

ORIGINAL PAPER

Open Access

# Explaining demand patterns during COVID-19 using opportunistic data: a case study of the city of Munich



Vishal Mahajan\* , Guido Cantelmo and Constantinos Antoniou

## Abstract

**Background:** The COVID-19 pandemic is a new phenomenon and has affected the population's lifestyle in many ways, such as panic buying (the so-called "hamster shopping"), adoption of home-office, and decline in retail shopping. For transportation planners and operators, it is interesting to analyze the spatial factors' role in the demand patterns at a POI (Point of Interest) during the COVID-19 lockdown viz-a-viz before lockdown.

**Data and Methods:** This study illustrates a use-case of the POI visitation rate or popularity data and other publicly available data to analyze demand patterns and spatial factors during a highly dynamic and disruptive event like COVID-19. We develop regression models to analyze the correlation of the spatial and non-spatial attributes with the POI popularity before and during COVID-19 lockdown in Munich by using lockdown (treatment) as a dummy variable, with main and interaction effects.

**Results:** In our case-study for Munich, we find consistent behavior of features like stop distance and day-of-the-week in explaining the popularity. The parking area is found to be correlated only in the non-linear models. Interactions of lockdown with POI type, stop-distance, and day-of-the-week are found to be strongly significant. The results might not be transferable to other cities due to the presence of different city-specific factors.

**Conclusion:** The findings from our case-study provide evidence of the impact of the restrictions on POIs and show the significant correlation of POI-type and stop distance with POI popularity. These results suggest local and temporal variability in the impact due to the restrictions, which can impact how cities adapt their transport services to the distinct demand and resulting mobility patterns during future disruptive events.

**Keywords:** COVID-19, Demand patterns, POIs, Spatial-temporal, Crowdsensed data, Machine learning

## 1 Introduction

People undertake different activities during the day to satisfy their needs and wants. The performance of such activities depends on various population and environmental factors. The interaction of land-use and transport systems in activity generation has also been widely researched and modeled [1]. Places well connected with transport systems and dense neighborhoods will generate more trips

than less connected and sparse neighborhoods. The interaction between land-use and transport and individual constraints is captured by the concept of accessibility [2, 3]. Apart from the above "structural" factors, planned or unplanned special events [4], weather conditions [5, 6], and others, can also influence in the short-term, where, when, and how people move. Cities strive to plan, design, and operate their transportation systems based on the forecasted demand, derived out of the activities due to these factors.

In case of disruptive and highly dynamic events, such as natural or human-made hazards, people tend to adapt their short-term [7] and long-term mobility behavior

\*Correspondence: [vishal.mahajan@tum.de](mailto:vishal.mahajan@tum.de)

Chair of Transportation Systems Engineering, Department of Civil, Geo and Environmental Engineering, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany

[8, 9] to the prevailing conditions. For example, travelers might be reluctant to enter the underground metro after an earthquake or go near the sea-coast in case of a cyclone or tsunami warning. COVID-19, one of the most severe pandemics in the last 100 years, has affected almost the entire world in unprecedented ways. To control COVID-19 transmission, guidelines such as social distancing, masks, and movement restrictions were recommended or enforced. In response to these measures, people not only reduce their mobility [10], but they also adapt their travel patterns in order to limit their exposure by avoiding places with a high number of cases [11]. The anticipation or announcement of movement restrictions due to COVID-19 causes specific changes in people's lifestyles, routines, and consumption patterns, such as panic buying during the early pandemic or lockdown stage [12], working from home, or a decline in non-essential retail consumption [13]. With people spending on average, around 40% less, these new trends also risk generating an economic slowdown that could last for a long time [13]. Such changes in the behavior and attitudes, if significant, can reveal a pattern exhibited through changes in the number, types, and spatio-temporal extent of the activities. For example, crowding at some locations or imbalanced use of transport facilities, like roads and transport modes, can be observed. Planners must understand these behavioral changes and, more importantly, the spatio-temporal patterns of the population's activities for an effective response. The scale and speed of these changes have left cities, transport operators, and research communities with several unanswered questions on how to respond so that a basic service level is efficiently maintained.

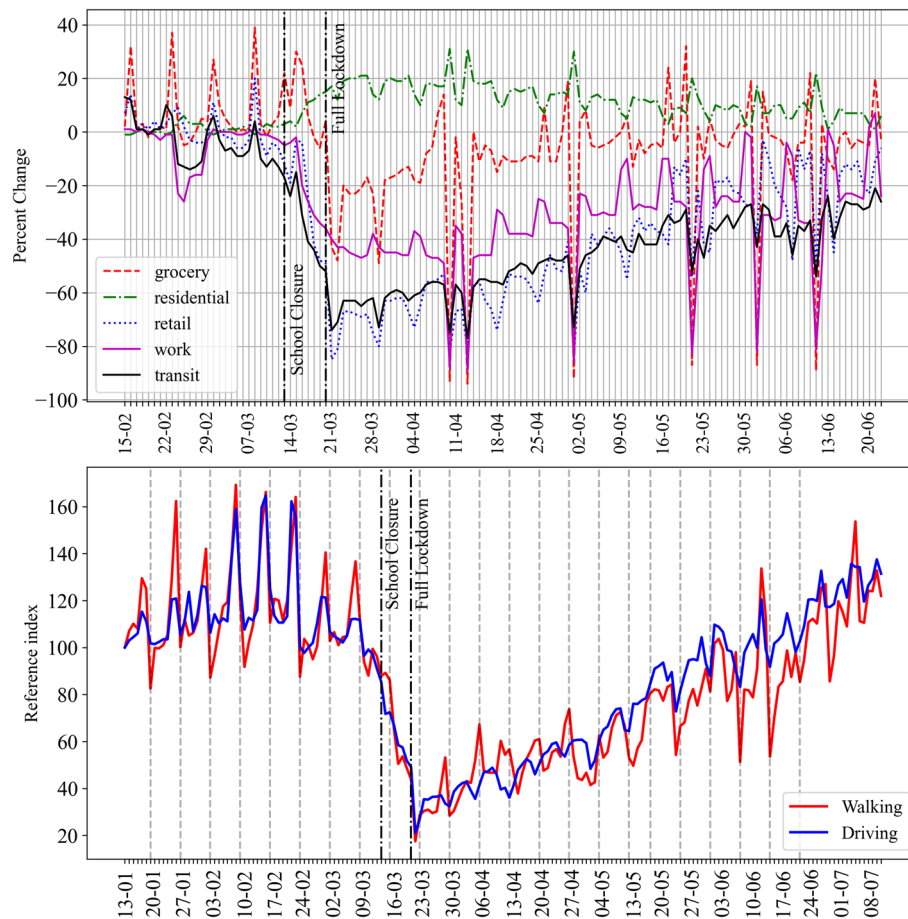
The study of human activity and travel behavior is traditionally (and commonly) based on the data from Stated and Revealed preference surveys. Emerging sources of data [14], such as social media [15] or mobile phones [16], have pushed the use of data-centric approaches to study activity patterns. Alternate data sources can play a crucial role, especially during situations like COVID-19, when responses or policies have to be adopted faster, whereas surveys take some time in planning and execution. This aspect became prominent during COVID-19 when several organizations came forward by making some of their data publicly available to help governments and citizens understand the changes in activity patterns and travel behavior. Some of the prominent examples are COVID-19 Community Mobility Reports [17] and Apple Mobility Reports [18]. For instance, we obtain activity and mobility trends for Munich (in Bavaria, Germany), which provide information about the overall changes in activity and travel mode patterns in a region, respectively (Fig. 1) from [17, 18]. In the figure, the overall activity and mobility trends confirm some expected behavioral

patterns during COVID-19 such as a decline in transit mode use, drop in retail and workplace-related visits, and increase in stays at residences. The grocery related visits can be seen to gradually recover from the initial drop in visits.

During COVID-19, Mobile phone data emerged as a potential source to understand and respond to the pandemic, as it provides a large spatio-temporal information [19]. A study using the mobile phone data found that the lockdown in France caused a 65% reduction in the performed trips, especially work-related trips during peak hours and long trips [10]. Researchers in the US and China also applied the mobile phone data to establish that the social distancing and decreased mobility (due to restrictions or lockdowns) is positively correlated to the reduced growth in COVID-19 cases [20, 21]. These few examples illustrate the potential of passively collected data during the COVID-19 for the informed policy decisions. Apart from the data source, the level of data aggregation also decides its usability. The aggregated datasets (at the city or county level) do not provide detailed information at finer geographical scales such as the POI level and thus have limited applications. On the other hand, disaggregate datasets (at finer geographical scales) could provide richer information for understanding heterogeneity across POIs and demographics [22]. Thus, disaggregate data allows for analysis at a local level for understanding the mechanisms between activity patterns and environmental factors, e.g., at the level of a shop or a transit stop.

Mobile devices or sensors equipped with wireless communication or internet can help understand when, where, and if people are crowding. Crowdsensed information, for example, consists of large datasets built with the help of a large group of people. In Mobile Crowdsensing, individuals "collectively share data and extract information" with the help of a sensing device (like smartphones) towards a common goal [23], such as identifying spatio-temporal patterns of a phenomenon. Crowdsensed data from mobile phones and social media platforms [24, 25], such as Twitter data, can, for example, help to study highly dynamic and disruptive events [26, 27]. Some of these data are available on the internet and could be exploited to create the first line of defense against this pandemic and to develop policies that can mitigate the pandemic's impact on transport systems and local businesses.

One such exciting and seemingly potential data set is the crowdsensed check-in rate or busyness at the POIs. The crowdsensed check-in rate is the representation (in absolute or relative terms) of the number of people or customers visiting a specific establishment at a given point of time and thus shows its busyness. These data are primarily collected from smartphone applications, in which the user's location history is enabled, such as geotagged data or Location-Based Social Networks (LBSNs). Geo-



**Fig. 1** (Top) Activity patterns in Bavaria and (Bottom) travel mode patterns in Munich, data source: [17, 18]

tagged tweets [27], Foursquare check-ins [28], and popularity trends [29, 30] are some examples of such data that capture the spatial-temporal evolution of the demand and have already shown their utility in previous studies. Specifically, popularity trends have been used, to predict venue popularity [30], calculate demand expansion factors [31], classify activities [29] and investigate consumer behaviors [32]. These varied applications of POI demand patterns from popularity trends suggest their potential in other unexplored avenues, such as disruptive events. Certain crowdsensed data such as popularity trends provide only relative or normalized values of the demand for certain activities in certain locations [29]. This is a limitation because it prevents one from inferring the corresponding number of absolute check-ins for which the data is recorded, and thus should be considered when designing methodologies that leverage this information as well as when processing the results. More details about how this limitation is addressed in this study will be provided in the next section, where we introduce both the data and the methodology for this study.

This paper shows that crowdsensed information could also provide some useful insights into the spatial-temporal distribution of the activities or demand during the pandemic situation. Subsequently, we propose a model that breaks down POI demand patterns into a set of crucial spatial and other attributes, which are assumed to explain the POI demand. This paper is structured as follows: Section 2 presents the study's methodology; Sections 3 and 4 concern data collection and data analysis respectively, Section 5 presents the results. Section 6 discusses the findings and limitations, and lastly, we provide a set of conclusions from the research.

## 2 Methodology

We use POI visitation data (response variable, denoted by  $P$ ) and check their correlation with the spatial and other attributes (explanatory variables) of the POI. Firstly, we define a bounding box for the study area and identify the POIs within that area. For these identified POIs, we collect the historical popularity data and live popularity data on different days to capture the before-lockdown

and during-lockdown situation, respectively. Some of the prominent spatial attributes that could affect customer visits at a POI are population density [33], parking facilities [34] and public transit stops [33] nearby the POI. As we aim to capture the spatial variability among the POIs, the selected spatial attributes should capture the local variation, e.g., threshold distance for calculating population around a POI should not be too large. Because of this, a square bounding-box of side 600 m (two times the assumed neighborhood distance of 300 m) is used to calculate the population living within the catchment of a POI. Similarly, for the parking area, the catchment corresponds to a square bounding-box of side 50 m. Here the distance threshold is selected to characterize short walking distance because parking far from the supermarket is discouraging for the customers [34]. We adopt the same threshold for all POI types, but in doing so, we ignore the effect of the POI type on the catchment distance, as it is treated uniformly for all POIs. To compute the average distance between a POI and transit stops, we identify the stops within a straight-line distance (“as the crow flies”) 400 m of a POI. This selection of straight-line distance threshold could even result in walking distances greater than 400 m in some cases because the actual route length may be longer depending on the street network. Commonly transit agencies use a walking distance of 400 m as a thumb rule for measuring neighborhood accessibility and transit accessibility, as reflected in previous studies in accessibility research [35, 36]. However, it is relevant to point out that the walking behavior is determined by a multitude of factors such as trip purpose, built-up environment, mode type, and population demographics [37, 38], and thus a detailed sensitivity analysis is beyond the scope of this paper. Average transit stop distance is the average distance of all the identified stops from a POI. Finally, weather features specific features such as temperature and precipitation can also be relevant for studying demand [39].

Non-spatial attributes are the POI type (e.g., supermarket or chemist); the number of reviews and ratings of a POI, e.g., the supermarket’s temporal demand pattern, could be different from that of a fast-food outlet shown in [29]. Further, if a POI has a high positive rating, it could imply its high likeability or customer satisfaction. Similarly, a large number of reviews by customers could be indicative of latent characteristics of a POI. These features such as rating and number of posted reviews are also used in modeling the demand trend modeling [30, 32]. It is pertinent to mention that other demographic factors, such as average income in the locality, could also play an essential role in the retail consumption [33], but the same is not considered in this study because such a dataset was not available.

## 2.1 Data sources

POI data is collected from Open Street Maps (OSM) via Overpass-Turbo [40]. Google’s Popular time graph data [41] is collected as a measure of the demand patterns at all the identified POIs. A popular time graph shows the busyness (workload or saturation) of a POI during the day, relative to the busiest time during the week [41]. Historical busyness is quantified on a relative scale of [0,100], where 100 indicates the busiest hour. This information is derived from the anonymized and aggregated data from the POI visitor’s location history [41]. As per Google, if the number of such users (who have opted for the location service) visiting a POI is not sufficient, then the popular time graph and the place’s live visit data may not be available [41]. For a given POI, a Popular time graph is averaged over the last few months [41], which could be referred to as “historical popular time”. Live visit data shows the popularity in real-time, which in some cases could be greater than 100 depending on its busyness or crowding relative to the past trends. The popular time data for a particular POI is publicly viewable on the Google Maps website [42]. Due to the smartphone-based passive data collection, Popular time data could also suffer from sample bias. As mentioned above, Popular time data is relative information and cannot infer the actual number of visitors without extrinsic information. Based on the above, we argue in this paper that live data could be an important indicator for measuring changes in the demand as, for each POI, they provide a measure of the deviation between the current and the average venue popularity.

Population data are obtained from the publicly available High-Resolution Population Density Maps provided by Facebook [43]. Facebook used state-of-the-art Computer vision techniques to process satellite imagery and generate this data. Population data provide human population distribution at 30-meter spatial resolution. Parking area (size and locations) and transit stop locations are obtained from the OSM data (obtained via Overpass [40]) and GTFS [44], respectively. Python library OSMNX [45] is used for the processing and analyzing OSM data.

## 2.2 Modeling approach

The study aims to examine the effect of the lockdown restrictions on the popularity of POIs. This is a problem of causal inference framework, where lockdown acts as a treatment variable. With pre-treatment and post-treatment data, a preferred modeling approach based on causal inference framework could be adopted by controlling for the treatment (lockdown) and confounding (day-of-the-week, POI type) variables. Herein we check the significance of covariates in explaining the day-specific popularity before and during the lockdown. Previous studies such as [32] and [30] have also modeled popula-

rity as a dependent variable in regression formulation. The dependent variable is the day-specific popularity of a POI, which is to be mapped to a set of explanatory variables, represented analytically as follows:

$$P_{i-d} = f(p_i, pa_i, sd_i, r_i, nr_i, type_i, L_d, D_d, T_d, Pr_d) + \epsilon_{i-d} \quad (1)$$

Where,

- $P_{i-d}$  is the response variable in terms of popularity of  $i^{th}$  POI on day  $d$
- $p_i$  is population within the defined catchment of  $i^{th}$  POI
- $pa_i$  is the total parking area within the defined catchment of  $i^{th}$  POI
- $sd_i$  is the average distance to the transit stops within the defined catchment of  $i^{th}$  POI
- $r_i$  is the rating of  $i^{th}$  POI
- $nr_i$  is total number of reviews of  $i^{th}$  POI
- $type_i$  is the dummy variable of  $i^{th}$  POI type namely, supermarkets, chemists and fast-food
- $L_d$  is the lockdown dummy variable, wherein during lockdown  $L_d = 1$ ; for historical data  $L_d = 0$
- $D_d$  is the dummy variable representing the day of the week e.g., Monday, Tuesday, and so on.
- $T_d$  and  $Pr_d$  are weather specific covariates for temperature and precipitation, respectively, on day  $d$
- $\epsilon_{i-d}$  is the residual term

POI type (*type*), lockdown (*L*) and day (*D*) are categorical variables, and are used as dummy variables after one-hot encoding e.g., for a supermarket POI, *supermarket* = 1, whereas *fast-food* and *chemist* are assigned 0 values. Similarly during the lockdown, *lockdown* = 1, else *lockdown* = 0; and on a Monday, only *monday* = 1, while other day-of-the-week dummy variables are equal to zero.

Linear regression models are simple and intuitive as they help to understand the average or global effects of the features. However, these models depend heavily on the explicit analytical formulation and thus could introduce model-bias. To counter this, we use regularized Gradient Boosting (GB) [46] for regression, inspired by previous studies [30]. GB, a machine learning technique, is based on training weak learners in an additive manner. Unlike linear models, GB models do not need an analytical specification and are also less sensitive to outliers. GB can work well with small data while avoiding overfitting. The use of regularized objective function in equation 2 helps to control overfitting. We refer to the regularized GB as Gradient Boosting Regression (GBR) model in this paper.

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t) \quad (2)$$

where  $\mathbf{x}_i$  denotes the  $i$ -th instance of the dataset of size  $n$ ;  $f_t$  is the current model fit;  $l$  is the loss function which measures the difference between the target ( $y_i$ ) and the prediction  $\hat{y}_i$ , at  $t$ -th iteration;  $\Omega(f)$  is the regularization term to check over-fitting. The details of the GBR are given in [47] and [46].

The best GBR model is selected based on the lowest Mean Squared Error (Eq. 3) on the training data (90% split), using 10-fold cross-validation. The main tunable parameters for the GBR model are the number of estimators and the tree depth [48]. To handle overfitting, we check the MSE on the test data (10%) to ensure that training and test errors are close.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

The interpretation of tree-based models, like the GBR model, is not straightforward since single coefficients (as in linear regression models) for attributes are not available. There are many tools for global interpretation, i.e., the average impact of the features on the model output. Recently, work has been done on the local explanations of these models to uncover the role of each feature for every model instance. The combined behavior of these local explanations can also infer the global behavior. In this regard, SHAP (SHapley Additive exPlanations) is a recently developed Python tool for explaining a machine learning model's outputs using the game-theoretic approach [49]. TreeExplainer method from SHAP, calculates classic Shapley values (a concept from the game theory [50]) and assigns importance or credit to the input features based on their role in the particular model prediction [51]. Similarly, local interaction effects are captured based on the Shapley interaction index from game theory by allocating the credit to a pair of features [49]. A novel advantage of TreeExplainer is that it can compute Shapley values for tree-based models in polynomial time [49], which makes them highly efficient for practical applications. For details on SHAP, we refer the reader to [49, 51].

We use Ordinary Least Squares (OLS) regression (linear regression), as a reference model for checking the consistency in the interpretation of the global effect of the features (Eq. 4). It can be seen that, in addition to the main effects, we also include interaction effects of lockdown ( $L_d$ ) with all the other variables. It is pointed out that in the linear model, the coefficient ( $\beta_8$ ) of *lockdown* gives the effect of lockdown on the *chemist* POIs, conditional on the other covariates. Thus, the coefficient ( $\beta_8$ ) actually represents interaction effect of *lockdown-chemist*. We do not include *chemist* dummy explicitly in Eq. 4, as it is highly negatively correlated with *supermarket*.

$$\begin{aligned}
 P_{i-d} = & \beta_0 + \beta_1 \cdot p_i + \beta_2 \cdot pa_i + \beta_3 \cdot sd_i + \beta_4 \cdot r_i \\
 & + \beta_5 \cdot nr_i + \beta_6 \cdot supermarket_i + \beta_7 \cdot fast-food_i \\
 & + \beta_8 \cdot L_d + \beta_9 \cdot Monday_d + \beta_{10} \cdot T_d + \beta_{11} \cdot Pr_d \\
 & + \beta_{12} \cdot L_d \cdot p_i + \beta_{13} \cdot L_d \cdot pa_i + \beta_{14} \cdot L_d \cdot sd_i \\
 & + \beta_{15} \cdot L_d \cdot r_i + \beta_{16} \cdot L_d \cdot nr_i \\
 & + \beta_{17} \cdot L_d \cdot supermarket_i + \beta_{18} \cdot L_d \cdot fast-food_i \\
 & + \beta_{19} \cdot L_d \cdot Monday_d + \beta_{20} \cdot L_d \cdot T_d \\
 & + \beta_{21} \cdot L_d \cdot Pr_d + \epsilon_{i-d}
 \end{aligned} \tag{4}$$

We also use the Robust regression or Robust Linear Model (RLM or M-Estimation) with Huber objective function [52]. This objective function uses two different formulations: least squares (in the center) and least absolute values (in the tails), basically underweighting the high influence observations or outliers in the dependent variable. Finally, it is noteworthy to refer to a recently published study that also uses both the GBR and linear regression for modeling and SHAP to explain building energy performance [53], due to inherent similarity in our modeling approaches. Modeling is done using Python programming language using the following libraries: statsmodels [54] and XGBoost [48].

### 3 Data collection and processing

We select Munich (the Free State of Bavaria’s capital city in Germany) as the study area. Even though many countries in the world are affected by COVID-19, the extent of the impact depends on multiple factors, such as first COVID-19 incidence [55], rate of spread, travel restrictions [56], testing and contact tracing & containment [57], amongst many others. Therefore, the data for before-during scenarios were collected based on the restriction or lockdown timeline. In Germany, the need for social distancing was announced on 12.03.2020, followed by the announcement of the temporary closure of schools on 14.03.2020 and non-essential travel ban on 18.03.2020 [58]. The Federal States took up state-specific measures depending on their needs, as the imposition of a full lockdown in Bavaria on 20.03.2020 [58]. Therefore, it can be concluded that

the second and third weeks of March were the transition period from pre-lockdown to the lockdown period. We are also interested in exploring how the demand pattern at a POI evolves during different stages of the lockdown (e.g., during the early lockdown in the third week of March viz-a-viz during the late lockdown in the last week of April).

3283 number of POIs were initially identified in the bounding box around Munich [*latitude* : 48.137585 ± 0.1125, *longitude* : 11.575444 ± 0.175]. The POI attributes, namely location (latitude and longitude), type, rating (on a scale of 1-5), and the number of reviews, are collected. For these POIs, we use python library [59] to obtain hourly historical data (Table 1). The live data is collected bi-hourly, e.g., 1200 H, 1400 H, 1600 H, etc. Not all of these POIs are found to have live popularity information during the lockdown, possibly due to the temporary closure of such POIs due to restrictions or insufficient users visiting such POIs. The availability of the live data varies per hour-day. Therefore, for the subsequent analysis and modeling, POIs without live data during 1200-1800 hours are dropped. The analysis period of 1200-1800 is chosen to represent the consistent working time for all the POIs, away from the opening (around 0800-1000 H) and closing hours (1900-2000 H). Only POI types with at least five samples are selected to ensure representativeness, which leaves a total of about 180 POIs for three categories (Fig. 2), namely supermarket, fast-food, and chemist (Table 2). The low number of POIs makes sense because several retail and leisure POIs, such as restaurants, stores, and barber shops, were closed and severely affected due to the lockdown restrictions, and that is why we suppose no popular time data were available for such POIs.

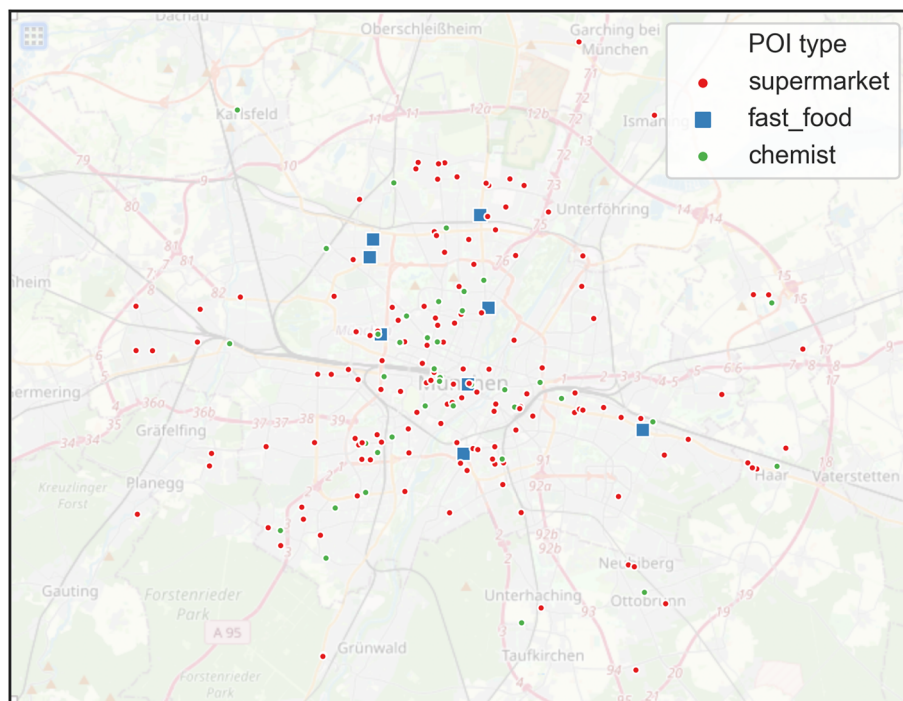
### 4 Data analysis

The hourly trends of average historical popularity in three types of POIs, namely supermarkets, chemists, and fast-foods, are shown in Fig. 3. In the historical trend, supermarkets show a prominent peak during the evening hours, coinciding with the evening commute. Chemists

**Table 1** Popular time data collection

Date	Cumulative COVID-19 Cases*	Type of data (Period)	Description
13-02-2020	—	Historical average (0000-2400)	Before lockdown
20.03.2020	878	Live (1200-1400)	Start of lockdown
03.04.2020	3304	Live (1200-1400)	Middle of lockdown
14.04.2020	4714	Live (0800-2000)	Late lockdown
22.04.2020	5332	Live (0800-2000)	Late lockdown
27.04.2020	5607	Live (0900-2100)	Late lockdown

\* source: [63]



**Fig. 2** Spatial distribution of POIs with Live data<sup>1</sup> Basemap source: Open Street Maps

also show a similar pattern. The trend is absent on Sunday, as most supermarkets and chemists are closed on Sundays in Munich. The fast-food category trend shows two prominent peaks during the weekdays, which can be attributed to busyness during lunch and late-evening hours. The demand trend on the weekend shows a high demand from lunch to late-evening hours.

Figure 3 shows the average live trends on two days of the week during the lockdown, 22nd April (Wednesday) and 27th April (Monday). Compared to the average historical popularity, the drop in the peak popularity and the general trend is evident. Interestingly, the drop in the fast-food category is more significant and characterizes lockdown’s adverse effect on similar POIs. It can also be recognized that the shape of the historical popularity trend is different on Monday and Wednesday for both supermarkets and chemists, indicating variations during the week. The trend of the afternoon (1400 H) popularity on few selected weekdays also shows the effect of lockdown on the three POI categories (Fig. 4). Again, supermarkets and chemists show similar trends with average historical popularity at around 40-50% of maximum popularity, but markedly increasing on 20th March, i.e., the day lockdown was announced. This increase (57% for supermarkets and 10% for chemists) could result from panic buying for groceries and health retail because of the uncertainty in the early

days of the lockdown and pandemic. During the later lockdown period in April, a gradual recovery of the popularity of the supermarkets and the chemists’ POIs is observed. The fast-food category trend is distinct by fall in its popularity, which did not wholly recover in April, although it shows small signs of recovery. It can also be seen that there is no panic buying in the fast-food category on the day of the lockdown announcement, unlike the other two types of POIs.

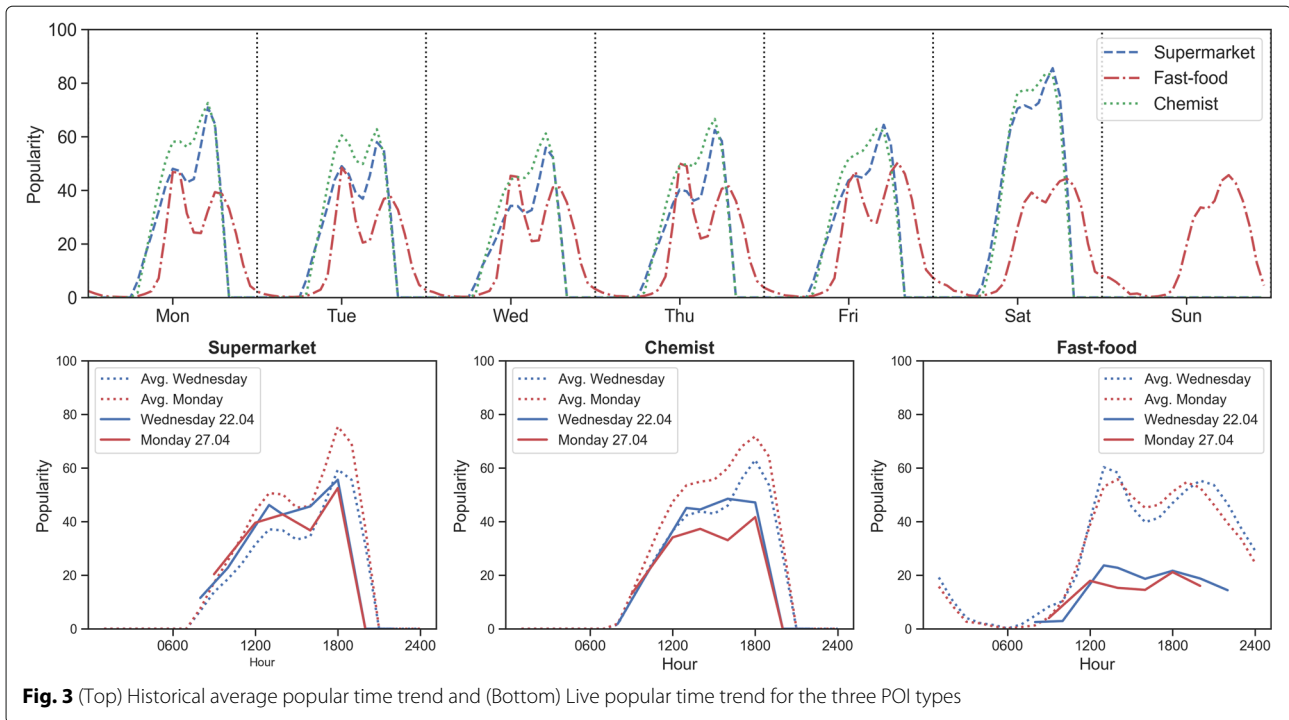
The summary of the explanatory variables is given in Table 3. The parking area locations in OSM correspond to a mix to different parking types such as surface parking, multi-level parking and underground parking. The composition of the parking areas in our sample is surface (70%), underground (3%), multi-storey (1%) and missing label (26%). We use the historical data, and live data (on 22.04.2020 and 27.04.2020) for modeling pre-lockdown and during lockdown scenarios, respectively (Table 1).

**Table 2** Number of identified POIs with historical data and live data

POI type	With historical data	With live data*
Supermarkets	262	137
Fast-food	170	8
Chemist	73	35

\*live data availability varies per day

<sup>1</sup>Fast-food POI’s symbol is enhanced for better visibility



**Fig. 3** (Top) Historical average popular time trend and (Bottom) Live popular time trend for the three POI types

The response variable in the regression models is the average of the popularity at two-hour interval over the period of 1200-1800 H, as follows:

$$P_{i-d} = (P_{i-d}^{1200} + P_{i-d}^{1400} + P_{i-d}^{1600} + P_{i-d}^{1800}) / 4 \tag{5}$$

where,  $P_{i-d}^t$  is the popularity at time  $t$

The features such as rating and number of reviews change with time as new users rate and review a specific POI. In our case, the change is found to be marginal, i.e., the mean percentage change in rating and reviews during the analysis period is found to be 0.2% and 0.0%, respectively. We do not control the weather-specific covariates due to the panel's limited dimension (two days of live data). The weather for these two days was similar as characterized by sunny or partial cloudy [60], which makes it reasonable to not control for weather-specific covariates. In the case of sufficient panel data, we recommend controlling for weather covariates for precise model estimation.

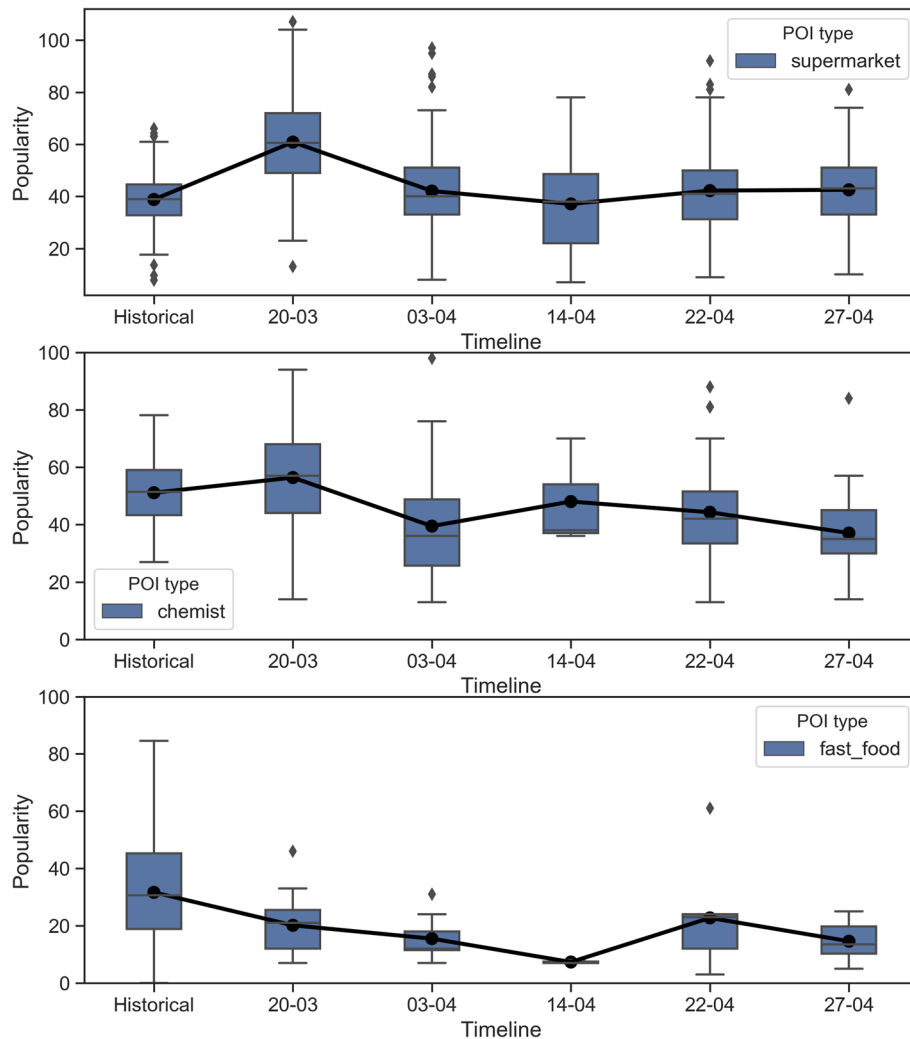
### 5 Results

Using cross-validation, we identify the best parameters for the GBR model (number of estimators: 20, maximum tree depth: 4). With these parameters, the model achieves an  $R^2$  of 0.63. The MSEs obtained on the training (7.4), and test data (9.6) are close, which implies no over-fitting. In the SHAP summary plot (Fig. 5), the feature impact on the output of the GBR model is shown with the distribution of SHAP values. In these plots, each point corresponds to one POI instance in the dataset

and corresponding SHAP values of the features. The color represents the feature value (blue for low value and pink for high value). The features in these sub-plots are ordered by the sum of the SHAP values' magnitude over the training dataset. If high SHAP values are observed for corresponding high values of the feature, it means an increase in that particular feature results in an increase in popularity and vice-versa. If SHAP values for a feature are concentrated near 0, that particular feature does not play much importance in predicting the popularity.

The main effects of *lockdown*, *monday* and *fast-food* are clear due to distinct distribution SHAP values for low and high feature values. The *lockdown* feature is found to be correlated with the drop in popularity. The popularity on Monday is found to be higher than that on Wednesday. It is interesting to note that the POI type plays an important role, especially for fast-foods. The *fast-food* attribute is found to be correlated with low SHAP values (i.e., *fast-food* = 1), which pushes the popularity to the lower side. The impact of the *population* and the number of reviews (*rating\_n\_x*) is not clearly correlated with the popularity value, as evident by overlapping pink and blue points. The low values of the *parking area* feature show low SHAP values, whereas high *parking area* is associated with high SHAP values (albeit with some overlap); i.e., it pushes the POI popularity to the higher side. Similarly, the type of POI, namely *chemist*, is correlated with the decrease in popularity as evident from negative SHAP values. Hence,





**Fig. 4** Live popular time trend at 1400 H on different days during the lockdown

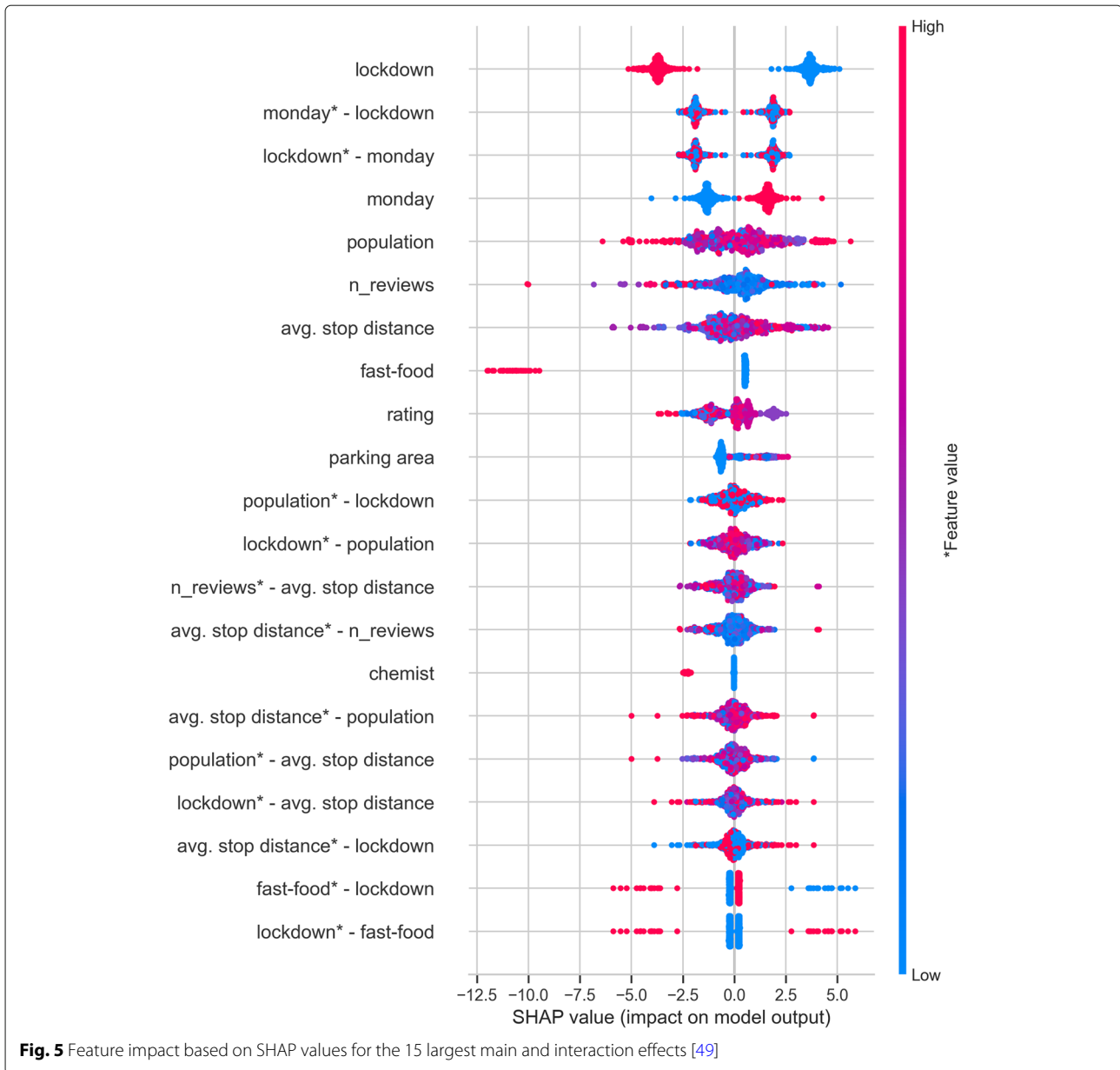
the features viz. *lockdown*, *day-of-the-week*, *POI-type*, and *Parking area* show a clear correlation with popularity.

Figure 5 also shows the interaction effects, where the superscript \* indicates which feature is represented by the color bar. The interaction effects of *lockdown* and *fast-food* features also show clear effects, implying the adverse

effect of lockdown on the fast-food POIs in terms of popularity, also seen in Fig. 3. Spatial factors, *population*, and *avg. stop distance* are found to have mixed effects (overlap of pink and blue points), and thus their global effects on popularity are not clear in Fig. 5. However, the interaction effect of *lockdown - avg. stop distance* (see feature

**Table 3** Summary of the explanatory variables

	Population (< 300 m)	Parking area m <sup>2</sup> (< 50 m)	Transit stops (< 400 m)	Avg. Stop distance (m)	Rating (1 – 5)	Reviews
minimum	195	0	1	105.2	2.5	8
mean	2614	414.3	8	324.8	3.9	371
maximum	4340	4503.7	28	794.0	4.9	4742
$\sigma$	727	850.1	5	87.4	0.3	551

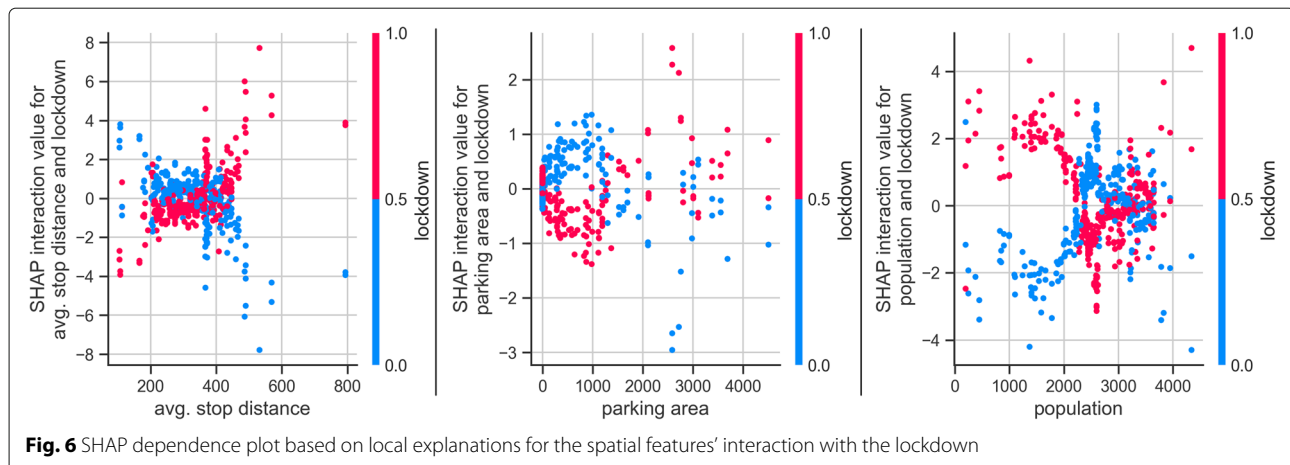


**Fig. 5** Feature impact based on SHAP values for the 15 largest main and interaction effects [49]

*avg. stop distance\* - lockdown*) shows high SHAP values for some longer stop distances, and vice-versa. This effectively means that POIs close to the transit stops had lower popularity than those farther from a stop during a lockdown. This is even more clear in the local explanation plot in Fig. 6, wherein the interaction effects of *lockdown - avg. stop distance* are inverted during the lockdown.

The fit of the OLS and RLM models is not as good as that of the GBR model, as evidenced from lower Adjusted  $R^2$  values (Table 4), which also justifies the use of the GBR model as it introduces less bias as compared to the linear models. Nevertheless, the results of the OLS model (the sign and magnitude of the coefficients) show that the

average behavior of the features is consistent with that of the GBR models. In both the OLS and RLM model, the intercept term is found to be significant, with a value close to 50. The main effect of *lockdown* is not found to be significant, unlike in Fig. 5. In the linear model (Eq. 4), it represents interaction of *lockdown-chemist*. The main effects of only *monday* and *average stop distance* are found to be significant. The positive coefficient of *monday* shows that the average historical popularity of POIs on Monday is more than that on a Wednesday due to variations in the daily demand patterns (Fig. 3). The negative coefficient of the *average stop distance* implies that popularity decreases with an increase in the distance to a transit stop. The



*lockdown* has significant interactions with other features. The popularity during lockdown depends on the type of POI, as the interaction of *fast-food* type POI has a greater negative coefficient than the other two types of POIs, whereas the popularity of the supermarkets is marginally greater than the chemist type POIs. During the lockdown, Monday's popularity is lower than on Wednesday, which provides evidence of the daily temporal variations in popularity during the lockdown. An interesting observation is that popularity is positively correlated with the *lockdown - avg. stop distance* interaction. A possible explanation is a drop in transit ridership during the lockdown, which can be seen in Fig. 1. Specifically, passenger ridership dropped by around 70% during April 2020 in Munich [61], and thus POIs closed to transit stops observed a greater reduction in popularity than others located far from the transit stops. The *lockdown-population* interaction also has a negative coefficient, albeit with a weak significance. One thing to note is that in the linear models, main and interaction coefficients of *parking area* are not found to be significantly correlated.

## 6 Discussion

There is a lack of existing research on the effects of COVID-19 on demand patterns at the POI level because this phenomenon is new and not experienced at the same scale in the last 100 years. Activity patterns uncovered in this study match the expectations of viz-a-viz restrictions during the COVID-19 lockdown in Munich. We explained the effect of features in the GBR regression model using SHAP. The behavior of coefficients is consistent with previous studies to some extent, wherein transit stop connectivity is associated with the demand at retail locations [33, 34]. The *population* is not found significant in our model, which could seem counter-intuitive. However, it should be kept in mind that the response variable (POI popularity) is a relative value instead of the absolute value

of demand. Significance of *POI type* (fast-food) during the COVID-19 confirms the dominance of POI type in explaining the lockdown impact, possibly as the lockdown was directed to reduce non-essential retail consumption and crowding. POI types are significant in explaining the dip in the POI's popularity, as *POI-type* captures latent consumer behavior. It is pointed out that the role of the spatial factor might vary depending on city-specific factors like the effectiveness of the lockdown and fall in transit ridership. The behavior (sign) of coefficients in GBR and linear models is similar for most of the features, with some exception such as *parking area*, which furthers the case for the use of advanced models with explainable tools such as SHAP.

The study is naturally not without its limitations. As stated above, relative popularity or demand fails to capture the population's effect around the POIs. Adding more features, like land-use type (residential vs. workplace), could improve the results, mainly because during COVID-19, generally, work from home was recommended. We also found that live popularity is not available for most of the POIs during the lockdown, limiting the data for modeling and adding to sampling bias. Sensitivity analysis on the effect of sampling variation and feature threshold could be an interesting topic for the future. We do not account for the marketing strategies, which influence consumers. The marketing decisions could be motivated by a complex set of factors such as weather, time, day, and month, and thus to some extent, the overall effects can be captured by collecting time-series data and controlling for an hour, month, and day. But at an individual POI level, marketing-specific data could be hard to collect as the marketing strategies could be diverse and highly dynamic even across similar POI types. Further research should be done on data to infer latent features such as consumer preferences and socializing behavior during disruptive events.

**Table 4** Results of Linear Regression

	<i>Dependent variable: <math>P_{i-d}</math></i>	
	1:OLS	2: RLM
Intercept	48.74*** (6.93)	53.37*** (7.04)
fast-food	-2.56 (3.75)	-3.57 (3.81)
lockdown	3.66 (9.82)	-3.50 (9.98)
lockdown:fast-food	-19.47*** (5.31)	-16.73*** (5.39)
lockdown:monday	-16.45*** (1.52)	-16.24*** (1.55)
lockdown:average stop distance/1000	19.49** (8.93)	20.19** (9.07)
lockdown:number of reviews/1000	0.13 (1.88)	-0.97 (1.91)
lockdown:parking area/1000	0.18 (0.95)	0.13 (0.96)
lockdown:population/1000	-1.93* (1.10)	-1.93* (1.12)
lockdown:rating	-1.68 (2.26)	-0.10 (2.30)
lockdown:supermarket	4.53** (2.02)	4.84** (2.05)
monday	11.20*** (1.08)	11.29*** (1.09)
average stop distance/1000	-10.52* (6.32)	-11.40* (6.42)
number of reviews/1000	-1.52 (1.33)	-1.41 (1.35)
parking area/1000	0.67 (0.67)	0.64 (0.68)
population/1000	1.18 (0.78)	1.20 (0.79)
rating	-0.40 (1.60)	-1.41 (1.62)
supermarket	-1.96 (1.43)	-2.12 (1.45)
Observations	718	718
$R^2$	0.34	
Adjusted $R^2$	0.32	
Residual Std. Error	10.18	3.21
F Statistic	21.20***	

Note: \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

## 7 Conclusions

This research uses a data-driven approach to analyze the activity and demand patterns at the POIs in Munich. We show that POI check-ins are a potential source of information during dynamic events like COVID-19. The use of POI level data and features helps to understand the underlying interactions of spatial and non-spatial features in detail and identify the spatial variability (if any) and the influencing factors thereof. The results are also of interest to the transport planners and operators as they provide insights on the effect of transport variables such as parking area and transit-stop distance on the POI popularity. We provide empirical evidence of the disproportionate effect of the lockdown restrictions on the POIs in Munich, depending on their distance from a transit stop. Businesses near or in the transit hubs are more vulnerable to these disruptions due to reduced commuters, potential customers. This outcome could be due to reduced travel (home-office) or changed travel behavior (customers avoiding public transport). These insights point to the lack of resilience of transit-near POIs due to excessive dependence on commuting customers. Policy-makers can look into or even adapt the transit-oriented development principles to diversify the customers of near-transit POIs. It is again highlighted that this study's findings might not hold for other cities due to the presence of different city-specific factors.

The use of publicly available data sources increases the transferability of this study to other study areas. A time-series crowdsensed data over a longer duration is suitable for causal inference, to conduct the policy impact evaluation of lockdowns, to evaluate the crowding or busyness, and their correlation with the spread of the pandemic. This data could help measure crowding patterns at the POIs, especially when there is an increased need to reduce mobility and contacts. Such patterns could also be correlated with other factors such as public transport schedules, weather, number of infections, and tests to check their influence on crowding behavior. Post-lockdown, the researchers could use the crowdsensed information to analyze if the POI visitation trend has stabilized and returned to normal levels. After all, a pandemic is a kind of disruptive event, and thus similar data could be the potential to study other planned, unplanned or disruptive events. Such insights could help cities to monitor demand patterns and devise effective responses to such events.

### Abbreviations

COVID-19: COrona Virus disease 2019; GBR: Gradient boosting regression; LBSNs: Location-based social networks; OLS: Ordinary least squares; POI: Point of interest; RLM: Robust linear model; RMSE: Root mean squared error; SHAP: SHapley Additive exPlanations

### Acknowledgements

We want to thank the two anonymous reviewers for their valuable comments, which helped to improve the manuscript. This work was supported by the

Technical University of Munich within the Open Access Publishing Funding Programme.

#### Authors' contributions

All authors were involved during the study conception. VM and GC analyzed the data and developed models. All authors interpreted the results. VM wrote the initial manuscript, which was edited and reviewed by GC and CA. All authors read and approved the final manuscript.

#### Funding

This research was supported by the German Research Foundation DFG (TRAMPA Project, Grant 415208373), and European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754462. Open Access funding enabled and organized by Projekt DEAL.

#### Availability of data and materials

The popular time data analyzed in this study are not publicly available due to data restrictions. Other datasets analyzed during the current study are publicly available at [43, 44] and [62].

#### Declarations

#### Competing interests

The authors declare that they have no competing interests.

Received: 30 July 2020 Accepted: 17 March 2021

Published online: 12 April 2021

#### References

- Acheampong, R.A., & Silva, E. (2015). Land use–transport interaction modeling: A review of the literature and future research directions. *Journal of Transport and Land Use*, 8(3). <https://doi.org/10.5198/jtlu.2015.806>.
- Hansen, W.G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2), 73–76. <https://doi.org/10.1080/01944365908978307>.
- Geurs, K.T., & van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, 12(2), 127–140. <https://doi.org/10.1016/j.jtrangeo.2003.10.005>.
- Dunn, W. (2007). *Managing Travel for Planned Special Events Handbook*. New York: US Department of Transportation. [https://ops.fhwa.dot.gov/eto\\_tim\\_pse/preparedness/pse/handbook.htm](https://ops.fhwa.dot.gov/eto_tim_pse/preparedness/pse/handbook.htm).
- Sabir, M. (2010). *Impact of Weather on Daily Travel Demand*. Amsterdam: VU University, Department of Spatial Economics. Monograph. <https://edepot.wur.nl/306452>.
- Cantelmo, G., Kucharski, R., Antoniou, C. (2020). Low-dimensional model for bike-sharing demand forecasting that explicitly accounts for weather data. *Transportation Research Record*, 2674(8), 132–144. <https://doi.org/10.1177/0361198120932160>.
- Yabe, T., Sekimoto, Y., Tsubouchi, K., Ikemoto, S. (2019). Cross-comparative analysis of evacuation behavior after earthquakes using mobile phone data. *PLOS ONE*, 14(2), 1–12. <https://doi.org/10.1371/journal.pone.0211375>.
- Gray, C.L., & Mueller, V. (2012). Natural disasters and population mobility in bangladesh. *Proceedings of the National Academy of Sciences of the United States of America*, 109(16), 6000–6005. <https://doi.org/10.1073/pnas.1115944109>. 22474361 [pmid].
- Yamamura, E., Tsutsui, Y., Yamane, C., Yamane, S. (2014). Effect of major disasters on geographical mobility intentions: The case of the fukushima nuclear accident. ISER Discussion Paper 903, Osaka. <http://hdl.handle.net/10419/127069>. Accessed on 20 July 2020.
- Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S., Colizza, V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the covid-19 epidemic in france under lockdown: a population-based study. *The Lancet Digital Health (XXXX)*. [https://doi.org/10.1016/S2589-7500\(20\)30243-0](https://doi.org/10.1016/S2589-7500(20)30243-0).
- Brinkman, J., & Mangum, K. (2020). Travel behavior and the coronavirus outbreak. *Economic Insights*, 5(3), 23–26.
- Arafat, S.M.Y., Kar, S.K., Kabir, R. (2020). Possible controlling measures of panic buying during covid-19. *International Journal of Mental Health and Addiction*, 1–3. <https://doi.org/10.1007/s11469-020-00320-1>. PMC7241580 [pmcid].
- Nicola, M., Alsaifi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, 78, 185–193. <https://doi.org/10.1016/j.ijsu.2020.04.018>.
- Antoniou, C., Balakrishna, R., Koutsopoulos, H.N. (2011). A synthesis of emerging data collection technologies and their impact on traffic management applications. *European Transport Research Review*, 3(3), 139–148.
- Rashidi, T., Abbasi, A., Maghrebi, M., Hasan, S., Waller, T. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, 75, 197–211. <https://doi.org/10.1016/j.trc.2016.12.008>.
- Järv, O., Ahas, R., Witlox, F. (2014). Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C: Emerging Technologies*, 38, 122–135. <https://doi.org/10.1016/j.trc.2013.11.003>.
- Community Mobility Reports. <https://www.google.com/covid19/mobility>. Accessed on 25 July 2020.
- Mobility Trends Reports. <https://www.apple.com/covid19/mobility>. Accessed on 25 July 2020.
- Grantz, K.H., Meredith, H.R., Cummings, D.A.T., Metcalf, C.J.E., Grenfell, B.T., Giles, J.R., Mehta, S., Solomon, S., Labrique, A., Kishore, N., Buckee, C.O., Wesolowski, A. (2020). The use of mobile phone data to inform analysis of covid-19 pandemic epidemiology. *Nature Communications*, 11(1), 4961. <https://doi.org/10.1038/s41467-020-18190-5>.
- Badr, H.S., Du, H., Marshall, M., Dong, E., Squire, M.M., Gardner, L.M. (2020). Association between mobility patterns and covid-19 transmission in the usa: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(11), 1247–1254. [https://doi.org/10.1016/S1473-3099\(20\)30553-3](https://doi.org/10.1016/S1473-3099(20)30553-3).
- Fang, H., Wang, L., Yang, Y. (2020). Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china. *Journal of Public Economics*, 191, 104272. <https://doi.org/10.1016/j.jpubeco.2020.104272>.
- Roy, A., & Kar, B. (2020). Characterizing the spread of covid-19 from human mobility patterns and sociodemographic indicators, In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Advances in Resilient and Intelligent Cities. ARIC '20*. <https://doi.org/10.1145/3423455.3430303> (pp. 39–48). New York: Association for Computing Machinery.
- Liu, J., Shen, H., Zhang, X. (2016). A survey of mobile crowdsensing techniques: A critical component for the internet of things, In *2016 25th International Conference on Computer Communication and Networks (ICCCN)*. <https://doi.org/10.1109/ICCCN.2016.7568484> (pp. 1–6).
- Efthymiou, D., & Antoniou, C. (2012). Use of social media for transport data collection. *Procedia - Social and Behavioral Sciences*, 48, 775–785. <https://doi.org/10.1016/j.sbspro.2012.06.1055>. Transport Research Arena 2012.
- Chaniotakis, E., Antoniou, C., Pereira, F. (2016). Mapping social media for transportation studies. *IEEE Intelligent Systems*, 31(6), 64–70.
- Bagrow, J.P., Wang, D., Barabási, A.-L. (2011). Collective response of human populations to large-scale emergencies. *PLOS ONE*, 6(3), 1–8. <https://doi.org/10.1371/journal.pone.0017680>.
- Chaniotakis, E., Antoniou, C., Pereira, F.C. (2017). Enhancing resilience to disasters using social media, In *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. <https://doi.org/10.1109/MTITS.2017.8005602> (pp. 699–703).
- D'Silva, K., Noulas, A., Musolesi, M., Mascolo, C., Sklar, M. (2018). Predicting the temporal activity patterns of new venues. *EPJ Data Science*, 7(1), 13. <https://doi.org/10.1140/epjds/s13688-018-0142-z>.
- Capponi, A., Vitello, P., Fiandrino, C., Cantelmo, G., Kliazovich, D., Sorger, U., Bouvry, P. (2019). Crowdsensed data learning-driven prediction of local businesses attractiveness in smart cities, In *2019 IEEE Symposium on Computers and Communications (ISCC)*. <https://doi.org/10.1109/ISCC47284.2019.8969771> (pp. 1–6).
- Timokhina, S., Sadrani, M., Antoniou, C. (2020). Predicting venue popularity using crowd-sourced and passive sensor data. *Smart Cities*, 3(3), 818–841. <https://doi.org/10.3390/smartcities3030042>.
- MacKenzie, D., & Cho, H. (2020). Travel demand and emissions from driving dogs to dog parks. *Transportation Research Record*, 2674(6), 291–296. <https://doi.org/10.1177/0361198120918870>.
- Möhring, M., Keller, B., Schmidt, R., Dacko, S. (2020). Google popular times: towards a better understanding of tourist customer patronage behavior. *Tourism Review (ahead-of-print)*. <https://doi.org/10.1108/TR-10-2018-0152>.

33. Rolph, I.K. (1932). The population pattern in relation to retail buying: As exemplified in Baltimore. *American Journal of Sociology*, 38(3), 368–376.
34. van der Waerden, P., Borgers, A., Timmermans, H. (1998). The impact of the parking situation in shopping centres on store choice behaviour. *GeoJournal*, 45(4), 309–315. <https://doi.org/10.1023/A:1006987900394>.
35. Aultman-Hall, L., Roorda, M., Baetz, B.W. (1997). Using GIS for evaluation of neighborhood pedestrian accessibility. *Journal of Urban Planning and Development*, 123(1), 10–17. [https://doi.org/10.1061/\(ASCE\)0733-9488\(1997\)123:1\(10\)](https://doi.org/10.1061/(ASCE)0733-9488(1997)123:1(10)).
36. Achuthan, K., Titheridge, H., Mackett, R.L. (2010). Mapping accessibility differences for the whole journey and for socially excluded groups of people. *Journal of Maps*, 6(1), 220–229. <https://doi.org/10.4113/jom.2010.1077>.
37. Daniels, R., & Mulley, C. (2013). Explaining walking distance to public transport: The dominance of public transport supply. *Journal of Transport and Land Use*, 6(2), 5–20. <https://doi.org/10.5198/jtlu.v6i2.308>.
38. Islam, S.R., Markus, M., Kumar, S.S. (2019). Walking to a public transport station: Empirical evidence on willingness and acceptance in Munich, Germany. *Smart and Sustainable Built Environment*, 9(1), 38–53. <https://doi.org/10.1108/SASBE-07-2017-0031>.
39. Horanont, T., Phithakitnukoon, S., Leong, T.W., Sekimoto, Y., Shibasaki, R. (2013). Weather effects on the patterns of people's everyday activities: A study using GPS traces of mobile phone users. *PLOS ONE*, 8(12), 1–14. <https://doi.org/10.1371/journal.pone.0081153>.
40. Raifer, M. Overpass-turbo. <http://overpass-turbo.eu/>. Accessed on 20 July 2020.
41. Popular Times, Wait Times, and Visit Duration. <https://support.google.com/business/answer/6263531?hl=en>. Accessed on 20 July 2020.
42. Google Maps. Google Maps. <https://www.google.com/maps/>. Accessed on 20 July 2020.
43. Facebook. Germany: High Resolution Population Density Maps + Demographic Estimates. [https://data.humdata.org/organization/facebook?q=germany&ext\\_page\\_size=25](https://data.humdata.org/organization/facebook?q=germany&ext_page_size=25). Accessed on 20 July 2020.
44. GTFS.DE. GTFS for Germany. <https://gtfs.de/de/feeds/>. Accessed on 20 July 2020.
45. Boeing, G. (2017). Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65, 126–139. <https://doi.org/10.1016/j.compenvurbsys.2017.05.004>.
46. Chen, T., & Guestrin, C. (2016). Xgboost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/2939672.2939785>.
47. Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.
48. XGBoost. Introduction to Boosted Trees. <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>. Accessed on 20 July 2020.
49. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56–67. <https://doi.org/10.1038/s42256-019-0138-9>.
50. Shapley, L.S. (1988) In A.E.E. Roth (Ed.), *A value for n-person games*, (pp. 31–40): Cambridge University Press. <https://doi.org/10.1017/CBO9780511528446.003>.
51. Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf> (pp. 4765–4774): Curran Associates, Inc.
52. Huber, P.J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.*, 1(5), 799–821. <https://doi.org/10.1214/aos/1176342503>.
53. Arjunan, P., Poolla, K., Miller, C. (2020). EnergyStar++: Towards more accurate and explanatory building energy benchmarking. *Applied Energy*, 276, 115413. <https://doi.org/10.1016/j.apenergy.2020.115413>.
54. statsmodels. Linear Regression. <https://www.statsmodels.org/stable/regression.html>. Accessed on 20 July 2020.
55. Böhmer, M.M., Buchholz, U., Corman, V.M., Hoch, M., Katz, K., Marosevic, D.V., Böhm, S., Woudenberg, T., Ackermann, N., Konrad, R., Eberle, U., Treis, B., Dangel, A., Bengs, K., Fingerle, V., Berger, A., Hörmansdorfer, S., Ippisch, S., Wicklein, B., Grahl, A., Pörtner, K., Müller, N., Zeitlmann, N., Boender, T.S., Cai, W., Reich, A., an der Heiden, M., Rexroth, U., Hamouda, O., Schneider, J., Veith, T., Mühlemann, B., Wölfel, R., Antwerpen, M., Walter, M., Protzer, U., Liebl, B., Haas, W., Sing, A., Drosten, C., Zapf, A. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *The Lancet Infectious Diseases*. [https://doi.org/10.1016/S1473-3099\(20\)30314-5](https://doi.org/10.1016/S1473-3099(20)30314-5).
56. Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M.E., Longini, I.M., Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489), 395–400. <https://doi.org/10.1126/science.aba9757>.
57. Lorch, L., Trouleau, W., Tsirtsis, S., Szanto, A., Schölkopf, B., Gomez-Rodriguez, M. (2020). A Spatiotemporal Epidemic Model to Quantify the Effects of Contact Tracing, Testing, and Containment. [2004.07641](https://arxiv.org/abs/2004.07641).
58. Robert Koch Institute & Humboldt University of Berlin, n.A. COVID-19 Mobility Project. <https://www.covid-19-mobility.org/reports/first-report-general-mobility/>. Accessed on 20 July 2020.
59. m-wrzt, riedmaph (2018). populartimes. *GitHub*. Accessed on 20 July 2020.
60. Time and Date: April 2020 Weather in Munich. <https://www.timeanddate.com/weather/germany/munich/historic?month=4&year=2020>. Accessed on 20 Nov 2020.
61. Münchner Verkehrsgesellschaft mbH (MVG). Together Against Corona: Passenger Traffic. <https://www.mvg.de/services/aktuelles/coronavirus.html>. Accessed on 15 Jan 2021.
62. OSM. Open Street Maps. <https://www.openstreetmap.org/>. Accessed on 20 July 2020.
63. StadtMuenchen. <https://twitter.com/StadtMuenchen>. Accessed on 20 Nov 2020.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)