Seoul Bike Sharing Demand Data

Christopher Vandenburgh, Yashas Guddada

Sreenivas, Samuel Jackson

OPIM 5604 Pankaj Prakash Final Project

December 4, 2024

Table of Contents

Executive Summary

This analysis considers how to accurately predict the demand for bike rentals, enabling the Seoul Public Bike system to optimize its operations, balance supply and demand effectively while maximizing both profitability and user satisfaction. With urbanization rapidly increasing, transportation systems must evolve to support sustainable growth, minimize reliance on non-renewable resources, and reduce the spatial burden of personal vehicles. Seoul's population density of approximately 17,000 people per square kilometer (2) highlights the importance of innovative transportation solutions like bike-sharing services, which cater to citizens lacking personal storage space for bikes and seeking affordable, eco-friendly alternatives to motorized travel.

The project utilized a comprehensive dataset from Seoul Bike, capturing bike rentals per hour using variables that might influence demand, such as weather conditions, time, season, and holiday schedules. Our first models were less powerful than we desired. After significant changes to preprocessing, variable creation, and modeling techniques we were able to make drastic improvements to our predictive power. By creating a manageable combination of hourly and seasonal variables, the model was able to predict demand more precisely without overwhelming computational resources. This innovation demonstrated that even in the absence of detailed weather forecasts, knowing the season and time of day could serve as strong proxies for estimating rental behavior. Our most powerful model that utilized temporal variables was able to explain around 90 percent of variability in bike rental demand. Without weather related variables we could still explain around 65 percent of variability in bike rental demand. Rainfall emerged as the single most impactful variable, highlighting the importance of weather conditions in shaping demand patterns.

Through this analysis, several strategic insights emerged. Seoul Bike should adopt a dynamic resource allocation strategy, adjusting the number of bikes and stations seasonally to align with demand fluctuations. For example, winter months consistently showed lower demand,

necessitating a reduction in available bikes, while summer experienced peak usage, warranting an expansion of resources. Additionally, the system must prioritize maintaining operational uptime, as outages have a profound impact on rental activity and user satisfaction. Incorporating real-time weather forecasting into the system's predictive framework could further enhance demand estimation, allowing for more responsive and adaptive operations. Incorporating a dynamic pricing option could enable Seoul Bike to maximize revenue while also managing supply during periods of extreme demand.

This analysis illustrates the potential of data-driven decision making to revolutionize urban transportation systems. By leveraging a robust analytical approach, this study provides actionable recommendations for optimizing resource allocation, improving user experience, and promoting sustainability in urban mobility. Future directions could include exploring advanced machine learning techniques to refine predictive accuracy further, as well as integrating additional external data sources, such as commuter patterns and traffic congestion levels. Such advancements would not only enhance the operational efficiency of Seoul Bike but also position it as a model for other cities seeking to implement similar systems in the quest for sustainable urban development.

Business Understanding

Currently, more than half of the global population lives in urban environments, and this is projected to reach around two-thirds by 2050 according to the United Nations (1). Transportation has been a growing concern for urban environments as countries try to limit the usage of non-renewable resources while moving toward cities that are not centered around cars. As environmental and spatial concerns become more prevalent alternative sources of transportation must be considered. In highly dense cities, space is seen as the most valuable and scarce resource. Seoul, South Korea has a population density of around 17,000 people per square kilometer, which is double the population density of New York City (2). With a population density that high many citizens do not even have the space required to store a bike that they own themselves.

Bike rental services are becoming more popular throughout cities as a method of transportation that does not occupy an individual's scarce living space. The global bike and scooter rental market is expected to experience a growth rate of 17.1% per year between 2024 and 2030 with bike rentals holding around 64% of that market share (3). Cities are also trying to reduce their reliance on oil by moving towards cleaner methods of transportation such as biking. South Korea, specifically, is trying to reduce their reliance on Russian oil due to the war in Ukraine as they lack the domestic natural resource (4). The reduction of oil usage also makes for cleaner cities with improved air quality that will benefit both the people and the planet.

With the numerous benefits of bike rentals in urban environments many cities are looking to implement bike sharing as a transportation staple. Seoul Public Bike is an un-manned bike rental system that can be used anywhere in Seoul, at any time, and by anyone (5). However, Seoul Bike needs to find the correct balance between the demand for bicycles and their supply if they wish to become profitable. If someone wants to rent a bike and there are none available due to Seoul Bike not having enough rental stations, they will look towards alternative transportation methods and consider bike sharing to be too unreliable in the future. If Seoul Bike has too many rental stations, they will spend too much on space and be unable to make a profit. Seoul Bike needs to be able to accurately predict the number of bikes rented to find the correct balance between supply and demand.

Seoul Bike needs to understand the impact that different variables will have on their bike rentals. The weather will play a significant role as the rider is exposed to the elements, and demand is expected to vary between seasons. The days of the week, holidays, and functional workdays will have an impact on bike demand. Seoul Bike needs to understand where their rentals come from whether it be tourists, people travelling for work or school, or people traveling to run errands.

Data Dictionary

Data source: https://archive.ics.uci.edu/dataset/560/seoul+bike+sharing+demand

The Seoul bike sharing demand data set being used for this report contains data collected from the company Seoul Bike between December 2017 and November 2018. Seoul Bike is a public bike rental system that operates in the capital of South Korea, Seoul.

Rented Bike count - Indicates the number of bikes rented during a specific hour for each day. This is the dependent variable that we are trying to predict.

Date – Represents the date of observation in day/month/year format.

Hour – Specifies the hour of the day when the data was recorded, using a 24-hour format from hour 0 starting at midnight to hour 23 starting at 11 P.M. Consult appendix 1 for hour visualization.

Temperature - Captures the temperature at the time of observation in the Celsius scale, we will be changing this to Fahrenheit. A temperature between 45 and 80 F would be considered good

Humidity - Measures the atmospheric humidity in percentage. A humidity between 30 and 60% is considered healthy.

Wind speed - Records the wind speed in meters per second (m/s). Any wind speed greater than 8m/s would be considered a high wind speed

Visibility – Represents the clarity of the atmosphere, measured in units of 10 meters(10m). A good visibility would be 10,000 meters or more (over 1000 for this data set)

Dew point temperature – Indicates the temperature the air needs to be cooled to (at constant pressure) in order to achieve a relative humidity (RH) of 100%.

Solar Radiation – measures the intensity of solar energy received at the surface, give in megajoules per square meter(MJ/m^2). This is essentially how much sunlight reaches the earth and is heavily correlated with cloud cover.

Rainfall – captures the amount of precipitation in millimeters (mm) during the recorded hour. A good rainfall would be 0mm.

Snowfall – Represents the amount of snowfall recorded during the observation measured in centimeters (cm). A good snowfall would be 0mm.

Seasons – Indicates the season during which the data was collected as a categorical variable (Winter, Spring, Summer, Autumn). Consult appendix 2 for season visualization.

Holiday – Categorizes whether the observation falls on Holiday or not.

Functional day – Describes whether Seoul Bike system was functional or down due to errors.

<u>Initial Variable Creation / Manipulation</u>

First, we will change the temperate from Celsius to Fahrenheit as we are presenting to an American audience. We do not believe that changing the meters units is necessary as these are more easily relatable. After that we will make dummy indicator columns for our categorical variables of season, holiday, and functioning day.

After that we want to make a new column called "Good hour" this column will be a categorical column that will tell us if the hour has a good temperature (40-85), humidity (30-60), wind speed (less than or equal to 8), visibility (greater than or equal to 1000), rainfall (0), and snowfall (0). Consult appendix 3 for good hour visualization.

We will also convert the date variable into a categorical to describe if the day is a weekend or not. If we converted this into each day of the week that would create too much noise for our model, and we believe that there would only be a significant difference between the week and weekend. Now the date variable is being accurately described by our weekend and season variables, and we will not include date it in the model.

Before we model, we do not expect there to be any target leakage as there are no variables that would be perfect predictors of the rented bike count. We also expect some multicollinearity between variables that are related to weather as those variables can be highly correlated.

<u>Modeling</u>

A validation column was created with half of the data in the training set and half in the validation.

Multiple linear regression model 1

Functioning Day_No	128.836		0.00000

Functioning Day_No	-929.4888	37.09645	-25.06	<.0001*

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.5738	421.62	4380
Validation Set	0.5580	428.21	4380

After running the model for the first time we noticed that the functioning day variable had the most significant impact on our predictions. When functioning day was "no" the entire system was down, and no bikes could be rented. Every nonfunctioning day record has a rented bike count of 0, and since this is a perfect predictor, we will remove the variable from the model. Consult appendix 4 for the full estimates.

After we removed the functioning day variable from the model our r-square significantly dropped and our error increased. We believe that nonfunctioning days should not have an impact on our model as it was not possible to rent bikes at all on those days. Records with nonfunctioning days should also be removed so that they do not have an impact on the other variables.

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.5125	450.94	4380
Validation Set	0.4921	459.03	4380

Multiple linear regression model 2: functioning day variable and nonfunctioning days removed

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.5677	422.90	4234
Validation Set	0.5537	428.51	4231

Our r-squared significantly increased by around 6% and our error decreased after removing nonfunctioning days from the model. Before we remove variables with insignificant p-values we wanted to see if we could manipulate them to be useful for our model. We saw that snowfall had an insignificant p-value but believed that snowfall would have an impact. We decided to convert snowfall from a numeric variable into a categorical variable where any snowfall that is not 0 would be categorized. Consult appendix 5 for parameter estimates.

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.5684	422.57	4234
Validation Set	0.5545	428.11	4231

Now snowfall is showing as significant, and the r-squared slightly increased. We decided to test the model to see if rainfall as a categorical variable would improve performance. When people decide to bike, they do not care about how much it is raining or snowing just whether it is raining or snowing at all.

Multiple linear regression model 3: rainfall and snowfall transformed from numeric to categorical

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.5888	412.45	4234
Validation Set	0.5729	419.21	4231

The r-squared value significantly improved by about 2% and the error decreased after changing rain to a categorical variable. Now we will remove variables with insignificant p-values to reduce the noise of our model. Consult appendix 6 for parameter estimates.

Multiple linear regression model 4: Insignificant p-values removed

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.5885	412.58	4234
Validation Set	0.5721	419.58	4231

First, we removed the dew point temperature and saw that it did not have a significant impact on the model performance. Then Visibility was still showing as insignificant, so we removed that variable, and it also did not have a significant effect on the model performance. We noticed that two variables had a variance inflation factor that was slightly over 4. Summer had a VIF of 6.7, but we will leave it in as Autumn and Spring do not have a high VIF. We also saw that Temperature had a high VIF of 5.3, but that was not too extreme and made sense as many of the variables are weather related. At this point we tried creative variable engineering to help improve our model. First, we created a time-of-day categorical variable for Afternoon (11-15), Evening (16-20), Night (21-5), and Morning (6-10). Consult appendix 7 for parameter estimates.

Multiple linear regression model 5: Time of day categorical variable created

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.6216	395.67	4234
Validation Set	0.6117	399.70	4231

Our model significantly improved in R-squared value (~4%), along with slightly reducing the error. Then we thought about combining Hour with Season to better capture seasonal differences, but this would create too many categorical variables. We then decided to combine our new Time of day variable with Seasons to create a manageable combination of 16 different categories. Consult appendix 8 for parameter estimates.

Multiple linear regression model 6: Time of day and Season combined

Our model significantly improved in terms of r-squared (~3%) and error. Consult appendix 9 for parameter estimates.

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.6555	377.55	4234
Validation Set	0.6447	382.35	4231

Multiple linear regression model 7: insignificant p-values removed.

First, we decided to remove wind speed as it has the least significant p-value. Removing wind speed had virtually no effect on our model performance. Then we saw that the snow variable was insignificant, and before removing snow we combined it with rain to make a categorical that was for any amount of either rain or snow. This reduced the r-squared by about 2%, so we put rain back in and just removed snow. Removing snow barely influenced the model's performance, but helped reduce noise. Consult appendix 10 for parameter estimates.

Crossvalidation			
Source	RSquare	RASE	Freq
Training Set	0.6552	377.71	4234
Validation Set	0.6443	382.53	4231

When discussing this model, we realized that weather variables would not be available to a high degree of accuracy within a reasonable time of modeling. We decided to build a model with all weather-related variables removed. To do this we decided to combine seasons and hours to create a categorical variable with 96 categories so that we could have competitive predictive power. Despite the high number of categories, we believed this to be the only way to achieve a strong model without any weather-related variables. We also included weekend, day of the week, and holiday variables as these would also be readily available years prior.

Multiple linear regression model 8: Weather variables removed

Crossvalidation			
Source	RSquare	RASE	Freq
Training Set	0.6343	388.95	4234
Validation Set	0.6262	392.15	4231

We removed the Holiday variable for being insignificant without any loss to model performance. We were surprised to see that the r-squared value only fell by around 2%, with a slight increase in error. The seasons by hour was able to capture the weather-related variables relatively well. Consult appendix 11 to see the most impactful variables in terms of p-value along with positive and negative coefficients.

Multiple linear regression model 9: Weather variables restored

We decided to put the weather variables back into the model to see the change in model performance and how much of rented bike count can be attributed to weather specific variables. We can account for around 63% of the variability in bike rental count without any weather variables. Insignificant p-values of visibility, dew point temperature, snow, and wind speed were removed without any loss of model performance. Temperature and solar radiation both had a VIF of around 5, but we decided to keep them in the model as this does not represent very high multicollinearity. Consult appendix 12 to see the parameter estimates.

Source	Logworth	PValue
Rain_0	113.801	0.00000
Temperature(F)	95.234	0.00000
SeasonHour_Autumn18	74.369	0.00000
SeasonHour_Summer18	59.858	0.00000

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.7667	310.66	4234
Validation Set	0.7624	312.67	4231

After adding the weather variables along with the season by hour category we were able to achieve a model that was significantly better than our first attempts. The r-squared of 76% is significantly higher than any of our models. We can also see that weather alone accounts for around 13% of variability in bike rental count, taking the difference from the 63% r-squared without weather variables. Weather is highly correlated with the season and time of day, so from just knowing the season and time we can estimate the weather conditions to a high degree of accuracy. We see that rain is the most impact variable overall, followed by temperature. Both variables have a greater impact than any time of the day per season. In our model 7 we saw that temperature (F) was the most impactful weather-related variable. This model contains 104 variables used to predict bike rent count, and this is drastically lower than our sample size. The large number of variables for the past two models are worth the extra complexity due to the significant predictive power that they provide.

We believe that our two final multiple linear regression models can be used in conjunction with each other. The model without weather variables can be used to predict broad season and hourly demand far into the future. This model can be used to determine the impact that seasons and hours will have on demand. While the model with weather variables can be used on a weekly basis to predict demand for the week based on current weather projections. This model will show the impact that weather variables will have on demand. Now we will try other modeling techniques to maximize model performance.

Lasso 1: No weather variables

Validation	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
Training	Rented Bike Count Prediction Formula	Fit Generalized Lasso		0.6297	391.40	259.76	4234
Validation	Rented Bike Count Prediction Formula	Fit Generalized Lasso		0.6356	387.20	257.26	4231

Lasso 2: Weather variables

Validation	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
Training	Rented Bike Count Prediction Formula	Fit Generalized Lasso		0.7633	312.92	218.88	4234
Validation	Rented Bike Count Prediction Formula	Fit Generalized Lasso		0.7678	309.11	218.55	4231

Ridge 1: No weather variables

Validation	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
Training	Rented Bike Count Prediction Formula 2	Fit Generalized Ridge		0.6271	392.76	260.02	4234
Validation	Rented Bike Count Prediction Formula 2	Fit Generalized Ridge		0.6379	385.98	255.85	4231

Ridge 2: Weather variables

Validation	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
Training	Rented Bike Count Prediction Formula 2	Fit Generalized Ridge		0.7631	313.06	219.49	4234
Validation	Rented Bike Count Prediction Formula 2	Fit Generalized Ridge		0.7681	308.85	218.67	4231

Elastic Net 1: No weather variables

Validation	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
Training	Rented Bike Count Prediction Formula 3	Fit Generalized Elastic Net		0.6279	392.34	260.66	4234
Validation	Rented Bike Count Prediction Formula 3	Fit Generalized Elastic Net		0.6374	386.24	256.72	4231

Elastic Net 2: Weather variables

Validation	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
Training	Rented Bike Count Prediction Formula 3	Fit Generalized Elastic Net		0.7641	312.40	218.44	4234
Validation	Rented Bike Count Prediction Formula 3	Fit Generalized Elastic Net		0.7670	309.60	218.68	4231

Ridge, lasso, and elastic net did not have a significant effect on model performance with r-squared values changing by less than 1%. Variable estimates and impacts remained relatively unchanged. There were no variables that had substantially high impact on the model compared with the other variables as both models had around 100 variables with most of the variables making contributions to the prediction.

Bagging 1: No weather variables

	RSquare	RASE	N	Number of Trees
Training	0.682	362.47866	4234	
Validation	0.650	379.53764	4231	24

Bagging 2: Weather Variables

	RSquare	RASE	N	Number of Trees
Training	0.877	225.971	4234	
Validation	0.809	280.44354	4231	13

Boosting 1: No weather variables

	RSquare	RASE	N
Training	0.687	359.7791	4234
Validation	0.644	382.69821	4231

Boosting 2: Weather variables

	RSquare	RASE	N
Training	0.925	176.52976	4234
Validation	0.880	222.52349	4231

The ensemble models for weather variables gained a significant amount of predictive power, especially the boosting model. However, in the process complexity was added and interpretability was lost. We also did not have problems with overfitting or underfitting that these ensembles could have mitigated.

Neural Net 1: No weather variables

We decided to run neural nets testing different combinations of layers and nodes on our final two models to see if they could improve the overall predictive power.

Model
NTanH(3)NLinear(3)NGaussian(3)NTanH2(3)NLinear2(3)NGaussian2(3)

Training — **Rented Bike Count**

Measures	Value
RSquare	0.6722799
RASE	368.21348
Mean Abs Dev	237.88584
-LogLikelihood	31025.064
SSE	574050661
Sum Freq	4234

Validation — **Rented Bike Count**

Measures	Value
RSquare	0.6575213
RASE	375.3698
Mean Abs Dev	245.34165
-LogLikelihood	31084.523
SSE	596158418
Sum Freq	4231

The validation r-squared improved by around 3% when predicting the rented bike count without weather variables, and we do not believe this is worth it for the loss of interpretability.

Neural Net 2: Weather variables

Model NTanH(3)NTanH2(3)NLinear2(3)NGaussian2(3)

Training		Validation	
Rented Bike Count		Rented Bike Count	
Measures	Value	Measures	Value
RSquare	0.9382077	RSquare	0.897012
RASE	159.88774	RASE	205.84291
Mean Abs Dev	103.72193	Mean Abs Dev	137.94533
-LogLikelihood	27493.1	-LogLikelihood	28542.545
SSE	108238349	SSE	179272984
Sum Freq	4234	Sum Freq	4231

After trying many different combinations of layers and nodes this model had the best validation results. This model has a very significant improvement over with a 13% increase in validation r-squared over the final multiple linear regression model. Both of our neural nets are experiencing slight overfitting. However, the error has not experienced significant improvement as the model is overpredicting by about 200 bikes.

<u>Final Model</u>

If we had to choose one model to represent our final model it would be the Neural Net 2 with weather variables model. We believe that the variables in our model can be explained by visual representations to accurately capture their impact more easily than our model estimates. We also believe that we can run the model with accurate weather predictions either at the beginning of every week or every day to create a dynamic pricing option for rented bikes. We believe this neural network to be worth the loss of interpretability due to the extremely high predictive power since it can account for around 90% of the variability in the rented bike count.

Evaluation

The model results demonstrate that around 90 percent of the variability in rented bike count can be explained with a neural network after a significant amount of variable creation. However, this prediction can only be applied to a short-term time frame of less than a week for accurate weather modeling. This can be useful if the business wants to deploy dynamic pricing. Dynamic pricing can help increase the ROI of Seoul Bike while also managing demand. If a day is projected to have an extremely high rented bike count the hourly rate could be increased to ensure there are enough available bikes. Without any weather variables the model can still predict demand to a valuable degree of accuracy of around 65 percent. This can be useful if Seoul Bike wants to gradually scale back their rental for the Winter while slowly reintroducing rental spaces during the Spring. This can help to drastically reduce operating costs for valuable rental space.

Deployment

The Seoul Bike Sharing Demand project demonstrates a robust approach to optimizing resource allocation by aligning operations with demand patterns influenced by seasonal and temporal factors. For deployment, scaling back operations during the winter months and gradually increasing availability through spring, peaking in summer, aligns well with observed trends. This ensures efficient resource utilization, minimizes overhead costs during low-demand periods, and enhances user satisfaction during peak seasons. Dynamic pricing will also be deployed to encourage rentals during times of low predicted bike rent counts and generate extra revenue during times of high predicted bike rent counts. However, there are ethical risks associated with this strategy. Some people could be living on a fixed income and would not have the ability to pay for rentals during peak hours. This can impact their job or school performance and cause user dissatisfaction and distrust. To mitigate this risk, we would propose a monthly pass option for frequent users who rely on our service. This monthly pass can then be further used to help predict demand and ensure locals who require our bikes for transporation have adequate access. However, this can increase the demand for bike rentals among tourists and infrequent renters.

References

Data: https://archive.ics.uci.edu/dataset/560/seoul+bike+sharing+demand

(1). https://www.un.org/development/desa/pd/content/urbanization-0#:~:text=The%20world%20is%20becoming%20increasingly,around%20two%2Dthirds%20in%202050.

(2). https://populationstat.com/south-korea/seoul

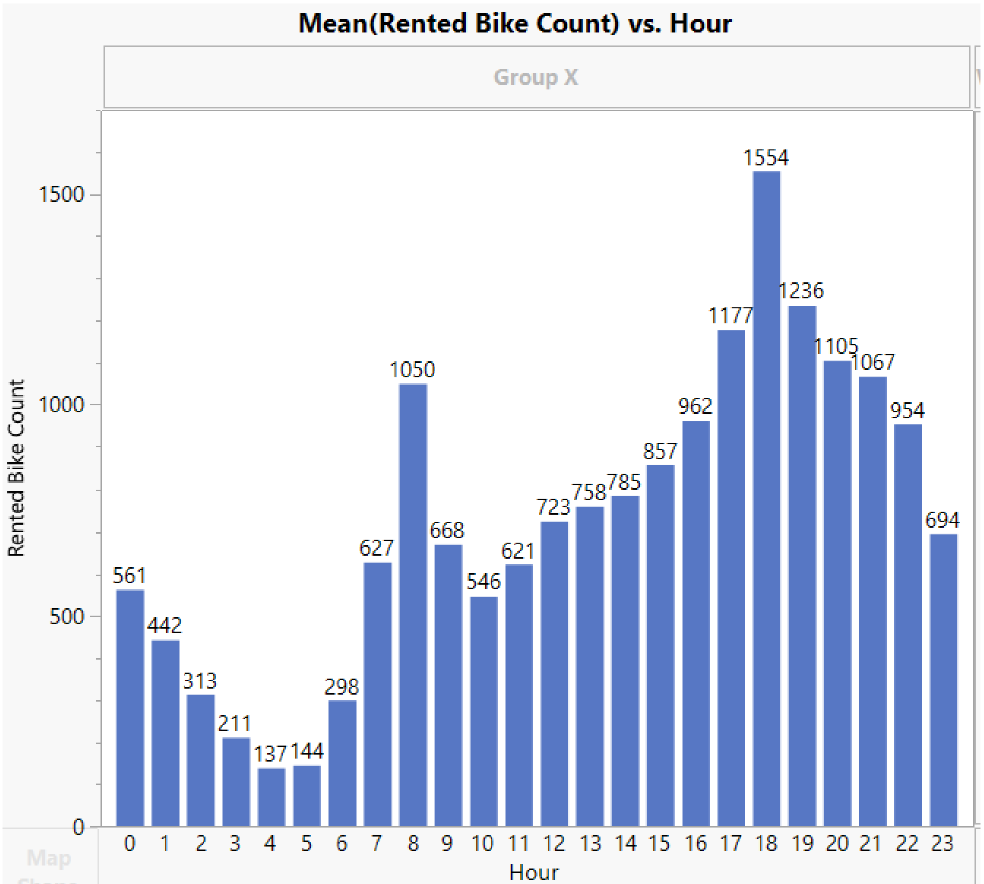
(3). https://www.grandviewresearch.com/industry-analysis/bike-scooter-rental-market-report#:~:text=The%20global%20bike%20%26%20scooter%20rental,of%20over%2063.6%25%20in%202023.

(4). https://www.bloomberg.com/news/articles/2024-05-07/russian-oil-product-cargoes-stranded-as-south-korea-cracks-down

(5). https://english.seoul.go.kr/service/movement/seoul-public-bike/1-seoul-public-bike/#:~:text=Seoul%20Public%20Bike%2C%20Ttareungyi%20is,of%20life%20for%20Seoul%20citizens.

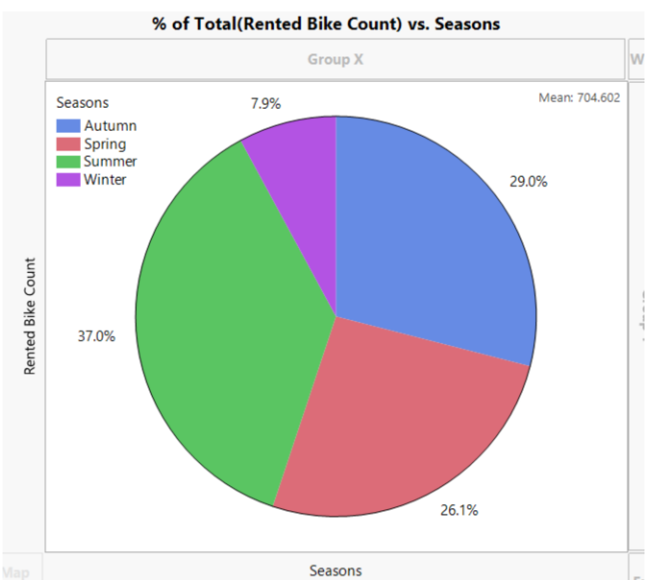
Appendix

Appendix 1: Hour visualization



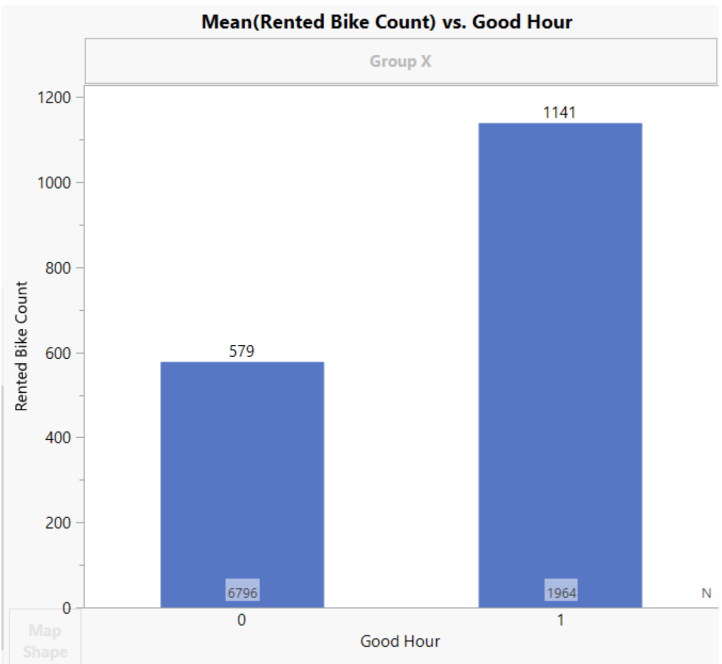
Rented bike count spikes before work/school at 8 A.M. and then peaks after work/school at 6 P.M.

Appendix 2: Season Visualization



This chart shows the percentage of bike rentals that occur during each Season. Winter clearly has significantly less rentals than the other seasons and rental stations should be drastically reduced for the Winter and slowly increased as the Seasons in change with a maximum in the Summer.

Appendix 3: Good hour visualization



This chart shows that good hours will on average have around twice the number of hourly bike rentals as hours that are not good. This also shows the number of good and not good hours on the bottom of the graph and around 22% of hours are considered good.

Appendix 4: multiple linear regression 1

Source	Logworth	PValue
Hour	135.677	0.00000
Functioning Day_No	128.836	0.00000
Good Hour_0	30.573	0.00000
Seasons_Autumn	28.313	0.00000
Rainfall(mm)	23.206	0.00000
Solar Radiation (MJ/m2)	14.810	0.00000
Seasons_Spring	11.307	0.00000
Weekend_0	10.310	0.00000
Humidity(%)	8.726	0.00000
Seasons_Summer	7.972	0.00000
Holiday_Holiday	4.623	0.00002
Temperature(F)	3.710	0.00019
Wind speed (m/s)	2.654	0.00222
Snowfall (cm)	0.722	0.18977
Dew point temperature (F)	0.540	0.28868
Visibility (10m)	0.484	0.32804

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	203.97961	123.4188	1.65	0.0985
Weekend_0	93.48414	14.18416	6.59	<.0001*
Hour	26.33637	1.022115	25.77	<.0001*
Good Hour_0	-217.8553	18.57478	-11.73	<.0001*
Humidity(%)	-8.765383	1.455859	-6.02	<.0001*
Seasons_Autumn	313.86876	27.86042	11.27	<.0001*
Seasons_Summer	222.42376	38.81178	5.73	<.0001*
Functioning Day_No	-929.4888	37.09645	-25.06	<.0001*
Holiday_Holiday	-122.4348	28.94237	-4.23	<.0001*
Wind speed (m/s)	21.737864	7.101059	3.06	0.0022*
Seasons_Spring	181.72373	26.23435	6.93	<.0001*
Snowfall (cm)	22.282608	16.99079	1.31	0.1898
Visibility (10m)	-0.013568	0.013871	-0.98	0.3280
Rainfall(mm)	-52.36407	5.159985	-10.15	<.0001*
Dew point temperature (F)	3.1568853	2.974972	1.06	0.2887
Solar Radiation (MJ/m2)	-84.95099	10.61562	-8.00	<.0001*
Temperature(F)	10.596006	2.841675	3.73	0.0002*

Appendix 5: Multiple linear regression 2

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	228.39675	124.6172	1.83	0.0669	.
Weekend_0	97.369683	14.36171	6.78	<.0001*	1.0172966
Hour	27.357179	1.041667	26.26	<.0001*	1.2213989
Temperature(F)	10.703516	2.866821	3.73	0.0002*	91.856142
Good Hour_0	-237.6168	19.0102	-12.50	<.0001*	1.4586438
Humidity(%)	-9.137684	1.468345	-6.22	<.0001*	21.861627
Seasons_Autumn	300.26828	28.19022	10.65	<.0001*	3.3505898
Seasons_Summer	205.03114	39.27179	5.22	<.0001*	6.9700109
Holiday_Holiday	-123.2495	29.9314	-4.12	<.0001*	1.0318802
Wind speed (m/s)	21.788399	7.252338	3.00	0.0027*	1.2931798
Seasons_Spring	166.15704	26.50408	6.27	<.0001*	3.1326165
Solar Radiation (MJ/m2)	-86.39389	10.78646	-8.01	<.0001*	2.0318959
Visibility (10m)	-0.018153	0.014221	-1.28	0.2018	1.7506327
Dew point temperature (F)	3.4416142	2.999861	1.15	0.2513	121.70028
Rainfall(mm)	-52.58671	5.202304	-10.11	<.0001*	1.0829261
Snowfall (cm)	25.076036	17.05093	1.47	0.1415	1.1290452

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	314.98195	128.5192	2.45	0.0143*	.
Weekend_0	96.333398	14.35602	6.71	<.0001*	1.0180816
Hour	27.368771	1.04003	26.32	<.0001*	1.2194703
Temperature(F)	10.950734	2.860956	3.83	0.0001*	91.623907
Good Hour_0	-236.4151	18.98694	-12.45	<.0001*	1.4573542
Humidity(%)	-9.254067	1.460901	-6.33	<.0001*	21.674432
Seasons_Autumn	303.96614	28.16627	10.79	<.0001*	3.3501364
Seasons_Summer	204.92599	39.18948	5.23	<.0001*	6.9516886
Holiday_Holiday	-122.9871	29.89966	-4.11	<.0001*	1.0313048
Wind speed (m/s)	20.601416	7.257797	2.84	0.0046*	1.2971551
Seasons_Spring	171.9789	26.49186	6.49	<.0001*	3.1346272
Solar Radiation (MJ/m2)	-88.15803	10.79227	-8.17	<.0001*	2.037271
Snow_0	-96.48708	32.58145	-2.96	0.0031*	1.2086529
Visibility (10m)	-0.01826	0.014206	-1.29	0.1987	1.7497756
Dew point temperature (F)	3.4544962	2.991081	1.15	0.2482	121.17835
Rainfall(mm)	-52.34512	5.195288	-10.08	<.0001*	1.0816981

Appendix 6: Multiple linear regression 3

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	-546.4179	137.4321	-3.98	<.0001*	.
Weekend_0	95.581896	14.00519	6.82	<.0001*	1.0170549
Hour	27.758111	1.015183	27.34	<.0001*	1.2196073
Temperature(F)	16.381735	2.817409	5.81	<.0001*	93.26928
Good Hour_0	-235.8266	18.52893	-12.73	<.0001*	1.4568277
Humidity(%)	-4.688706	1.460368	-3.21	0.0013*	22.734366
Seasons_Autumn	304.11328	27.49081	11.06	<.0001*	3.3498926
Seasons_Summer	212.27197	38.25442	5.55	<.0001*	6.9529114
Holiday_Holiday	-115.5046	29.18652	-3.96	<.0001*	1.0315052
Wind speed (m/s)	27.036245	7.097229	3.81	0.0001*	1.3020031
Seasons_Spring	185.50672	25.86824	7.17	<.0001*	3.1372351
Solar Radiation (MJ/m2)	-87.83125	10.53386	-8.34	<.0001*	2.0372776
Snow_0	-91.39754	31.79038	-2.88	0.0041*	1.2078264
Rain_0	528.50082	29.72772	17.78	<.0001*	1.2538506
Visibility (10m)	-0.018389	0.013864	-1.33	0.1848	1.7493622
Dew point temperature (F)	-2.543022	2.948555	-0.86	0.3885	123.60591

Appendix 7: Multiple linear regression 4:

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	-492.2716	60.28293	-8.17	<.0001*	.
Weekend_0	93.660776	13.94206	6.72	<.0001*	1.0077554
Hour	27.883609	1.012295	27.54	<.0001*	1.212495
Temperature(F)	14.043943	0.670504	20.95	<.0001*	5.2817321
Good Hour_0	-229.6975	18.1286	-12.67	<.0001*	1.3943465
Humidity(%)	-5.505793	0.453661	-12.14	<.0001*	2.1935878
Seasons_Autumn	299.10621	27.14861	11.02	<.0001*	3.2665244
Seasons_Summer	201.96653	37.66756	5.36	<.0001*	6.7402089
Holiday_Holiday	-116.4692	29.18055	-3.99	<.0001*	1.0309281
Wind speed (m/s)	26.32993	7.065403	3.73	0.0002*	1.2901585
Seasons_Spring	186.27186	25.85082	7.21	<.0001*	3.1325395
Solar Radiation (MJ/m2)	-82.91268	10.04563	-8.25	<.0001*	1.8525247
Snow_0	-92.99428	31.76481	-2.93	0.0034*	1.2057027
Rain_0	523.00691	29.23707	17.89	<.0001*	1.2126212

Appendix 8: Multiple linear regression 5:

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-550.6733	58.55618	-9.40	<.0001*
Temperature(F)	13.169657	0.663455	19.85	<.0001*
Good Hour_0	-225.7995	17.40824	-12.97	<.0001*
Humidity(%)	-4.621591	0.443644	-10.42	<.0001*
Wind speed (m/s)	13.166082	6.949778	1.89	0.0582
Solar Radiation (MJ/m2)	-30.75447	13.20876	-2.33	0.0199*
Rain_0	529.24442	28.37225	18.65	<.0001*
Snow_0	-85.77114	30.48284	-2.81	0.0049*
Hour	22.748284	1.083048	21.00	<.0001*
Weekend_0	98.746486	13.38095	7.38	<.0001*
Holiday_Holiday	-117.5078	28.02319	-4.19	<.0001*
Seasons_Autumn	299.79048	26.24879	11.42	<.0001*
Seasons_Spring	184.53894	25.15348	7.34	<.0001*
Seasons_Summer	204.69348	36.77984	5.57	<.0001*
Time_Afternoon	-43.99425	25.77587	-1.71	0.0879
Time_Evening	289.74806	20.54954	14.10	<.0001*
Time_Morning	161.99535	17.85299	9.07	<.0001*

Appendix 9: Multiple linear regression 6:

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-472.1267	57.31937	-8.24	<.0001*
Temperature(F)	12.902548	0.636163	20.28	<.0001*
Good Hour_0	-209.6397	16.79317	-12.48	<.0001*
Humidity(%)	-4.563128	0.426493	-10.70	<.0001*
Wind speed (m/s)	7.5699226	6.673895	1.13	0.2568
Solar Radiation (MJ/m2)	-31.92584	13.24934	-2.41	0.0160*
Rain_0	532.88595	27.1489	19.63	<.0001*
Snow_0	-43.85414	29.23652	-1.50	0.1337
Hour	23.018546	1.035173	22.24	<.0001*
Weekend_0	96.159679	12.78376	7.52	<.0001*
Holiday_Holiday	-112.6948	26.7815	-4.21	<.0001*
SeasonAndTime_AutumnAfternoon	173.84041	42.17496	4.12	<.0001*
SeasonAndTime_AutumnEvening	583.5858	39.47966	14.78	<.0001*
SeasonAndTime_AutumnMorning	368.51826	36.74252	10.03	<.0001*
SeasonAndTime_AutumnNight	86.931109	33.0313	2.63	0.0085*
SeasonAndTime_SpringAfternoon	106.39286	42.06112	2.53	0.0115*
SeasonAndTime_SpringEvening	393.14423	39.01243	10.08	<.0001*
SeasonAndTime_SpringMorning	220.73397	35.71108	6.18	<.0001*
SeasonAndTime_SpringNight	10.771667	31.93585	0.34	0.7359
SeasonAndTime_SummerAfternoon	-87.75833	50.05478	-1.75	0.0796
SeasonAndTime_SummerEvening	590.45213	47.11481	12.53	<.0001*
SeasonAndTime_SummerMorning	152.98216	43.59657	3.51	0.0005*
SeasonAndTime_SummerNight	101.89413	40.81234	2.50	0.0126*
SeasonAndTime_WinterAfternoon	-119.6559	34.18643	-3.50	0.0005*
SeasonAndTime_WinterEvening	-164.0937	33.0585	-4.96	<.0001*
SeasonAndTime_WinterMorning	131.98159	31.63065	4.17	<.0001*

Appendix 10: Multiple linear regression 7:

Source	Logworth	PValue
Hour	105.365	0.00000
Temperature(F)	86.933	0.00000
Rain_0	81.463	0.00000
SeasonAndTime_AutumnEvening	47.652	0.00000
SeasonAndTime_SummerEvening	35.751	0.00000
Good Hour_0	34.722	0.00000
Humidity(%)	25.828	0.00000
SeasonAndTime_SpringEvening	23.781	0.00000
SeasonAndTime_AutumnMorning	22.076	0.00000
Weekend_0	13.512	0.00000
SeasonAndTime_SpringMorning	8.703	0.00000
SeasonAndTime_WinterEvening	6.110	0.00000
SeasonAndTime_WinterMorning	4.553	0.00003
Holiday_Holiday	4.547	0.00003
SeasonAndTime_AutumnAfternoon	4.234	0.00006
SeasonAndTime_WinterAfternoon	3.244	0.00057
SeasonAndTime_SummerMorning	3.235	0.00058
SeasonAndTime_AutumnNight	1.919	0.01204
SeasonAndTime_SummerNight	1.891	0.01286
SeasonAndTime_SpringAfternoon	1.851	0.01410
Solar Radiation (MJ/m2)	1.439	0.03639
SeasonAndTime_SummerAfternoon	1.123	0.07538
SeasonAndTime_SpringNight	0.080	0.83121

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-493.8718	49.00091	-10.08	<.0001*
Weekend_0	97.334813	12.77082	7.62	<.0001*
Hour	23.171067	1.029138	22.52	<.0001*
Temperature(F)	12.653635	0.622661	20.32	<.0001*
Good Hour_0	-210.5116	16.78565	-12.54	<.0001*
Humidity(%)	-4.488277	0.417983	-10.74	<.0001*
Holiday_Holiday	-112.1721	26.76589	-4.19	<.0001*
SeasonAndTime_AutumnAfternoon	169.4097	42.10729	4.02	<.0001*
SeasonAndTime_AutumnEvening	583.70888	39.42919	14.80	<.0001*
SeasonAndTime_AutumnMorning	361.53438	36.56704	9.89	<.0001*
SeasonAndTime_AutumnNight	82.755136	32.94506	2.51	0.0120*
SeasonAndTime_SpringAfternoon	102.83258	41.87476	2.46	0.0141*
SeasonAndTime_SpringEvening	395.89693	38.50478	10.28	<.0001*
SeasonAndTime_SpringMorning	213.25928	35.46991	6.01	<.0001*
SeasonAndTime_SpringNight	6.7641297	31.7323	0.21	0.8312
SeasonAndTime_SummerAfternoon	-88.99343	50.03527	-1.78	0.0754
SeasonAndTime_SummerEvening	596.28219	46.82396	12.73	<.0001*
SeasonAndTime_SummerMorning	150.00219	43.57273	3.44	0.0006*
SeasonAndTime_SummerNight	101.5078	40.78827	2.49	0.0129*
SeasonAndTime_WinterAfternoon	-117.7502	34.14733	-3.45	0.0006*
SeasonAndTime_WinterEvening	-162.9443	32.92784	-4.95	<.0001*
SeasonAndTime_WinterMorning	132.62801	31.62503	4.19	<.0001*
Rain_0	532.09456	27.10173	19.63	<.0001*
Solar Radiation (MJ/m2)	-27.13007	12.96068	-2.09	0.0364*

Appendix 11: multiple linear regression 8

Source	Logworth	PValue
SeasonHour_Summer18	109.244	0.00000
SeasonHour_Autumn18	99.579	0.00000
SeasonHour_Summer19	81.764	0.00000
SeasonHour_Summer20	81.604	0.00000
SeasonHour_Spring18	71.762	0.00000
SeasonHour_Summer21	69.622	0.00000
SeasonHour_Autumn17	60.823	0.00000
SeasonHour_Summer22	57.594	0.00000
SeasonHour_Summer17	51.468	0.00000
SeasonHour_Autumn19	45.376	0.00000
SeasonHour_Autumn8	43.625	0.00000
SeasonHour_Autumn20	43.436	0.00000
SeasonHour_Autumn16	42.247	0.00000
SeasonHour_Autumn21	39.616	0.00000
SeasonHour_Summer8	35.464	0.00000
SeasonHour_Summer16	34.743	0.00000
SeasonHour_Summer23	30.499	0.00000
SeasonHour_Spring16	30.023	0.00000
SeasonHour_Spring19	30.020	0.00000
SeasonHour_Autumn15	29.697	0.00000
SeasonHour_Spring17	28.038	0.00000

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
SeasonHour_Autumn18	1923.9564	88.02814	21.86	<.0001*	1.9334575
SeasonHour_Summer18	1922.8205	83.71865	22.97	<.0001*	2.1477306
SeasonHour_Summer19	1663.3912	84.52384	19.68	<.0001*	2.0990262
SeasonHour_Summer20	1602.6324	81.52118	19.66	<.0001*	2.2877362
SeasonHour_Summer21	1552.8683	85.91166	18.08	<.0001*	2.028548
SeasonHour_Spring18	1529.8016	83.29318	18.37	<.0001*	2.1697287
SeasonHour_Summer22	1422.037	86.93886	16.36	<.0001*	1.9816711
SeasonHour_Autumn17	1408.79	83.69285	16.83	<.0001*	2.1464072
SeasonHour_Summer17	1310.1184	84.97186	15.42	<.0001*	2.0757167
SeasonHour_Autumn19	1270.5011	88.03234	14.43	<.0001*	1.9336419
SeasonHour_Autumn8	1236.6749	87.47466	14.14	<.0001*	1.9577082
SeasonHour_Autumn20	1175.3705	83.32833	14.11	<.0001*	2.17156
SeasonHour_Autumn21	1161.3675	86.40604	13.44	<.0001*	2.0047198

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
SeasonHour_Winter5	-120.5806	84.53159	-1.43	0.1538	2.0994114
SeasonHour_Winter4	-120.1899	84.97089	-1.41	0.1573	2.0756694
SeasonHour_Winter3	-86.42024	85.94117	-1.01	0.3147	2.0299418
SeasonHour_Winter6	-80.98233	83.30606	-0.97	0.3311	2.1703998
SeasonHour_Spring4	-58.77088	80.31574	-0.73	0.4644	2.3827845
SeasonHour_Spring5	-51.63241	90.57864	-0.57	0.5687	1.8389101
SeasonHour_Winter2	-41.90447	84.11887	-0.50	0.6184	2.1236483
SeasonHour_Autumn5	-12.36557	89.25709	-0.14	0.8898	1.8867795
SeasonHour_Autumn4	-12.22792	88.02957	-0.14	0.8895	1.9335201
SeasonHour_Winter1	-11.29839	82.92498	-0.14	0.8916	2.1939534
SeasonHour_Winter0	-2.408427	84.54667	-0.03	0.9773	2.1001604

Appendix 12: Multiple linear regression 9

Source	Logworth		PValue
Rain_0	113.801		0.00000
Temperature(F)	95.234		0.00000
SeasonHour_Autumn18	74.369		0.00000
SeasonHour_Summer18	59.858		0.00000
SeasonHour_Spring18	48.178		0.00000
SeasonHour_Summer20	45.582		0.00000
SeasonHour_Summer19	44.278		0.00000
SeasonHour_Autumn8	40.948		0.00000
SeasonHour_Summer21	40.392		0.00000
Good Hour_0	33.631		0.00000
SeasonHour_Autumn17	33.077		0.00000
SeasonHour_Summer22	31.688		0.00000
SeasonHour_Autumn19	30.729		0.00000
SeasonHour_Autumn20	26.566		0.00000
SeasonHour_Autumn21	21.739		0.00000
Weekend_0	19.530		0.00000
SeasonHour_Spring19	17.132		0.00000
Humidity(%)	16.434		0.00000
SeasonHour_Spring8	15.776		0.00000

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
SeasonHour_Autumn18	1373.0976	73.36222	18.72	<.0001*	2.1019502
SeasonHour_Summer18	1218.1891	72.97932	16.69	<.0001*	2.5545888
SeasonHour_Summer19	1045.2233	73.35765	14.25	<.0001*	2.4747783
SeasonHour_Spring18	1027.206	68.97105	14.89	<.0001*	2.3286604
SeasonHour_Summer20	1025.6836	70.89951	14.47	<.0001*	2.7085535
SeasonHour_Summer21	1003.8831	73.93208	13.58	<.0001*	2.3514347
SeasonHour_Autumn8	972.47288	71.10803	13.68	<.0001*	2.0249143
SeasonHour_Summer22	887.30278	74.21224	11.96	<.0001*	2.2601684
SeasonHour_Autumn19	853.91441	72.58055	11.77	<.0001*	2.0573967
SeasonHour_Autumn17	849.96653	69.50815	12.23	<.0001*	2.3173567
SeasonHour_Autumn20	750.83041	68.88581	10.90	<.0001*	2.3229081
SeasonHour_Autumn21	699.90062	71.35997	9.81	<.0001*	2.1402322
SeasonHour_Spring19	610.87985	70.63713	8.65	<.0001*	2.146511
SeasonHour_Spring21	578.90911	73.09475	7.92	<.0001*	1.9805873
SeasonHour_Summer17	560.84105	74.39827	7.54	<.0001*	2.4907465
SeasonHour_Autumn22	559.54869	72.31712	7.74	<.0001*	2.0424893
SeasonHour_Spring8	550.26586	66.47533	8.28	<.0001*	2.381069
SeasonHour_Summer8	545.41247	72.64558	7.51	<.0001*	2.3225555
Rain_0	539.05521	22.95954	23.48	<.0001*	1.2908427

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	-411.1424	62.83027	-6.54	<.0001*	.
SeasonHour_Summer4	-404.4934	70.27805	-5.76	<.0001*	2.5152338
SeasonHour_Summer5	-324.6297	72.54716	-4.47	<.0001*	2.316266
SeasonHour_Spring4	-268.4297	65.41133	-4.10	<.0001*	2.4738619
SeasonHour_Summer3	-239.4507	72.08155	-3.32	0.0009*	2.3380408
SeasonHour_Spring5	-236.8572	73.4581	-3.22	0.0013*	1.8931029
SeasonHour_Summer11	-233.924	75.80469	-3.09	0.0020*	2.585806
SeasonHour_Autumn5	-214.4002	72.44592	-2.96	0.0031*	1.9455817
SeasonHour_Autumn4	-211.7949	71.363	-2.97	0.0030*	1.9889493
Good Hour_0	-175.7675	14.24918	-12.34	<.0001*	1.4870027
SeasonHour_Summer10	-167.8993	75.20057	-2.23	0.0256*	2.3768059
SeasonHour_Spring3	-160.8984	68.74766	-2.34	0.0193*	2.173508
SeasonHour_Autumn3	-158.5919	72.62895	-2.18	0.0290*	1.9554248
SeasonHour_Summer2	-146.7664	73.0146	-2.01	0.0445*	2.2934356
SeasonHour_Summer13	-142.01	79.50796	-1.79	0.0742	2.4061418
SeasonHour_Summer12	-140.8637	75.33433	-1.87	0.0616	2.7221198
SeasonHour_Winter13	-127.8015	69.87159	-1.83	0.0675	2.1002369
SeasonHour_Summer6	-120.6924	71.38135	-1.69	0.0909	2.3935985
Holiday_Holiday	-112.224	22.38845	-5.01	<.0001*	1.0475608