BUAN 6312.501 - Applied Econometrics and Time Series Analysis

Impact of environmental and seasonal factors on bike rental demand

Submitted to:

Prof. Quanquan Liu

Naveen Jindal School of Management

Submitted by:

Group 1

Abhishek Dubey	AXD220145
Aditi Tiwari	AXT230041
Aparna Mishra	AXM230011
Jayesh Saini	JXS230007
Jefferson Nguithi Kimani	JNK230000
Manmohan Dash	MXD220045
Neeti Mishra	NXM230008
Santhosh Sekar	SXS230156
Umang Singhal	UXS230011

Introduction

The concept of bike-sharing systems has turned urban transportation into an eco-friendly and efficient mode of transport. These systems address the increasing demand for eco-friendly transit, congestion, and pollution challenges that characterize urban settings. However, bike rentals depend on various factors like environmental conditions and seasonal variations, posing a challenge in effective resource planning and operational management.

While the earlier studies have identified the role of some variables, such as temperature and precipitation, in determining the demand for bike rentals, for instance, higher temperatures are associated with increased outdoor activities and hence more rentals, whereas extreme weather conditions, heavy rain, or strong winds discourage use. However, an integrated econometric analysis that incorporates environmental and seasonal factors to quantify their combined effect on bike rental patterns has yet to be done.

This study tries to fill this lacuna by relying on the UCI Bike Sharing Dataset, which contains over 17,000 observations ranging in attributes from temperature and humidity to wind speed and holiday/workday. Using econometric approaches such as multiple linear regressions, including hypothesis testing, this paper tries to bring out the essential predictors of the demand for bike rentals and give actionable insights. These will further offer insights into the decision of bike-sharing companies and, therefore, further operational efficiency for probably other industries reliant on the use of short-distance transport.

Literature Review

Bike-sharing systems have become an important topic in research because they help with urban transportation, environmental issues, and public health. Many studies have looked at how this data can be used to analyze patterns in city mobility, detect unusual events, and improve urban planning. Our study builds on these ideas and adds new insights.

Connection to Previous Research: A key paper which was published by Fanaee-T and Gama (2014) focuses on detecting and labeling events in bike-sharing data. They used methods like ensemble methods that will combine with external data sources like search engines, to identify unusual events. Their main goal was to find anomalies and unusual patterns. They faced challenges like false alarms and reliance on specific detection methods.

While Fanaee-T and Gama's work is important for event detection, it focuses more on identifying anomalies rather than exploring the patterns and trends in bike rentals under normal conditions.

Our Contribution: Our study builds on their work but looks at the data from a different angle. Instead of just finding anomalies, we focus on understanding what influences bike rentals on a daily basis and predicting future trends. Here’s how our work adds to the existing research:

- Improved Data Analysis: We include more weather-related details and normalize the data to better understand how weather, seasons, and other factors affect bike rentals.
- Seasonal Trends: We focussed on how bike usage changes with seasons and time

Our study focuses on predicting bike usage trends instead of just identifying unusual events. This makes our findings more useful for improving bike-sharing services, designing better city infrastructure, and promoting sustainable transportation.

In simple terms, while we build on previous research, our work goes further by focusing on everyday bike usage and providing tools to help cities make better decisions about bike-sharing programs.

Data

This dataset contains daily bike rental counts from the Capital Bikeshare system for the years 2011 and 2012. It includes various factors that influence bike usage, such as weather, seasons, and time, making it a great resource for understanding how people move around cities and how different conditions affect bike rentals.

What the Data Includes:

- **Time Details:** Information about the year, month and day to analyse rental patterns over time.
- **Weather Data:** Details like temperature, humidity, wind speed, and weather conditions (e.g., sunny, cloudy, or rainy) to see how weather affects bike usage.
- **Seasonal Effects:** Data grouped by seasons (spring, summer, fall, winter) to explore changes in usage throughout the year.
- **Rental Numbers:** The number of bikes rented each day, separated into casual users (non-members) and registered users (members).
- **Holidays and Workdays:** Whether a specific day is a holiday or a regular workday, helping to understand social influences on bike rentals.

Key Facts on the dataset:

- **No Missing Values:** The dataset is complete and ready to use—no need to clean it for missing information.
- **No Duplicate entries**

Data Preprocessing

We cleaned and transformed the dataset using the steps below:

Loading the Data:

The data was loaded into a pandas dataframe to easily work with it.

Dealing with Categorical Data:

Columns like season, month, weekday, and weather were turned into multiple columns using one-hot encoding. For example, the season column was split into new columns like season_1, season_2, etc., with values of 1 or 0 to show which category applies to each row.

Making the Data Consistent:

The new one-hot encoded columns were made sure to only contain values of 0 or 1 for clarity and consistency.

Scaling the Bike Count:

The bike count column which is our target variable was scaled using Min-Max Scaler to a range between 0 and 1. This was done to ensure the values are consistent and easier to use in models that rely on scaled data.

Saving the Clean Data:

The cleaned and processed data was saved in two formats:

- A CSV file (day_preprocessed.csv) for general use.
- A Stata file (day_preprocessed.dta) for compatibility with Stata software.

These steps ensure the dataset is clean, easy to understand, and ready for further analysis which was done on STATA.

Empirical Methods

Multiple Linear Regression Analysis

- **Purpose**: To quantify the effect of independent variables (temperature, humidity, wind speed, seasonal indicators, and weather conditions) on the dependent variable (daily bike rentals).
- **Model Specification**: Ordinary Least Squares (OLS) regression is used, expressed as:
    - Bike Rentals = -0.35 + 0.77 * temp - 0.36 * hum - 0.41 * windspeed - 0.07 * holiday + 0.11 * season_2 + 0.08 * season_3 - 0.17 * season_4 + 0.00 * mnth_2 - 0.01 * mnth_3 - 0.01 * mnth_4 + 0.00 * mnth_5 + 0.02 * mnth_6 - 0.12 * mnth_7 - 0.05 * mnth_8 - 0.04 * mnth_9 - 0.01 * mnth_10 - 0.03 * mnth_11 + 0.02 * weekday_1 + 0.03 * weekday_2 + 0.04 * weekday_3 + 0.03 * weekday_4 + 0.04 * weekday_5 + 0.03 * weekday_6 - 0.27 * weathersit_2 - 0.22 * weathersit_3 + u

Result

Our analysis is focused on understanding the key factors that will influence the bike rentals and how they can be used to predict future trends. Using multiple regression, we examined the effects of weather, seasons, and time-related variables on bike usage. Below are the main takeaways from our findings.

Model Performance

Our model performed well, with R-squared equals 58.84% which is explaining about of the variation in bike rentals. This means the factors we analyzed—like weather, seasons, and time—have a strong impact on bike usage. After adjusting for the number of variables in the model, the adjusted R-squared is 57.32%, showing that the model remains reliable even with multiple factors included.

```
. reg cnt temp hum windspeed holiday season_2 season_3 season_4 mnth_2 mnth_3 mnth_4 mnth_5 mnth_6 mnth_7 mnth_8 mnth_9 mnth_10 mnth_11 mnth_12 weekday_1 weekday_2 weekday_3 weekday_4 wee
> weekday_6 weathersit_2 weathersit_3
```

Source	SS	df	MS			
				Number of obs	=	731
				F(26, 704)	=	38.70
Model	21.3351394	26	.820582286	Prob > F	=	0.0000
Residual	14.9256688	704	.021201234	R-squared	=	0.5884
				Adj R-squared	=	0.5732
Total	36.2608082	730	.04967234	Root MSE	=	.14561

cnt	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
temp	.7726386	.0769717	10.04	0.000	.6215171	.9237601
hum	-.3562541	.0546155	-6.52	0.000	-.4634828	-.2490254
windspeed	-.4093587	.076778	-5.33	0.000	-.5601	-.2586173
holiday	-.0731707	.0340712	-2.15	0.032	-.1400641	-.0062773
season_2	.1089069	.0339662	3.21	0.001	.0422196	.1755942
season_3	.0842354	.0403383	2.09	0.037	.0050376	.1634333
season_4	.1729981	.0342536	5.05	0.000	.1057466	.2402497
mnth_2	.0006175	.0271901	0.02	0.982	-.0527659	.054001
mnth_3	.022758	.0312544	0.73	0.467	-.0386049	.084121
mnth_4	-.0104636	.0467579	-0.22	0.823	-.1022652	.081338
mnth_5	.0015735	.0504752	0.03	0.975	-.0975264	.1006734
mnth_6	-.0579482	.0530401	-1.09	0.275	-.1620839	.0461875
mnth_7	-.1153496	.0590426	-1.95	0.051	-.2312703	.0005711
mnth_8	-.0536435	.0568496	-0.94	0.346	-.1652585	.0579715
mnth_9	.0467746	.0499864	0.94	0.350	-.0513656	.1449149
mnth_10	.021549	.0456856	0.47	0.637	-.0681474	.1112454
mnth_11	-.0288041	.0436679	-0.66	0.510	-.1145391	.056931
mnth_12	-.0153707	.0344799	-0.45	0.656	-.0830665	.052325
weekday_1	.0224243	.0207206	1.08	0.280	-.0182573	.063106
weekday_2	.0303566	.0202772	1.50	0.135	-.0094544	.0701677
weekday_3	.0397922	.0203339	1.96	0.051	-.0001302	.0797146
weekday_4	.0342191	.0203471	1.68	0.093	-.0057292	.0741673
weekday_5	.0406223	.0202909	2.00	0.046	.0007843	.0804603
weekday_6	.0463488	.0201678	2.30	0.022	.0067525	.085945
weathersit_2	-.0272187	.014532	-1.87	0.061	-.05575	.0013125
weathersit_3	-.2245019	.0372132	-6.03	0.000	-.297564	-.1514398
_cons	.3464778	.0437407	7.92	0.000	.2606	.4323556

Weather Effects

- **Temperature:** Warmer weather encourages bike usage, as shown by its positive coefficient of 0.772. Higher temperature is leading to more bike rentals which is likely because it makes riding more comfortable.
- **Humidity:** High humidity discourages bike rentals, as it can make riding unpleasant. It is shown by its coefficient of -0.356.
- **Wind Speed:** Strong winds reduce the bike usage as windy conditions will make riding more difficult as shown by its coefficient of -0.409.

Seasonal Trends

Seasons play an important role in bike rentals:

- **Spring:** Rentals increase during spring, as the weather becomes more favorable for outdoor activities.
- **Fall:** Fall shows the highest increase in bike rentals, likely due to cooler temperatures and enjoyable riding conditions.

Summer also sees more rentals compared to winter, but the effect is less noticeable than in spring and fall.

Monthly and Weekly Patterns

Bike rentals vary across months and days:

- **July and August:** Rentals slightly decrease during these months, possibly due to extreme heat or people going on vacations.
- **Fridays:** Rentals tend to be higher on Fridays compared to other weekdays, likely because people use bikes for leisure as the weekend approaches.

Impact of Holidays

- **Holidays:** Bike rentals drop slightly on holidays, which could be because fewer people commute to work or prefer other activities on those days.
- **Working Days:** Weekdays show consistent usage, driven by commuting and regular activities.

Weather Conditions

The type of weather also matters:

- Clear and sunny weather will lead to more rentals, while light snow or rain significantly reduce usage.

Key Insights

1. **Temperature and Seasons:** Warmer weather and seasons like spring and fall leads to higher bike rentals usage.
2. **Weather Challenges:** Unfavourable weather challenges like high humidity, strong winds, and rain will lead to lower bike rentals usage.
3. **Social Factors:** Bike usage patterns align well with commuting habits and leisure activities, with more rentals on working days and Fridays.

In our analysis, most coefficients make sense and match what we expected. For example, bike rentals increase when it's warmer (positive temperature coefficient) and decrease with higher humidity or windspeed (negative coefficients). Many of these relationships are statistically significant, meaning they likely aren't due to chance. However, some variables aren't significant. This is due to overlapping effects between variables.

Conclusion

The bike-sharing dataset analysis has shown some important determinants of rental demand. Among the environmental variables, temperature was the most important and positively correlated with rentals, while humidity and windspeed were negatively correlated, as intuitively expected. Seasonal and monthly trends were also important. Spring-summer with peak months like May and August, contributed to the increased demand. Winter conditions reduced the usage of bike rentals. Rain or snow also did the same thing. This suugests that good weather was an essential ingredient for biking. Countering adverse weather and seasonal downturns, some possible strategies involve testing demand for alternative services-for example, cars or autonomous rides like Waymo-improving bike station availability for shorter-distance trips and offering proper winter wear with bike rentals to appeal to customers in colder months.

The results also indicate that middle-range temperatures may attract more riders, thus the need to emphasize marketing efforts during such periods. Further bike-sharing infrastructure expansion in high-demand zones will enable better access and serve more localized customer needs. A data driven approach can help in predicting the impact of weather on usage. During winter selective promotions and partnerships to provide winter biking gear could encourage usage. Additionally, incentivizing the use of bike-sharing for short trips in urban areas during adverse weather can help retain customers and be competitive against other transportation options such as Uber or Waymo. The results are very useful for integrating environmental, seasonal, and behavioural factors into decision-making processes for sustainable growth in the bike-sharing market.

Appendix

Exploratory Data Analysis

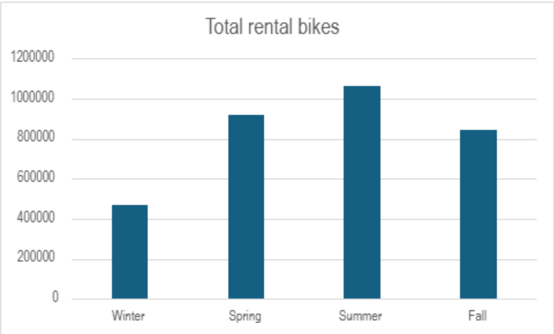

Figure 1

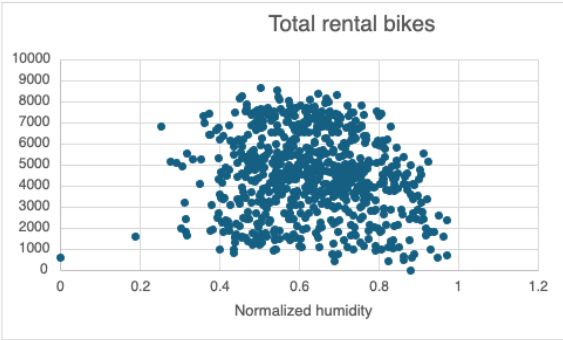

Figure 2

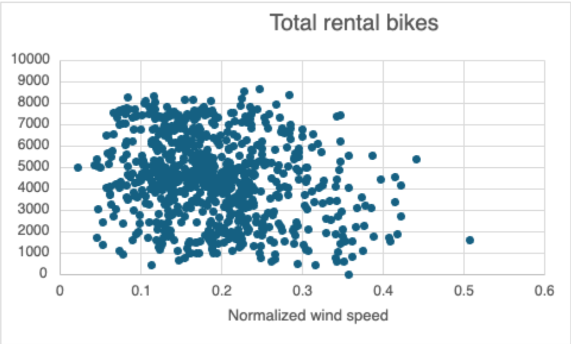

Figure 3

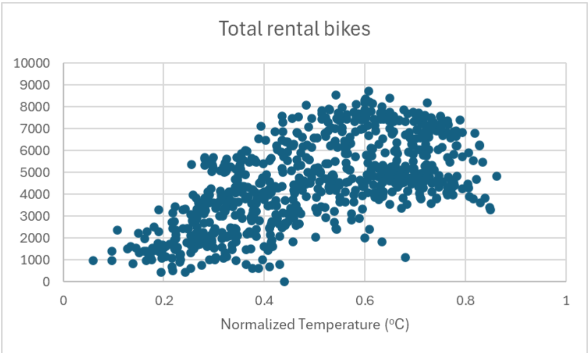

Figure 4

Figure 1: Total Rental bikes distributed by season

Figure 2: Total Rental bikes vs Normalized Humidity

Figure 3: Total Rental bikes vs Normalized Wind speed

Figure 4: Total Rental bikes vs Normalized Temperature (ºC)

References

- Fanaee-T, H., & Gama, J. (2013). Event labeling combining ensemble detectors and background knowledge