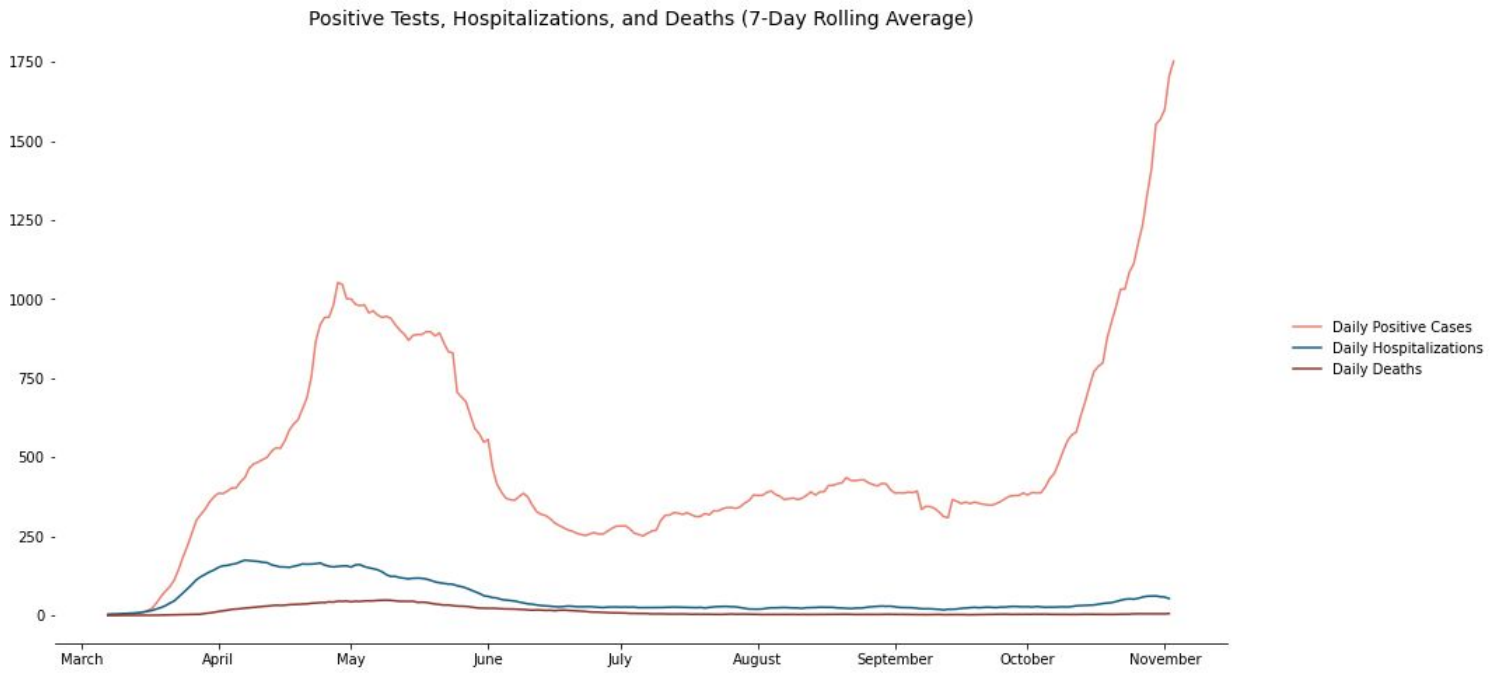
Chicago COVID-19 ZIP Code Analysis

Jamil Mirabito

Chicago is in the midst of a second wave of COVID-19 cases, now having surpassed the total number of cases during the initial wave.



Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020); Chicago COVID-19 Daily, Cases, Deaths, and Hospitalizations (3/01/2020 - 11/03/2020).

Challenge

With the massive uptick in positive cases, it's important to **understand which communities are most affected by the virus** and how certain mitigation efforts may be working to **reduce the city's overall positivity rate through the coming weeks.**

Analyses

ZIP-Code Analysis:

Using zip-code-level demographic and economic characteristics to understand which variables are most related to positivity rates and testing.

Projecting Chicago's Positivity Rate:

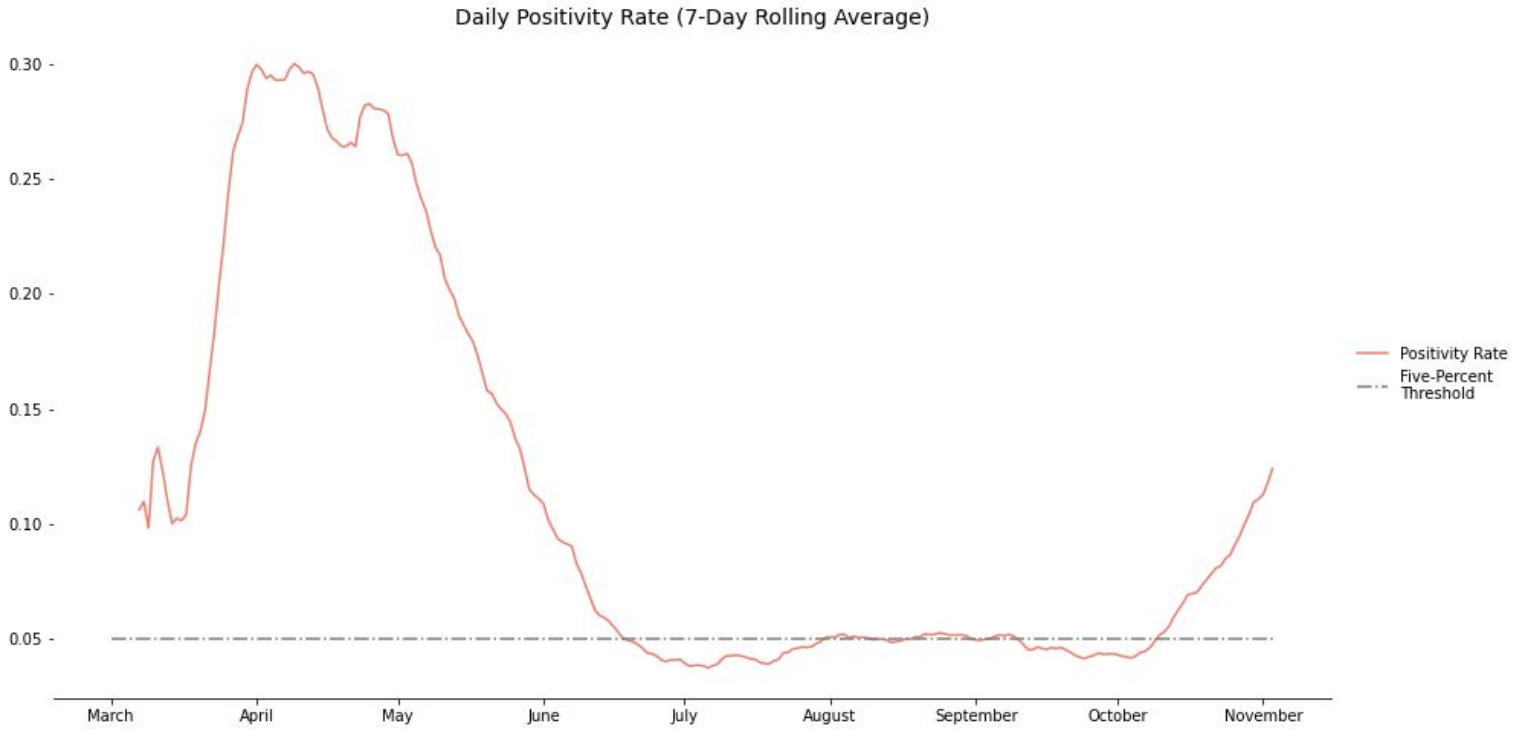
Using time-variant characteristics to model the city's positivity rate for the next two weeks.

Positivity Rates

Positivity rate is the percentage of total tests administered in a given day that come back positive. This does not include rapid tests, or antibody tests.

The World Health Organization (WHO) recommends that the **positivity rate in a particular area remain below 5 percent** for at least 2 weeks before loosening restrictions.

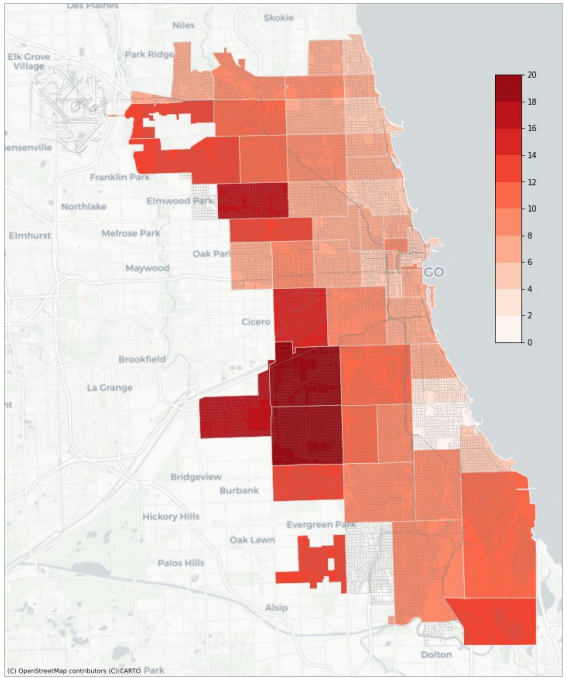
As recently as the start of November, the city-wide positivity rate has approached 15 percent - far exceeding the 5 percent threshold



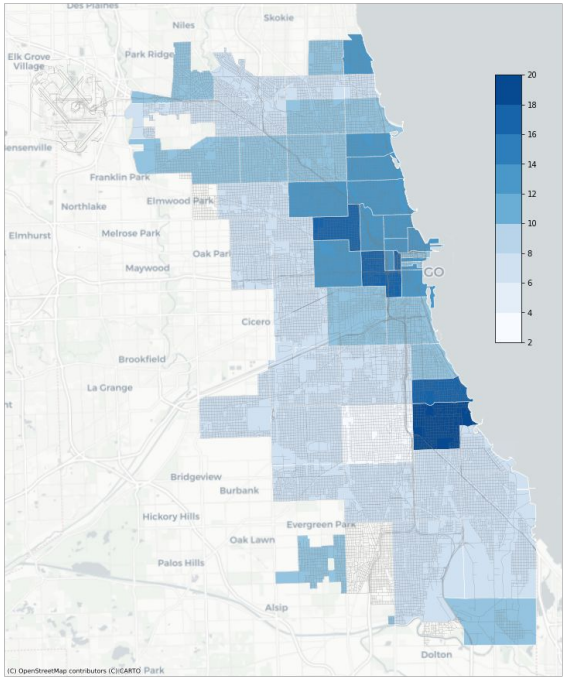
Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020). See Appendix section **Cases, Testing & Positivity Rates** for trendlines disaggregated by age group and race.

Primarily Black and Latinx communities have the highest 2-week average positivity rates while testing seems to be focused near universities

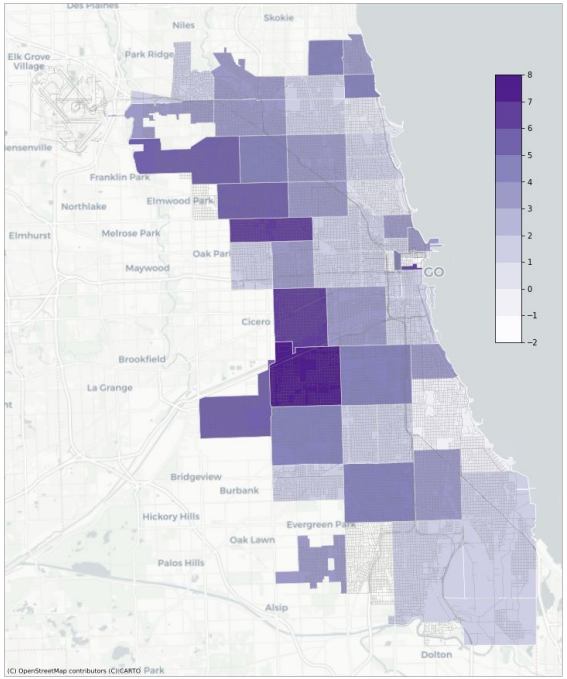
COVID-19 Positivity Rates by ZIP Code (10/18 - 10/31)



COVID-19 Testing Rates by ZIP Code (10/18 - 10/31)



COVID-19 Percentage-Point Change in Positivity Rate from Prior 2-Week Period (10/4 - 10/17)



Sources: US Census Data, Chicago COVID-19 Cases, Deaths, and Tests (3/01/2020 - 10/31-2020) at the zip-code level. Data sourced from different dataset with ZIP-code-level positivity rates available on a weekly basis. See **Appendix Slide 26** for list of top 5 ZIP codes for positivity rates and testing.

ZIP Code Analysis

ZIP Code Analysis

Data Sources:

- US Census Data
- City of Chicago Cases, Testing and Deaths by ZIP Code
- City of Chicago Testing Sites

Target Variables:

- COVID-19 Positivity Rates
- Testing Rate

Variables included in **Robust Linear Model (RLM):**

- Median Age
- Median Household Income
- Median Household Size
- Percent of a ZIP Code that is Hispanic/Latinx
- Percent Black
- Percent White
- Percent Undocumented & Percent Undocumented Foreign Born Latin America (FBLA)
- Percent Uninsured
- Percent Unemployed
- Percent Healthcare Workers
- Percent Essential Workers
- Distance from Testing Site

The following features were found to be significantly correlated with COVID-19 positivity rates & testing controlling for all other features

Target: COVID-19 Positivity Rates

- Mean Household Size (+0.021)
- Percent Hispanic (+0.124)
- Percent Uninsured (-0.309)
- Percent Health Workers (-0.696)
- Percent Essential Workers (+0.184)

Target: Testing Rate

- Median Age (-0.002)
- Median Household Income (-3.16e-07)
- Mean Household Size (-0.013)
- Percent Uninsured (-0.246)
- Percent Unemployed (-0.244)

Sources: US Census Data, Chicago COVID-19 Cases, Deaths, and Tests (3/01/2020 - 10/31-2020) at the zip-code level. See **Appendix - Regression Summaries** for model details. See **Appendix - Individual Scatter Plots** for individual variable correlations with targets.

Predicting Positivity Rates

Time-Series Modeling

Data Sources:

- Chicago Open Data - Daily Tests by Test
- NOAA Historical Weather Data

Models Tested:

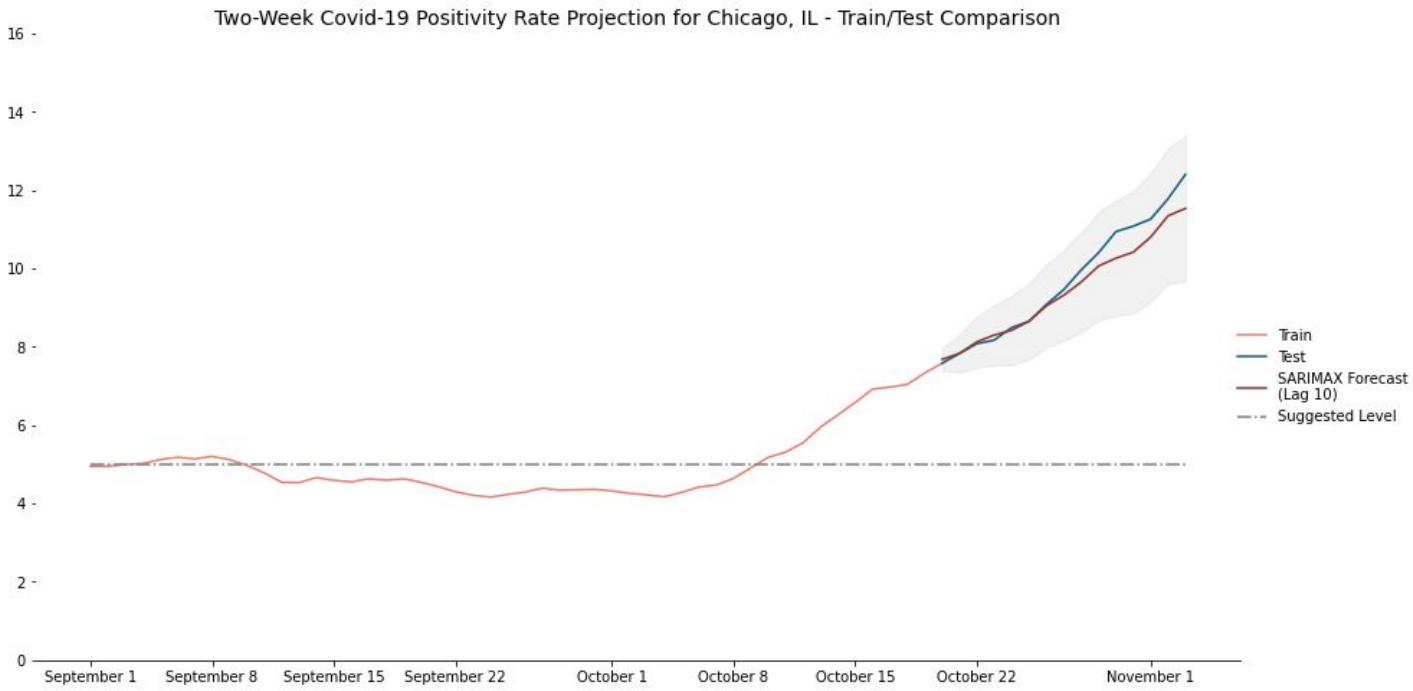
- Naive Untrained Model
- Moving Average Model
- Triple Exponential Smoothing Model
- SARIMA Model
- Various iterations of a SARIMAX model

Training Period: 03/01/2020 - 10/20/2020

Testing Period: 10/20/2020 - 11/3/2020

Model Type	RMSE
Naive Untrained Model	2.582
Moving Average Model	3.112
Triple Exponential Smoothing Model	1.252
SARIMA Model	1.149
Final SARIMAX model	0.393

The best-performing model was a SARIMAX Model trained on data from 05/15/2020 - 10/20/2020



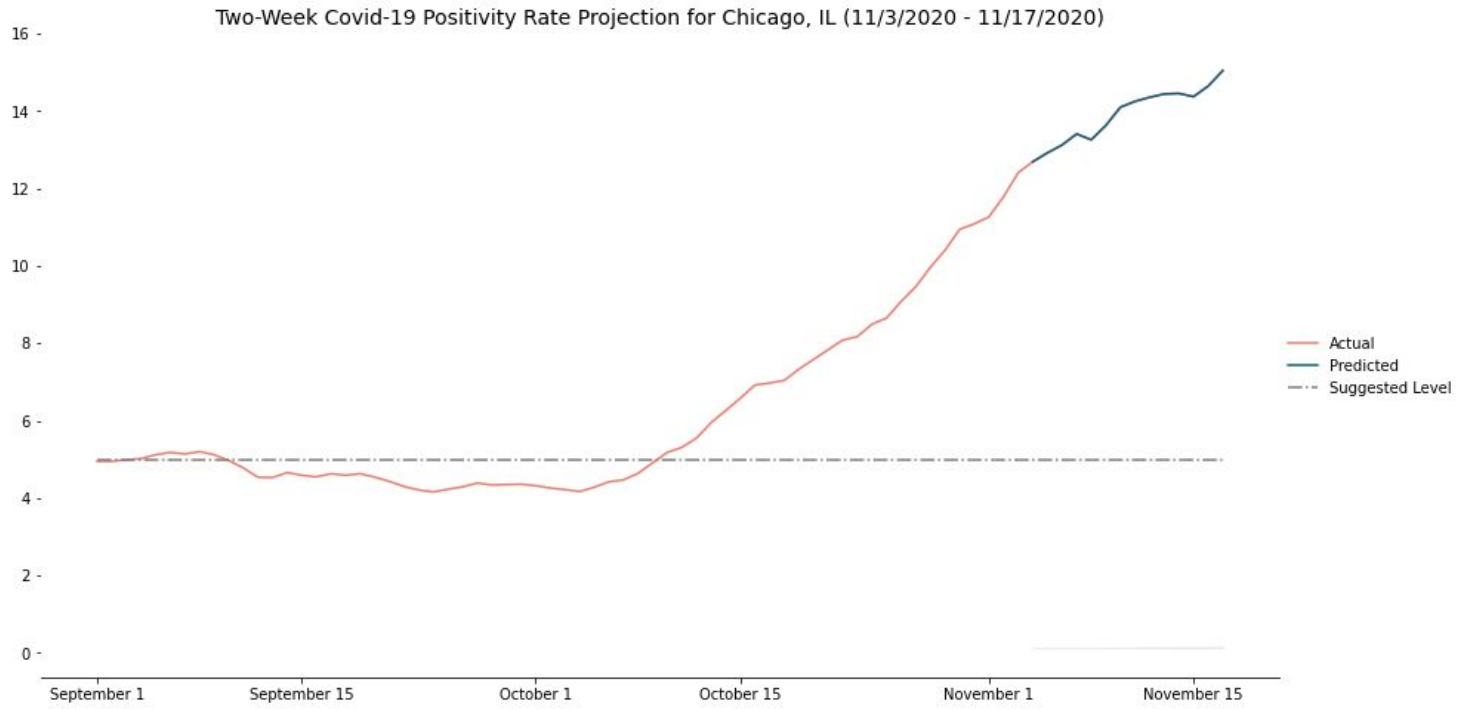
RMSE: 0.393

Exogenous variables used in model - all 10-day lag variables:

- Holidays
- Tightening Restrictions
- Loosening Restrictions
- Average Daily Temperature
- Precipitation

Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020). See appendix section **Time Series Models** for additional models and model performance.

Optimal model predicts steady increase in daily positivity rates of roughly one percentage point per week for the next 2 weeks



Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020)

Summary & Recommendations

Summary:

Over the past month, COVID-19 positivity rates were highest in **Latinx communities** and areas with a high percentage of **essential workers**.

In the coming weeks, it's projected that the City's overall **positivity rate will decrease by roughly 5 percentage points**, likely driven by increased testing among younger people.

While testing is readily available for **younger students**, communities with a higher percentage of **uninsured individuals are less likely to be tested**.

Recommendations:

Work with community leaders and city stakeholders to **ensure that all community residents are aware of free testing regardless of insurance coverage.**

Consider **mandating regular testing** for all essential and non-essential workers

Continue **stressing the importance of social distancing** in mitigating the spread of the virus particularly in communities with higher positivity rates (i.e., Hispanic communities)

Questions?

Appendix

Data Sources

Census Data (API):

- American Community Survey 2018 Estimates - zip-code level
- ACS Occupation Lookup - Table S2401

Chicago Open Data:

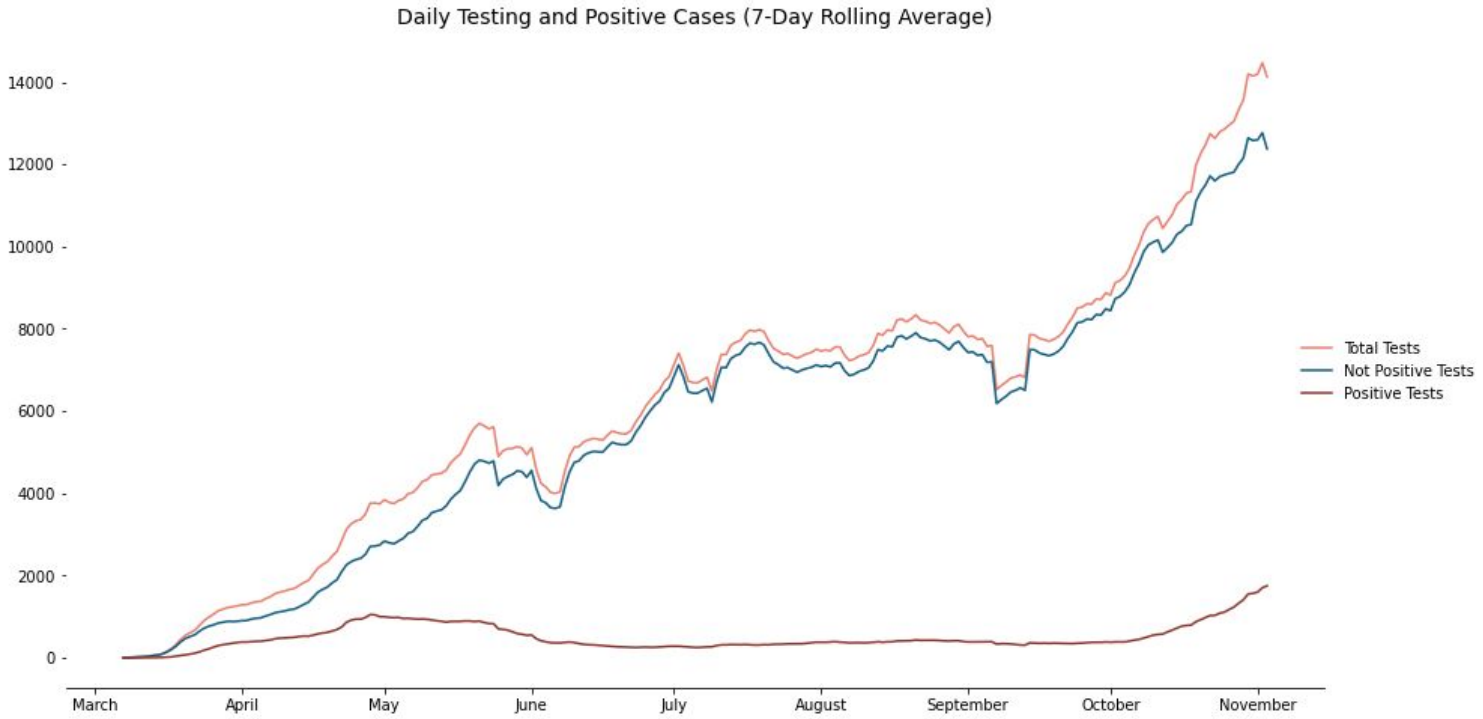
- COVID-19 Cases, Tests, and Deaths by ZIP Code
- COVID-19 Testing Sites
- COVID-19 Daily Cases, Deaths, and Hospitalizations
- COVID-19 Daily Testing by Test
- Chicago ZIP Codes Shapefile

NOAA:

- Average Daily Temperature and Precipitation (Historical Data)

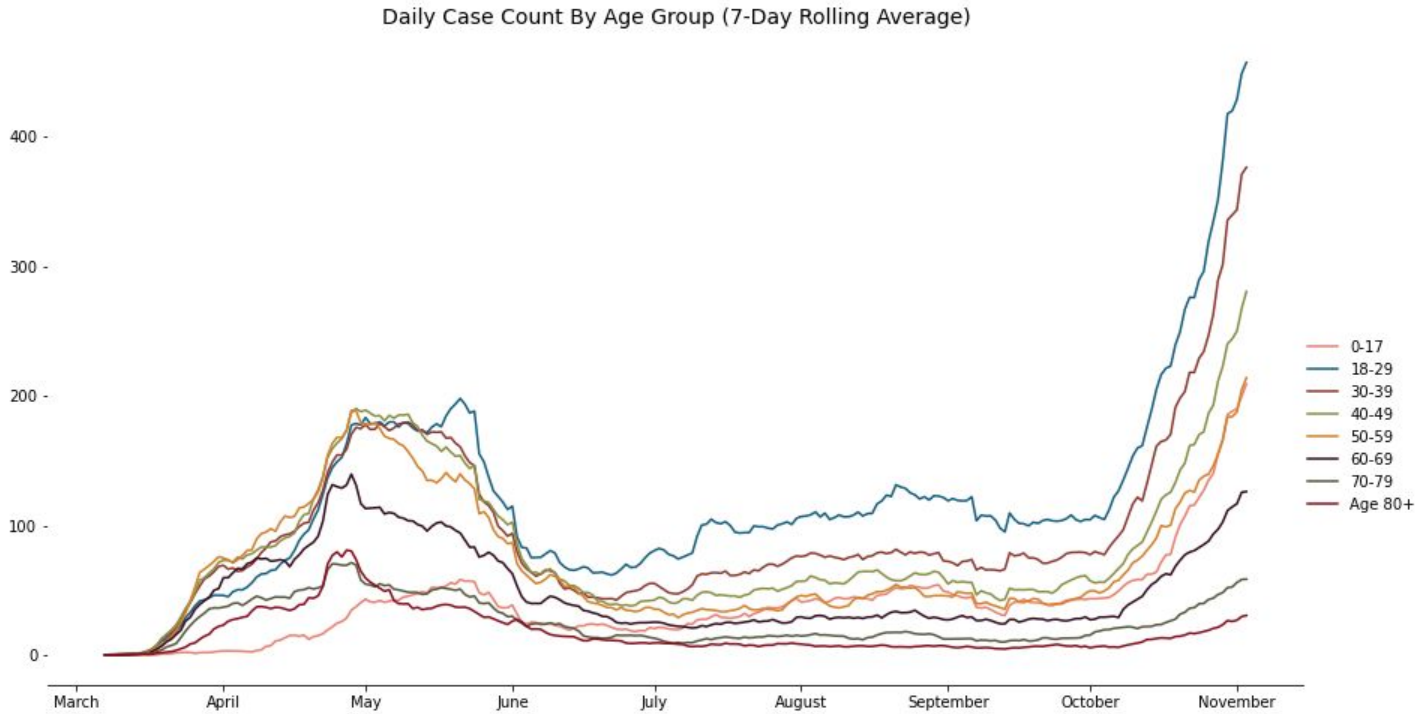
Cases, Testing & Positivity Rates

While total positive cases have surpassed the maximum number of cases in March, positivity rates appear much lower due to expanded testing



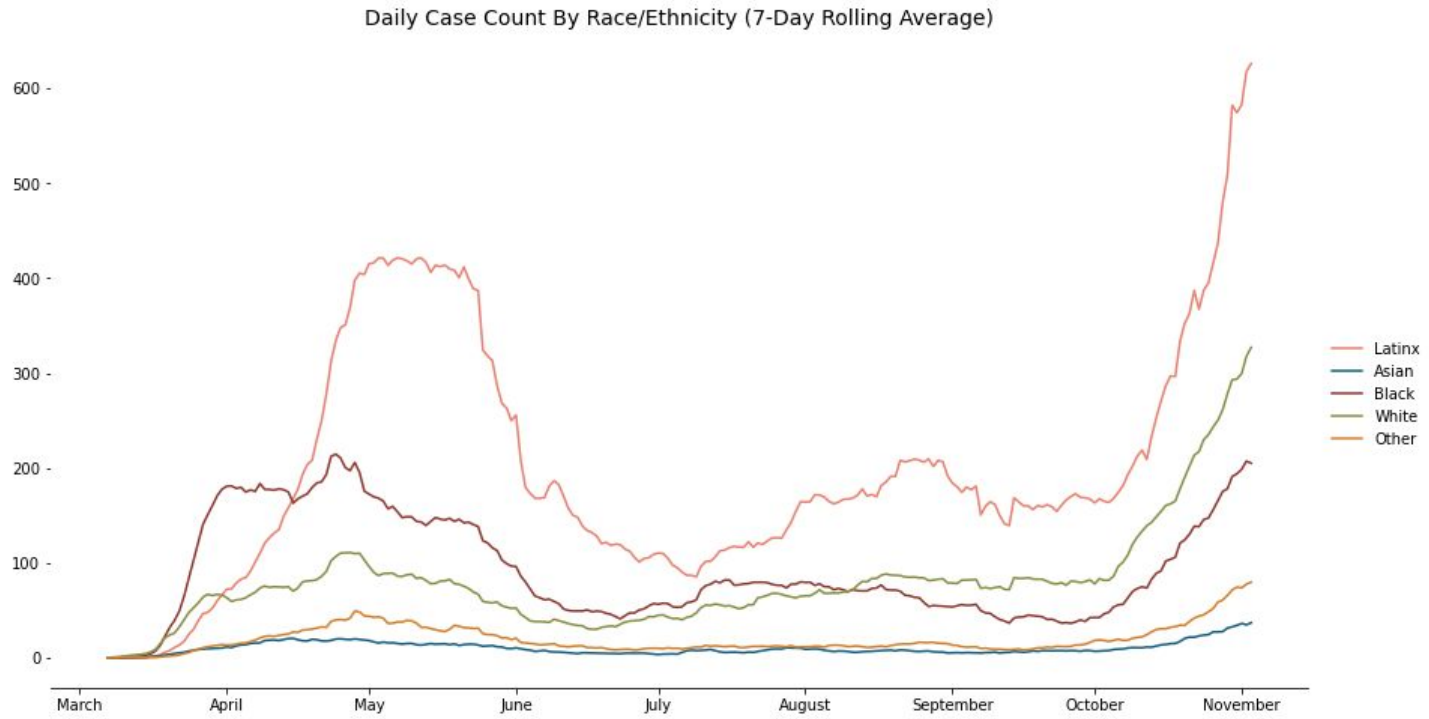
Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020)

Individuals lead the City in total number of daily cases (7-day moving average)



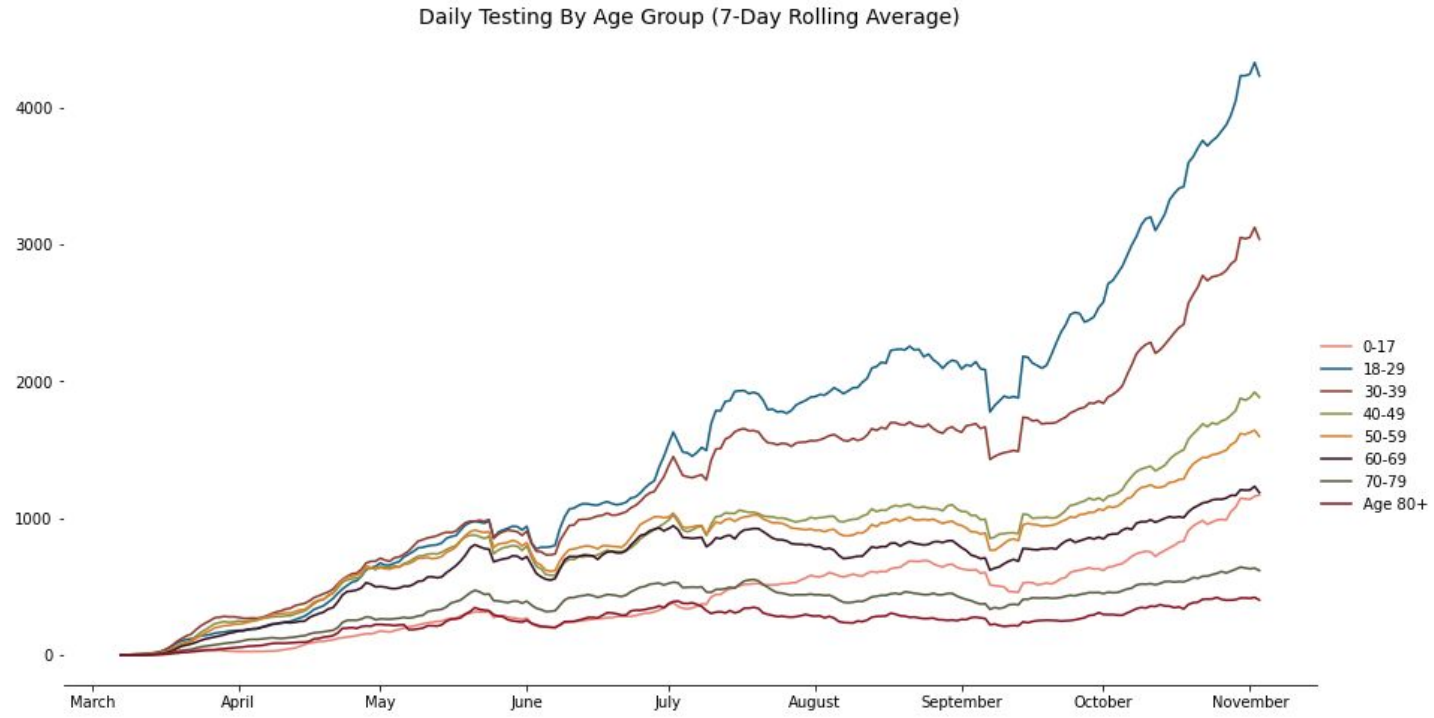
Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020)

Latinx individuals also account for the highest number of daily cases consistently throughout the pandemic



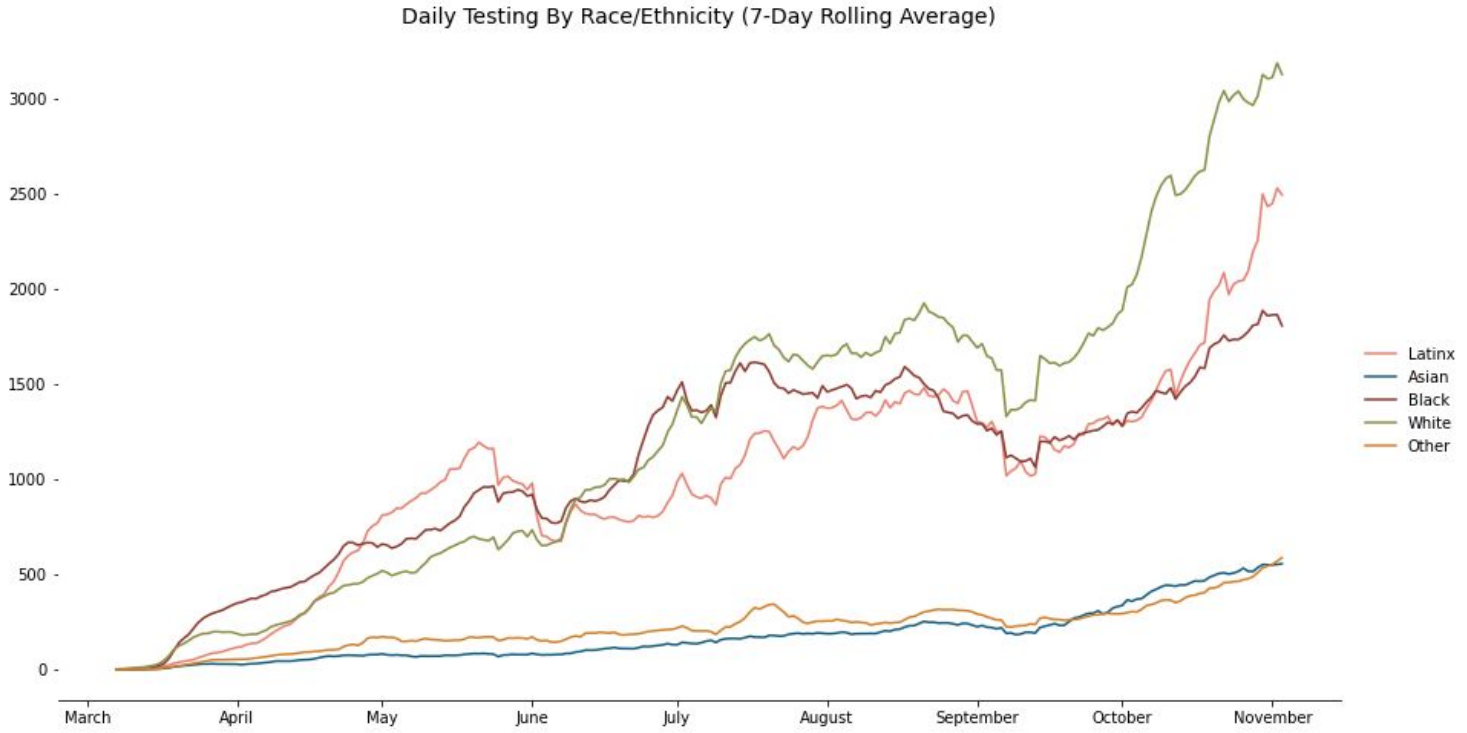
Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020)

Individuals between the ages of 18 and 29 are tested at the highest rate



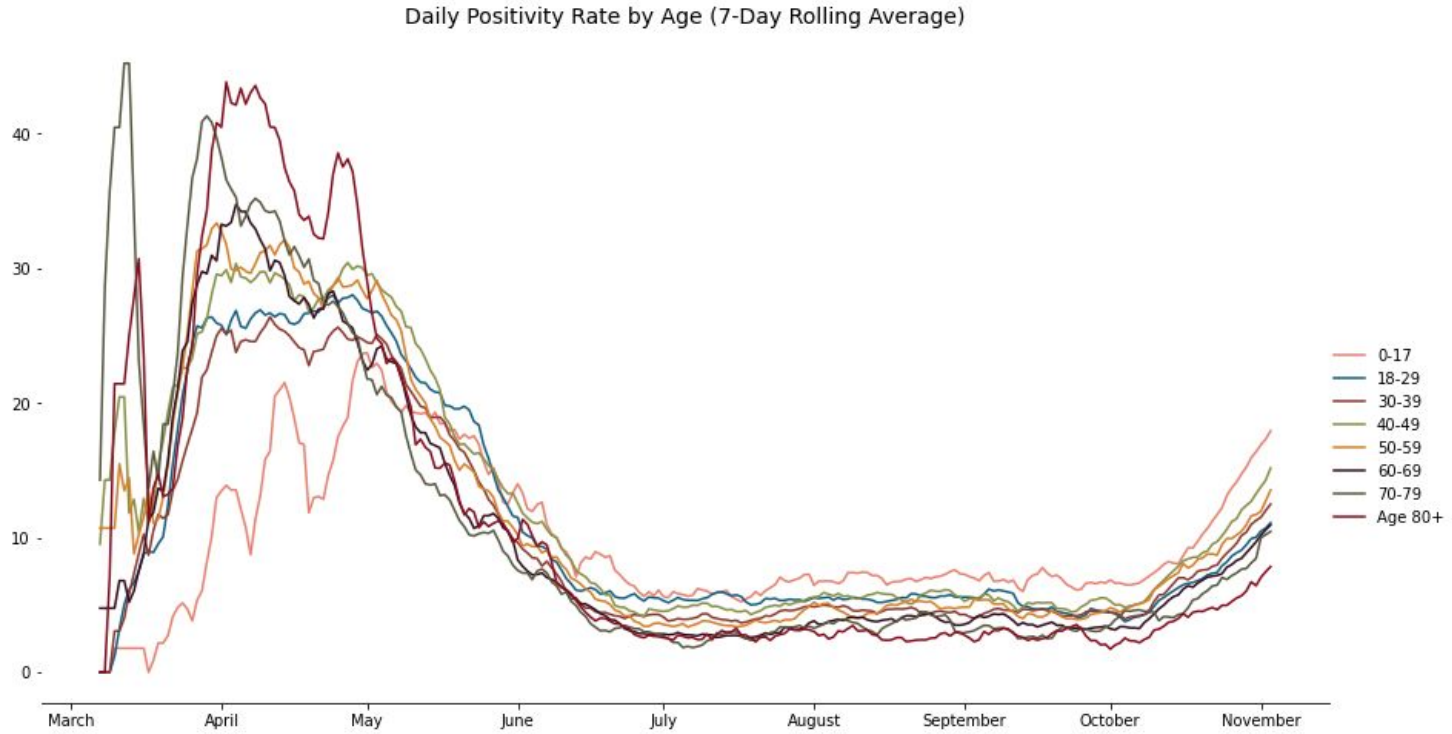
Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020)

White Chicagoans tend to get tested more than other ethnic groups



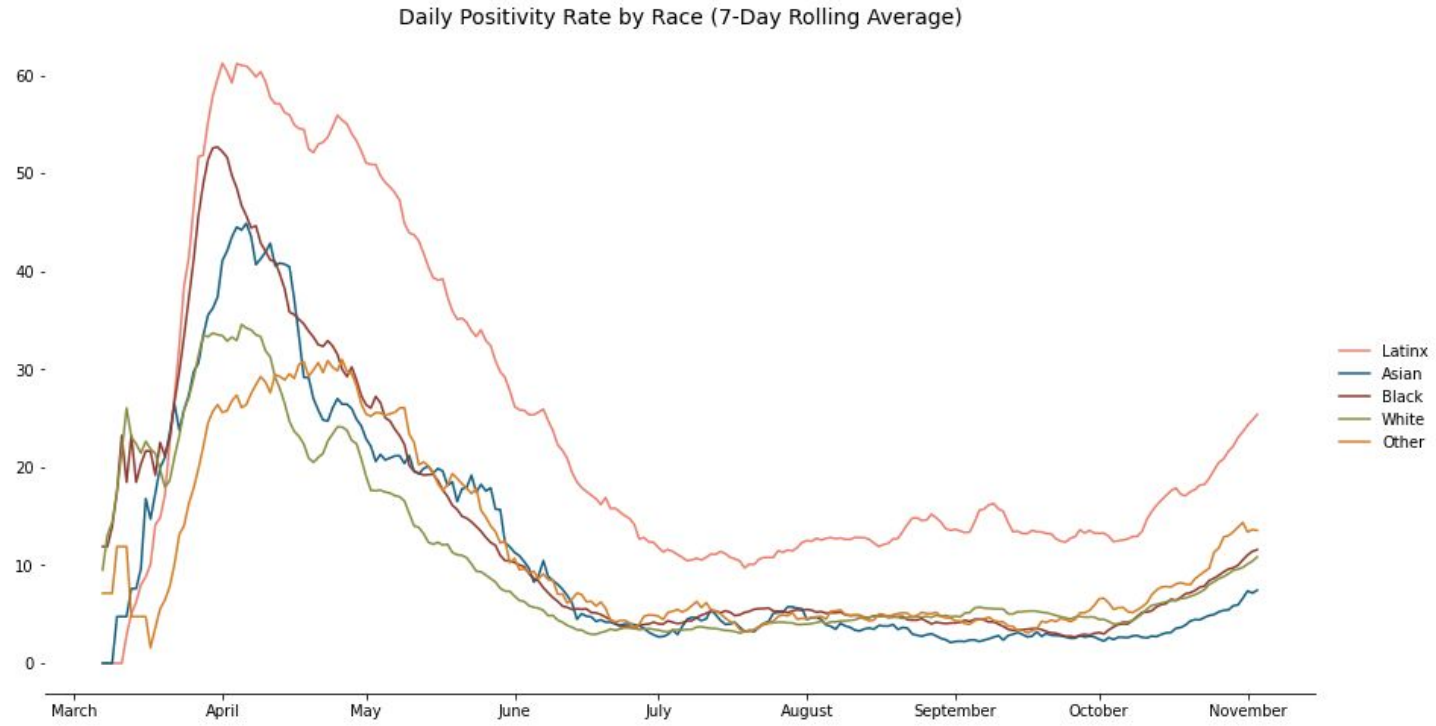
Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020)

Individuals younger than 17 seem to be leading the city in positivity rates



Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020)

Latinx Chicagoans lead all other groups in positivity rates



Source: Chicago Daily Testing By Test (3/01/2020 - 11/03/2020)

Top 5 ZIP Codes for both Positivity Rates and Testing

Positivity Regression Analysis:

1. **Brighton Park (60632):** Primarily Latinx Neighborhood on Southwest Side (84 percent Latinx) with 19.9 percent positivity rate.
2. **Chicago Lawn (60629):** Primarily Latinx neighborhood in Chicago (71 percent Latinx) with 18.4 percent positivity rate.
3. **Clearing (60638):** Primarily White & Latinx community near Midway Airport with 30.6 percent essential workers and a positivity rate of 17.9 percent.
4. **Belmont Cragin (60639):** Primarily Latinx neighborhood (77.9 percent latinx) with 31.4 percent essential workers and a positivity rate of 17 percent.
5. **Little Village (60623):** Primarily Mexican neighborhood on the Southwest Side (66.1 percent Latinx) with positivity rate of 15.6 percent.

Testing Regression Analysis:

1. **Hyde Park/Woodlawn (60637):** Primarily Black neighborhoods home to the UChicago Campus and student body. The testing rate here is 18.2 percent.
2. **West Loop (60607):** Upscale neighborhoods directly west of downtown home to UIC and the Illinois Medical District. The testing rate here is 11.1 percent.
3. **Loop/West Loop (60606):** Wealthy area in the financial district of the city where most residents are White and few unemployed. Testing per capita is 10.6 percent.
4. **East Hyde Park (60615):** Neighborhood directly north of Hyde Park and the UChicago campus. The testing rate in this area is 10.2 percent.
5. **Wicker Park/West Town (60622):** Home to a number of students, recent graduates, and families. The testing rate here is 10 percent.

Regression Summaries

Target: Positivity Rates

Robust linear Model Regression Results

Dep. Variable:	posRate	No. Observations:	56
Model:	RLM	Df Residuals:	44
Method:	IRLS	Df Model:	11
Norm:	HuberT		
Scale Est.:	mad		
Cov Type:	H1		
Date:	Thu, 12 Nov 2020		
Time:	22:02:19		
No. Iterations:	42		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0486	0.041	-1.180	0.238	-0.129	0.032
medianAge	0.0017	0.001	1.897	0.058	-5.61e-05	0.003
medianHHInc	1.87e-07	1.83e-07	1.022	0.307	-1.72e-07	5.46e-07
mean_HHsize	0.0206	0.010	2.136	0.033	0.002	0.040
pctHispanic	0.1235	0.027	4.531	0.000	0.070	0.177
pctBlack	0.0260	0.026	0.991	0.322	-0.025	0.077
pctUndocumented	0.0639	0.058	1.092	0.275	-0.051	0.179
pctUninsured	-0.3086	0.130	-2.372	0.018	-0.564	-0.054
pctUnemployed	0.2272	0.224	1.013	0.311	-0.212	0.667
pctHealthWorkers	-0.6956	0.258	-2.695	0.007	-1.201	-0.190
pctEssential	0.1836	0.089	2.074	0.038	0.010	0.357
dist	-0.6104	0.379	-1.609	0.108	-1.354	0.133

Target: Testing Rate

Robust linear Model Regression Results

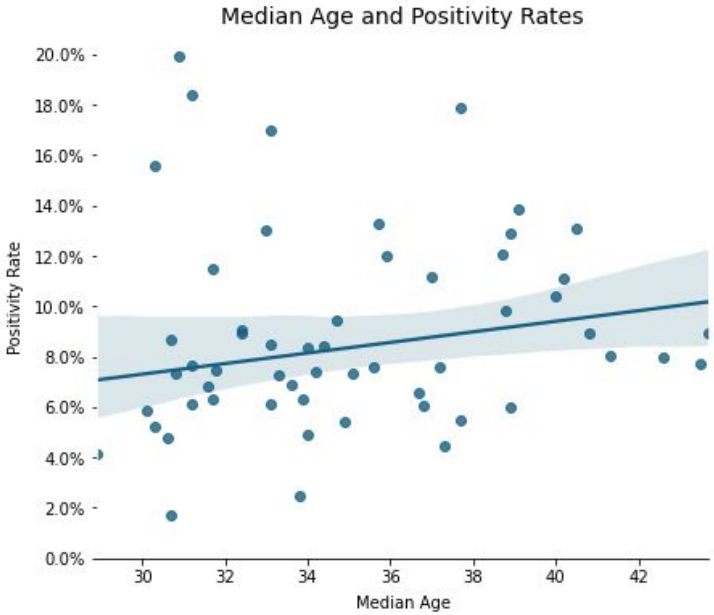
Dep. Variable:	testpercent	No. Observations:	56
Model:	RLM	Df Residuals:	44
Method:	IRLS	Df Model:	11
Norm:	HuberT		
Scale Est.:	mad		
Cov Type:	H1		
Date:	Thu, 12 Nov 2020		
Time:	22:27:18		
No. Iterations:	50		

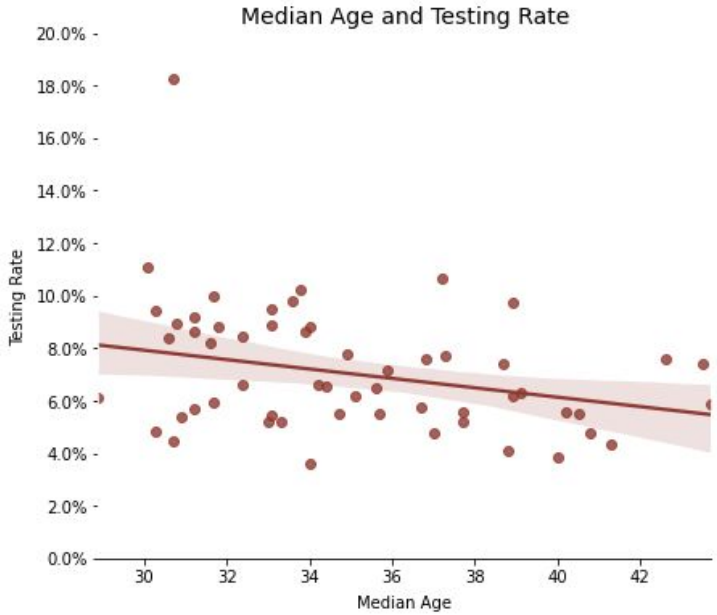
	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2239	0.022	10.262	0.000	0.181	0.267
medianAge	-0.0020	0.000	-4.334	0.000	-0.003	-0.001
medianHHInc	-3.159e-07	9.7e-08	-3.258	0.001	-5.06e-07	-1.26e-07
mean_HHsize	-0.0132	0.005	-2.579	0.010	-0.023	-0.003
pctHispanic	0.0226	0.014	1.564	0.118	-0.006	0.051
pctBlack	-0.0175	0.014	-1.261	0.207	-0.045	0.010
pctUndocumented	-0.0331	0.031	-1.069	0.285	-0.094	0.028
pctUninsured	-0.2463	0.069	-3.573	0.000	-0.381	-0.111
pctUnemployed	-0.2437	0.119	-2.050	0.040	-0.477	-0.011
pctHealthWorkers	0.1154	0.137	0.844	0.399	-0.153	0.383
pctEssential	-0.0037	0.047	-0.079	0.937	-0.096	0.088
dist	0.0828	0.201	0.412	0.680	-0.311	0.477

Individual Scatter Plots

Median Age

r = 0.121
p = 0.373

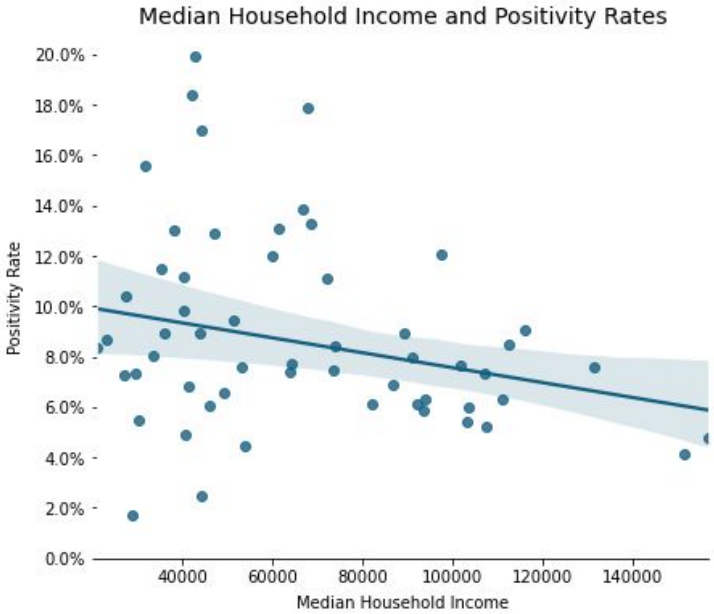


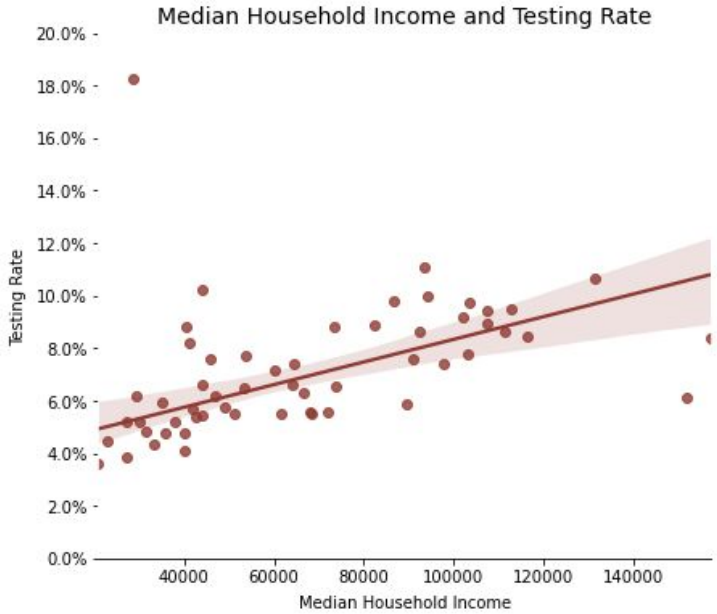


r = -0.334
p = 0.012*

Median Household Income

r = -0.281
p = 0.036*

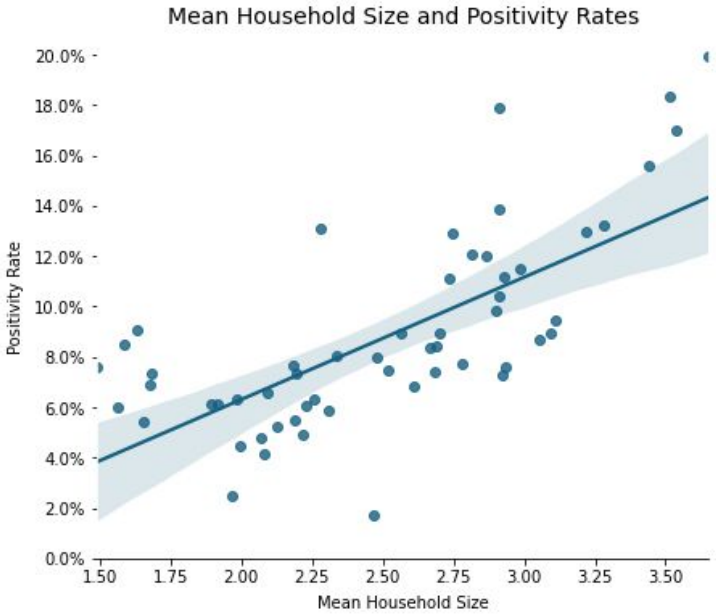


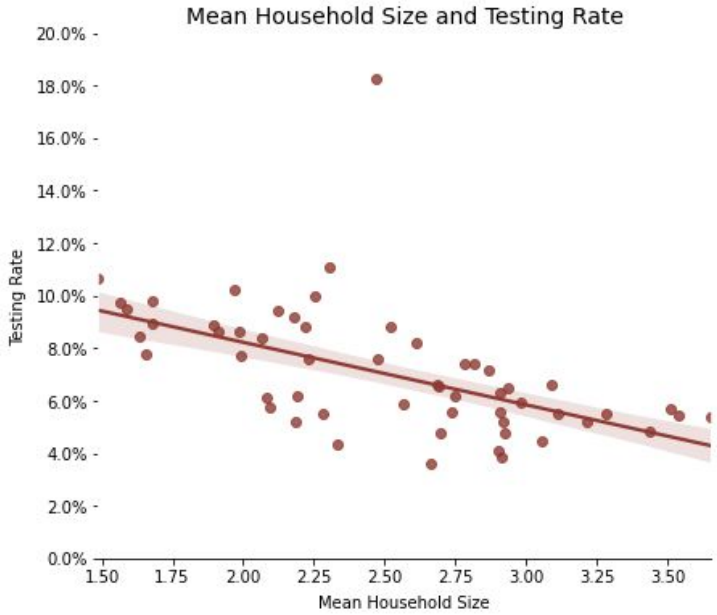


r = 0.411
p = 0.002**

Mean Household Size

r = 0.719
p = 0.000***

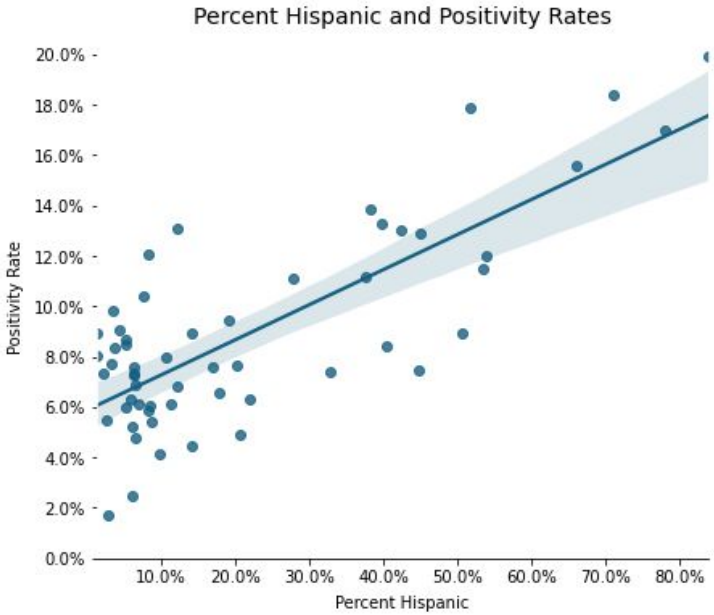


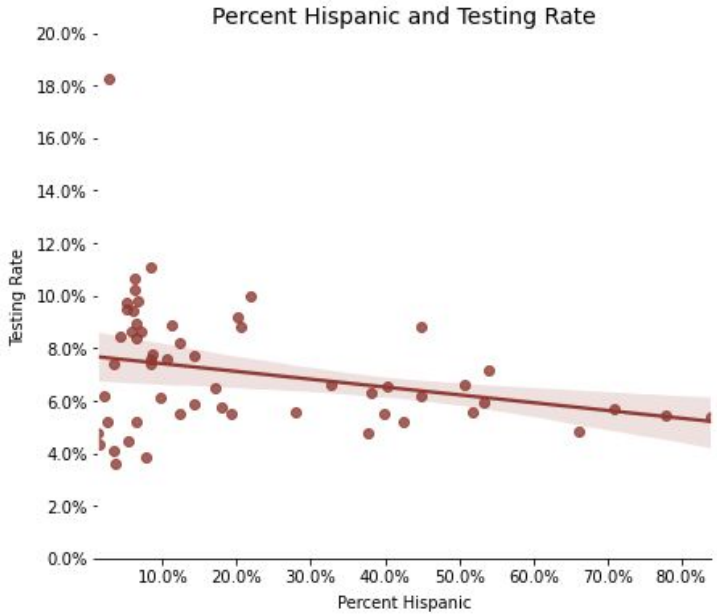


r = -0.549
p = 0.000***

Percent Hispanic

r = 0.776
p = 0.000***

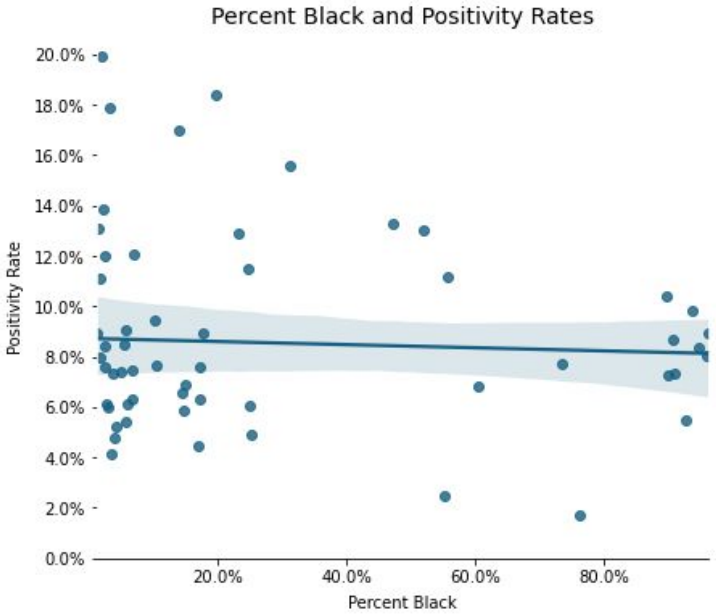


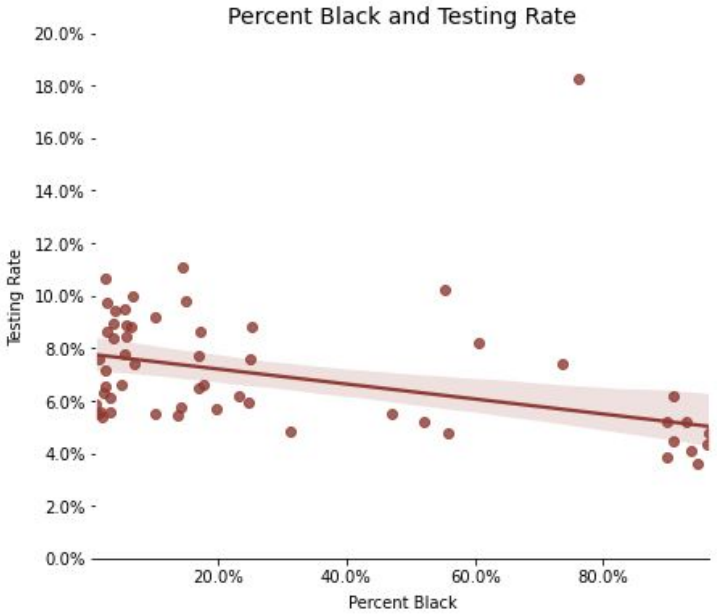


r = -0.297
p = 0.026*

Percent Black

r = -0.122
p = 0.371

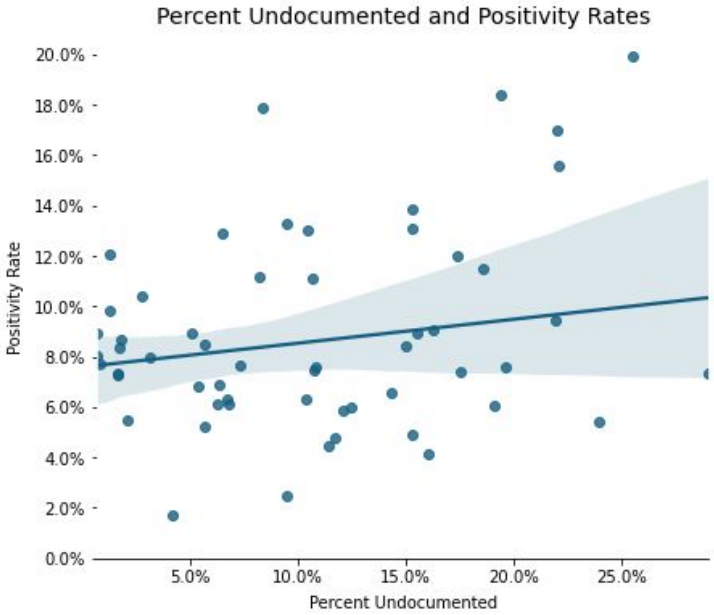


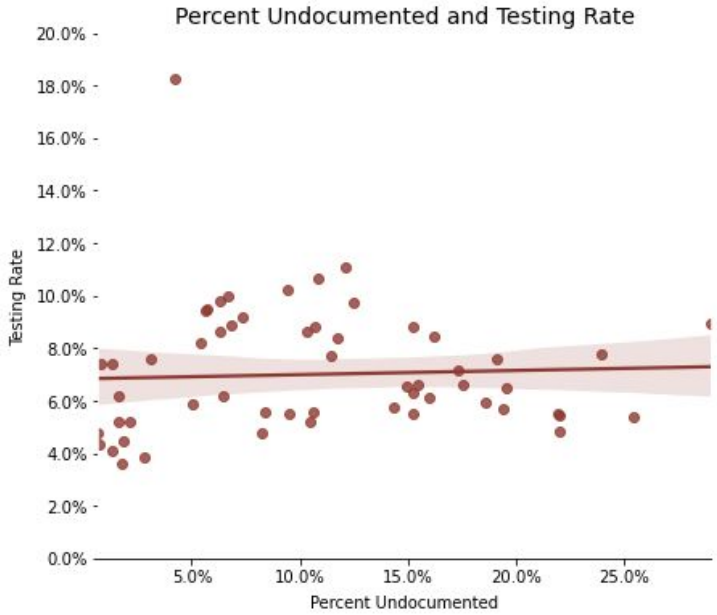


r = -0.287
p = 0.032*

Percent Undocumented

r = 0.280
p = 0.037*

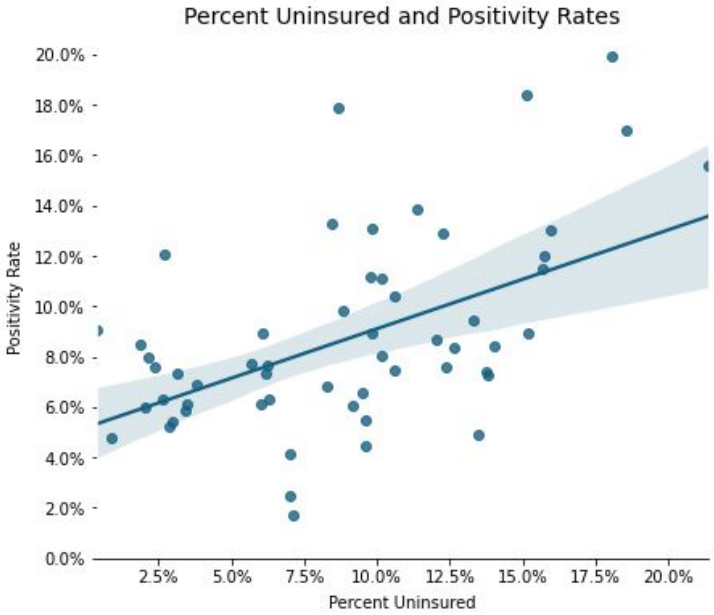


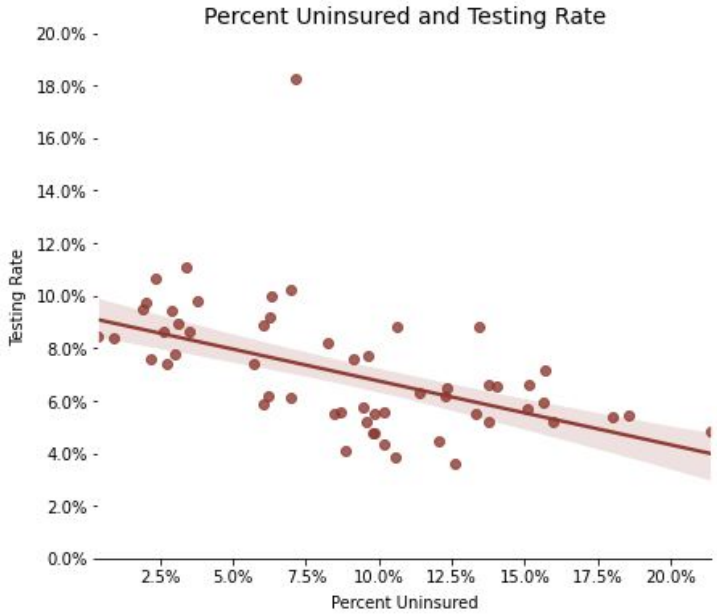


r = -0.005
p = 0.972

Percent Uninsured

r = 0.552
p = 0.000***

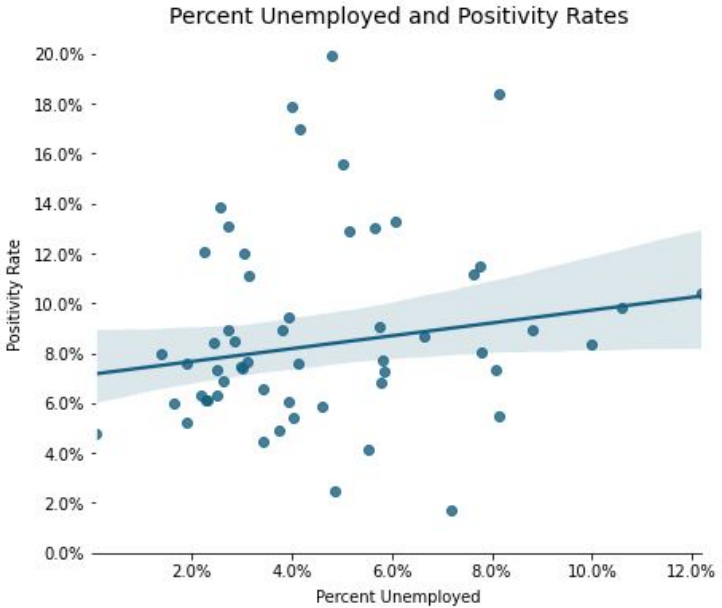


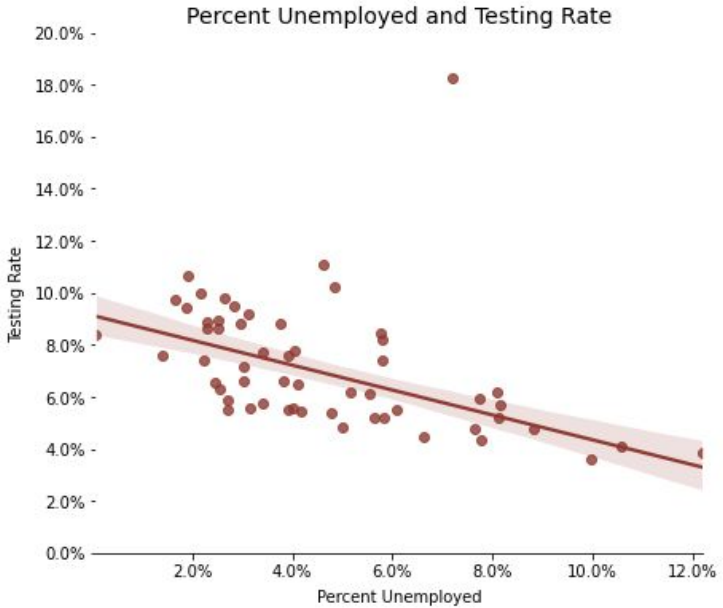


r = -0.528
p = 0.000***

Percent Unemployed

r = 0.160
p = 0.238

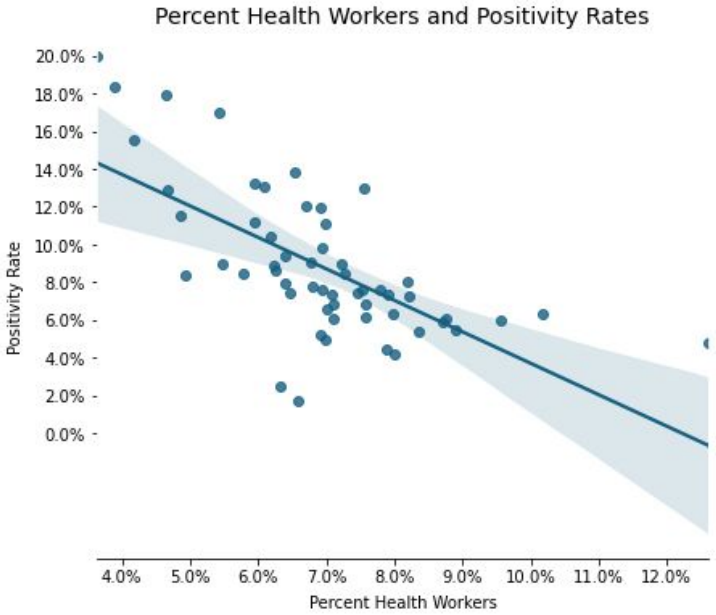


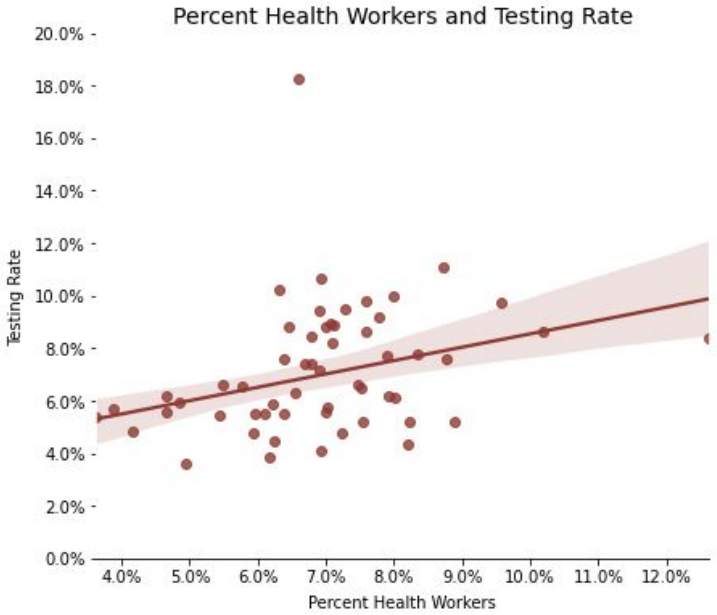


r = -0.421
p = 0.001**

Percent Health Workers

r = -0.664
p = 0.000***

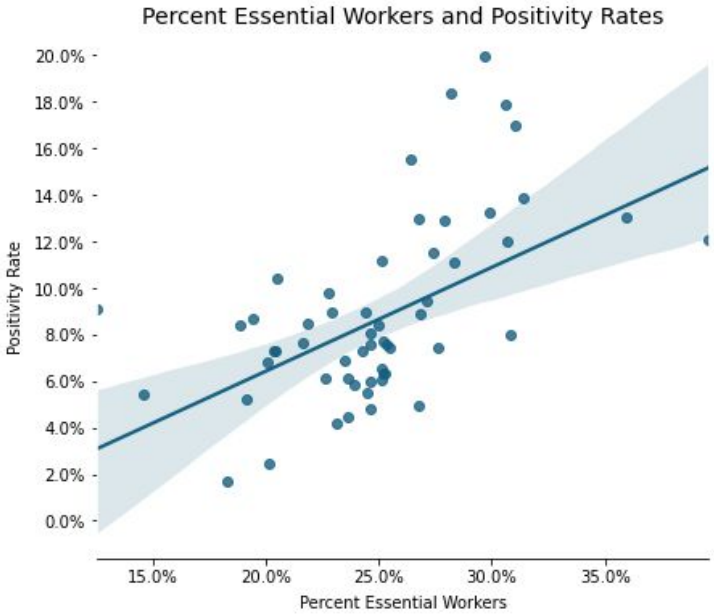


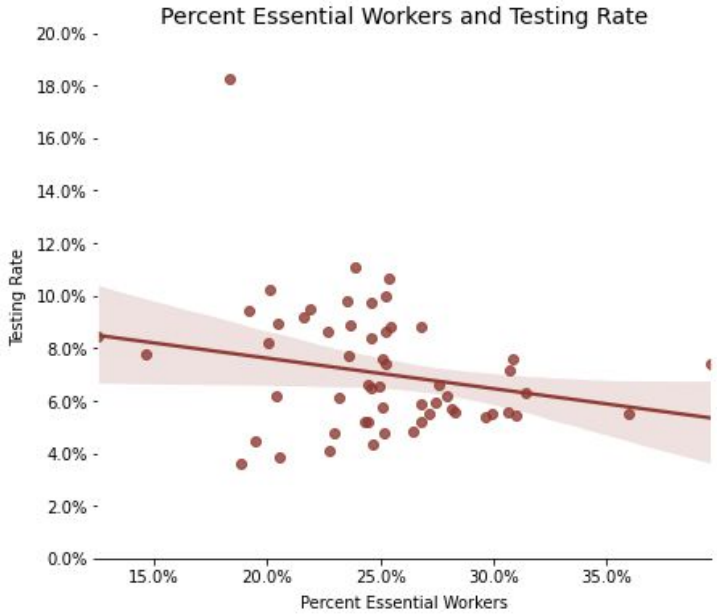


r = 0.306
p = 0.022*

Percent Essential Workers

r = 0.571
p = 0.000***

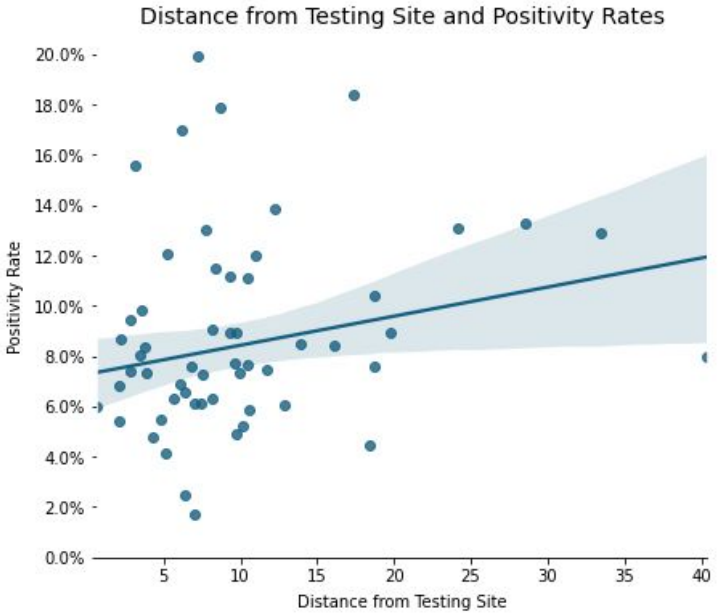


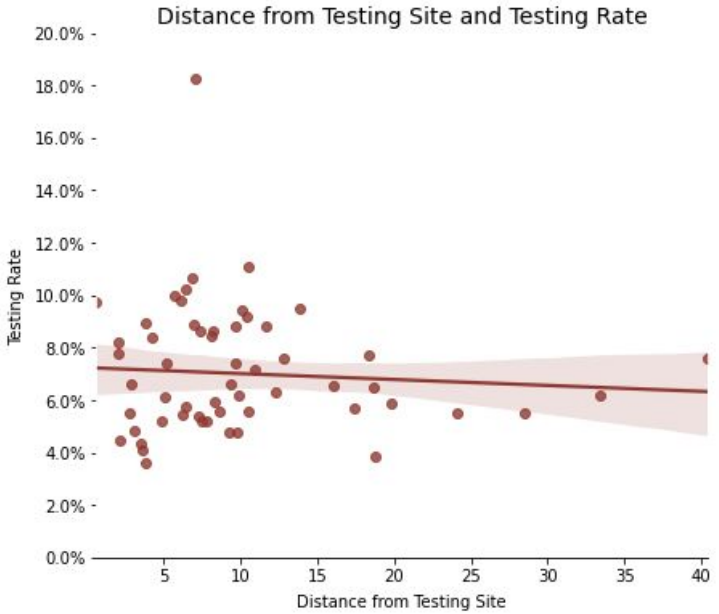


r = -0.277
p = 0.038*

Distance From Testing Site

r = 0.210
p = 0.121

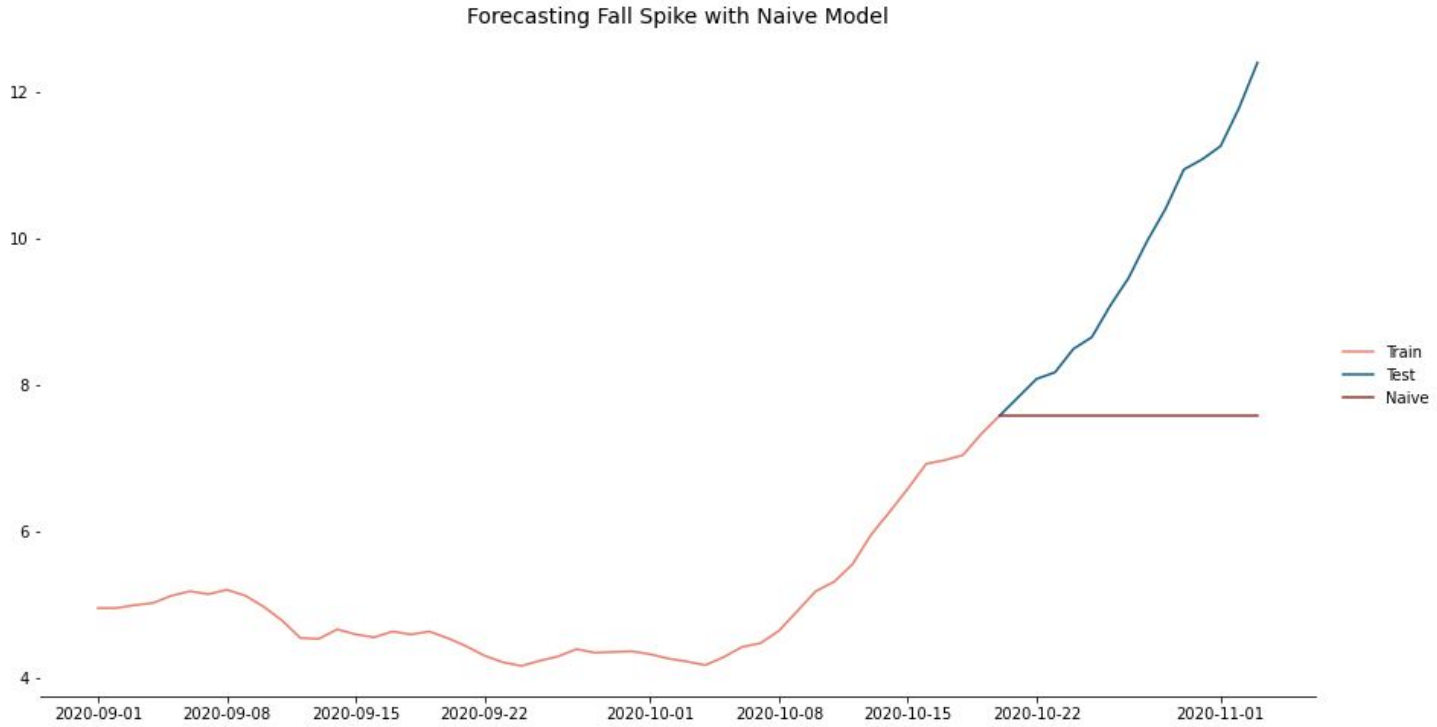




r = -0.094
p = 0.492

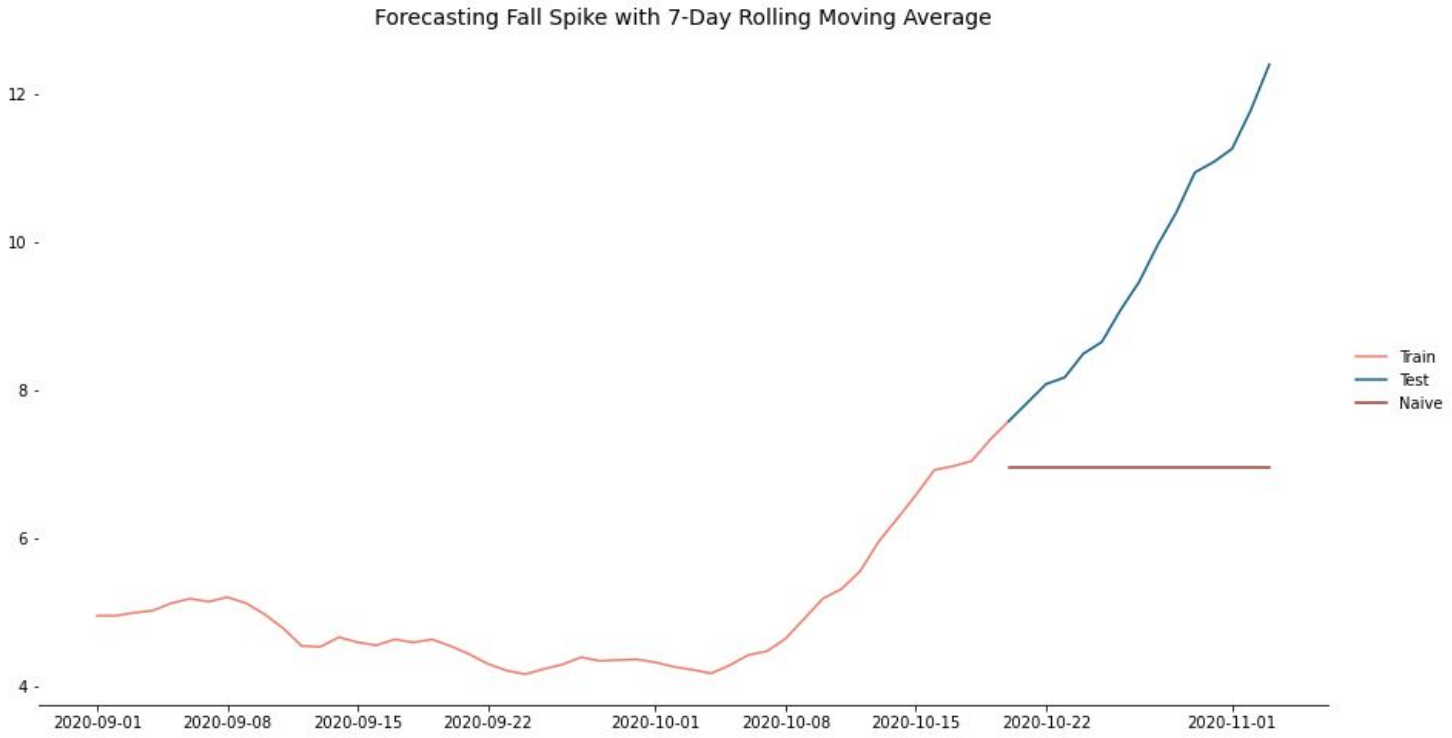
Time Series Models

Naive Model Forecast



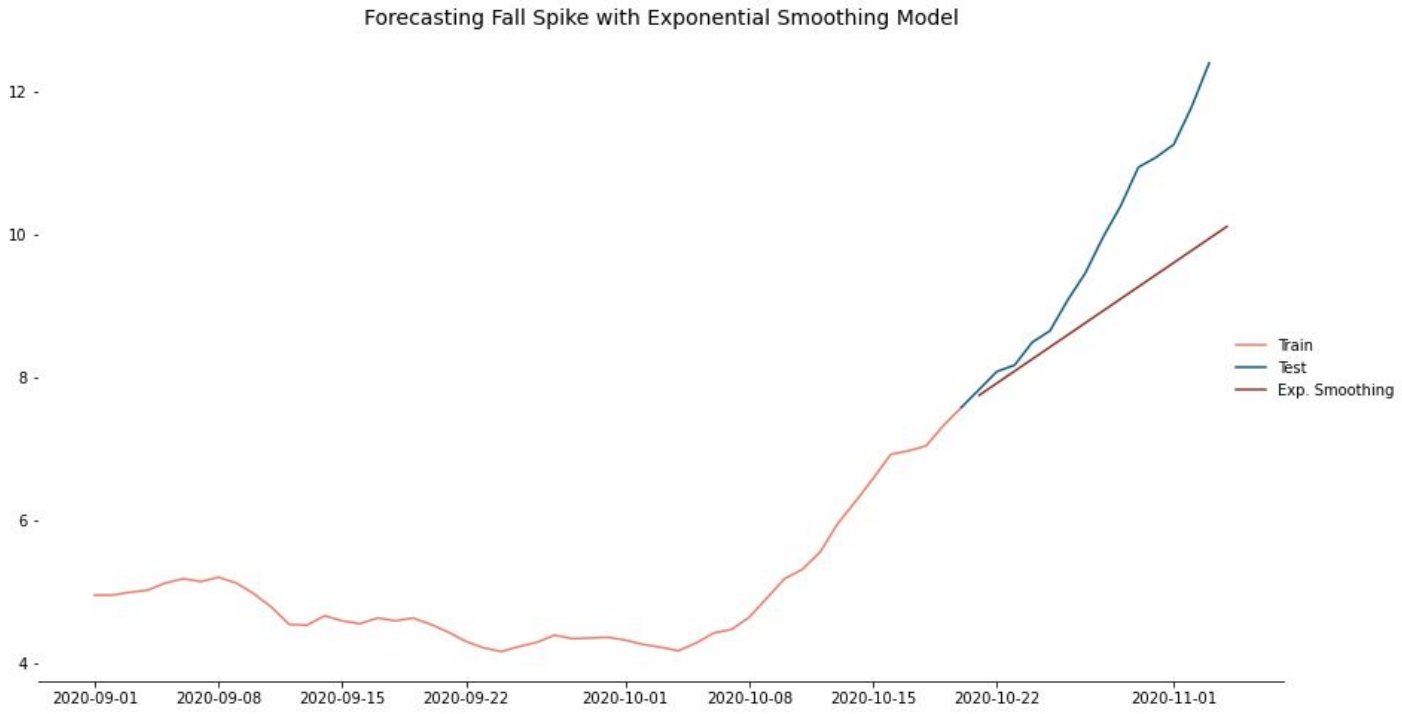
RMSE: 2.582

Moving Average Model



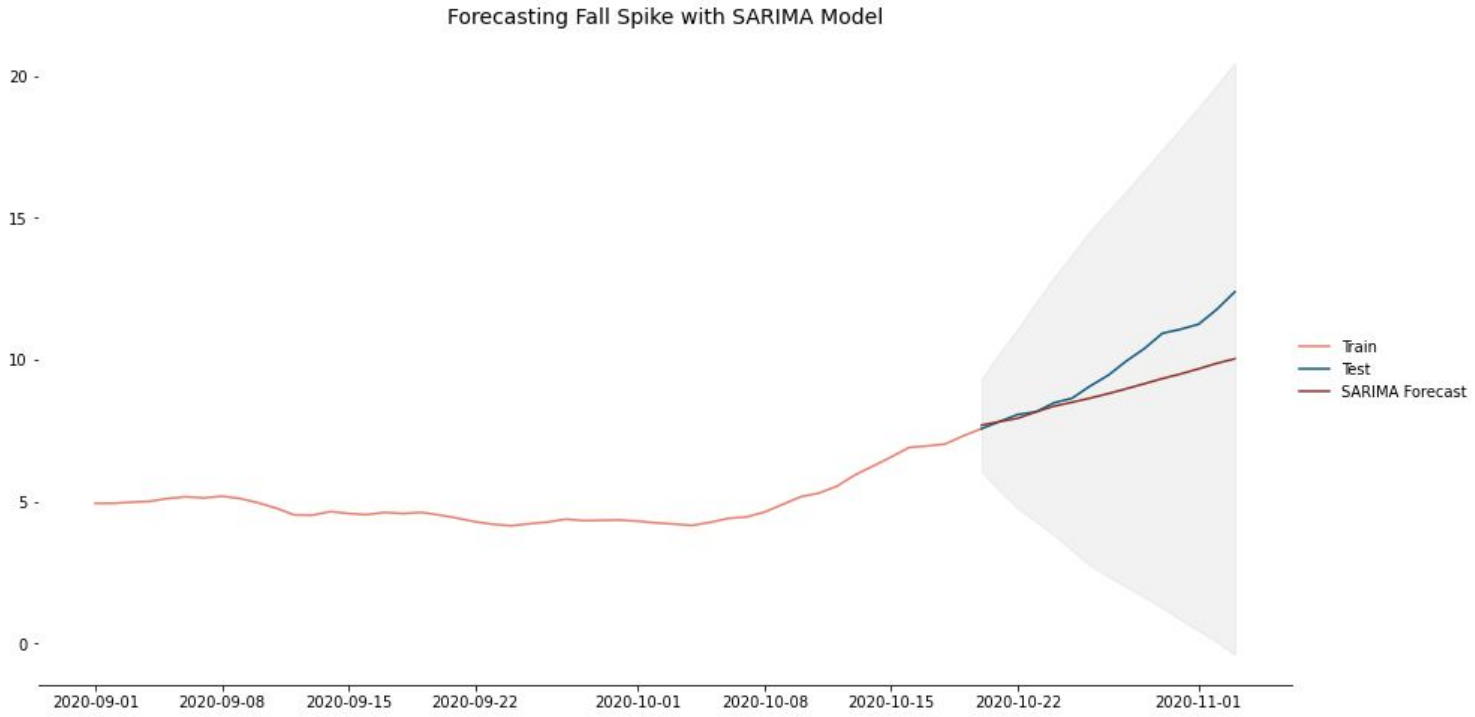
RMSE: 3.112

Triple Exponential Smoothing Model



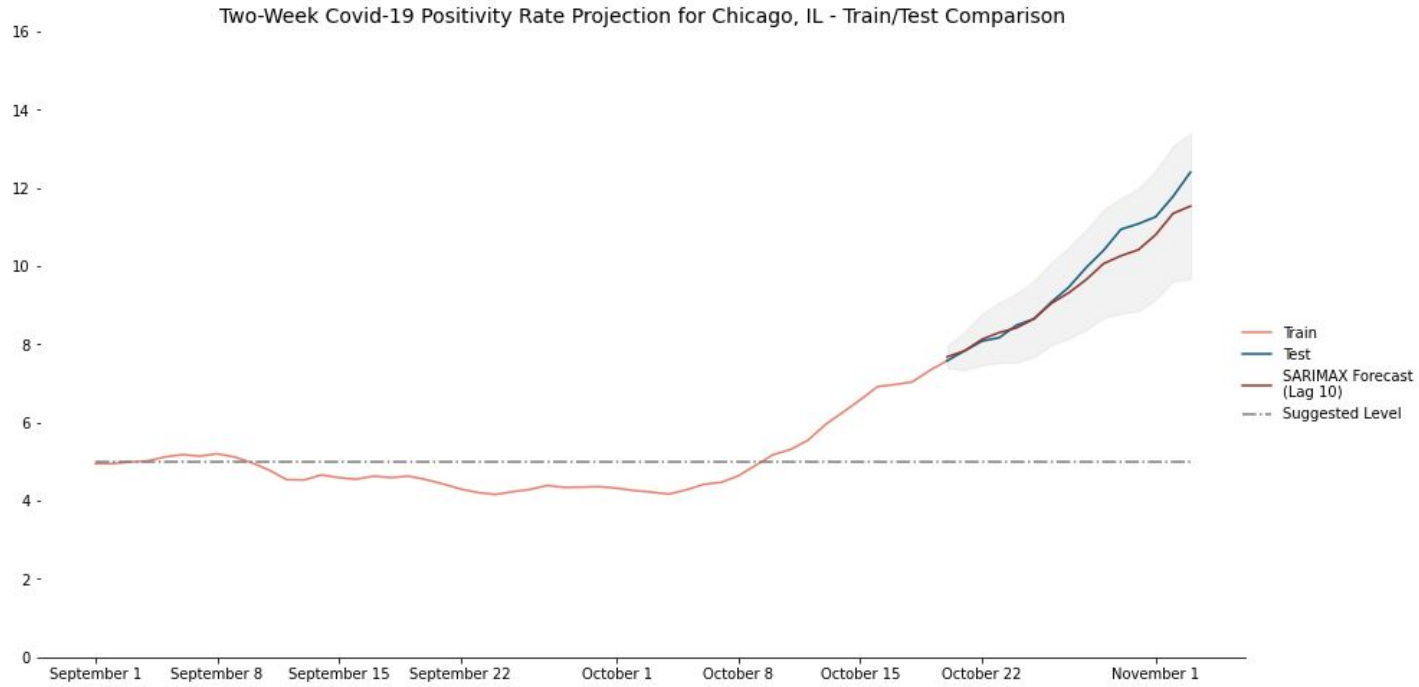
RMSE: 1.252

SARIMA Model



RMSE: 1.149

SARIMAX Model – Final Model



RMSE: 0.393