# NYC Homeless Population Forecasting

**Presented by:**
**Dae Han**

# Predicting future demand improves service quality

I used **Department of Homeless Service (DHS) daily report**
to create **a time series model** that forecasts the number of individuals
sheltered by DHS.

This is to predict the demand for shelter spaces
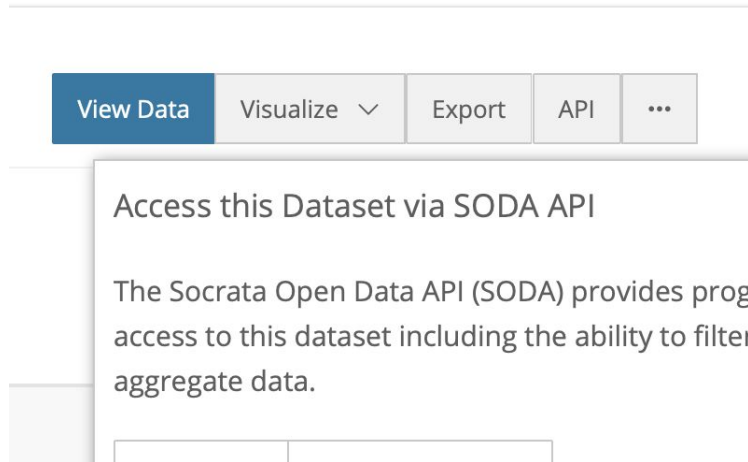in order to better **support homeless communities**.

# Demand for shelter spaces is higher than it ever was

"Homelessness in New York City has reached the highest levels since the Great Depression in the 1930s"

- Coalition for the homeless

# DHS Daily Report is scraped using SODA API

**View Data** | Visualize ∨ | Export | API | ⋯

Access this Dataset via SODA API

The Socrata Open Data API (SODA) provides prog
access to this dataset including the ability to filter
aggregate data.

## DHS Daily Report

- Statistics about homelessness

- Daily updated from 2013

- Assessed through Socrata Open Data (SODA) API

# DHS Daily Report is scraped using SODA API
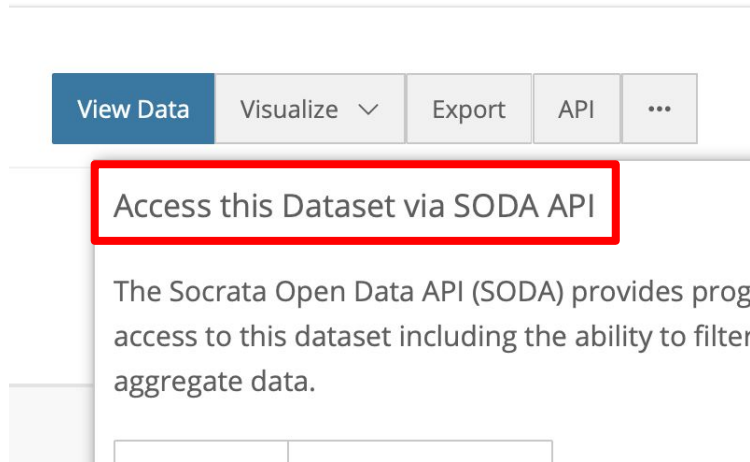


## DHS Daily Report

- Statistics about homelessness

- Daily updated from 2013
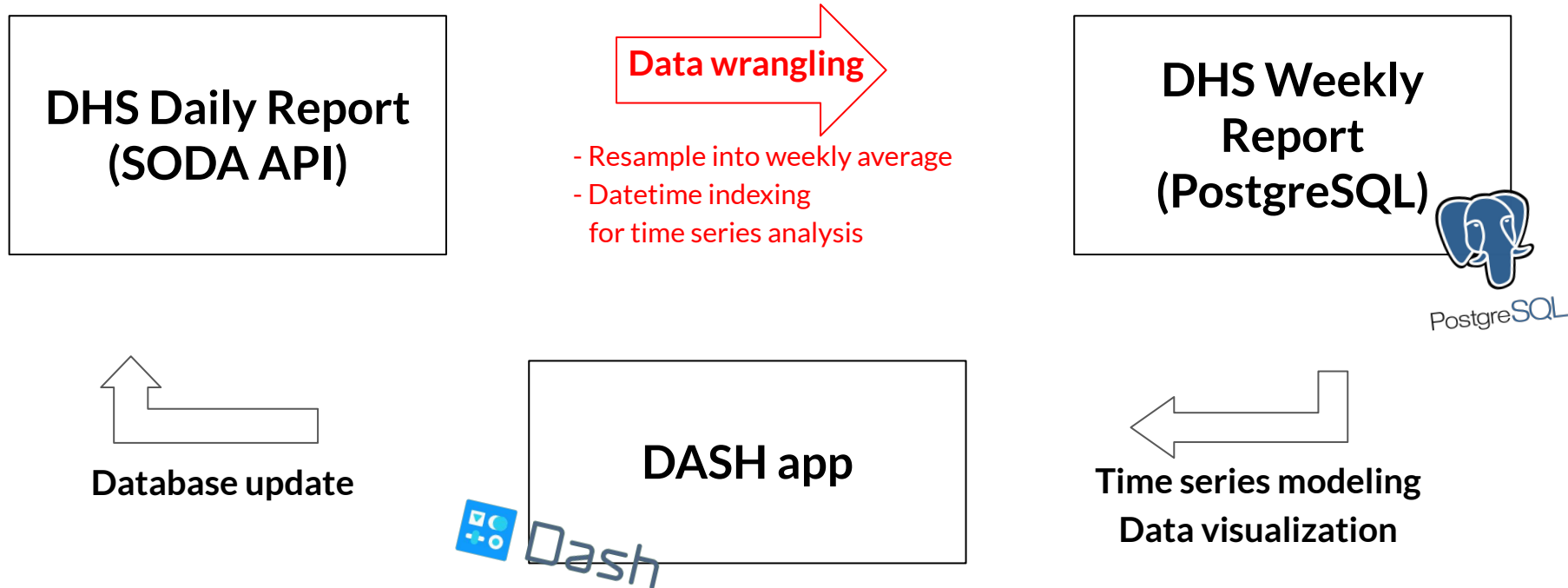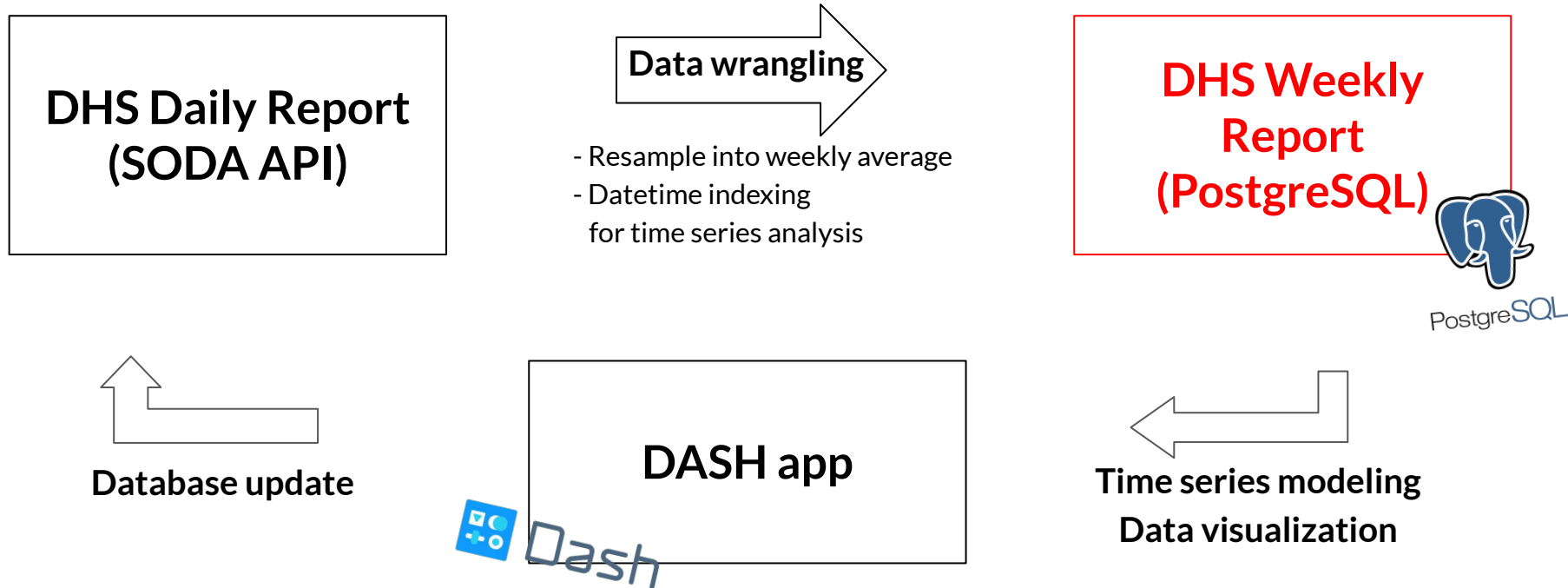
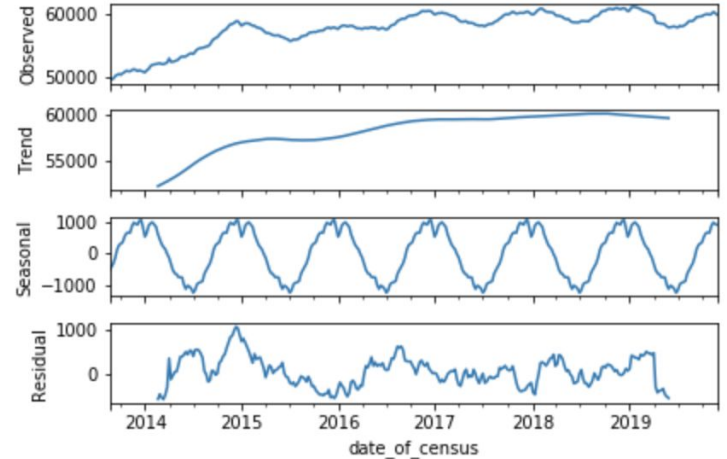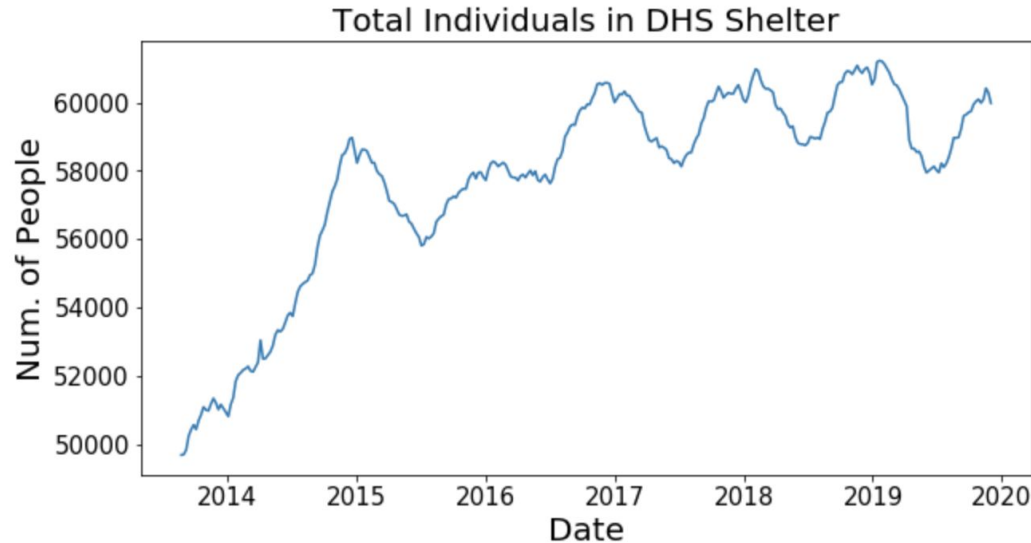- Assessed through Socrata Open Data (SODA) API
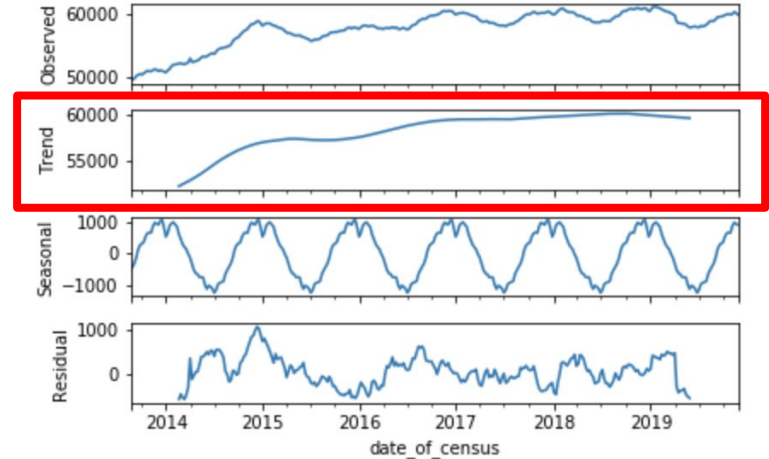
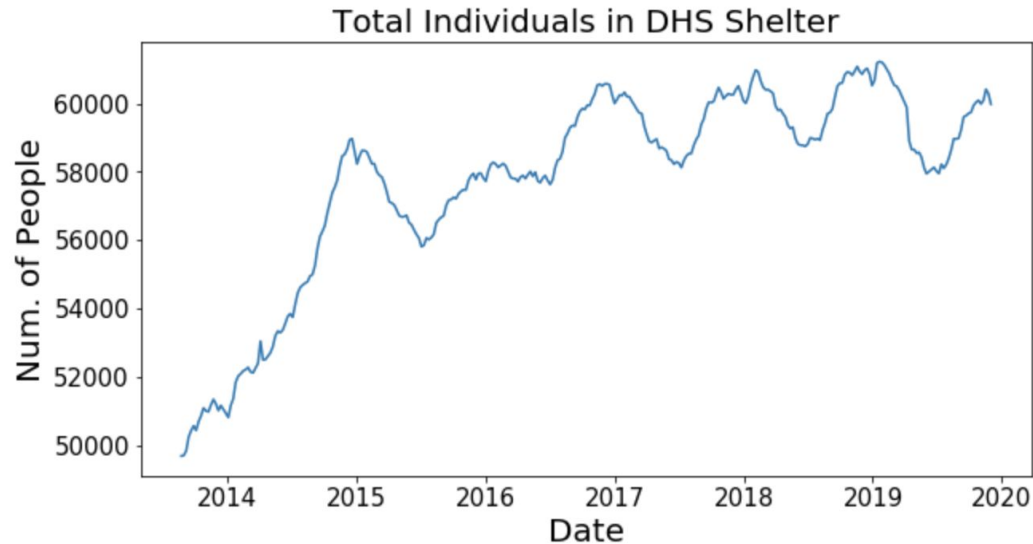# Database was used to accumulate cleaned data

**DHS Daily Report (SODA API)**

**Data wrangling**

- Resample into weekly average
- Datetime indexing
  for time series analysis

**DHS Weekly Report (PostgreSQL)**

PostgreSQL

**Database update**

**DASH app**

Dash

**Time series modeling**
**Data visualization**

# Database was used to accumulate cleaned data

**DHS Daily Report (SODA API)**

**Data wrangling**

- Resample into weekly average
- Datetime indexing
  for time series analysis

**DHS Weekly Report (PostgreSQL)**

PostgreSQL

**Database update**

**DASH app**

Dash

**Time series modeling**
**Data visualization**

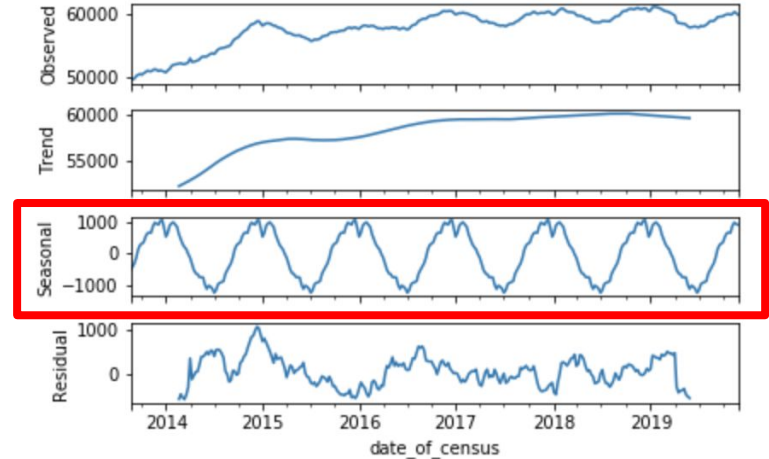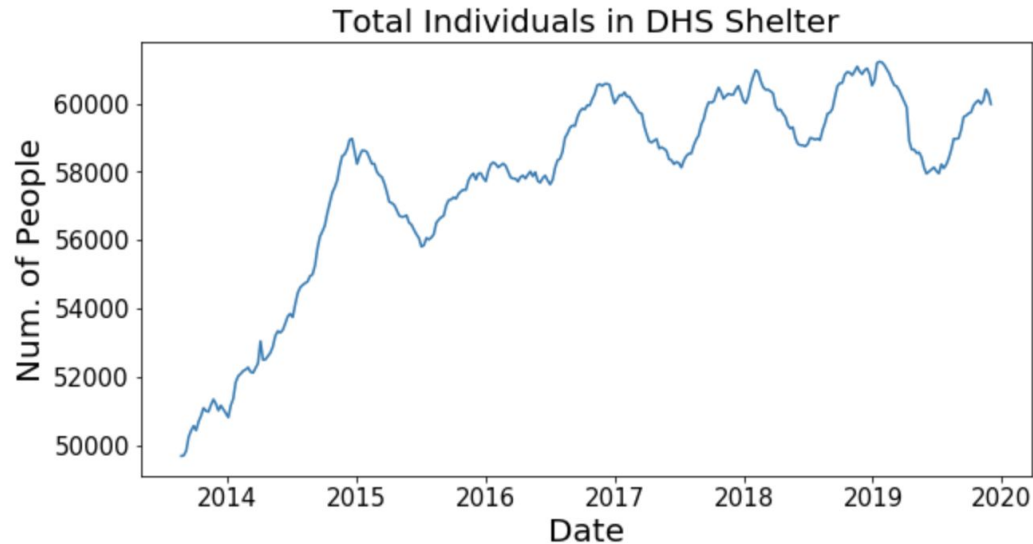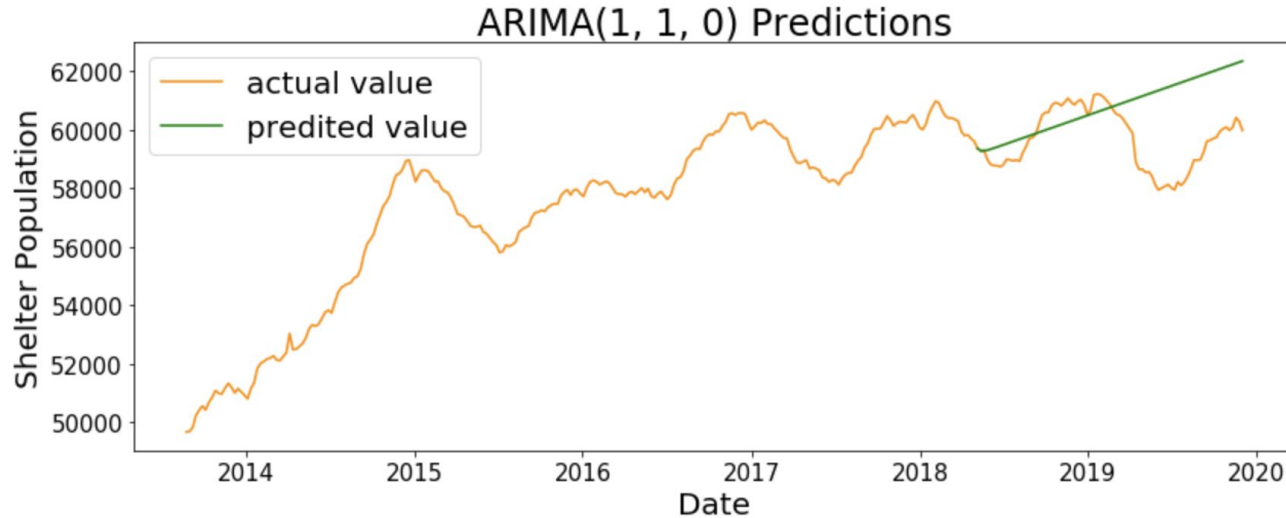# Shelter population has an upward trend and seasonality

# Shelter population has an upward trend and seasonality



Total Individuals in DHS Shelter
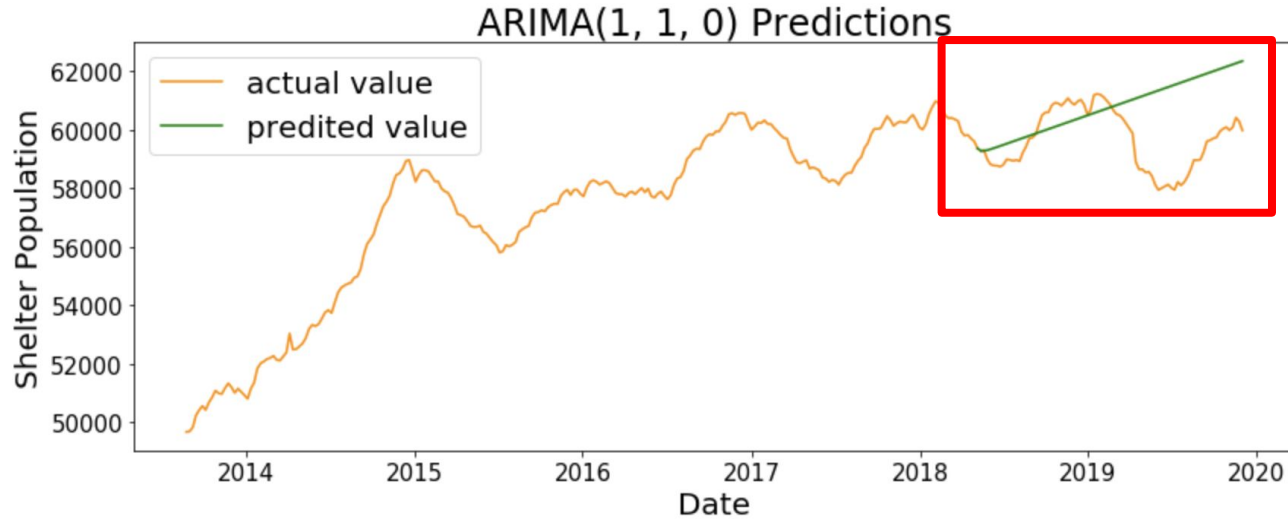
# Shelter population has an upward trend and seasonality

# ARIMA model did not make meaningful prediction



ARIMA(1, 1, 0) Predictions

**Root mean square error: 1834**

# ARIMA model did not make meaningful prediction



ARIMA(1, 1, 0) Predictions

**Root mean square error: 1834**

# ARIMA model did not make meaningful prediction



Weekly Difference in The Total Individuals in Shelter

**Root mean square error: 1834**

# ARIMA model did not make meaningful prediction



Weekly Difference in The Total Individuals in Shelter
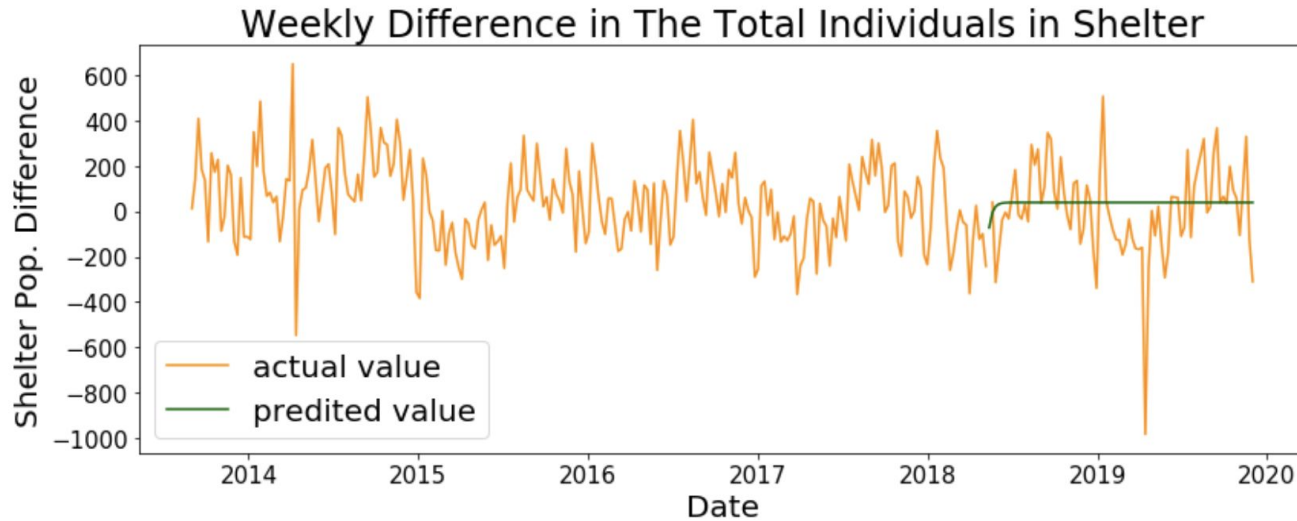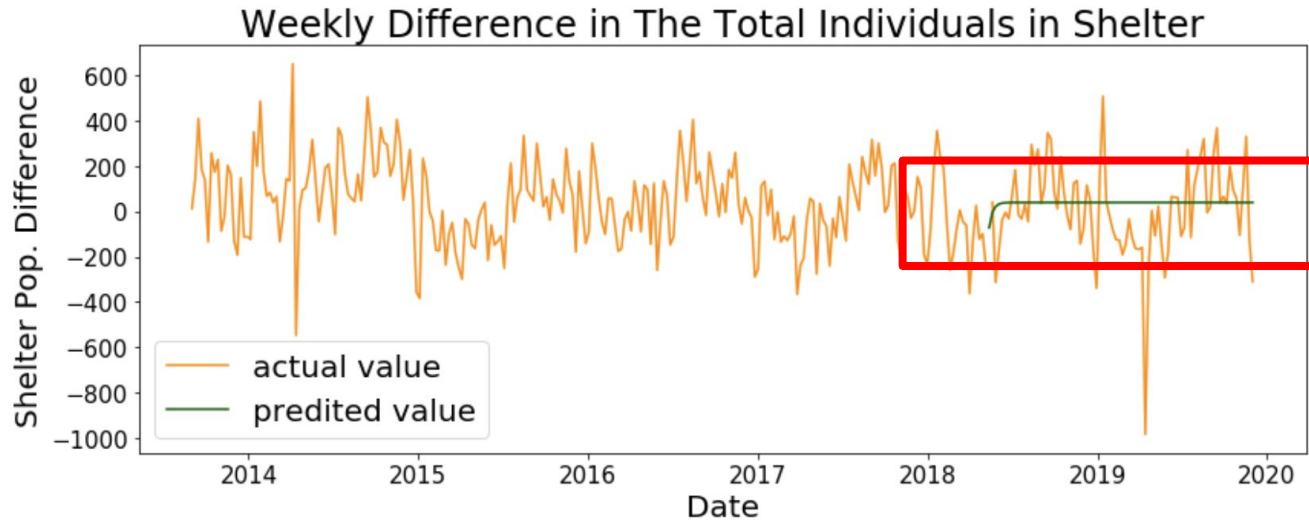
**Root mean square error: 1834**

# Seasonality terms significantly improves performance



SARIMA(1, 1, 0) x (2, 2, 0, 51) Predictions
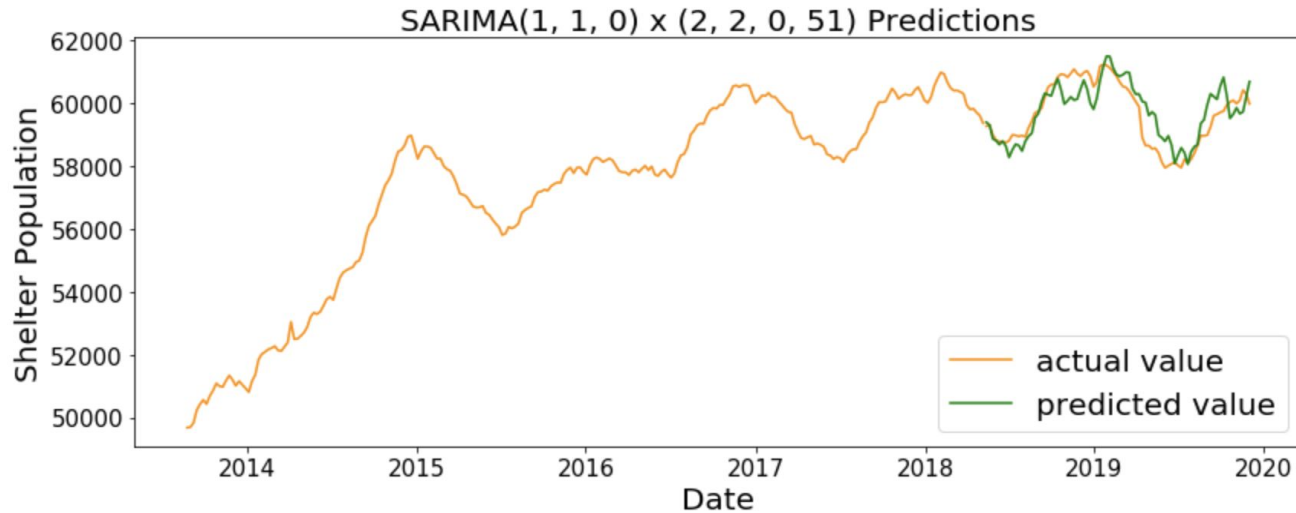
**Root mean square error: 560**

# Seasonality terms significantly improves performance



SARIMA(1, 1, 0) x (2, 2, 0, 51) Predictions
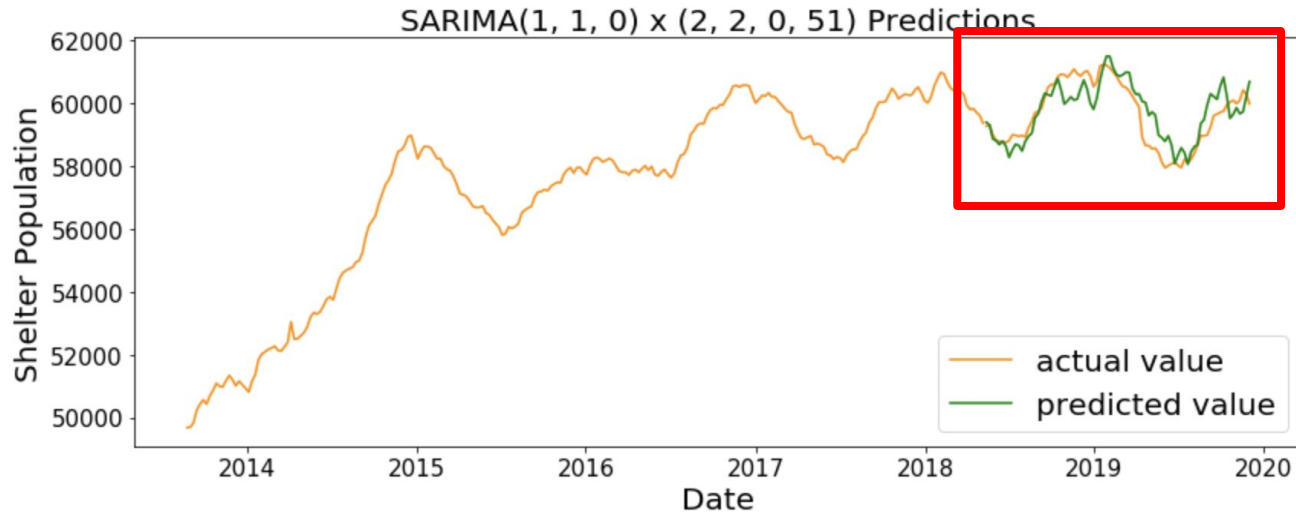
**Root mean square error: 560**

# Final Product!

**DHS Daily Report (SODA API)**

**Data wrangling**

- Resample into weekly average
- Datetime indexing
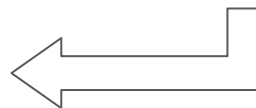  for time series analysis

**DHS Weekly Report (PostgreSQL)**

PostgreSQL

**Database update**

**DASH app**

Dash

**Time series modeling**
**Data visualization**

# *Conclusion*

## Next steps

- **Forecast demographic breakdowns**
- **Try exogenous features / VAR**
  - **Housing affordability**
  - **Medical debt**
  - **HUD funding**
  - **Natural language analysis on social media posts discussing homelessness**

## Issue

- **Model requires reliable data update**

→

## Solution

- **Data scraper that automates web-scraping and make the scraper part of the app (Shell scripting)**

# *Source*

1. http://www.nationastate.com

2. https://www.coalitionforthehomeless.org