Modeling of the daily dynamics in bike rental system using weather and calendar conditions: A semi-parametric approach

Christopher Odoom, Alexander Boateng, Sarah Fobi Mensah, Daniel Maposa

Scientific African 24 (2024) e02211

Contents lists available at ScienceDirect

Scientific African

journal homepage: www.elsevier.com/locate/sciaf

Modeling of the daily dynamics in bike rental system using weather and calendar conditions: A semi-parametric approach

Christopher Odoom [a], Alexander Boateng [b], Sarah Fobi Mensah [c], Daniel Maposa [d,*]

[a] *Department of Mathematics and Statistics, University of Massachusetts, Amherst, USA*
[b] *Department of Mathematics and Computer Science, Modern College of Business and Science, Muscat, Sultanate of Oman*
[c] *Department of Mathematical Sciences, Montana State University, Bozeman, USA*
[d] *Department of Statistics and Operations Research, South Africa*

ARTICLE INFO

Editor: DR B Gyampoh

Keywords:
Bike rentals
Semi-parametric
Casual
Registered
Washington D.C

ABSTRACT

This study proposes a more robust methodological approach to modeling the effect of weather and calendar variables on the number of bike rentals. We employ penalized splines quasi-Poisson regression (a semi-parametric model), which involves some form of regularization, like those used in lasso, ridge, and other types of parametric regularization models. We demonstrate that this modeling approach reveals hidden relationships that a pure parametric model fails to identify.

The findings show that visibility, windspeed, season, working day, and year all significantly impact bike rentals. Increased rentals are associated with increased visibility and lower wind speed. Rentals are negatively affected by the spring and winter seasons, while working days and the year show positive trends except in a few cases. The analysis of rentals by registered and casual users reveals similar patterns, though the magnitudes of the effects differ. These findings highlight the importance of considering weather and calendar variables when managing and promoting bike-sharing services. The study has implications for bike-sharing system operators and policymakers, suggesting strategies such as improving visibility and wind protection, seasonally tailoring promotional campaigns, targeting non-working days for casual users, and adapting to changing user demands. The study adds to our understanding of the factors that influence bike rentals and provides suggestions for improving the utilization and accessibility of bike-sharing systems.

Introduction

Bike rental systems have gained significant popularity in urban areas as a sustainable and convenient mode of transportation [1,2]. Understanding the daily dynamics of bike rentals is crucial for optimizing resource allocation and meeting demand effectively. Weather conditions and calendar events play a pivotal role among the factors influencing these dynamics. Weather conditions, including temperature, precipitation, wind speed, and humidity, directly impact bike rental decisions [3–6]. Favourable weather, such as mild temperatures and clear skies, tends to encourage bike rentals, while adverse weather conditions can deter potential riders [7]. Calendar events, such as public holidays, weekends, and festivals, also significantly influence bike rental patterns [8].

Abbreviations: GAM, generalized additive model; GAMs, generalized additive models; GLM, generalized linear model; P-IRS, penalized iterative re-weighted least squares; EDF, effective degrees of freedom; GCV, generalized cross-validation.

* Corresponding author.
E-mail address: daniel.maposa@ul.ac.za (D. Maposa).

https://doi.org/10.1016/j.sciaf.2024.e02211
Received 9 September 2023; Received in revised form 21 March 2024; Accepted 5 April 2024
Available online 6 April 2024

While previous studies have investigated the impact of weather and calendar events on bike rental demand [2–4,9–11], research gaps still need to be addressed. Most importantly, many studies have employed a parametric model, neglecting the advantages of combining parametric and non-parametric approaches. Also, most studies have focused on specific cities or regions, limiting the generalizability of their findings. Additionally, there is a need for more comprehensive analyses that consider a wider range of weather variables and calendar events to capture their subtle effects on bike rental patterns for registered and casual membership types.

This study aims to bridge these research gaps by employing a generalized additive model (GAM) approach that combines the flexibility of non-parametric models with the interpretability of parametric models. Generalized additive models (GAMs) represent a semi-parametric version of generalized linear models within the parametric model framework. The field of GAMs has witnessed significant advancements. [12] discuss in their paper how GAM models are fitted using scatter plot smoothers. [13] also discuss the use of GAMs in ecological modeling, providing insights into how related statistics are employed to select predictors and perform model diagnostics and evaluation. The application of GAMs in environmental and meteorological data is studied by [14], demonstrating that this approach works well for non-normal data.

In the context of predicting electricity grid load, where the suitability of generalized additive models is evident, the practical application is hindered by the substantial size of the dataset, rendering their utilization challenging with existing methods. [15] introduces practical fitting methods for generalized additive models designed for large datasets, specifically addressing cases where smooth terms in the model are represented through penalized regression splines.

The findings will assist us in identifying the conditions necessary for establishing a successful bicycle rental system in other places and provide a more appropriate approach to modeling bike rental data.

The paper's outline is as follows: literature review and materials and methods are presented in Sections 2 and 3, respectively. The results and discussion are presented in Sections 4 and 5, respectively, while Section 6 presents the conclusion and implications of the results for further research.

Literature review

In recent years, there has been a lot of interest in analyzing daily dynamics in bike rental systems utilizing parametric models. Weather conditions are one of the most important variables influencing bike rental demand. Feng did a time series study in Chicago and discovered that meteorological conditions, such as temperature, precipitation, and wind speed, substantially impacted daily bike-sharing demand. Similarly, [10] conducted a systematic review, underlining the relevance of sunny skies and pleasant temperatures in encouraging bike rentals. [16] expanded on this research by investigating the function of smart bike-sharing systems in urban transportation, highlighting the potential of technological developments to improve bike rental services.

Calendar events and weather factors have been identified as significant predictors of bike rental trends. [11] conducted a case study and found that calendar events like public holidays and festivals substantially impacted bike-sharing system demand. These findings agreed with those of Weng et al., who evaluated the impact of calendar events on bike-sharing demand in Beijing. [17] investigated the potential of bicycle transportation to boost urban tourism in Warsaw.

Furthermore, some research has looked into many aspects of bike-sharing systems. [10] examined the elements influencing bike-sharing demand, revealing consumer preferences and system design considerations. [18] concentrated on the efficient inventory management of bike-sharing stations to improve such systems' operational efficiency. Macioszek et al. conducted a case study on a bike-sharing system in Poland, emphasizing the importance of such systems in promoting sustainable mobility. Modeling methodologies were used by [19,20] to study bike-sharing system demand and the factors that influence it. [21] investigated renting behavior using linear regression.

While earlier research, like the ones described in this paper, has shed light on the impact of weather and calendar events on bike rental patterns, research gaps remain. The potential advantages of merging parametric and non-parametric models have also gone untapped.

The current work intends to solve these research gaps by employing a generalized additive models (GAMs) technique to analyze the daily dynamics of bike rental systems. Generalized additive models (GAMs) have become an elegant and practical option in model building, and so much research has been conducted to investigate its usage in many fields, extensions of the area, and related areas. A generalized additive model (GAM) fits a response variable Y by sum of smooth functions of the explanatory variables, X_i for $i = 1, \ldots, q$ [12]. In practice Estimation of the smooth GAM component usually makes use of penalized iterative re-weighted least squares (P-IRLS) algorithms ([22], p. 145–217). The history of GAMs dates back to [23]; Hastie et al. who introduced the concept to overcome the curse of dimensionality, which is a problem in non-parametric regression. It is that the precision of the estimates obtained from a non-parametric regression is inversely related to the number of explicate variables included in the model [24,25] proposed a method for direct generalized additive modeling with penalized likelihood. In their study, they emphasized the importance of incorporating penalized likelihood in the modeling process ([25], p. 193–209). [22] provides a comprehensive introduction to generalized additive models using the R programming language. Several authors have some rich literature on GAMs, curious readers can visit [23,26]. This approach allows us to capture all non-linear relationship between number of bike rentals and the calendar and weather variables. It also allows us to incorporate a bivariate function which help us model the effect of location on the number of bike rentals. The findings will have consequences for bike rental system operators, urban planners, and policymakers in implementing effective resource planning, demand forecasting, and sustainable mobility initiatives.

Materials and methods

This section first describes Washington, DC, the area under study in this paper and then describes the dataset used. The main goal of this study was to understand the relationship between calendar and weather conditions on the number of bicycle rentals at the station level and compare the effect of these predictors on the number of rentals by membership type (registered and casual). This section also presents the proposed method as a robust technique to model "count" data instead of the traditional parametric count data models. In addition, we show the model (semi-parametric Quasi-Poisson regression) we are building to achieve the goal of this study.

The geographical area covered by the study

Washington, D.C., was founded in 1790 and is the capital of the United States of America. This city houses the president of the United States and other prominent members of the Country. According to the U.S. Census Bureau, it has a land area of 61.13 m^2iles and a population size of 689,545 as of 2020. Many nations and cities are trying to fight climate change, as is Washington, DC. Many of these climate change problems are caused by excessive carbon released into the atmosphere, primarily because of burning fuel. The state's most available transport systems, including trains, buses, domestic aeroplanes, and individual vehicles, are known to contribute to climate change.

To reduce traffic congestion which may lead to more carbon growth, most cities in the world have increased the use of bicycles for short distances. Washington, DC, also shares the same faith as other cities. Fig. 1 shows the numerous bicycle rental stations in Washington, DC. Capital Bikeshare system stations are shown in Fig. 1 with the blue location symbol, and this system has the potential to help save energy, promote good climate conditions, and promote good health in the city's people.

Dataset

Capital Bikeshare data

Capital Bikeshare is Metro DC's bike-share system, with over 5000 bikes available at 650 regional stations. The system, which opened in September 2010, offers residents and visitors a convenient, enjoyable, and cost-effective mode of transportation. You can rent a bike from one of the hundreds of stations located throughout the metro D.C. area. Capital Bikeshare is a District-owned transportation system with thousands of bicycles and over 300 stations in Washington, DC, and hundreds more in Arlington, Alexandria, Fairfax, Falls Church, Prince George's County, and Montgomery County. According to the Executive Office of the Mayor, all D. C. residents will be eligible for free Capital Bikeshare memberships beginning in October 2021. Users can either register (registered customers) to rent the bike or use a bike when they want to without registering as members (casual customers). The services are available 365 days, all time, and offer "registered" membership (Annual Member, 30-Day Member, or Day Key Member) and a "casual" use (Single Trip, 24-Hour Pass, 3-Day Pass, or 5-Day Pass) [9].

Fig. 3 shows how the membership type number of rentals compares from a distributional point of view. The distribution of the registered looks more like the combined distribution, which may suggest more rental by registered users than casual users. There is a

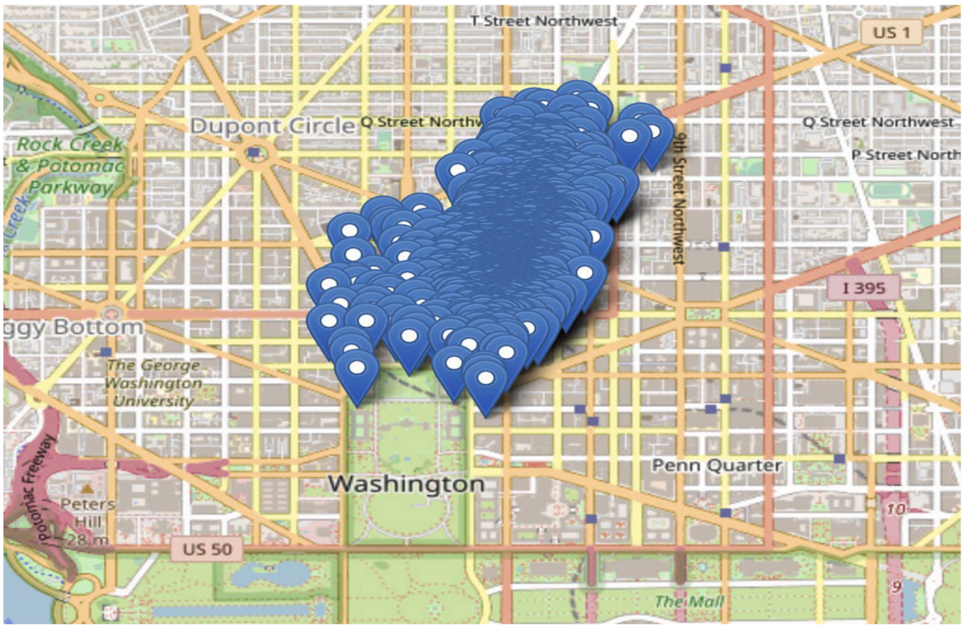

Fig. 1. Map of Bike rental Stations in Washington DC.

mobile App where people can access bike availability at various stations. Capital bike-share system started operating in 2010, and they have all the data on the rentals from 2010 to date. [9] used the 2011 and 2012 rental datasets in his study to understand the Meteorological barriers to bike rental demands. However, our research uses a more recent capital bike-share dataset, allowing us to capture the most recent changes in the bike rental system. In addition, [9] used only two years, but our study used four years from 2020 to 2023. Fig. 2 shows the dynamics in the total bike rentals in the region for variation seasons and years. The number of rentals is higher in fall and summer for 2022 and lower in winter for all years. The highest for this period was recorded in summer 2022, with about 19, 000 daily rentals.

Calendar data

We accessed the database of all public holidays from April 2020 to May 2023 and created a new calendar variable called working day, an indicator. It assumes 0 if the day is a weekend or public holiday and 1 otherwise. This variable allows us to understand bike transportation's role in developing a city. Fig. 4 shows this new variable's dynamics in the number of rentals by membership type. It is clear from Fig. 4 that registered members have the highest average rentals on a working day, which may suggest that this membership type is more for people who go to work on a bike or go through the work hours using a bike as opposed to the casual members who used bike more often on holidays and weekend probably for workout purposes or take fresh air. This study hopes to see this effect captured in a model that can be used to predict daily rental numbers. We hope to see if people ride bikes more often during work or when they are off work. This dataset was obtained from the Mayor of Washington, DC, executive office.

Weather data

The *timeanddate* website provides data on historical weather conditions, including daily average temperature, humidity, visibility, etc. We extracted the ones of relevance to this study, as shown in Table 1. Table 2 shows the summary statistics of these weather variables; the minimum daily average temperature recorded in the entire period is 14.86 F, and the highest is 88.71 F. The most humid day recorded average humidity of 97.34%, and the lowest was 24.12%.

The core data set for this study is related to the four-year historical data corresponding to years 2020, 2021, 2022 and 2023 from Capital Bikeshare system, Washington D.C., USA, which is publicly available on their website [27]. We aggregated the data daily and then extracted and added the corresponding weather and calendar information. Weather information is extracted from (timeanddate 2023). Table 1 contains the variables obtained from the Capital bike share system website and the weather and calendar variables.

Modeling the number of bike rental

This study is focused on modeling the number of bike rentals; in so doing, we can quantify the effect of weather dynamics and

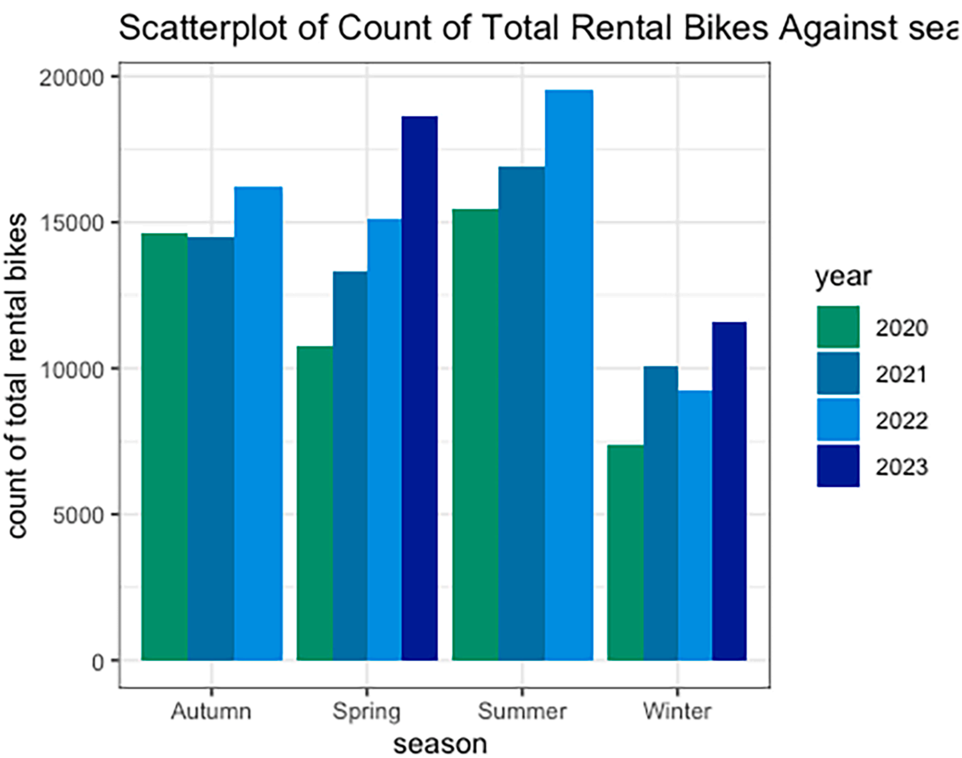

Fig. 2. Number of bike rentals for season and years.

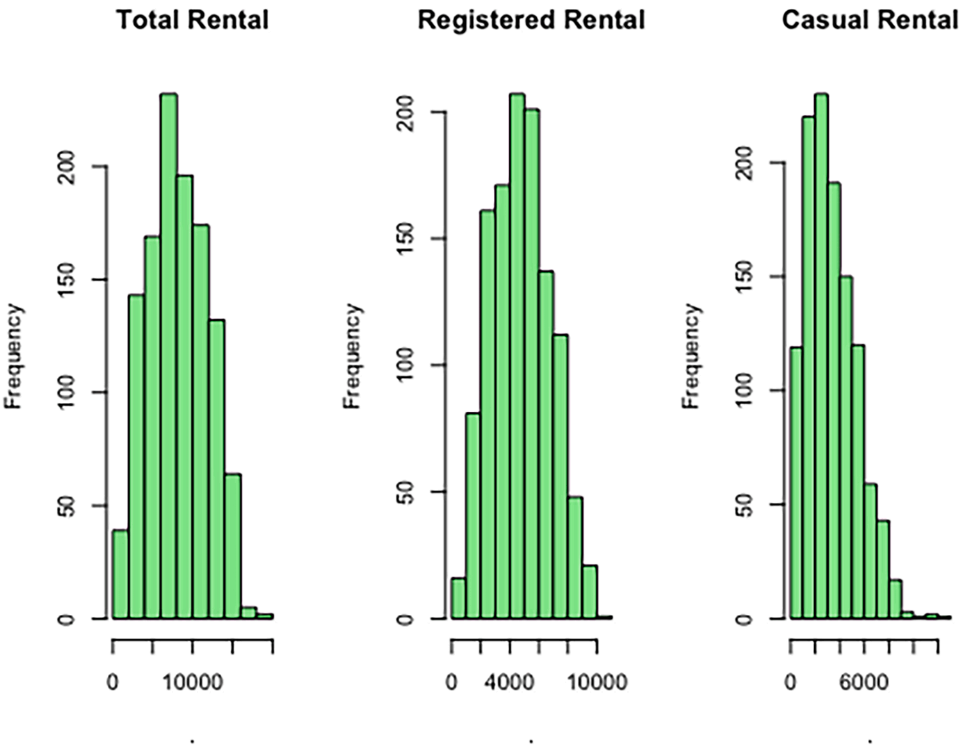

Fig. 3. Histogram of the number of rentals by membership type.

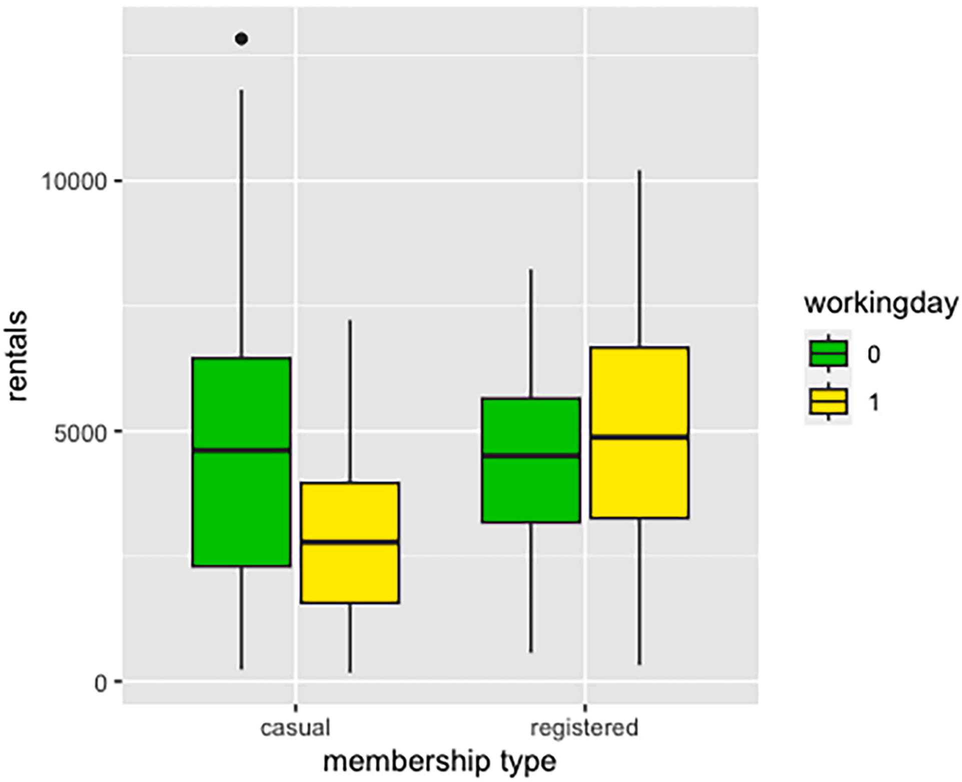

Fig. 4. A boxplot of the number of bike rentals for each member type by working day.

Table 1
Variable description.

Variable	Description
Total Rentals	count of total rental bikes, including both casual and registered
Registered	count of rentals by registered users
Casual	count of rentals by casual users
Year	(2020, 2021, 2022, 2023)
Humidity	humidity measured in percentage.
Visibility	is the measure of the distance at which an object or light can be discerned in km
Windspeed	wind speed measured in km/h
Temp	The raw temperature in Fahrenheit.
Working day	if the day is neither weekend nor holiday is 1, otherwise it is 0.
Month	1 to 12 (1: January, …, 12: December)
Season	spring, summer, autumn(fall) and winter
Latitude, Longitude	Station Location coordinates

Table 2
Summary of Weather Variables.

Summary	Total Rental	Temp	Windspeed	Humidity	Visibility
Min	497	14.86	3.714	24.12	2.304
Median	7964	60.39	8.429	64.13	10.000
Mean	8157	59.95	9.055	63.91	9.466
Max	19,531	88.71	20.667	97.34	10.000

calendar variables on the overall number of rentals and compare this effect for membership type number of rentals. In this model, the response variable is the number of daily bike rentals, and the predictors are the weather and calendar variables. We also wanted to capture station variability, so we used the stations' latitude and longitude as our location variables. This will allow us to see the location or station dynamics in the rental system. As shown in Fig. 5, the number of rentals differs by location and station. The south-central part of the city has the highest number of rentals, and the lowest number are in stations located in the northeast and southwest

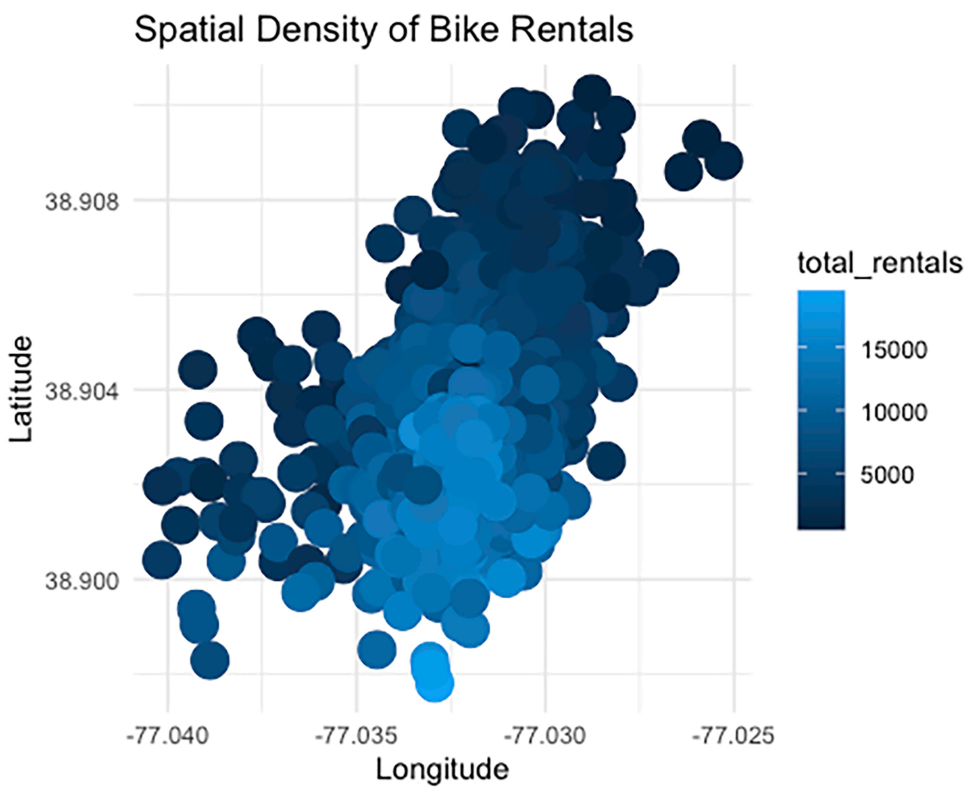

Fig 5. Dynamics of bike rental by station.

of the city.

In addition to accounting for location disparities, the target variable is count data. Poisson regression is mostly used for modeling count dependent variable. It is assumed that dependent variable follows a Poisson distribution, However, one unrealistic assumption of the Poisson distribution is that the mean and variation are equal. This assumption is shown mathematically in Eq. (1).

Suppose X follows a Poisson distribution

$$P(x,\ \mu\ ) = \frac{e^{-\mu}\mu^{x}}{x!},\ x = 0,\ 1,\ 2,\ \ldots$$

where, μ is an average number of bikes rental, e is the base of the logarithm and x is the random variable. then the following holds.

$$Var(X) = E(X) = \mu \tag{1}$$

When the means and the variances of daily bike rentals' overall counts are calculated, the ratio of the variance to the means indicates that the variance is about five times greater than the means in all instances [11]. In addition, we performed dispersion test on the poison model and found the presence of overdispersion. This result suggests that the Poisson assumption is unrealistic for this data type; hence, we have the problem of "over-dispersion". [11] proposes a negative binomial regression approach to overcome this problem. However, this approach also assumes a linear relationship between the log of the mean response and the predictors, such as temperature. Still, it is flawed since the relationship between temperature and the number of bike rentals is not linear, as shown in Fig. 6. When the temperature increases, the number of rentals also increases; when the temperature gets too hot, the number of rentals starts to reduce. To capture all non-linear relationships well with limited assumptions, we propose a penalized splines quasi-Poisson regression which will overcome the flaws of the model proposed by [11]. This model framework allows more flexibility and hence can capture all non-linear relationships without making or thinking about necessary transformations to satisfy the assumption of linearity in the classical parametric approach. We discuss this new approach briefly in the next section.

Generalized additive models (GAMs)

The analysis employs a semi-parametric approach called generalized additive models (GAMs) to capture linear and non-linear relationships between variables. The GAM framework allows for flexible modeling of the relationships between the target variable and the weather and calendar variables. A semi-parametric model has two parts; a parametric part to capture all the linear relationships and a non-parametric part to capture the non-linear relationships. When the dependent variable is not normally distributed, generalized linear models are often used by choosing the appropriate distribution usually from exponential family of distributions and

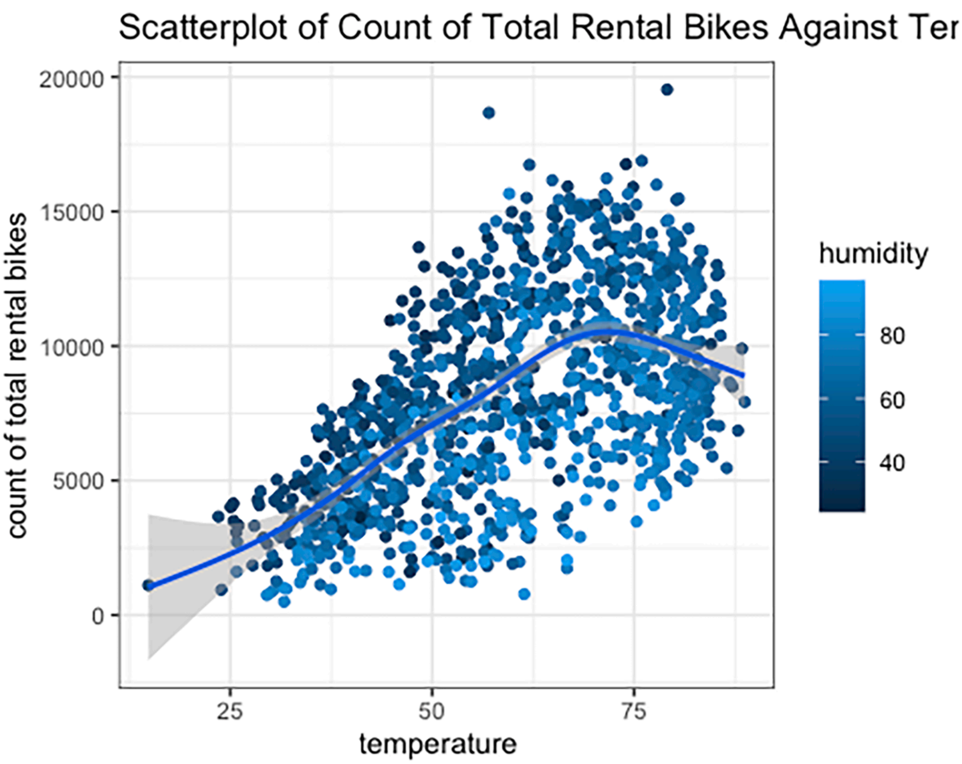

Fig. 6. Scatterplot of temperature against the number of rentals by humidity.

a link function $g(.)$ which is monotone and two times differentiable is employed to transform the model into linear in the parameters. The GAM generalizes the GLM by introducing smooth functions for the predictor variables. GAMs provide useful approximations to regression surfaces while maintaining the richness of GLMs but relaxing the assumption of linearity in the structure of the additive effect.

A typical penalized spline regression model is given in Eq. (2)

$$g(Y_i|X_{1i}, X_{2i}) = X_{1i}\beta + f(X_{2i}) + \varepsilon_i, \tag{2}$$

where, $(x_i,\ y_i)$ is the predictor/response pair, *f(.)* is a smooth function, and the errors, ε_i, are independent random variables with $E(\varepsilon_i) = 0$. In Eq. (2), the parametric part is the $X_{1i}\beta$ and the non-parametric part is the $f(X_{2i})$, which is assumed to be a smooth function, and basis functions characterize these functions. X_1 is a design matrix containing $1's$ in the first column and other variables considered to have a linear association with the response variable. X_2 is any variable which is believed to have some form of non-linear relationship with the response variable, however X_2 can be available even if its relationship with the response is linear because the smooth function will capture the actual relationship. Note that, the specification in (2) is when identity link is used while the specification in (3) is more general with any given link function. A more general GAM model is fit a response variable Y by sum of smooth functions of the explanatory variables, $X_j,\ j = 1, \ldots,\ p$.

$$g(\mu) = g(E(Y)) = \alpha + \sum_{j=1}^{p} f_j(X_j), \tag{3}$$

where, $g(.)$ is a link function (log, logit, probit etc.), α is the intercept in the model and f_j is the jth smooth function for the jth predictor variable. It is also possible to have a bivariate smooth function $(f_b(X_iX_j),\ i \neq j)$ for two predictor variables such as longitude and latitude which signifies location.

Penalized spline smoothing

One way to estimate the smooth function f is using penalized spline smoothing. One can view Penalized Spline regression from the non-parametric point. A typical non-parametric regression model is given by.

$$y_i = g(x_i) + \epsilon_i,\ \ i \leq i \leq\ n, \tag{4}$$

where, $(x_i,\ y_i)$ is the predictor/response pair, g is a smooth function and the errors, ϵ_i, are independent random variables with $E(\epsilon_i) = 0$.

A penalized spline regression defines the smooth function comprising parametric and non-parametric components, as shown in Eq. (5).

$$g(x) = \gamma_0 + \gamma_1 x + \sum_{k-1}^{K} u_k (x - K_k)_+, \tag{5}$$

where

$$(x - K_k)_+ = \begin{cases} x - K_k, & x < K_k \\ 0, & x \geq K_k \end{cases} \text{ and } 1 \leq k \leq K, \tag{6}$$

where, $K_k^{'}s$ are called knots and symbolize where we believe the slope will change for a non-linear smooth function. One common practice is to select evenly spaced knots.

A more general model for g is to replace $(x - K_k)_+$ with $Z_k(x)$, this allows us to define other types of basis functions. A more popular choice is $p = 3$, also known as the piecewise cubic basis, because it has a continuous second derivative but some computational instability [28]. One can use other basis functions, including thin-plate basis, B-spline basis, tensor product basis and many more. However, this study uses the thin-plate basis, which has good convergence properties and is computationally stable.

$$g(x) = \gamma_0 + \gamma_1 x + \sum_{k-1}^{K} u_k Z_k(x) \tag{7}$$

$$y_i =\ \gamma_0 + \gamma_1 x + \sum_{k-1}^{K} u_k Z_k(x) + \epsilon_i,\ \ i = 1, \ldots,\ n \tag{8}$$

The model parameters $\gamma_0,\ \gamma_1$ *and* $u_k^{'}s$, as shown in Eq. (7) are estimated by solving the optimization problem in Eq. (9).

$$\text{Minimize}\left[\ \sum_{i=1}^{n} \{y_i - \widehat{g}(x)\}^2 + \lambda \int_{-\infty}^{\infty} \widehat{g}^{(2)}(x)^2\ dx \right], \tag{9}$$

where, $\widehat{g}(x)$ is an estimated smooth function for $g(x)$, but we penalize the second derivative of the function to control the curvature of the smooth function. λ is a smoothing parameter, and it controls the smoothness of the function.

The second derivative is penalized even in non-parametric density estimation. The reason for this choice is to penalize the roughness of the function g, so we called this penalty term the roughness penalty term. The tuning parameters one needs to consider when estimating g are the number of Knots(K), type of basis function (e.g. thin-plate, B-splines, cubic-spline etc.) and smoothing parameter (λ). A good choice for each of these tunning parameters determines the goodness and appropriateness of the fit.

Selecting the smoothing parameter (λ). The choice of smoothing parameter significantly impacts the resultant penalized spline that is supported [28]. There are known algorithms for selecting the desired smoothing parameter. The R package "mgcv" uses data and a method is known as generalized cross-validation to choose the smoothing parameter . Using the magnitude of the smoothing parameter to assess the complexity of the penalized spline fit is inefficient because this value depends on the units of measurement of the data used for the regression. Instead, effective degrees of freedom (EDF) measure the model's complexity.

The effective degrees of freedom (EDF) is used to measure the complexity of the model; it is a monotone transformation of the smoothing parameter to obtain a measure with meaning.

Given the design matrix denoted as X and a matrix D which are given by

$$X = \begin{pmatrix} 1 & x_1 & z_1(x_1) & \cdots & Z_K(x_1) \\ & & \vdots & \ddots & \vdots \\ 1 & x_n & z_1(x_n) & \cdots & Z_K(x_n) \end{pmatrix} \tag{10}$$

$$D = \begin{bmatrix} 0_{2\times 2} & 0_{2\times K} \\ 0_{K\times 2} & I_K \end{bmatrix} \tag{11}$$

The closed-form solution to the optimization is shown in Eq. (12). The solution is based on a selected smoothing parameter.

$$\widehat{g}(X;\lambda) = X\left(X^T X + \lambda D\right)^{-1} X^T y \tag{12}$$

From (12), Eqs. (13) and (14) can be obtained.

$$EDF(\lambda) = tr\left\{ X\left(X^T X + \lambda D\right)^{-1} X^T \ \mathrm{X}\right\} \tag{13}$$

$$RSS(\lambda) = \sum_{i=1}^{n} \{y_i - \widehat{g}(x_i;\lambda)\}^2 \tag{14}$$

The GCV finds the smoothing parameter $\widehat{\lambda}_{GCV}$ which minimizes Eq. (14)

$$GCV(\lambda) \equiv \frac{RSS(\lambda)}{\left\{1 - \frac{EDF(\lambda)}{n}\right\}^2} \tag{15}$$

In this study, we will use the GCV approach to select the smoothing parameter for all the models, and we also use the number of knots ($K \leq 35$). Most signal in practice is captured with knots close to 35.

Quasi-Poisson regression

A quasi-Poisson regression is a more general form of Poisson regression used when there is a problem of overdispersion in the data under discussion.

$$\log(E(Y|X)) = X\beta \tag{16}$$

As given in Eq. (3), a Poisson regression model makes a strict assumption of equal variance and mean, which, when violated, will undermine estimates and statistical inference. In quasi-poisson regression, the model is the same as in Eq. (3). However, here it assumes that the variance is linearly related to the mean, as given in Eq. (4)

$$Var(Y) = \phi\mu, \tag{17}$$

where, ϕ is known as the dispersion parameter, and μ is the mean.

Penalized splines quasi-Poisson regression

We propose a quasi-Poisson regression that combines parametric and non-parametric components. The model we proposed is shown mathematically in Eq. (18) below. This model has the same assumption as shown in Eq. (17).

$$\log(E(Y|X_1,\ X_2) = X_1\beta + f(X_2) \tag{18}$$

However, we also capture the station dynamics using the latitude and the longitude. We do so by including a bivariate smooth function, as shown in Eq. (19)

$$\log(E(Y|X_1,\ X_2, lat,\ long) = X_1\beta + f(X_2) + f(lat, long), \quad (19)$$

where, Y is the number of rentals (either total rentals or rentals by registered users or rentals by casual users), X_1 is a vector of variables known to have linear relationships, so the parametric model is appropriate for them. In our study X_1 vector of variables, including visibility, windspeed, season, working day, year, and a constant for the intercept. It makes sense to make this choice upon inspecting the relationship between these variables and response. X_2 is any variable which shows a non-linear relationship with the response, which also includes temperature, humidity, the interaction between temperature and working day and month (smooth as a random effect) and finally, the bivariate smooth function of *lat* (latitude) and *long* (longitude) for location effect.

Results

This section presents the modeling results and some descriptive to support the findings. This section first discusses some relevant graphical summaries and then presents model results.

Descriptive analysis

Fig. 7 illustrates the scatter plot of total rentals against temperature, color-coded by visibility. Generally, rentals increase with higher temperatures, but at extremely high temperatures, a decline is observed. Enhanced visibility is associated with increased rentals. Two notable instances in Fig. 7 involve lower visibility (around 60°F) with a daily rental of approximately 18,000, and higher visibility and temperatures resulting in elevated rentals. Despite different weather conditions, these two points exhibit similar total rental numbers. The study aims to uncover the dynamics prompting higher bike rentals on the less visible day, possibly due to unavoidable events or efforts to avoid traffic congestion. Further insights into rental dynamics are pursued in this study.

Fig. 8 displays the fluctuating bike rental numbers across months and seasons. Rentals rise from January to July and decline from August to December. Seasonal impact is evident, with higher rentals in late spring, summer, and early autumn, contrasting lower numbers in winter and late fall. To capture these variations, Fig. 8 informs our modeling approach by treating seasonality and the month effect as a random effect smooth due to the nuanced relationship with total rentals.

In Fig. 9, a correlation plot explores associations among continuous variables. This helps gauge predictor-response relationships and identify potential multicollinearity issues. Variables like temperature, wind speed, humidity, and visibility are linked to bike rentals. Despite humidity and visibility correlating (due to reduced visibility in humid weather), both factors contribute uniquely to bike rental dynamics, justifying their inclusion for a comprehensive analysis.

The insight from the description in this section guided us in the model process. We present the results of this study in the next

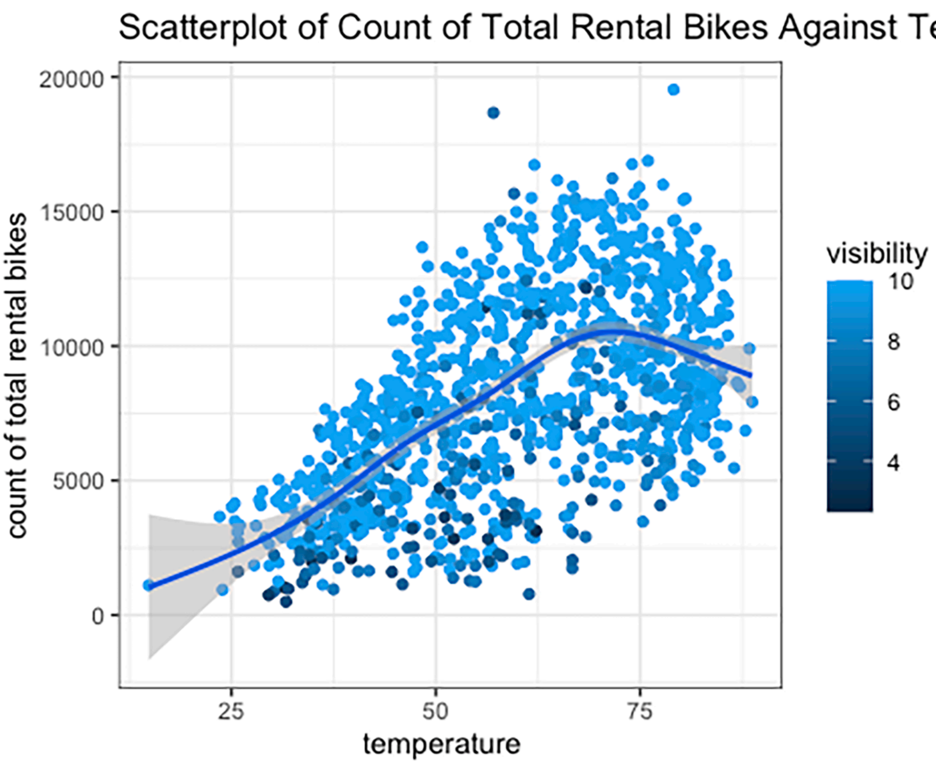

Fig. 7. Scatterplot of temperature against the number of rentals by visibility.

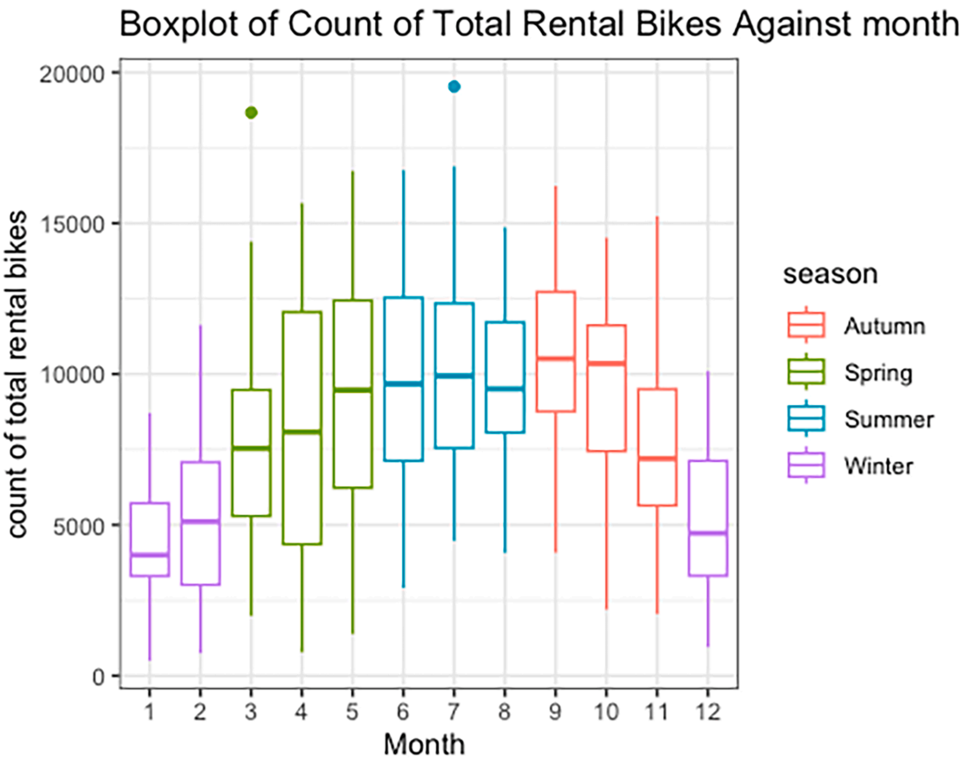

Fig. 8. Boxplot of the number of rentals for each month by season.

section.

Model results

In this section, we present all the findings from our study and discuss them in detail.

Model 1: total number of rentals

We built three separate models, and this section presents the first model's results. This model uses the total number of bike rentals as the target variable. This variable combines the number of rentals by both registered and casual users, as described in Table 1. In this model, we have the estimates for all relevant variables' effect on the total number of rentals in all the stations in Washington, D.C.

Visibility, with a positive estimate of 0.0281 (Pr(>|t|) = 0.003588), correlates significantly with increased bike rentals, attracting more users. Conversely, windspeed, with a negative estimate of −0.01862 (Pr(>|t|) = 6.742e-17), discourages rentals in windy conditions, aligning with [9] findings. Season is a categorical variable with three indicators, and their corresponding estimates indicate the impact of the three seasons compared to fall (reference season). Spring, with an estimate of −0.2248 (Pr(>|t|) = 5.397e-05), sees decreased rentals, while summer's −0.1083 estimate (Pr(>|t|) = 0.05825) is not statistically significant. Winter, at −0.2798 (Pr(>|t|) = 1.668e-06), significantly reduces rentals compared to fall.

The working day variable (estimate −0.08952, Pr(>|t|) = 7.02e-11) associates with fewer rentals on working days, suggesting higher demand on weekends or holidays. Yearly trends show a significant increase in rentals in 2021 (estimate 0.1368, Pr(>|t|) = 1.651e-12) and a further rise in 2022 (estimate 0.3263, Pr(>|t|) = 1.178e). In summary, visibility, windspeed, seasons, working days, and years are influential predictors in bike rentals.

The model includes yearly indicators for 2020, 2021, 2022, and 2023, comparing each to the reference year (2020). The estimate 0.1368 signifies increased rentals in 2021 (Pr(>|t|) = 1.651e-12), and 0.3263 indicates a further rise in 2022 (Pr(>|t|) <0.0001). Visibility maintains a positive association with bike rentals, attracting more users. Conversely, windspeed negatively influences rentals, discouraging bike usage. Seasons contribute, with spring and winter significantly decreasing rentals compared to fall, while non-working days exhibit higher demand.

In addition to parametric estimates (Table 4), Figs. 10, and 11 present the non-linear effects of several predictors. Months reveal varying patterns, highlighting non-linear associations with rentals. Geographical coordinates (longitude and latitude) underscore location importance, while temperature exhibits a non-linear effect, with an optimal range for higher rentals. The interaction between temperature and working days presents distinctive patterns with significant non-linear impacts. Humidity also demonstrates non-linear associations with bike rentals.

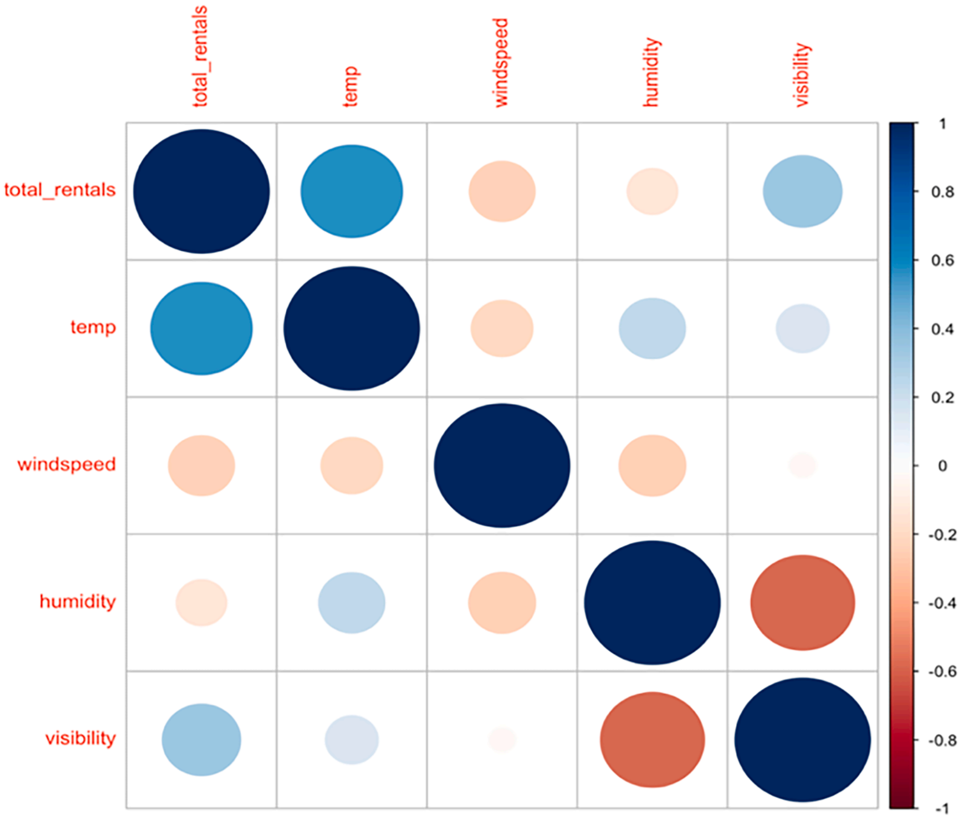

Fig. 9. Pairs plot for the continuous variables.

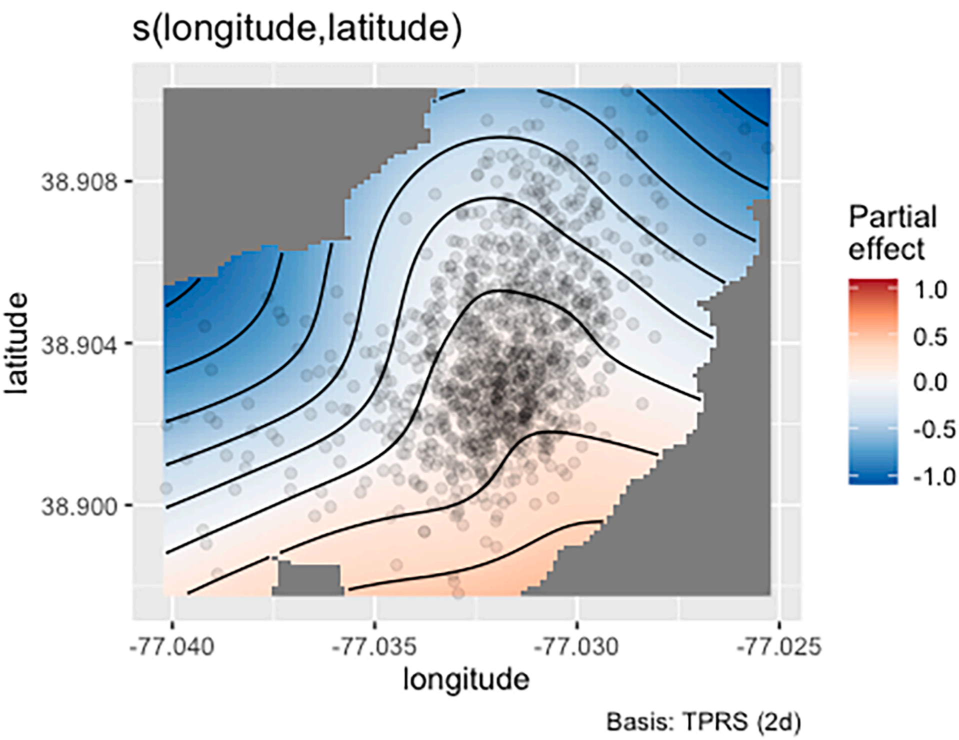

Fig. 10. Location effect on total rentals.

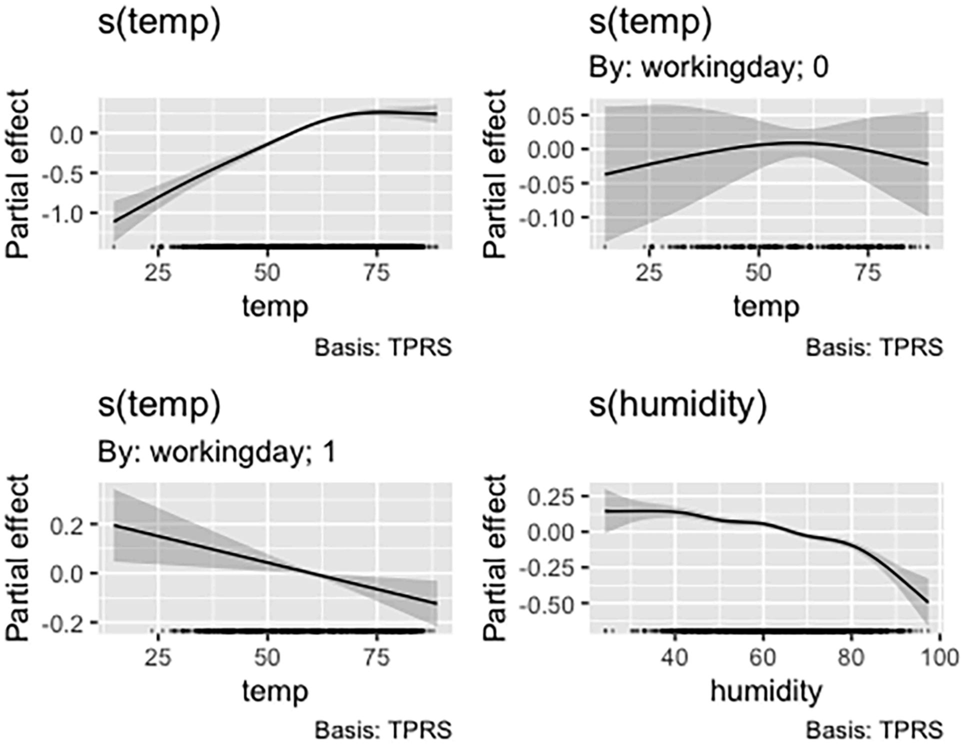

Fig. 11. Smooth functions for non-parametric estimates.

Fig. 10 illustrates the bivariate smooth for location effects, revealing higher rentals in the southern part of Washington D.C. with fewer in the north and west. Rentals are relatively high in the middle, especially pronounced in the southernmost area. Southeastern and northwest regions lack stations, resulting in no impact on recorded rentals. Fig. 11 presents smooth functions for other predictors, showcasing non-linear relationships with total bike rentals.

Temperature influences rentals non-linearly, with an overall increase. However, at extremely high temperatures (around 75°F, as shown in Fig. 11), scorching conditions deter bike rentals, causing a curve bend. Additionally, we explored the temperature and working day interaction. Fig. 11 reveals that temperature's impact on bike rentals depends on whether it's a working day. Working days show decreased rentals with rising temperatures, possibly due to people avoiding harsh conditions affecting their workday. Conversely, non-working days exhibit a similar effect as temperature rises, suggesting flexibility in bike usage regardless of environmental conditions.

Model 2: number of bike rentals by registered users

We present a model to compare how weather and calendar variables affect rentals by membership type. Table 5 shows the effects of these variables on rentals by registered users. Model 2 focuses on this group. The intercept is 8.04, the expected rentals when all predictors are zero. This baseline is lower than Model 1, which includes casual users. Visibility has a positive effect of 0.03189, larger than Model 1. Higher visibility means more rentals by registered users ($Pr(>|t|) = 0.0007055$). Windspeed has a negative effect of -0.01665, smaller than Model 1. Higher windspeed means fewer rentals by registered users ($Pr(>|t|) = 1.495e\text{-}14$). Table 5 also shows the season effects, compared to autumn. Spring, summer, and winter have negative effects of -0.3047, -0.1961, and -0.3142, respectively. These effects are larger and more significant than Model 1. All seasons have fewer rentals by registered users than autumn ($Pr(>|t|) < 0.001$). An interesting finding is the positive effect of working day (0.09564), unlike Model 1. Working days have more rentals by registered users than non-working days. This makes sense because registered users may use bikes to commute.

We compare the effects of different years on rentals by membership type. Table 5 shows the effects of binary variables for 2021, 2022, and 2023 on rentals by registered users, using 2020 as the reference year. Model 2 focuses on this group. The 2021 estimate is 0.2697, indicating more rentals by registered users in 2021 than 2020. This is significant, showing an upward trend in 2021. The 2022 estimate is 0.4803, indicating even more rentals by registered users in 2022 than 2020. This is highly significant ($Pr(>|t|) = 1.668e\text{-}88$), showing a continued upward trend in 2022. The 2023 estimate is the highest, indicating the most rentals by registered users compared to 2020. The year effects are larger and more significant in Model 2 than Model 1, suggesting that rentals have increased over time and more so among registered users.

Table 6 shows the non-linear effects of predictors on rentals by registered users. The edf for the smooth function of the month is 7.363, showing a non-linear relationship between month and rentals. This means that rentals vary by month, with some months having more impact than others.

The smooth function of longitude and latitude has an edf of 16.89, showing a non-linear relationship between location and rentals. This means that some locations in Washington, D.C. affect rentals more than others. The F-value of 9.87 and a p-value of less than 0.0001 show that this relationship is significant. The edf for the smooth function of temperature is 4.43, showing a non-linear relationship between temperature and rentals. This means that there is an optimal temperature range for higher rentals. Too high or low temperatures may reduce rentals. The F-value of 44.69 and a p-value of less than 0.0001 show that this relationship is significant. The smooth function of the interaction between temperature and working day has an edf of 0.000159 for working day 0 and 1 for working day 1. This means that there is no non-linear effect of this interaction on rentals. The high p-values show that this effect is not significant. The smooth function of humidity has an edf of 6.461, showing a non-linear relationship between humidity and rentals. This means that some humidity levels affect rentals more than others. The F-value of 13.39 and a p-value of less than 0.0001 show that this relationship is significant.

Fig. 12 shows how location affects rentals by registered users. This is different from Fig. 11. Rentals are still higher in the south, but not as much as in Fig. 11. Rentals are lower in the northeast and southwest. Temperature affects rentals like in Model 1. We observed a straight-line effect for temperature and non-working day. This means that rentals are constant for all temperatures on non-working days, but decrease with higher temperatures on working days. The non-parametric estimates show the non-linear effects of predictors on rentals by registered users. These results show the importance of non-linear relationships and the effects of months, locations, temperature, and humidity on rentals by registered users. Bike-sharing system operators can use this information to adjust their strategies, such as offering promotions in some months, optimizing bike availability in some locations, and considering temperature and humidity levels to match the needs and wants of registered users.

Model 3: number of bike rentals by casual users

Finally, we considered model 3, where the response variable is the number of rentals by casual users. This section presents the results for this model and compares the effects with the registered user's model.

Table 7 shows the results of Model 3, focusing on rentals by casual users. We compare them with Model 2 for registered users. The intercept is 8.148, the expected rentals when all predictors are zero. This is the baseline for casual users. Visibility has a positive effect of 0.02656, meaning more visibility might lead to more rentals by casual users. This is significant ($\Pr(>|t|) = 0.0244$), showing that better visibility is associated with increased number of casual users. Higher windspeed is also associated with a reduction in number of rentals by casual users. This relationship is significant ($\Pr(>|t|) = 8.698\text{e-}17$), showing that windy conditions discourage casual users. Spring has a negative effect of -0.0973, but this is not significant ($\Pr(>|t|) = 0.1421$), meaning spring season might not affect rentals by casual users much. Summer has a positive effect of 0.01005, but this is also not significant ($\Pr(>|t|) = 0.8823$), meaning summer seasons might not affect number of rentals by casual users. Winter has a negative effect, meaning fewer rentals by casual users in winter than in fall. Working day has a negative effect of -0.353, meaning casual users rent less on working days than non-working days.

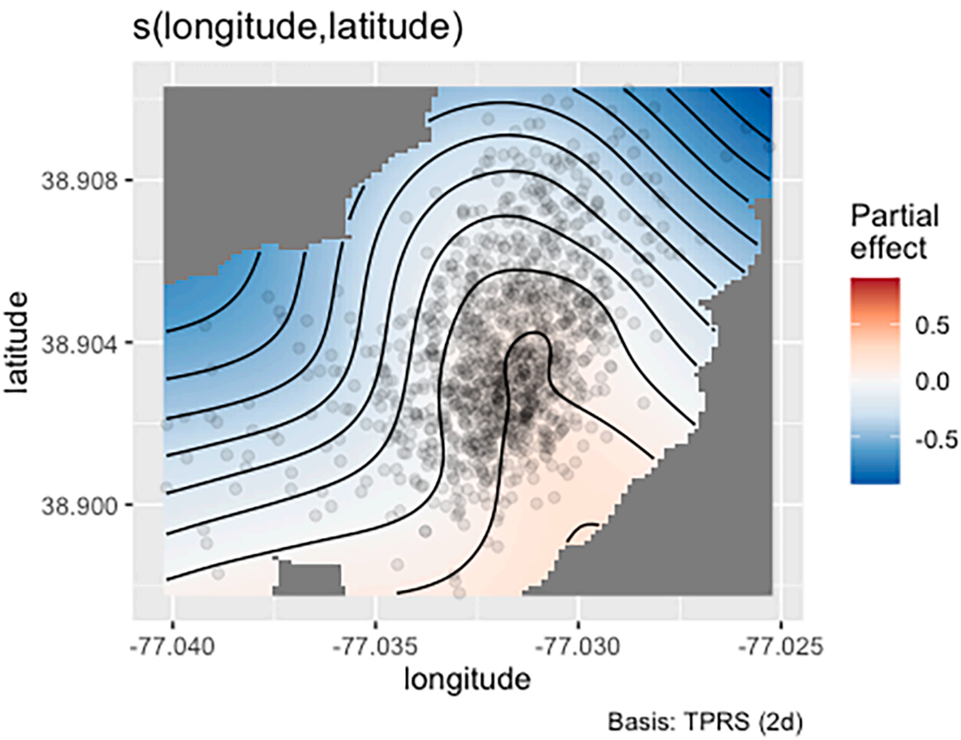

Fig. 12. Location effect on total rentals by registered users.

The years 2022 and 2023 have positive coefficients, meaning more rentals by casual users in these years than in 2020. These relationships are significant, showing an upward trend in 2022 and 2023.

Table 8 shows the non-linear effects of predictors on rentals by casual users. The edf for the smooth function of the month is 7.219, showing a non-linear relationship between month and rentals. This means that some months affect rentals more than others. The F-value of 5.005 and a p-value of less than0.0001 show that this relationship is significant. The smooth function of longitude and latitude has an edf of 18.8, showing a non-linear relationship between location and rentals. This means that some locations in Washington, D.C. influence rentals more than others. This matches the visual from Fig. 5. The edf for the smooth function of temperature is 3.982, showing a non-linear relationship between temperature and rentals. This means that there is an optimal temperature range for higher rentals. Extreme temperatures may lower rentals. The F-value of 28.72 and a p-value of less than 0.0001 show that this relationship is significant. The smooth function of the interaction between temperature and working day has an edf of 0.9225 for working day 0 and 1 for working day 1. This means that there is a small non-linear effect of this interaction on rentals. But the p-values (0.2683 and 0.4544) show that this effect is not significant. The smooth function of humidity has an edf of 6.352, showing a non-linear relationship between humidity and rentals. This means that some humidity levels affect rentals more than others. The F-value of 14.73 and a p-value of less than 0.0001 show that this relationship is significant. The non-parametric estimates show the non-linear effects of predictors on rentals by casual users. These results show the importance of non-linear relationships and the effects of months, locations, temperature, and humidity on rentals by casual users. Bike-sharing system operators can use this information to improve their services and marketing strategies by considering factors like month, location, temperature, and humidity to meet the needs and wants of casual users.

Comparing model 2 and model 3

We compare the effects of weather and calendar on rentals by membership type. We hypothesize that bike rental systems run both types reasonably. We use Fig. 16 to test this hypothesis. It shows the effect bars for registered and casual users. In 2022 and 2023, both types have more rentals than in 2020. In 2021, there is a difference between 2020 and 2021 rentals. This difference is negative for casual users and positive for registered users. Bike rentals increase for registered users from 2020 to 2023, as shown by the rising bars in Fig. 16. But casual users' growth rate is not consistent. It decreases in 2021 and increases in 2022 and 2023.

The working day effect is different for both types. Registered users rent more on working days, while casual users rent more on non-working days. Windspeed and visibility affect both types similarly in magnitude, but differently in direction. Fig. 16 shows that more visibility increases rentals, while more windspeed decreases rentals. The season effect is also different for both types. Spring and winter have negative effects for both types, but summer has a negative effect for registered users and a positive effect for casual users, compared to fall.

Discussion

The results of our study, which revealed a nuanced non-linear relationship between temperature, humidity, and month and the number of bike rentals, carry significant implications for both the bike industry and customers. While our findings align with previous research on the positive correlation between temperature and bike rentals, our study adds a layer of complexity by uncovering that this relationship holds true only within specific temperature ranges. Beyond these ranges, the number of rentals peaks and declines, indicating a critical insight for the industry. In addition, similar non-linear relationship was seen with humidity and month with the help of the GAM model. We also discovered an interaction between temperature and working days variable. We showed that the effect of temperature on number of bike rentals depends on whether the day is a working day or not which is a new finding in this area.

Previous studies have not explored the effect of the potential dependency between temperature and whether the day is a working day or not on the number of bike rentals. Since increasing temperature during working days is associated with a decreasing number of total rentals, it gives bike operators insight into the type of customers they are dealing with. It is more reasonable to say that customers use the bikes to commute to work, and since most people go to work in the morning, it makes sense to see a decline during the late hours when temperatures are higher. On the other hand, during weekend, regardless of the temperature the effect on bike rentals holding other variables constant does not change thereby insignificant. This makes this finding novel in the sense that it enlightens bike operators to prioritize optimizing operations on working days.

For the bike industry, this implies a need for strategic temperature management in optimizing rental services. Businesses can leverage this information to adapt their operations based on weather conditions, ensuring resource allocation and marketing efforts are aligned with the identified patterns. Recognizing the non-linear nature of the relationship allows for more effective planning and the ability to capitalize on peak rental periods while navigating less favorable conditions. For customers, the implication lies in the potential for an enhanced rental experience. A bike rental service that adjusts its operations based on our insights can offer improved reliability, ensuring that bikes are available when demand is high and optimizing the overall customer experience. This nuanced understanding of temperature-related demand patterns contributes to a more efficient and customer-centric approach in the bike rental industry.

Businesses can optimize bike rental operations by incorporating temperature-related demand patterns into various strategies. To begin, they can implement flexible pricing by offering discounts during peak heat periods in order to encourage rentals during cooler hours. Second, targeted marketing efforts can emphasize the convenience of biking in pleasant weather in order to attract customers. Third, businesses can diversify their product offerings by providing bikes suitable for different temperatures, such as electric bikes or those with adjustable gears. Fourth, localized operations can be accomplished by utilizing weather forecasts to allocate bikes where demand is highest based on temperature trends. Furthermore, organizing evening events with incentives, such as group rides during cooler hours, can increase rental rates.

Improving customer comfort by providing amenities such as shaded rest areas and hydration stations for a more enjoyable biking experience in hot weather is also critical. Collaborating with nearby businesses on joint promotions can help attract customers during peak heat periods. Finally, technological investments, such as developing apps for real-time weather updates and bike availability, can help customers plan.

We have uniquely captured the influence of location on bike rentals by employing a bivariate smooth surface that incorporates the longitude and latitude of the stations. Our analysis revealed a distinct spatial pattern, indicating that a higher frequency of bike rentals occurs in the southern part of the city, while the eastern and northern regions experience lower rental activity. Additionally, there is a moderate number of rentals in the central part of the city.

This spatial insight not only enhances our understanding of rental patterns but also holds significant implications. For the bike-sharing industry, it suggests the need for targeted resource allocation and marketing strategies. By recognizing the spatial variations in rental demand, companies can optimize bike distribution and promotional efforts in the high-activity southern areas, potentially increasing revenue. Conversely, adjustments can be made in the less active eastern and northern regions to ensure efficient resource utilization.

Furthermore, for urban planning and infrastructure development, acknowledging these spatial trends in bike rentals can inform decisions related to station placement and the expansion of bike-sharing services. This insight contributes to the creation of more effective and tailored strategies for both the bike-sharing industry and urban development initiatives.

Finally, our analysis demonstrates a clear distinction in bike usage patterns between registered and casual customers. Registered customers predominantly utilize bikes on working days, suggesting a focus on commuting and regular weekday activities. In contrast, casual customers exhibit a different trend, with a notable preference for bike usage during weekends and holidays. This divergence in usage patterns highlights the distinct preferences and behaviors of these customer segments, providing valuable insights for tailoring marketing strategies, service offerings, and operational considerations based on the differing needs of registered and casual users. The methodology utilized in this study is effective, particularly for datasets with a relatively large sample size, making it applicable to various regression problems. Generalized Additive Models (GAMs) boast a broad spectrum of applications, ensuring the generalizability of our methodology to diverse datasets.

Generalized Additive Models (GAMs) provide several advantages in statistical modeling. They offer flexibility and nonlinear modeling capabilities, enabling the capture of complex relationships that go beyond linear assumptions. GAMs effectively handle noisy data by incorporating smoothing functions, which result in smoother estimates of underlying relationships. Additionally, GAMs have automatic feature selection, which reduces overfitting and improves generalization performance. Their interpretability allows researchers to easily understand and communicate individual predictor effects, while their ability to model predictor interactions aids in the capture of sophisticated dependencies within the data. Furthermore, GAMs exhibit robustness to outliers, which contributes to

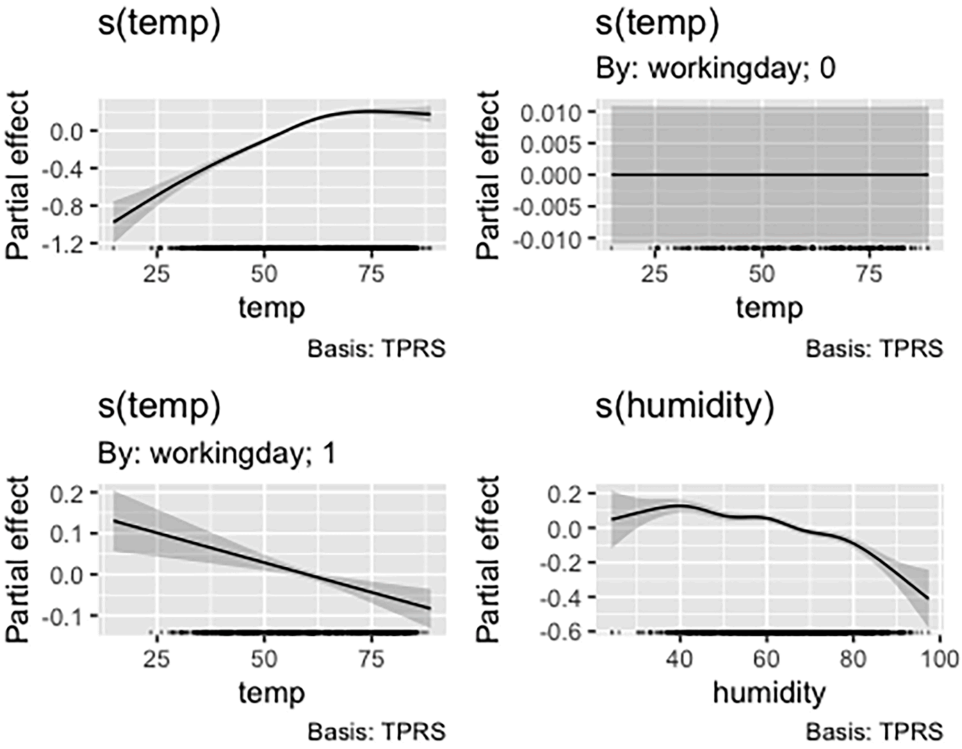

Fig. 13. Smooth functions for non-parametric estimates registered.

increased model stability. Finally, tuning GAMs with cross-validation ensures reliability on previously unseen data, increasing the model's overall robustness and validity.

While we successfully navigated through this study, there are several limitations that should be acknowledged. Firstly, our analysis is based on a relatively short time frame, encompassing only three years (2021, 2022, and 2023). As a result, generalizing the findings to other years may be constrained, and a more extended historical dataset would have strengthened the robustness of our conclusions. However, obtaining and processing a more extensive dataset poses computational challenges.

Additionally, our study focuses on a single city with unique weather and calendar dynamics. Consequently, the results may not be directly applicable to other cities with distinct environmental and temporal characteristics. It is crucial to recognize that variations in weather patterns and calendar dynamics across different locations could yield different outcomes. Therefore, while our findings may be relevant to areas sharing similar weather and calendar variables, caution should be exercised when extrapolating these results to locations with markedly different situations (Figs. 13–15, Tables 3–8).

Conclusion

In conclusion, our study has unearthed hidden non-linear relationships between temperature, humidity, and month with the number of bike rentals, presenting a nuanced understanding that significantly impacts the bike industry and customers. The identification of specific temperature ranges where rental patterns peak and decline offers a crucial insight, revolutionizing temperature management strategies for bike rental services. Moreover, our exploration of humidity, month, and their interactions with temperature on working days unveils novel findings, enhancing the comprehensiveness of our insights.

The spatial dimension adds another layer to our understanding, highlighting distinct rental patterns across different city regions. This spatial insight not only informs targeted resource allocation and marketing for the bike-sharing industry but also provides valuable input for urban planning and infrastructure development.

The discernment of usage patterns between registered and casual customers during weekdays and weekends emphasizes the importance of tailoring services to different customer segments. This revelation guides marketing strategies, service offerings, and operational considerations, acknowledging the differing needs of these customer groups.

While our methodology effectively addresses the identified gaps and proves applicable to various regression problems, it's imperative to recognize the limitations. The constrained time frame and focus on a single city may limit the generalizability of our findings. However, these limitations offer opportunities for future research endeavors, urging scholars to explore longer historical datasets and diverse city contexts to enhance the robustness and applicability of similar studies.

In essence, our study not only contributes methodologically but also offers practical implications for the bike industry, urban planning, and customer-centric service optimization. It encourages a broader perspective on weather-related demand patterns, spatial considerations, and customer segmentation, providing readers with valuable insights for both research and real-world applications.

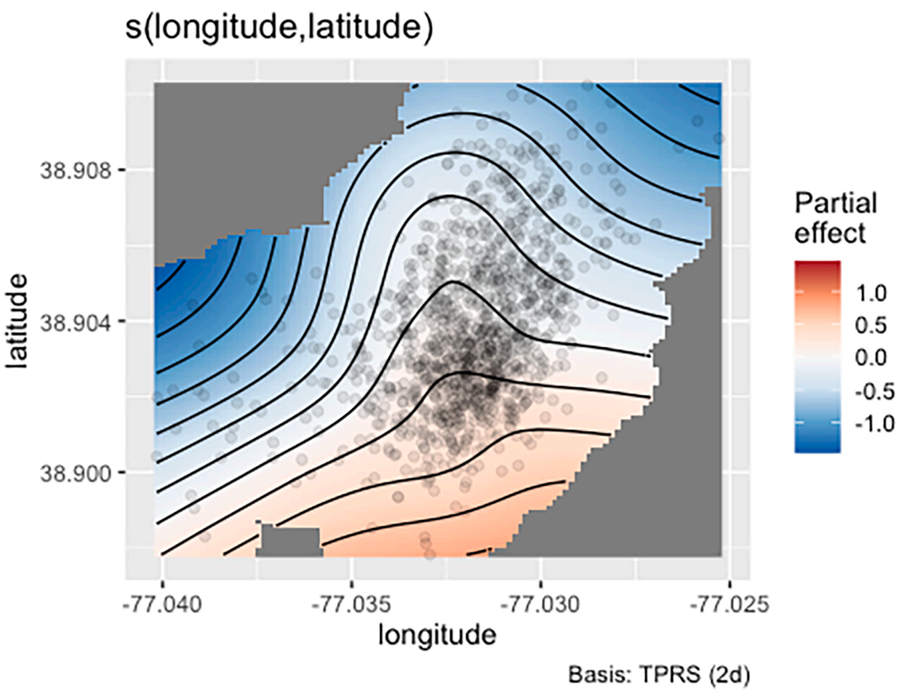

Fig. 14. Location effect on the number of bike rentals by Casual users.

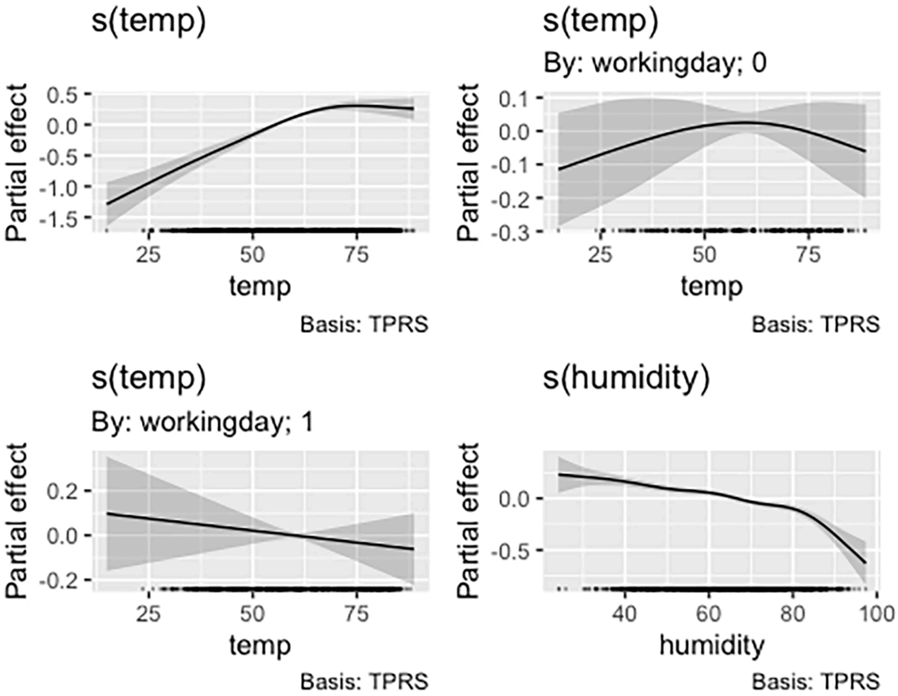

Fig. 15. Smooth functions for non-parametric estimates-Casual.

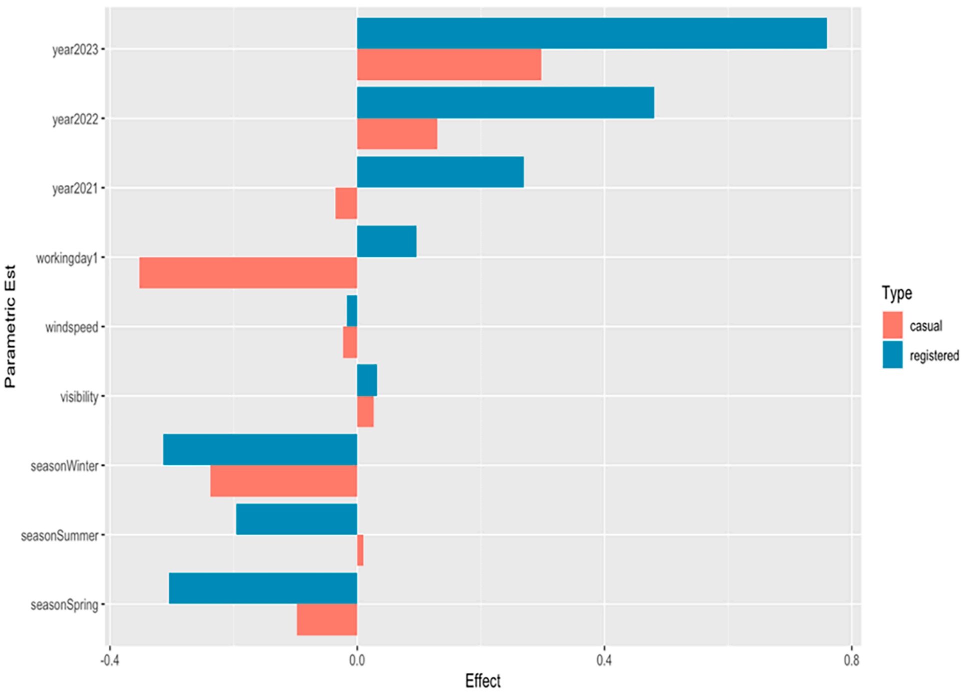

Fig. 16. Comparing Parametric effects for models 2 and 3.

CRediT authorship contribution statement

Christopher Odoom: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Alexander Boateng:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing –

Table 3
Parametric estimates for predictors' effects.

Characteristic	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.8	0.1059	83.13	0
visibility	0.0281	0.009628	2.918	0.003588
windspeed	−0.01862	0.002194	−8.487	6.742e-17
season Spring	−0.2248	0.05545	−4.054	5.397e-05
season summer	−0.1083	0.05715	−1.896	0.05825
season winter	−0.2798	0.05811	−4.816	1.668e-06
workingday1	−0.08952	0.0136	−6.585	7.02e-11
year2021	0.1368	0.01916	7.143	1.651e-12
year2022	0.3263	0.0217	15.04	1.178e-46
year2023	0.561	0.02781	20.17	2.58e-77

Table 4
Non-Parametric estimates for predictors' effects.

Characteristic	edf	Ref. df	F	p-value
s(month)	7.276	8	6.515	0
s(longitude, latitude)	18.71	23.58	15.54	0
s(temp)	4.237	5.315	35.5	0
s(temp): workingday0	0.4895	0.8286	0.02039	0.8966
s(temp): workingday1	1.001	1.001	6.77	0.009383
s(humidity)	6.217	7.42	16.22	0

Table 5
Parametric estimates for predictors effects (response: registered).

Characteristic	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.04	0.1046	76.88	0
visibility	0.03189	0.009387	3.397	0.0007055
windspeed	−0.01665	0.002137	−7.793	1.495e-14
season spring	−0.3047	0.05753	−5.297	1.419e-07
season summer	−0.1961	0.05943	−3.3	0.0009979
season winter	−0.3142	0.05973	−5.261	1.718e-07
workingday1	0.09564	0.01364	7.012	4.071e-12
year2021	0.2697	0.01936	13.93	8.731e-41
year2022	0.4803	0.02195	21.88	1.668e-88
year2023	0.7595	0.02792	27.21	2.716e-125

Table 6
Non-Parametric estimates for predictors effects (response: registered).

Characteristic	edf	Ref.df	F	p-value
s(month)	7.363	8	9.013	0
s(longitude, latitude)	16.89	21.9	9.87	0
s(temp)	4.43	5.545	44.69	0
s(temp):workingday0	0.000159	0.0003113	0.01497	0.9983
s(temp):workingday1	1	1	12.29	0.0004721
s(humidity)	6.461	7.645	13.39	0

review & editing, Supervision, Methodology, Visualization. **Sarah Fobi Mensah:** Methodology, Writing – original draft, Writing – review & editing, Conceptualization, Visualization, Data curation, Formal analysis. **Daniel Maposa:** Writing – original draft, Writing – review & editing, Supervision, Conceptualization, Methodology, Visualization, Data curation, Formal analysis.

Declaration of competing interest

On behalf of the authors, I declare that we have no conflicts of interest to disclose for the submitted original research article entitled "Modeling of the Daily Dynamics in Bike Rental System Using Weather and Calendar Conditions: A Semi-Parametric Approach" for consideration by the *Scientific African*. I confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

Table 7
Parametric estimates for predictors effects (response: casual).

Characteristic	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.148	0.1285	63.43	0
visibility	0.02656	0.01178	2.254	0.0244
windspeed	−0.02265	0.002679	−8.455	8.698e-17
seasonspring	−0.0973	0.06624	−1.469	0.1421
season summer	0.01005	0.06785	0.1481	0.8823
season winter	−0.238	0.07022	−3.39	0.0007241
workingday1	−0.353	0.01638	−21.55	2.653e-86
year2021	−0.03514	0.02224	−1.58	0.1144
year2022	0.1293	0.02512	5.146	3.139e-07
year2023	0.2976	0.03261	9.128	3.229e-19

Table 8
Non-Parametric estimates for predictors effects (response: casual).

Characteristic	edf	Ref.df	F	p-value
s(month)	7.219	8	5.005	1.175e-06
s(longitude, latitude)	18.8	23.54	22.83	0
s(temp)	3.982	5.011	28.72	0
s(temp):workingday0	0.9225	1.422	1.07	0.2683
s(temp):workingday1	1.001	1.001	0.5605	0.4544
s(humidity)	6.352	7.544	14.73	0

Data availability

To access the data and code for this paper, use the GitHub link below: https://github.com/codoom1/Modelling-of-the-Daily-Dynamics-in-Bike-Rental-System-Using-Weather-and-Calendar-Conditions.git

Acknowledgements

The authors are grateful to the colleagues who shared their experience in the bike rental system and the methodology. We also extend our gratitude to the Editor and the two anonymous reviewers for their constructive criticism and review.

References

[1] E. Macioszek, M. Cieśla, External environmental analysis for sustainable bike-sharing system development, Energies 15 (3) (2022) 791.
[2] J. Corcoran, T. Li, D. Rohde, E. Charles-Edwards, D. Mateo-Babiano, Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events, J. Transp. Geogr. 41 (2014) 292–305.
[3] K. Gebhart, R.B. Noland, The impact of weather conditions on bikeshare trips in Washington, DC, Transportation 41 (2014) 1205–1225.
[4] T. Thomas, R. Jaarsma, B. Tutert, Exploring temporal fluctuations of daily cycling demand on Dutch cycle paths: the influence of weather on cycling, Transportation 40 (2013) 1–22.
[5] JDAntos:CapitalBikeshare:UsageandWeather. http://jdantos.wordpress.com/2012/03/01/capital-bikeshare-usage-and-weather/(2012).
[6] L.F. Miranda-Moreno, T. Nosal, Weather or not to cycle: temporal trends and impact of weather on cycling in an urban environment, Transp. Res. Rec. 2247 (1) (2011) 42–52.
[7] H. Ding, N.N. Sze, Y. Guo, Y. Lu, Effect of the ultra-low emission zone on the usage of public bike sharing in London, Transp. Lett. 15 (7) (2022) 1–9.
[8] C. Lyu, J. Bie, H. Wang, Discovery of temporal variations of public bike ridership using internet of things for building smart city, Microprocess. Microsyst. (2021) 104320.
[9] D. Kumar, Meteorological barriers to bike rental demands: a case of Washington DC using NCA approach, Case Stud. Transp. Policy 9 (2) (2021) 830–841.
[10] E. Eren, V.E. Uz, A review on bike-sharing: the factors affecting bike-sharing demand, Sustain. Cities Soc. 54 (2020) 101882.
[11] K. Kim, Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations, J. Transp. Geogr. 66 (2018) 309–320.
[12] T. Hastie, R. Tibshirani, Generalized additive models: some applications, J. Am. Stat. Assoc. 82 (398) (1987) 371–386.
[13] A. Guisan, T.C. Edwards Jr, T. Hastie, Generalized linear and generalized additive models in studies of species distributions: setting the scene, Ecol. Model. 157 (2–3) (2002) 89–100.
[14] K. Ravindra, P. Rattan, S. Mor, A.N. Aggarwal, Generalized additive models: building evidence of air pollution, climate change and human health, Environ. Int. 132 (2019) 104987.
[15] S.N. Wood, Y. Goude, S. Shaw, Generalized additive models for large data sets, J. R. Stat. Soc. Ser. C Appl. Stat. 64 (1) (2015) 139–155.
[16] P. Midgley, The role of smart bike-sharing systems in urban mobility, Journeys 2 (1) (2009) 23–31.
[17] M. Roman, M. Roman, Bicycle transport as an opportunity to develop urban tourism–Warsaw example, Procedia Soc. Behav. Sci. 151 (2014) 295–301.
[18] T. Raviv, O. Kolka, Optimal inventory management of a bike-sharing station, IIE Trans. 45 (10) (2013) 1077–1093.
[19] T.D. Tran, N. Ovtracht, B.F. d'Arcier, Modeling bike sharing system using built environment factors, Procedia CIRP 30 (2015) 293–298.
[20] I. Frade, A. Ribeiro, Bicycle sharing systems demand, Procedia Soc. Behav. Sci. 111 (2014) 518–527.
[21] Y. Feng, S. Wang, A forecast for bicycle rental demand based on random forests and multiple linear regression, in: Proceedings of the IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), IEEE, 2017, pp. 101–105.
[22] S.N. Wood, Generalized Additive models: An Introduction with R, CRC Press, 2017.
[23] C.J. Stone, C.Y. Koo, Additive splines in statistics, in: Proceedings of the American Statistical Association 45, 1985, p. 48. *Original pagination is p*.

[24] S. Amodio, J. Meulman, Generalized Boosted Additive Models, University of Naples Federico II, 2011.
[25] B.D. Marx, P.H. Eilers, Direct generalized additive modeling with penalized likelihood, Comput. Stat. Data Anal. 28 (2) (1998) 193–209.
[26] A. Buja, T. Hastie, R. Tibshirani, Linear smoothers and additive models, Ann. Stat. 17 (2) (1989) 453–510.
[27] Capital Bike Share Data: http://capitalbikeshare.com/system-data (Accessed date 14 June 2023).
[28] J. Harezlak, D. Ruppert, M.P. Wand, Semiparametric Regression with R, Springer, New York, 2018.