FINAL PROJECT

BICYCLE RENTAL BUSINESS ANALYSIS

SUBHASH SHANMUKA PRAKTHESWARA

GEN BUS 704 – GORDON ENDERLE

Wisconsin School of Business

Contents

Introduction

we are given a dataset BIKE_SHARING that tells us about the bicycle rentals in London for the past 2 years.

The primary goal of this project is to come up with analysis and suggestions as to help the business.

This project will provide some interesting observations, statistics & graphs to help us predict future bicycle rentals.

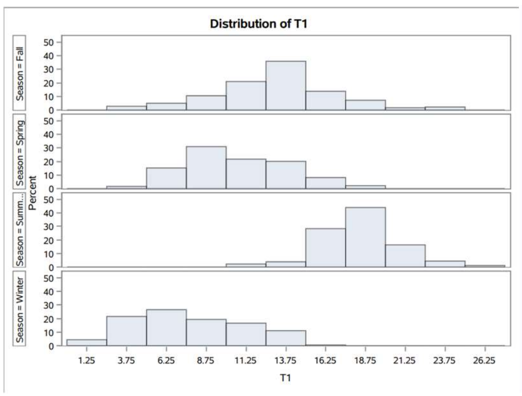
Objective

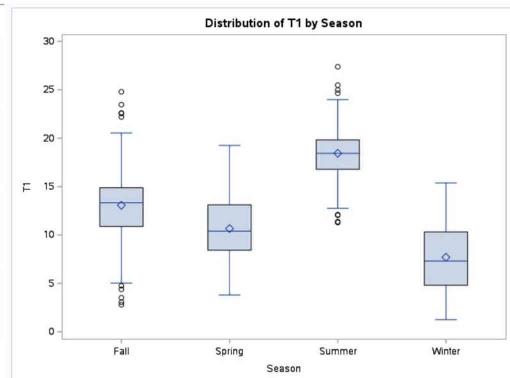
The Objective of this project is to understand and analyze given dataset and come up with practical solutions to tackle problems in the Bike rental business and business improvement strategies. Additionally a Regression Model is also built to help predict future rentals.

Data Exploration

Let us explore the data first, to understand relationships among variables and identify any trends (or) patterns.

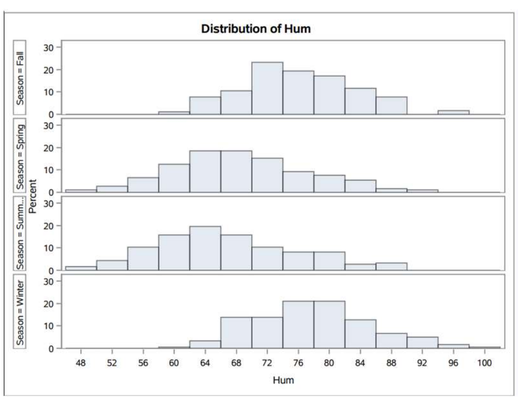
Distribution of T1 across seasons

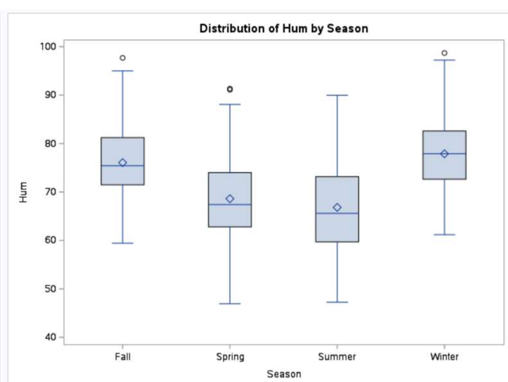




From these graphs we see that Summer is hottest followed by Fall, spring and winter.

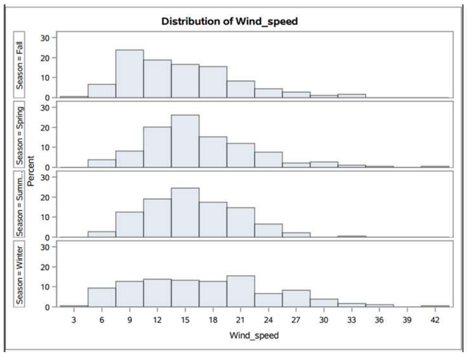
Distribution of humidity across seasons

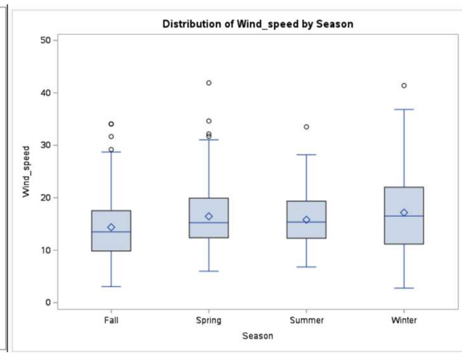




We see that Humidity is high during Fall and winter comparatively.

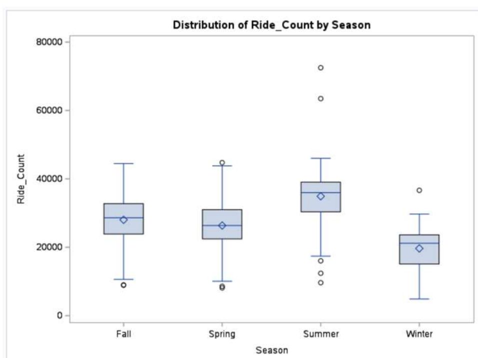
Distribution of Wind speed across seasons





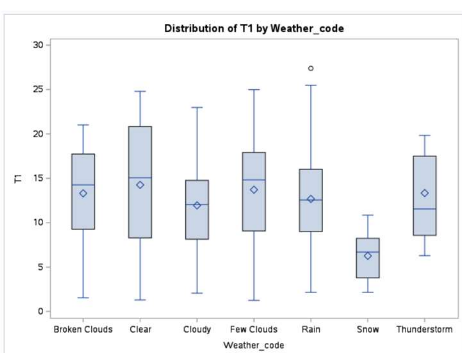
wind speed is higher in winter than in other seasons.

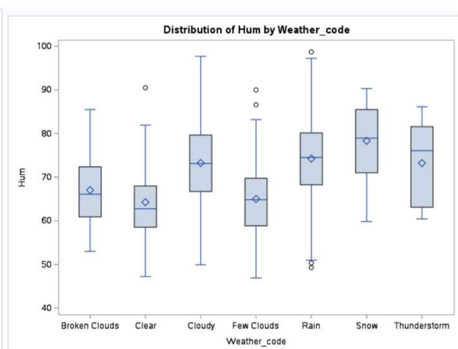
Ride distribution by season



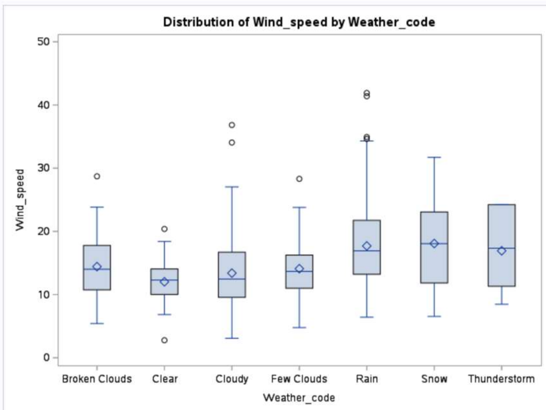
we see that Rides are highest during summer and least during winter.

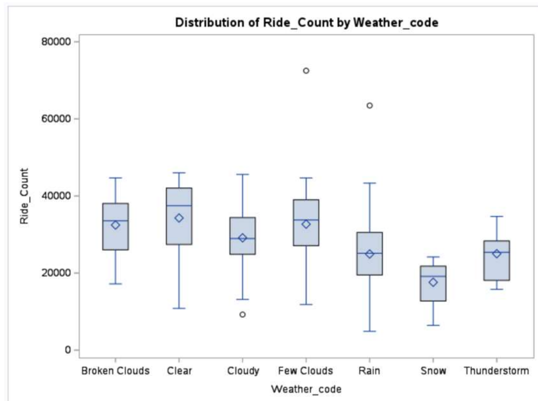
Distribution of T1 and humidity over weather





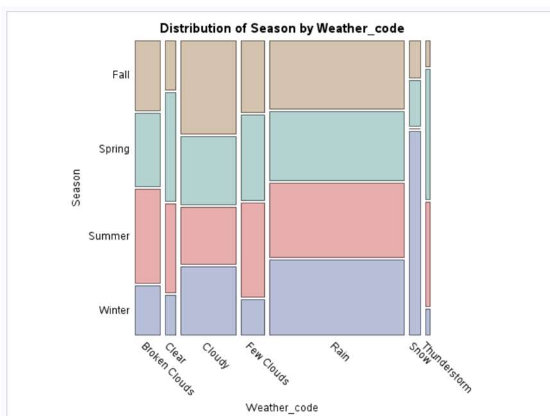
The graphs discuss the distribution of TI over weather and Humidity over weather.





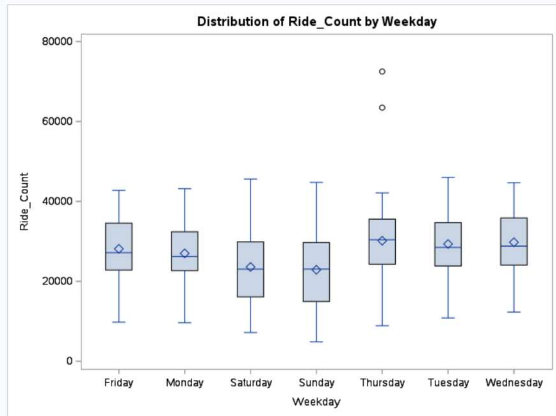
The graphs discuss wind and ride distributions over weather.

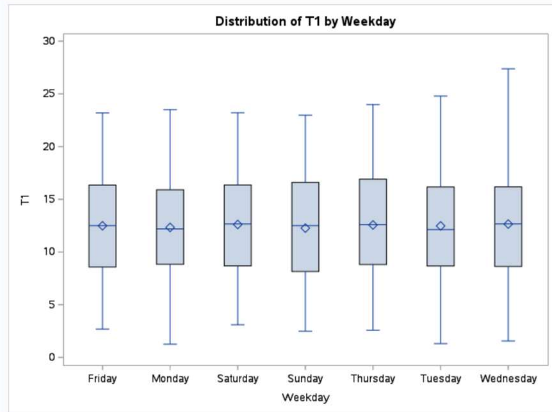
<u>Season vs weather</u>

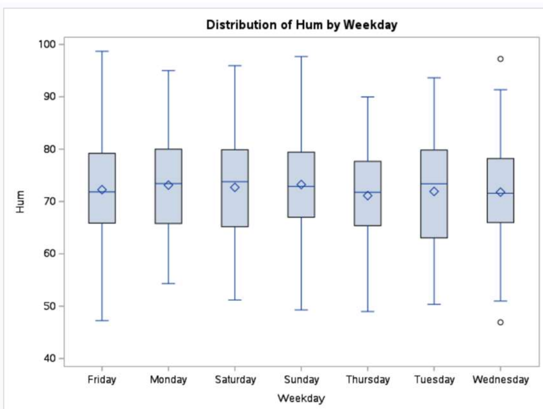


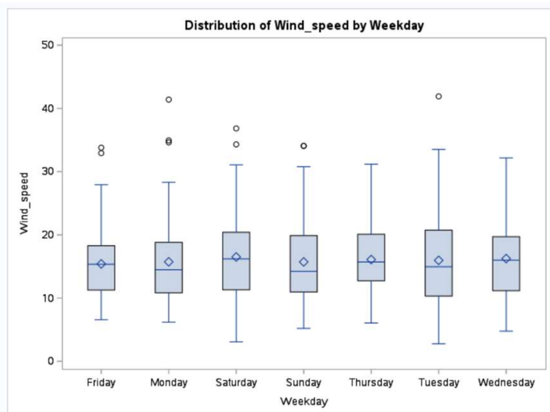
The graph displays distribution of weather over each season.

Comparison by Weekday:



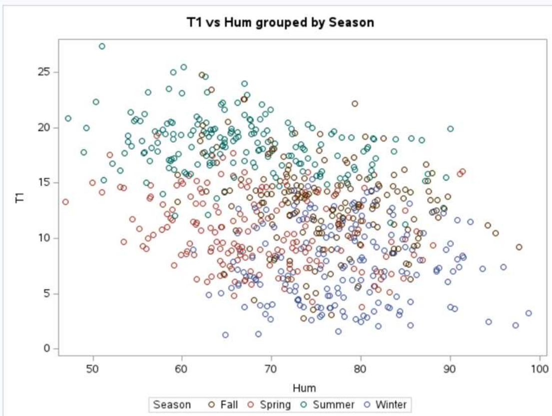


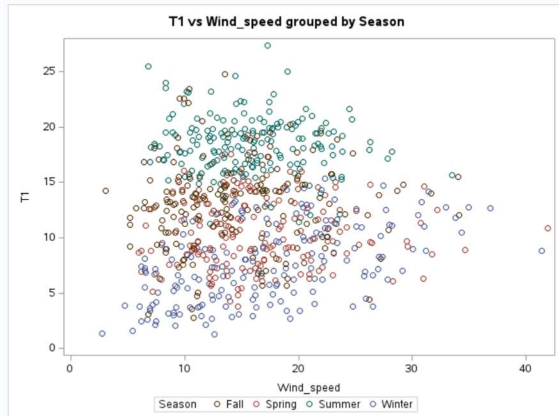


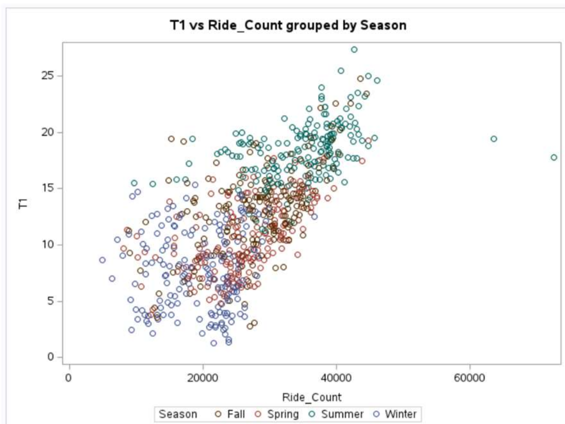


The graphs display distribution of Rides, T1, Humidity and wind speed against weekday.

Additional Plots:







The graphs capture the distribution of T1 and season against humidity, windspeed and ride count.

Statistical Analysis

Let us perform some primary statistical analysis on our data.

<u>Summary Statistics:</u>

Variable	Mean	Std Dev	Minimum	Maximum	Median	N
Date	20456.67	211.1201615	20092.00	20822.00	20456.50	730
Ride_Count	27268.45	8607.70	4869.00	72504.00	27011.50	730
T1	12.4818219	5.1373876	1.2500000	27.3800000	12.5000000	730
T2	11.5353836	6.1533934	-2.4000000	27.4400000	12.3800000	730
Hum	72.3095616	9.5010850	46.9200000	98.6900000	72.4800000	730
Wind_speed	15.9548767	6.2217214	2.7700000	41.9000000	15.1700000	730
Holiday	0.0219178	0.1465156	0	1.0000000	0	730
Weekend	0.2863014	0.4523419	0	1.0000000	0	730

<u>Correlation analysis:</u>

8 Variables:	Date T1 T2 Hum Wind_speed Holiday Weekend Ride_Count

Pearson Correlation Coefficients, N = 730

	Date	T1	T2	Hum	Wind_speed	Holiday	Weekend	Ride_Count
Date	1.00000	0.14398	0.15500	0.17956	-0.15934	0.02613	0.00173	0.11600
T1	0.14398	1.00000	0.99271	-0.37035	0.03631	-0.04607	-0.00563	0.67739
T2	0.15500	0.99271	1.00000	-0.33583	-0.00850	-0.04331	-0.00901	0.67613
Hum	0.17956	-0.37035	-0.33583	1.00000	-0.18634	0.04846	0.04403	-0.56400
Wind_speed	-0.15934	0.03631	-0.00850	-0.18634	1.00000	-0.00430	0.01584	-0.19749
Holiday	0.02613	-0.04607	-0.04331	0.04846	-0.00430	1.00000	-0.09481	-0.15314
Weekend	0.00173	-0.00563	-0.00901	0.04403	0.01584	-0.09481	1.00000	-0.29641
Ride_Count	0.11600	0.67739	0.67613	-0.56400	-0.19749	-0.15314	-0.29641	1.00000

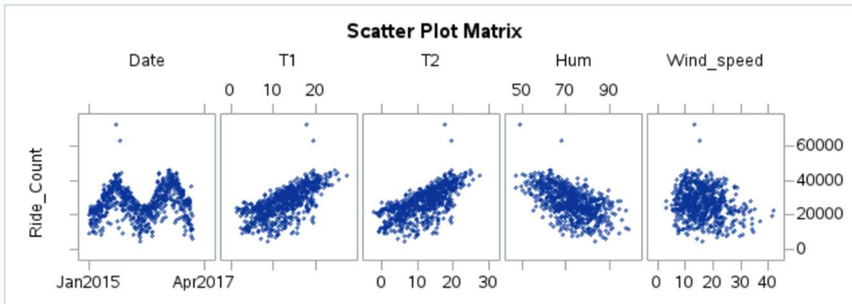
The correlation analysis reveals good relationships among T1, T2, ride count, Humidity. Another In-depth analysis of Ride count is presented below.

Ride Count vs other variables – Correlation analysis:

1 With Variables:	Ride_Count
7 Variables:	Date T1 T2 Hum Wind_speed Holiday Weekend

Pearson Correlation Coefficients, N = 730

	Date	T1	T2	Hum	Wind_speed	Holiday	Weekend
Ride_Count	0.11600	0.67739	0.67613	-0.56400	-0.19749	-0.15314	-0.29641



Correlation analysis reveals T1, Humidity have strong (medium) relation with Ride_count. Scatter plots show seasonality trend for rentals, strong relation with T1, T2 and Humidity against rentals. Our Regression model will include these variables.

Confidence Interval analysis for Ride distribution:

Using the SAS data analysis functions, I was able to obtaine the Confidence Intervals for Ride_Count by season and weekday. The below images show the same.

Weekday=Friday

Analysis Variable : Ride_Count					
Season	N Obs	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
Fall	25	29713.56	4605.77	27812.39	31614.73
Spring	26	27336.77	6312.00	24787.30	29886.24
Summer	26	33882.12	6546.84	31237.79	36526.44
Winter	26	21663.96	4462.58	19861.49	23466.43

Weekday=Sunday

Analysis Variable : Ride_Count					
Season	N Obs	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
Fall	26	23323.50	7060.77	20471.59	26175.41
Spring	27	22736.93	8664.55	19309.34	26164.51
Summer	26	31655.54	8036.98	28409.33	34901.74
Winter	26	13860.15	5599.45	11598.49	16121.82

Weekday=Monday

Analysis Variable : Ride_Count					
Season	N Obs	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
Fall	26	28419.96	5416.05	26232.37	30607.55
Spring	26	25983.62	5494.49	23764.34	28202.89
Summer	27	32387.44	7447.03	29441.50	35333.39
Winter	26	20990.96	4067.20	19348.19	22633.74

Weekday=Thursday

Analysis Variable : Ride_Count					
Season	N Obs	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
Fall	26	31814.62	4825.15	29865.69	33763.54
Spring	26	28554.46	6230.47	26037.92	31071.00
Summer	26	38459.27	9914.32	34454.79	42463.75
Winter	26	21823.81	4729.53	19913.51	23734.10

Weekday=Saturday

Analysis Variable : Ride_Count					
Season	N Obs	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
Fall	26	22186.54	7260.27	19254.05	25119.02
Spring	26	23456.58	7279.32	20516.40	26396.76
Summer	26	34796.50	6657.82	32107.35	37485.65
Winter	26	13946.54	3180.30	12661.99	15231.09

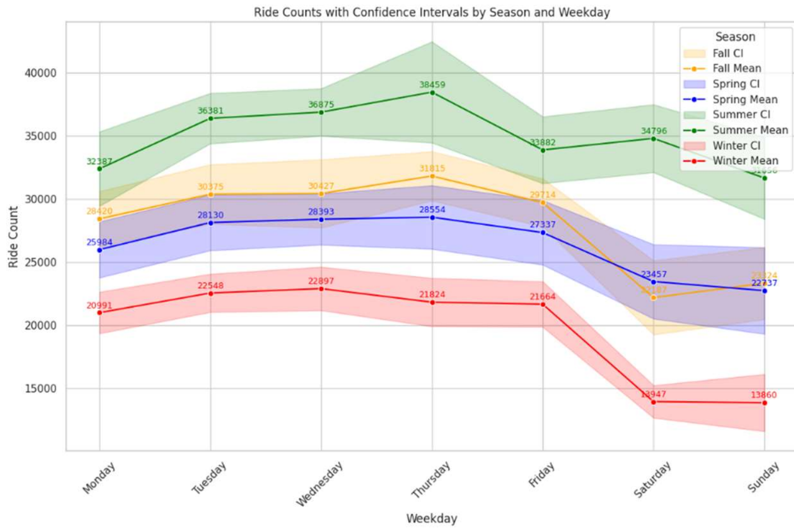
Weekday=Tuesday

Analysis Variable : Ride_Count					
Season	N Obs	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
Fall	26	30374.96	5822.31	28023.28	32726.64
Spring	27	28130.19	5573.88	25925.23	30335.14
Summer	26	36380.92	4942.07	34384.78	38377.07
Winter	26	22547.92	3744.96	21035.30	24060.54

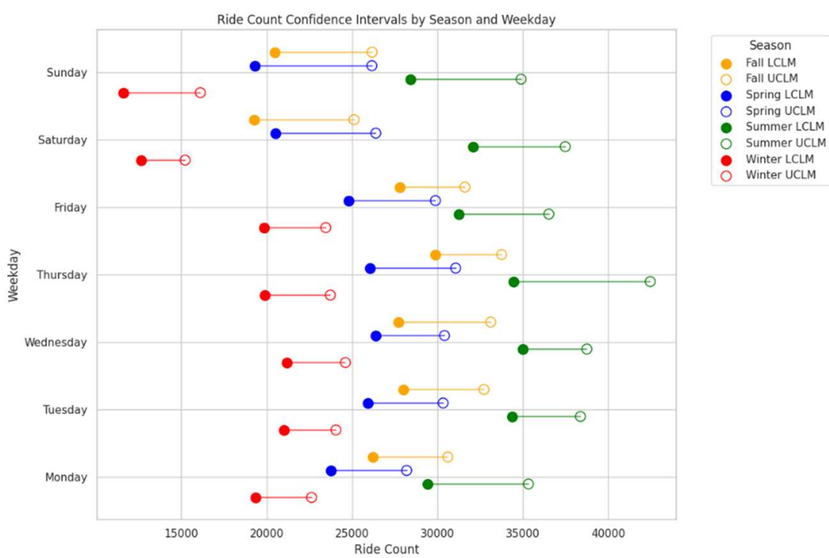
Weekday=Wednesday

Analysis Variable : Ride_Count					
Season	N Obs	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
Fall	26	30426.62	6680.91	27728.14	33125.09
Spring	26	28392.58	5015.78	26366.66	30418.50
Summer	27	36874.63	4748.90	34996.03	38753.23
Winter	25	22896.80	4164.96	21177.59	24616.01

Here is an interesting visualization of the Ride counts Confidence Interval, by season and weekday. I have provided mean, upper bound (upper shaded region line) and lower bound (lower shaded region line). The legend shows the season. We see summer season is very good for the business and winter season the worst. This was done using confidence interval data and python programming



Another interesting visual called Dumbbell plot. This plot shows the upper & Lower bounds of the intervals in the form of a dumbbell.



Modelling

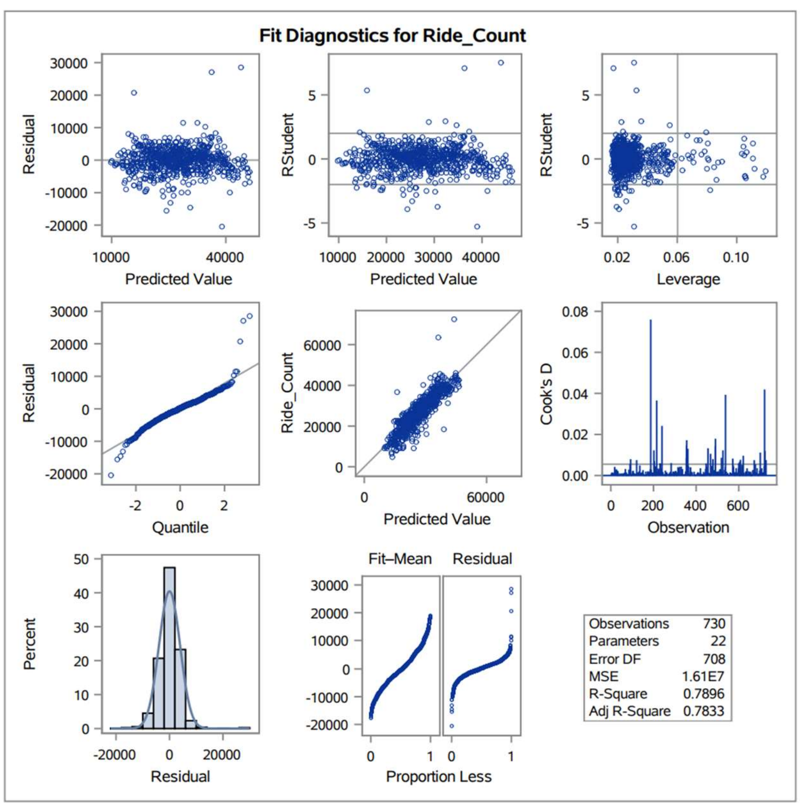
Now that we have an understanding of our data and variables, let us try to build a linear regression model to predict future bicycle rentals based on the available variables.

Initial Model (Pre – selection):

The Initial SAS linear regression analysis using all variables except Ride_count as independent variables and Ride_count as dependent variables yielded the following results.

Root MSE	4006.55285
Dependent Mean	27268
R-Square	0.7896
Adj R-Sq	0.7833
AIC	12865
AICC	12867
SBC	12234

Parameter Estimates									
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation	95% Confidence Limits	
Intercept	Intercept	B	-4139.15788	15621	-0.26	0.7911	0	-34808	26530
Date	Date	1	2.31107	0.76879	3.01	0.0027	1.19635	0.80169	3.82045
T1	T1	1	506.32719	296.82681	1.71	0.0885	105.60311	-76.43892	1089.09330
T2	T2	1	196.51738	240.22446	0.82	0.4136	99.23151	-275.12017	668.15494
Hum	Hum	1	-314.21269	21.88351	-14.36	<.0001	1.96321	-357.17704	-271.24834
Wind_speed	Wind_speed	1	-267.01318	28.69222	-9.31	<.0001	1.44722	-323.34519	-210.68118
Holiday	Holiday	1	-7153.06781	1059.61589	-6.75	<.0001	1.09459	-9233.43319	-5072.70243
Weekend	Weekend	B	-6373.95842	558.48928	-11.41	<.0001	2.89834	-7470.45176	-5277.46508
Weather_code Broken Clouds	Weather_code Broken Clouds	B	4922.40139	1333.49820	3.69	0.0002	7.01071	2304.31732	7540.48546
Weather_code Clear	Weather_code Clear	B	4794.99340	1454.73169	3.30	0.0010	3.67128	1938.88915	7651.09765
Weather_code Cloudy	Weather_code Cloudy	B	5154.64603	1275.46933	4.04	<.0001	12.13755	2650.49121	7658.80084
Weather_code Few Clouds	Weather_code Few Clouds	B	4431.21667	1341.36046	3.30	0.0010	6.54456	1797.69646	7064.73688
Weather_code Rain	Weather_code Rain	B	1920.82972	1244.11738	1.54	0.1231	17.58651	-521.77118	4363.43062
Weather_code Snow	Weather_code Snow	B	2975.39523	1451.67392	2.05	0.0408	3.89684	125.29437	5825.49608
Weather_code Thunderstorm	Weather_code Thunderstorm	0	0	.	.	.	.	.	.
Season Fall	Season Fall	B	2097.42749	507.94347	4.13	<.0001	2.18785	1100.17178	3094.68321
Season Spring	Season Spring	B	1350.35114	476.34658	2.83	0.0047	1.94532	415.13025	2285.57204
Season Summer	Season Summer	B	3041.76354	673.43427	4.52	<.0001	3.88809	1719.59638	4363.93070
Season Winter	Season Winter	0	0	.	.	.	.	.	.
Weekday Friday	Weekday Friday	B	-1453.32112	563.45961	-2.58	0.0101	1.74971	-2559.57280	-347.06945
Weekday Monday	Weekday Monday	B	-1598.63189	566.92385	-2.82	0.0049	1.79992	-2711.68499	-485.57879
Weekday Saturday	Weekday Saturday	B	801.04511	557.80858	1.44	0.1514	1.72868	-294.11180	1896.20201
Weekday Sunday	Weekday Sunday	0	0	.	.	.	.	.	.
Weekday Thursday	Weekday Thursday	B	148.11040	558.47048	0.27	0.7909	1.73278	-948.34603	1244.56683
Weekday Tuesday	Weekday Tuesday	B	-337.91886	561.04058	-0.60	0.5472	1.76276	-1439.42122	763.58349
Weekday Wednesday	Weekday Wednesday	0	0	.	.	.	.	.	.



Final Model (Post Forward Selection):

The Forward selection method helps us identify the final model. We can observe the following changes in our new model:

1. T2 and weekend are removed, as T2 has high p-value and is redundant in analysis given T2 is based on humidity, T1, and actual wind speed. Weekend is also redundant to weekday variable.

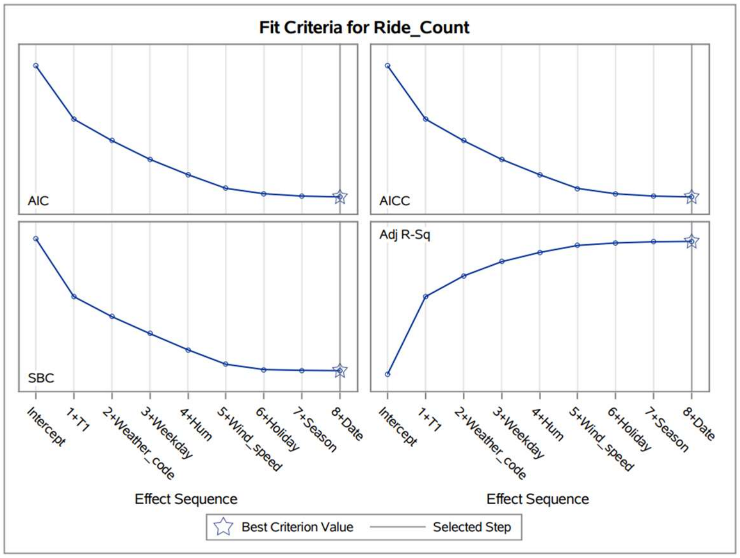
2. Adj. R^2 hasn't decreased (0.7833 → 0.7834)

3. AIC and SBC have both improved slightly.

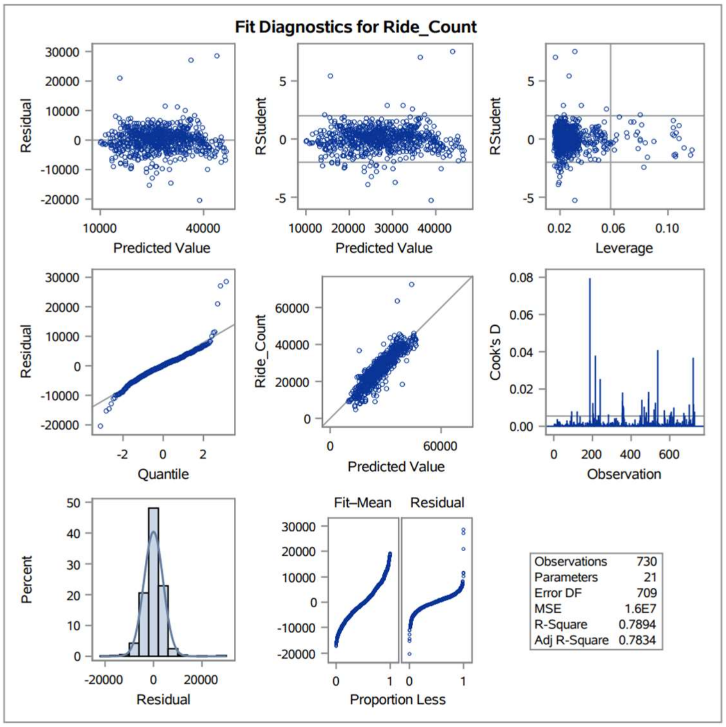
4. RMSE has also reduced.

Considering all the above changes, we can consider the selection based model as more suitable for analysis.

Root MSE	4005.61811
Dependent Mean	27268
R-Square	0.7894
Adj R-Sq	0.7834
AIC	12864
AICC	12865
SBC	12229



Parameter Estimates							
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	Intercept	B	-4464.40360	15612	-0.29	0.7750	0
T1	T1	1	746.01511	47.52015	15.70	<.0001	2.70788
Hum	Hum	1	-309.90753	21.23634	-14.59	<.0001	1.84967
Wind_speed	Wind_speed	1	-273.81811	27.45352	-9.97	<.0001	1.32558
Holiday	Holiday	1	-7143.10587	1059.29872	-6.74	<.0001	1.09444
Weather_code Broken Clouds	Weather_code Broken Clouds	B	5026.26216	1327.13090	3.79	0.0002	6.94716
Weather_code Clear	Weather_code Clear	B	4828.05882	1453.83080	3.32	0.0009	3.66844
Weather_code Cloudy	Weather_code Cloudy	B	5242.64745	1270.62832	4.13	<.0001	12.05121
Weather_code Few Clouds	Weather_code Few Clouds	B	4512.72257	1337.34298	3.37	0.0008	6.50845
Weather_code Rain	Weather_code Rain	B	1982.73215	1241.52432	1.60	0.1107	17.52145
Weather_code Snow	Weather_code Snow	B	2889.29608	1447.51578	2.00	0.0463	3.87635
Weather_code Thunderstorm	Weather_code Thunderstorm	0	0	.	.	.	.
Season Fall	Season Fall	B	2125.67795	506.64996	4.20	<.0001	2.17773
Season Spring	Season Spring	B	1385.51989	474.29197	2.92	0.0036	1.92948
Season Summer	Season Summer	B	2979.14263	668.91346	4.45	<.0001	3.83785
Season Winter	Season Winter	0	0	.	.	.	.
Weekday Friday	Weekday Friday	B	-1459.28283	563.28103	-2.59	0.0098	1.74942
Weekday Monday	Weekday Monday	B	-1597.94839	566.79097	-2.82	0.0049	1.79992
Weekday Saturday	Weekday Saturday	B	-5588.15188	556.75028	-10.04	<.0001	1.72293
Weekday Sunday	Weekday Sunday	B	-6371.17202	558.34860	-11.41	<.0001	1.74670
Weekday Thursday	Weekday Thursday	B	141.73157	558.28576	0.25	0.7997	1.73244
Weekday Tuesday	Weekday Tuesday	B	-323.81345	560.64473	-0.58	0.5637	1.76110
Weekday Wednesday	Weekday Wednesday	0	0	.	.	.	.
Date	Date	1	2.27852	0.76758	2.97	0.0031	1.19315



Model Description:

Let us discuss the impact of the variables on our final model:

1. T1 (Temperature) – Positive effect.
   More temperature = More bike rides.

2. Hum (Humidity) – negative effect.
   More humidity = lesser ride counts.

3. wind speed – Negative effect
   More wind = lesser ride counts.

4. Holiday – negative effect
   Fewer rides on holidays.

5. Weather code - Positive impact from clear/partly cloudy and negative impact from snow/thunderstorm.

6. Season - warmer the season, better the bike rental.

7. weekday - High on weekdays, low on weekends.

8. Date - slight positive effect reflecting a positive growing adoption.

Bicycle rental improvement strategies

Benefits for the Business:

My Analysis provides a 360° perspective of a Bike rental business (based on the given data).

A Bicycle rental business will able to understand the tends and patterns of their data based on my analysis.

Furthermore, a business can also benefit from visualizations provided in the project. The Regression model developed, helps the business predict it's future rentals. The solutions given at the end also helps the business move forward.

Business Improvement strategies:

Some strategies the company can undertake are:

1. Business should target warm, low humidity days heavily and run subsidies on other days to boost rentals.

2. Increase fleet availability and brand visibility during peak seasons like summer and Fall.

3. Weekdays bring in more rentals and hence, the business should do different promotion strategies like weekend cycling events to boost rentals for weekends.

4. Holidays see a loss in rentals and hence special packages can be run to boost rentals.

Additional Information that maybe needed:

Some additional data that can help:

1. Time of rentals - to better understand day usage.

2. Distance travelled - to understand usage patterns.

3. User Demographics - to understand customer preferences.

4. Traffic data - to understand rental ride counts.

5. Bike availability - to get real time data about bike usage.

Final Thoughts

This project involved a comprehensive analysis of data and a rental prediction model. Furthermore business improvement strategies were also discussed for better rentals. This analysis can be refined further by adopting an even more complex model.