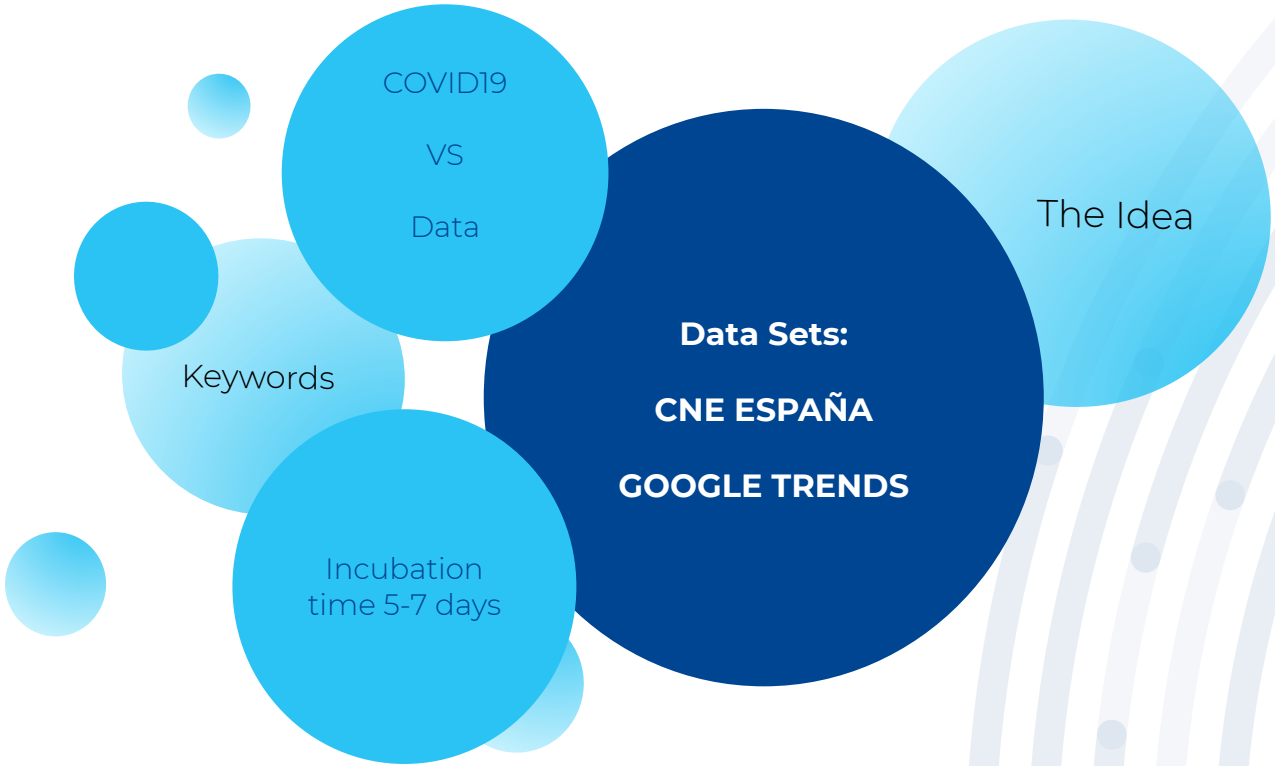
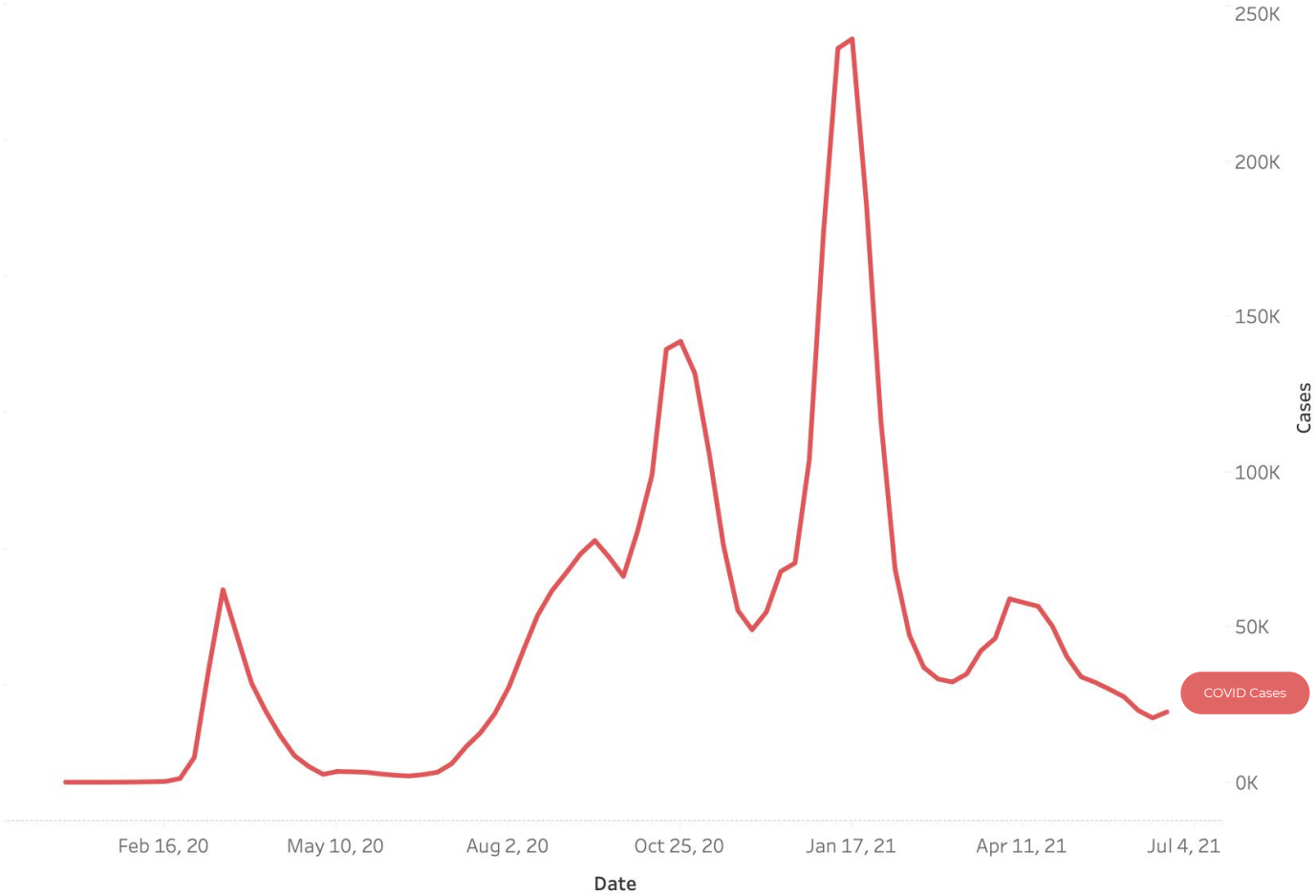
How Google Data could have predicted COVID-19 cases

WALTER CONDE

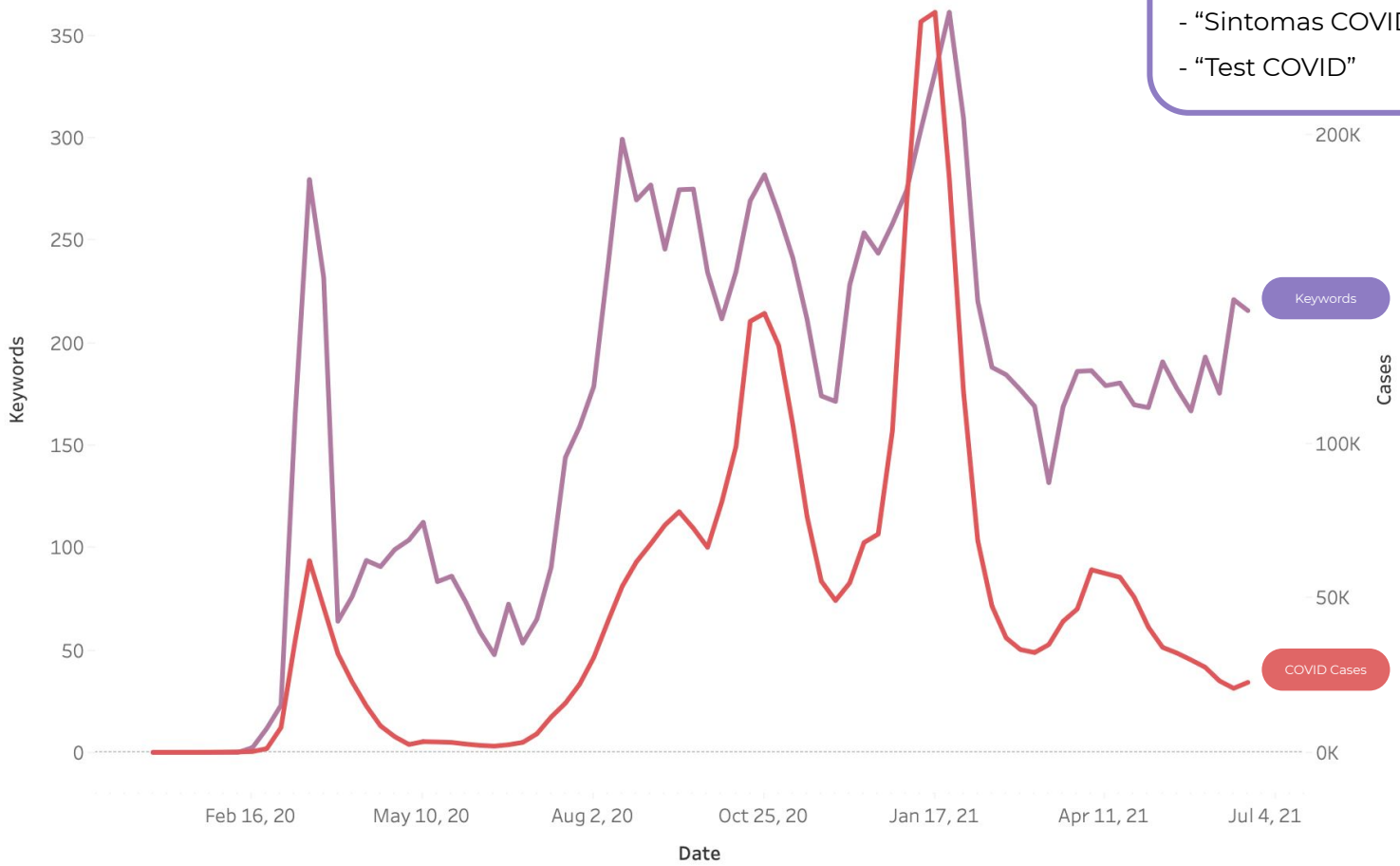
Don’t underestimate google data



Keywords


250K
200K
150K
100K
50K
0K
Cases
COVID Cases
Feb 16, 20
May 10, 20
Aug 2, 20
Oct 25, 20
Jan 17, 21
Apr 11, 21
Jul 4, 21
Date

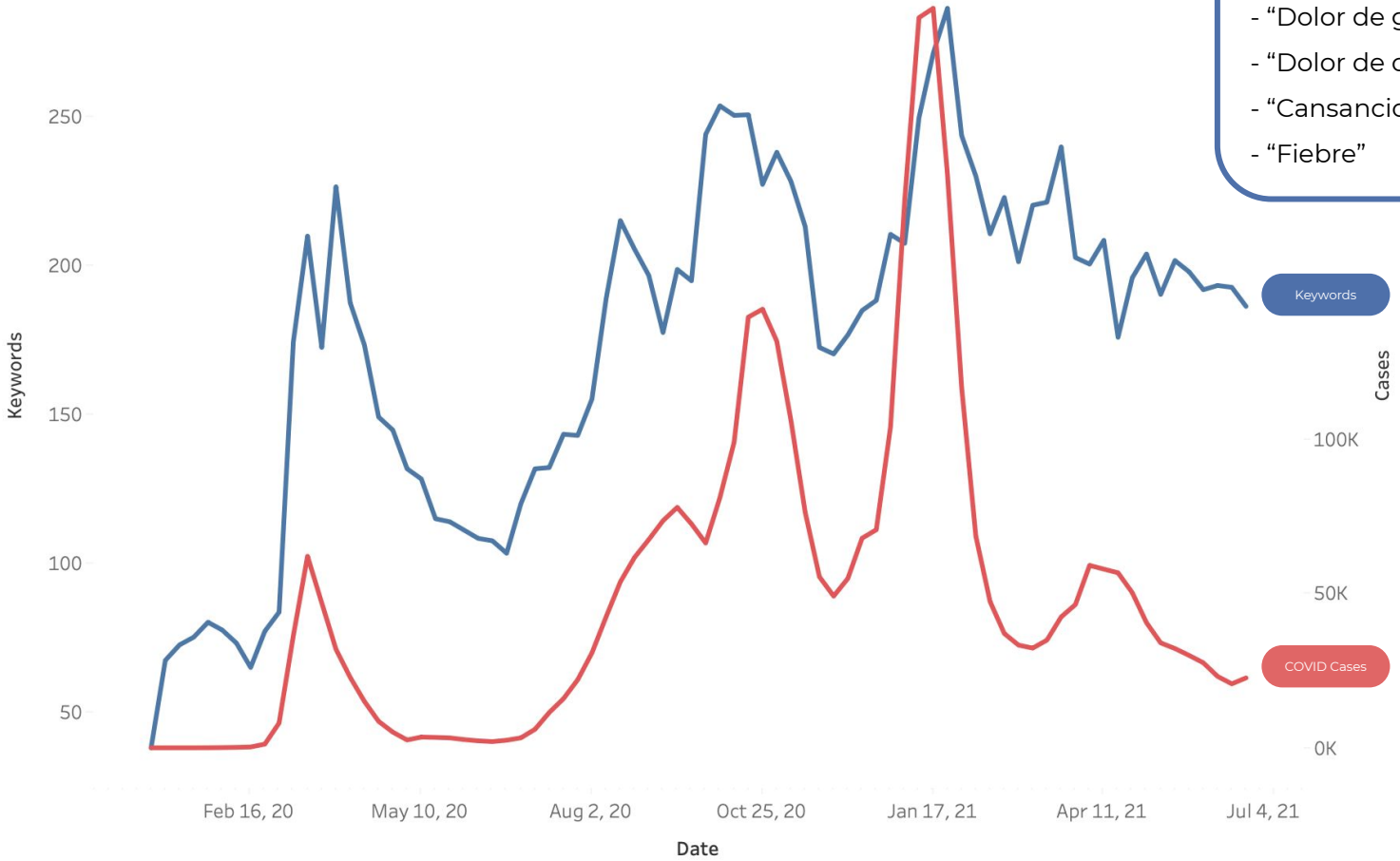
Keywords



Google searches:

- “Como saber si tengo COVID”
- “Sintomas COVID”
- “Test COVID”

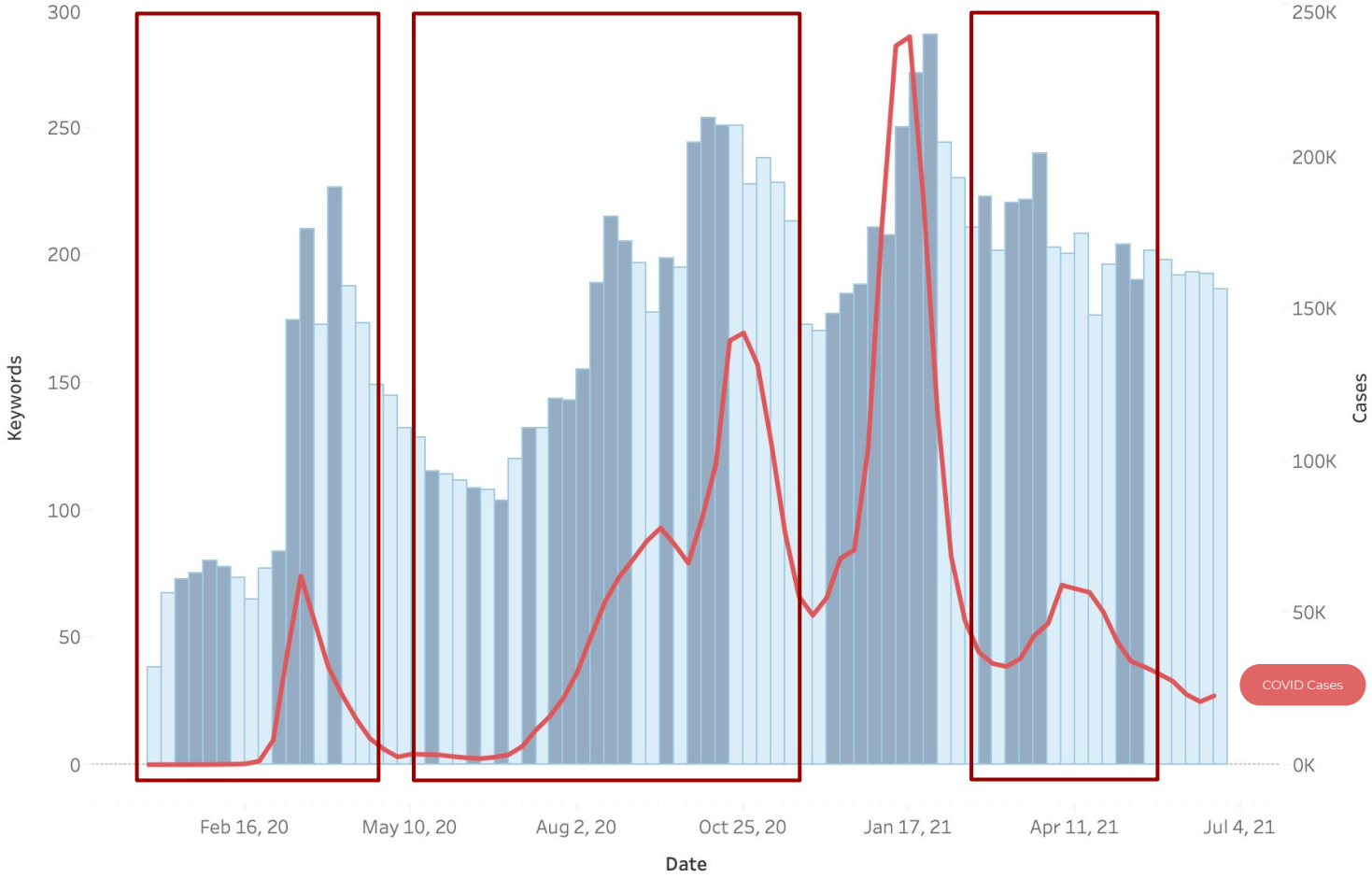
Keywords



Google searches:

- “Tos”
- “Dolor de garganta”
- “Dolor de cabeza”
- “Cansancio”
- “Fiebre”

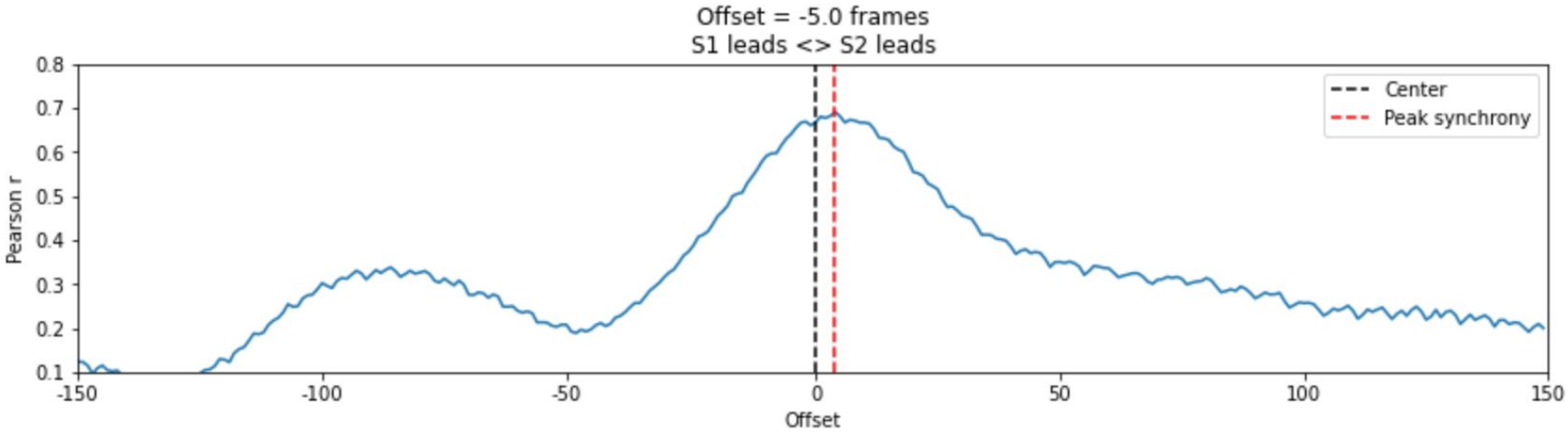
Keywords


300
250
200
150
100
50
0
Keywords
250K
200K
150K
100K
50K
0K
Cases
COVID Cases
Feb 16, 20
May 10, 20
Aug 2, 20
Oct 25, 20
Jan 17, 21
Apr 11, 21
Jul 4, 21
Date

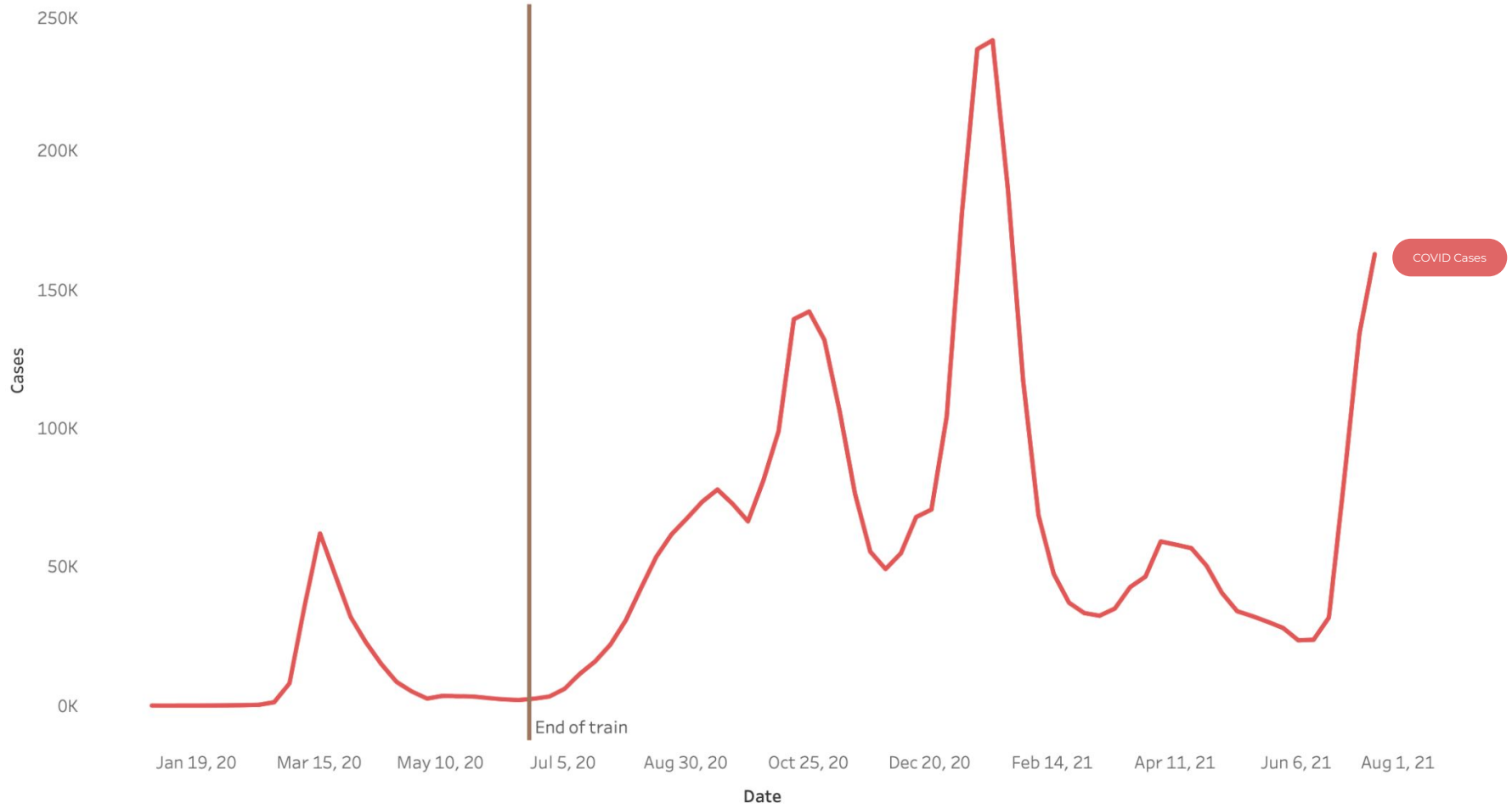
Time Lagged Cross Correlation

Is measured by incrementally shifting one time series vector and repeatedly calculating the correlation between two signals.

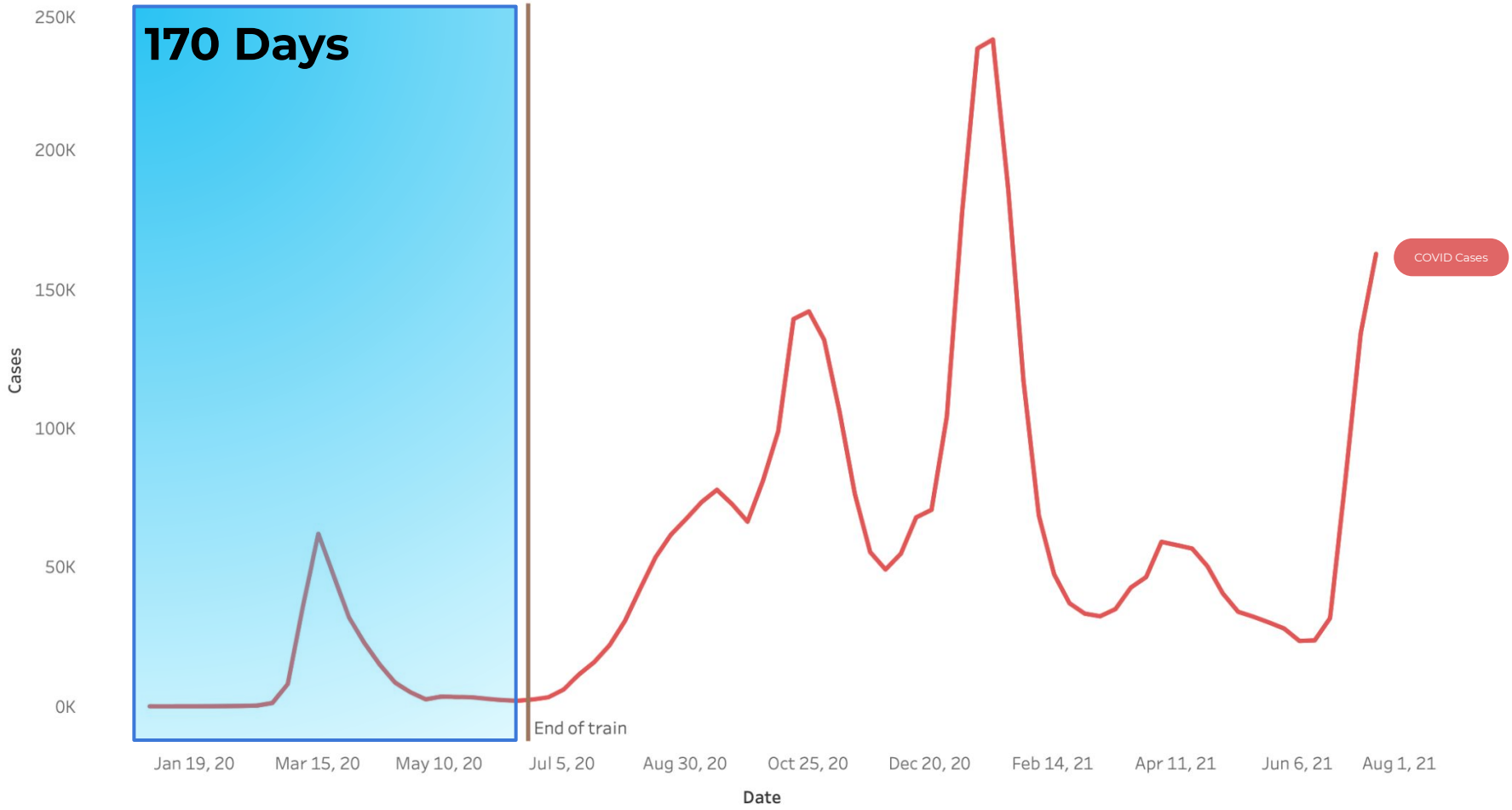
COVID/Keywords: Correlation is maximized when “COVID Cases” is pulled back by 5 frames (or 5 days)



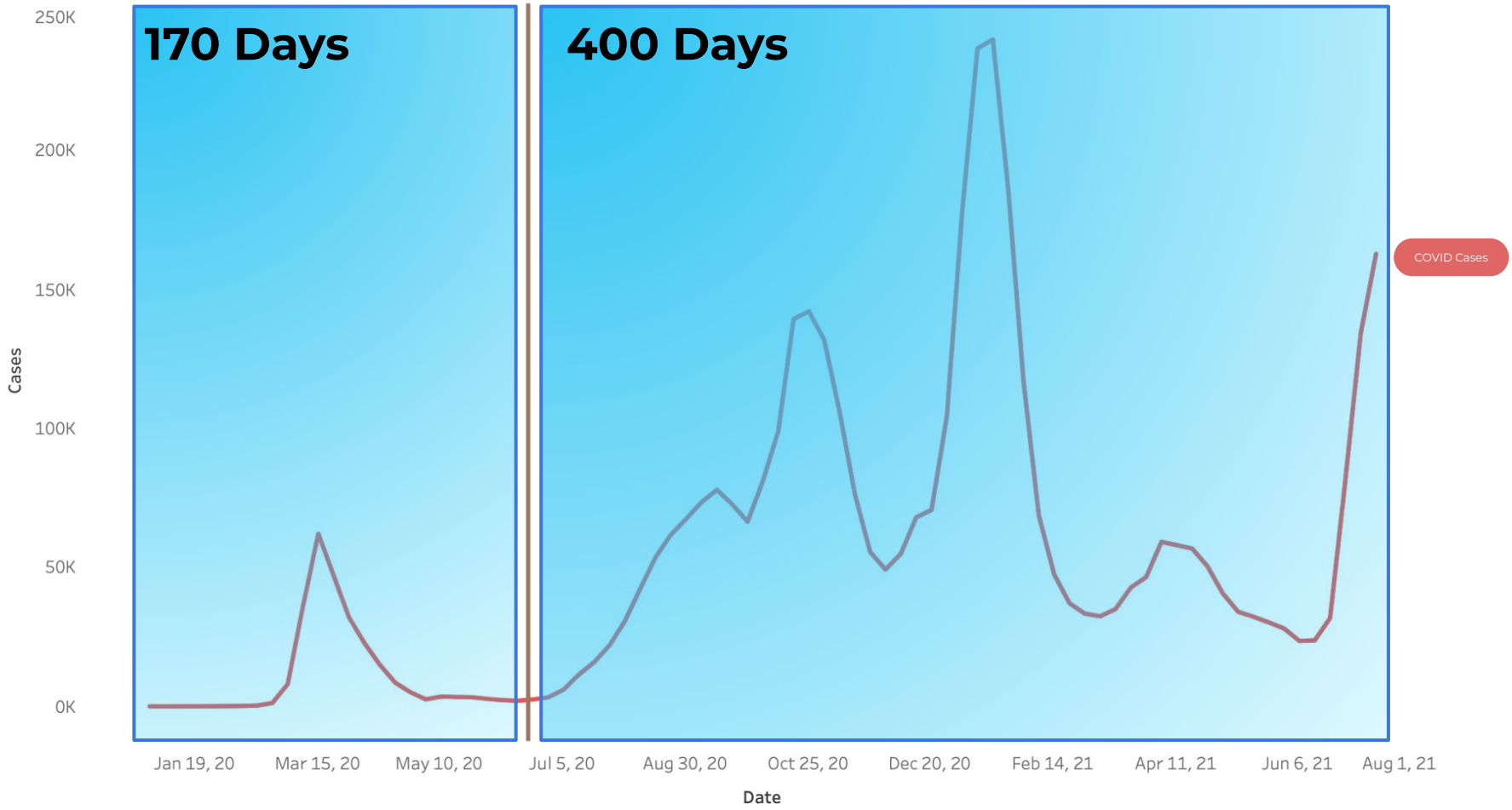
Predictions



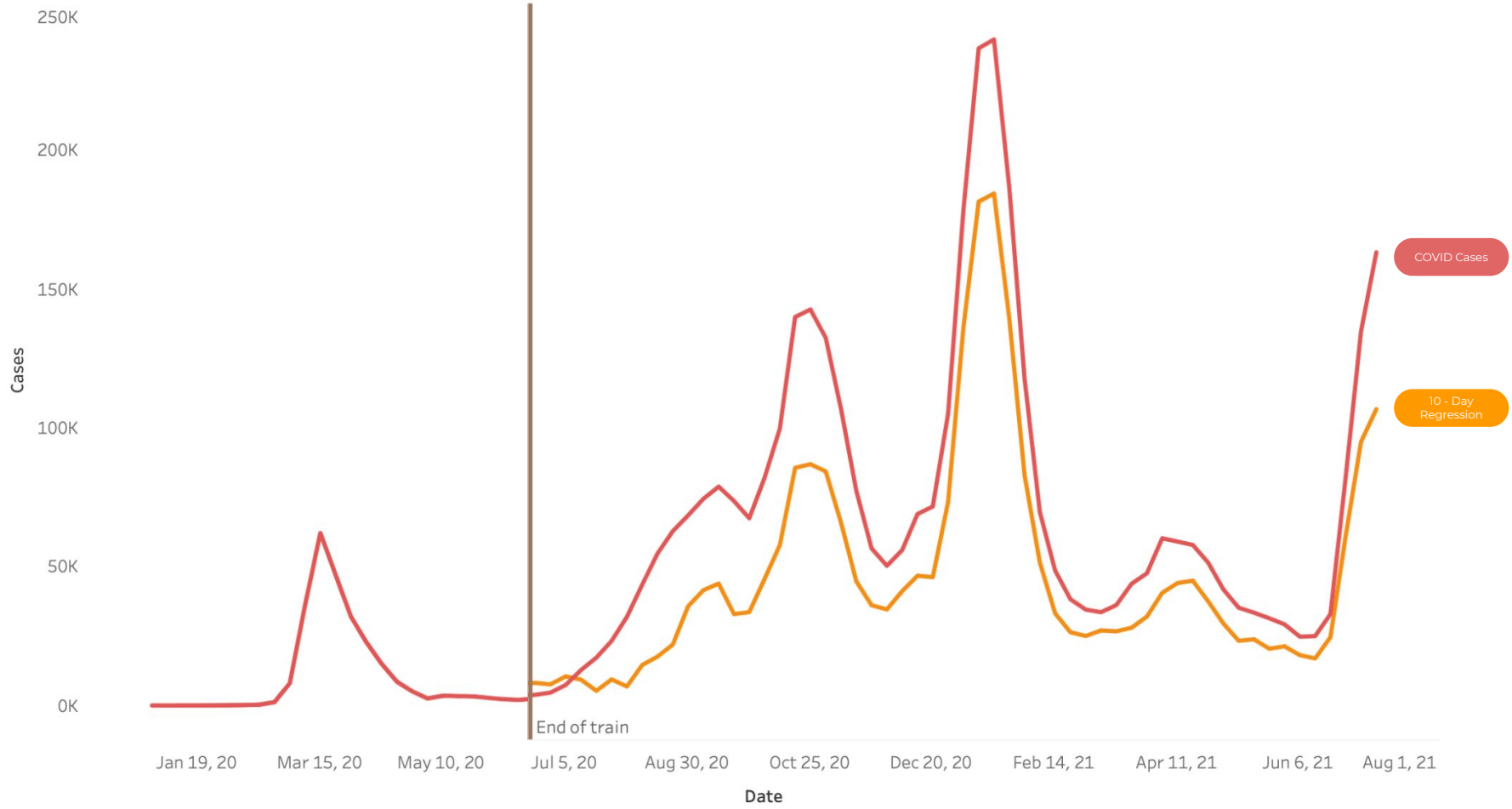
Predictions



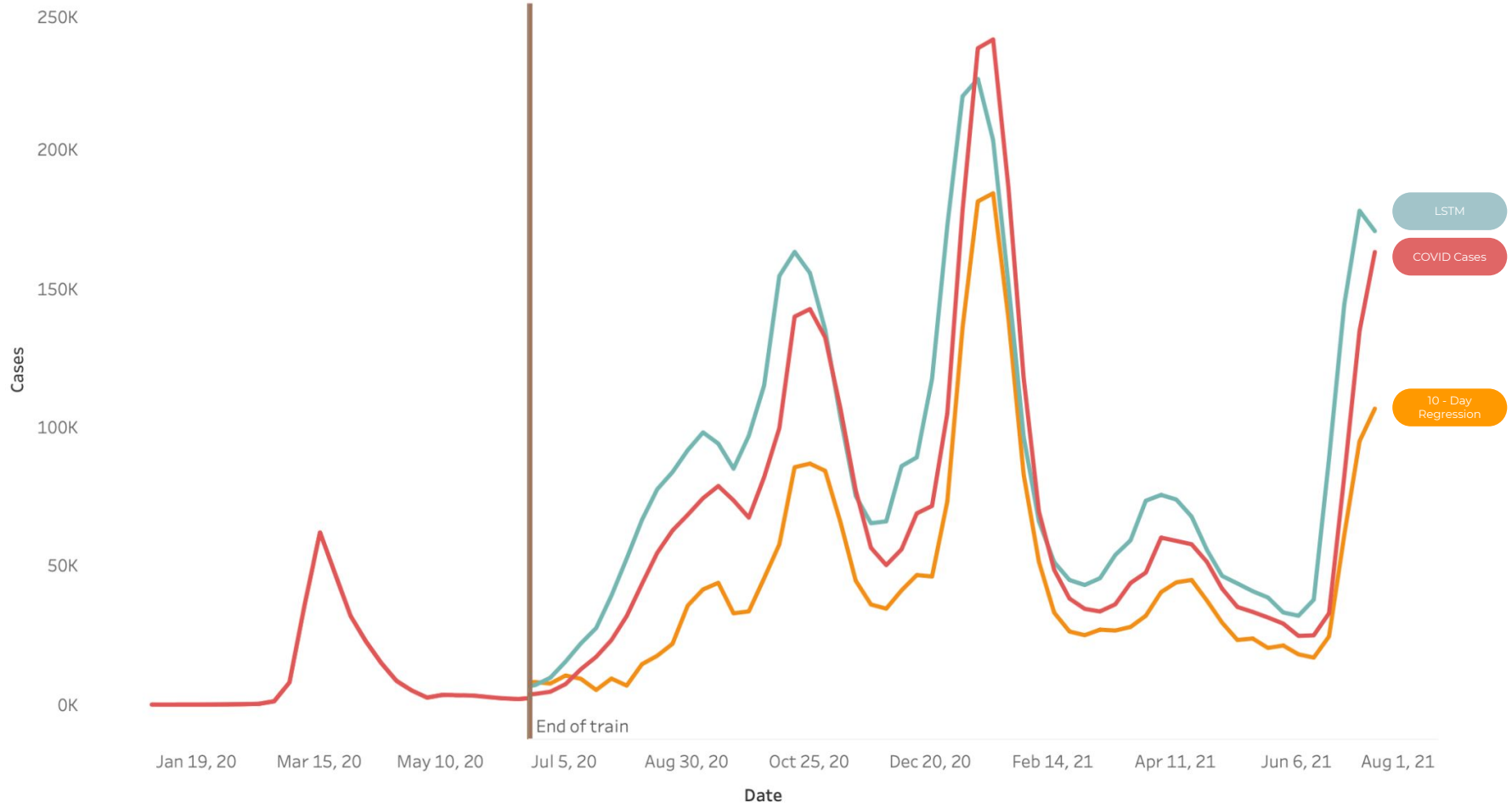
Predictions



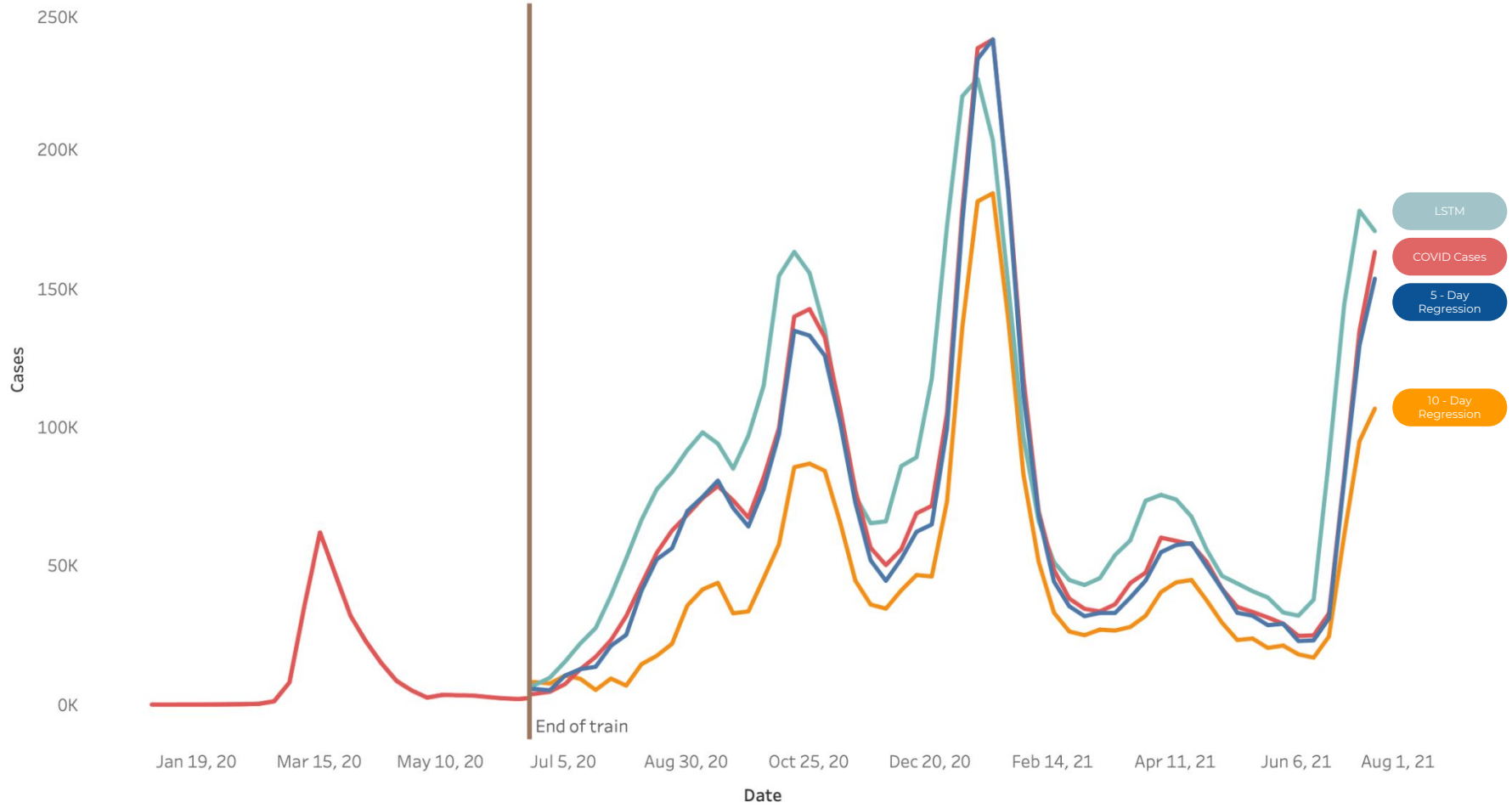
Predictions



Predictions



Predictions



Comparing the RMSE of the models

	RMSE AVG	RMSE 5 Days	RMSE 10 Days
Model 1 - Linear Regression	**4.671**	**3.364**	**5.978**
Model 2 - Time series LSTM	**4.729**		

27.661

New cases of COVID in Spain today (30/07/2021)

Thanks!

Any questions?