Research paper

A machine learning approach to predicting bicycle demand during the COVID-19 pandemic

Carolina Baumanis [*], Jennifer Hall, Randy Machemehl

The University of Texas at Austin, Department of Civil, Architectural and Environmental Engineering, 301 E. Dean Keeton Street, Stop C1761, Austin, TX, 78702, USA

ABSTRACT

Count-based bicycle demand models have traditionally focused on estimation rather than prediction and have been criticized for lacking a direct causal relationship between significant variables and the activity being modeled. Because they are not *choice*-based models, they are doubted for their ability to forecast well. The rise of machine learning techniques has given researchers tools to build better predictive models, and the tools to evaluate predictiveness. Extensive previous work in the statistics and machine learning field has shown that the best predictive model is not synonymous with the most *true* (or explanatory) model. The non-motorized demand modeling community could leverage these lessons learned to develop better count-based predictive models.

The rise of the COVID-19 pandemic has clearly affected travel patterns, and the broad data collection has opened-up an opportunity to leverage machine learning techniques to build a predictive bicycle demand model. This study uses bicycle count data, COVID-19 data, and weather data to develop a LASSO regression model for three facilities in Austin, TX. The COVID-19 variables included both state- and local-level data between March 15, 2020, and January 31, 2021. The final model selects six variables out of 28 variables and reveals that the increase of statewide COVID-19 fatalities, statewide molecular positivity rate, and local precipitation cause a decrease in bike ridership, meanwhile maximum temperature causes an increase. The LASSO model also has a lower prediction MSE during cross validation compared to the full model. This paper aims to bring to light that our present-day demand and volume forecasting efforts would benefit tremendously from a predictive modeling approach rather than valuing the most explanatory models as the only strong forecasters of demand. In the end, modelers can use this approach to improve the forecasting ability of any count-based bicycle demand model.

1. Introduction

The pandemic has been a catalyst for cities to invest more toward equitable, healthy, and sustainable transportation options, like bicycling and walking. At the start of the pandemic, the CDC labeled public transportation as high risk for contracting and spreading the COVID-19 virus, and bicycling and walking as an overall low risk activity (Centers for Disease Control and Prevention, 2019a; Centers for Disease Control and Prevention, 2022). A study conducted by People for Bikes highlights that the US saw a surge in bicycling rates in 2020 across all demographics and approximately 200 cities responded by changing street functionality to accommodate the increased outdoor activity (People For Bikes, 2021). Planning efforts seeking to implement bicycle facilities must accurately predict bicycle ridership to maximize project benefits. Predicting future bicycle demand helps decision-makers pick appropriate facilities and prioritize projects according to benefit/cost ratios.

In the end, having an *accurate* demand forecasting model can save planners money in construction costs, maximize health and safety benefits for users, and facilitate decision-making. Traditionally, count-based demand models have been doubted for their ability to predict bicycle ridership for various reasons, including their lack of connection to direct causal factors, but they are widely used in practice (National Academies of Sciences, 2014).

Previous research in machine learning has taught us that the model that best represents the data-generating process is not necessarily the most predictive (Shmueli, 2010; Yarkoni & Westfall, 2017). In general, transportation demand estimation has almost always only focused on choice-based models because it seems that prediction would require explanation. However, machine learning has shown how a model that fails to capture all causal mechanisms can outperform a more complex, more accurate model in terms of prediction (Brand et al., 2020; Shmueli, 2010) because of a phenomenon known as *overfitting*. With increasing

access to data and data collection methods, count-based demand models could become very powerful predictive tools by implementing this lesson learned in machine learning. This paper aims to bring to light that our present-day demand and volume forecasting efforts would benefit tremendously from a predictive modeling approach rather than valuing the most explanatory models as the only strong forecasters.

Since the start of the COVID-19 pandemic, state, local, and territorial health departments have been collecting case data that can be leveraged to help predict bicycle volumes. To understand which COVID-19 variables are best at predicting bicycle ridership, this paper uses a machine learning method called LASSO regression. For this study, the variables considered for building a predictive model are COVID-19 case data, vaccination data, government ordinances (city and state level), and weather. Both COVID-19 and weather data are routinely collected and should be easily accessible to anyone. Other cities and transportation authorities can adopt this method to select a combination of variables out of all available variables that are best suited for predicting bicycle ridership at locations of interest.

2. Literature review

Heightened anxiety over public transportation and an increased focus on getting exercise has led to a boom in bicycle demand and bicycle riding during the COVID-19 pandemic (United States Census Bureau, 2021). Traveling on public transportation is considered high-risk for contracting and spreading COVID-19 due to close contact and exposure to frequently touched surfaces (Centers for Disease Control and Prevention, 2019a). On the other hand, the Centers for Disease Control and Prevention (CDC) states that bicycling and walking outdoors are among the lowest risk activities in terms of contracting and spreading COVID-19 (Centers for Disease Control and Prevention, 2019b). Bicycling provides a way for people to both travel and get exercise with a very low risk of getting the virus.

Using COVID-19 data in a predictive bicycle demand model could help with understanding how the pandemic influences future bicycle ridership at bicycle facilities. Conventionally, count-based demand modeling approaches have not really been evaluated for their actual predictive ability (National Academies of Sciences, Engineering, and Medicine, 2014). Evaluating predictive ability requires some sort of cross validation step where the model is tested with data it has never seen before to see how well it performs, which is not currently standard practice. The remainder of this literature review summarizes previous findings on bicycle demand and the pandemic, and then explains the value of focusing on building predictive rather than explanatory models for demand forecasting purposes.

2.1. Bicycling and the pandemic

Since the start of the pandemic, the communication between public health officials and the public on the risk of getting and spreading the virus has influenced whether people decide to participate in lower or higher risk activities. Public health officials continuously update and communicate risk-based guidelines based on the waxing and waning of COVID-19 cases. Stay-at-home orders and awareness of the spread of the virus has caused people to make changes in their daily travel patterns and exercise habits.

2.1.1. Pandemic bicycling: bicycle boom
The COVID-19 pandemic caused a bicycle renaissance that has triggered significant spikes in bicycle sales and ridership in the US and abroad. As a low-risk activity for contracting the virus and an overall beneficial activity for mental and physical health, bicycle sales and ridership soared between the months of April and May 2020 (Powell, 2020). Within those months, bicycle sales saw their biggest spike in the U.S. since the oil crisis of the 1970s (Sharp & Chan, 2020). Even as the pandemic reached September 2020, the bike boom still did not let up. In the U.S., bicycle ridership was still up 11% by September as the typical driver rate continued to be down by 6.5% (Wilson, 2020). According to recent research, 1 in 10 American adults reported having ridden a bike for the first time in a year (or longer) since the onset of COVID-19 (People For Bikes, 2021). Not only that, approximately 200 U.S. cities changed the functionality of their streets to accommodate increased outdoor activity (People For Bikes, 2021). The growth of U.S. bicycle sales declined a 4% in 2021 and is still 45% higher overall than it was in 2019 (Sorenson, 2022). On the international side, Italy even rewarded the purchase of new bikes or bike maintenance expenditures in a bid to encourage ridership during the pandemic. The Italian Government in May 2020 offered a voucher of 500 euros for the purchase of a bicycle, including pedal assisted bicycles, and vehicles for electric micro-mobility such as scooters, hoverboards and segways for the use of sharing mobility services (Ministry of Sustainable Infrastructure and Mobility).

2.1.2. Pandemic bicycling: demand
An exploratory study of outdoor physical activity in Houston, Seattle, and New York City found that people continued to use trails during stay-at-home orders (Doubleday et al., 2021). In Houston, nearly all locations saw significantly more daily bicycle and pedestrian counts during the stay-at-home period. The results in Seattle varied by local context and trail type (commuter versus recreational). Meanwhile, New York City, did see increases in bicycle counts just prior to the stay-at-home order, but decreases once the order was in place. The differing results are likely because the level of transmission in spread in New York City was much worse compared to Houston and Seattle during the analysis period.

Cities across Europe also experienced an increase in bicycle ridership (Bucsky, 2020; Möllers, Specht, & Wessel, 2021). Researchers have reviewed automatic counting stations located in 10 major cities within Germany during the pandemic. This study focused on measuring and analyzing the impact of COVID-19 and governmental interventions on active mobility (Möllers, Specht, & Wessel, 2021). In the end, almost all cities saw an increase in bicycle ridership but cities that do *not* already have a large cycling presence (large cycling modal share) within the transportation infrastructure experienced a larger increase in bicycle ridership than cities that already had a large presence of cycling (Möllers, Specht, & Wessel, 2021). Researchers hypothesized that the cities with a minor cycling presence experienced a larger increase in ridership because of public transportation being their main form of modal share (as noted as a place of high-risk for contracting COVID-19), therefore, people shifted their mode of travel to cycling to mitigate that risk. Clearly, COVID-19 has had a tremendous impact on bicycle riding both in the U.S. and internationally. COVID-19 data along with bicycle counts could be used to develop a predictive demand model for cyclists. Clearly, COVID-19 has had an influential impact on bicycle ownership and ridership.

2.1.3. Pandemic bicycling: policy relevancy
Since the pandemic bicycle boom, many cities have committed to shifting temporary cycling infrastructure to permanent (Institute for Transportation and Development Policy, 2022). With that, bicycle demand forecasting can help with prioritizing *which* temporary infrastructure changes should be made permanent and even further enhanced based on benefit-cost analyses. Ultimately, this type of information would guide investments in these facilities that improve the safety and health of bicyclists, therefore increasing the associated economic benefits (University of Minnesota, 2021). However, currently count-based bicycle models, which are widely used in practice as a way for forecasting, do not systematically prioritize nor evaluate predictive ability. Just as in traditional inference-based regression modeling, count-based modeling focuses on the significance of estimates and relationship between the dependent (bicycle counts) and independent variables rather than measuring the true predictability of the model. Predictive power is measured by quantifying how well a model preforms using data it has

never seen before with cross validation. The following section will provide evidence as to why it is time to shift to a prediction mindset rather than estimation for forecasting bicycle facility demand.

2.2. Count-based bicycle demand models

Count-based bicycle and pedestrian demand models have traditionally relied on regression techniques to relate independent (explanatory) variables to counts, and have been doubted for their ability to forecast demand (National Academies of Sciences, Engineering, and Medicine, 2014). These count-based or *direct demand* models have been praised for being good at explaining levels of non-motorized activity based on factors that describe the local context but criticized because the variables that are significant many times have little causal relationship with the activity being modeled. The standard four-step models used for regional planning and forecasting, the other *choice-based* demand modeling option, are based on Traffic Analysis Zones (TAZs), which really are not small enough to capture non-motorized travel (National Academies of Sciences, Engineering, and Medicine, 2014). TAZs are a unit of geography that are commonly used for travel demand modeling. These units represent the spatial distribution of trip-origins and destinations, as well as the population, employment and other spatial attributes that generate or otherwise influence travel demand (Miller, 2021). Utilizing the same population restraints as the U.S Census Bureau's block groups, TAZs contain a population of more than 1200 people but less than 3,000, and typically range between 0.25 to one square mile in area (Florida Department of Transportation Systems Planning Office, 2007). In the end, the tools for forecasting and predicting non-motorized travel are limited.

2.2.1. Prediction versus estimation

A major goal in transportation demand modeling is to understand human behavior, which many times has translated into trying to explain behavior rather than truly trying to predict it. The four-step model has really been the trusted way for planners to model the causal mechanisms that lead people to decide among travel alternatives, but this does not work well for forecasting non-motorized travel because, as mentioned before, TAZ's are not granular enough to capture these short trips (National Academies of Sciences, Engineering, and Medicine, 2014). Meanwhile, the count-based modeling approach has not really been considered a serious contender in the forecasting toolkit because demand modelers generally value the understanding of the behaviors leading to travel decisions more than anything. It really does seem that explanation and prediction are interlaced – how can one happen without the other? But, in a statistical and practical sense, explanation and prediction sometimes are two opposing forces (Shmueli, 2010; Yarkoni & Westfall, 2017). There is a difference statistically-speaking between *explanation* – finding all variables that are significant – and *prediction* – accurately forecasting behaviors that have not been observed.

The model that most closely resembles the data-generating process will not necessarily be the best at predicting real-world outcomes (Shmueli, 2010; Yarkoni & Westfall, 2017). Because of overfitting, a more accurate complex model many times can be outperformed by a simpler, less comprehensive model (Shmueli, 2010; Wu et al., 2007; Yarkoni & Westfall, 2017). Overfitting is when a model fits the model too well and learns the dataset's minor details and noise, therefore does not work well to outside of the data used to build the model. Other disciplines have applied the lessons learned in machine learning and see a great benefit to prioritizing prediction and leave explanation as a secondary goal (Yarkoni & Westfall, 2017). Of course, explanation will always remain an important research goal, but it should not be the only goal. Choosing models that do a thorough job of explaining behavior seems great in theory, but it does not guarantee that they will predict well (Shmueli, 2010; Yarkoni & Westfall, 2017). Because ideal explanatory science is not ideal predictive science, then one must make a deliberate choice to explain or to predict (Yarkoni & Westfall, 2017).

Most disciplines dealing with human behavior have really opted to fall on explanatory approaches over *black box* machine learning approaches, but fortunately there are some machine learning techniques that are interpretable. Machine learning tools have not really been popular in the past because 1) understanding of machine learning science was poor and 2) there was not too much availability of large-scale datasets. Now, that data availability and machine learning understanding has evolved, it makes sense that transportation modelers start trying to prioritize predictive approaches that can lead to simpler, more predictive models. One method that is especially attractive for direct demand modeling is the LASSO method, which will be discussed in further detail in the Methodology section of this paper. The technique was first introduced by Tibshirani in 1996 and later used in air pollutant prediction, epidemiological studies (Fung et al., 2021), and it is also picking up traction in psychology (Yarkoni & Westfall, 2017).

3. Methodology

To develop a predictive count-based bicycle demand model, we use the LASSO regression technique to reduce model complexity and increase prediction accuracy by preventing overfitting. The LASSO regression method helps pinpoint the variables that are the most predictive of bicycle counts by only keeping relevant variables and shrinking the rest to zero. Since there are up to 28 COVID-19 and weather variables that could work for a bicycle volume model, using a tool that increases model interpretability and prediction ability has a tremendous benefit. A 10-fold cross validation was used as part of finding a solution that minimizes the mean square error (MSE). In general, 10-fold cross validation is recommended because of its relatively low bias and variance (Han, 2012). For comparison purposes, the results from ordinary least-squares results are also provided. The following sub-sections review the data used for this study and explain LASSO regression.

4. Study area

The study area for the bicycle demand model is the City of Austin, Texas. There are 26 permanent counters and 41 short duration counters currently operating within the city. Only three locations were chosen because they were the only ones that collected counts during the COVID-19 pandemic, and this study aims to create a predictive bicycle demand model from COVID-19 data and weather. Fig. 1 shows the following counter locations and names used to build the predictive bicycle demand model:

- Counter 1: Ann and Roy Butler Trail at Mopac Crenshaw Bridge
- Counter 2: Duval Street North of E. 32nd Street
- Counter 3: Manor Road at Alamo Street

Referring to Fig. 1, the areas adjacent to the Ann and Roy Butler Trail Counter (Counter 1) are Lady Bird Lake, the Warehouse and Art District, Downtown and the Tarrytown residential area. The counter is located near Lady Bird Lake connects to the Colorado River and is primarily used for recreational activities, like paddle boarding and kayaking, and helps the city with flood control. The Warehouse and Art District is home to office buildings, refurbished warehouses, bars, and trendy restaurants. Like the Warehouse District, Downtown Austin is home to office spaces, high-rise residential developments, shops, museums, restaurants, bars, and entertainment venues. One of the more iconic attractions within the Downtown area is the Texas Capitol. Finally, to the west of the counter location is Tarrytown. Established in 1915, Tarrytown is known as one of the most prestigious and expensive residential neighborhoods in Austin. It is also home to the Laguna Gloria Contemporary Art Museum.

Counter 2, which is located on Duval Street North of E. 32nd Street resides in the North Campus residential area. This area is north of the University of Texas at Austin and apart from homes and small apartment
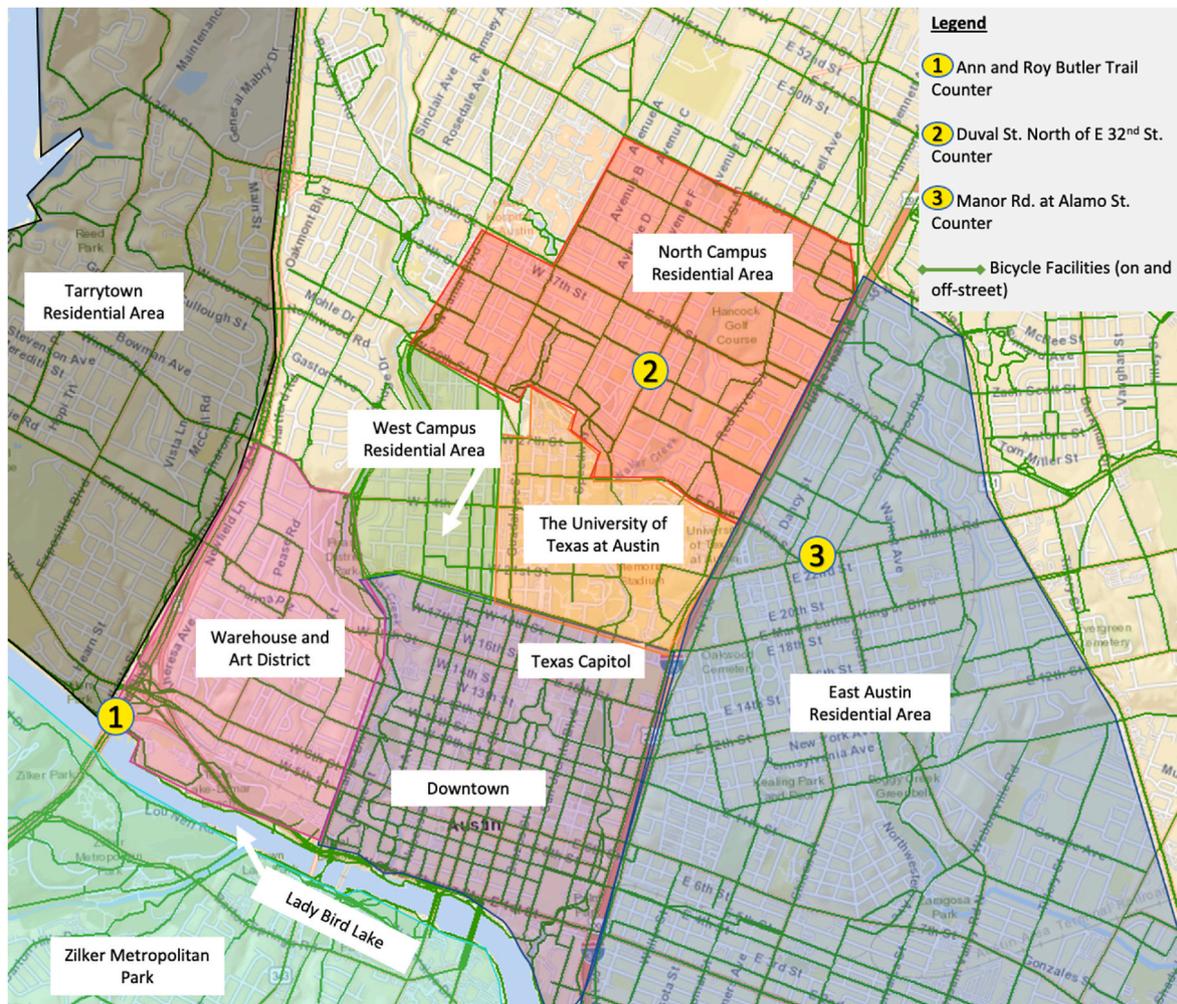
Fig. 1. Counter locations and major areas in Central Austin, TX

complexes, it houses a few restaurants and small bars. North Campus is considered a very popular choice for graduate student living because it is easy to get to the UT campus via bike or public transportation.

Finally, to the east of the two other counter locations is Counter 3 in the North Campus residential area. Located at Manor Road and Alamo Street, this counter is within an up-and-coming neighborhood that experienced a massive boom in terms of the housing market during the pandemic. The East Austin area has undergone gentrification, so there is a combination of older homes and new developments in the area. Not only does this area contain homes and apartments, but also relatively new bars, restaurants, breweries, event spaces, shops, and other local businesses.

4.1. Data summary

The City of Austin uses Eco-Visio counters to capture long-term pedestrian and bicyclist counts. The Eco-Visio counting system is a permanent counting system called Urban Multi, which monitors and differentiates between pedestrians and bicyclists. This system operates by combining PYRO (passive infrared sensors), ZELT inductive loops, and a smart connected subsystem in order to seamlessly transfer the collected count data to the Eco-Counter Server (Eco-Counter). The count data is collected at a 15 min or 1-h time interval and is then displayed onto the Eco-Visio online Eco-Counter Server.

The analysis time period covers the period between the first reported COVID-19 case in Travis County, March 15, 2020, through the last date where count data is available for all three trails, January 31, 2021.

Table 1 lists all the variables considered by the LASSO model to create a prediction-focused bicycle demand model for the three chosen bicycle facilities in Austin, TX. Table 2 is an attribute summary of all the data collected for this effort that includes separate sections for COVID-19 data and weather data.

The table summary shows the min, mean, max, and standard deviation for the numerical variables considered. The overall daily average number of bikes observed is about 270, with the lowest number of bikes observed being 10 and the highest 1074.

Clearly, there's strong correlation between many of the potential predictor variables, which does not present an issue for pure prediction (SpringerLink, 1997, pp. 176–180). This strong correlation would present serious interpretation issues if the goal were to build an *inference* model (SpringerLink, 1997, pp. 176–180). Fig. 2 shows the correlation between COVID and weather variables for informational purposes.

The strongest correlations are those with coefficients of ±0.50 and ±1, and medium correlations are those with values that lie between ±0.30 and ±0.49. Some of the strongest correlations occur within the following combinations of variables:

- all wind-related variables (WSF5, WSF2, WDF5, WDF2, AWND),
- min and max temperature for the day (TMIN, TMAX),
- COVID-related case, hospitalization, fatality, and Texas Executive Order variables, and
- Vaccination variables (Texas_Doses, Texas_OneDose, Texas_FullVac).

If the goal of the model is purely prediction, it is not necessarily

Table 1
Initial test variables.

Variable Title	Description
Daily Confirmed COVID-19 Cases in Travis County, TX	The number of people who have been confirmed to have COVID-19 in Travis County per day.
Daily COVID-19 Fatalities in Travis County, TX	The number of people who have died from COVID-19 in Travis County per day.
Daily Confirmed COVID-19 Cases in TX	The number of people who have been confirmed to have COVID-19 in the State of Texas per day.
Daily COVID-19 Fatalities in TX	The number of people who have died from COVID-19 in Texas per day.
Daily No. of COVID-19 Vaccine Doses Administer in TX	This is the daily aggregated number of people vaccinated with one dose and the number of people fully vaccinated in the state of Texas. The data comes from vaccination records submitted by healthcare providers as of 11:59 the previous night.
Daily No. of People Vaccinated with at Least One Dose of the COVID-19 Vaccine in TX	The number of people who have received at least one dose of COVID-19 vaccine per day.
Daily No. of People who are Fully Vaccinated with the COVID-19 Vaccine in TX	The number of people who have completed the full series of the COVID-19 vaccine (2 doses of Moderna or Pfizer or 1 dose of Janssen vaccine), as outlined by the CDC's Advisory Committee on Immunization Practices (ACIP).
Daily No. of COVID-19 TX Texas Executive Orders	The total number of COVID-19 related executive orders enacted by the Governor per day.
Running Total No. COVID-19 TX Disaster Declarations	Each time a disaster is declared the running total increases by 1.0 meaning the state has yet to declare an emergency, 1 is the first COVID-19 state of emergency declaration by Governor Greg Abbot. Each day after the first state of emergency declaration is labeled as 1 as well until the 2nd state of emergency declaration. Every day after the 2nd state of emergency declaration is labeled as 2. This goes on until the end of the analysis time period where the 12th emergency declaration is reached.
Daily Molecular Positivity Rate (Previous 7-days) in TX	The sum of the total number of positive COVID-19 test results in the previous 7-days divided by the total number of COVID-19 test results in the previous 7-days multiplied by 100.
Average Wind Speed	The average wind speed per day in Camp Mabry, Austin, Texas.
Precipitation	Amount of precipitation per day in Camp Mabry, Austin, Texas.
Maximum Temperature	Maximum temperature per day in Camp Mabry, Austin, Texas (℉).
Minimum Temperature	Minimum temperature per day in Camp Mabry, Austin, Texas (℉).
WDF2	Direction of fastest 2-min wind per day in Camp Mabry, Austin, Texas.
WDF5	Direction of fastest 5-min wind per day in Camp Mabry, Austin, Texas.
WSF2	Fastest 2-min speed per day in Camp Mabry, Austin, Texas.
WSF5	Fastest 5-min speed per day in Camp Mabry, Austin, Texas.
Daily No. of COVID-19 Hospitalizations in TX	Daily number of people admitted into a hospital due to COVID-19.
Active COVID-19 Cases in Travis County, TX	Daily number active COVID-19 cases within Travis County.
Active COVID-19 Cases in TX	Daily number active COVID-19 cases Texas.
Total COVID-19 Fatalities in Travis County, TX	The daily aggregate number of COVID-19 related fatalities in Travis County.

Table 1 (*continued*)

Variable Title	Description
Total COVID-19 Fatalities in TX	The daily aggregate number of COVID-19 related fatalities throughout time in Texas.
Daily No. of COVID-19 Tests in Travis County	Daily number of people who got tested for COVID-19 in Travis County.
Daily No. of COVID-19 Tests in TX	Daily number of people who got tested for COVID-19 in Texas.
Ann and Roy Butler Daily Bicycle Counts	Categorical variable that = 1 if the counts come from Ann Roy and Butler and 0 otherwise. The Duval St. location is the reference location.
Manor Rd. Daily Bicycle Counts	Categorical variable that = 1 if the counts come from Manor Rd. and 0 otherwise. The Duval St. location is the reference location.
Day	Each day of the analysis period, beginning with March 15, 2020.

Table 2
Attribute summary.

Variable	Min	Mean	Max	St. Dev
Total Daily Bikes	10	269.50	1074	188.29
Hospitalizations	0	5764.75	14218	4012.54
Daily Molecular Positivity	5.74	11.75	20.80	4.24
Texas Daily Active Cases	0	116878.24	376507	102728.83
Texas Daily Confirmed Cases	0	6353.63	28020	5609.23
Texas Daily COVID Fatalities	0	127.03	387	99.14
Texas Disaster Declarations	0	4.98	10	3.13
Texas Daily Doses Administered	0	8806.88	148652	26522.54
Texas Executive Orders	0	24.27	31	8.78
Texas Daily Full Vaccination	0	1900.65	61340	7685.50
Texas Daily One Dose	0	6906.45	93254	20016.59
Texas Daily No. of COVID Tests	0	6314776.00	20247888	5774309.47
Texas Total Fatalities	1	13618.30	41195	11672.64
Travis Daily Active Cases	0	1846.82	6682	1603.41
Travis Confirmed Daily COVID Cases	0	210.39	1461	225.49
Travis Daily COVID Fatalities	0	2.54	12	2.65
Travis Daily No. of COVID Tests	0	255341.86	876487	248641.66
Travis Total Fatalities	0	319.88	825	226.00
Average Wind Speed	0.45	4.49	10.07	1.84
Precipitation	0	0.09	2.37	0.33
TMAX	43	84.64	108	14.00
TMIN	30	61.75	83	13.86
WDF2	10	156.89	360	102.47
WDF5	10	163.14	360	101.87
WSF2	6	12.79	23.90	3.22
WSF5	10.1	23.42	40.90	6.27

straightforward for an analyst to know which variables to keep. Traditional inference-based thinking focuses on removing multicollinearity and keeping significant variables, which does not actually address the goal of prediction. The strong correlation seen in the predictive variables (see Fig. 2) is not a concern for this comparison because multicollinearity alone does not affect the prediction ability of a model (Farrar & Glauber, 1967).

4.2. LASSO regression

The acronym *LASSO* stands for Least Absolute Shrinkage and Selection Operator, and is a variable selection and regularization technique in machine learning (Tibshirani, 1996, 2011) that avoids the problem of overfitting. Overfitting presents a serious issue for prediction-focused
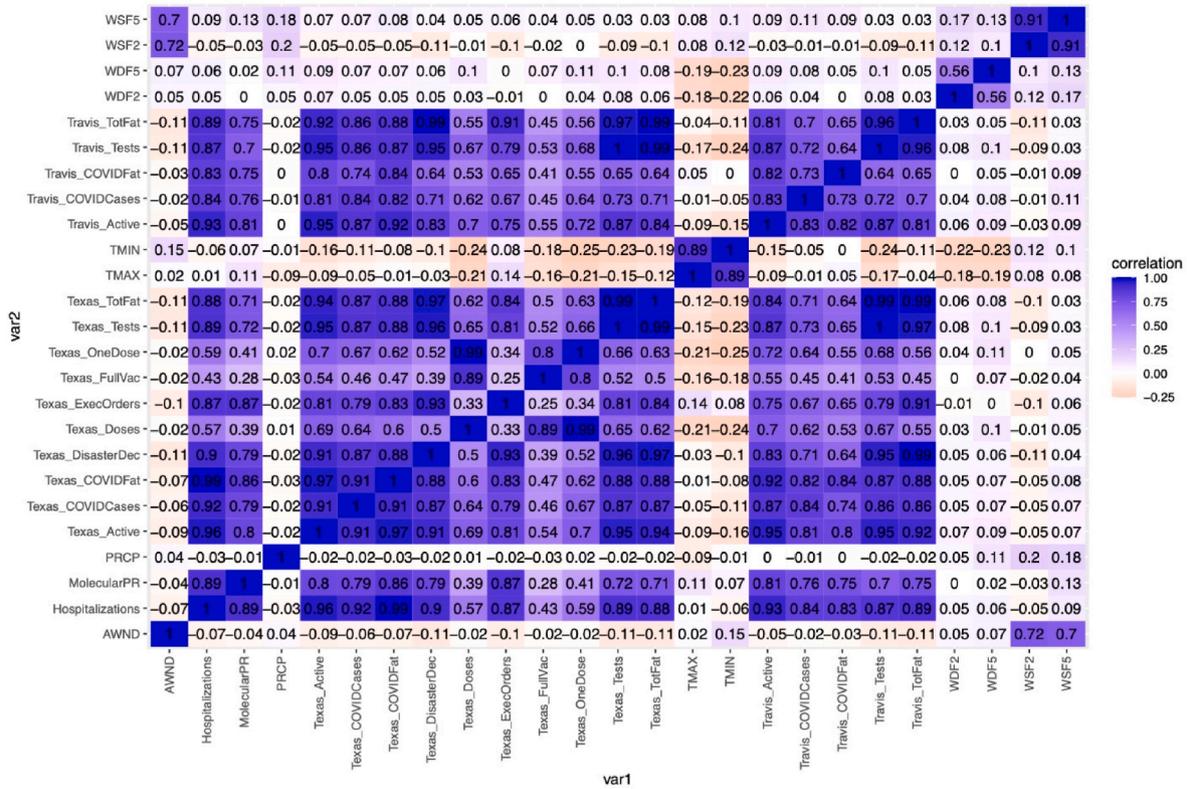
Fig. 2. Correlation matrix for potential predictor variables.

models because the resulting model will fit the data too closely, meaning it will perform poorly when asked to perform on new data. On the other hand, regularization penalizes models as they become more complex (i. e., including more variables) and only keeps the variables that are most predictive. In many ways, LASSO regularization beats other techniques for improving OLS estimates, like subset selection and ridge regression, because of its prediction accuracy and interpretability (Tibshirani, 1996). Keeping only the most predictive variables also generally takes care of multicollinearity issues.

LASSO's objective function is the same as in OLS regression, with the exception the LASSO solution also minimizes the sum of the absolute value of the model coefficients. The minimization of the absolute value of the predictor variable coefficient estimates helps penalize the model as it becomes more complex. When the shrinkage penalty λ gets sufficiently large, it will shrink coefficients down to zero, therefore removing variables from the model. The objective function for LASSO is

$$\min_{\beta}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - X_i^T\beta\right)^2\right\} \quad subject\ to\ \sum_{j=1}^{p}\left|\beta_j\right| \leq \lambda, \tag{1}$$

where the coefficients represent:

- N = the number of observations,
- y_i = the outcome measures (bicycle counts),
- β = the predictor variable coefficients (COVID-19 and weather data points),
- X = the predictor variable matrix,
- p = set of predictor variables, and
- λ = the amount of penalty (shrinkage penalty) and chosen through cross validation.

LASSO is also helpful if p is large and when p is much larger than N. Inference models easily fail in situations where $p > N$ because there are infinitely many solutions to the least square problem (Tibshirani, 2017).

One of the great benefits of the LASSO technique is that it only keeps the variables that are really driving prediction and shrinks the rest down to zero (Tibshirani, 1996). Meaning, it is especially helpful when there are many potential predictors. The more non-zero coefficients kept by the model, the more variability explained, but at the same time, the larger the penalty. When the λ penalty goes to zero, the model retains all predictor variables in the exact way regular OLS regression would. At the other extreme, when λ gets very large (λ_{max}), then the number of variables picked by LASSO goes to zero. Since neither extreme of λ prioritizes predictive ability, cross validation will run through a sequence of λ's to find an optimal λ that minimizes the mean of squared differences between the observed value and the predicted value (MSE). An optimal λ will improve prediction accuracy for a model and keep the number of variables small, therefore helping overall interpretability.

4.3. K-fold cross validation

An optimal λ is picked through k-fold (usually 5 or 10-fold) cross validation (CV). For the LASSO approach, an array of potential λ's undergoes CV to find an optimal model. The array of potential λ values are generated by the glmnet package in R (Friedman et al., 2010). The array is computed based on nlambda (number of λ values) and lambda.min. ratio (ratio between the maximum and minimum λ value) Here, we used the values commonly used in practice – nlambda = 100 and lambda.min. ratio = 10^{-6} (Friedman et al., 2010).

For each λ, k-fold cross validation (CV) splits a dataset up into two sets: one for building the model and the other reserved only for testing the model and calculates the corresponding MSE. In this case, the k-fold CV breaks the data up into 10 folds (a fold is the same as a subset) and repeats this process of training the model with some data and saving the other part for testing. The algorithm for k-fold CV is explained below. For every $i = 1, \ldots, K$:

1. Train the model on every fold *except* the i th fold,

2. Test the model by calculating the test error (in this case, MSE) on the *i* th fold, and

3. Compute the mean of all the MSE or other test errors.

K-fold CV is the most common way of understanding how well the proposed model works on data it has never seen before (the data in the *i* th fold), which is why it is sometimes called out-of-sample testing. CV is repeated for each potential value of λ. In the end, the best predictive model is the one that contains the combination of variables that provides the lowest MSE during CV.

4.3.1. Model set-up

Friedman et al. developed a fast algorithm for fitting generalized linear models with various penalties, and this study used their glmnet R package. The LASSO regression approach was used to develop a model with only the variables that are most predictive of bicycle demand. The potential variables are listed in Table 1. The Duval St. location serves as the reference location, so whenever a count comes from Duval St., the Ann Roy and Butler and Manor Rd., those variables are set to zero. COVID-19 data was taken from the Texas Department of State Health Services and includes both county-level and state-level data. Weather data was taken from the Camp Mabry weather station in Austin, Texas (National Centers for Environmental Information, 2021; Texas Department of State Health Services, 2021).

5. Data analysis and results

This section gives a qualitative and quantitative overview of what happened to bicycle activity in Travis County during the pandemic. The first subsection will provide a qualitative overview of the general trends

seen on the three trails of interest pre and post pandemic. The second subsection provides a count-based bicycle demand forecasting model developed from the LASSO regression technique which includes COVID-19 data and weather data.

5.1. Bicycle activity pre- and post-pandemic

The bicycle count data that was used for the model included daily counts from the years of 2019, 2020, and into 2021. For Ann and Roy Butler trail, the last count date is January 31st, 2021, while Duval and Manor's last count date is March 31st, 2021. Figs. 3, 6 and 7 compare the difference in bicycle volumes between 2019 and 2020 for all three routes. The black vertical line within the three figures represents the first confirmed COVID-19 case in Travis County (March 15th, 2020), while the green vertical line represents the first day of classes in the Fall semester at the University of Texas at Austin (August 26th, 2020).

5.1.1. Ann and Roy Butler Trail ridership trends

The counter on Ann and Roy Butler Trail is the only counter that is along an off-street trail. It is also the only counter that has more counted daily riders in 2020 (also slightly into 2021), than in 2019 (see Fig. 3). The blue boxes represent daily number of 2020 riders, the orange circles 2019 riders, and the yellow triangles are the 2021 riders.

Ann and Roy Butler Trail reached its peak of daily volume on April 27, 2020, which is nearly a month after the first confirmed case of COVID-19 in Travis County. During April, Governor Greg Abbot issued six COVID-19 related executive orders, where most of these orders involve reopening plans for businesses and services, and travel restrictions for Texans. Executive Order GA-18, which was coincidentally signed on April 27, 2020, is Phase One to reopen Texas. This is one of
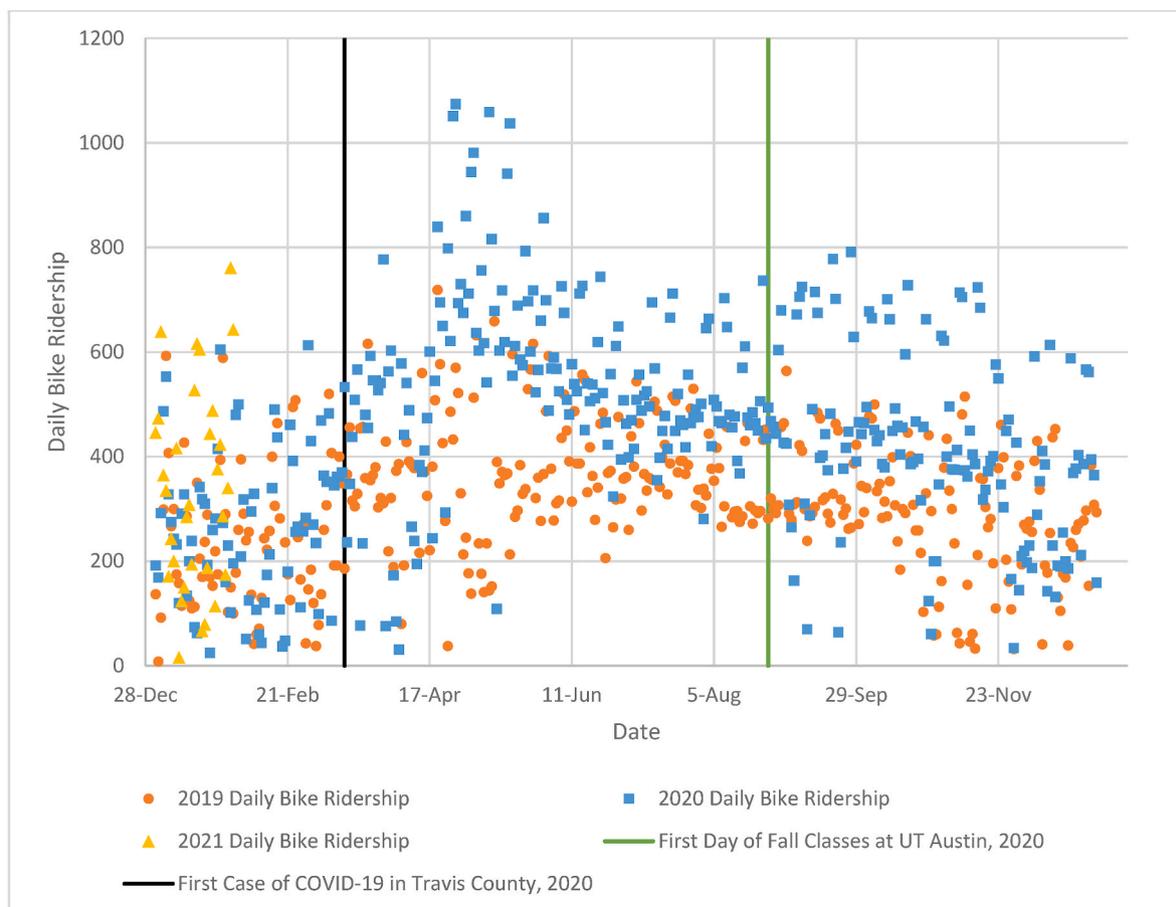

Fig. 3. Ann and Roy Butler trail at Mopac Crenshaw Bridge daily ridership counts for 2019, 2020, and 2021.

Gov. Greg Abbot's first steps toward reopening businesses that were not already labeled as essential businesses and gave business the capability to operate at 25% capacity or have the option to provide pickup or delivery services (Executive Order, 2020). When the pandemic began in Texas in early March, the rise of cases and fatalities began to drastically accumulate into the month of April after the first documented COVID-19 related death on March 15, 2020. Fig. 4 shows the fitted trendline of confirmed COVID-19 cases throughout the state of Texas from March 6, 2020 to April 30, 2020. This graph reveals the growth and spread of the virus at the beginning of the pandemic, nearly 21 confirmed cases per day in Texas. Fig. 5 shows that 1 person per day died of COVID-19 between March 6, 2020 and April 30, 2020.

Around April 2020, the University of Texas decided to provide students with classes that were completely online and limited on-campus operations to mitigate the spread of the virus in the Spring semester of 2020 (The University of Texas at Austin). Employees were also told to work remotely until further notice. Once the Fall semester began the University decided that more than 75% of the actual seats in classes would be online and only 5% would be fully in person. The remaining 20% were offered through hybrid courses that mix online and in-person elements (The University of Texas at Austin, 2020). The University also suggested that employees who can work remotely should continue doing so if possible. The slight increase in ridership at the beginning of the Fall semester, reflect people even outside of UT who were beginning to work and go to school in a hybrid way.

Finally, the daily ridership of Ann Roy and Butler Trail into 2021 has an upward trend. During this time the first doses of COVID-19 vaccines are being distributed. The first round of vaccines, according to the Texas Department of Health data source was documented on December 14, 2021. Overall, these trends allude to the possibility of these riders being mostly composed of recreational bikers.

5.1.2. Duval St. North of E. 32nd st.

The counting station on Duval St. North of E. 32nd St. did not have more daily riders in 2020 than in 2019. As mentioned before, this counter is located north of the UT Austin campus and is a major hub for student housing. The 2020 ridership trend experiences a rise at the beginning of the year yet when the first case of COVID-19 appeared in Travis County, the daily number of riders plummeted. As mentioned before, UT Austin began online classes near this time. Like Ann and Roy Butler Trail, Duval experienced a rise in ridership once the Fall semester began at UT Austin and began hybrid classes (25% of all students were taking hybrid classes).

As 2021 began, the daily ridership was rising once again but still mimicking the trends of 2020. Also, UT Austin was still conducting the hybrid class format into the Spring Semester of 2021. This data most likely represents UT Austin students who ride a bike to and from campus.

5.1.3. Manor Rd. At Alamo St.

The last counter, Manor Rd. and Alamo St., is also close to UT Austin's campus. Located east of the campus, and near Gonzalo High School, Campbell Elementary School, and several apartments and shopping units, this counter location is another prime location for students and young families. Like Duval, this counter did *not* see a greater number of daily riders in 2020 than in 2019. These trends are extremely like Duval where at the beginning of 2020, ridership was growing, yet when COVID-19 reached Travis County, ridership plummeted. Yet bike ridership began its upward climb once more as the 2020 Fall semester began.

The beginning of 2021, ridership was on the rise, yet like Duval, was mimicking the trends of 2020. The riders along this trail are most likely young families, or UT students.

5.2. LASSO modeling results and discussion

The LASSO regression method was preformed to select the COVID-19 and weather variables that minimize the MSE and increase model prediction accuracy for bicycle demand at the three locations. Fig. 8 shows a CV curve for choosing an optimal λ, the first vertical dotted line represents the λ value that gives the minimum mean CV error, while the second dotted line is that value of λ within one standard error of the minimum. The numbers shown at the very top of the figure show the number of non-zero variables kept by the LASSO model for a given λ. The optimal model comes from λ within one standard error of the minimum to obtain the six model coefficients for the bicycle demand model, however, one could also pick the λ value represented by the first line. The *one-standard error rule* for picking λ is common practice to ensure that the model that is chosen is the most parsimonious. This way, the errors are no more than one standard error above the error of the best model ((Friedman et al., 2010), (Krstajic et al., 2014)). Since LASSO is not used for inference, there is no hypothesis test used.

We compare the prediction ability of the full model to the LASSO model using k-fold CV on both models and show that the full model suffers from overfitting. The 10-fold CV process calculates prediction MSE for both models to measure their predictive ability (See Table 3).
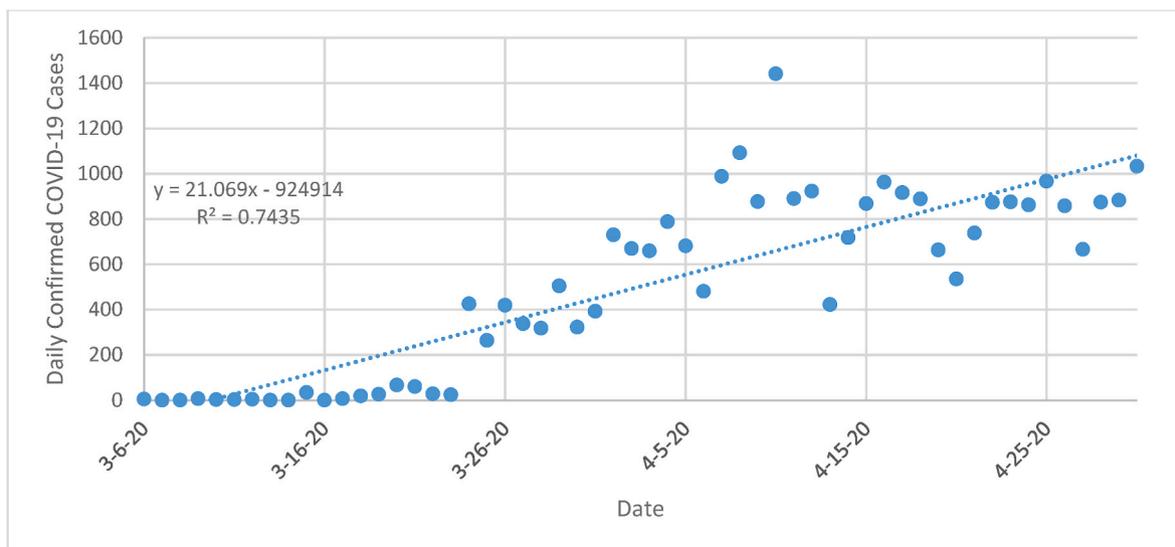

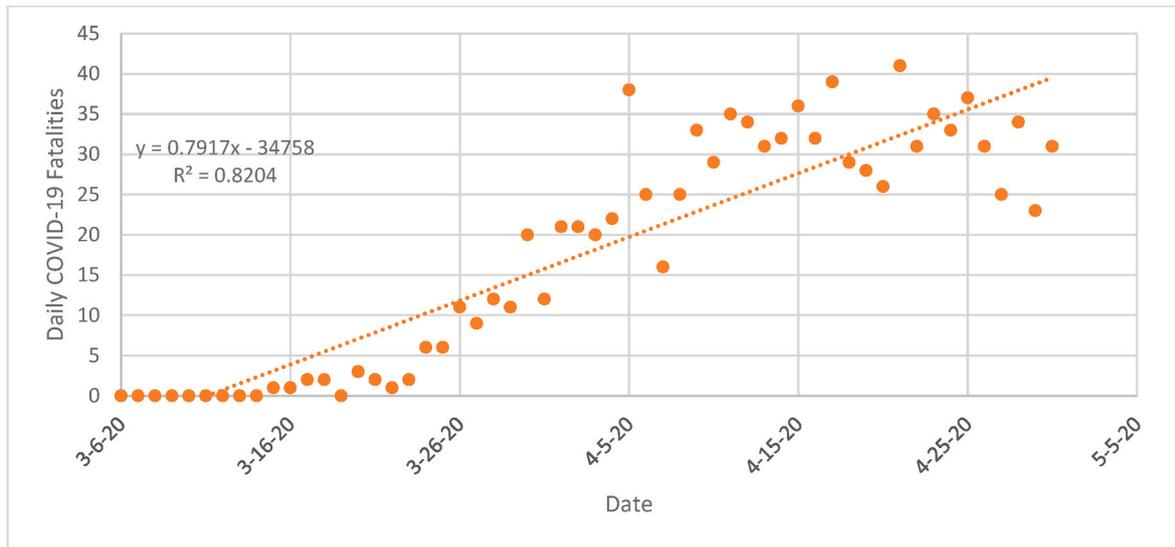**Fig. 4.** Daily confirmed COVID-19 cases throughout Texas (3/6/2020–4/30/2020).

Fig. 5. Daily confirmed COVID-19 fatalities throughout Texas (3/6/2020–4/30/2020).
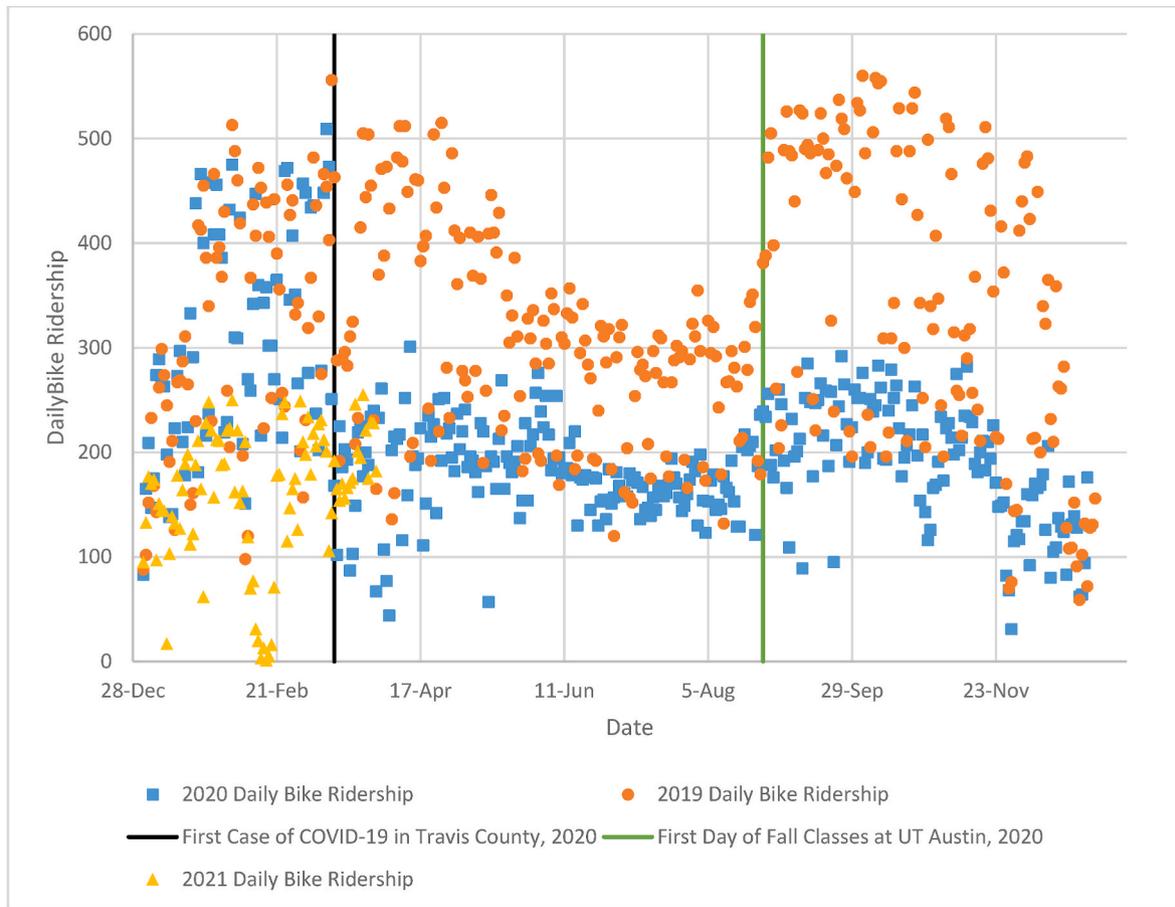

Fig. 6. Duval St. North of E. 32nd St. Daily ridership counts for 2019, 2020, and 2021.

The regular MSE (from in-sample testing) of the full model is 9837.209, but upon testing the full model's ability to predict through CV the prediction MSE (from out-of-sample testing) is 12007.68. The larger prediction MSE indicates that the model does not predict very well. As we would expect, we have evidence of the full model overfitting because the MSE is much lower than its prediction MSE. That is, it does well when fitting its own data points, but it fails to perform as well when shown

data not used during the model building process. In the end, the final LASSO model's prediction MSE value is smaller (= 11660.42) than the full model, and therefore better at predicting future bicycle ridership. In this study, we did not include any variables that capture UT's changes in attendance policy. Including such a variable could impact LASSO's final choice of most predictive variables.

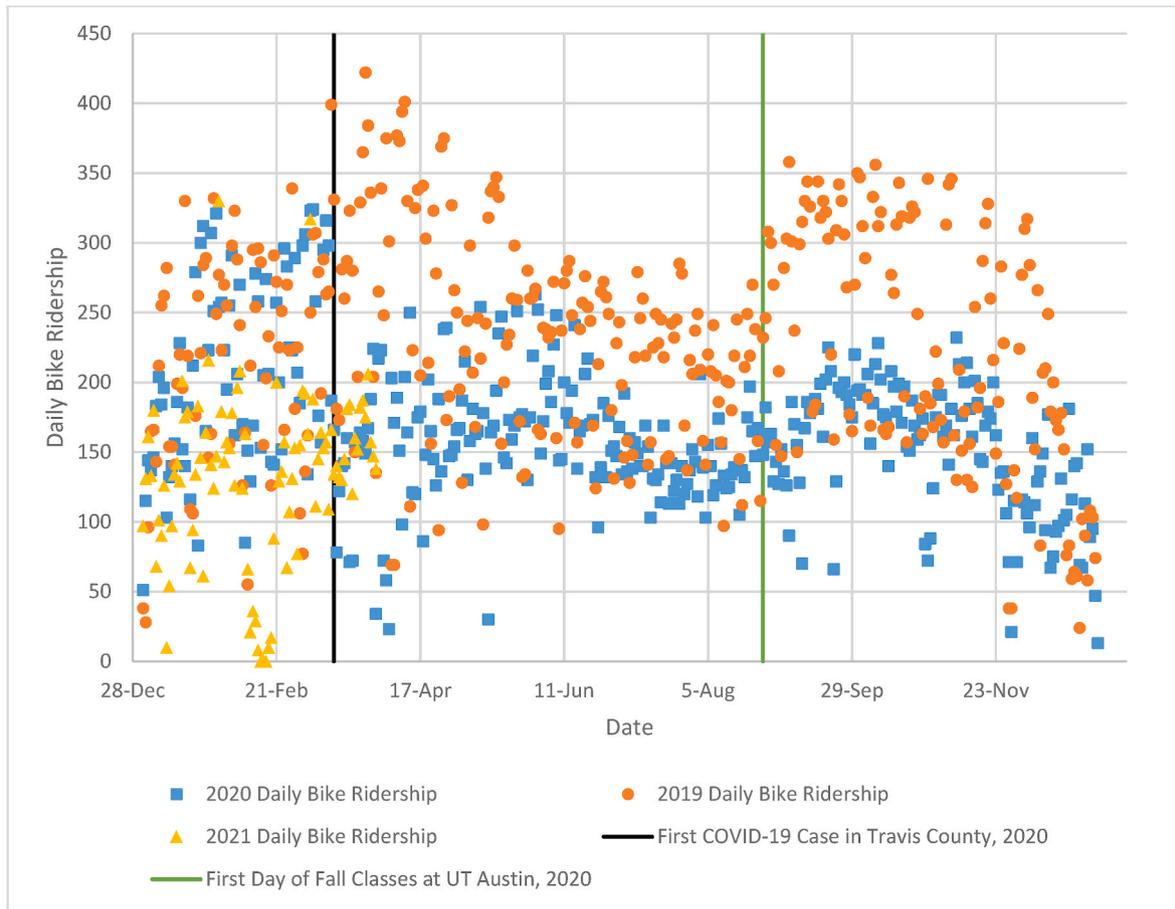The LASSO model picked six variables as the most important vari-

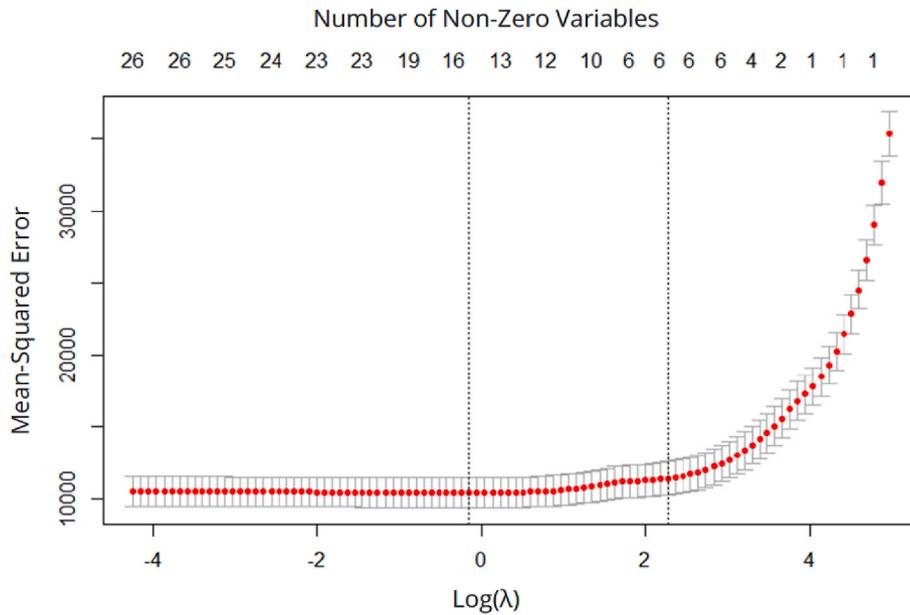**Fig. 7.** Manor Rd. at Alamo St. Daily ridership counts for 2019, 2020, and 2021.

Fig. 8. Plotted λ cross validation curve.

ables for predicting COVID-19. Table 4 shows the comparison between LASSO regression and traditional OLS regression. The LASSO approach shows that the number of COVID-19 fatalities in Texas ($\beta = -15.06$), Molecular Positivity Rate ($\beta = -7.458$), precipitation ($\beta = -17.53$), max temperature ($\beta = 27.73$), and counter location ($\beta_{AnnRoy} = 131.6$ and

$\beta_{Manor} = -6.846$, Duval St. is the reference location) are the most important predictors of bicycle demand. The only variables LASSO and OLS agree on are precipitation, max temperature, and location. While a modeler could pick out the most significant variables or those that increase R^2, those would not necessarily be the most *predictive* variables

Table 3

OLS and LASSO K-fold cross validation results.

Model	Prediction MSE
OLS	12007.68
LASSO	11660.42

Table 4

LASSO and OLS regression model output.

Variables	LASSO Estimates	OLS Estimates	OLS 95% CI	p-value
Daily Confirmed COVID-19 Cases in Travis County, TX	–	0.0039	(-0.0460, 0.0539)	0.9
Daily COVID-19 Fatalities in Travis County, TX	–	−3.221	(-7.834, 1.392)	0.2
Daily Confirmed COVID-19 Cases in TX	–	0.0005	(-0.0023, 0.0032)	0.7
Daily Confirmed COVID-19 Fatalities in TX	−15.06	−0.3076	(-0.7861, 0.1709)	0.2
Daily No. of COVID-19 Vaccine Doses Administered in TX	–	−1.301	(-8.352, 5.749)	0.7
Daily No. of People Vaccinated w/at Least One Dose of COVID-19 Vaccine in TX	–	1.301	(-5.750, 8.351)	0.7
Daily No. of People Fully Vaccinated with the COVID-19 Vaccine in TX	–	1.302	(-5.749, 8.352)	0.7
Daily No. of COVID-19 Texas Executive Orders	–	1.686	(-7.746, 11.12)	0.7
Running Total No. of COVID-19 Texas Disaster Declarations	–	2.845	(-21.89, 27.58)	0.8
Daily Molecular Positivity Rate (Previous 7-days) in TX	−7.458	0.7083	(-4.036, 5.453)	0.8
Average Wind Speed	–	1.673	(-3.881, 7.227)	0.6
Precipitation	−17.53	−77.66	(-99.40, −55.91)	**<0.001**
Maximum Temperature	27.73	5.365	(4.217, 6.513)	**<0.001**
Minimum Temperature	–	−2.669	(-3.888, −1.450)	**<0.001**
WDF2	–	−0.0278	(-0.1065, 0.0508)	0.5
WDF5	–	0.0101	(-0.0667, 0.0868)	0.8
WSF2	–	0.5691	(-5.211, 6.349)	0.8
WSF5	–	−1.275	(-4.247, 1.697)	0.4
Daily No. of COVID-19 Hospitalizations in Texas	–	−0.0164	(-0.0338, 0.0010)	0.064
Active COVID-19 Cases in Travis County, TX	–	0.0590	(0.0391, 0.0789)	**<0.001**
Active COVID-19 Cases in TX	–	−0.0004	(-0.0010, 0.0003)	0.3
Total COVID-19 Fatalities in Travis County, TX	–	0.2497	(-0.6426, 1.142)	0.6
Total COVID-19 Fatalities in TX	–	0.0018	(-0.0301, 0.0336)	>0.9
Daily No. of COVID-19 Tests in Travis County	–	−0.0012	(-0.0024, 0.0001)	0.065
Daily No. of COVID-19 Tests in TX	–	0.0001	(0.0000, 0.0001)	0.063
Ann and Roy Butler Daily Bicycle Counts	131.6	292.7	(277.2, 308.3)	**<0.001**
Manor Rd. Daily Bicycle Counts	−6.846	−28.29	(-43.87, −12.72)	**<0.001**
Day	–	−1.181	(-3.793, 1.431)	0.4

nor fit the primary purpose of the study. Again, prediction and inference are different, sometimes opposing forces, and they each have their place. However, an accurate prediction of bicycle demand is critical to building a useful benefit-cost analysis of bicycle facility alternatives.

5.2.1. Predictive ability

If the primary goal is to increase predictive power, modelers should focus on removing variables that inflate the prediction MSE. The traditional approach to building a count-based bicycle demand model claims to predict riders, but uses inference-based techniques to build the model. Inference tells us about the relationship between the variables and the number of riders, but that does not mean the model will perform well when shown new data. LASSO differs from the traditional approach in that it evaluates out-of-sample predictive ability with k-fold cross validation and keeps only the most predictive variables.

5.2.2. Variable selection

Because of the λ penalty, LASSO only selects the variables that are strong predictors and displays no predetermined favoritism over any variables. This effectively removes the *pet variable dilemma* where a researcher focuses on a favored explanatory variable and *p-hacks* their way into getting a statistically significant result. P-hacking occurs sometimes even unconsciously and does not prioritize predictive ability (Beecham et al., 2020).

5.2.3. Interpretability

Since LASSO tends to remove redundant variables, it increases the interpretability of the model. It favors subsets of the predictor variables that have less collinearity. Since prediction is not harmed by multi-collinearity, other prediction-focused methods do not exclude highly correlated variables the way LASSO does.

6. Limitations

Comparing both the OLS and LASSO estimates, both models clearly show that location matters but neither model was given interaction terms to represent time-varying local conditions. In general, UT's attendance policies were driven by Travis County COVID-19 metrics. Therefore, since all counters are located in the same county, county-level COVID effects could differ among the count stations. LASSO can only pick from the variables given, and it was not given any interaction terms to explore this time-varying station-specific condition.

The LASSO method is not specifically designed to find interactions, but we can give it interaction terms to pick from to create a new predictive model. From a classical model building standpoint, the issue with giving LASSO interaction terms is that it might drop main effects while keeping the interaction term. This can present a problem if the goal is to maintain interpretability, since interpreting interaction terms without main effects is problematic. In other words, LASSO just selects what is informative, and it might find main effects to be uninformative, but interactions to be informative.

If we add the interaction terms between each location and Travis County COVID cases, for example, LASSO does not retain the main feature of Travis County COVID cases and only picks up the interaction term Travis_COVIDCases*trail_AnnRoy in addition to the terms it picked in Table 4. LASSO does not pick up the Travis_COVIDCases main effect because LASSO is neutral to the pattern in which features exist (Jain & Xu, 2021). We get a slightly better MSE (=11362.41) when giving it these two additional interaction terms to pick from, but at the expense of interpretability. In a case where you want an interpretive model, the analyst might prefer a different method that keeps the hierarchy of the model. However, testing different machine learning techniques that maintain model hierarchy when exploring interaction terms is outside the scope of this paper.

7. Conclusions

Moving forward, demand forecasting models should use techniques that focus on prediction accuracy, and avoid the common error of conflating explanatory methods as predictive methods. Explanation and prediction are not the same, and in a statistical sense, can act as opposing forces. Direct demand forecasting models should avoid standard OLS and logit models and use penalized regression techniques like LASSO. LASSO is a very interpretable, readily available technique that helps models predict well out of sample, and only selects the most important predictors. LASSO helps modelers focus on the most powerful predictors and prevents the issue of adding in marginally significant variables that help the model from an explanatory perspective but could hurt its actual forecasting ability. To showcase the benefits of the LASSO technique, this paper develops a LASSO model for predicting bicycle demand at three bicycle facilities in Austin, TX based on COVID-19 and weather data. Only one of the three bicycle facilities (Ann and Roy Butler Trail) experienced an overall daily increase in bicycle ridership in 2020 (slightly into 2021) compared to 2019. This alludes to the boom in recreational bike riding during the pandemic, due to CDC stating the outdoor activity as low-risk for contracting and spreading COVID-19. All three bicycle facilities experienced an initial drastic decrease in ridership once the first case of COVID-19 was confirmed in Travis County (March 15th, 2020), and a slight increase in daily ridership as UT Austin began the Fall 2020 semester. The LASSO model selects six variables as the most powerful predictors out of total of 28 potential variables, which successfully minimizes the MSE. On the other hand, the traditional OLS method selects up to 10 potentially significant variables, and out of those ten only three agree with LASSO's picks. The COVID-19 variables that most significantly affect future bicycle demand are state-level fatalities, molecular positivity rate, precipitation, max temperature, and facility location.

CRediT authorship contribution statement

Carolina Baumanis: Methodology, Software, Validation, Writing – review & editing, Project administration. **Jennifer Hall:** Conceptualization, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization, Project administration, Dr. **Randy Machemehl:** Writing – review & editing, Supervision.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Beecham, R., Williams, N., & Comber, A. (2020). Regionally-structured explanations behind area-level populism: An update to recent ecological analyses. *PLoS One, 15* (3). https://doi.org/10.1371/JOURNAL.PONE.0229974

Brand, R., Timme, S., & Nosrat, S. (2020). When Pandemic Hits: Exercise Frequency and Subjective Well-Being During COVID-19 Pandemic. *Frontiers in Psychology*, vol. 11. https://doi.org/10.3389/FPSYG.2020.570567

Bucsky, P. (2020). Modal share changes due to COVID-19: The case of budapest. *Transportation Research Interdisciplinary Perspectives* (Vol. 8), Article 100141. https://doi.org/10.1016/J.TRIP.2020.100141

Centers for Disease Control and Prevention. Requirement for face masks on public transportation conveyances and at transportation hubs. https://www.cdc.gov/coronavirus/2019-ncov/travelers/face-masks-public-transportation.html. (Accessed 26 July 2021).

The University of Texas at Austin. Protect Texas Together Annoucements. https://protect.utexas.edu/announcements/.

Centers for Disease Control and Prevention. Choosing safer activities. https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/participate-in-activities.html. (Accessed 6 July 2021).

Doubleday, A., Choe, Y., Isaksen, T. B., Miles, S., & Errett, N. A. (2021). How did outdoor biking and walking change during COVID-19?: A case study of three U.S. Cities. *PLoS One, 16*(1), Article e0245514. https://doi.org/10.1371/JOURNAL.PONE.0245514

People For Bikes. (2021). How bicycling changed during A pandemic. https://www.peopleforbikes.org/news/how-bicycling-changed-during-a-pandemic.

Centers for Disease Control and Prevention. How to Be Physically Active While Social Distancing. Physical Activity. https://www.cdc.gov/physicalactivity/how-to-be-physically-active-while-social-distancing.html. (Accessed 22 June 2022).

Executive Order, GA-18.. https://gov.texas.gov/uploads/files/press/EO-GA-18_expanded_reopening_of_services_COVID-19.pdf.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics, 49*(1), 92. https://doi.org/10.2307/1937887

Florida Department of Transportation Systems Planning Office. A recommended approach to delineating traffic analysis Zones in Florida. https://www.fsutmsonline.net/images/uploads/reports/fr1_fdot_taz_white_paper_final.pdf.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1. https://doi.org/10.18637/jss.v033.i01

Fung, P. L., Zaidan, M. A., Timonen, H., Niemi, J. V., Kousa, A., Kuula, J., & Hussein, T. (2021). Evaluation of white-box versus black-box machine learning models in estimating ambient black carbon concentration. *Journal of Aerosol Science, 152*, Article 105694. https://doi.org/10.1016/J.JAEROSCI.2020.105694

Institute for Transportation and Development Policy. (2022). *The year of the bicycle*. https://www.itdp.org/2022/01/19/2022-the-year-of-the-bicycle/.

Jain, R., & Xu, W. (2021). HDSI: High dimensional selection with interactions algorithm on feature selection and testing. *PLoS One, 16*(2). https://doi.org/10.1371/JOURNAL.PONE.0246159

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics, 6*(1), 1–15. https://doi.org/10.1186/1758-2946-6-10

Han, J. (2012). Classification: Basic Concepts. *Data mining concepts and techniques /* (pp. 327–391). Elsevier. https://doi.org/10.1016/B978-0-12-381479-1.00008-3.

Eco-Counter. Urban MULTI. https://www.eco-counter.com/produits/multi-range/urban-multi/. (Accessed 20 June 2022).

Miller, E. J. (2021). *Traffic analysis zone definition: Issues & guidance*. https://tmg.utoronto.ca/files/Reports/Traffic-Zone-Guidance_March-2021_Final.pdf.

Möllers, A., Specht, S., & Wessel, J. (2022). The impact of the Covid-19 pandemic and government intervention on active mobility. *Transportation research. Part A, Policy and practice, 165*, 356–375. https://doi.org/10.1016/j.tra.2022.09.007.

Ministry of Sustainable Infrastructure and Mobility. Sustainability: Up to 500 Euros per Mobility Voucher. Italian Government. https://www.mit.gov.it/comunicazione/news/sostenibilita-fino-500-euro-per-buono-mobilita. (Accessed 26 July 2021).

Sharp, D. & Chan, K. (2020). Pandemic Leads to a bicycle boom, and shortage, around world. The Associated Press.https://apnews.com/article/technology-lifestyle-business-ap-top-news-europe-39b3691b8e1ea3d74d1280a75e451a36.

National Academies of Sciences, Engineering, and Medicine. (2014). Estimating Bicycling and Walking for Planning and Project Development: A Guidebook. Washington, DC: The National Academies Press. https://doi.org/10.17226/22330.

National Centers for Environmental Information. Austin Camp Mabry, TX daily summaries. https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00013958/detail. (Accessed 16 July 2021).

Powell, M.. *Cycling Industry Sales Growth Accelerates in April, Up 75%, Generating an Unprecedented $1 Billion for the Month, According to NPD*. The NPD Group. https://www.npd.com/news/press-releases/2020/cycling-industry-sales-growth-accelerates-in-april-up-75-generating-an-unprecedented-1-billion-for-the-month-according-to-npd.

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. https://doi.org/10.1214/10-STS330

Sorenson, D. (2022). *Four Ways the Cycling Industry Can Set Its Sights on Success*. The NPD Group. https://www.npd.com/news/blog/2022/four-ways-the-cycling-industry-can-set-its-sights-on-success/.

SpringerLink. (1997). The problem of multicollinearity. *Understanding Regression Analysis*. Boston, MA: Springer. https://doi.org/10.1007/978-0-585-25657-3_37.

Texas Department of State Health Services. Texas COVID-19 data. https://dshs.texas.gov/coronavirus/AdditionalData.aspx. (Accessed 16 July 2021).

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, 58*(1).

Tibshirani, R. (2011). Regression shrinkage and selection via the Lasso: A retrospective. *Journal of the Royal Statistical Society: Series B, 73*(3), 273–282. https://doi.org/10.1111/J.1467-9868.2011.00771.X

Tibshirani, R., Wasserman. (2017). Sparsity, L., & the Lasso. and Friends. Carnegie Mellon University. https://www.stat.cmu.edu/~ryantibs/statml/lectures/sparsity.pdf.

The University of Texas at Austin. (2020). UT Austin projects more than 75% of class seats will Be online this fall. UT News.https://news.utexas.edu/2020/08/11/ut-austin-projects-more-than-75-of-class-seats-will-be-online-this-fall/.

United States Census Bureau. (2021). *Consumers Turn To Biking for Safe Fun and Exercise During Pandemic*. https://www.census.gov/library/stories/2021/06/consumers-turn-to-biking-for-safe-fun-exercise-during-pandemic.html.

University of Minnesota. How biking infrastructure has changed during the pandemic. https://twin-cities.umn.edu/news-events/how-biking-infrastructure-has-changed-during-pandemic.

Wilson, K.. Study: The biggest COVID-19 bike booms weren't where you think. Streetsblog USA. https://usa.streetsblog.org/2020/11/25/study-the-biggest-covid-19-bike-booms-werent-where-you-think/.

Wu, S., Harris, T. J., & Mcauley, K. B. (2007). The use of simplified or misspecified models: Linear case. *Canadian Journal of Chemical Engineering, 85*(4), 386–398. https://doi.org/10.1002/CJCE.5450850401

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393