Forecasting and Monitoring Hospitalization Trends from Respiratory Viruses in Kenya: A Time Series Analysis by County

Group 10

Okech Maxmillan

Diana Ogeto

Lydia Mangoa

Eric Miriti

Christine Ambasa

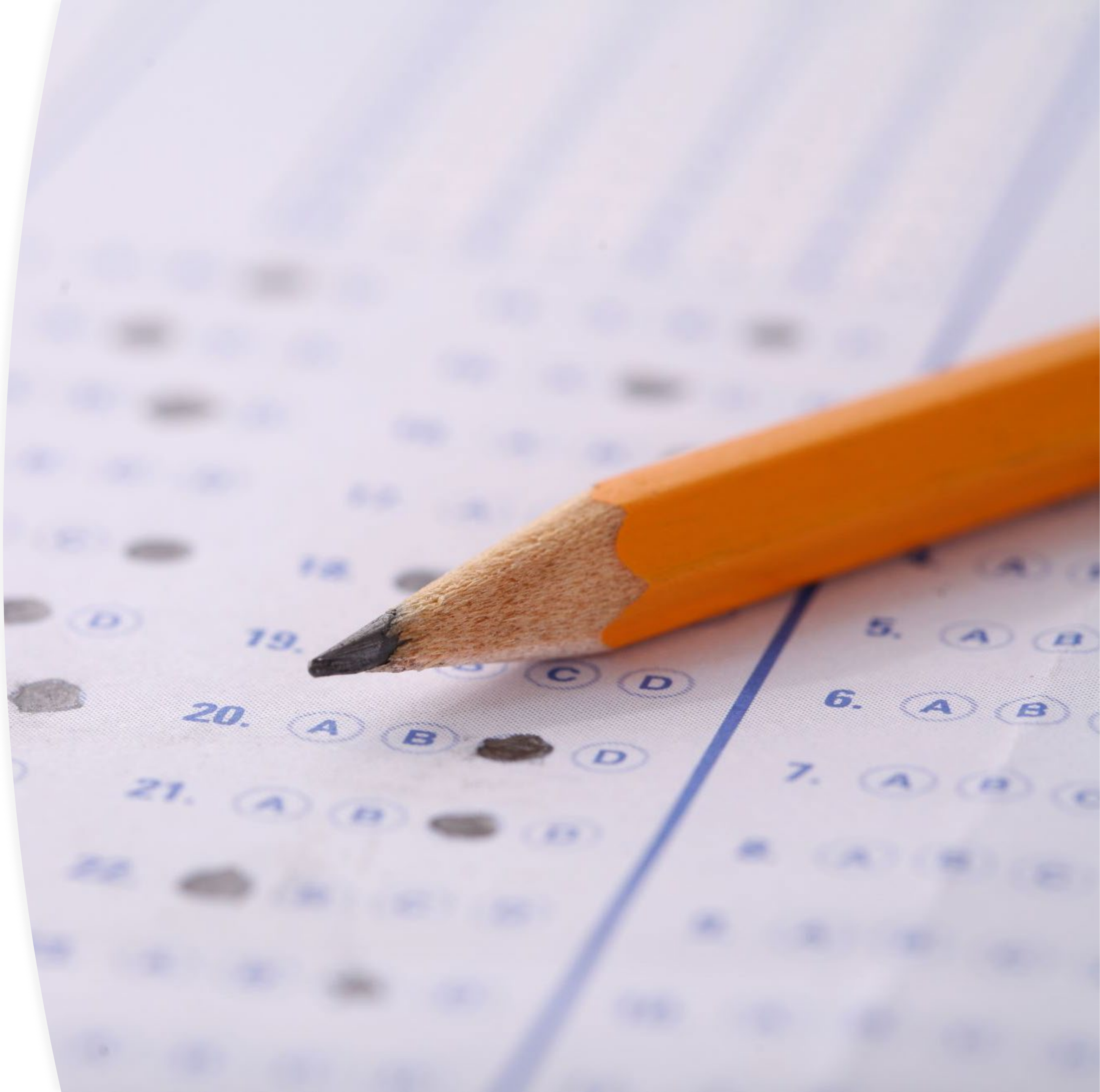
Business Understanding

Demonstrate how data-driven forecasting using U.S. hospitalization data can help Kenya anticipate respiratory virus surges, improve resource allocation, and strengthen public health response once local data systems are in place.

Problem Statement

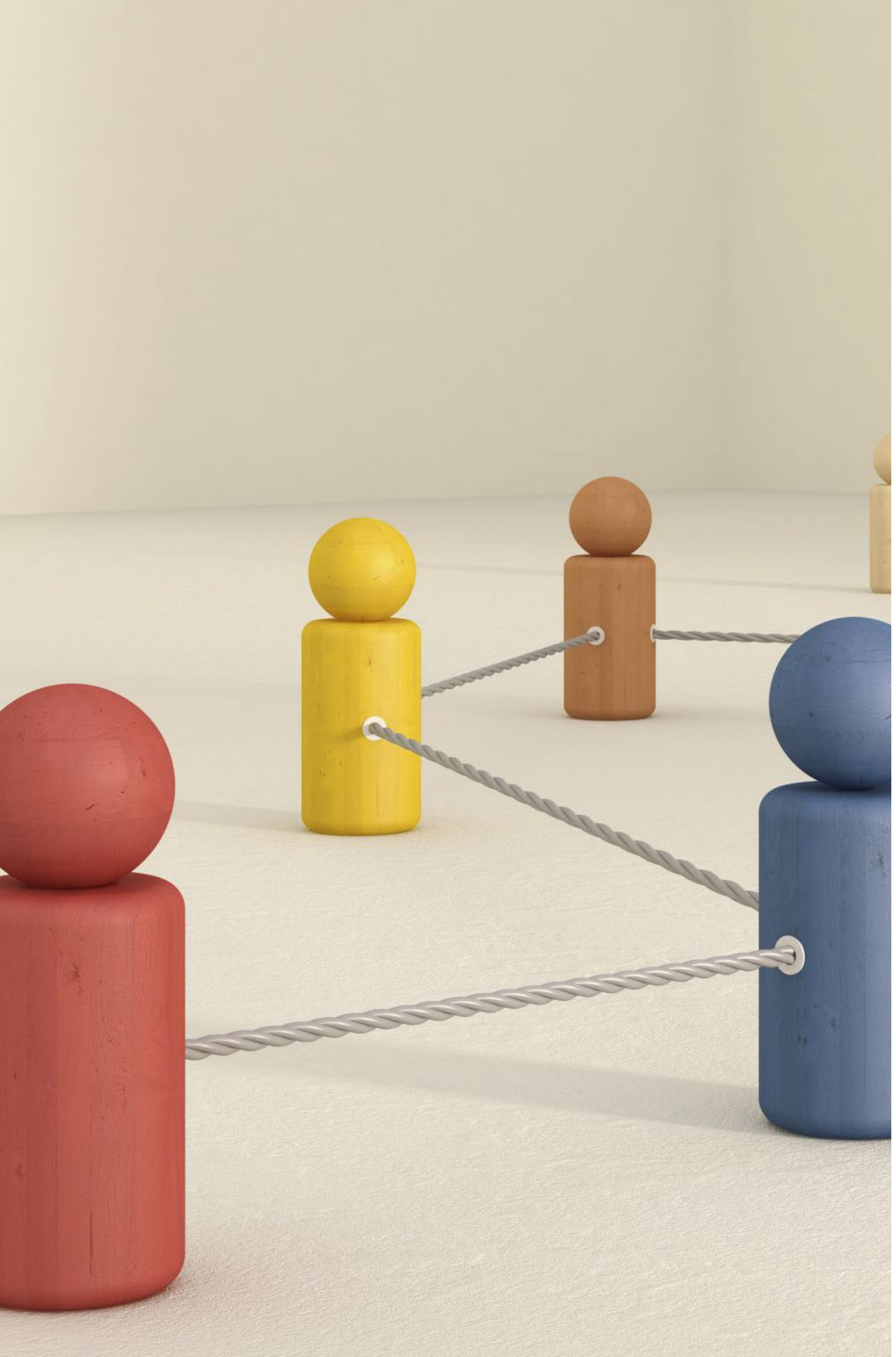
Kenya lacks real-time surveillance for respiratory virus hospitalizations.

This project uses U.S. data to demonstrate how forecasting can guide future public health response and planning once local data systems are established.



Objectives

- Monitor, model, and forecast respiratory virus-related hospitalizations across U.S. regions
- Support healthcare planning, outbreak response, and policy-making
- Detect seasonal patterns in hospitalization trends
- Forecast future spikes or declines in respiratory cases
- Provide real-time dashboards for public health decision-making



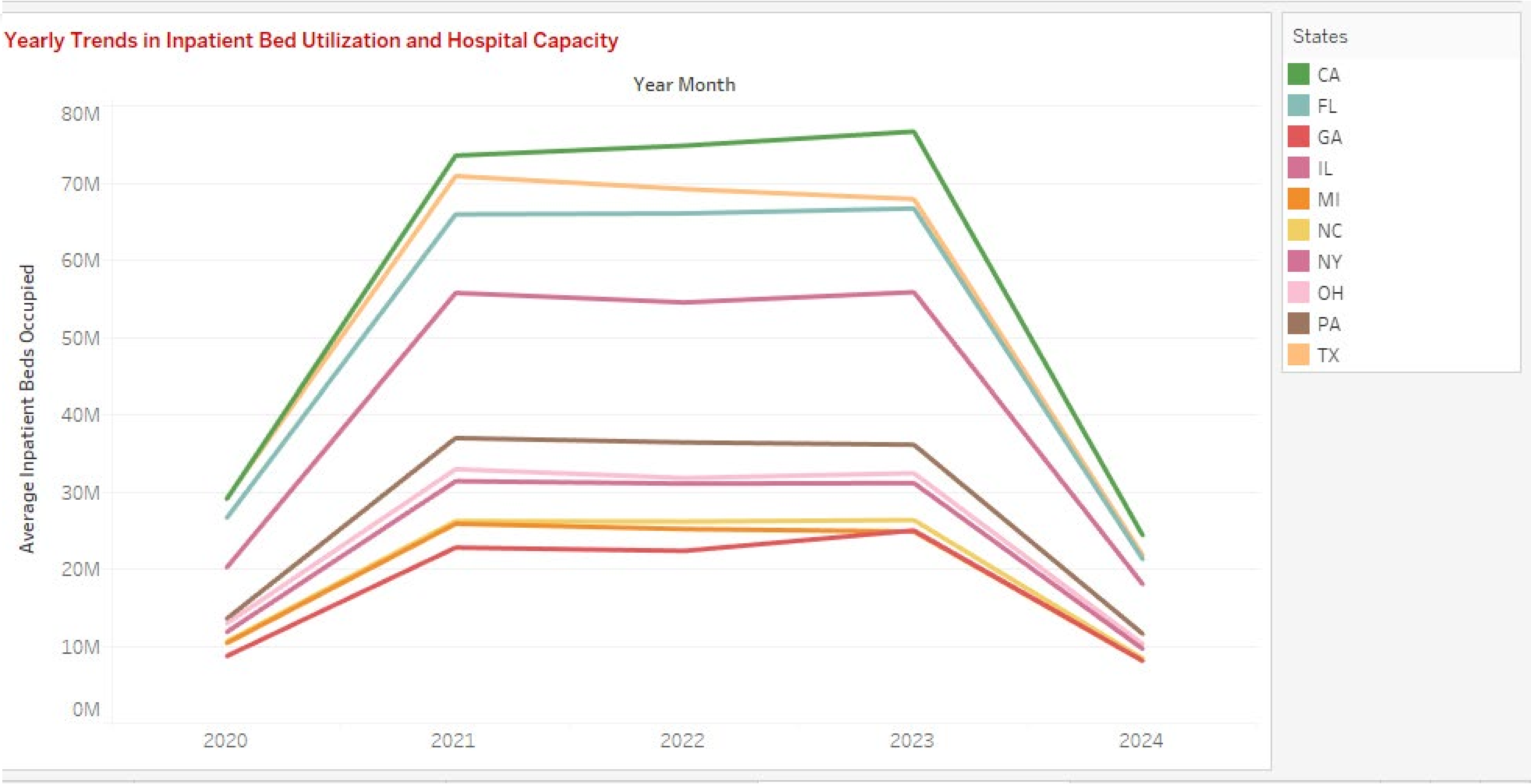
Stakeholders

1. MOH and local health departments
2. Hospital administrators and planners
3. Epidemiologists and public health researchers
4. Policy-makers and emergency response teams

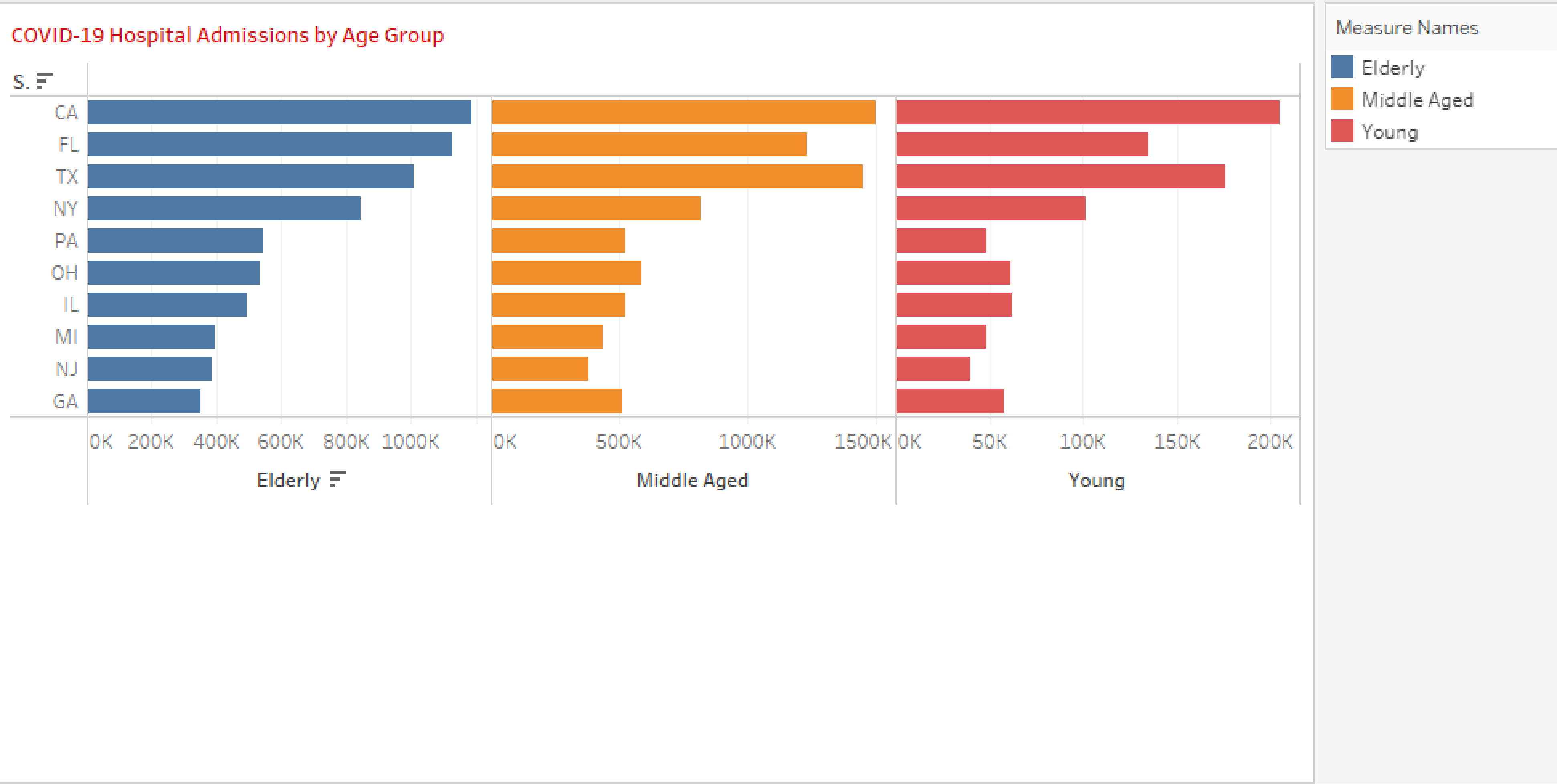
Data Understanding

- This project leverages a dataset derived from publicly available U.S. health records on the US.gov repository.
- **Weekly COVID-19 hospital admissions – used as the primary target variable.**
- **Handling missing values and skewness of the dataset through imputation using the mean and median.**

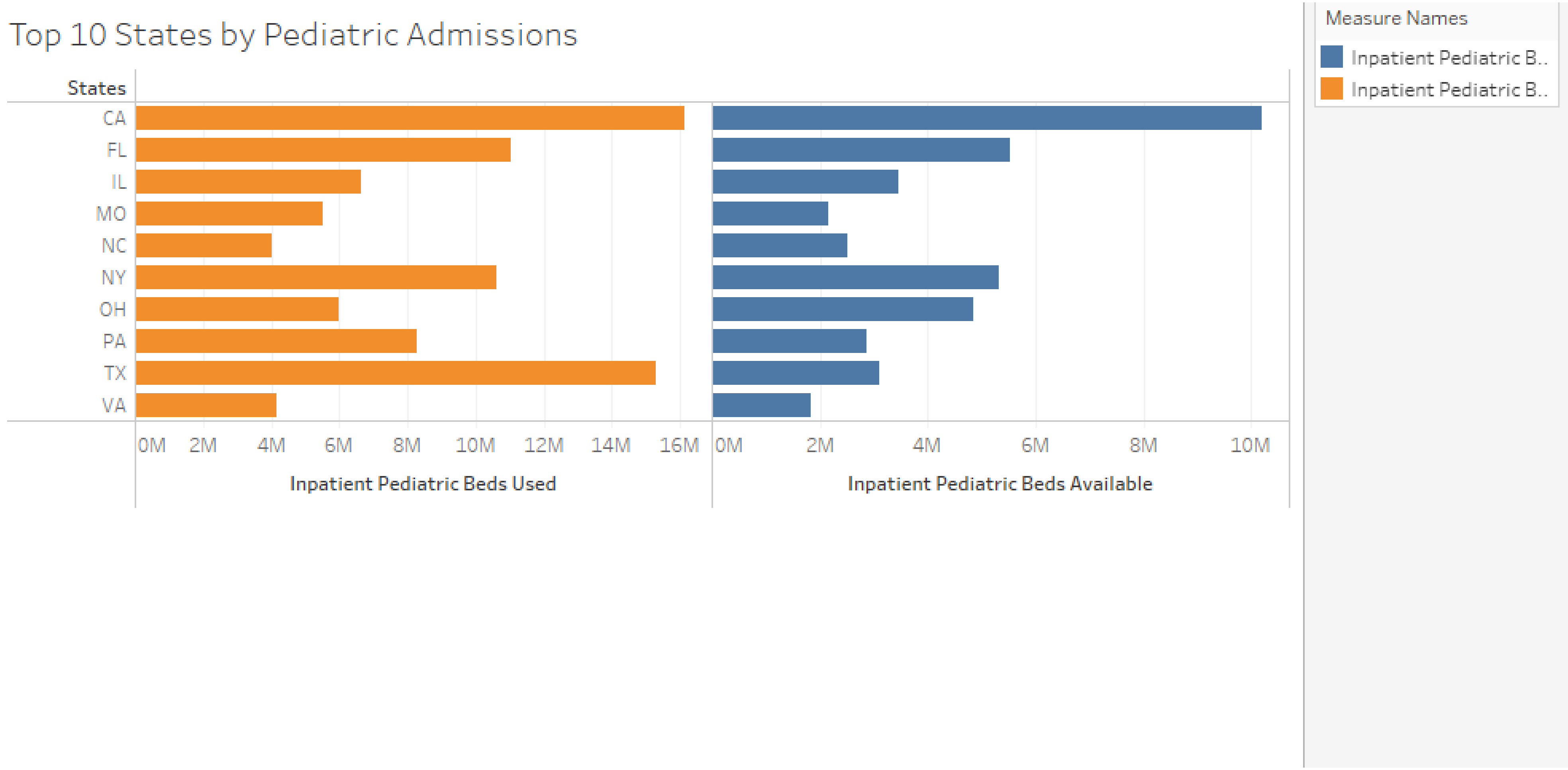
EDA (Explorative Data Analysis)



EDA Cont...

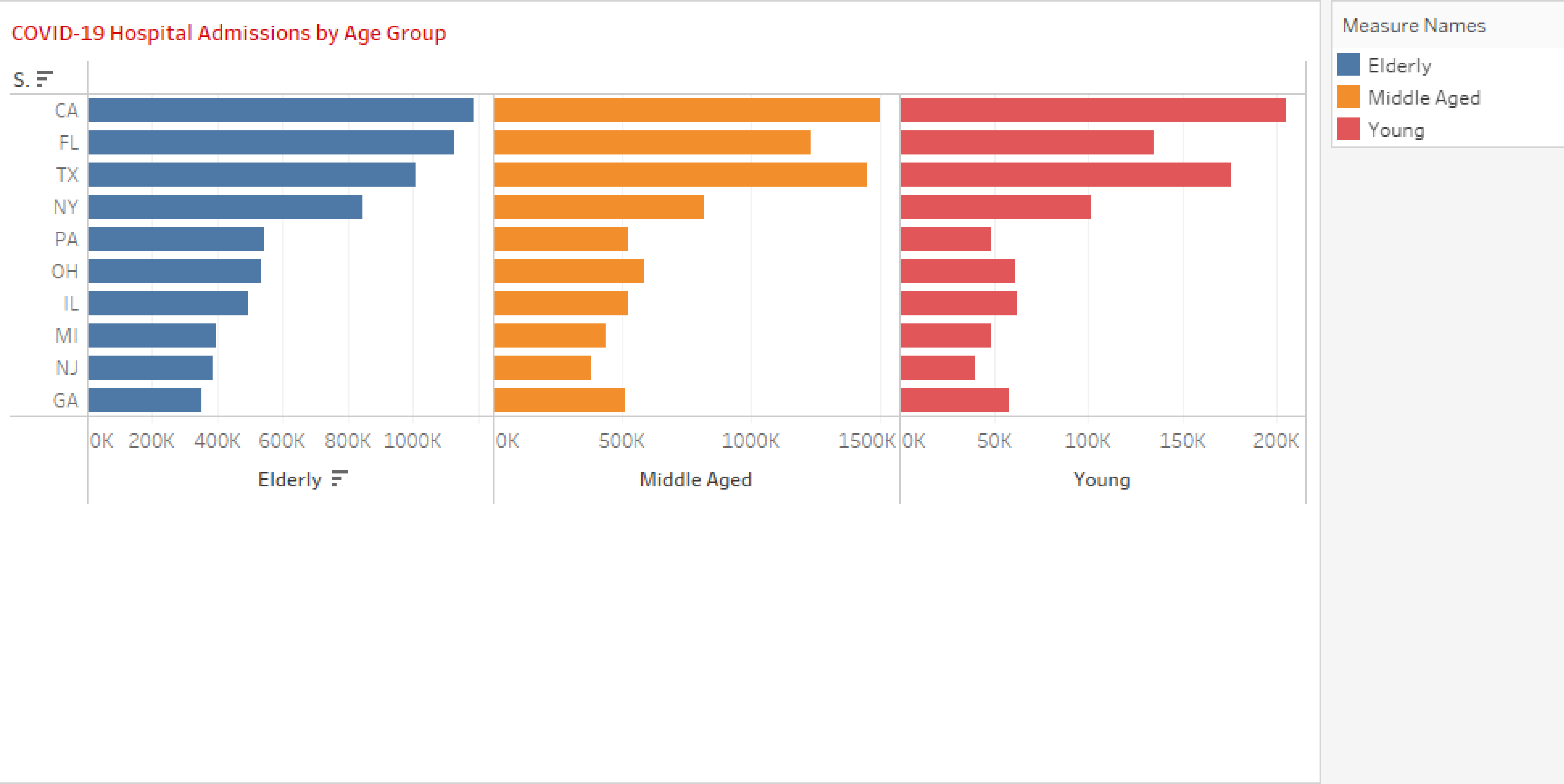

COVID-19 Hospital Admissions by Age Group
S.
CA
FL
TX
NY
PA
OH
IL
MI
NJ
GA
0K 200K 400K 600K 800K 1000K
Elderly
0K 500K 1000K 1500K
Middle Aged
0K 50K 100K 150K 200K
Young
Measure Names
Elderly
Middle Aged
Young

EDA cont...



Top 10 States by Pediatric Admissions
States
CA
FL
IL
MO
NC
NY
OH
PA
TX
VA
0M 2M 4M 6M 8M 10M 12M 14M 16M
Inpatient Pediatric Beds Used
0M 2M 4M 6M 8M 10M
Inpatient Pediatric Beds Available
Measure Names
Inpatient Pediatric B..
Inpatient Pediatric B..

EDA cont...



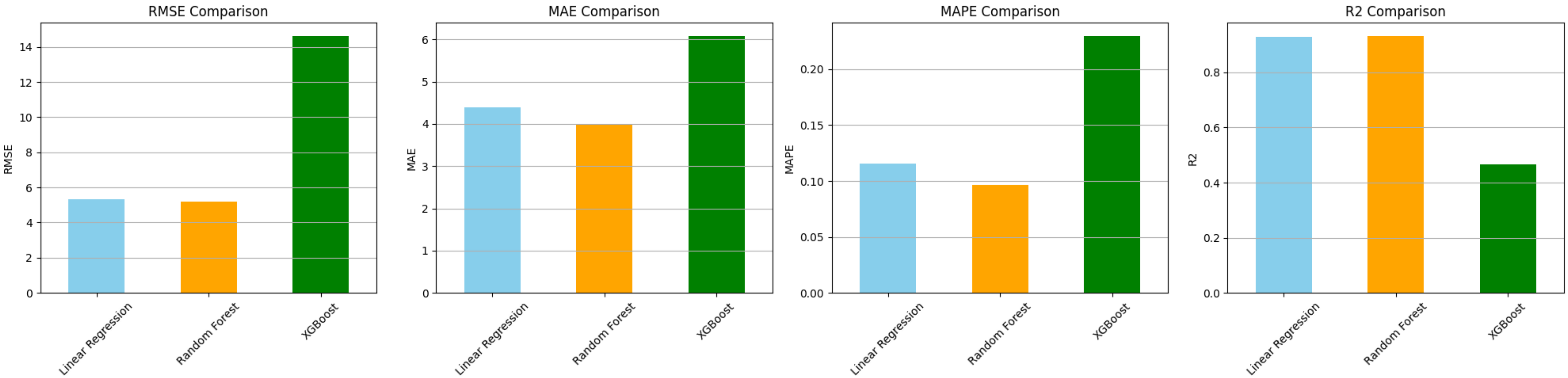
COVID-19 Hospital Admissions by Age Group
S.
CA
FL
TX
NY
PA
OH
IL
MI
NJ
GA
0K
200K
400K
600K
800K
1000K
Elderly
0K
500K
1000K
1500K
Middle Aged
0K
50K
100K
150K
200K
Young
Measure Names
Elderly
Middle Aged
Young

Feature Engineering and preprocessing

Comprehensive preprocessing included:

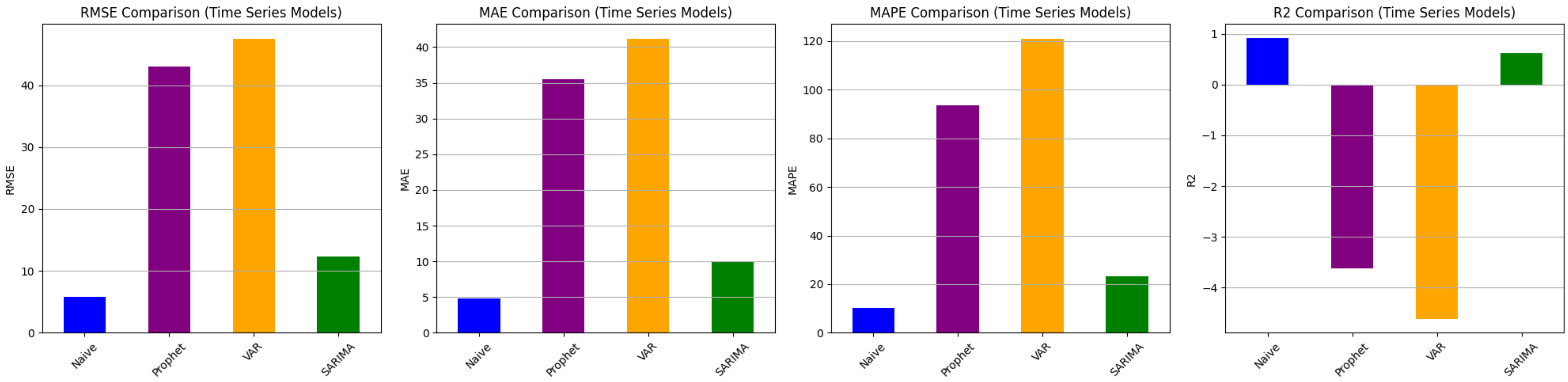
- Date time indexing
- ADF test – Check for stationarity for modeling (p-value <0.05)
- PACF and ACF test for determining the correct lag features and guide towards ARIMA/SARIMA and the parameters to be used in modeling
- Lag features (1 through to 5)
- Moving averages & Exponential smoothing - capture short term, medium term & long-term trends and filter out the noise.

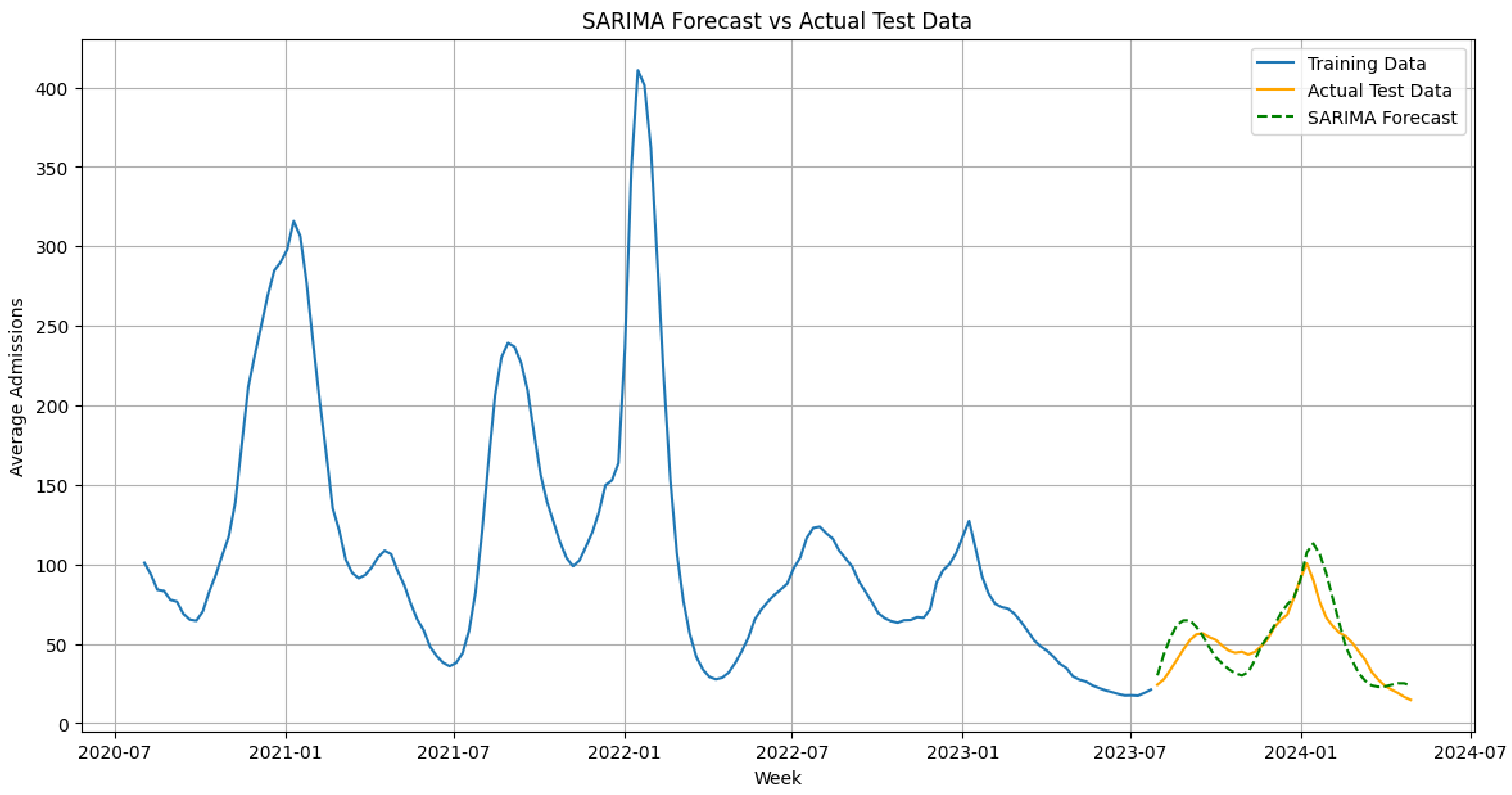
Modeling



RMSE Comparison
RMSE
0
2
4
6
8
10
12
14
MAE Comparison
MAE
0
1
2
3
4
5
6
MAPE Comparison
MAPE
0.00
0.05
0.10
0.15
0.20
R2 Comparison
R2
0.0
0.2
0.4
0.6
0.8
Linear Regression
Random Forest
XGBoost

Modeling Cont...





SARIMA Forecast vs Actual Test Data
Training Data
Actual Test Data
SARIMA Forecast
Average Admissions
Week
0
50
100
150
200
250
300
350
400
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01
2024-07

Evaluation Metrics

- The main evaluation metric is MAE which is complemented by RMSE.
- MAE still registers the error without letting it dominate overall metrics.
- RMSE heavily penalizes mispredictions during surges and this gives extra weight to bigger mistakes.

Deployment

• Increasing **q** smooths out noise by averaging more past errors, but too large can lag sudden changes.

Seasonal ARIMA Parameters

P (Seasonal AR) • Number of seasonal lag terms (e.g. how many years/weeks back). • Higher **P** lets model revisit seasonal spikes further in the past but may overfit seasonal noise.

D (Seasonal Diff) • Seasonal differencing order to remove seasonal trends. • Increasing **D** flattens seasonality more aggressively—helpful if patterns drift over time.

Q (Seasonal MA) • Moving-average on seasonal errors. • Larger **Q** smooths seasonal fluctuations, at the risk of missing sharp spikes.

S (Seasonal Period) • Length of

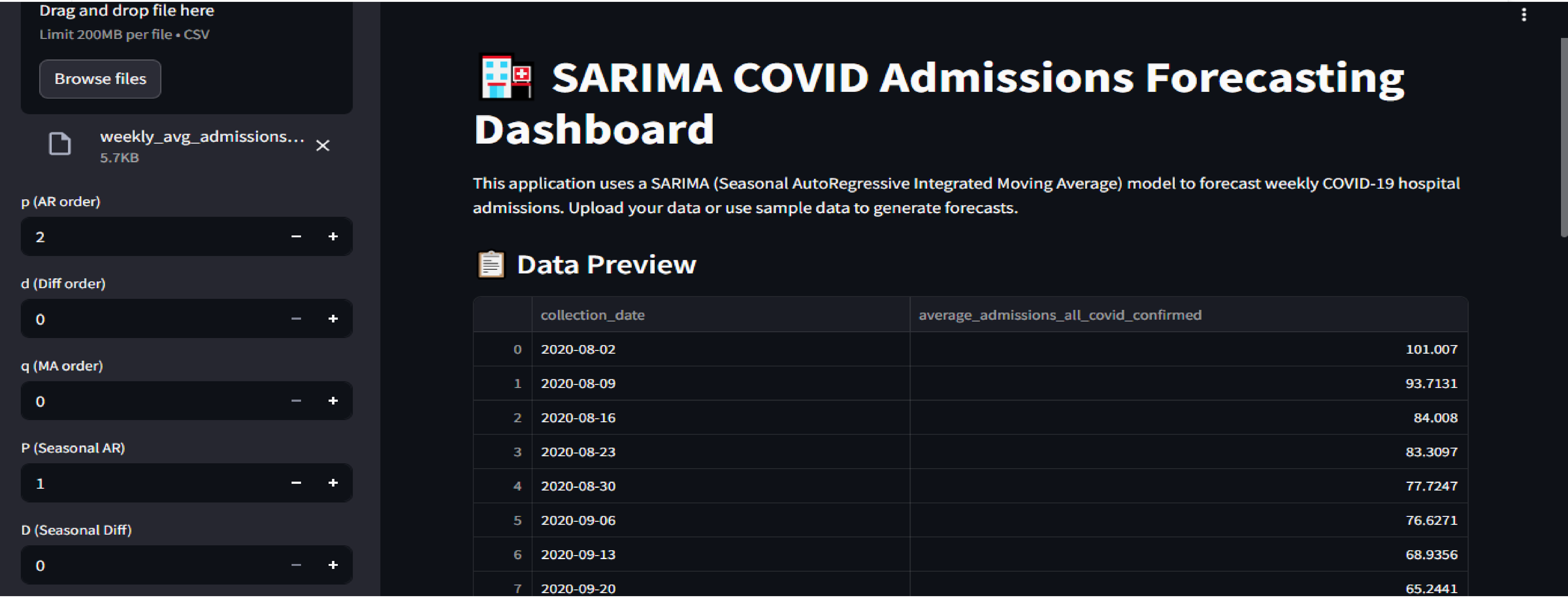
SARIMA COVID Admissions Forecasting Dashboard

This application uses a SARIMA (Seasonal AutoRegressive Integrated Moving Average) model to forecast weekly COVID-19 hospital admissions. Upload your data or use sample data to generate forecasts.

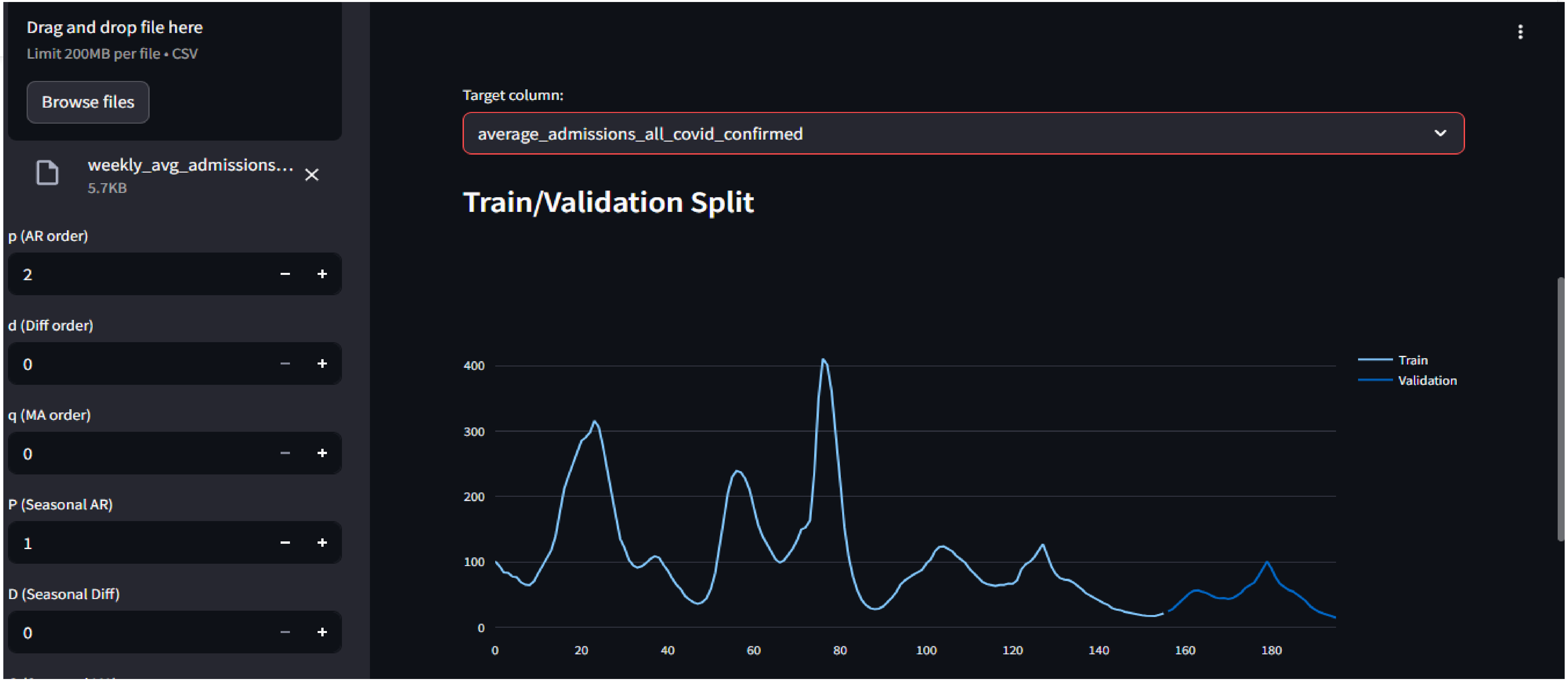
Upload a CSV to get started.

Tip: Run this in Colab with ngrok to share!

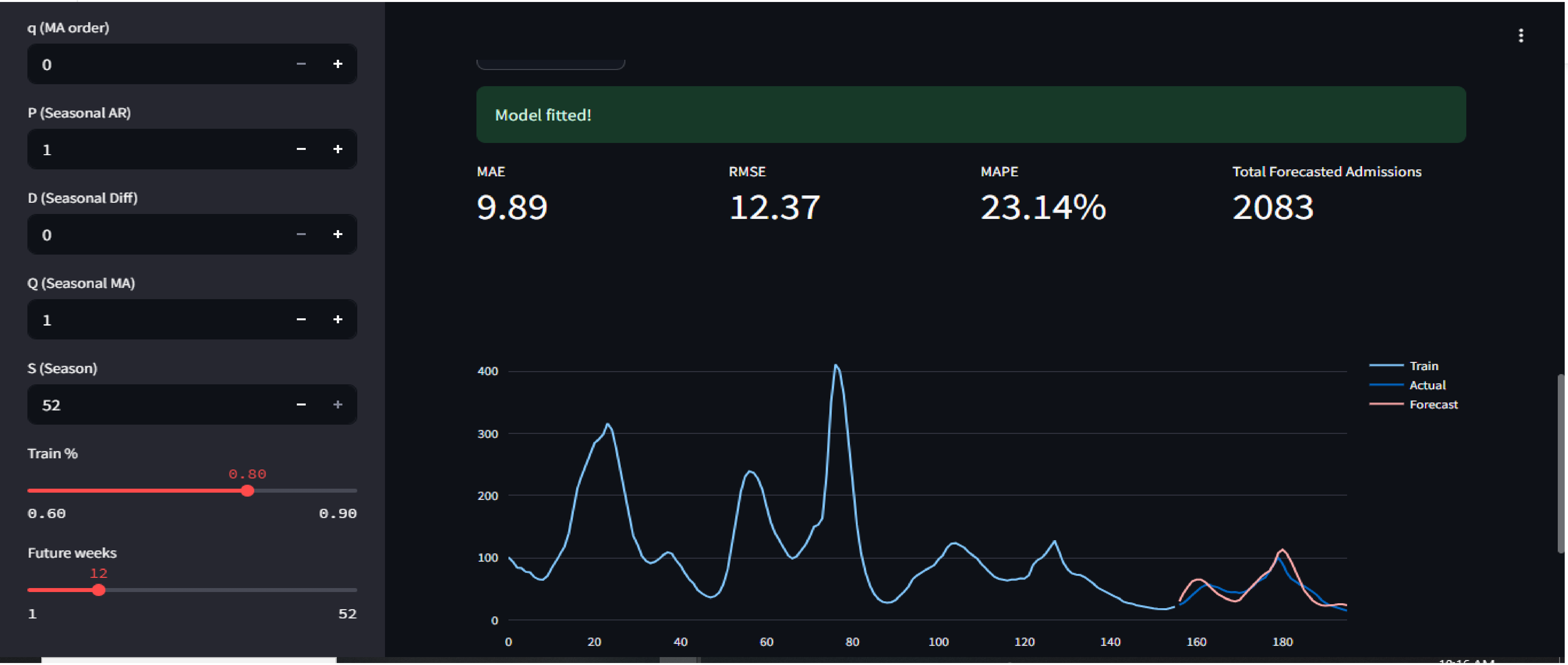
Deployment Cont...



Deployment Cont...



Deployment Cont...



Conclusion & Recommendations

Conclusion

- Successfully integrated monitoring, modeling, and forecasting of COVID-19 hospitalizations.
- Developed real-time dashboards for timely communication with public health stakeholders.
- Produced short- and medium-term forecasts to support healthcare resource planning and policy decisions.
- Identified seasonal patterns and trends across U.S. states.

Recommendations

- Kenya should prioritize developing a centralized, real-time health data collect
- Institutionalize Seasonal Surveillance and Early Warning Systems
- Optimize Resource Allocation Based on Forecast Insights
- Training for Local Public health agencies

THANK YOU!