Peertechz

MATHEMATICS

Computational Mathematics and its Applications

OPEN ACCESS JOURNAL

DOI: https://dx.doi.org/10.17352/cma

Research Article

Causal inference of Seoul bike sharing demand

Quan Yuan*, Zhixin Yang and Yayuan Xiao

Department of Mathematical Sciences, Ball State University, Muncie, IN 47304, USA

Received: 08 April, 2024
Accepted: 29 April, 2024
Published: 30 April, 2024

***Corresponding author:** Quan Yuan, Assistant Teaching Professor, Department of Mathematical Sciences, Ball State University, Muncie, IN 47304, USA,
E-mail: qyuan@bsu.edu

ORCiD: https://orcid.org/0000-0001-7024-5984

Keywords: Bike-sharing; Demand analysis; Causal inference; Multiple linear regression

Abstract

The global surge in environmental consciousness has significantly boosted the demand for rental bikes, particularly in metropolitan areas such as Seoul. This study delves into the causal relationships affecting this demand using a dataset from Seoul's bike-sharing system. Unlike previous research focusing predominantly on predictive analytics, this work innovatively applies multiple linear regression models to uncover causal inferences, offering insights that extend beyond mere forecasting. The challenges addressed include dealing with non-linear relationships and heteroscedasticity by employing the logarithmic transformation of rental counts. This approach not only aids in normalizing the data but also enhances the interpretability of the regression outcomes, emphasizing the changes in demand as a function of various environmental and temporal variables. Recent developments in causal inference methodologies have allowed for more robust and detailed analysis, paving the way for this study's contribution to the field. The findings underscore the significant influence of factors such as hour of the day, humidity, and seasonal changes on bike rental volumes, which can inform policy-making and operational strategies in urban transport planning.

Introduction

In recent years, the demand for rental bikes has been steadily increasing in metropolitan areas worldwide, driven by a growing global trend towards environmental protection and sustainable transportation [1-3]. Bike-sharing systems offer a convenient and eco-friendly alternative to traditional modes of urban mobility, allowing users to rent bicycles for short trips and return them to designated docking stations [4,5]. However, providing cities with a stable supply of rental bikes to meet the fluctuating demand has become a major challenge for bike-sharing operators [6,7]. Understanding the factors that influence bike rental demand is crucial for optimizing fleet management, improving user satisfaction, and promoting sustainable urban mobility [8-10]. While rental bikes serve as a key component of urban mobility, it is important to consider alternative options such as public transportation, private vehicles, walking, and other micro-mobility solutions like scooters [11-13]. Despite the presence of these alternatives, bike sharing remains a dominant force in the realm of sustainable transportation [14,15].

The topic of bike sharing demand has attracted significant attention from researchers in recent years [16-18]. Numerous studies have explored various aspects of bike-sharing systems, including demand prediction [19,20], user behavior analysis [21,22], and system optimization [23,24]. These studies have employed a wide range of methodologies, such as linear regression [25], time series analysis [26], and machine learning techniques like neural networks [27,28]. However, the majority of these works focus primarily on accurate demand prediction rather than causal inference [29,30]. While accurate prediction is undoubtedly valuable for operational planning, understanding the causal relationships behind bike rental demand is crucial for designing effective interventions and policies to encourage sustainable transportation.

Citation: Yuan Q, Yang Z, Xiao Y (2024) Causal inference of Seoul bike sharing demand. Comput Math Appl. 2(1): 005-009.
DOI: https://dx.doi.org/10.17352/cma.000005

Causal inference is a statistical approach that aims to identify the true causal effects of variables on an outcome of interest, going beyond mere correlations [9,10]. In the context of bike sharing demand, causal inference can help uncover the factors that directly influence user behavior and rental patterns, such as weather conditions, time of day, or bike infrastructure [3,15,20]. Several studies have applied causal inference techniques to investigate bike-sharing demand. For example, [11] used a difference-in-differences approach to evaluate the impact of a policy change on bike rental demand, while [12] employed a regression discontinuity design to estimate the effect of weather on bike usage. However, there remains a need for more comprehensive studies that apply causal inference methods to large-scale bike-sharing datasets, considering a wide range of potential causal factors [29,30].

To address this gap, this paper uses a dataset of Seoul bike-sharing demand and attempts to identify the key factors that contribute to the demand for rental bikes. By employing multiple linear regression models and analyzing the causal relationships between various independent variables and bike rental demand, this study aims to provide valuable insights for policymakers and bike-sharing operators. The findings can inform strategies to optimize bike fleet management, improve user experience, and promote sustainable urban mobility in Seoul and beyond [6,8,13,24].

Data and empirical strategy

The dataset used in this paper is from the UCI Machine Learning Repository [31], which records the number of rental bikes in Seoul every hour from December 1, 2017, at 0:00 to November 30, 2018, at 23:00, containing a total of 8,465 observations. Table 1 lists the variables in the dataset and their descriptions.

Among these variables, we do not use the “Functioning Day” variable because when the rental station is closed, the number of rental bikes is 0. Therefore, we deleted 295 observations where “Functioning Day” is “No”.

Table 1: Variables in the dataset and their description.

Variable	Description
Rented Bike Count	The integer number of bikes rented every hour (from 12:00 am to 11:59 pm)
Hour	Takes 24 integers from 0 to 23
Temperature	The temperature in every hour (similarly hereinafter, Celsius scale)
Humidity	percentage
Wind speed	m/s
Visibility	10 meters
Dew point temperature	Celsius scale
Solar Radiation	MJ/m^2
Rainfall	mm
Snowfall	cm
Seasons	Spring, Summer, Autumn, and Winter
Holiday	Yes or No
Functioning Day	Yes or No (The bike rental station open or close)

We employed multiple linear regression models to make causal inferences about the factors influencing bike rental demand [32]. The key assumption behind this approach is that the regression models can adequately capture the causal relationships between the independent variables and the dependent variable by controlling for multiple potential confounding factors simultaneously. By estimating the coefficients of the independent variables and assessing their statistical and economic significance, the models aim to identify the factors that have a causal impact on bike rental demand.

We used the natural logarithm of bike rental counts as the dependent variable in the regression models. We made this choice based on the following reasons: First, taking the natural logarithm can help transform a potential nonlinear relationship between bike rental counts and the influencing factors into a linear one, making the linear regression model more applicable. Second, it can reduce heteroscedasticity, which occurs when the conditional variance of the dependent variable varies with the levels of the independent variables. Third, taking the natural logarithm can improve the normality of the residuals, as the distribution of bike rental counts may be right-skewed. Fourth, when the dependent variable is log-transformed, the interpretation of the coefficients becomes more intuitive, representing the percentage change in bike rental counts for a one-unit change in the independent variable. Finally, it can reduce the differences in scales among variables, making the coefficients more comparable.

We considered several multiple linear regression models. We used the natural logarithm of bike rental counts as the dependent variable; the independent variables are Hour, Temperature, Humidity, Wind speed, Visibility, Dew point temperature, Solar Radiation, Rainfall, and Snowfall. The values of these variables are numeric. For the Season and Holiday variables, they can be considered as random samples across multiple periods. Therefore, we also introduced 3 dummy variables for the season (Spring, Summer, and Autumn) and 1 dummy variable for holiday. Table 2 lists the variables used in the models.

The following 4 models are considered

Model (1): Using Hour, Temperature, and Humidity as independent variables.

$$(1)\ y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+u$$

Model (2): Using Hour, Temperature, Humidity, Wind speed, Visibility, and Dew point temperature as independent variables.

$$(2)\ y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\beta_6x_6+u$$

Model (3): Adding Solar Radiation, Rainfall, and Snowfall as independent variables.

$$(3)\ y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\beta_6x_6+\beta_7x_7+\beta_8x_8+\beta_9x_9+u$$

Model (4): Including all dummy variables.

$$(4)\ y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\beta_6x_6+\beta_7x_7+\beta_8x_8+\beta_9x_9+\beta_{10}d_1+\beta_{11}d_2+\beta_{12}d_3+\beta_{13}d_4+u$$

Citation: Yuan Q, Yang Z, Xiao Y (2024) Causal inference of Seoul bike sharing demand. Comput Math Appl. 2(1): 005-009.
DOI: https://dx.doi.org/10.17352/cma.000005

Results

Table 3 shows the results obtained using R software. The numbers outside the parentheses are the estimated coefficient values; the numbers in parentheses are standard errors. Estimates with ** are statistically significant at the 1% level, and those with * are statistically significant at the 5% level.

Analysis

The analysis of variables is shown in Table 4.

Interpretation:

- From Table 3, it can be seen that the R-squared of all 4 models exceeds 50%, so all multiple linear models are good.
- All dummy variables are statistically and economically significant. This means that Seasons and Holiday are very important factors.
- Hour, Humidity, Dew point temperature, and Rainfall are important factors affecting bike demand. They are similar across different models and have statistical and economic significance.
- Visibility and Solar Radiation are not important factors.
- Some factors such as Wind speed and Snowfall are significant in some models but not in others. This paper speculates that the reason is these factors are related to Season.

Conclusion and prospects

In conclusion, this paper investigates the factors influencing the demand for rental bikes in Seoul using a dataset of Seoul bike sharing demand. By employing multiple linear regression models and analyzing the statistical and economic significance of the estimated coefficients, this study identifies several key factors that have a causal impact on bike rental demand, such as Hour, Humidity, Dew point temperature, Rainfall, and dummy variables for Season and Holiday. The results suggest that these factors play a crucial role in determining the demand for rental bikes in Seoul. Furthermore, the paper highlights the importance of considering causal relationships rather than solely focusing on prediction accuracy when analyzing bike-sharing demand.

The methodology and findings of this study have potential applications beyond Seoul. Bike-sharing programs are becoming increasingly popular in cities around the world as a sustainable mode of transportation. Future research could apply similar causal inference techniques to analyze bike-sharing demand in other regions and countries, taking into account local contextual factors. This could provide valuable insights for policymakers and bike-sharing operators looking to optimize their systems and promote sustainable urban mobility.

Table 3: Results of the 4 models obtained by R.

	Dependent Variable: Natural log of Rented Bike Count			
Independent variables	(1)	(2)	(3)	(4)
Constant	6.0043**	8.025**	7.338**	7.403**
	(0.0343)	(0.1709)	(0.1650)	(0.1572)
Hour	0.0379**	0.0410**	0.0432**	0.0440**
	(0.0013)	(0.0013)	(0.0013)	(0.0013)
Temperature	0.0563**	-0.0283**	-0.0067	-0.0268**
	(0.0008)	(0.0064)	(0.0065)	(0.0062)
Humidity	-0.0184**	-0.0413**	-0.0323**	-0.0363**
	(0.0005)	(0.0019)	(0.0018)	(0.0017)
Wind speed		-0.0522**	-0.0423**	-0.0144
		(0.0093)	(0.0091)	(0.0088)
Visibility		0.0001**	0.00004*	-0.00002
		(0.00002)	(0.00002)	(0.00002)
Dew point temperature		0.0905**	0.0672**	0.0744**
		(0.0069)	(0.0068)	(0.0065)
Solar Radiation			-0.0044	-0.0034
			(0.0137)	(0.013)
Rainfall			-0.229**	-0.227**
			(0.0077)	(0.0073)
Snowfall			-0.0408*	-0.0066
			(0.0197)	(0.0189)
Spring				0.4713**
				(0.0316)
Summer				0.5073**
				(0.0479)
Autumn				0.8101**
				(0.0335)
Holiday				-0.3636**
				(0.0375)
R-squared	0.5034	0.5177	0.563	0.6064
Observations	8465	8465	8465	8465

Table 2: Variables used in models.

Variable	Letter	Independent/Dependent
Natural Log of Rented Bike Count	y	Independent
Hour	x_1	Dependent
Temperature	x_2	Dependent
Humidity	x_3	Dependent
Wind speed	x_4	Dependent
Visibility	x_5	Dependent
Dew point temperature	x_6	Dependent
Solar Radiation	x_7	Dependent
Rainfall	x_8	Dependent
Snowfall	x_9	Dependent
Spring	d_1	Dependent/Dummy
Summer	d_2	Dependent/Dummy
Autumn	d_3	Dependent/Dummy
Holiday	d_4	Dependent/Dummy

Citation: Yuan Q, Yang Z, Xiao Y (2024) Causal inference of Seoul bike sharing demand. Comput Math Appl. 2(1): 005-009.
DOI: https://dx.doi.org/10.17352/cma.000005

Table 4: Analyzes each variable.

Variable	Analysis
Hour	In all models, the coefficients are statistically significant and economically significant. On average, 1 unit change causes about a 4% increase in Rented Bike Count. So, it is an important factor.
Temperature	In models (1), (2), and (4), the coefficients are statistically significant. In (1) the coefficient is positive but in (2) and (4) the coefficients are negative. We think it is not an important factor.
Humidity	In all models, the coefficients are statistically significant and economically significant. On average, 1 unit change causes about a 3% decrease in Rented Bike Count. So, it is an important factor.
Wind speed	In models (2) and (3), the coefficients are statistically and economically significant. On average, 1 unit change causes about a 4% decrease in the Rented Bike Count. However, in model (4), the coefficient is not statistically significant. We think it is not as important as hour and humidity.
Visibility	Not an important factor since it is not economically significant.
Dew point temperature	In models (2), (3), and (4) the coefficients are statistically significant and economically significant. On average, 1 unit change causes about a 7% increase in Rented Bike Count. So, it is an important factor.
Solar Radiation	Not an important factor since it is not statistically significant.
Rainfall	In models (3), and (4) the coefficients are statistically significant and economically significant. On average, 1 unit change causes about a 20% decrease in Rented Bike Count. So, it is a very important factor.
Snowfall	In model (3), the coefficients are statistically significant. On average, 1 unit change causes about a 4% decrease in Rented Bike Count. However, in model (4), the coefficient is not statistically significant. We think it is not an important factor.
Dummy variables (Spring, Summer, Autumn, Holiday)	All the coefficients are statistically significant and economically significant. So, they are very important factors. Especially for Autumn, I find in Autumn the Rented Bike Count increases by about 120%. It seems that people in the Seoul area like to ride bikes in Autumn.

The insights gained from this study can serve as a foundation for further research and policy decisions aimed at enhancing bike-sharing systems and encouraging sustainable transportation. By understanding the key factors that influence bike rental demand, policymakers and operators can develop targeted strategies to improve system efficiency, user satisfaction, and overall ridership.

Moreover, the causal inference approach employed in this study can be extended to investigate the impact of other potential factors on bike sharing demand, such as the built environment, public transit integration, or socio-economic characteristics of users. As cities continue to grapple with the challenges of congestion, air pollution, and climate change, bike sharing offers a promising solution for promoting active, low-carbon mobility. By leveraging the insights from this research, cities can create more resilient and adaptable bike-sharing systems that contribute to the broader goals of sustainable development.

Declarations

Authors' contributions: Quan Yuan conceived the idea and performed the programming; Zhixin Yang and Yayuan Xiao contributed to the writing and revision of the manuscript.

Acknowledgment

The authors would like to express their gratitude to the reviewers and editors for their valuable comments and suggestions, which have greatly improved the quality of this manuscript.

References

1. Faghih-Imani A, Eluru N, El-Geneidy AM, Rabbat M, Haq U. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. Journal of Transport Geography. 2014; 41: 306–14. doi: 10.1016/j.jtrangeo.2014.01.013
2. Fishman E, Washington S, Haworth N. Bike share: a synthesis of the literature. Transport Reviews. 2013;33(2):148–65. doi: 10.1080/01441647.2013.775612
3. Gebhart K, Noland RB. The impact of weather conditions on bikeshare trips in Washington, DC. Transportation. 2014;41(6):1205–25. doi: 10.1007/s11116-014-9540-7
4. Zhang Y, Thomas T, Brussel M, van Maarseveen M. Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China. Journal of Transport Geography. 2017; 58: 59–70. doi: 10.1016/j.jtrangeo.2016.11.014
5. Frade I, Ribeiro A. Bicycle sharing systems demand. Procedia - Social and Behavioral Sciences. 2014;111:518–27. doi: 10.1016/j.sbspro.2014.01.085
6. Borgnat P, Abry P, Flandrin P, Robardet C, Rouquier JB, Fleury E. Shared bicycles in a city: A signal processing and data analysis perspective. Advances in Complex Systems. 2011;14(03):415–38. doi: 10.1142/S0219525911002950
7. Lin L, He Z, Peeta S. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. Transportation Research Part C: Emerging Technologies. 2018; 97: 258–76. doi: 10.1016/j.trc.2018.10.011
8. Xu C, Ji J, Liu P. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. Transportation Research Part C: Emerging Technologies. 2018; 95, 47-60. doi: 10.1016/j.trc.2018.07.013
9. Pearl J. Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press. 2009. doi: 10.1017/CBO9780511803161
10. Imbens GW, Rubin DB. Causal inference in statistics, social, and biomedical sciences: An introduction. Cambridge University Press. 2015. doi: 10.1017/CBO9781139025751
11. Bao J, He T, Ruan S, Li Y, Zheng Y. Planning bike lanes based on sharing-bikes' trajectories. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017;1377–86. doi: 10.1145/3097983.3098056
12. An R, Zahnow R, Pojani D, Corcoran J. Weather and cycling in New York: The case of Citibike. Journal of Transport Geography. 2019; 77: 97–112. doi: 10.1145/3097983.3098056
13. Caulfield B, O'Mahony M, Brazil W, Weldon P. Examining usage patterns of a bike-sharing scheme in a medium sized city. Transportation Research Part A: Policy and Practice. 2017; 100: 152–61. doi: 10.1016/j.tra.2017.04.023
14. de Chardon CM, Caruso G, Thomas I. Bicycle sharing system 'success' determinants. Transportation Research Part A: Policy and Practice. 2017; 100: 202–14. doi: 10.1016/j.tra.2017.04.020

Citation: Yuan Q, Yang Z, Xiao Y (2024) Causal inference of Seoul bike sharing demand. Comput Math Appl. 2(1): 005-009.
DOI: https://dx.doi.org/10.17352/cma.000005

15. El-Assi W, Salah Mahmoud M, Nurul Habib K. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. Transportation. 2017; 44(3): 589–613. doi: 10.1007/s11116-015-9669-z

16. Wang K, Akar G, Chen Y-J. Bike sharing differences among Millennials, Gen Xers, and Baby Boomers: Lessons learnt from New York City's bike share. Transportation Research Part A: Policy and Practice. 2018; 116: 1–14. doi: 10.1007/s11116-015-9669-z

17. Noland RB, Smart MJ, Guo Z. Bikeshare trip generation in New York City. Transportation Research Part A: Policy and Practice. 2016; 94: 164–81. doi: 10.1016/j.tra.2016.08.030

18. Faghih-Imani A, Eluru N. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. Journal of Transport Geography. 2016; 54: 218–27. doi: 10.1016/j.jtrangeo.2016.06.008

19. Hyland M, Frei C, Frei A, Mahmassani HS. Riders on the storm: Exploring weather and seasonality effects on commute mode choice in Chicago. Travel Behaviour and Society. 2018; 13: 44–60. doi: 10.1016/j.tbs.2018.05.001

20. Guo Y, Zhou J, Wu Y, Li Z. Identifying the factors affecting bike-sharing usage and degree of satisfaction in Ningbo, China. PLoS One. 2017 Sep 21;12(9):e0185100. doi: 10.1371/journal.pone.0185100. PMID: 28934321; PMCID: PMC5608320.

21. Ashqar HI, Elhenawy M, Rakha HA. Modeling bike counts in a bike-sharing system considering the effect of weather conditions. Case Studies on Transport Policy. 2019; 7(2): 261–268. doi: 10.1016/j.cstp.2019.02.011

22. Shen Y, Zhang X, Zhao J. Understanding the usage of dockless bike sharing in Singapore. International Journal of Sustainable Transportation. 2018;12(9):686–700. doi: 10.1080/15568318.2018.1429696

23. Sun F, Chen P, Jiao J. Promoting public bike-sharing: A lesson from the unsuccessful Pronto system. Transp Res D Transp Environ. 2018 Aug;63:533-547. doi: 10.1016/j.trd.2018.06.021. Epub 2018 Jun 28. PMID: 37928131; PMCID: PMC10624502.

24. Böcker L, Anderson E, Uteng TP, Throndsen T. Bike sharing use in conjunction to public transport: Exploring spatiotemporal, age and gender dimensions in Oslo, Norway. Transportation Research Part A: Policy and Practice. 2020; 138: 389–401. doi: 10.1016/j.tra.2020.06.019

25. Li R, Sinniah KG, Li X. The Factors Influencing Resident's Intentions on E-Bike Sharing Usage in China. Sustainability. 2022, 14, 5013. doi: 10.3390/su14095013

26. Morton C. Appraising the market for bicycle sharing schemes: Perceived service quality, satisfaction, and behavioural intention in London. Case Studies on Transport Policy. 2018; 6(1): 102–111. doi: 10.1016/j.cstp.2017.11.003

27. de Chardon MC. The contradictions of bike-share benefits, purposes and outcomes. Transportation Research Part A: Policy and Practice. 2019; 121: 401–19. doi: 10.1016/j.tra.2019.01.031

28. Eren E, Uz VE. A review on bike-sharing: The factors affecting bike-sharing demand. Sustainable Cities and Society. 2020; 54: 101882. doi: 10.1016/j.scs.2019.101882

29. Reck DJ, Axhausen KW. Who uses shared micro-mobility services? Empirical evidence from Zurich, Switzerland. Transportation Research Part D: Transport and Environment. 2021; 94: 102803. doi: 10.1016/j.trd.2021.102803

30. Gu T, Kim I, Currie G. To be or not to be dockless: Empirical analysis of dockless bikeshare development in China. Transportation Research Part A: Policy and Practice. 2019; 119: 122–47. doi: 10.1016/j.tra.2018.11.007

31. UCI Machine Learning Repository. Seoul Bike Sharing Demand Data Set. UCI Machine Learning Repository" 2020. http://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand doi: 10.24432/C5F62R

32. Wooldridge JM. Introductory Econometrics: A Modern Approach. 5th ed. Mason, OH: South-Western Cengage Learning. 2013. https://archive.org/details/introductoryecon0000wool_c3l8.

Citation: Yuan Q, Yang Z, Xiao Y (2024) Causal inference of Seoul bike sharing demand. Comput Math Appl. 2(1): 005-009.
DOI: https://dx.doi.org/10.17352/cma.000005