Predicting Covid Case Counts per Ward in Washington, DC

Gabriel Cano

January 26, 2021

Overview

Problem Statement

Data Background

Data Visualizations

Modeling

Interpretation

Conclusions

Problem Statement

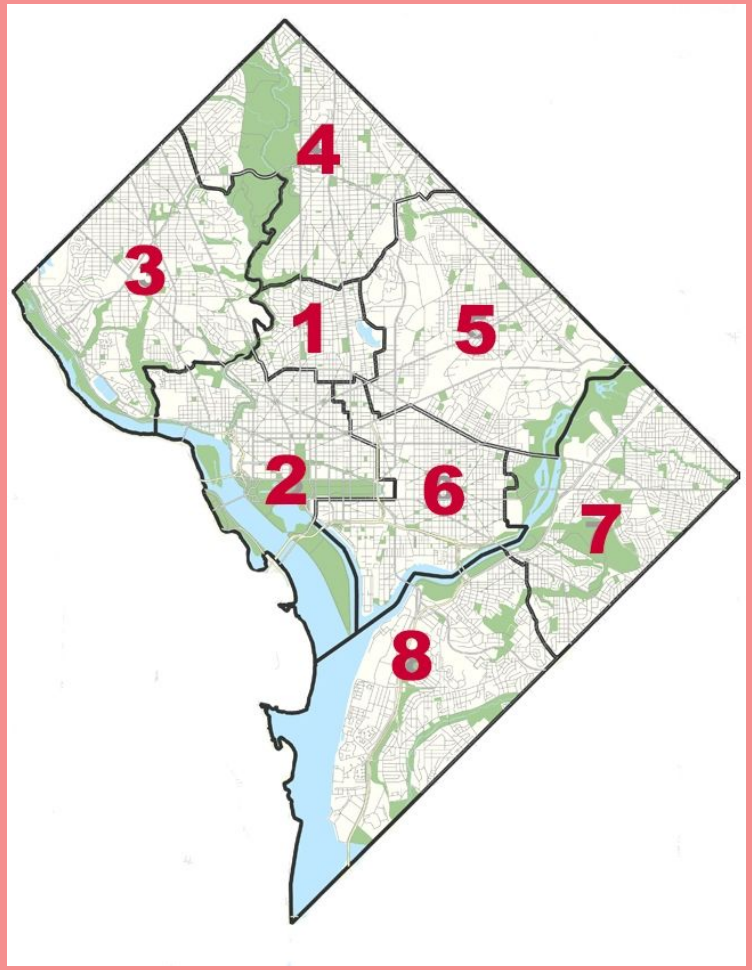
Although the U.S. Government has begun to distribute a promising vaccine to the public, Coronavirus cases are still spiking -- this is especially true in the DC area, where ICU beds are over 75% full. If healthcare administrators could better predict which areas of DC were going to receive more positive cases, they could better allot resources to provide infected patients the attention they need.

With this in mind, I endeavor to use an array of time series models to forecast the count of daily cases in each of DC's 8 Wards.

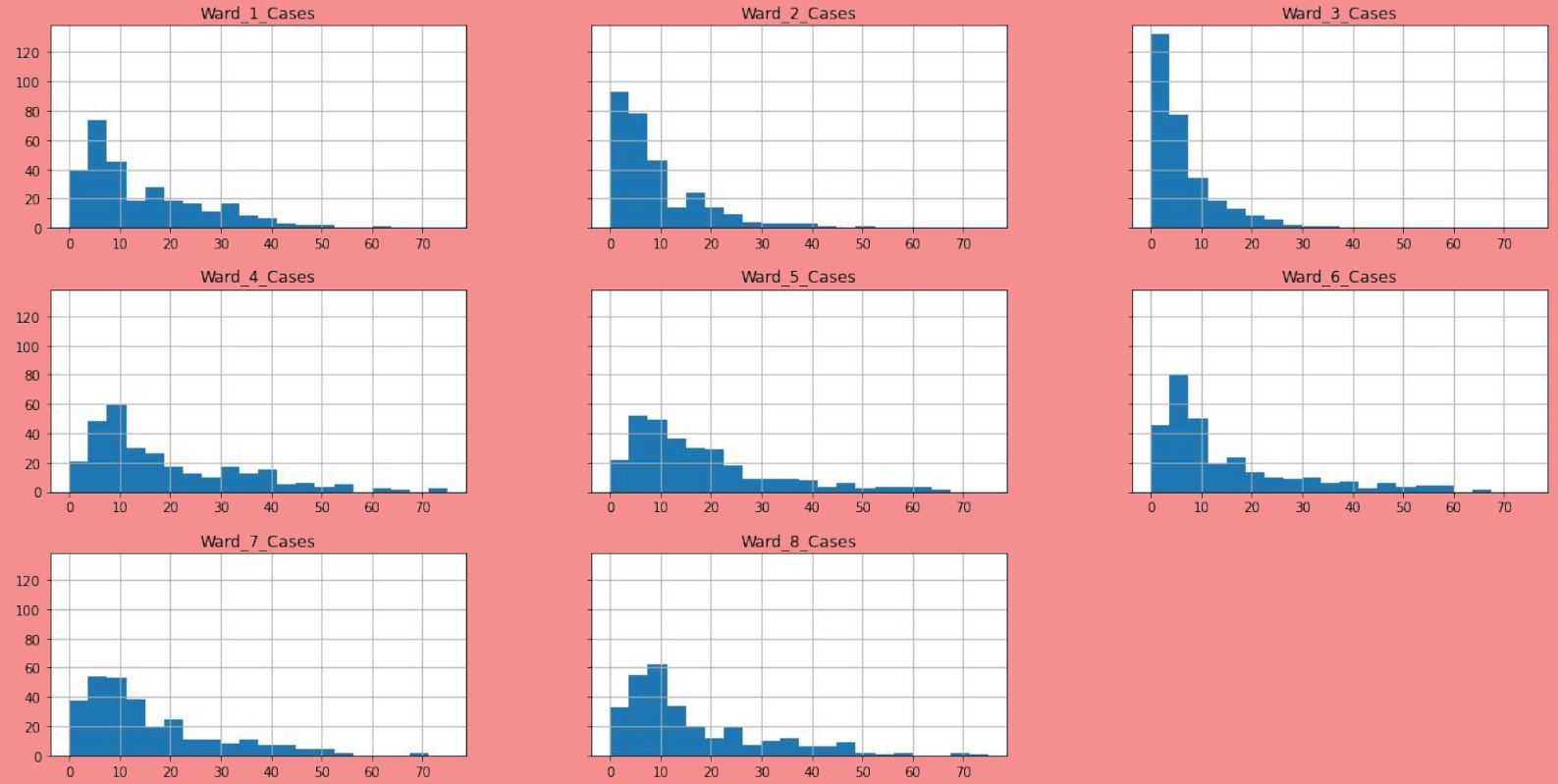
Data Background

- Open Data DC updates Covid-related data regularly
  - Dataset spans from 3/31/2020 - 01/18/2021
  - Cumulative positive cases reported per Ward

- Weather Underground provides data on daily average temperature for a given region

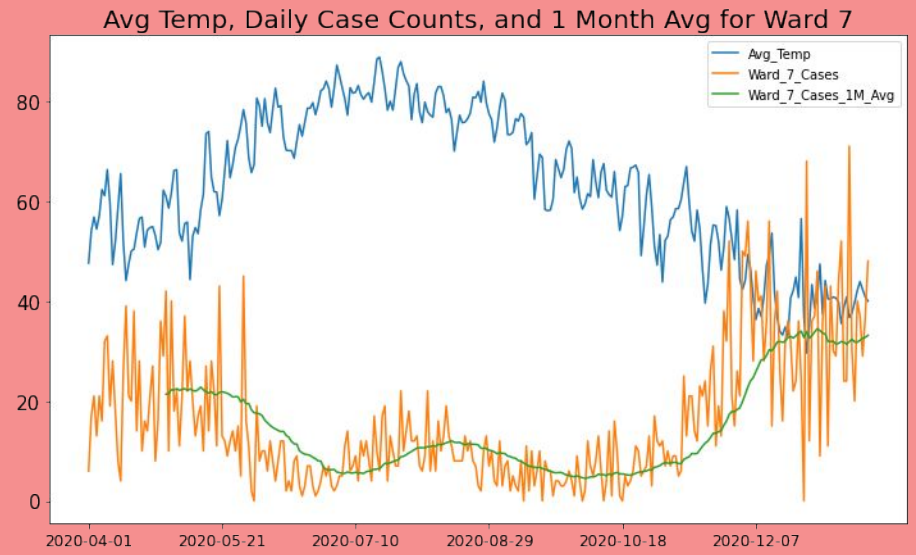


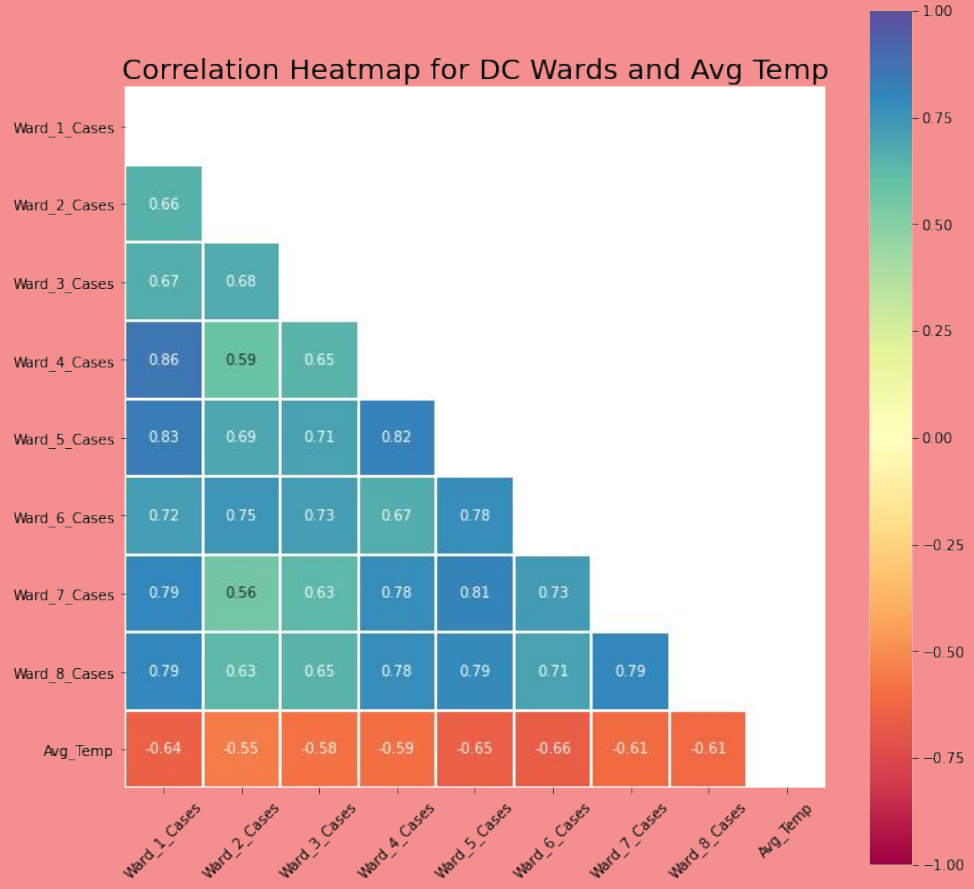
Frequency of Net-New Daily Case Counts



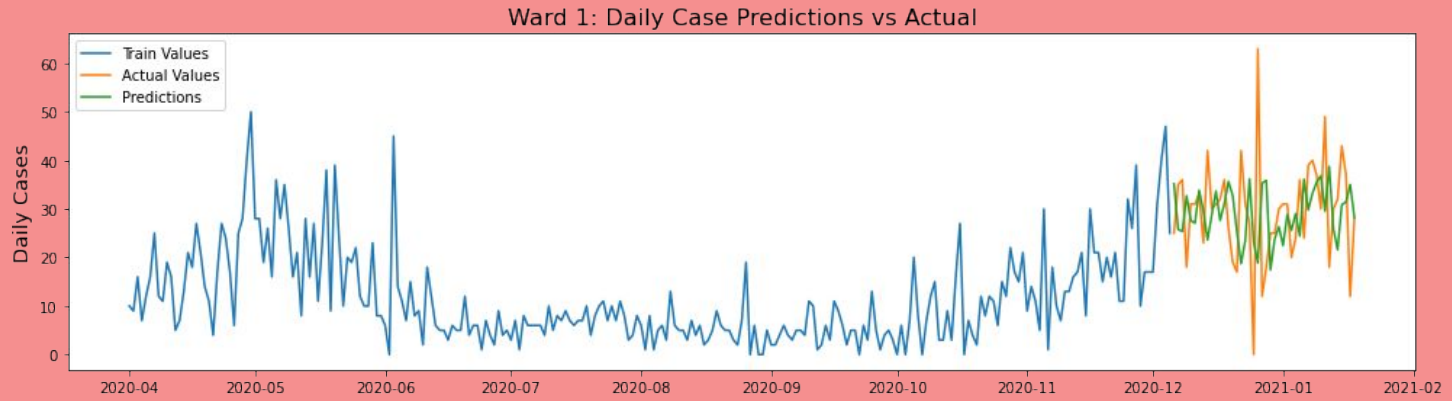
Average Temperature

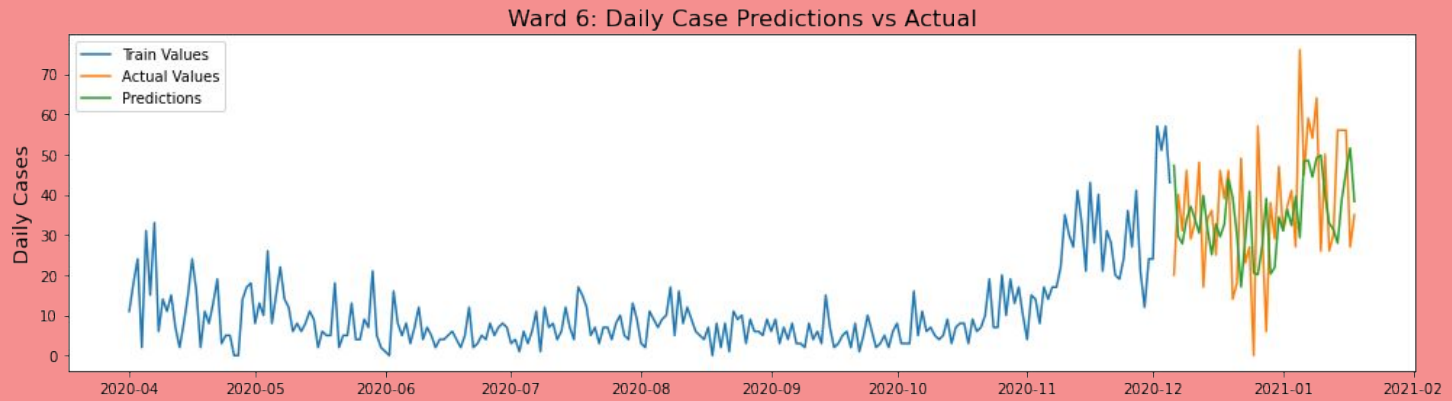
- A moderate negative correlation exists between average temperature and the daily new cases per Ward.



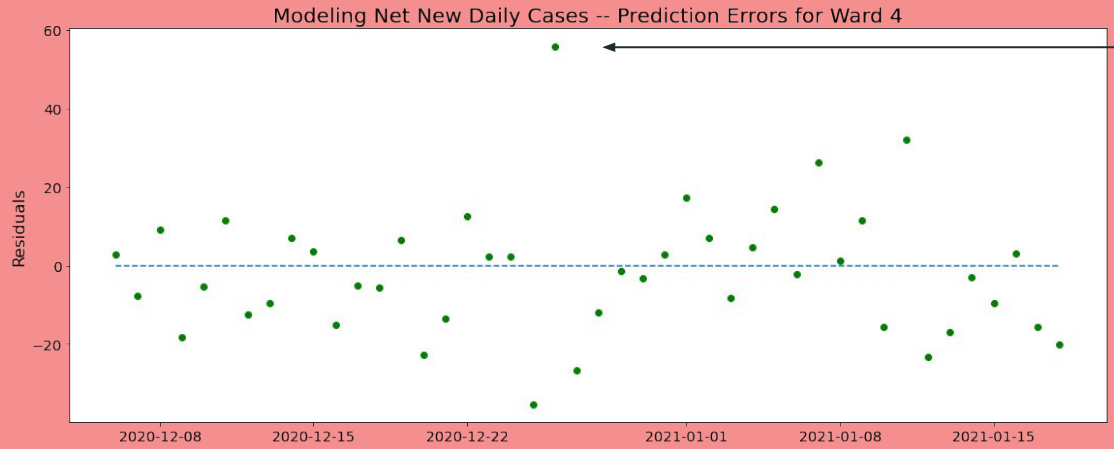


Model Results Summary

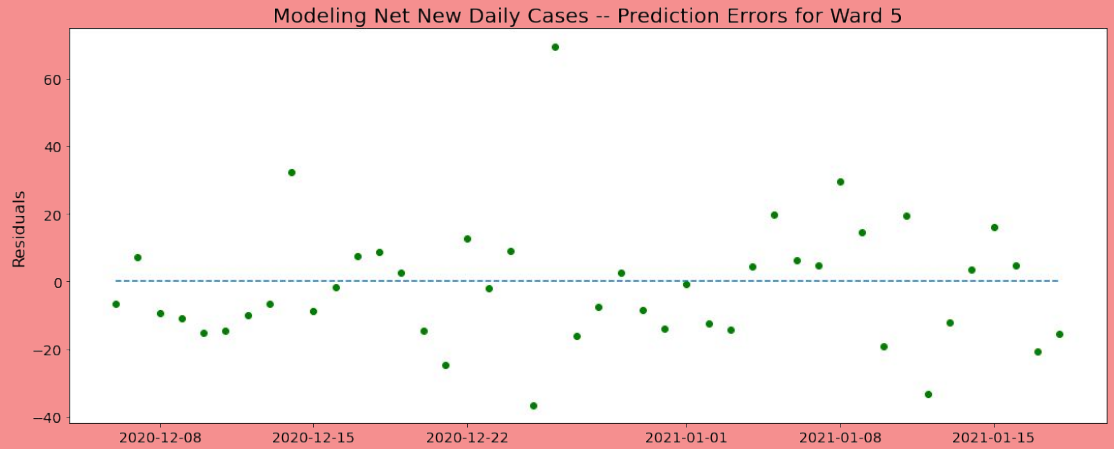




Model Results Summary



High residual on Christmas Day



Model Results Summary

- Both VAR and ARIMAX models give lower RMSE scores than the baseline model
- The ARIMAX model gives lower RMSE scores across all Wards, except for Ward 5

DC Ward	Baseline RMSE	VAR RMSE	ARIMAX RMSE (Rolling Forecast)	ARIMAX RMSE (One-Time Forecast)
Ward 1	17.0	13.2	12.6	16.6
Ward 2	19.4	13.7	12.3	14.0
Ward 3	9.1	7.0	6.7	7.1
Ward 4	23.9	18.7	16.3	19.9
Ward 5	24.4	18.0	18.4	25.4
Ward 6	22.1	17.3	16.4	19.1
Ward 7	23.1	17.0	15.7	24.1
Ward 8	20.8	17.4	15.4	16.1

Conclusions and Future Considerations

- Given Covid case counts and average temperature as variables, the rolling ARIMAX model gives lower prediction errors than the VAR model

- Health administrators in DC should leverage the ARIMAX model to identify which Wards are more likely to have spikes in new cases, in order to more efficiently allocate the healthcare resources at hand.

Looking forward...

- Include average income per Ward as a variable
- Spend more time tinkering with hyperparameters
- Make a Tableau dashboard!

Thank you!

Any questions?

Sources

https://khn.org/morning-breakout/critical-bed-shortages-in-u-s-hospitals/

https://www.nbcwashington.com/news/local/coronavirus-in-dc-maryland-virginia-what-to-know-on-dec-22/2517745/

https://machinelearningmastery.com/persistence-time-series-forecasting-with-python/#:~:text=Persistence%20Algorithm%20(the%20%E2%80%9Cnaive%E2%80%9D%20forecast)&text=The%20equivalent%20technique%20for%20use,step%20(t%2B1).

https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/)