Article

Understanding the Drivers of Mobility during the COVID-19 Pandemic in Florida, USA Using a Machine Learning Approach

Guimin Zhu [1,*], Kathleen Stewart [1], Deb Niemeier [2] and Junchuan Fan [3]

1   Center for Geospatial Information Science, Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA; stewartk@umd.edu
2   Department of Civil & Environmental Engineering, Maryland Transportation Institute, University of Maryland, College Park, MD 20742, USA; niemeier@umd.edu
3   Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA; fanj@ornl.gov
*   Correspondence: gmzhu@umd.edu

Abstract: As of March 2021, the State of Florida, U.S.A. had accounted for approximately 6.67% of total COVID-19 (SARS-CoV-2 coronavirus disease) cases in the U.S. The main objective of this research is to analyze mobility patterns during a three month period in summer 2020, when COVID-19 case numbers were very high for three Florida counties, Miami-Dade, Broward, and Palm Beach counties. To investigate patterns, as well as drivers, related to changes in mobility across the tri-county region, a random forest regression model was built using sociodemographic, travel, and built environment factors, as well as COVID-19 positive case data. Mobility patterns declined in each county when new COVID-19 infections began to rise, beginning in mid-June 2020. While the mean number of bar and restaurant visits was lower overall due to closures, analysis showed that these visits remained a top factor that impacted mobility for all three counties, even with a rise in cases. Our modeling results suggest that there were mobility pattern differences between counties with respect to factors relating, for example, to race and ethnicity (different population groups factored differently in each county), as well as social distancing or travel-related factors (e.g., staying at home behaviors) over the two time periods prior to and after the spike of COVID-19 cases.

Keywords: COVID-19; mobility; random forest; spatial modeling; Florida

1. Introduction

Since January 2020, when the first confirmed case of the SARS-CoV-2 coronavirus disease (COVID-19) was reported in the United States, the pandemic has ravaged the United States, with the number of confirmed cases and deaths at over 30.2 million and 551,000, respectively, as of March 2021 [1]. Questions about how to best slow or stop the spread of this highly infectious disease, including what are the key factors that have enabled the spread of the virus and what can be done to impede its deadly progress, remain under study. The movement of people as they go about their daily lives or travel over larger spatial extents (e.g., travel by air) has been a key focus of study, throwing a spotlight on the role of mobility in sustaining the level of infection and transmission [2,3]. Tracking the movement of individuals as they undertake daily activities using the expanding location-based services via applications that apply passive tracking technologies [4–6] allows us to dig deeper into the role of mobility in infectious disease modeling.

In this paper, we investigate mobility patterns, i.e., mean inflow trip patterns, during a peak period of the pandemic, May, June, and July 2020, for three Florida counties, Miami-Dade, Broward, and Palm Beach. We use a random forest regression model to determine how a set of more than 30 different factors, including sociodemographic (e.g., median household income, age, race, and ethnicity), travel (e.g., mean travel time to work, percent of the population working from home), and built environment factors (e.g., road network density, street intersection density), as well as the changing number of COVID-19 positive

cases, relates to changing levels of mobility across the tri-county region. Our study is at the detailed granularity of census tracts, highlighting how human behaviors relating to mobility across tracts and between counties varied over time and space, and providing insights for planning, as well as possible consequences for pandemic outcomes.

Florida's unique attractions (highly regarded oceanside beaches, hotels and resorts, and year-round warm weather) make the state a draw for tourists and travelers year-round, giving Florida a unique status of possibly being a driver for virus transmission beyond its borders [7]. Local population groups of diverse race and ethnicity succumbed to high levels of infection, which, combined with the high number of elderly residents, contributed to over 2 million confirmed cases and 33,000 deaths as of March 2021 [8].

Machine learning algorithms [9], and random forest models in particular [10], are widely used in geospatial modeling by providing determinant-specific spatial contexts. These models have been especially useful for identifying explanatory variables and assessing the importance of these variables with respect to dependent variables, such as transport mode choice decision prediction, transportation mode recognition, travel demand system prediction, and explanation of drivers for forest change [11–14]. A random forest regression model is a meta-estimator that fits a number of decision trees to various subsamples of the dataset, and uses averaging to improve the predictive accuracy and control over-fitting [15,16]. Generally, random forest models are a good choice for regression and classification tasks based on their advantages, e.g., little preprocessing (rescaling or transforming) of the data is required, the modeling can be parallelizable, are compatible with high dimensional data, and are typically robust to outliers and unbalanced data [17]. Comparisons of random forest models with other machine learning algorithms (e.g., linear regression, decision tree, artificial neural network, and support vector machine) for geospatial modeling find that the random forest model performance, in terms of both computation time and prediction accuracy, is generally positive [18,19].

We used a random forest model for examining explanatory factors (i.e., sociodemographic, travel-related, built environment, and health factors) and their relative importance for revealing drivers underlying patterns of mobility based on inflow trips in the context of rising COVID-19 cases in three key counties in Florida.

2. Related Work

Studies published since the pandemic began to show the effect that COVID-19 has had on employment, education, and the economy. Franch-Pardo et al. conducted a systematic review of scientific articles on geospatial and spatial-statistical analysis of COVID-19 using perspectives drawn from spatiotemporal analysis, health and social geography, environmental variables, data mining, and web-based mapping [20]. New mobility platforms using mobile device data from SafeGraph, Google mobility reports, and Descartes Labs [6,21–23] have shown the dynamic nature of mobility data at different granularities, e.g., county, metropolitan area, and state. The University of Maryland's COVID-19 Impact Analysis Platform reports daily updated mobility-related data products (e.g., social distancing index and trip distances) [24]. Facebook, in partnership with academic institutions, created a global COVID-19 symptom survey that invites users to report on COVID-19 related symptoms, social distancing behaviors, and vaccine acceptance on a daily basis [25].

Mobility restrictions have been posited to be effective for constraining disease transmission within and between communities [26], and mobility data that have been collected from mobile devices and location-based applications can be measured against a baseline from pre-pandemic times to provide insights for policymakers and epidemiologists interested in monitoring social distancing and the spread of COVID-19 [5,27]. Investigations of mobility trends indicate that stay-at-home orders were largely effective [28].

Numerous researchers have examined the relationship between human mobility and COVID-19 infection rates. For example, analysis using mobile device location data from across the U.S. and a simultaneous equations model (SEM) found a positive relationship between inflow trips for each U.S. county and COVID-19 infections, which may be useful for

gauging the relationship between mobility and COVID-19 transmission risks [4]. Gao et al. examined the association between the rate of human mobility changes of mobile phone users (i.e., change rates of median travel distance and median home dwelling time), and the rate of confirmed COVID-19 cases in 50 U.S. states and the District of Columbia, finding that social distancing mandates were associated with the slowing of COVID-19 spread, especially when stay-at-home orders were to be lifted and states were planning for reopening their economies [6]. Other dimensions were also studied, including socioeconomic factors, such as population, household income [29], age, race, and ethnicity. A multinational study investigated the relationship between the severity of COVID-19, mobility changes, and lockdown measures, and found that lockdown measures were significant with respect to encouraging people to maintain social distancing, while the severity of socioeconomic and institutional factors (e.g., median age, percentage of the population employed in services, and percentage of health expenditure) may have limited effects to sustain social distancing [30]. It has also been demonstrated that COVID-19 case positivity during spring break in New York City was independently associated with mobility, and largely driven by residents' socioeconomic status, including proportion of population living in households with more than three inhabitants and proportion of the 18- to 64-year-old population that is uninsured [31]. Behavioral changes, measured by multiple mobility metrics for March to May 2020, also seem to matter, with senior communities reacting faster and longer in response to the stay-at-home orders compared to younger communities [32]. Research by Lou et al. involved a comparative analysis of responses between lower-income and upper-income groups, and assessed their relative exposure to COVID-19 risks at the county level [33]. Analysis results showed that higher incomes were related to an improvement in social distancing behavior [34]. This research informed our study such that levels of income and poverty were included in the random forest model as explanatory variables.

A variety of regression models and algorithms have been used to predict or explain the occurrence of COVID-19. Mollalo et al. modeled over 50 environmental, socioeconomic, topographic, and demographic candidate explanatory variables, as well as age-adjusted mortality rates of several disease factors at the county level across the U.S. using geographically weighted regression (GWR) and machine learning algorithms, such as artificial neural network (ANN). The interest was in identifying significant explanatory variables (e.g., median household income, income inequality, and age-adjusted mortality rates of ischemic heart disease) and hotspots of COVID-19 incidence [35,36].

3. Materials and Methods

3.1. Data and Study Area

The study area for this research comprises three counties in Florida, Miami-Dade, Broward, and Palm Beach, located in the southeastern tip of Florida. One of the unique characteristics of Florida is the large population of retirees (over 65 years), approximately 18% of the state's total population. The southeastern part of Florida also has a diverse population with respect to race and ethnicity; for example, Hispanics comprise 68% of Miami-Dade and 30% of Broward counties, respectively, Blacks represent approximately 29% of Broward County, and White non-Hispanics represent 55% of Palm Beach County (Table 1), based on the 2019 American Community Survey (ACS) [37].

Table 1. Demographics of Miami-Dade, Broward, and Palm Beach counties.

	Miami-Dade		Broward		Palm Beach	
# of Census Tracts	519		362		338	
Total population	2,699,428		1,926,205		1,465,027	
Race and ethnicity						
Black	469,202	17.38%	551,097	28.61%	273,384	18.66%
White	2,028,500	75.15%	1,170,083	60.75%	1,077,422	73.54%
Non-Hispanic	850,503	31.51%	1,351,916	70.19%	1,137,087	77.62%
Black Non-Hispanic	426,336	15.79%	530,990	27.57%	266,676	18.20%
White Non-Hispanic	356,026	13.19%	698,805	36.28%	799,422	54.57%
Hispanic	1,848,925	68.49%	574,289	29.81%	327,940	22.38%
Black Hispanic	42,866	1.59%	20,107	1.04%	6,708	0.46%
White Hispanic	1,672,474	61.96%	471,278	24.47%	278,000	18.98%
Gender						
Male	1,311,459	48.58%	938,043	48.70%	710,241	48.48%
Female	1,387,969	51.42%	988,162	51.30%	754,786	51.52%
Median household income (USD)	52,669		57,433		62,571	
Age group						
0–19	615,919	22.82%	451,353	23.43%	313,436	21.39%
20–39	736,246	27.27%	501,570	26.04%	338,567	23.11%
40–59	765,800	28.37%	539,530	28.01%	373,605	25.50%
60–79	459,748	17.03%	349,128	18.13%	331,428	22.62%
80 and above	121,715	4.51%	84,624	4.39%	107,991	7.37%

We used mobility data provided by the Maryland Transportation Institute (MTI) at the University of Maryland. These data included origin–destination trips data computed from mobile device locations that capture travel patterns at the granularity of census tracts for four time periods per day (6a.m.–10a.m., 10a.m.–2p.m., 2p.m.–6p.m., and 6p.m.–6a.m.) [4]. The origin and destination trips data were aggregated into inflow (the number of trips per person flowing into a specific census tract from all other places) and outflow (the number of trips per person flowing out of a specific census tract to all other tracts). As there was very little difference in the patterns of inflow and outflow trips per person per census tract, i.e., when there is a trip flowing into a specific census tract there is usually a trip going out, the number of inflow trips per person per tract was used to analyze mobility in this study (Figure 1). Inflow trips per person per unit have also been used in other studies for analyzing mobility [4,28].

As of March 2021, these three counties had the highest COVID-19 severity in the state of Florida, contributing a total of approximately 38% of the total positive cases and approximately 33% of total deaths [8], while these three counties comprise over 28% of the total population of Florida. Miami-Dade County was the first county to implement a stay-at-home order among all Florida counties (March 2020), and was the last to lift the order and enter a reopening phase (May 2020). During this March–May 2020 stay-at-home order period, the cumulative COVID-19 cases reached a total of over 31,000 in the three counties; the number of cases in Florida during the same period reached over 55,000 [1]. After the stay-at-home order was lifted, COVID-19 cases remained low for the month of May, and then, in mid-June, cases began to increase. We examined data for May, June, and July 2020 (a total of 92 days).
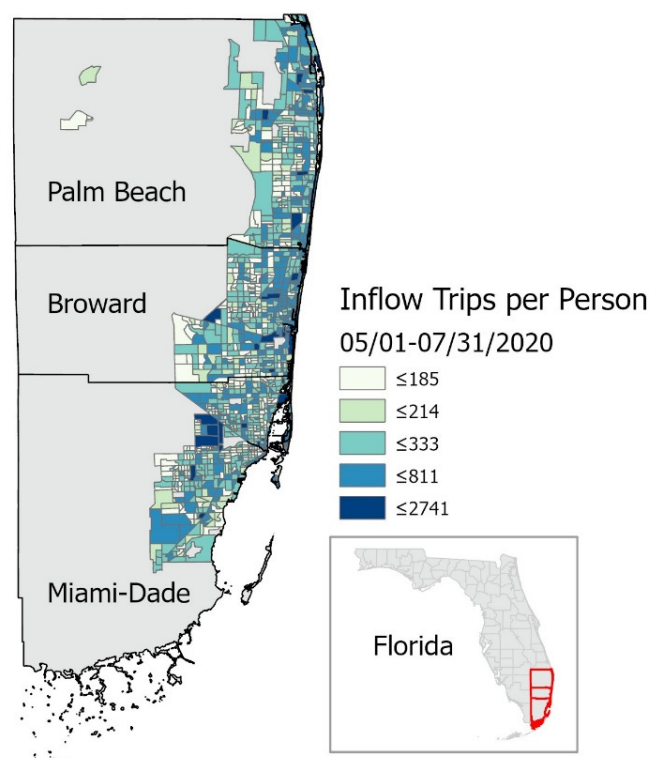
Figure 1. Inflow trips per person per census tract 05/01–07/31/2020 in Miami-Dade, Broward, and Palm Beach counties.

County-level data were available from 2 March 2020, when the first COVID-19 case was reported in Florida; ZIP code level COVID-19 case number data were made available from the Florida Department of Health (DOH) public dashboard from 18 May 2020 [38].

The first two weeks of May were extrapolated based on the overall COVID-19 trend at county level. To be consistent with the other study variables, the ZIP code level data were converted to census tracts using the HUD USPS ZIP Code Crosswalk provided by the U.S. Department of Housing and Urban Development's Office of Policy Development and Research [39]. The relationship between the daily median inflow trips per person per census tract and daily new COVID-19 cases shows an increase in the number of cases in all three counties after the middle of June 2020 (Figure 2). We divided the 3-month period into two time segments, i.e., 1 May to 15 June 2020, and 16 June to 31 July 2020 (both 46 days), and ran random forest models separately for these two periods in order to investigate any changes in the factors that might underlie mobility during these times.
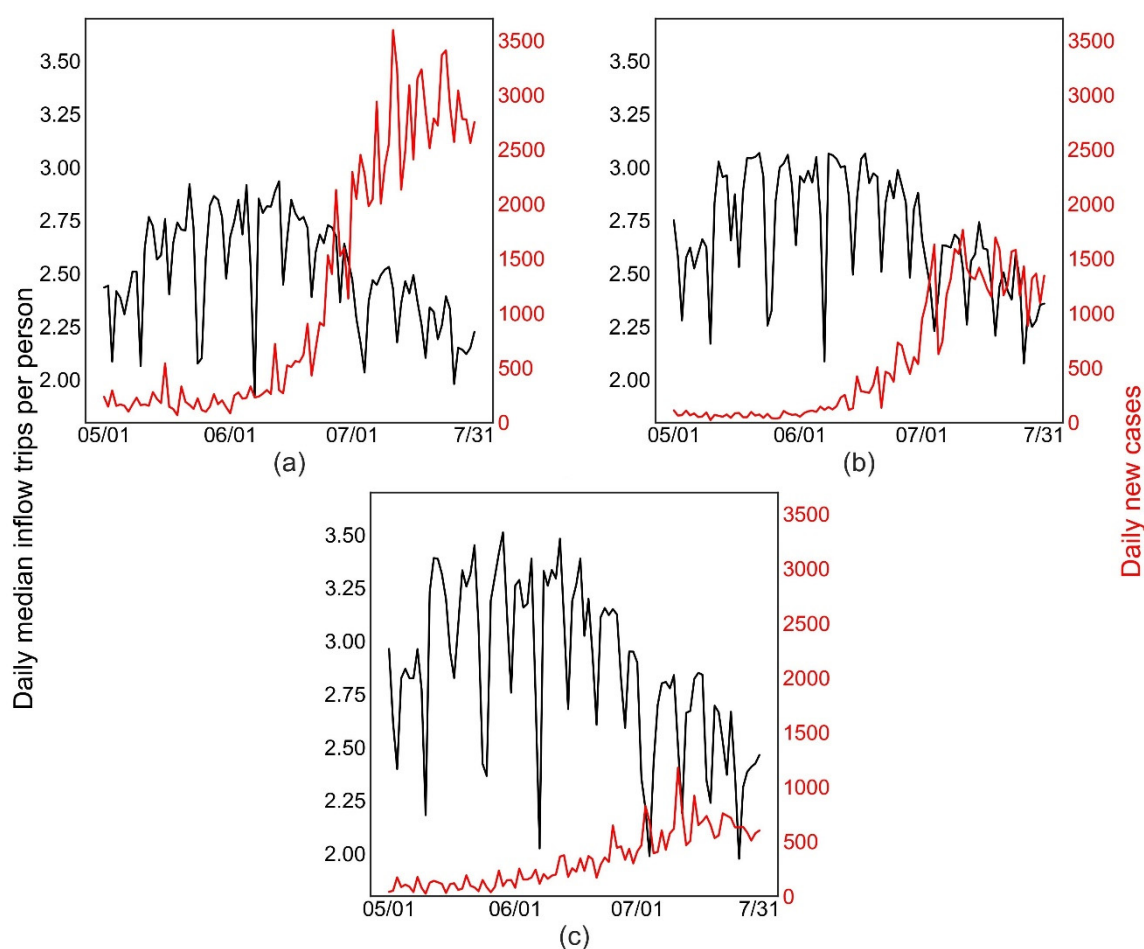
Figure 2. Median daily inflow trips per person and daily new COVID-19 cases (05/01–07/31/2020) for (**a**) Miami-Dade County, (**b**) Broward County, and (**c**) Palm Beach County.

We collected additional explanatory variables across three different categories: sociodemographic, travel, and built environment. Sociodemographic factors refer to sociological and demographic population characteristics collected from 2019 ACS, including income, employment, education, race and ethnicity (Figure 3), gender, age, and work-related measures. These variables were collected and processed at census tract level. Population demographic details have already been listed in Table 1. In this paper, Black non-Hispanic populations refer to Black, and White non-Hispanic populations refer to White. Based on previous studies finding that different income groups respond differently to the COVID-19 outbreak in terms of practicing social distancing [33,34], a factor representing essential workers was included in the model using 2019 ACS data and calculated based on a ratio of service and production occupations, transportation, and material moving occupations to all occupations.
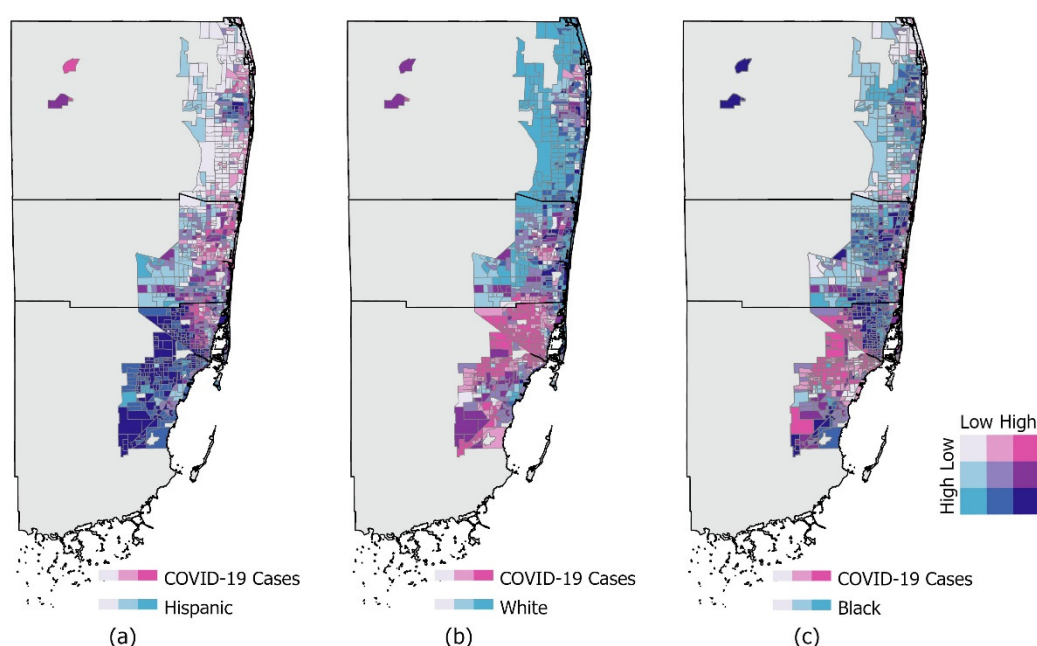
Figure 3. Bivariate mappings of COVID-19 cases per 10,000 people and (**a**) percent of Hispanic population, (**b**) percent of White population, and (**c**) percent of Black population.

Travel-related factors included human mobility behavioral changes impacted by stay-at-home orders, work travel movements, travel distance to beaches, etc. The principal beaches in each county (i.e., Miami Beach, Fort Lauderdale Beach, and Palm Beach) attract both tourists and local people, and we assumed these points of interest play an important role in daily mobility patterns during the COVID-19 pandemic. For this reason, the Euclidean distance from census tracts to their corresponding nearest beaches was calculated as one of the travel-related factors. To capture how people's behaviors changed under social distancing requirements, SafeGraph's Social Distancing Metrics dataset consisting of three different variables: percent of time dwelling at home, percent of devices completely at home, and percent of both full-time and part-time work behaviors (defined as devices spending over 3 h at a location other than their home from 8am to 6pm) at census block group level were used in this study [40]. The data were generated using GPS locations from anonymous mobile devices to census tract level for consistency. In addition, SafeGraph also provided POI daily visit pattern data at census block group level. Among all the POIs, bars (NAICS code = 722410) and restaurants (NAICS code = 722511) are typically correlated with higher exposure to COVID-19, and limits on bar and restaurant operations have been considered one of the most effective social distancing implementations [41]. The numbers of bar- and restaurant-related POIs for the three counties during May–July 2020 vary by county (Table 2). The numbers of bars open in all three counties were likely lower than normal due to COVID-19 business closures. We processed and aggregated the mean daily bar and restaurant visits by census tract for processing in the random forest model.

Table 2. Numbers of bars and restaurants in Miami-Dade, Broward, and Palm Beach counties during May–July 2020 (from SafeGraph).

POI	Miami-Dade	Broward	Palm Beach
Bars	68	48	45
Restaurants	5609	3725	2605

Built environment factors were obtained from the Smart Location Database, which is a nationwide geographic data resource for measuring location efficiency maintained by the United States Environmental Protection Agency [42]. Among the more than 90 attributes

summarizing characteristics, e.g., neighborhood design, transit service, and employment, a set of four spatial and built environmental variables that are most relevant to this study were selected: gross employment density, road network density, street intersection density, and distance to the nearest transit stop. The dataset was available at the census block group level, which was processed to census tract level for the random forest model. Details of the explanatory and dependent variables used in this analysis, and data sources for the variables are provided (Table 3).

Table 3. Explanatory variables and the dependent variable used in this study.

Category	Variables	Sources
	Explanatory variables	
Sociodemographic	Median household income	2019 ACS
	Unemployment rate	
	Average household size	
	Percent of population with low, medium, and high wages	
	Percent of population with high school degree	
	Percent of population with bachelor's degree or above	
	Percent of the Black population	
	Percent of the White population	
	Percent of the Hispanic population	
	Sex ratio (number of males per 100 females)	
	Age groups 0–19, 20–39, 40–59, 60–79, 80+	
	Percent of the population working from home	
	Percent of population defined as essential workers	
Travel-related	Mean travel time to work	2019 ACS and SafeGraph
	Distance to beach	
	Percent of time dwelling at home	
	Percent of devices completely at home	
	Percent of full-time and part-time work behaviors	
	Mean bar/restaurant visits	
Built environment	Gross employment density	Smart Location Database
	Total road network density	
	Street intersection density	
	Distance from centroids to the nearest transit stop	
COVID-19	Cumulative COVID-19 positive cases (05/01–07/31/2020) per 10,000 people	Florida DOH
	Dependent variable	
Mobility	Inflow trips per person per census tract (05/01–07/31/2020) at census tract level	MTI

3.2. Random Forest Model

We used Python as the processing language and Scikit-learn as the Python machine learning package. Before splitting the dataset into training and testing sets, extreme observations were filtered out in order for these values not to influence the regression model. This included census tracts with a total population less than 500 and population density less than 0.0001, as these were considered to be not representative (e.g., tracts containing the Miami International Airport and the Everglades National Park). Moreover, outliers in the daily trips per person (i.e., the dependent variable), exceeding the 90th percentile, were removed to avoid the influence of extreme and unusual values skewing the models. The remaining data contained 1065 observations at census tract level, which were randomly divided into two subsets. A training set comprising 80% of the data was used to develop the random forest model with 5-fold cross-validation (we also tested with 10-fold cross-validation), and a testing set comprising 20% of the data was used to assess model performance. To analyze the effect of the training and testing set split ratios, other split ratios, including 60–40%, 70–30%, and 75–25%, were also tested to understand the impact on model performance. Four evaluation measures were used to assess the

model performance: (1) Pearson correlation coefficient (*r*) between the observed values and predicted values, (2) the coefficient of determination (R^2), (3) root mean square error (*RMSE*), and (4) mean absolute error (*MAE*). *RMSE* and *MAE* are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (2)$$

While parameter tuning is often applied to avoid overfitting, this step also seeks the optimal combination of given parameters for the best model performance. Four parameters were tuned, including the number of trees (*n_estimators*), maximum depth of trees (*max_depth*), the number of features considered when looking for the best split (*max_features*), and the minimum number of samples required to be at a leaf node (*min_samples_leaf*). Then, each combination of parameters was trained with 5-fold cross-validation while the optimal parameters were selected, and the best model performance was returned.

Overfitting occurs when the model is overly trained, resulting in a good fit for a limited set of data, but performs unsatisfactorily when it comes to the unseen out-of-bag testing samples. To prevent overfitting, several techniques were applied in this study, including recursive feature elimination (RFE), which is a feature selection algorithm, parameter tuning, oversampling [43,44], and adding cost-complexity pruning (CCP) for regularization.

After the optimal model was trained and tested, the contributions of explanatory variables for mobility patterns (i.e., inflow trips) in each county were assessed by visualizing a ranked list of feature importance. In this study, we used the Gini importance to evaluate the feature importance [45]. Gini importance is computed as the (normalized) total reduction of a criterion, i.e., the function to measure the quality of a split of randomized decision trees (i.e., the random forest) brought about by a specific feature. We use mean squared error (*MSE*) as the criterion, and the function was computed by the Sci-kit learn package. The three counties were trained first as one model, and then a model for each county was trained separately for the two time periods so that any differences with respect to feature importance could be compared, and county patterns and trends could be identified.

4. Results

4.1. Mobility Patterns and Related Sociodemographic Factors in the Three Counties

Our primary interest was in investigating how mobility patterns changed across the three counties during a time in the pandemic when cases were rising, and what were the driving factors underlying these changes. At the county level, the pattern of COVID-19 daily new cases with daily median inflow trips per person (Figure 2) showed an increase in the number of cases beginning in mid-June 2020 and continuing into July. In contrast, mobility changes from the first time period to the second declined by −6.07%, −6.29%, and −10.62% for Miami-Dade, Broward, and Palm Beach counties, respectively (Table 4). Prior to mid-June 2020, Palm Beach and Broward counties experienced higher inflow trips per person than Miami-Dade County, and Palm Beach County experienced the largest decrease in mobility overall from the first time period to the second compared to the other two counties. Palm Beach County maintained the highest inflow trips per person and the lowest COVID-19 case numbers in the second time period.

Table 4. Total inflow trips for 05/01–06/15/2020 and 06/16–07/31/2020 for Miami-Dade, Broward, and Palm Beach counties.

County	05/01–06/15	06/16–07/31	Change (%)
Miami-Dade	388,724,381	365,125,529	−6.07
Broward	280,165,073	262,556,430	−6.29
Palm Beach	219,750,854	196,404,838	−10.62

Pearson correlation coefficients were computed to determine the relationships between inflow trips per person and sociodemographic variables, including median household income and age, with significance levels of $p < 0.05$, $p < 0.01$, and $p < 0.001$ (Table 5). For the first time period, for Miami-Dade and Palm Beach counties, the correlation between mobility and median household income was weakly positive, while, for Broward County, it was weakly negative. For the second time period when COVID-19 cases were spiking, Miami-Dade dipped to a weakly negative correlation with median household income, while Palm Beach (with fewer new COVID-19 cases) remained weakly positive (relationship for Broward County did not change). Examining the relationships between mobility and age groups showed that younger aged groups tended to be negatively correlated with mobility, both before and after the peak in cases, while, for older age groups (over 60 years), there was a weak positive correlation in Miami-Dade and Broward counties and a weak negative correlation in Palm Beach County. For the second period where COVID-19 was higher, these relationships continued to hold, suggesting that, in Palm Beach County, there was more concern about the increase in COVID-19 among older-aged individuals.

Table 5. Pearson correlation analyses between inflow trips per person and median household income and age groups for Miami-Dade, Broward, and Palm Beach counties for 05/01–06/15/2020 and 06/16–07/31/2020.

	Miami-Dade		Broward		Palm Beach	
	05/01–06/15	06/16–07/31	05/01–06/15	06/16–07/31	05/01–06/15	06/16–07/31
Income	0.0957 *	−0.0268	−0.0301	−0.1097 *	0.1570 **	0.0247
Age group						
0–19	−0.0701	−0.0717	−0.1742 ***	−0.1593 **	0.0379	0.0517
20–39	0.0576	0.1256 **	−0.0069	0.0303	0.1434 **	0.1732 **
40–59	−0.0965 *	−0.1566 ***	0.1398 **	0.1146 *	0.1296 *	0.1262 *
60–79	0.0344	0.0148	0.0652	0.0386	−0.0993	−0.1244 *
80 or above	0.0973 *	0.0771	0.0081	0.0068	−0.1338 *	−0.1404 **

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.2. Mobility Patterns and Travel-Related Behaviors

The stay-at-home orders for these three counties were issued at similar times: Miami-Dade County on March 26, and Broward County and Palm Beach County on March 27. Palm Beach County lifted its stay-at-home order on May 11, while Miami-Dade and Broward counties were part of the reopening phase on May 18. Two variables that related to how individuals responded to restrictions in travel, median percent of time dwelling at home (Figure 4a) and percent of population staying completely at home (Figure 4b), were analyzed at county level. The figures suggest that, after the stay-at-home orders were lifted, the percent of time people spent dwelling at home decreased and remained relatively low through mid-June, when COVID-19 cases began to spike in this part of Florida and continued to be relatively low compared to the stay-at-home period through the end of July (Figure 4a). Miami-Dade County had the highest overall percent of the population who stayed at home throughout the three-month period (Figure 4b), while Palm Beach County had the lowest percent.
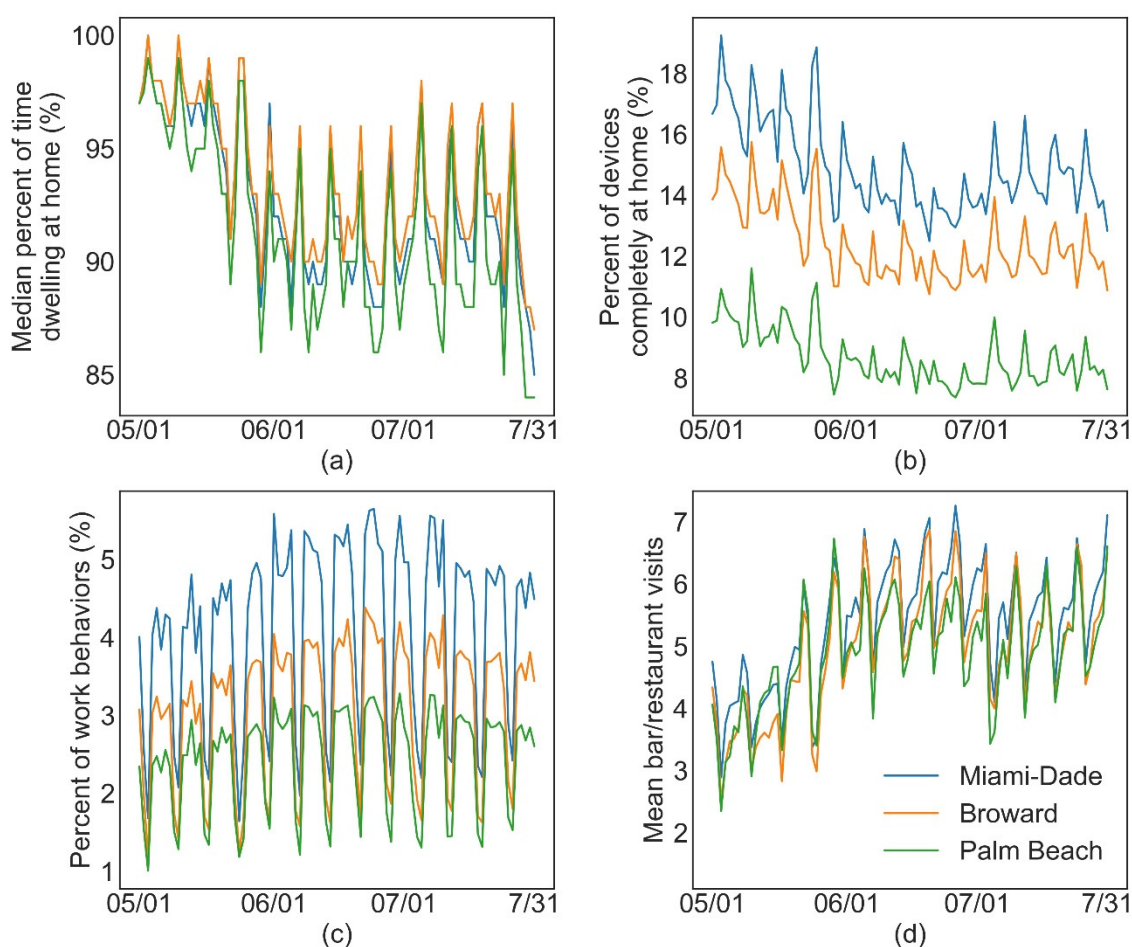
Figure 4. Mobility-related behaviors during 05/01–07/31/2020, including (**a**) median percent of time dwelling at home, (**b**) percent of devices completely at home, (**c**) percent of both full-time and part-time work behaviors (i.e., devices spending over three hours away from home), and (**d**) mean bar and restaurant visits.

Patterns associated with either full-time and/or part-time work behaviors were captured through tracking mobile devices that spent more than 3 h per day away from home (Figure 4c). While all three counties had similar patterns with respect to the percent of devices that spent more than 3 h per day away from home, steadily increasing from early May to mid-June followed by a decrease from mid-June to the end of July, Miami-Dade County had the highest proportion of devices with such pattern, suggesting either full-time and/or part-time work behaviors, while Palm Beach County had the lowest, suggesting different rates of work-related behaviors in the three counties.

While there was an overall lower level of mean bar and restaurant visits for the three counties due to COVID-19-related closures, our analysis showed that there was a steady increase in bar and restaurant visits until mid-June, when these types of outings showed a sudden decrease followed by a subsequent increase again in early July (Figure 4d).

4.3. Random Forest Models

4.3.1. Model Performance

Thirty explanatory variables (Table 3) were trained separately for each of the two time periods as features for the random forest regression models. The performance of all random forest models was assessed using the measures of r, R^2, $RMSE$, and MAE (Table 6). We found some interesting variations between the models for each of the counties. With respect to values of r, i.e., the correlation between the observed values and predicted values that reflect how well the predictive model performed, the Palm Beach model returned the highest r values (0.6781 and 0.6766, respectively), followed by Broward and Miami-

Dade. This suggests perhaps that the set of analyzed variables performed slightly better for Palm Beach when it came to being able to predict mobility patterns than for the other two counties.

Table 6. Random forest model performance for 05/01–06/15/2020 and 06/16–07/31/2020 for Miami-Dade, Broward, and Palm Beach counties.

	Miami-Dade		Broward		Palm Beach	
	05/01–06/15	06/16–07/31	05/01–06/15	06/16–07/31	05/01–06/15	06/16–07/31
r	0.5104	0.6068	0.5496	0.6712	0.6781	0.6766
R^2	0.2555	0.3549	0.2964	0.3666	0.4358	0.4415
RMSE	34.03	33.31	44.22	42.67	37.27	37.80
MAE	27.21	26.64	36.61	35.48	31.14	28.89

The coefficient of determination (R^2) that measures the percentage of the response variable variation that is explained by the random forest model was also found to be highest for Palm Beach County, while the R^2 values for both Miami-Dade and Broward counties for the second time period (when cases were rising) were higher than that of the first time period. As we were not able to collect and include all the variables that could be impactful for mobility, for example, changes in employment due to the pandemic and COVID-19 mortality and hospitalization data, it is not completely surprising that the models showed room for improvement. In terms of prediction errors, Broward County had the highest *RMSE* and *MAE*, although the values were similarly strong across all models. In general, the model performance for the second time period was better than that of the first time period with higher *r* values and lower error values.

4.3.2. Feature Contributions for the Period Prior to the Rise in COVID-19 Cases

Feature importance scores for the three counties were analyzed to obtain an understanding of how the different factors ranked in importance according to the random forest model, with respect to the number of inflow trips per person. During the first time period (05/01–06/15/2020), when mobility was relatively high, COVID-19 cases were still relatively low, the number of new COVID-19 cases was ranked 7th in importance in Broward and 8th in Miami-Dade, while, for Palm Beach County, this variable was not among the top 15 factors ranked by importance scores. While COVID-19 cases were not so high, the importance scores for both the built environment factors and travel-related factors ranked higher overall than sociodemographic factors (Figure 5). Gross employment density was ranked very highly for all three counties (1st for Broward and Palm Beach, and 2nd for Miami-Dade). Other built environment factors, e.g., street intersection density and road network density, were also present in the top 15 factors for all three counties. With respect to travel factors for the first period, these were highly ranked in all three counties, with mean bar and restaurant visits ranked 1st for Miami-Dade, 2nd for Palm Beach, and 5th for Broward. Time spent completely at home, full-time and part-time work behaviors (based on devices being away from home for more than 3 h), median percent of time dwelling at home, and other social distancing factors were also in the top 15 factors for all three counties, suggesting that the population was also sensitive to the ongoing COVID-19 situation in their region.
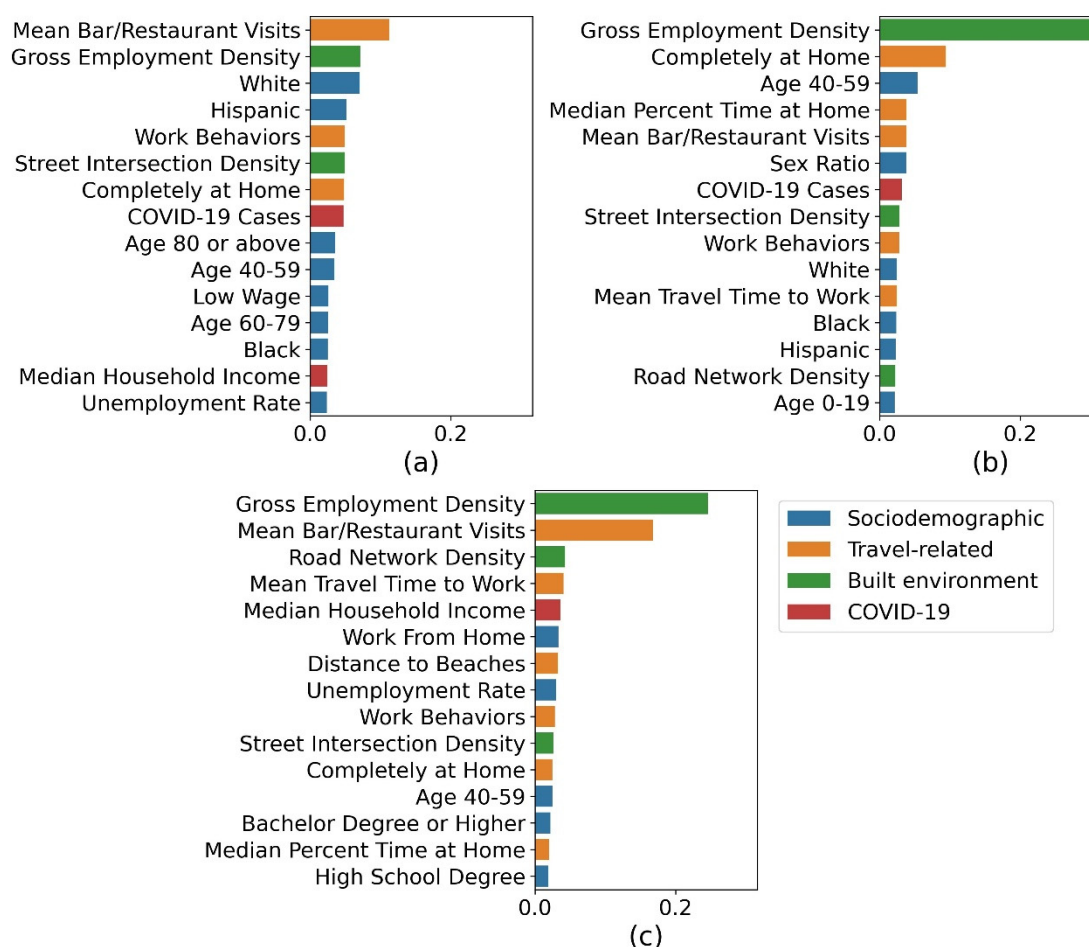
Figure 5. The relative importance of the top 15 variables to the number of inflow trips per person (05/01–06/15/2020) using random forest models for (**a**) Miami-Dade County, (**b**) Broward County, and (**c**) Palm Beach County.

With regard to sociodemographic factors during the first time period for Miami-Dade County, the percent of White and Hispanic population was ranked 3rd and 4th, respectively, for Miami-Dade County. White and Hispanic populations contribute, respectively, approximately 13% and 68% of the total population for Miami-Dade (Figure 5a). In Broward County, the percent of both Black and White populations were also in the top 15 rankings, albeit not as highly ranked (positions 9 and 12, respectively), and the percent of Hispanic population was 13th in the rankings (Figure 5b). For Palm Beach County, the results were different, with important sociodemographic factors relating to income (median household income ranked 5th), employment (general unemployment levels ranked 8th), and education (bachelor's degree and high school degree ranked 13th and 15th, respectively) rather than race and ethnicity (not one of the top 15 factors) (Figure 5c). These intercounty differences in the model results relating to sociodemographic factors are interesting to note and underscore the kinds of population differences that exist between the counties.

4.3.3. Feature Contributions for the Period Following the Rise in COVID-19 Cases

As the number of new COVID-19 cases began to spike in mid-June 2020, the second period captured some changes in the ranking of variables based on importance scores. Factors that ranked highest in importance during this period continued to be those related to travel and built environment (Figure 6). Both gross employment density (1st for all three counties) and the mean number of bar and restaurant visits (2nd for all three counties) continued to be top factors for all the models. In Palm Beach County, the importance scores for these two factors were much higher than for the other counties (Figure 6c). Built environment factors, e.g., street intersection density and road network density, were

still present in the rankings. Job- and work-related factors, i.e., mean travel time to work and full-time and part-time work behaviors, were most important in Palm Beach County (ranked 3rd and 4th, respectively), while, for Miami-Dade County, full-time and part-time work behaviors were ranked 6th and, for Broward County, they ranked 10th. Mean travel time to work ranked 3rd in Palm Beach, 12th in Miami-Dade, and 15th in Broward County, underscoring how work-related factors seemed to continue as strong drivers in Palm Beach County, even with cases rising. Travel distance to beaches was ranked 5th for Broward and 8th for Palm Beach, while this factor was not in the top 15 for Miami-Dade County.
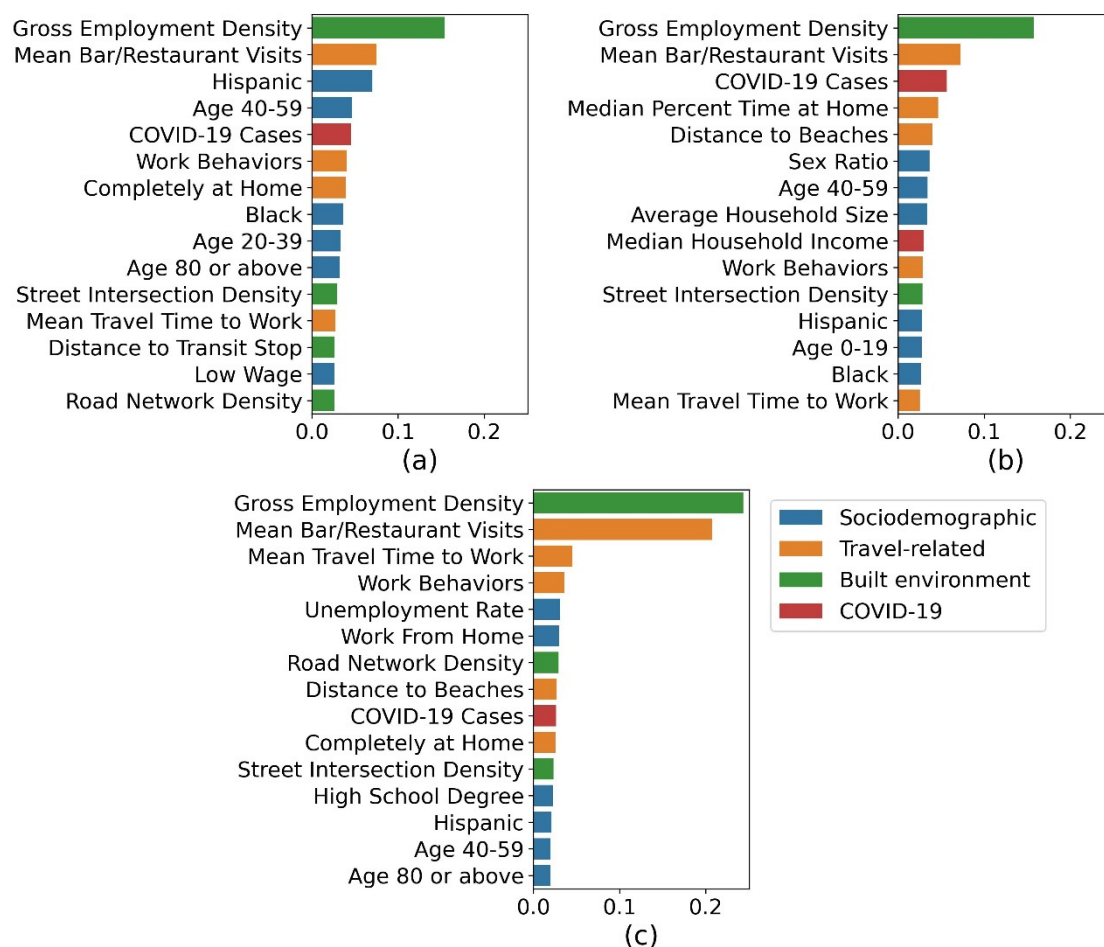

Figure 6. The relative importance of the top 15 variables for the inflow trips per person (06/16–07/31/2020) using random forest models for (**a**) Miami-Dade County, (**b**) Broward County, and (**c**) Palm Beach County.

With respect to sociodemographic factors for the second time period, the percent of Hispanic population was a factor in all three county models, but was much more of a factor for Miami-Dade County, where it ranked 3rd, while it was 12th in Broward and 13th in Palm Beach. Black population was 8th in importance in Miami-Dade and 14th in Broward County (not present in the Palm Beach rankings). The age group 40–59 years was another common factor, but with different importance, as it ranked 4th for Miami-Dade, 7th for Broward, and 14th for Palm Beach, although the percent population corresponding to ages 40–59 was similar across the three counties (approximately 28%, 28%, and 26%, respectively). The factor of age 80 or above ranked at 10 in Miami-Dade and 15 in Palm Beach County. Conversely, the youngest age group (0–19 years) appeared only in Broward County and at rank 13.

The most noticeable change between the two time periods was that the factor representing the number of new COVID-19 cases was much higher ranked for the second time period, being 5th, 3rd, and 9th for Miami-Dade, Broward, and Palm Beach counties,

respectively. The random forest model was able to discern that the increase in COVID-19 was increasingly important for mobility, even in Palm Beach County where, for the first period of time, COVID-19 cases were not in the top 15 factors explaining inflow mobility.

We also analyzed a random forest model trained using all three months together. The Palm Beach model returned the highest *r* value (0.6672), followed by Broward and Miami-Dade (0.5774 and 0.4946, respectively), which is similar to the order of model performance for the two separate time periods. The results showed that the rankings of important features were similar to the period from mid-June to late July (i.e., the second time period), with mean bar and restaurant visits, gross employment density, and the percent of Hispanic population being the top three factors for Miami-Dade. These three factors were within our expectations, since Miami-Dade County is different from the other two counties in terms of race and ethnicity. Gross employment density, mean bar and restaurant visits, and median percent of time dwelling at home were the top three factors for the Broward model. Similarly, mean bar and restaurant visits, gross employment density, and mean travel time to work were the top three factors for the Palm Beach model. The time spent dwelling at home for Broward County and the mean travel time to work factor for Palm Beach County both relate to social distancing, and suggest local county populations were sensitive to the changing COVID-19 situation and how that affected work travel decisions. In this model, new COVID-19 cases were ranked 4th for Broward, 5th for Miami-Dade, and 12th for Palm Beach, reflecting the situation that, with the lowest number of new COVID-19 cases, mobility in Palm Beach County was not as influenced by COVID-19 cases, while Miami-Dade and Broward counties experienced higher numbers of new COVID-19 cases, and mobility appeared to be sensitive to this situation. The increasing importance of COVID-19 cases as a driver for changing mobility patterns is evident in our models, demonstrating that the pandemic was indeed impacting mobility.

5. Discussion

For this research, we used random forest models to understand mobility patterns during the COVID-19 pandemic in three Florida counties, including Miami-Dade, Broward, and Palm Beach counties, and examined a set of sociodemographic, travel, and built environment explanatory factors, and their relative importance for explaining patterns of mobility in the context of rising COVID-19 cases. Much of the recent research investigating mobility under COVID-19 is at county-level or state-level across the U.S. [4,6,35,36], or at nation-level [3,30]. However, this research was undertaken at census-tract granularity to discover finer-grained patterns of mobility, as well as the drivers for mobility based on the number of inflow trips for each county.

Using a random forest model, we were able to compare the contributions of the explanatory variables over the three counties and over the two time periods. A changing relationship between important features was identified. Previous research suggested an association with COVID-19 cases, and reductions in mobility were correlated with the slowing of COVID-19 spread [4,6,46]. The results of our random forest model analysis indicated that new COVID-19 cases did have an overall impact on mobility for the three counties we analyzed. In Palm Beach County, for example, this factor was much less important until COVID-19 case numbers started to rise, when this factor shifted to become increasingly important for mobility. Other studies showed that socioeconomic and institutional factors (e.g., median age, percentage of the population employed in services, and percentage of health expenditure) may have limited effects for sustaining social distancing and reduced mobility [30], and studies have also indicated a noticeable correlation between mobility and socioeconomic factors [6,32,33]. Our random forest models revealed that sociodemographic factors (e.g., race, ethnicity, and age groups) did affect the number of inflow trips (e.g., the percent of the Hispanic population in Miami-Dade County, the age group of 40–59 in Broward County, and income and employment factors in Palm Beach County) and that, based on this result, this group of factors should be considered by decisionmakers and

healthcare providers when considering strategies to reach different population groups during a spike in infections.

Due to not being able to collect and include all the variables that could be impactful for mobility, the model performance and overfitting issues could perhaps be improved by including more dimensions of data, e.g., COVID-19 mortality and hospitalization data that are strongly related to healthcare resource availability [47,48] and changes in employment due to the pandemic. In addition, estimates for essential workers were made using subcategories of occupation data in the 2019 ACS, while 2020 estimates might differ, which might also affect the random forest model results.

6. Conclusions

As the COVID-19 pandemic impacted the daily lives of individuals, this research found that, based on tracking inflow trips at census tract level for three counties in Florida, mobility was indeed impacted by COVID-19, especially when compared to mobility during the pre-COVID period (i.e., in 2019). In addition, during a summertime spike in COVID-19 cases, there were further impacts on the number of trips being made in each county. The set of key explanatory factors revealed by the random forest model were travel-related factors (e.g., social distancing and work travel-related variables) and built environment factors (e.g., gross employment density and street and road network density), while sociodemographic factors (race and ethnicity, age, household income) were also present. These three counties represent an urban region in the United States that has had a very high number of COVID-19 cases and that has high Black and Hispanic populations that have been particularly vulnerable to COVID-19 infections, as well as a significant population of individuals over the age of 65, also vulnerable to this infectious disease. These different factors that affect the number of trips made across this tri-county region (e.g., social distancing, work travel-related variables, and gross employment density) may be helpful for local officials and public health experts as they review steps and strategies, such as stay-at-home orders and business restrictions or closures. It is also important to note that counties have their unique local characteristics (sociodemographic, economic, points of interest), and our analysis showed how these different characteristics resulted in different sets of factor rankings for each county. While this study focused on counties in Florida, the methodology is generalizable to other locations across the U.S. and other regions. Future research could focus on the model performance improvement and overfitting elimination by including more variables that may be impactful on mobility, e.g., changes in employment during the pandemic, mortality and testing data if available, and trips to additional POIs. Further research on modified random forest approaches, e.g., geographically weighted random forest, could offer new opportunities for improved spatial data handling.

Author Contributions: Conceptualization, Kathleen Stewart, Guimin Zhu, Deb Niemeier; data collection, Guimin Zhu, Deb Niemeier; formal analysis, Guimin Zhu, Junchuan Fan, Kathleen Stewart; interpretation, Kathleen Stewart, Deb Niemeier, Guimin Zhu; writing—original draft, Guimin Zhu; writing—review and editing, Guimin Zhu, Kathleen Stewart, Deb Niemeier, Junchuan Fan. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Centers for Disease Control and Prevention CDC COVID Data Tracker. Available online: https://covid.cdc.gov/covid-data-tracker/#datatracker-home (accessed on 30 March 2021).
2. Kraemer, M.U.G.; Yang, C.-H.; Gutierrez, B.; Wu, C.-H.; Klein, B.; Pigott, D.M.; du Plessis, L.; Faria, N.R.; Li, R.; Hanage, W.P.; et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **2020**, *368*, 493–497. [CrossRef]
3. Nouvellet, P.; Bhatia, S.; Cori, A.; Ainslie, K.E.C.; Baguelin, M.; Bhatt, S.; Boonyasiri, A.; Brazeau, N.F.; Cattarino, L.; Cooper, L.V.; et al. Reduction in mobility and COVID-19 transmission. *Nat. Commun.* **2021**, *12*, 1–9. [CrossRef]

4. Xiong, C.; Hu, S.; Yang, M.; Luo, W.; Zhang, L. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 27087–27089. [CrossRef] [PubMed]

5. Kishore, N.; Kiang, M.; Engø-Monsen, K.; Vembar, N.; Balsari, S.; Buckee, C. Mobile phone data analysis guidelines: Applications to monitoring physical distancing and modeling COVID-19. *OSF Prepr.* **2020**. [CrossRef]

6. Gao, S.; Rao, J.; Kang, Y.; Liang, Y.; Kruse, J.; Dopfer, D.; Sethi, A.K.; Mandujano Reyes, J.F.; Yandell, B.S.; Patz, J.A. Association of Mobile Phone Location Data Indications of Travel and Stay-at-Home Mandates With COVID-19 Infection Rates in the US. *JAMA Netw. Open* **2020**, *3*, e2020485. [CrossRef]

7. Mangrum, D.; Niekamp, P. College Student Contribution to Local COVID-19 Spread: Evidence from University Spring Break Timing. *SSRN Electron. J.* **2020**. [CrossRef]

8. Florida Department of Health Florida's COVID-19 Data and Surveillance Dashboard. Available online: https://experience.arcgis.com/experience/96dd742462124fa0b38ddedb9b25e429 (accessed on 30 March 2021).

9. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

10. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.

11. Rasouli, S.; Timmermans, H.J.P. Using ensembles of decision trees to predict transport mode choice decisions: Effects on predictive success and uncertainty estimates. In Proceedings of the 17th International Conference of Hong Kong Society for Transportation Studies, HKSTS 2012: Transportation and Logistics Management, Hong Kong, China, 15–17 December 2012; Volume 14, pp. 515–522.

12. Ghasri, M.; Hossein Rashidi, T.; Waller, S.T. Developing a disaggregate travel demand system of models using data mining techniques. *Transp. Res. Part. A Policy Pract.* **2017**, *105*, 138–153. [CrossRef]

13. Santos, F.; Graw, V.; Bonilla, S. A Geographically Weighted Random Forest Approach for Evaluate Forest Change Drivers in the Northern Ecuadorian Amazon. *PLoS ONE* **2019**, *14*, e0226224. [CrossRef]

14. Jahangiri, A.; Rakha, H.A. Transportation Mode Recognition Using Mobile Phone Sensor Data. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2406–2417. [CrossRef]

15. Hao, J.; Ho, T.K. Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. *J. Educ. Behav. Stat.* **2019**, *44*, 348–361. [CrossRef]

16. Chen, C.; Liaw, A.; Breiman, L. Using Random Forest to Learn Imbalanced Data. *Discovery* **2004**, *666*, 1–12.

17. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* **2015**, *71*, 804–818. [CrossRef]

18. Nguyen, K.A.; Chen, W.; Lin, B.-S.; Seeboonruang, U. Comparison of Ensemble Machine Learning Methods for Soil Erosion Pin Measurements. *Isprs Int. J. Geo-Inf.* **2021**, *10*, 42. [CrossRef]

19. Hagenauer, J.; Omrani, H.; Helbich, M. Assessing the performance of 38 machine learning models: The case of land consumption rates in Bavaria, Germany. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 1399–1419. [CrossRef]

20. Franch-Pardo, I.; Napoletano, B.M.; Rosete-Verges, F.; Billa, L. Spatial analysis and GIS in the study of COVID-19. A review. *Sci. Total Environ.* **2020**, *739*, 140033. [CrossRef]

21. Kang, Y.; Gao, S.; Liang, Y.; Li, M.; Rao, J.; Kruse, J. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. *Sci. Data* **2020**, *7*, 1–13. [CrossRef]

22. Gao, S.; Rao, J.; Kang, Y.; Liang, Y.; Kruse, J. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *arXiv* **2020**, arXiv:2003.14228.

23. Warren, M.S.; Skillman, S.W. Mobility Changes in Response to COVID-19. *arXiv* **2020**, arXiv:2003.14228.

24. Zhang, L.; Ghader, S.; Pack, M.L.; Xiong, C.; Darzi, A.; Yang, M.; Sun, Q.Q.; Kabiri, A.A.; Hu, S. An interactive COVID-19 mobility impact and social distancing analysis platform. *medRxiv* **2020**, 1–14. [CrossRef]

25. Kreuter, F.; Barkay, N.; Bilinski, A.; Bradford, A.; Chiu, S.; Eliat, R.; Fan, J.; Galili, T.; Haimovich, D.; Kim, B.; et al. Partnering with Facebook on a university-based rapid turn-around global survey. *Surv. Res. Methods* **2020**, *14*, 159–163. [CrossRef]

26. Espinoza, B.; Castillo-Chavez, C.; Perrings, C. Mobility Restrictions for the Control of Epidemics: When Do They Work? *Ssrn Electron. J.* **2020**, 1–14. [CrossRef]

27. Chang, M.-C.; Kahn, R.; Li, Y.-A.; Lee, C.-S.; Buckee, C.O.; Chang, H.-H. Variation in human mobility and its impact on the risk of future COVID-19 outbreaks in Taiwan. *medRxiv* **2020**. [CrossRef]

28. Lee, M.; Zhao, J.; Sun, Q.; Pan, Y.; Zhou, W.; Xiong, C.; Zhang, L. Human mobility trends during the early stage of the COVID-19 pandemic in the United States. *PLoS ONE* **2020**, *15*, e0241468. [CrossRef]

29. Huang, X.; Lu, J.; Gao, S.; Wang, S.; Liu, Z.; Wei, H. Staying at home is a privilege: Evidence from fine-grained mobile phone location data in the U.S. during the COVID-19 pandemic. *Ann. Am. Assoc. Geogr.* **2021**. [CrossRef]

30. Rahman, M.M.; Thill, J.-C.; Paul, K.C. COVID-19 Pandemic Severity, Lockdown Regimes, and People's Mobility: Evidence from 88 Countries. *SSRN Electron. J.* **2020**, 1–17. [CrossRef]

31. Lamb, M.R.; Kandula, S.; Shaman, J. Differential COVID-19 case positivity in New York City neighborhoods: Socioeconomic factors and mobility. *Influenza Other Respir. Viruses* **2021**, *15*, 209–217. [CrossRef] [PubMed]

32. Kabiri, A.; Darzi, A.; Zhou, W.; Sun, Q.; Zhang, L. How different age groups responded to the COVID-19 pandemic in terms of mobility behaviors: A case study of the United States. *arXiv* **2020**, arXiv:2007.10436.

33. Lou, J.; Shen, X.; Niemeier, D. Are stay-at-home orders more difficult to follow for low-income groups? *J. Transp. Geogr.* **2020**, *89*, 102894. [CrossRef] [PubMed]
34. Sun, Q.; Zhou, W.; Kabiri, A.; Darzi, A.; Hu, S.; Younes, H.; Zhang, L. COVID-19 and Income Profile: How People in Different Income Groups Responded to Disease Outbreak, Case Study of the United States. *arXiv* **2020**, arXiv:2007.02160.
35. Mollalo, A.; Vahedi, B.; Rivera, K.M. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci. Total Environ.* **2020**, *728*, 138884. [CrossRef]
36. Mollalo, A.; Rivera, K.M.; Vahedi, B. Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4204. [CrossRef]
37. 2019 American Community Survey Single-Year Estimates. Available online: https://www.census.gov/newsroom/press-kits/2020/acs-1year.html (accessed on 6 May 2021).
38. Florida Department of Health Florida Department of Health Open Data. Available online: https://open-fdoh.hub.arcgis.com/ (accessed on 30 March 2021).
39. United States Department of Housing HUD USPS ZIP CODE CROSSWALK FILES. Available online: https://www.huduser.gov/portal/datasets/usps_crosswalk.html (accessed on 30 March 2021).
40. SafeGraph SafeGraph Social Distancing Metrics. Available online: https://docs.safegraph.com/v4.0/docs/social-distancing-metrics (accessed on 30 March 2021).
41. Wellenius, G.A.; Vispute, S.; Espinosa, V.; Fabrikant, A.; Tsai, T.C.; Hennessy, J.; Dai, A.; Williams, B.; Gadepalli, K.; Boulanger, A.; et al. Impacts of US State-Level Social Distancing Policies on Population Mobility and COVID-19 Case Growth during the First Wave of the Pandemic. *arXiv* **2020**, arXiv:2004.10172.
42. United States Environmental Protection Agency Smart Location Database. Available online: https://www.epa.gov/smartgrowth/smart-location-mapping (accessed on 30 March 2021).
43. Lemaitre, G.; Nogueira, F.; Aridas, C.K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2015**, *40*, 1–5.
44. Branco, P.; Ribeiro, R.P.; Torgo, L.; Krawczyk, B.; Moniz, N. SMOGN: A Pre-processing Approach for Imbalanced Regression. *Proc. Mach. Learn. Res.* **2017**, *74*, 36–50.
45. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.
46. Badr, H.S.; Du, H.; Marshall, M.; Dong, E.; Squire, M.M.; Gardner, L.M. Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *Lancet Infect. Dis.* **2020**, *20*, 1247–1254. [CrossRef]
47. Baud, D.; Qi, X.; Nielsen-Saines, K.; Musso, D.; Pomar, L.; Favre, G. Real estimates of mortality following COVID-19 infection. *Lancet Infect. Dis.* **2020**, *20*, 773. [CrossRef]
48. Ji, Y.; Ma, Z.; Peppelenbosch, M.P.; Pan, Q. Potential association between COVID-19 mortality and health-care resource availability. *Lancet Glob. Health* **2020**, *8*, e480. [CrossRef]