Bike Sharing Project

Content Page

1. Introduction

Bike-sharing is the act of individuals renting bicycles from a central system, where they pick the bikes up from one location and return them at their destination. (Moon et al., 2020) The first bike-sharing system, "Witte Fietsen", started off in Amsterdam in 1965, where white bikes were placed on the streets for public use. (Smart Cities Dive) Over the last ten years, this bike-sharing system has developed into a new mode of public transportation. Renting bikes has been increasingly popular for the first-mile/last-mile connection to other modes of transportation. Bike sharing has significant effects on reducing greenhouse gas emissions, promoting public health, and mitigating traffic issues. These environmental and health benefits have appealed widely to the public and today, governments are finding ways to utilise this system and make it more accessible. (E et al., 2020) Given Bike-sharing's potential environmental and health benefits, it is in the government's interest to increase bike sharing practices among its population. This report will be focused on evaluating the factors of affecting rented bikes and its potential implications.

2. Dataset

The source of our data was taken from the UCI website based on Seoul's databanks Bike Sharing data. Apart from bike usage data, weather information is also included in the dataset by UCI. The dataset includes weather information (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information.

2.1 Variable Description

Refer to Appendix A

2.2 Data Preparation

Data within the winter season was removed as the bike sharing data will be exceptionally low in winter. The variable *Snowfall* is similarly removed given their obvious impact on bike usage. Rows with non-functioning days had no valuable data and were hence also removed.

For our hypothesis, two separate datasets were prepared for evaluation.

Dataset 1: Model for Environmental Analysis (For Hypothesis 1)

This model analyses the relationship between environmental factors and the usage of bikes. The original data was consolidated to provide average daily values rather than hourly values. Variables that were unnecessary (such as date) or were highly correlated to other variables (such as dew point temperatures) were removed. The season variables were converted to dummy variables, with the exception of the winter season.

Dataset 2: Model for time analysis (For Hypothesis 2)

This model analyses the relationship between hours and the usage of bikes. The data set used for model 2 uses the original data that was grouped by the hour.

2.3 Variable Investigation (EDA)

<u>Correlation Test</u>

Dewpoint temperature was correlated with temperature at a value of 0.95, as seen in Figure 1, and as such the variable was removed to prevent multi-correlation. The decision to remove dewpoint temperature instead of the regular temperature variable was because temperature is a more commonly understood metric, which will allow our work to be more readable and interpretable. Furthermore, windspeed and visibility have very low correlation with the number of rented bikes, -0.09 and 0.35 respectively, as such, we removed these from our multiple linear regression model.

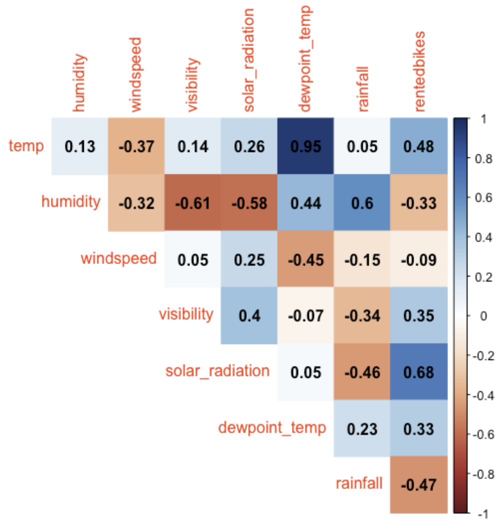

Figure 1: Multicollinearity Matrix

<u>Normality and Outlier Investigation</u>

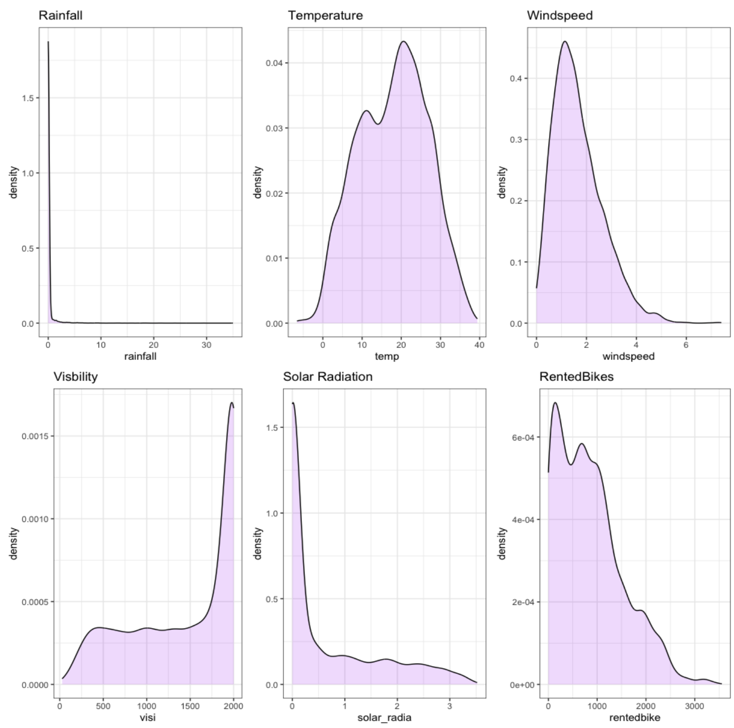

Figure 2: Density Plot for Continuous Variables

From the density plot in Figure 2, *solar_radia* and *rainfall* are extremely skewed and deviates away from a normal distribution. Hence, the team will use log transformation technique to standardize both variables.

3. Hypotheses

3.1 Hypothesis 1

Environmental factors affect the demand for bikes.

By evaluating the relationship between bike rentals and environmental conditions, we can identify causality and trends between these independent variables and the dependent variable. This information will be very helpful when discussing policies that increase the availability of bikes to the public, leading to better physical, social, and mental health.

3.1.1 Regression Modelling (Model 1)

$$\begin{aligned} rentedbikes = {} & \beta_0 + \beta_{temp} temp + \beta_{humidity} humidity + \beta_{\log\_solar\_radi} \log\_solar\_radi \\ & + \beta_{\log\_rainfall} \log\text{\_rainfall} + \beta_{summer} summer + \beta_{Spring} Spring \\ & + \beta_{is_{holiday}} is\_holiday + \varepsilon \end{aligned}$$

Referring to Appendix B for the summary of Model 1, the multiple R-squared value is 0.6576. It can be interpreted that the R squared value by stating that changes in the X variables explain 65.76% of changes in the number of rented bikes. This shows that the model has high explanatory power.

3.1.2 Causal Inference

Table 1: Statistically significant variables for Model 1

Temperature (*[1])	For every increase in temperature by **1 ºC**, the number of rented bikes increases by **10.62**, on average, holding all other variables constant. This is most likely due to people preferring to cycle in warmer weather conditions.
Log Solar Radiation (***[2])	A **1%** increase in solar radiation increases the number of bikes rented by **2.95**, on average, holding all other variables constant. This is most likely due to individuals wanting to cycle in safe conditions when it is bright.
Log Rainfall (*)	A **1%** increase in rainfall decreases the number of bikes rented by **0.46**, on average, holding all other variables constant. This is likely due to individuals not wanting to cycle in slippery conditions.
Is_holiday (*)	On holidays, the number of bikes rented are **300 less** compared to weekdays. This is most likely due to individuals renting bikes for work, which they do not do on weekends (generally).

Table 2: Non statistically significant variables for Model 1

Humidity	Generally, people do not see humidity as a factor for renting a bike, in Seoul, South Korea.
Summer	There is no difference between the number of bikes rented in summer compared to autumn.
Spring	There is no difference between the number of bikes rented in spring compared to autumn.

3.1.3 Gauss-Markov Assumption

To ensure that Model 1 fulfils the Best Linear Unbiased Estimator (BLUE), the analysis of the 5 Gauss-Markov assumptions is done below.

1. <u>Assumption of Linear Relationship between independent and dependent variables.</u>

Linearity and normality satisfied

[1] * indicates 0.05 significance level.

[2] *** indicates 0.001 significance level.

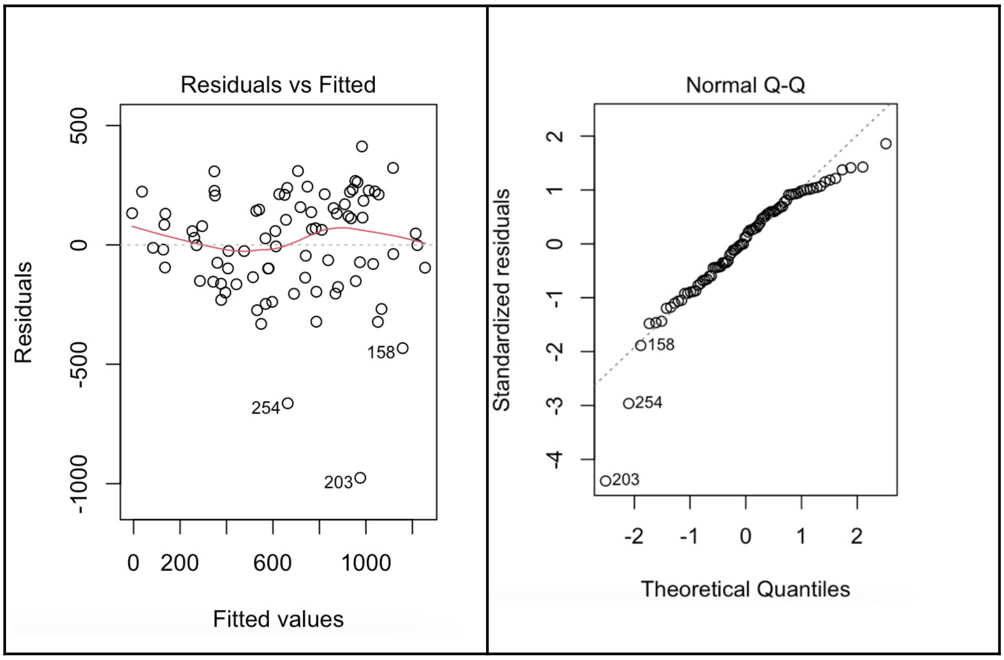

Figure 3: (a) Linearity Test for Model 1 (b) Normality Test for Model 1

As seen in Figure 3 (a), it can be observed that the trend shown by the red line does follow the horizontal zero line, additionally, the data points on the Q-Q plot follow the dotted straight line, as seen in Figure 3 (b).

2. Multicollinearity assumption

Multicollinearity is satisfied (as stated previously).

3. Homoscedastic assumption

Goldfeld-Quandt test output:

```
            Goldfeld-Quandt test

data:  model
GQ = 2.5653, df1 = 33, df2 = 33, p-value = 0.004163
alternative hypothesis: variance increases from segment 1 to 2
```

Figure 4: Goldfeld-Quandt Test for Homoscedasticity for Model 1

H0: Data is homoscedastic
H1: Data is heteroscedastic

P-value = 0.003836 < 0.05. So we have enough evidence to reject H0 and claim that the data is heteroscedastic.

4. No autocorrelation assumption

```
lag Autocorrelation D-W Statistic p-value
  1       0.1967373      1.600019   0.028
Alternative hypothesis: rho != 0
```

Figure 5: Autocorrelation Test for Model 1

H0: Autocorrelation doesn't exists
H1: Autocorrelation exists.

P-value = 0.003836 < 0.05. So we have enough evidence to reject H0 and claim that the data is autocorrelation.

5. Exogeneity assumption

Assumptions of Exogeneity will be explained under sources of exogeneity.

3.1.4 Evaluation

Out of the 4 statistically significant factors (temperature, solar radiation, holiday, rainfall), the holiday variable had the greatest effect on the number of bikes rented; there were significantly more bikes during weekdays than weekends. We can infer that the people in Seoul use cycling as a means of transportation as opposed to leisure (generally). Increasing incentives for cycling during weekdays, such as improving cycling roads and increasing the number of electronic bikes produced would most likely increase the number of people choosing to cycle as their mode of transportation.

Furthermore, temperature had the second greatest effect; that people prefer to cycle in warmer conditions. The government could forecast the future temperatures, increase the number of bikes available and reduce the prices for rental when the temperature is sufficiently warm enough.

These potential policies would hopefully lead to an increase in general health of individuals and boost their productivity at work. We can conclude that environmental factors do influence the number of bikes rented.

3.2 Hypothesis 2

Which time frames are significant in affecting the demand of rented bikes?

3.2.1 Regression Modelling (Model 2)

Investigating the causal relationship between the usage of bikes by the hour will allow us to see the trends in the demand for bikes throughout the day. Evaluating the increase in bike usage at a certain hour will allow us to look into the reasons that could explain the increase in demand as compared to certain hours. This could help determine if the implementation of public bike sharing in Seoul has reached its goals and objectives.
To investigate this causal relationship, the team has performed an OLS regression model to predict the demand of rented bikes with hours of the day as the predictors and adding environmental factors and seasons that were statistically significant in the previous hypothesis as controls in the model.

The equation of the model is given as:

$$
\begin{aligned}
rentedbikes = {} & \beta_0 + \beta_{hour1} hour1 + \beta_{hour2} hour2 + \beta_{hour3} hour3 + \beta_{hour4} hour4 \\
& + \beta_{hour5} hour5 + \beta_{hour6} hour6 + \beta_{hour7} hour7 + \beta_{hour8} hour8 + \beta_{hour9} hour9 \\
& + \beta_{hour10} hour10 + \beta_{hour11} hour11 + \beta_{hour12} hour12 + \beta_{hour13} hour13 \\
& + \beta_{hour14} hour14 + \beta_{hour15} hour15 + \beta_{hour16} hour16 + \beta_{hour17} hour17 \\
& + \beta_{hour18} hour18 + \beta_{hour19} hour19 + \beta_{hour20} hour20 + \beta_{hour21} hour21 \\
& + \beta_{hour22} hour22 + \beta_{hour23} hour23 + \beta_{\log\_solar\_radi} \log\_solar\_radi \\
& + \beta_{\log\_rainfall} log\_rainfall + \beta_{temp} temp + \beta_{summer} summer + \beta_{Spring} Spring + \varepsilon
\end{aligned}
$$

Referring to Appendix C for the summary of the regression model, the Multiple R-Squared of the model is 0.5719, which indicates that 57.19% of the variation in rented bikes, y, can be explained by the model. Given that the purpose of the hypothesis is to look at the causal relationship of rented bikes based on the hour, 57.19% provides reasonable explanatory power and we can therefore accept the model.

Furthermore, analysing the p-value of each regressor shows that *hour7*, *hour16* and *summer* are not statistically significant.

3.2.2 Causal Inference and Evaluation

The coefficient of the models of the model indicates the increases/decrease in the predicted demand in rented bikes, holding all other variables constant. For example, since the coefficient of hour1 is -157.055, we can interpret that relative to hour0 in season = Autumn, the demand of bikes will decrease by 157.055, holding all other variables constant. Hence, we investigate how each hour period will affect the demand of rented bikes by investigating the coefficients of the hour regressors.

Analysing the coefficients of the hours, there was a clear trend of time periods where the demand for rented bikes increases/decreases during a certain timeframe. To visualise this trend, we plotted a line plot of coefficients of regressors throughout the day, from 12am to 12am the next day.

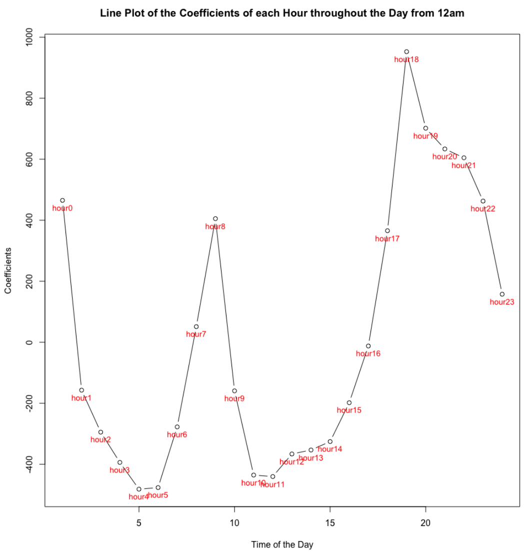

Figure 6: Line Plot of the Coefficient of the Hours Regression throughout the day

Based on Figure 6, observations can be made by the hours as below.

From hour1 to hour5 **(12am to 5am)**, there is a **decreasing trend** in the demand for rented bikes.

From hour6 to hour8 **(5am to 8am):** There is an **increasing trend** in the demand for rented bikes. From our research, bike sharing stations are often in high traffic areas like in front of offices and transportation stations. The increase in the demand for rented bikes can be attributed to commuters using bike sharing to get to their workplace in the early moas rning, either from their area of residence to transportation hubs or from transportation stations to their workplace. This goes to show that the government's objective of encouraging more environmentally friendly commuting methods were effective during the peak traffic timings of the day.

From hour9 to hour 10 **(8am to 10am):** Conversely, there was a **drastic decrease** in the demand for rented bikes relative to the (5am to 8am). This could be so as peak traffic has decreased and many would use normal public transportation as the traffic is lighter.

From hour15 to hour18 **(3pm to 6pm)**: There is an **increasing trend** in the demand for rented bikes. This can be attributed to the usage of bikes to get back from the users' workplace. Furthermore, the coefficient of hour18 is also much higher than the coefficient of hour8. This can be attributed as many users use the bike to commute back to their residence or use it for leisure purposes.

From hour19 to hour23 **(7pm to 11pm):** The demand for rented bikes **remains high** relative to hour0. This can be attributed to the users using bike sharing for leisure purposes such as cycling as a form of exercise.

The causal inferences from the regression model have shown that there are certain time frames during the day where the usage of bike sharing is higher/lower. The increase in the usage of bike sharing during the early commute hours and evening timing has shown that the government's objectives of encouraging more environmentally friendly commuting methods have been effective to a certain degree.

3.2.3 Gauss-Markov Assumptions

To ensure that Model 2 fulfils the Best Linear Unbiased Estimator (BLUE), the analysis of the 5 Gauss-Markov assumptions is done below.

1. <u>Assumption of Linear Relationship between independent and dependent variables.</u>

Linearity and Normality not satisfied.

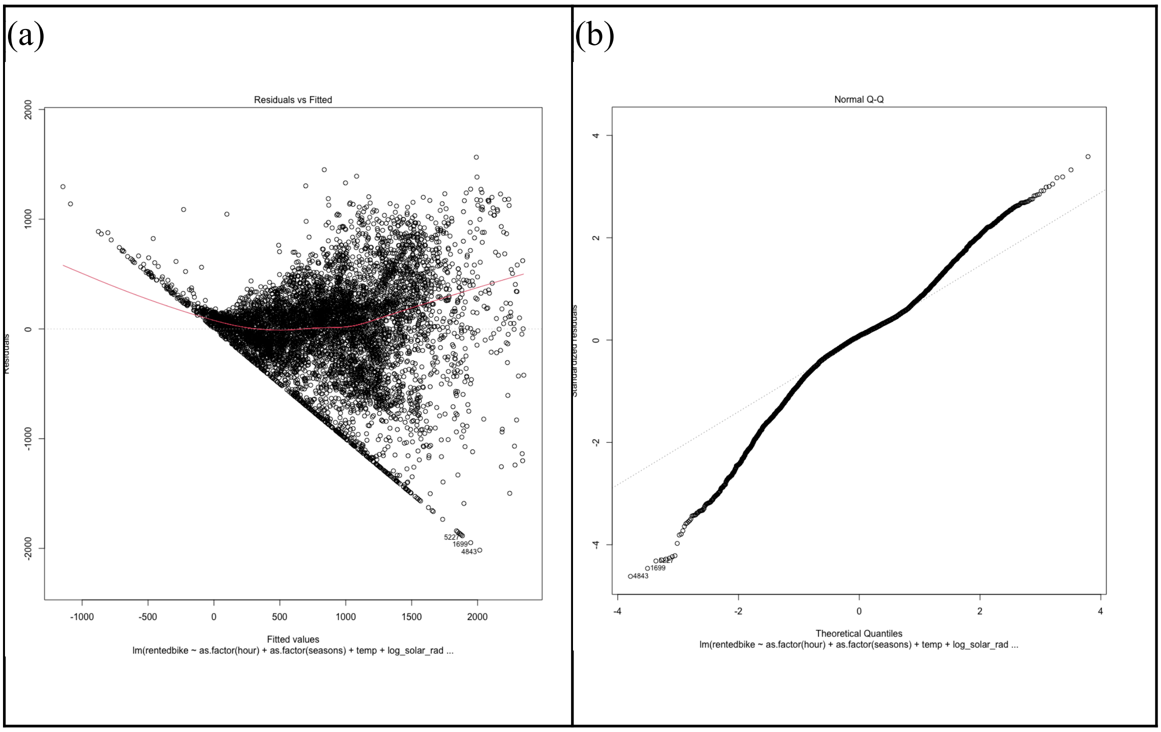

Figure 7: (a) Linearity Test for Model 1 (b) Normality Test for Model 1

As seen in Figure 7 (a), it can be observed that the trend shown by the red line does not follow the horizontal zero line, additionally, the data points on the Q-Q plot do not follow the dotted straight line, as seen in Figure 7 (b).

2. <u>Multicollinearity</u>

Multicollinearity is satisfied (as stated previously).

3. <u>Homoscedasticity</u>

The assumption of Homoscedasticity looks at the equal variance of residuals. Performing the Goldfeld-Quandt test, as seen in Figure 8, on the model above gives us a p-value of 0.03442. Given that the p-value is lesser than 0.05, we reject the null hypothesis that Homoscedasticity is present. Thus, the model does not satisfies the assumption of Homoscedasticity.

```
	Goldfeld-Quandt test

data:  model_hour
GQ = 1.1509, df1 = 3271, df2 = 3271, p-value = 2.95e-05
alternative hypothesis: variance increases from segment 1 to 2
```

Figure 8: Summary of Goldfeld-Quandt Test for Homoscedasticity for Model 2

 4. Assumption of Autocorrelation

Autocorrelation is defined by whether the dependent variables are correlated/inter-related to one another. Performing the Durbin-Waston test, as seen in Figure 9, we get a p-value of 0 which is smaller than 0.05. Hence, we have enough evidence to reject the null hypothesis that there is no autocorrelation. Hence, this implies that autocorrelation is present in the model.

Autocorrelation in the model could be due to the time series nature of the data as each row represents data at a certain hour. Hence, given that there is a trend in the demand for bikes throughout the day, there exists a pattern in each 24-hour block of data.

```
lag Autocorrelation D-W Statistic p-value
  1       0.8361066     0.3277669       0
Alternative hypothesis: rho != 0
```

Figure 9: Summary of Durbin-Waston Test for Model 2

 5. Assumption of Exogeneity

Assumptions of Exogeneity will be explained under sources of exogeneity.

3.2.4 Sub Hypothesis 2.1

Is there a difference in the trend of the usage of bikes between weekdays and weekends?

After understanding the causal relationship between hours and usage of rented bikes, the team also explored if there was a significant difference between the usage of rented bikes between weekdays and weekends as leisure usage and commuting patterns could change between weekdays and weekends. Hence, two models were ran based on two different subsets of the data, one for weekdays and one for weekends. The summary of the regression model for the weekday can be found in Appendix D while the summary of the regression model for weekend can be found in Appendix E.

To investigate if there is a difference in usage patterns, the coefficients of the hour regressors are used as the mode of comparison. To visualise the difference in its coefficient, a line plot is used.

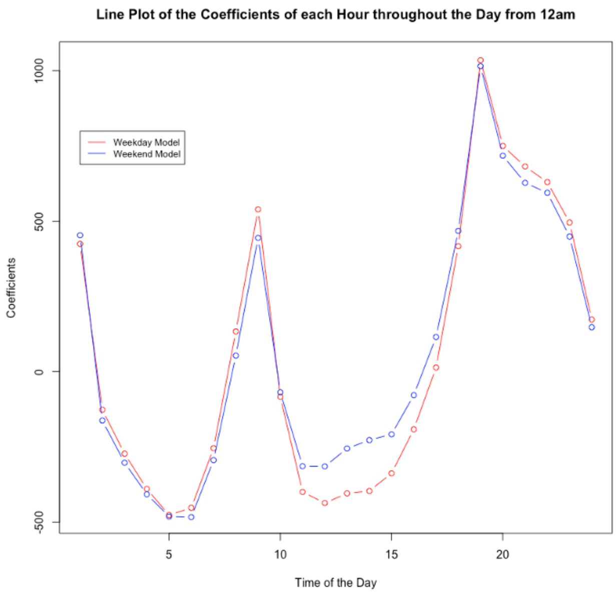

Figure 10: Line plot of the coefficient of the hour regressors for Weekday and Weekend models

Visually, from Figure 10, there is no difference in the coefficients of the hour regressors as the trend remains similar throughout the day between weekdays and weekends. Even though the team did not perform a cross model coefficient test to evaluate if the coefficients are statistically different as it is beyond the scope of this module, the line plot is sufficient to imply that there is no significant difference in the usage of bikes between weekdays and weekends.

3.4 Possible Sources of Exogeneity - Limitations

Omitted variable bias:
A potential omitted variable could be the age of the individual as people who are younger may be more likely to rent a bike. Another potential omitted variable could be the receptiveness of the South Korean public to the policy. As South Koreans come from a work centric society, the perception of the public towards bike rental could be skewed towards the convenience of commuting instead of the purpose of leisure.
Reverse Causality:
It is highly unlikely that the number of rented bikes would affect the independent variables in the model. For example, the number of rented bikes cannot affect solar radiation values or change the season. For our model, it is highly unlikely that reverse causality is a source of endogeneity.
Errors-in-variable bias:
There could be measurement errors that change the recorded value away from the true value. Given that the data was taken from Seoul's government database, the error-in-variable bias should be highly unlikely.

4. Policies and Implications

There are implications and policies that could be drawn from our investigation. Firstly, governments could collaborate with businesses to incentivise them to increase the cycling experience for the populations. An example would be by giving grants to businesses that make cycling easily accessible to all individuals, for example providing electric bikes. Another example includes providing subsidies or discount vouchers for renting bikes to the population to encourage more cycling habits. Workplaces can offer their employees similar subsidies or vouchers for renting bikes to incentivise them to cycle between their home and workplace.

Governments can also take action, independent of businesses, by increasing the amount of bikes accessible to individuals in high demand hours (morning and evening commutes). They could also achieve this by re-distributing the bikes available to the public during low demand hours (non-commute hours).

Finally, the government should promote the use of bikes in warm and sunny conditions by offering incentives to individuals, such as money back based on the number of miles they cycle.

5. Conclusion

In conclusion, there are significant timeframes that affect the increase in the usage of bikes. Namely, the early morning and evening commuting period saw a steep increase in the usage of bikes. Usage of bikes is also relatively much higher during the evening, potentially due to leisure usage as a form of exercise. This shows that the government implementation of bike sharing has been successful in meeting their objectives of a greener commuting alternative. Further promotions to encourage the usage of bike sharing can further increase the uptake of bike sharing as the increase in usage has evidently shown that users had found benefits in bike sharing.

Furthermore, the generalisation on the causal inference on how environmental factors can affect the demand of rented bikes, can be applied to other countries as well. Metropolitan cities with similar climates and seasons can utilise these causal inferences in its implementation of bike sharing policies. By understanding how climate affects the demand of bikes, countries can better adapt to differences in the demand of bikes.

6. References

Moon, C., Sharpin, A. B., Lanza, I. D. L., Khan, A., Re, L. L., & Maassen, A. (2020, February 1). *The evolution of bike sharing: 10 questions on the emergence of new technologies, opportunities, and risks*. World Resources Institute. Retrieved April 12, 2023, from https://www.wri.org/research/evolution-bike-sharing#:~:text=Bike%20sharing%20began%20in%20the,expanded%20over%20the%20last%20decade.

Don't miss Tomorrow's Smart Cities Industry News. Smart Cities Dive. (n.d.). Retrieved April 12, 2023, from https://www.smartcitiesdive.com/ex/sustainablecitiescollective/amsterdam-beijing-global-evolution-bike-share/1100421/#:~:text=In%201965%2C%20Amsterdam%20implemented%20what,the%20streets%20for%20public%20use.

E, S. V., Park, J., & Cho, Y. (2020, February 6). *Using data mining techniques for bike sharing demand prediction in Metropolitan City*. Computer Communications. Retrieved April 12, 2023, from https://www.sciencedirect.com/science/article/abs/pii/S0140366419318997

(n.d.). 서울 열린데이터광장. Retrieved April 12, 2023, from http://data.seoul.go.kr/

Seoul Bike Sharing Demand Data Set. (2020, March 1). UCI Machine Learning Repository. Retrieved April 12, 2023, from https://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand

Appendix

Appendix A: Variables from the original dataset

Variable Name	**Description**
Date	The date when the data was collected
Rented Bike Count	The number of bikes rented for that hour
Hour	The hour of the day, numbered from 0 to 23
Temperature (℃)	The temperature in ℃ for that hour of that date
Humidity (%)	The humidity percentage for that hour of that date
Wind Speed (m/s)	The wind speed for that hour of that date
Visibility (10m)	The visibility for that hour of that date
Dew point temperature (℃)	The dew point temperature for that hour of that date
Solar radiation (MJ/m2)	The solar radiation for that hour of that date
Rainfall (mm)	The amount of rainfall for that hour of that date
Snowfall (cm)	The amount of snowfall for that hour of that date
Seasons	The current season of that date as Spring, Summer, Autumn, Winter
Holiday	Whether it is the date is a holiday
Functioning day	Whether the rented bike sharing service is available. If no, the number of bikes shared is zero

Appendix B: Summary of Model 1 based on Environmental Factors

```
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)           596.169    234.162   2.546   0.0129 *  
temp                   10.619      5.295   2.006   0.0485 *  
humidity                2.755      3.198   0.861   0.3917    
log(solar_radiation)  295.444     49.120   6.015 5.88e-08 ***
log_rainfall          -45.720     20.435  -2.237   0.0282 *  
season_Summer         -25.285     80.530  -0.314   0.7544    
season_Spring        -118.895     73.629  -1.615   0.1105    
is_holiday           -299.939    143.689  -2.087   0.0402 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 235.2 on 76 degrees of freedom
Multiple R-squared:  0.6576,    Adjusted R-squared:  0.6261 
F-statistic: 20.85 on 7 and 76 DF,  p-value: 2.227e-15
```

Appendix C: Summary of Regression Model based on Time Hour Analysis

```
Call:
lm(formula = rentedbike ~ as.factor(hour) + as.factor(seasons) +
    temp + log_solar_radia + log_rainfall, data = bikeDB_Hour)

Residuals:
     Min       1Q   Median       3Q      Max
-2016.36  -198.05    35.42   222.41  1565.34

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               465.106     30.429  15.285  < 2e-16 ***
as.factor(hour)1         -157.055     37.311  -4.209 2.59e-05 ***
as.factor(hour)2         -294.585     37.313  -7.895 3.38e-15 ***
as.factor(hour)3         -393.812     37.323 -10.551  < 2e-16 ***
as.factor(hour)4         -481.528     37.331 -12.899  < 2e-16 ***
as.factor(hour)5         -476.353     37.342 -12.757  < 2e-16 ***
as.factor(hour)6         -277.437     37.359  -7.426 1.26e-13 ***
as.factor(hour)7           51.219     37.504   1.366 0.172083
as.factor(hour)8          405.184     38.590  10.500  < 2e-16 ***
as.factor(hour)9         -159.306     41.014  -3.884 0.000104 ***
as.factor(hour)10        -435.324     43.402 -10.030  < 2e-16 ***
as.factor(hour)11        -440.222     45.262  -9.726  < 2e-16 ***
as.factor(hour)12        -366.209     46.470  -7.881 3.79e-15 ***
as.factor(hour)13        -353.187     46.748  -7.555 4.75e-14 ***
as.factor(hour)14        -325.533     45.990  -7.078 1.61e-12 ***
as.factor(hour)15        -198.085     44.752  -4.426 9.74e-06 ***
as.factor(hour)16         -12.278     42.821  -0.287 0.774334
as.factor(hour)17         365.657     40.603   9.006  < 2e-16 ***
as.factor(hour)18         952.479     38.550  24.708  < 2e-16 ***
as.factor(hour)19         701.639     37.557  18.682  < 2e-16 ***
as.factor(hour)20         633.669     37.371  16.956  < 2e-16 ***
as.factor(hour)21         604.674     37.338  16.194  < 2e-16 ***
as.factor(hour)22         462.771     37.320  12.400  < 2e-16 ***
as.factor(hour)23         157.508     37.309   4.222 2.46e-05 ***
as.factor(seasons)Spring  -91.801     13.486  -6.807 1.08e-11 ***
as.factor(seasons)Summer  -28.603     17.921  -1.596 0.110521
temp                       16.699      1.014  16.465  < 2e-16 ***
log_solar_radia           448.040     27.015  16.585  < 2e-16 ***
log_rainfall             -434.247     18.228 -23.824  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 437.4 on 6571 degrees of freedom
Multiple R-squared:  0.5719,    Adjusted R-squared:  0.5701
F-statistic: 313.5 on 28 and 6571 DF,  p-value: < 2.2e-16
```

Appendix D: Summary of Weekday Regression Model based on Time Hour Analysis

```
Call:
lm(formula = rentedbike ~ as.factor(hour) + as.factor(seasons) +
    temp + solar_radia + rainfall, data = bikeDB_Hour_weekdays)

Residuals:
     Min       1Q   Median       3Q      Max
-1937.76  -223.61    47.01   239.96  2370.33

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               453.430     38.275  11.847  < 2e-16 ***
as.factor(hour)1         -162.479     46.895  -3.465 0.000536 ***
as.factor(hour)2         -302.538     46.900  -6.451 1.23e-10 ***
as.factor(hour)3         -407.491     46.906  -8.687  < 2e-16 ***
as.factor(hour)4         -481.590     46.918 -10.264  < 2e-16 ***
as.factor(hour)5         -483.341     46.934 -10.298  < 2e-16 ***
as.factor(hour)6         -294.147     46.946  -6.266 4.05e-10 ***
as.factor(hour)7           53.303     46.991   1.134 0.256718
as.factor(hour)8          444.874     47.410   9.384  < 2e-16 ***
as.factor(hour)9          -68.623     48.882  -1.404 0.160426
as.factor(hour)10        -314.264     51.015  -6.160 7.87e-10 ***
as.factor(hour)11        -314.861     53.289  -5.909 3.70e-09 ***
as.factor(hour)12        -255.366     54.799  -4.660 3.25e-06 ***
as.factor(hour)13        -227.439     55.519  -4.097 4.26e-05 ***
as.factor(hour)14        -208.341     54.530  -3.821 0.000135 ***
as.factor(hour)15         -78.090     52.980  -1.474 0.140561
as.factor(hour)16         115.156     50.722   2.270 0.023232 *
as.factor(hour)17         468.020     48.832   9.584  < 2e-16 ***
as.factor(hour)18        1015.399     47.543  21.357  < 2e-16 ***
as.factor(hour)19         717.898     47.060  15.255  < 2e-16 ***
as.factor(hour)20         627.781     46.964  13.367  < 2e-16 ***
as.factor(hour)21         594.658     46.925  12.673  < 2e-16 ***
as.factor(hour)22         449.150     46.901   9.577  < 2e-16 ***
as.factor(hour)23         147.833     46.891   3.153 0.001628 **
as.factor(seasons)Spring -103.986     16.868  -6.165 7.65e-10 ***
as.factor(seasons)Summer   33.293     22.376   1.488 0.136848
temp                       15.334      1.268  12.096  < 2e-16 ***
solar_radia               173.836     14.551  11.946  < 2e-16 ***
rainfall                  -76.744      4.971 -15.439  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 464.2 on 4675 degrees of freedom
Multiple R-squared:  0.5271,    Adjusted R-squared:  0.5243
F-statistic: 186.1 on 28 and 4675 DF,  p-value: < 2.2e-16
```

Appendix E: Summary of Weekend Regression Model based on Time Hour Analysis

```
Call:
lm(formula = rentedbike ~ as.factor(hour) + as.factor(seasons) +
    temp + solar_radia + rainfall, data = bikeDB_Hour_weekend)

Residuals:
    Min      1Q  Median      3Q     Max
-1874.0  -192.9    21.7   219.9  1268.9

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               424.814     53.063   8.006 2.06e-15 ***
as.factor(hour)1         -126.798     65.471  -1.937 0.052932 .
as.factor(hour)2         -272.600     65.488  -4.163 3.29e-05 ***
as.factor(hour)3         -389.829     65.497  -5.952 3.16e-09 ***
as.factor(hour)4         -476.673     65.515  -7.276 5.05e-13 ***
as.factor(hour)5         -452.488     65.535  -6.905 6.88e-12 ***
as.factor(hour)6         -254.295     65.555  -3.879 0.000108 ***
as.factor(hour)7          133.924     65.649   2.040 0.041489 *
as.factor(hour)8          539.585     66.417   8.124 8.10e-16 ***
as.factor(hour)9          -83.343     68.458  -1.217 0.223590
as.factor(hour)10        -399.729     71.889  -5.560 3.08e-08 ***
as.factor(hour)11        -436.545     75.424  -5.788 8.34e-09 ***
as.factor(hour)12        -404.459     78.297  -5.166 2.65e-07 ***
as.factor(hour)13        -396.398     78.459  -5.052 4.79e-07 ***
as.factor(hour)14        -337.662     76.652  -4.405 1.12e-05 ***
as.factor(hour)15        -191.571     74.142  -2.584 0.009847 **
as.factor(hour)16          13.975     71.364   0.196 0.844767
as.factor(hour)17         417.232     68.301   6.109 1.22e-09 ***
as.factor(hour)18        1035.552     66.517  15.568  < 2e-16 ***
as.factor(hour)19         750.386     65.788  11.406  < 2e-16 ***
as.factor(hour)20         682.212     65.670  10.388  < 2e-16 ***
as.factor(hour)21         630.163     65.556   9.613  < 2e-16 ***
as.factor(hour)22         495.953     65.509   7.571 5.80e-14 ***
as.factor(hour)23         173.241     65.483   2.646 0.008223 **
as.factor(seasons)Spring  -96.336     23.793  -4.049 5.36e-05 ***
as.factor(seasons)Summer -217.547     32.074  -6.783 1.58e-11 ***
temp                       22.698      1.778  12.767  < 2e-16 ***
solar_radia               236.837     20.848  11.360  < 2e-16 ***
rainfall                  -98.632     10.711  -9.209  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 411.5 on 1867 degrees of freedom
Multiple R-squared:  0.6064,    Adjusted R-squared:  0.6005
F-statistic: 102.7 on 28 and 1867 DF,  p-value: < 2.2e-16
```