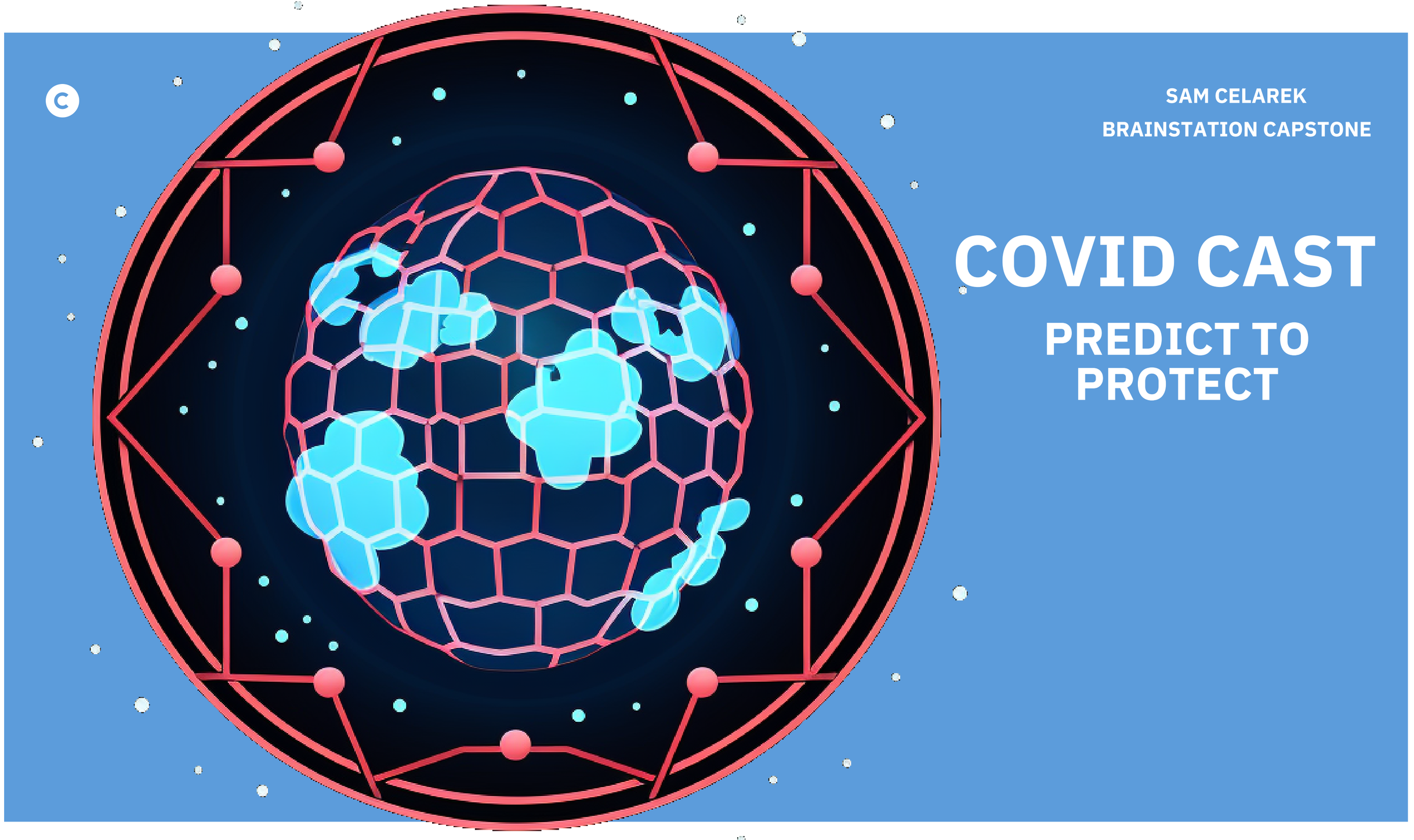
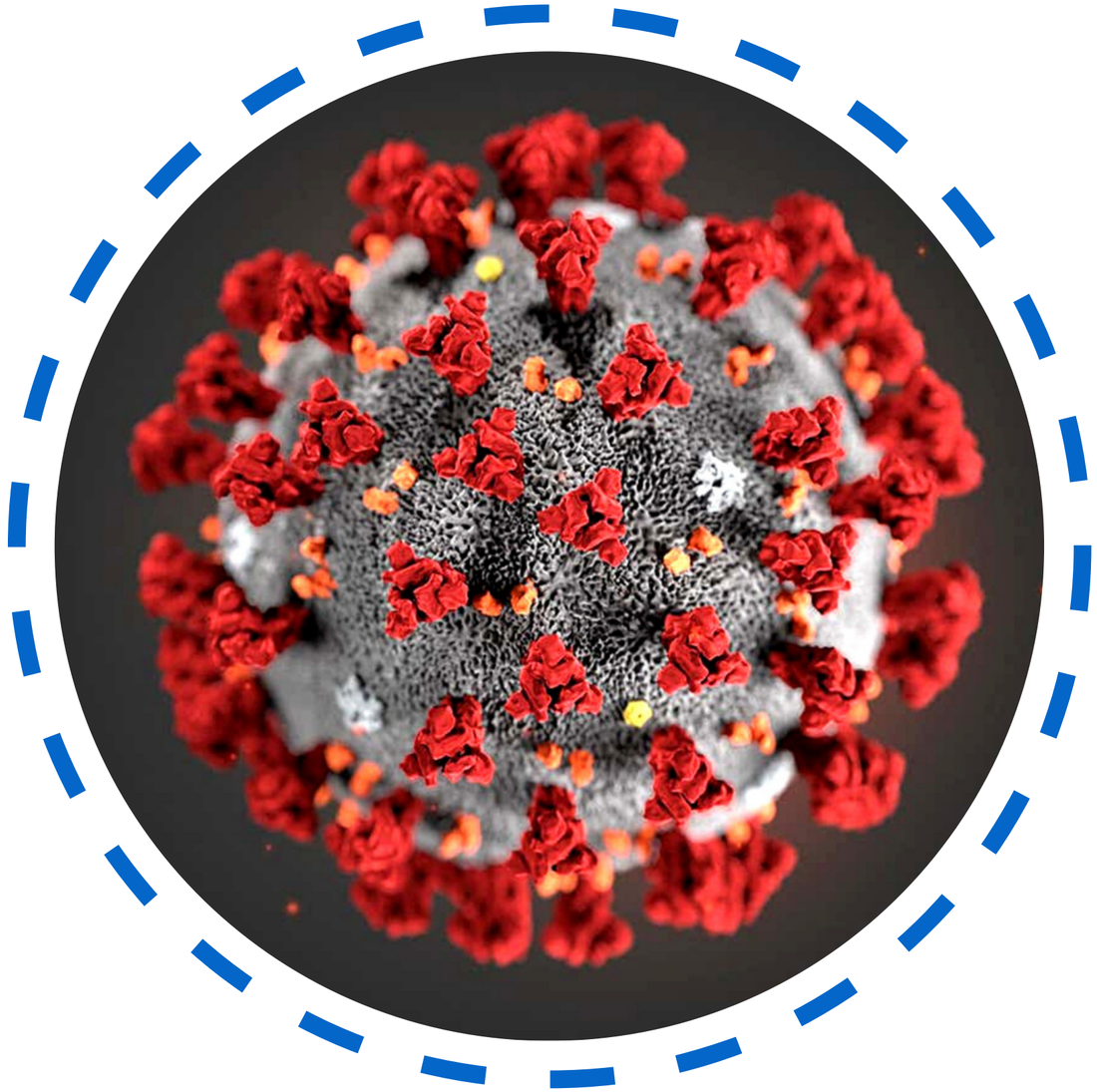


SAM CELAREK

BRAINSTATION CAPSTONE

# COVID CAST

## PREDICT TO PROTECT



# THE PROBLEM

**Pandemics are an existential threat.**

# COVIDCASTS GOAL

**Predict to Protect.**

# Time Line of Covid

**OUTBREAK**

First death in January 2020.

**January**
59 Cases
1 Dead

**March**
1m Cases
100k Dead

**PANDEMIC!**

By March 2020, COVID on every continent.

**WORLD UNDER SEIGE**

By July, 1/3 of the world in lockdowns. New Delta Variant,

**August**
30m Cases
1m Dead

**December**
100m Cases
2m Dead

**THE COMEBACK**

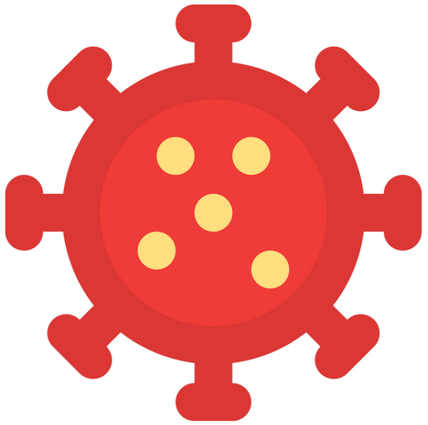
First vaccines in 2020 December.

**WHAT'S NEXT?**

**Today**
760m Cases
7 m Dead

# COVIDCast:

## SIRD MODEL



**REPRODUCTION RATE**

How many people get infected from one sick person!?


## Time Series Models

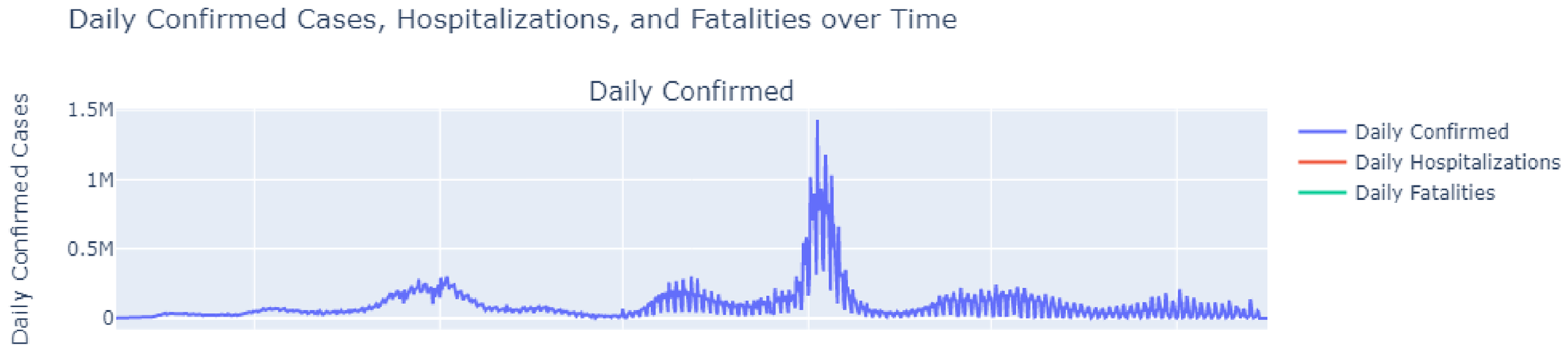
1. SARIMAX
2. PROPHET

# TARGET FEATURES

## NATIONAL US DATA

### NEW CONFIRMED 🎯

The number of new confirmed COVID-19 cases on the given date.



# DATA COLLECTION

## COVSIRPHY

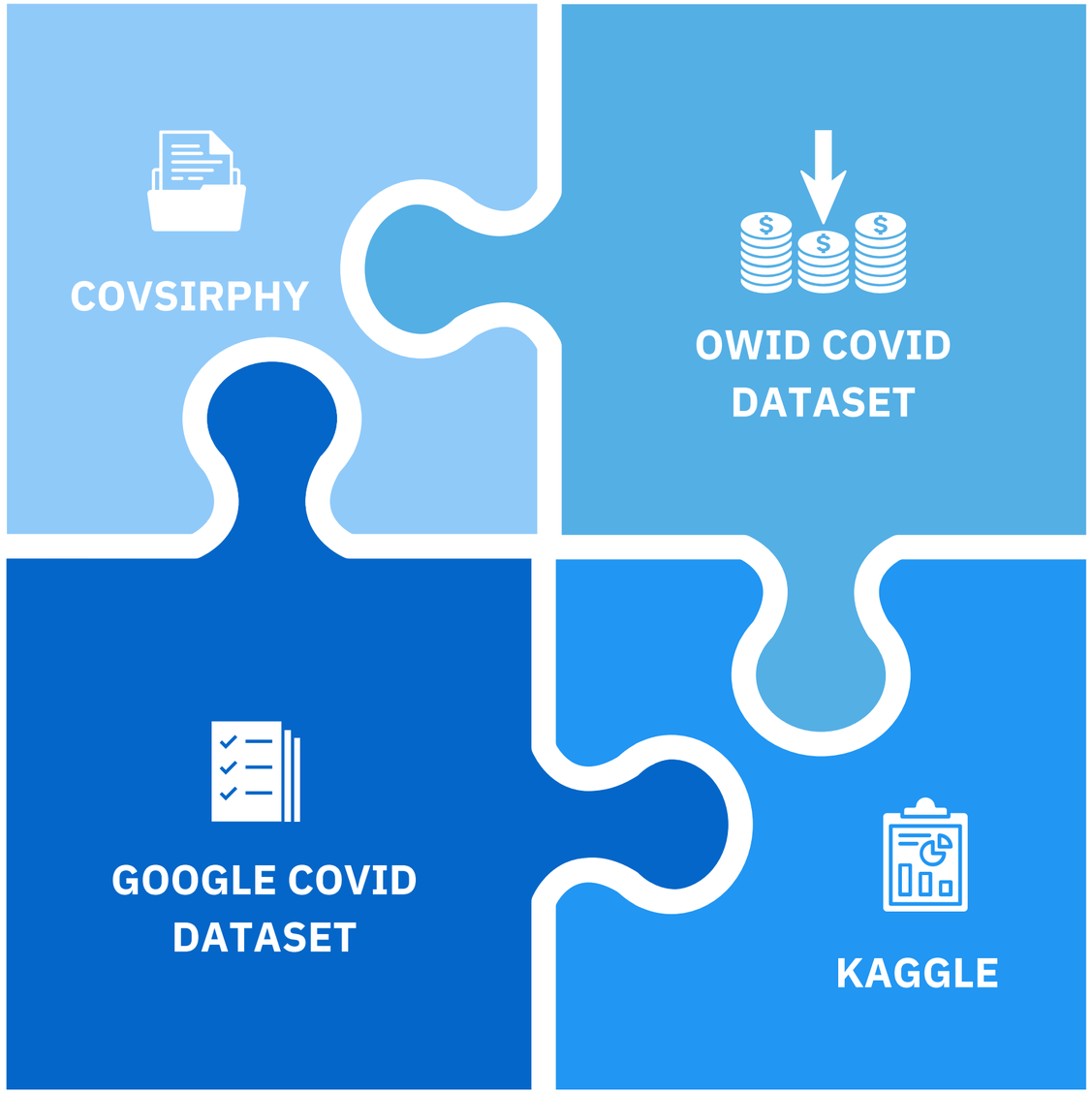
- Epidemiological Modeling.
- Real Time Global Covid Datasets.

## GOOGLE DATASET

- Weather, Mobility, Regulation
- Stopped 9/15/2022

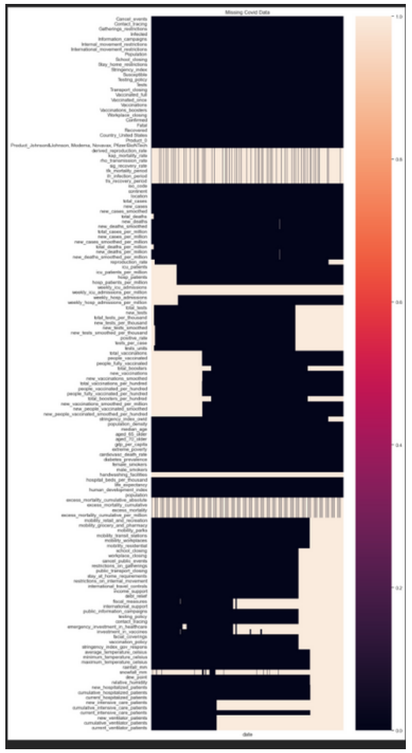
## OWID COVID DATASET

- Hospitalizations, Testing, Excess Morbidity.

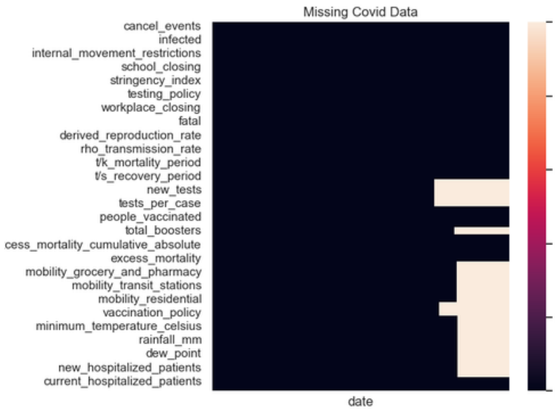


# The Stages of Cleaning

**1142 x 141 columns, 30% nulls**



**1142 x 59 columns, 7% nulls**




**1131 x 45 columns**

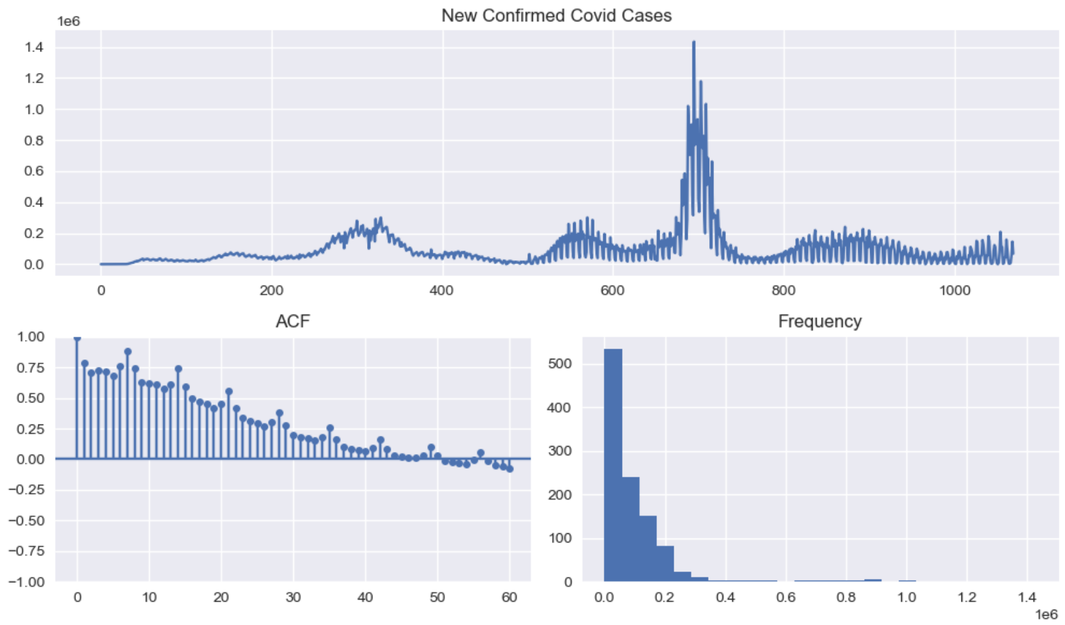

1. IMPUTING FROM OTHER DATASETS
2. INTERPOLATE
3. FILLNA AND DERIVE
4. REMOVE COLUMNS, NARROW RANGE

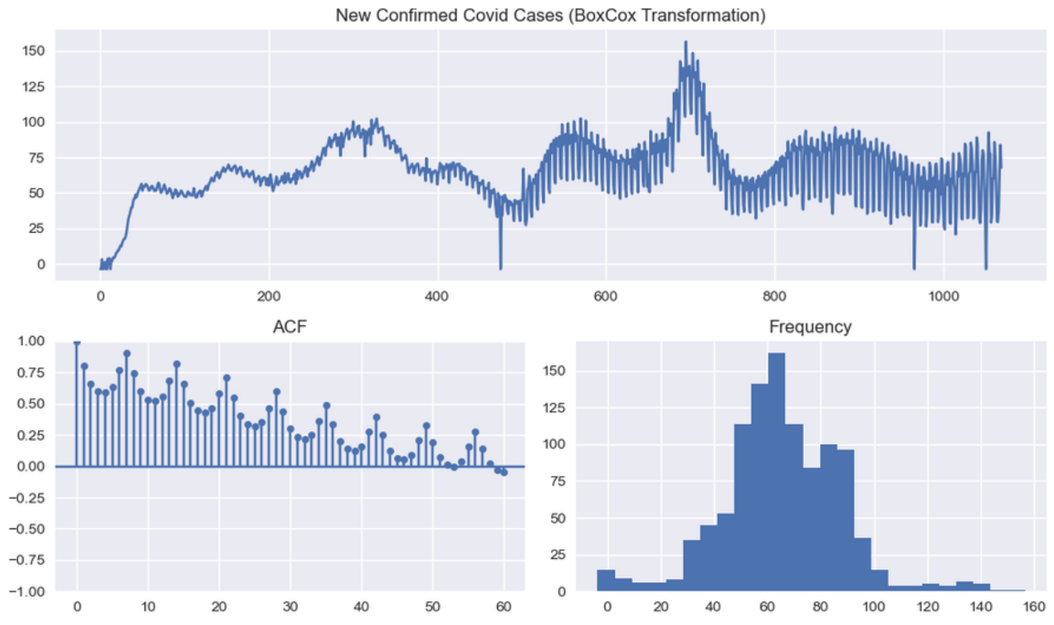
# TIMES SERIES PREPROCESSING FOR SARIMAX

**d:**
**Find the Difference**

**q, p:**
**ACF, PACF**

**P, D, Q, M:**
**Seasonality**



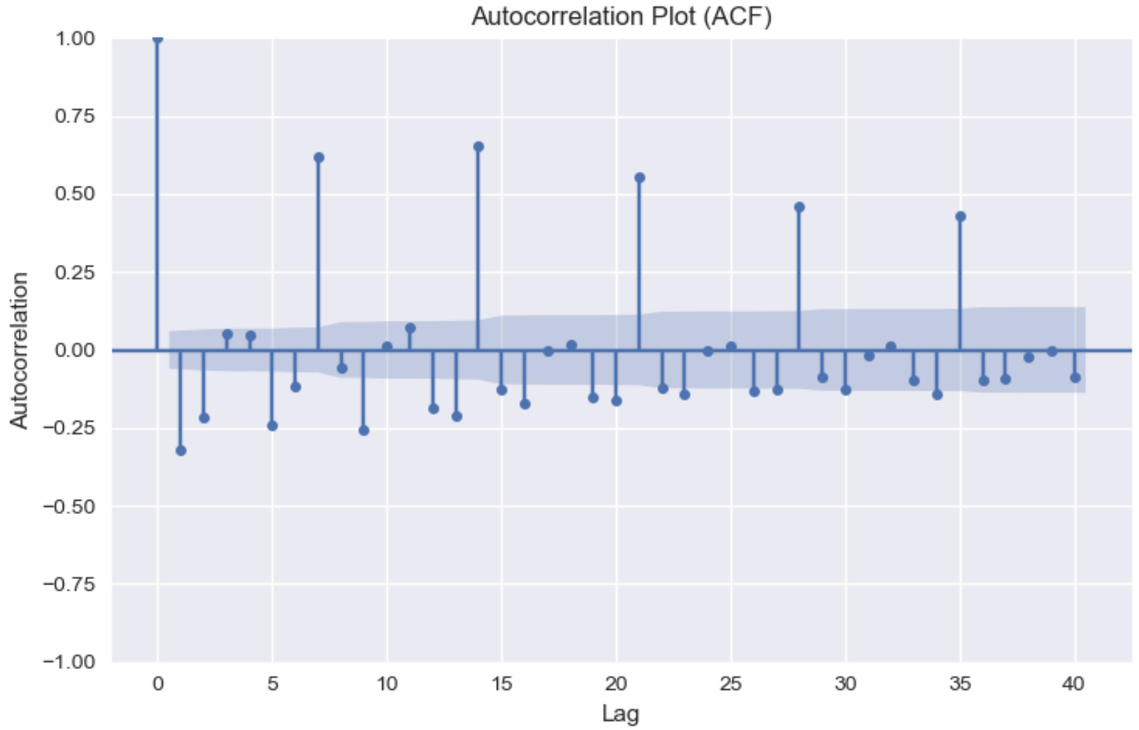
# FINDING THE ORDER OF ARIMA

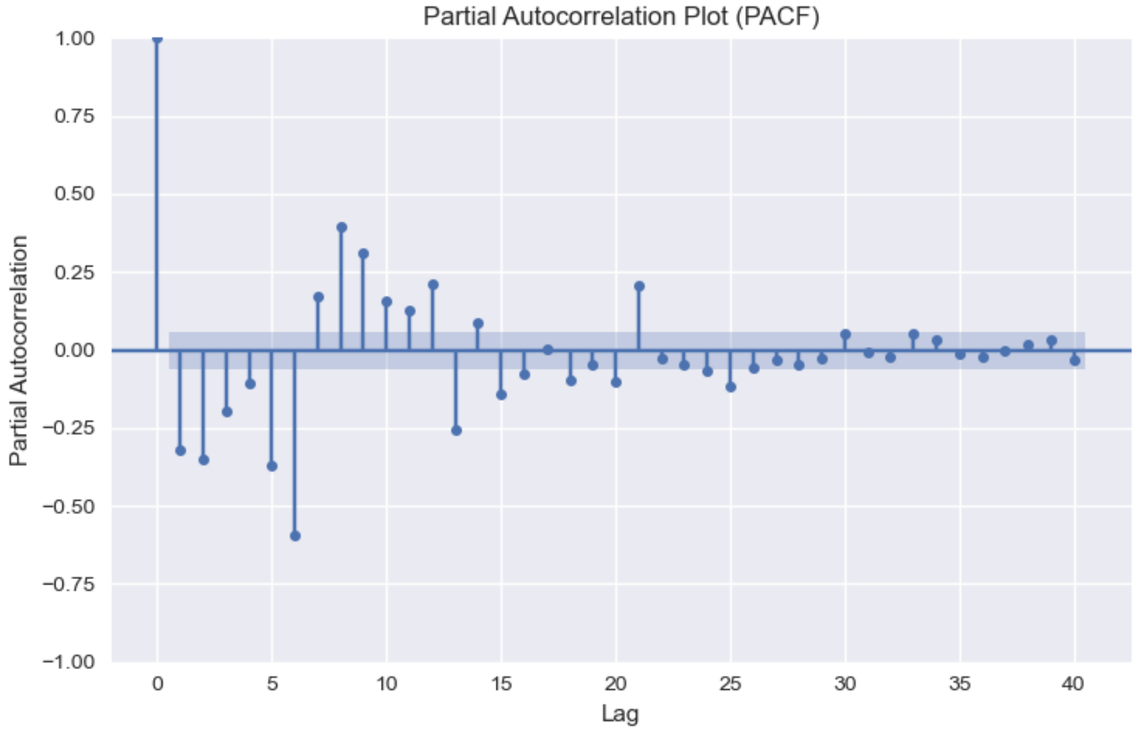
## AUTOCORRELATION

- Decays relatively quick
- q = 2-3 Moving Average Lags

## PARTIAL AUTOCORRELATION

- Remains significant out to several terms
- Seasonal effect likely important
- p = 5-6 Autoregression Lags



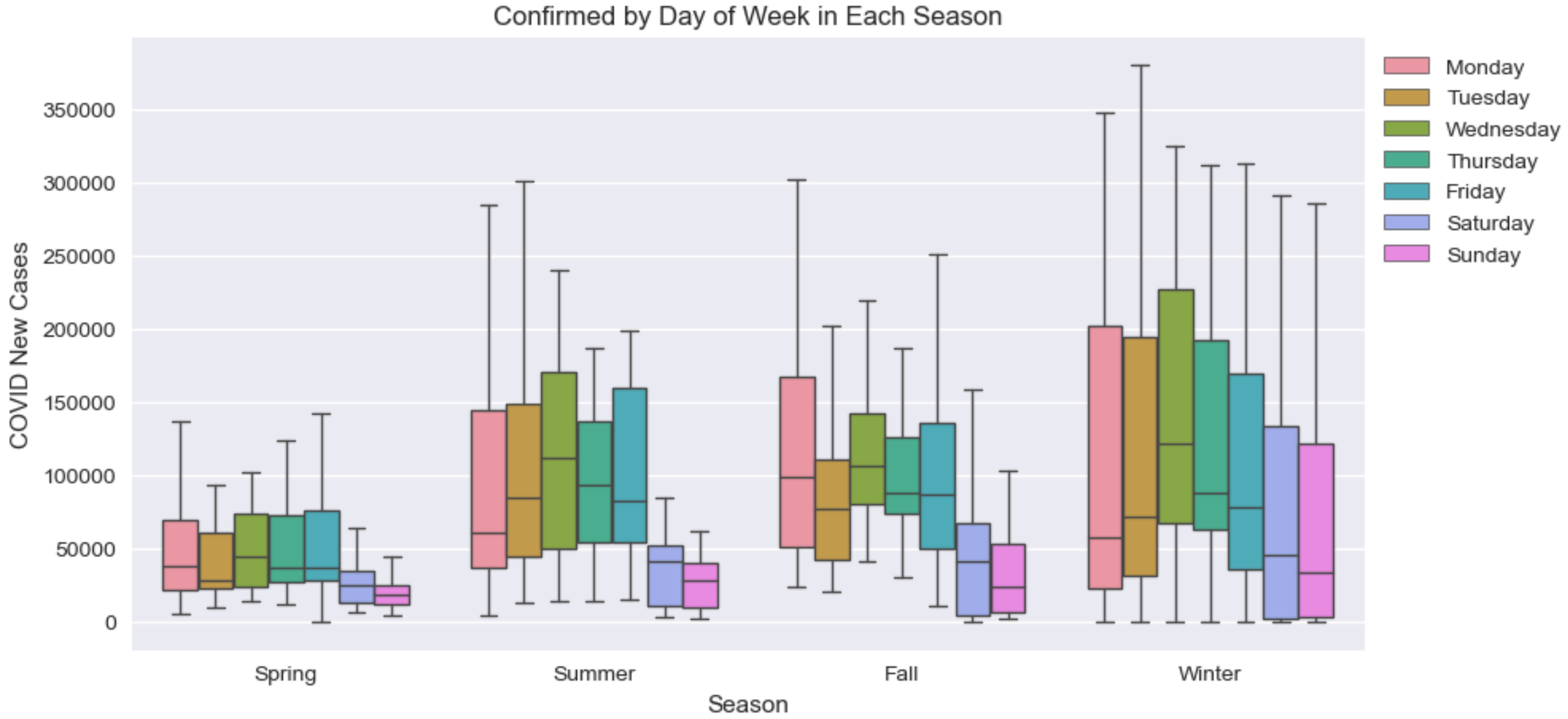
# SEASONALITY

## WEEKLY CORRELATION

- Rise and Fall throughout every week
- Wednesday is peak, Sunday is trough
- m = 7 day periods

## QUARTERLY CORRELATION

- Winter sees higher number and more variance
- Spring is consistently the lowest



# SARIMAX MODEL

**p, d, q**
**3, 0, 2**

## ARIMA ORDER

- Recent COVID data was predictive

**P, D, Q, M**
**2, 1, 1, 7**

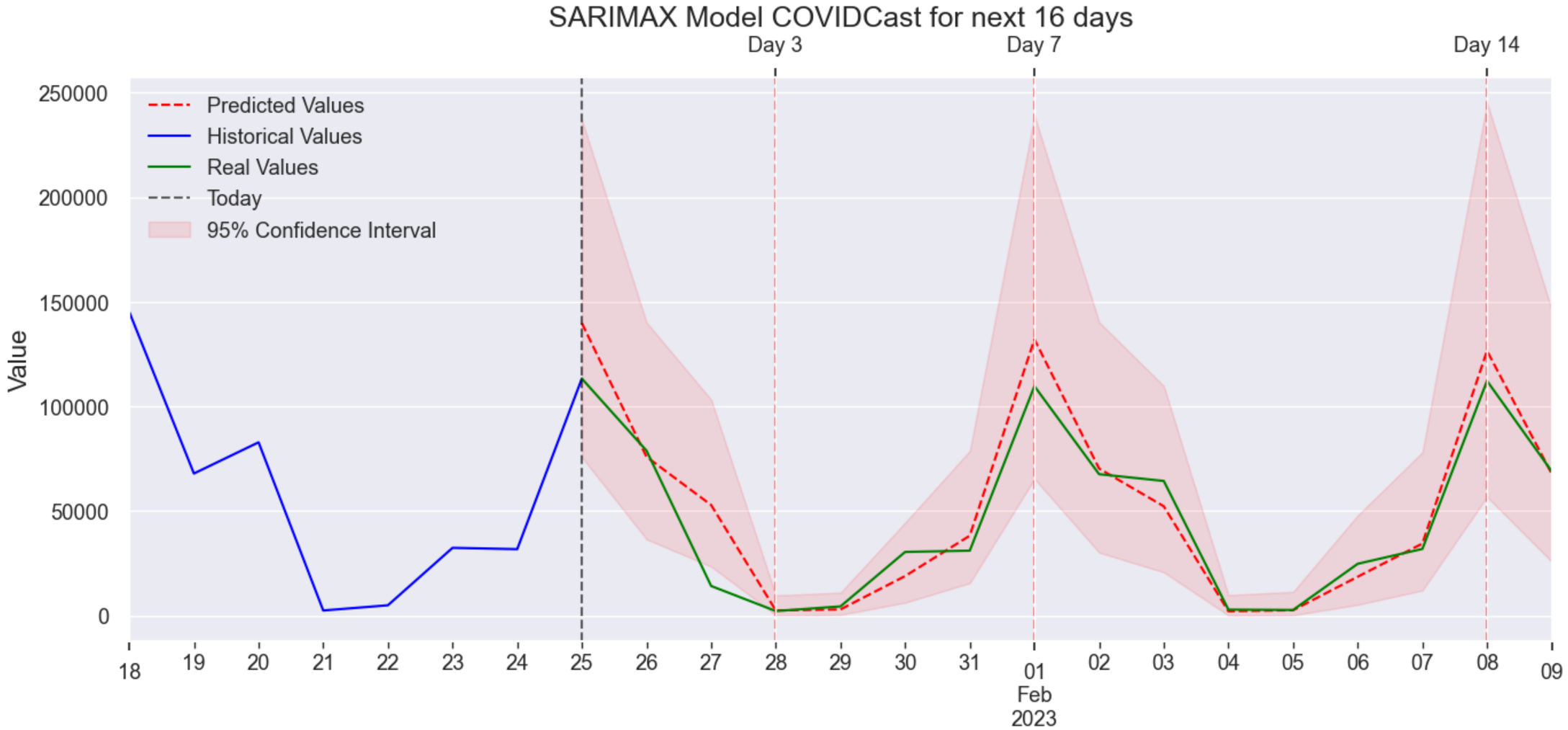
## SEASONAL ORDER

- Strong weekly effects

**6**

## EXOGENOUS VARIABLES

Stringency, r, vaccinations, hospitalized patients, mortality, month



# PROPHET MODEL

10

## CHANGE POINT

- Very Flexible

0.1

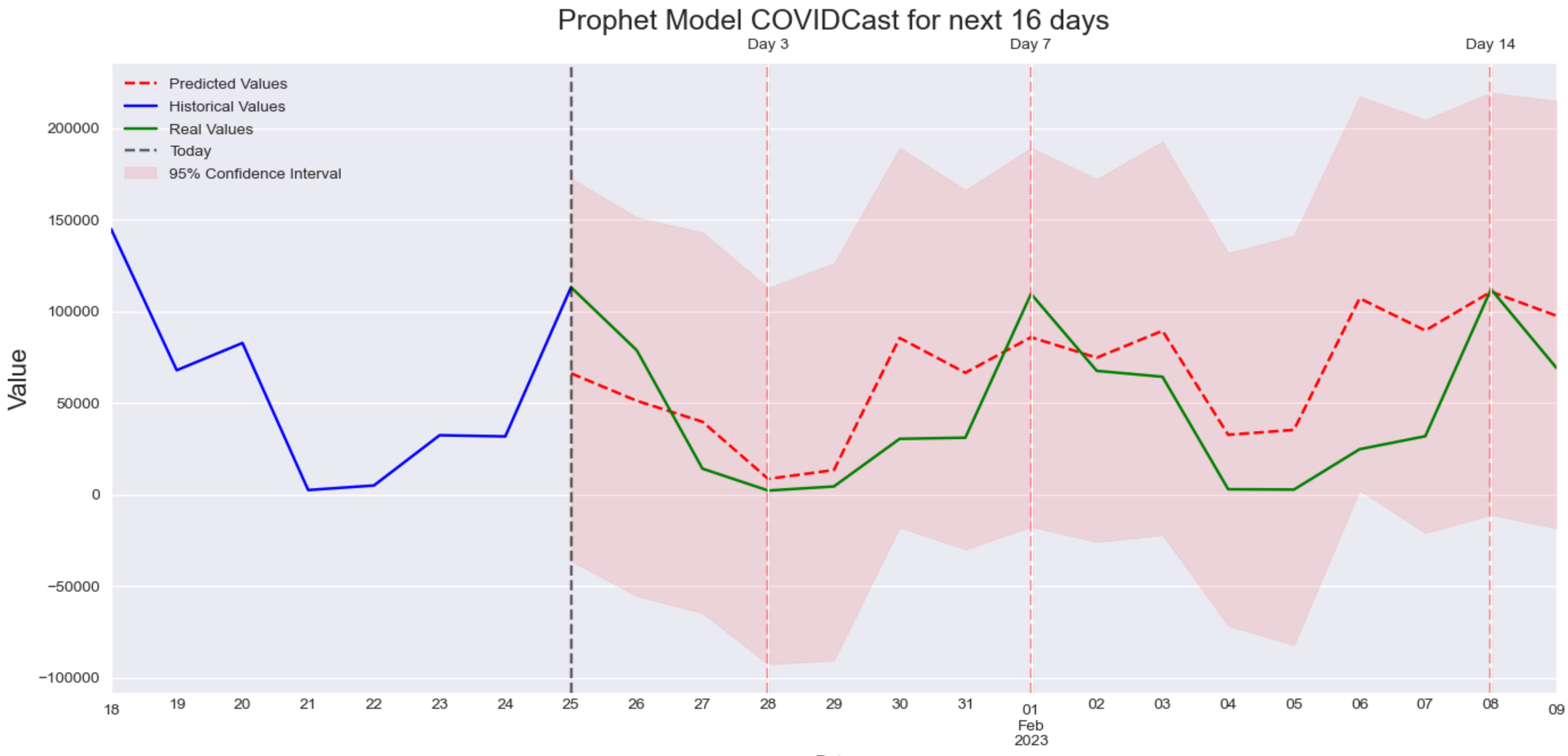
## SEASONALITY

- Strong weekly effects

19

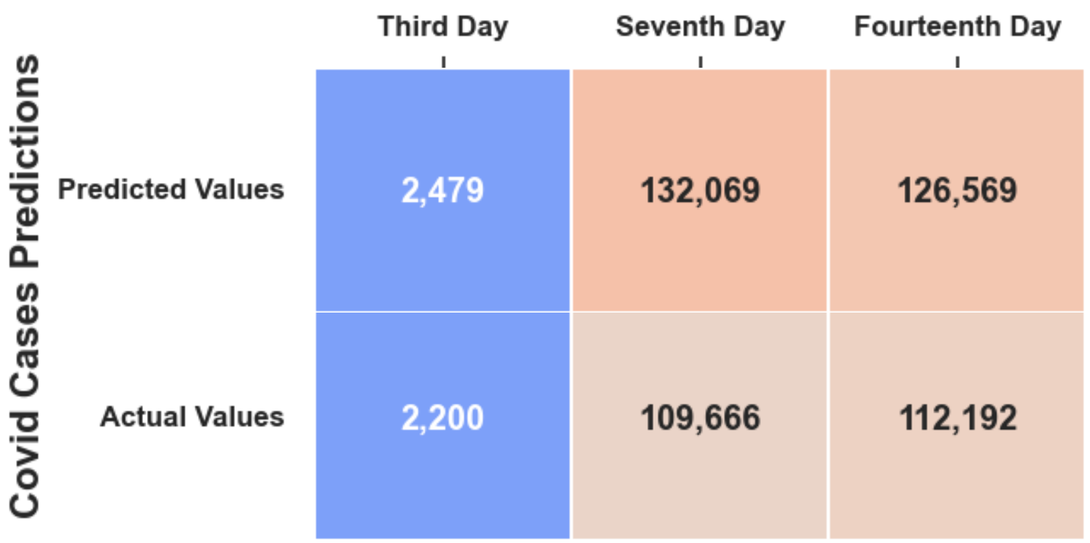
## EXOGENOUS VARIABLES

- Hospitalizations, and SIRD parameters



# COVID CAST

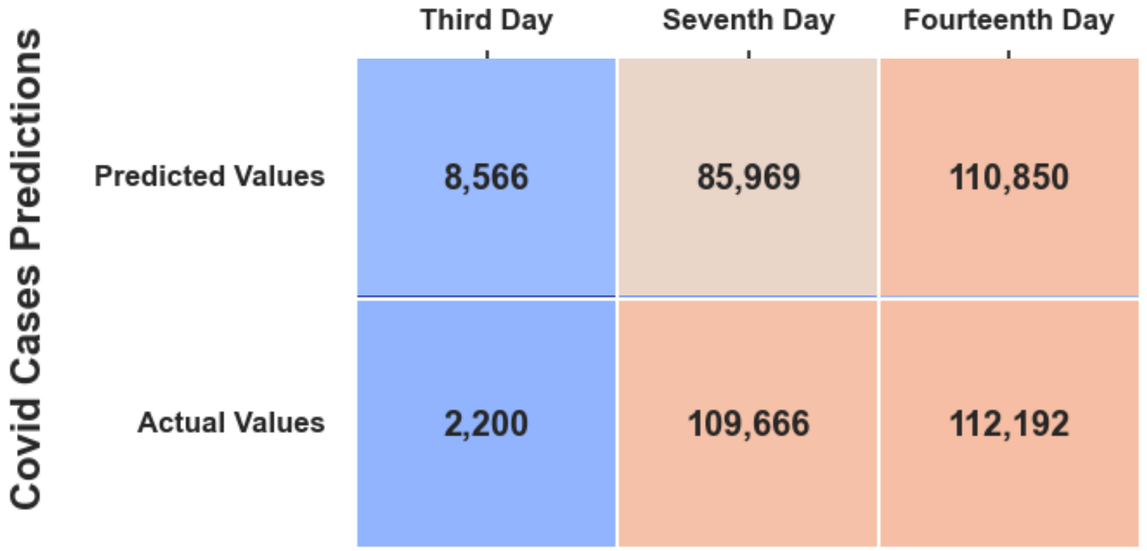
## SARIMAX MODEL




**Model's Cross Validation Scores over a 14 day Time Horizon**

**Mean Absolute Error:** 38,691 Covid Cases

## PROPHET MODEL




**Model's Cross Validation Scores over a 14 day Time Horizon**

**Model Coverage:** 75%

**Mean Absolute Error:** 56,582 Covid Cases

# HEAD TO HEAD

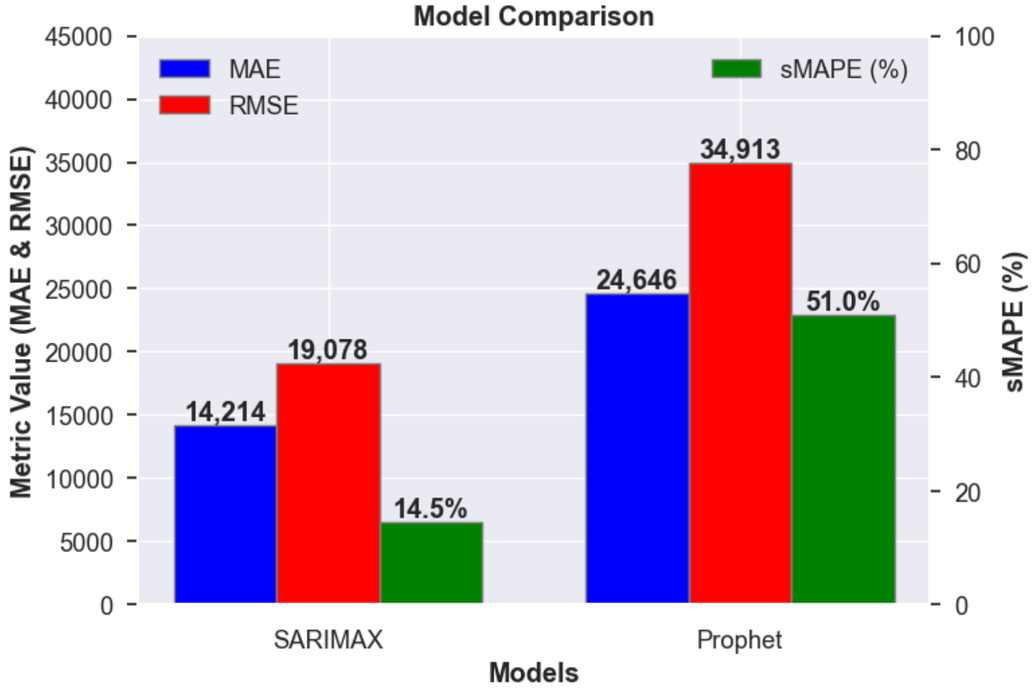
## Performance on Unseen Data

### *SARIMAX*

**Order**: (3, 0, 2)

**Season**: (2, 1, 1, 7)

**Exog:** 6



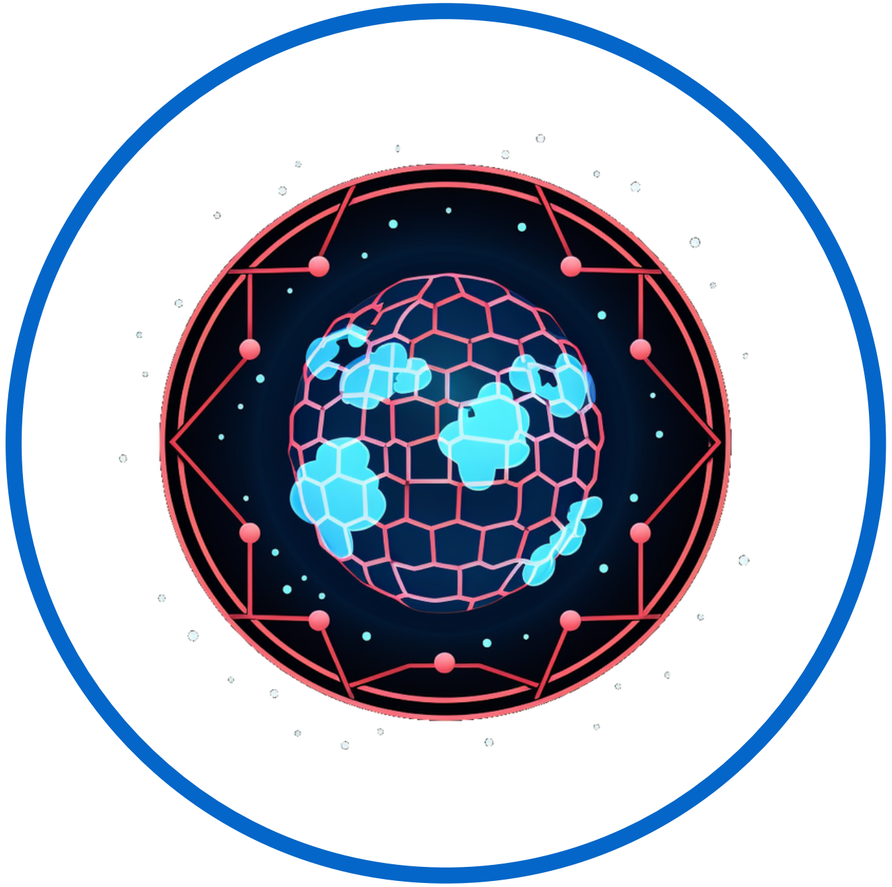

### *PROPHET*

**Change Points**: 10

**Seasonal**: 0.1

**Exog:** 19

# Next Steps...



1. DEPLOY
2. CHANGE TARGET VARIABLE
3. RNNS
4. PREDICT TO PROTECT

# THANK YOU

BRAINSTATION

## REACH OUT TO SAM:

SCELAREK@GMAIL.COM
LINKEDIN.COM/IN/SAM-CELAREK/
GITHUB.COM/SCELAREK