COVID-19 Effect: Mobility Report on Transportation and Transit Systems in California

Aris Briones

April 2, 2021

1 Introduction

In late December 2019, an outbreak of pneumonia of unknown origin in Wuhan occurred and set the stage for a global pandemic that changed the world. Since it's discovery, the virus has spread all around the world causing health and economic disasters leading to quarantine and stay-at-home orders. As a result, this directly affected the mobility and activity of businesses and transit systems. People are going out less, which means transit systems and transportation is seeing an all-time low in activity from people. Through machine learning and data analysis, we will investigate the activity of transit systems and what exactly contributes to the decrease or increase in activity of transportation. This will help transportation and transit systems prepare for the changes that come with COVID-19 and the aftermath.

2 Methodology

In this section, we will go over the process of everything that is done before analyzing the data and how the data is organized and compiled.

2.1 Data Summary: What is Mobility?

The data shows the changes in mobility of businesses and other places in California as a result of COVID-19. In other words, how visits to places, such as grocery stores and parks, are changing. It further shows how visits and length of stay at different places change for each date compared to the baseline times. This is done through percentage values with less activity giving negative values and increased activity giving positive values. Baseline times were determined through the median value during the 5-week period of Jan 3–Feb 6, 2020. It is calculated using the same kind of aggregated and anonymized data used to show popular times for places in Google Maps. Data also includes number of confirmed cases, new deceased cases, and new tested cases for each date. Refer to [1] for more info about the dataset.

2.2 Implementation Details

We first use the 'Epidemiology', 'Index', and 'Mobility' dataset from the GitHub repository [1]. We then join these three datasets. However, this contains data from all over the world but since we want to focus on only California we have to filter it by the state we want. After filtering our data and cleaning it so that we can work with it, we can begin visualizing and fitting our model. We will use Multiple Linear Regression, LASSO Regression, and LASSO Regression with Cross Validation for our machine learning algorithms.

To start fitting our model, we split our data into train and test sets. We take all the months before February to train our model, and use the month of February as our test set as our way to test our predictions. After getting our R^2 and MSE values we will compare the models and choose which machine learning algorithm performs best.

Predictor Variables:

- Number of confirmed cases, deaths, tested
- Mobility of: Workplaces, retail and recreation, parks, grocery stores and pharmacies, residential areas.

Response Variable:

- Mobility of transit systems

3 Exploratory Data Analysis

To give us a good idea of the variables we are working with, we take a look at the mobility of grocery stores and pharmacies, parks, transit stations, retail and recreation, residential areas, and work places.

Figure 1: Mobility of Places by Date

We also can take a look at the new confirmed cases, new deaths, and new tested for perspective on how increase in these cases can affect mobility.

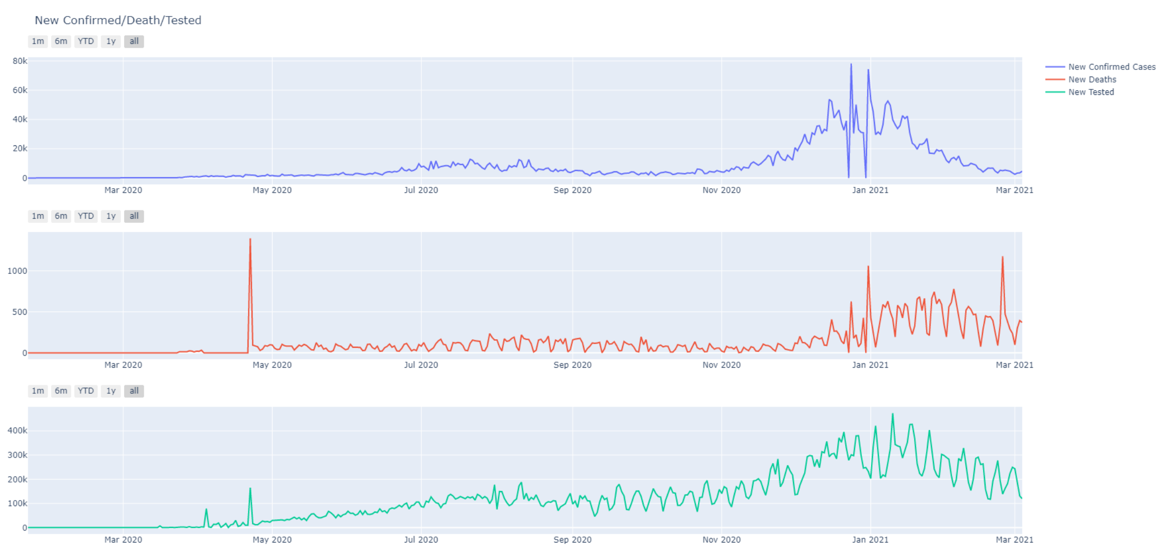

Figure 2: New Confirmed Cases, Deaths, and Tested by Date

4 Results and Interpretation

In this section, we will compare our results and interpret our findings.

4.1 Comparison of Models: Model Selection

MODELS	Multiple Linear Regression	LASSO	LASSO with Cross Validation
MSE	3.65	3.65	3.69
R-Squared Value	0.869	0.868	0.867

Figure 3: Comparison of Models: R-Squared and MSE

Higher R-Squared values are better whereas lower MSE values are better. After comparing our models, the three machine learning algorithms all perform very similarly with Multiple Linear Regression performing slightly better. Thus, the Multiple Linear Regression model is best at predicting transit system mobility.

4.2 Interpretation

```
Coefficients_______________:
new_confirmed: 3.692705651361406e-05
new_deceased: -0.0027233600801621037
new_tested: -2.7858516453619416e-05
mobility_grocery_and_pharmacy: 0.050015123086535905
mobility_parks 0.137666868973597
mobility_residential: 0.24167934967282423
mobility_retail_and_recreation: 0.47485093907693754
mobility_workplaces: 0.4130253463667927
```

Figure 4: Beta Coefficients for Multiple Linear Regression Model

Our coefficients tell us exactly how that variable affects the overall outcome of people's activity in transportation and transit stations. We can see here that the higher the number of confirmed cases, deaths, and tests from COVID-19 is, the lower mobility of transit systems are. This obviously makes sense. However, it's important to keep in mind that number of deceased is often a lot lower than number of confirmed cases, and number of tested is often a lot higher than number of confirmed cases.

The mobility of other venues such as grocery stores tell a different story however. Since mobility can give negative values just like our response, there is a linear relationship between the mobility of other areas and transit systems. This makes sense since the more that people are going out, the more likely transit systems are being used. For example, if grocery store activity is negative, then it will have a negative affect on transit systems as well. One thing to also note is the effect mobility of workplaces has on transit systems. It's one of the higher coefficients meaning that as people transition from Work-From-Home to offices, there will be a sharper increase in transit systems.

5 Conclusion

One thing that becomes obvious from this analysis is how much transit system mobility relies on other areas of mobility. Right now, transit systems are suffering at an all-time low, but once other areas of mobility increase then transit-systems will see a demand in use. It's important to recognize this pattern since transit systems and schedules need to be prepared for the transition to a post-pandemic period. As life returns back to normal, transit systems will only follow.

References

[1] O. Wahltinez et al. "COVID-19 Open-Data: curating a fine-grained, global-scale data repository for SARS-CoV-2". In: (2020). Work in progress. URL: https://goo.gle/covid-19-open-data.