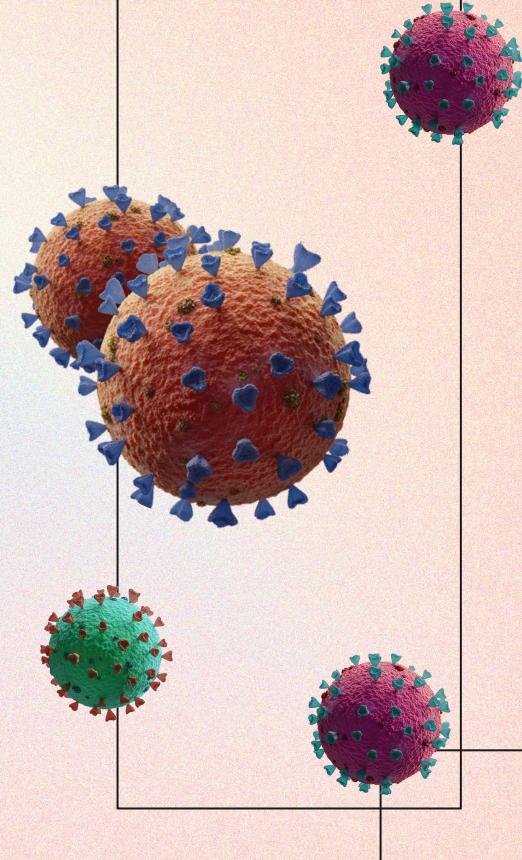
June Young Park

Yusuf Maung

Anmol Kalantri

Wooil Jung

Subhanjan Konwer



---

# TABLE OF CONTENTS

**01**

—  
Problem  
Statement

**02**

—  
Methodology

**03**

—  
Dataset Used

**04**

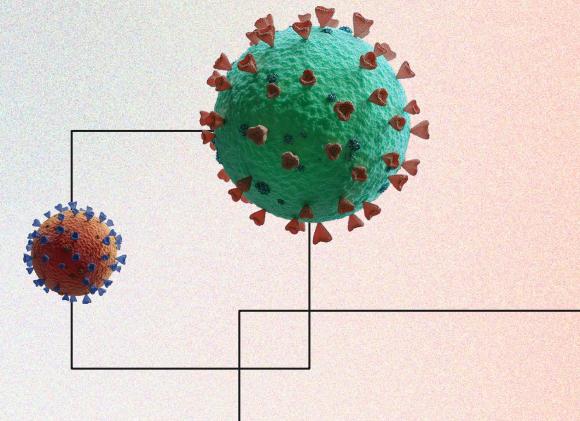
—  
Data wrangling &  
Feature engineering

**05**

—  
Correlation &  
Visualization

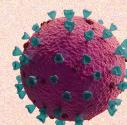
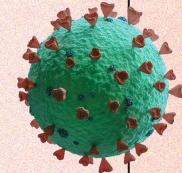
**06**

—  
Conclusion



# Problem Statement

- Since the declaration of COVID-19 (SARS-CoV-2) as a pandemic in 2020, it has caused significant morbidity and mortality worldwide. In this view, understanding real-time COVID-19 activity has grown increasingly challenging as ubiquitous surveillance systems - for instance, clinical case reporting, diagnostic testing volume and hospitalization/death metrics - have become incomplete or delayed.
- In these kind of rapidly evolving biosecurity landscapes, understanding the dynamics of any given disease through Environmental surveillance (ES) methods - like wastewater and weather monitoring is key to crafting diligent interventions.
- Our aim, with this project, is to integrate three complementary data streams—wastewater SARS-CoV-2 viral load (three major wastewater facilities across San Diego County), weekly COVID-19 epidemiological indicators (cases, hospitalizations, deaths), and weather patterns (temperature)—to investigate:
  - How reliably wastewater viral concentrations reflect real COVID-19 activity in the community.
  - How weather conditions influence wastewater viral signal dynamics.
  - How these relationships evolve across different COVID waves and over multiple years.
  - And lastly, how our inferences and physical implications may be applied to future epidemics.



# Methodology

## Data Acquisition & Integration



- Collected SARS-CoV-2 wastewater qPCR measurements from three wastewater treatment facilities (South Bay, Point Loma, Encina).
- Obtained weekly COVID-19 epidemiological indicators (cases, hospitalizations, deaths), as well as weather, temp, humidity data

## Data Wrangling and Feature Engineering

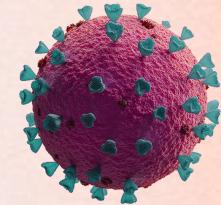


- Converted all timestamps to consistent datetime indices and handled missing/irregular sampling days.
- Aggregated daily wastewater values to epidemiological week boundaries using mean Z-scores.
- Smoothed noisy observations (centered rolling averages) to reduce variability from reporting artifacts.

## Correlation & Lag Analysis



- Quantified lead-lag relationships between wastewater and COVID outcomes.
- Performed rolling-window correlation during epidemic waves to evaluate dynamic coupling strength.



## Multi-Factor Interpretation & Final Synthesis



- Integrated results across wastewater, epidemiology, and weather to interpret community-level infection dynamics.
- Compared cross-site patterns to identify robustness of wastewater as a surveillance tool.
- Finally, summarized how these combined signals can inform early-warning public health systems.

## Weather Interaction Assessment

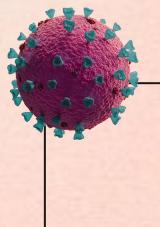
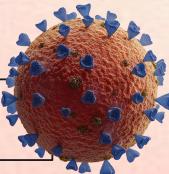


- Merged wastewater temperature data per site and examined correlations with Z-scores.
- Evaluated whether temperature influences viral detectability or wastewater-COVID alignment.

# Datasets

Our analysis integrates heterogeneous data sources that jointly describe community-level infection dynamics, environmental conditions, and wastewater viral shedding patterns.

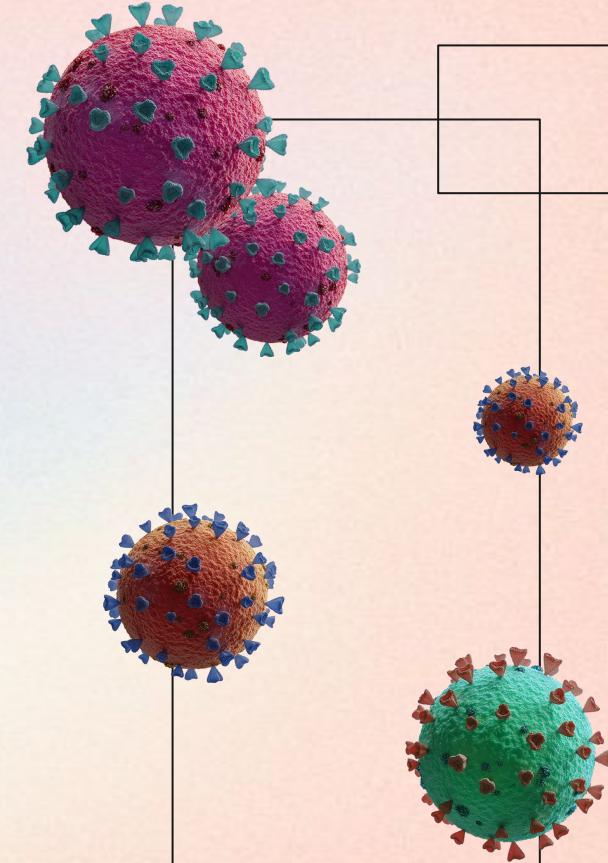
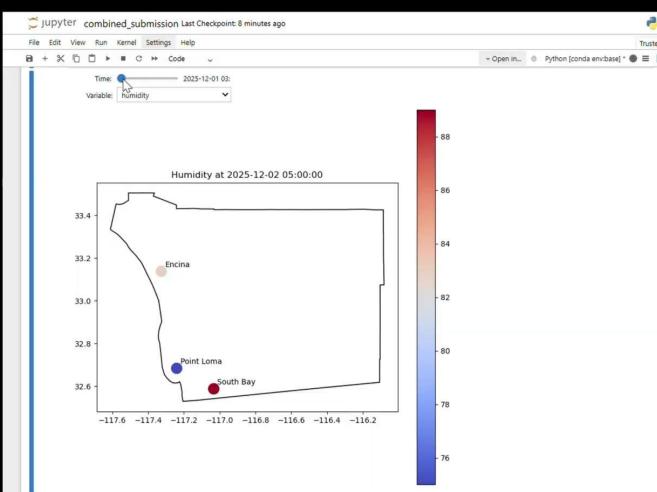
1. [https://www.sandiegocounty.gov/content/sdc/hhsa/programs/phs/community\\_epidemiology/dc/respiratoryviruses/surveillance.html](https://www.sandiegocounty.gov/content/sdc/hhsa/programs/phs/community_epidemiology/dc/respiratoryviruses/surveillance.html)
  - This is the county's official 'Respiratory Viruses Surveillance' page which houses counts for cases, hospitalizations, deaths grouped weekly (by onset/collection/report week) for the period of 2021-2025.
2. <https://searchcovid.info/dashboards/wastewater-surveillance/>
  - This dataset is a time-series of SARS-CoV-2 RNA concentrations in wastewater from major San Diego treatment plants, grouped by data and site, intended for population-level surveillance of COVID-19. It contains trend-analysis, early warning, and public health monitoring of viral spread in the region.
3. <https://open-meteo.com/>
  - Weather api pulled from using different coordinates for different cities to get weather information over time





# Weather Data

- Open Meteo API
- Geopandas Visualization
- Slider added for showing weather changes over time



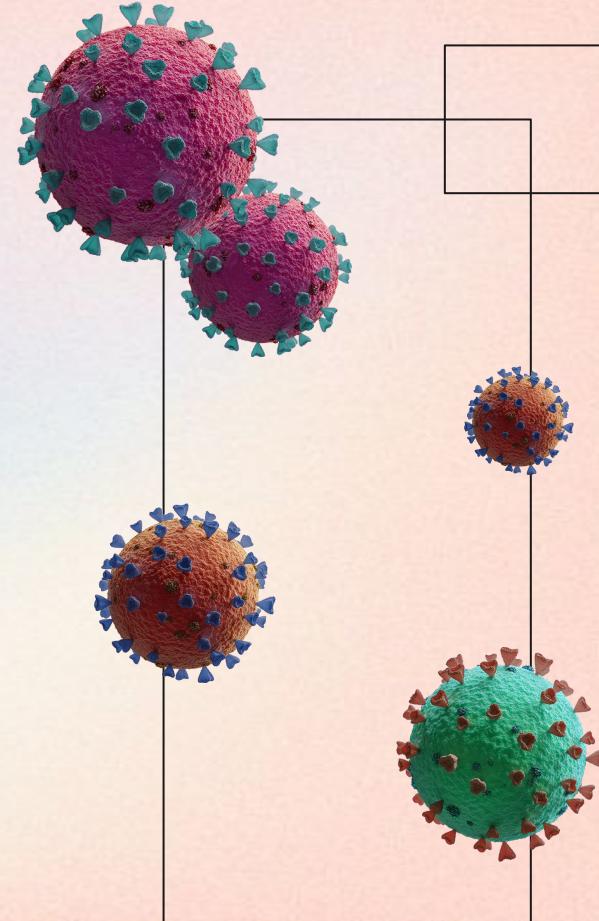


---

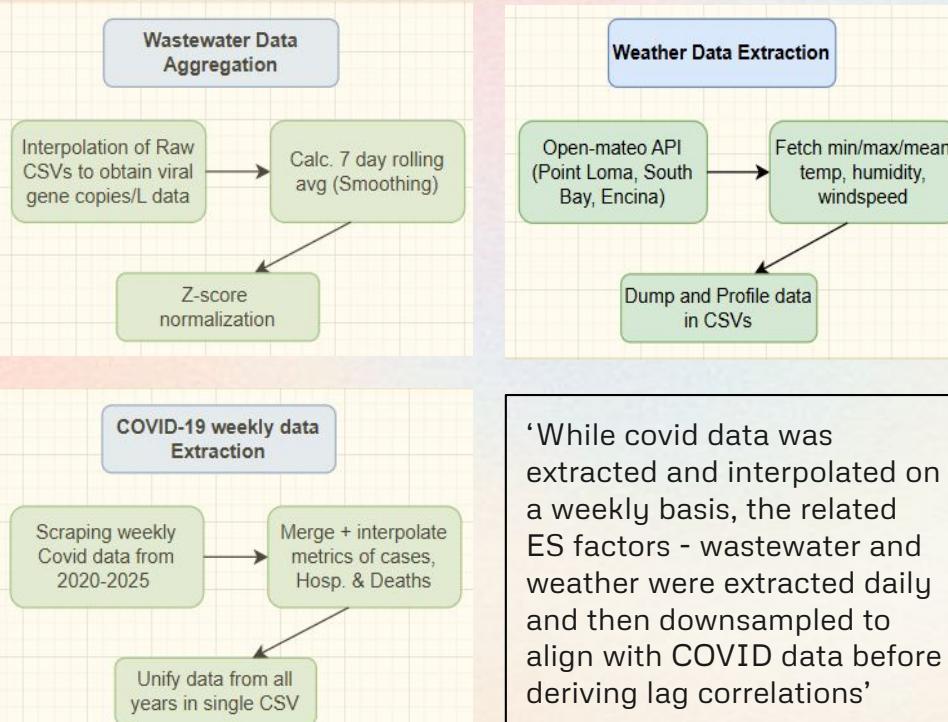
## Data wrangling & Feature engineering

---

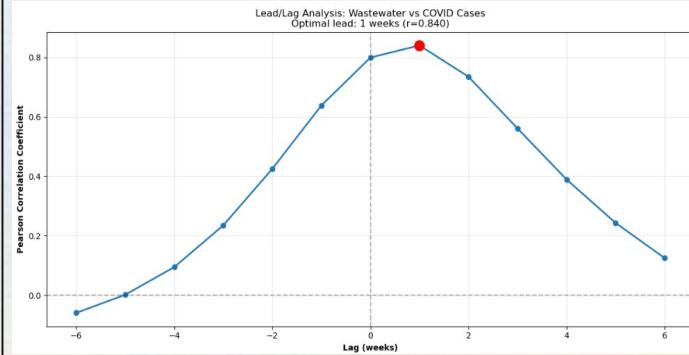
Transforming raw wastewater, weather, and epidemiological streams into meaningful signals required careful preprocessing and engineered features tailored to time-series correlation analysis.



## Data Wrangling



## Lag Correlation (wastewater v/s Covid)



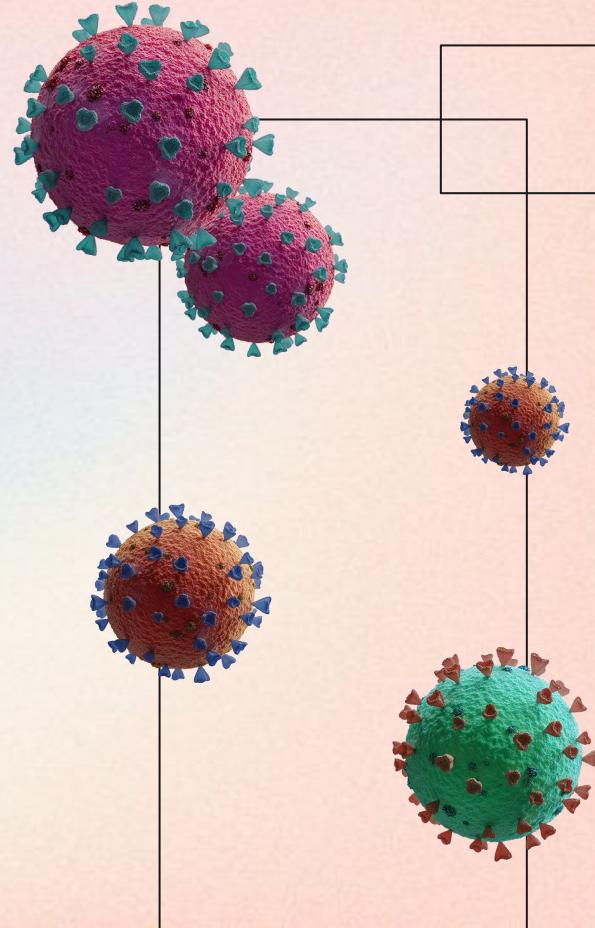
- In time series epidemiology, Lag/lead correlation analysis quantifies whether one signal consistently precedes or follows another.
- Mathematically,  $\rho(k)=\text{corr}(W(t-k), C(t))$
- The strong correlation ( $r \sim 0.84$ ) indicates synchronized trends: both signals track the same underlying epidemic

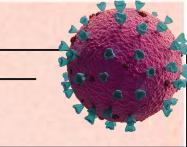
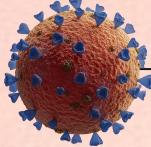


---

# Correlations & Visualizations

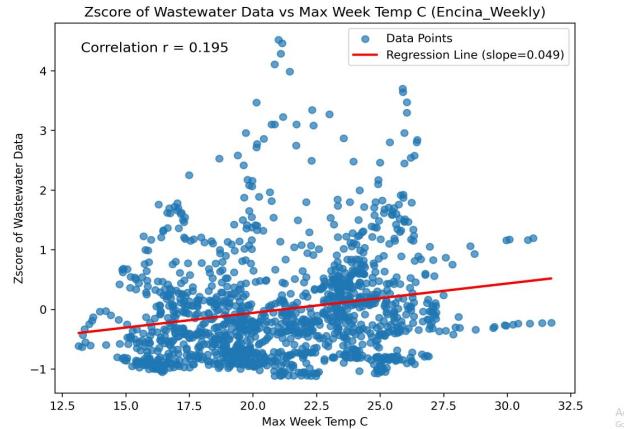
We now examine how these engineered features interact through a series of statistical correlations and visual analyses.



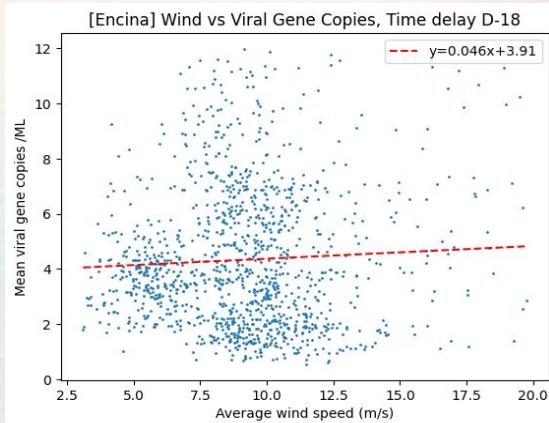


# Weather factors v/s Wastewater

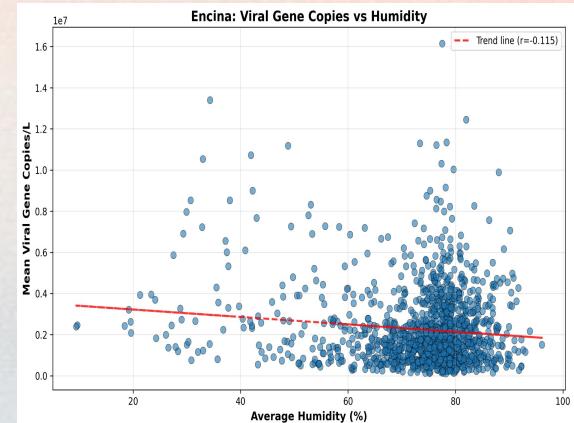
## Temperature vs Wastewater



## Wind Speed vs Wastewater



## Humidity vs Wastewater



- Computed correlations between wastewater Z-scores and daily min, max, and avg temperatures.
- Correlations were **very weak**, with Encina showing the highest values.
- Max temperature had the strongest (but still weak) correlation.
- Using 7-day rolling averages slightly increased correlations but overall trends remained weak (Graph shows highest correlation).

- Computed correlations between average wind speed and wastewater.
- Correlations were **very weak**, with Encina showing the highest values at a delay coefficient of 18.
- Pearson, Spearman, and Kendall correlation coefficients resulted in either very weak values or high p-values.

- Computed correlations between average humidity and wastewater.
- The correlations turned out to be **extremely weak**, essentially indicating no meaningful relationship.
- Among all sites, Encina showed the largest correlation in terms of magnitude

# Wastewater v/s Covid

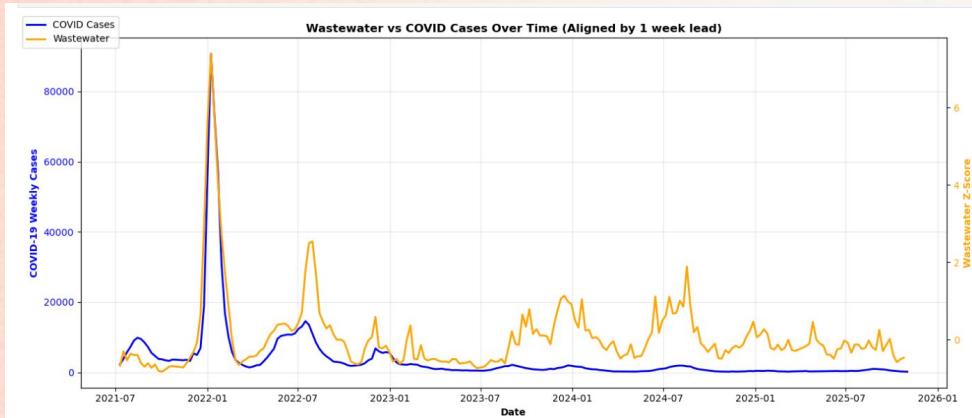
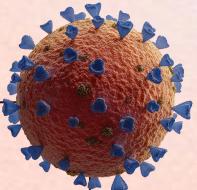


Fig. 1: While wastewater and COVID cases align closely during major outbreaks, wastewater continues to detect viral activity in later periods when clinical reporting declines, highlighting its superior sensitivity for ongoing surveillance.

This is an interesting insight - the **Omicron Variant (B.1.1.529)** - first identified in Nov 2021 - quickly became the dominant variant due to its extremely high transmissibility and ability to evade some vaccine immunity!



This inference is also backed by research! -  
<https://doi.org/10.1016/j.eti.2022.102667>

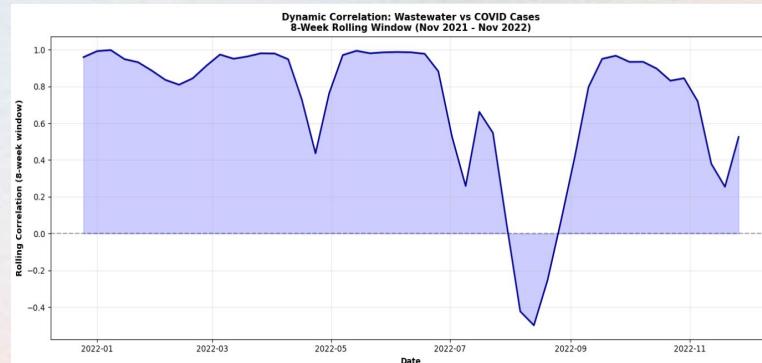
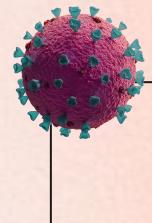
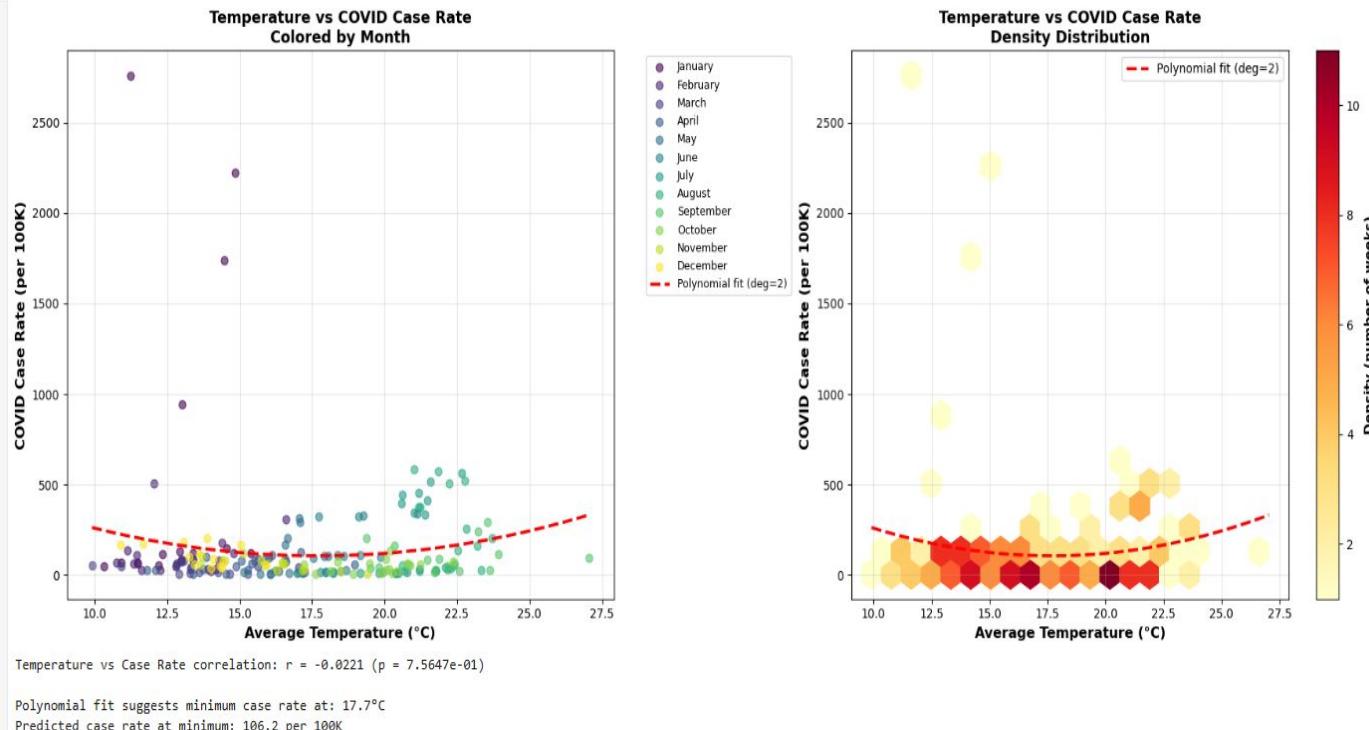


Fig. 2: Rolling window Dynamic Correlation



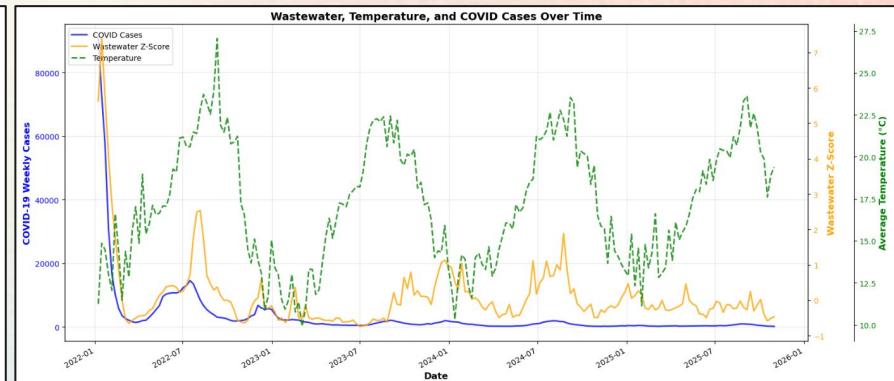
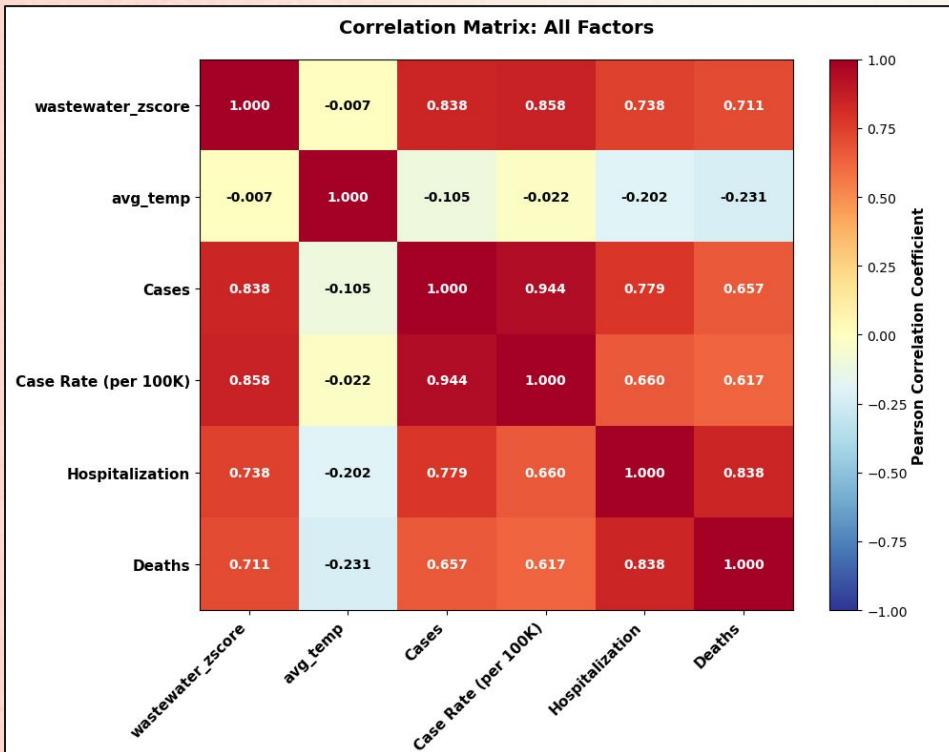
# Weather correlation with COVID-19 transmission



Winter months (Jan-Mar, 11-15 degrees ~avg): Highest case rates observed (some exceeding 2500 per 100K) - indoor crowding, reduced ventilation, increased susceptibility during cold season. Mild temperatures (15-22 degrees ~avg) are most optimal with lower case rates.

As a conclusion, we propose that weather alone is not the primary driver - The weak correlations suggests temperature is just one of many factors (the rest being behavior, immunity, variants, etc.)

# Combined Analysis: Wastewater, Weather, and COVID-19



**Key insight!!** 🧑

We observe that Wastewater correlations (0.71-0.86) are highly significant and practically meaningful. On the other hand, temperature correlations (-0.23 to -0.02) are too weak to be useful for prediction.

Hence, we can safely conclude that for future resource allocation, wastewater trends can be useful to forecast hospital bed needs or covid-related death trends!

# Conclusions & Results

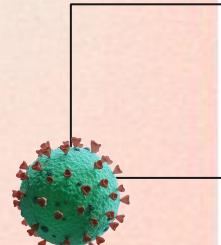
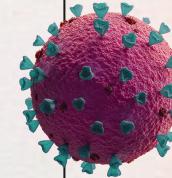
---

Wastewater data is highly correlated with covid cases, deaths, and resurgence.

- Reaffirms the ongoing surveillance of these statistics

Weather is weakly correlated with covid

- Not a significant predictor of covid cases, particularly compared to existing metrics
- Should not be used as a predictive measure



---

# THANKS!

Visit our project at:

<https://github.com/jpark0818/ECE143-Project-Group9>

Visualizations embedded in ipynb notebook:

[https://github.com/jpark0818/ECE143-Project-Group9/blob/main/combined\\_submission.ipynb](https://github.com/jpark0818/ECE143-Project-Group9/blob/main/combined_submission.ipynb)

