Riding the Trend: Mapping Bike Rental Demand

Yvanna Tchomnou, Sarang Deshpande, Majesti Davis, Eric Sheehan, Ethan Lozuke

University of North Carolina at Charlotte

INFO 3236 - Section 002

Dr. Dongsong Zhang

12/05/2023

Part 1. Understanding the Business and Data

Section 1: Introduction to the Bike Rental Business

Pedaling through the urban areas, bike rentals have brought together a perfect trifecta: convenience, sustainability, and accessibility. The bike rental industry, which falls under the service sector, “helps convert pedestrians into cyclists” by allowing temporary access to a variety of bikes, whether that be the classic city bike or more specialized options like mountain cycles or electric bikes (Normack et al., 2018). You will often find these bikes zipping around urban areas, tourist hotspots, and various recreational areas. They have garnered popularity among individuals as they provide a competitively priced, flexible and sustainable mode of transportation that is often placed in convenient locations allowing for maximum visibility and access for those willing to give them a try.

The history of the bike rentals system can be traced back to 1965 in Amsterdam, Netherlands in which an environmental organization planned to combat the traffic problem persisting in Amerstadam's inner city area (Shaheen et al., 2010). These bikes were painted white and left unlocked throughout the public area to encourage free-use. However this proved to be unsuccessful as these bikes were often stolen or left damaged. Nonetheless, as time passed and with the introduction of innovative technology and e-bikes, bike rental businesses & organizations alike saw a “sharp increase in both their prevalence and popularity worldwide” as they are now able regulate the use of their bikes whilst still incurring a profit (Fishman et al., 2013). For tourists, bike rentals serve as a means to explore new cities and scenic areas, offering a rather unique yet immersive experience. On the other hand, locals may opt to use rentals for recreational purposes, such as integrating them into a part of their daily commuting routine, or as an environmentally conscious alternative to owning a bike. This dual appeal contributes to the resilience and adaptability of bike rental businesses.

Section 2: Literature Review and Group Hypothesis

So how do these bike rentals work? In most urban areas, prospective riders can walk up to the biking station dock in which either a physical kiosk or mobile app can be used to purchase either as a pay-as-you-go price or a subscription to rent a bicycle, which will then unlock and renters can utilize the bikes for their desired ride time is over in which they will drop off and lock the bike at a nearby docking station (Sood, 2011). Riding subscriptions offer short-term options in which an individual can choose to ride from 30 minutes to several hours as well as some long-term subscriptions options that may be rented from a weekly to monthly basis, making it convenient for those utilizing these services.

The bike rental industry is growing at increasingly promising rates. In 2021, the bike rental market was reported to be valued at $2.1 billion and is estimated to “reach $11.3 billion by 2031,” growing at a compound annual growth rate of 18.5% (Bike Rental Market Size, Share, Competitive Landscape and Trend Analysis, 2022). The primary product offered by bike rental businesses is undoubtedly an assortment of bicycles that can be utilized by prospective customers. Nevertheless, some bike rental businesses also offer a variety of other products and services that may be less evident such as complementary helmets, brand merchandise, guided tours, biking gear, as well as fitting services to make minor adjustments to bikes to best tailor to each specific rider. Furthermore some companies even offer the opportunity for a customer to also potentially purchase a rental bike of their own. The diversification of their product base has widened their reach of potential customers but has also enabled opportunities for an increase in overall profitability of these companies and organizations.

As mentioned above, bike rental businesses provide an accessible and convenient mode of transportation to a wide range of individuals, but what segment of the consumer market do these businesses typically target? Bike rental businesses can range from traditional bike rental shops to modern bike sharing services in which self-service bike docking stations are placed in various locations. However these businesses tend to center themselves in urban city and touristic areas with population densities. The customer base for bike rental services exhibits a diverse composition, consisting of tourists who often turn to these bikes for recreational riding, as a means to explore new destinations at their own pace (Yglesias, 2014). On the other hand, within booming city centers, customers for these businesses tend to be locals who leverage these services as a means of transportation for their daily commute to work or school and as well as those looking for more environmentally conscious modes of transportation from Point A to B (Sood, 2011).

Understanding the factors that shape demand is pivotal for sustained success in the bike rental business. Research indicates that there are a multitude of factors that can influence the overall demand for bike rentals. The most common factors include the weather conditions at the time, location of rental station, temperature, the day of the week, humidity levels, as well as seasonal periods. As a case in point, a Poisson regression model analysis performed on a bike sharing system located in Park City, Utah found that higher daily temperatures were positively related to “higher rates of e-bike ridership” and that trips and overall bike rentals were generate on “weekdays” during more summer “months” (He et al., 2022). Furthermore, it was also found that the regression results showed that the volume of bike rentals made were often higher “near public transit, recreational centers," as well as areas with a high population density, which are often city centers and other urban areas (He et al., 2022).

In conclusion, bike rental businesses, which were once a simple service providing temporary access to bicycles, have now evolved into a dynamic industry at the intersection of sustainability and technology. Based on a comprehensive literature and research review, we infer that *casual, datetime, holiday season, windspeed, workingday, weather, temp, atemp, humidity, and registered* are the most influential factors affecting bike rental demand. Thus, with the "Bike Rental Dataset Fields.pdf" dataset we will be analyzing, we expect *datetime, season, workingday, weather, temp, atemp, humidity,* and *registered* as to be the independent variables that will serve as the most significant predictors affecting the overall demand of bike rentals. Furthermore we also infer that independent variables, *casual, windspeed and holiday,* will not be as accurate predictor variables towards bike rental demand.

Part 2. Building and Interpreting Analytic Models

Section 3: Data Preparation

For our bike dataset models, we introduced three new variables. For DemandType, divided the data into high and low demand categories based on a median demand count of 145. High demand records exceeded this median, while low demand records were those at or below 145. Additionally, we incorporated two independent variables derived from the record date: Time of Day (Morning: 4:00 am - 11:00 am, Afternoon: 12:00 pm - 3:00 pm, Evening: 4:00 pm - 9:00 pm, Night: 10:00 pm - 3:00 am) and Day of Week (Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday). Which is shown in the image below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count	DemandType	Time of Day	Day of Week
2	1/1/2011 0:00	1	0	0	1	9.84	14.395	81	0	3	13	16	low	Night	SAT
3	1/1/2011 1:00	1	0	0	1	9.02	13.635	80	0	8	32	40	low	Night	SAT
4	1/1/2011 2:00	1	0	0	1	9.02	13.635	80	0	5	27	32	low	Night	SAT
5	1/1/2011 3:00	1	0	0	1	9.84	14.395	75	0	3	10	13	low	Night	SAT
6	1/1/2011 4:00	1	0	0	1	9.84	14.395	75	0	0	1	1	low	Morning	SAT
7	1/1/2011 5:00	1	0	0	2	9.84	12.88	75	6.0032	0	1	1	low	Morning	SAT
8	1/1/2011 6:00	1	0	0	1	9.02	13.635	80	0	2	0	2	low	Morning	SAT
9	1/1/2011 7:00	1	0	0	1	8.2	12.88	86	0	1	2	3	low	Morning	SAT
10	1/1/2011 8:00	1	0	0	1	9.84	14.395	75	0	1	7	8	low	Morning	SAT
11	1/1/2011 9:00	1	0	0	1	13.12	17.425	76	0	8	6	14	low	Morning	SAT
12	1/1/2011 10:00	1	0	0	1	15.58	19.695	76	16.9979	12	24	36	low	Morning	SAT
13	1/1/2011 11:00	1	0	0	1	14.76	16.665	81	19.0012	26	30	56	low	Morning	SAT
14	1/1/2011 12:00	1	0	0	1	17.22	21.21	77	19.0012	29	55	84	low	Afternoon	SAT
15	1/1/2011 13:00	1	0	0	2	18.86	22.725	72	19.9995	47	47	94	low	Afternoon	SAT
16	1/1/2011 14:00	1	0	0	2	18.86	22.725	72	19.0012	35	71	106	low	Afternoon	SAT
17	1/1/2011 15:00	1	0	0	2	18.04	21.97	77	19.9995	40	70	110	low	Afternoon	SAT
18	1/1/2011 16:00	1	0	0	2	17.22	21.21	82	19.9995	41	52	93	low	Evening	SAT
19	1/1/2011 17:00	1	0	0	2	18.04	21.97	82	19.0012	15	52	67	low	Evening	SAT
20	1/1/2011 18:00	1	0	0	3	17.22	21.21	88	16.9979	9	26	35	low	Evening	SAT
21	1/1/2011 19:00	1	0	0	3	17.22	21.21	88	16.9979	6	31	37	low	Evening	SAT
22	1/1/2011 20:00	1	0	0	2	16.4	20.455	87	16.9979	11	25	36	low	Evening	SAT
23	1/1/2011 21:00	1	0	0	2	16.4	20.455	87	12.998	3	31	34	low	Evening	SAT
24	1/1/2011 22:00	1	0	0	2	16.4	20.455	94	15.0013	11	17	28	low	Night	SAT
25	1/1/2011 23:00	1	0	0	2	18.86	22.725	88	19.9995	15	24	39	low	Night	SAT
26	1/2/2011 0:00	1	0	0	2	18.86	22.725	88	19.9995	4	13	17	low	Night	SUN

When determining the set of proportions used for training and validation for the predictive models by the group, we set our data allocations to 40% Training, 30% validation, and 30% Test as indicated in the image below.

Data Set Allocations	
Training	40.0
Validation	30.0
Test	30.0

Section 4: Decision Models

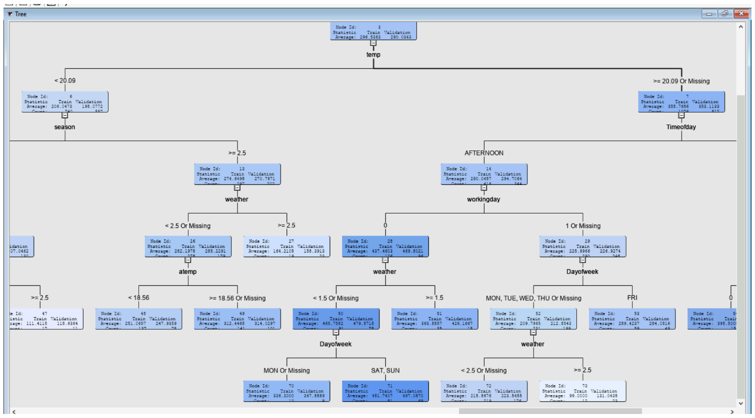
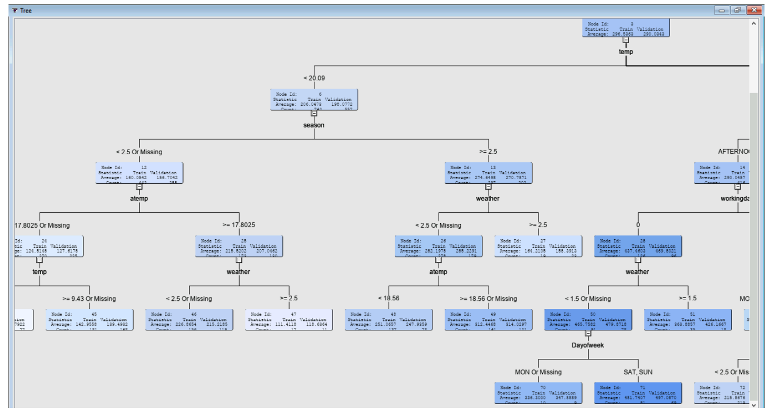
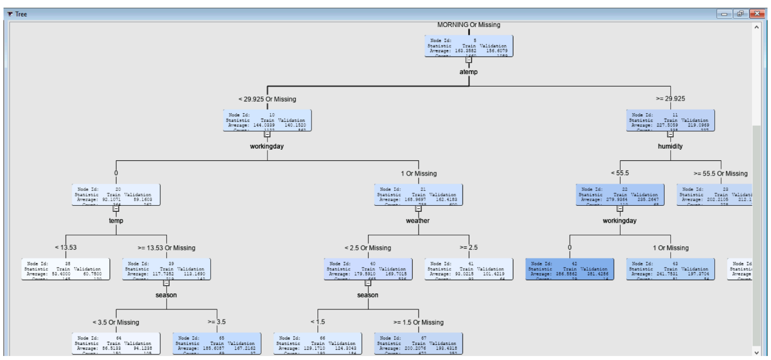
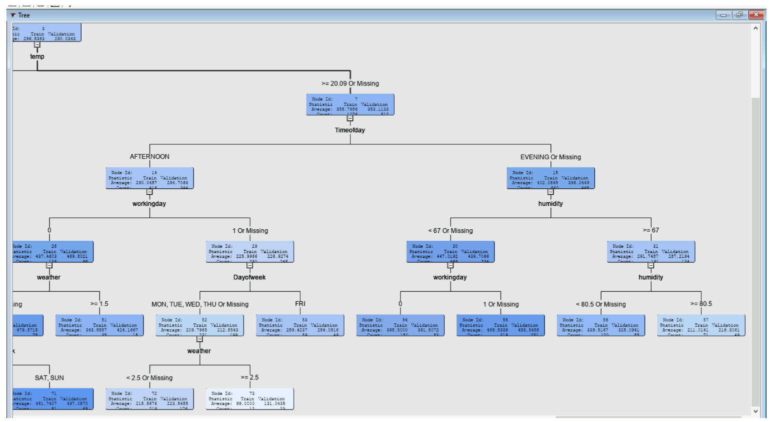
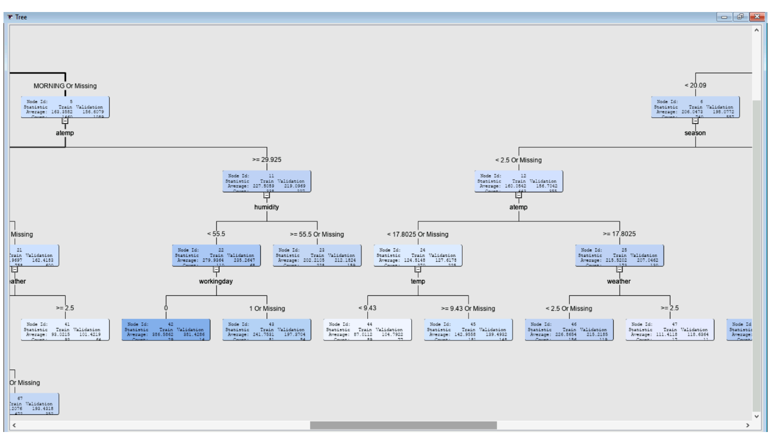
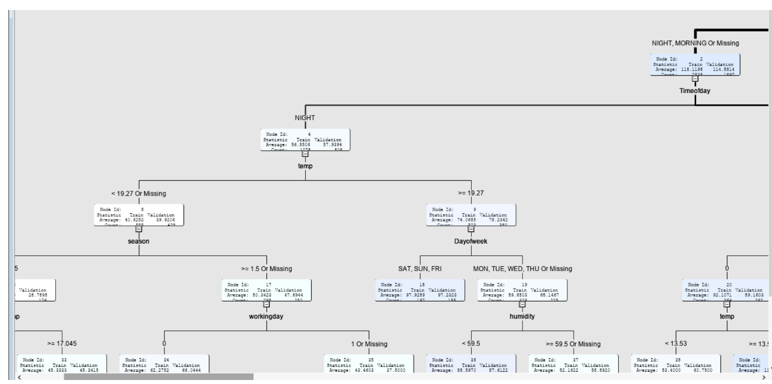
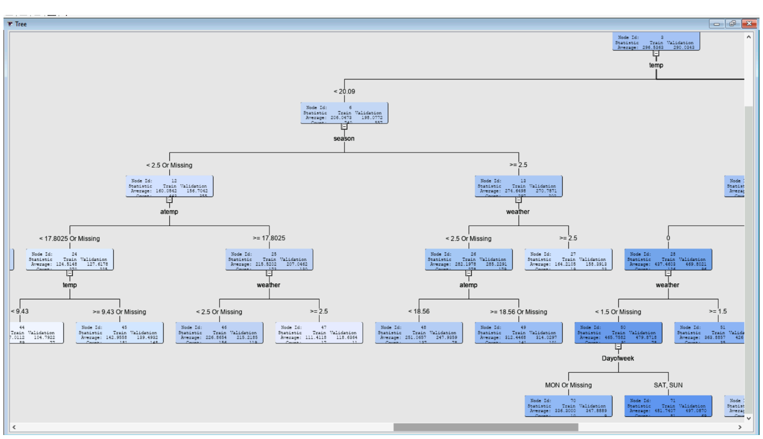
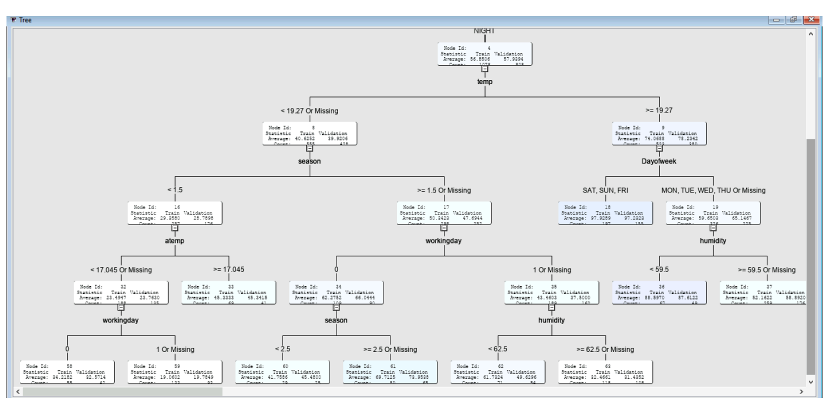
Decision Model #1.

Variable List

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
Dayofweek	Input	Nominal	No		No	.	.
DemandType	Rejected	Nominal	No		No	.	.
Timeofday	Input	Nominal	No		No	.	.
atemp	Input	Interval	No		No	.	.
casual	Rejected	Interval	No		No	.	.
count	Target	Interval	No		No	.	.
datetime	Time ID	Interval	No		No	.	.
holiday	Rejected	Binary	No		No	.	.
humidity	Input	Interval	No		No	.	.
registered	Rejected	Interval	No		No	.	.
season	Input	Interval	No		No	.	.
temp	Input	Interval	No		No	.	.
weather	Input	Interval	No		No	.	.
windspeed	Rejected	Interval	No		No	.	.
workingday	Input	Binary	No		No	.	.

With this first decision tree we wanted to focus on variables that we felt our research would confirm. Included in this are the variables of *datetime*, *atemp, holiday, humidity, registered, season, workingday, temp and weather*. These variables shown within our research proved to have the largest impact on the demand of bicycle rentals. From this group of variables we chose to keep *weather, temp, atemp, season* and *humidity* as our selected inputs for the DT and *datetime* fit in the role of TimeID because of the nature of the variable. The variables *casual, holiday* and *windspeed* were independent variables that we deemed unnecessary to include because of their inability to correspond to an effect in the demand of bike rentals. The *registered* and *demandtype* variables were variables that we assessed and saw that they would skew the data if included so these were also rejected in our variable list. Here we used *count* and *average square error* as the variables to complete our assessment measure.

Decision Model Results

Variables in my optimal decision tree:

- Time of Day : (Morning: 4:00 am - 11:00 pm, Afternoon: 12:00 pm - 3:00 pm, Evening: 4:00 pm - 9:00 pm, Night: 10:00 pm - 3:00 am)
- ATemp : "feels like" Temperature in Celsius
- Season: 1 = spring, 2 = summer, 3 = fall, 4 = winter
- Weather
    - 1: Clear, Few clouds, Partly cloudy, Partly cloudy
    - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
    - 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
    - 4: Heavy Rain + Ice Pellets + Thunderstorm + Mist, Snow + Fog
    - Humidity: Relative Humidity
    - Dependent Variable: Count is the variable that exists to represent the number of rentals.

Model Performance Results

Fit Statistics

Target=count Target Label=' '

Fit Statistics	Statistics Label	Train	Validation	Test
NOBS	Sum of Frequencies	4354.00	3266.00	3266.00
MAX	Maximum Absolute Error	638.79	634.79	633.79
SSE	Sum of Squared Errors	65910261.95	51835444.00	54621220.55
ASE	Average Squared Error	15137.86	15871.23	16724.19
RASE	Root Average Squared Error	123.04	125.98	129.32
DIV	Divisor for ASE	4354.00	3266.00	3266.00
DFT	Total Degrees of Freedom	4354.00	.	.

***Average Square Error*:** 15871.23

* There was no decision matrix in the output results which prevented us from completing accuracy calculations.

Variable Importance

Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
Timeofday		3	1.0000	1.0000	1.0000
temp		4	0.5021	0.5218	1.0392
workingday		6	0.3811	0.4259	1.1176
humidity		5	0.3224	0.2918	0.9053
season		5	0.2850	0.2837	0.9955
atemp		4	0.2600	0.2450	0.9426
weather		5	0.1836	0.1656	0.9018
Dayofweek		3	0.1231	0.1226	0.9958

Top 5 Variables:

1. Time of Day : (Morning: 4:00 am - 11:00 pm, Afternoon: 12:00 pm - 3:00 pm, Evening: 4:00 pm - 9:00 pm, Night: 10:00 pm - 3:00 am)
2. Temp : Temperature in Celcius
3. Working Day : whether the day is neither a weekend or a holiday
4. Humidity : relative humidity
5. Season : 1 = spring, 2 = summer, 3 = fall, 4 = winter

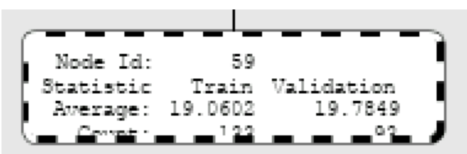
Best and Worst Rental Outcomes

Node with the Best Rental Outcome:

Rule: WHERE Timeofday AFTERNOON, EVENING AND temp >= 20.09 Or Missing AND Timeofday AFTERNOON AND workingday 0 AND weather < 1.5 Or Missing AND Daysofweek SAT, SUN

Node with the Worst Rental Outcome:

Rule: WHERE Timeofday NIGHT, MORNING Or Missing AND Timeofday NIGHT AND temp < 19.27 Or Missing AND season < 1.5 AND atemp < 17.045 Or Missing AND workingday 1 Or Missing



Interpretation of the data

It is important as a business to be able to look at the analytical and data driven side of their business to be able to take advantage of profits. This can be done in the best possible situations for demand or the worst as we are presented above. It can be inferred in our best outcome through the decision rule that demand rates in the afternoon or evening of nice weather weekends will tend to be higher. To take advantage of this I feel that offering higher rental rates on these days will be a successful angle. This is solely because of the fact that the demand is clearly present and if the conditions like that exist then the resulting demand will be present. On the other hand it appears that in mornings or nights where there are colder temperatures, in the spring on days that are Monday-Friday individuals are less inclined to rent bicycles. Here in our worst case rental outcome we are able to make the best out of this situation by offering certain incentives. For example for the working days as a whole we can offer a discount rate for renting a bicycle the entire week or discounts throughout days of the week. This can help to sustain sales and demand over the course of the days of the week that statistically have the lowest demand. In addition, shaping our business around the “Go Getters” and early commuter individuals can help to engage and capture more of the customers in that market.

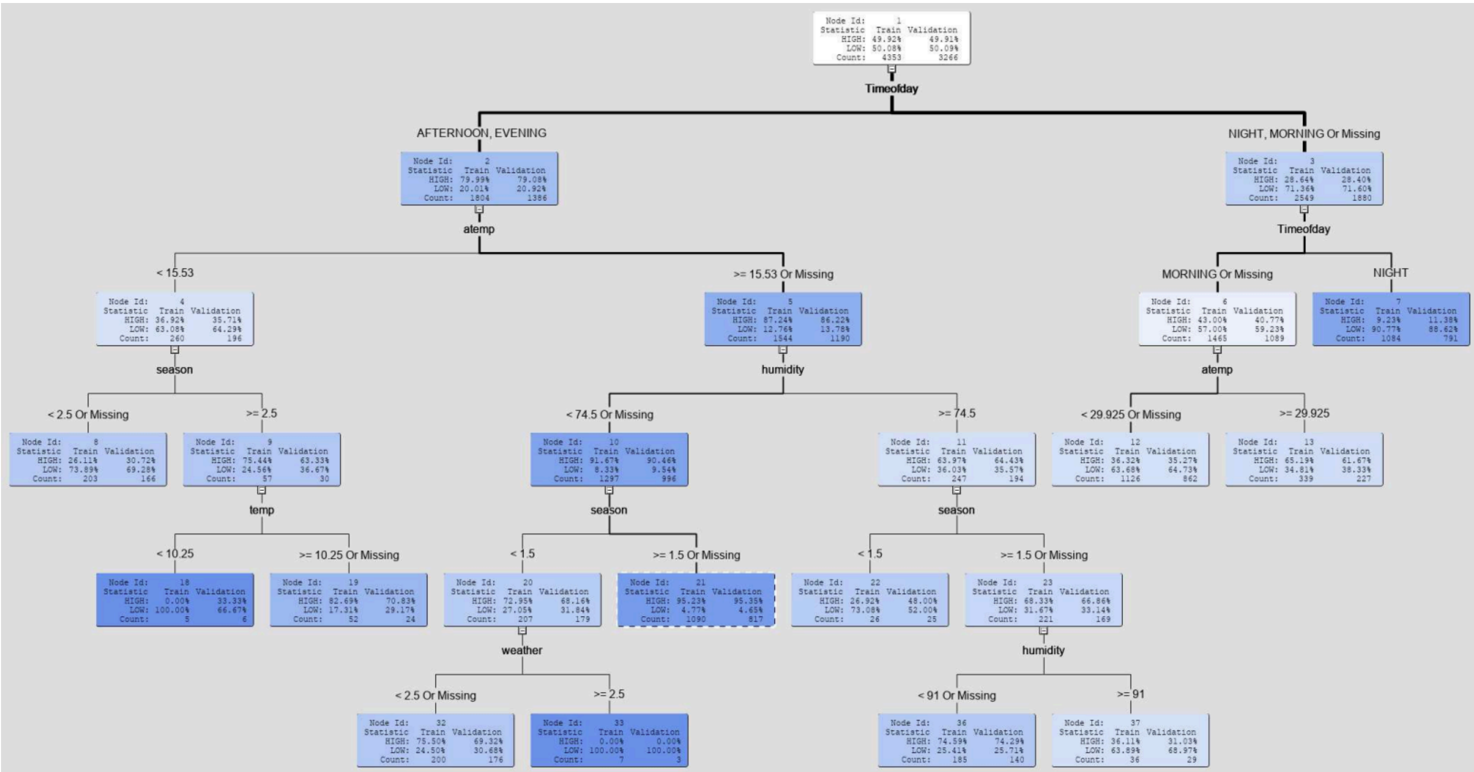
Decision Model #2.

Variable List

Name	Role	Level	Report	Order	Drop
Dayofweek	Input	Nominal	No		No
DemandType	Target	Nominal	No		No
Timeofday	Input	Nominal	No		No
atemp	Input	Interval	No		No
casual	Rejected	Interval	No		No
count	Rejected	Interval	No		No
datetime	Time ID	Interval	No		No
holiday	Rejected	Binary	No		No
humidity	Input	Interval	No		No
registered	Rejected	Interval	No		No
season	Input	Interval	No		No
temp	Input	Interval	No		No
weather	Input	Interval	No		No
windspeed	Rejected	Interval	No		No
workingday	Input	Binary	No		No

Based on the overall hypothesis in which we concluded in the first section. Our group thus, based on the bike rental dataset received along with a comprehensive literature review completed, inferred that *datetime, season, workingday, weather, temp, atemp, humidity,* and *registered* are the variables that would serve as the most significant predictors affecting the overall demand of bike rentals. Therefore we chose to have *season, workingday, weather, temp, atemp, humidity* as possible inputs and datetime as TimeID. Furthermore we also infer that independent variables, *casual, windspeed and holiday,* will not be very accurate predictor variables towards bike rental demand and therefore we chose to reject these variables. In catering towards my specific model I chose to reject *registered*, and *count* given that including these variables when included in our model skewed our data set and resulted in inaccurate possible decision nodes from our decision tree which didn't make sense in real life.

Decision Model Results



What variables are included in the optimal tree?

- Time of Day: (Morning: 4:00 am - 11:00 pm, Afternoon: 12:00 pm - 3:00 pm, Evening: 4:00 pm - 9:00 pm, Night: 10:00 pm - 3:00 am)
- ATemp : "feels like" Temperature in Celsius
- Season: 1 = spring, 2 = summer, 3 = fall, 4 = winter
- Weather
    - 1: Clear, Few clouds, Partly cloudy, Partly cloudy
    - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
    - 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
    - 4: Heavy Rain + Ice Pellets + Thunderstorm + Mist, Snow + Fog

- Humidity: Relative Humidity
- Dependant Variable: Demand Type: High where greater than 145 and Low Demand less than or equal to 145

Top 5 Variables in the order of decreasing importance

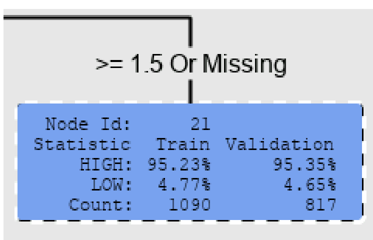
1.) Time of Day: (Morning: 4:00 am - 11:00 pm, Afternoon: 12:00 pm - 3:00 pm, Evening: 4:00 pm - 9:00 pm, Night: 10:00 pm - 3:00 am)
2.) Atemp: "feels like" Temperature in Celsius
3.) Season: 1 = spring, 2 = summer, 3 = fall, 4 = winter
4.) Humidity: Relative Humidity
5.) Weather:
    a.) 1: Clear, Few clouds, Partly cloudy, Partly cloudy
    b.) 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
    c.) 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
    d.) 4: Heavy Rain + Ice Pellets + Thunderstorm + Mist, Snow + Fog

Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
Timeofday		2	1.0000	1.0000	1.0000
atemp		2	0.4725	0.4778	1.0113
season		3	0.2590	0.2269	0.8759
humidity		2	0.2415	0.2518	1.0425
weather		1	0.1050	0.0689	0.6561
temp		1	0.0945	0.0213	0.2255

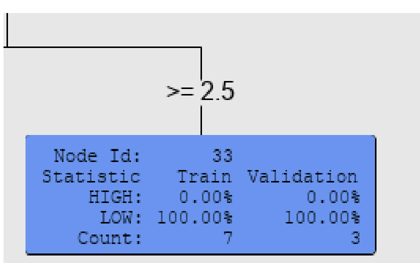
Best and Worst Rental Outcomes

Node with the Best Rental Outcome:



Rule: WHERE Timeofday AFTERNOON, EVENING AND atemp >= 15.53 Or Missing AND humidity < 74.5 Or Missing AND season >= 1.5 Or Missing

Node with the Worst Rental Outcome:



Rule: WHERE Timeofday AFTERNOON, EVENING AND atemp >= 15.53 Or Missing AND humidity < 74.5 Or Missing AND season >= 1.5 Or Missing AND weather >= 2.5

Decision Matrix Computations:

Sensitivity = TP/(TP+FN) = (1414)/(1414+222) = 86.43%

Specificity = TN/(TN+FP) = (1162)/(1162+468) = 71.23%

Accuracy = TP+TN/Total Predicted = (1414 + 1162)/(1414 + 1162 + 222 + 468) = 2576 / 3266 = 78.87%

Misclassification = FP+FN/Total Predicted = (468+222)/(222+1162+468+1414) = 690/3266 = 21.13%

```
Data Role=VALIDATE Target=DemandType Target Label=' '

  False       True        False       True
Negative    Negative    Positive    Positive

   222        1162         468        1414
```

Interpretation

In order to grow our businesses and increase bike sales, we suggest following the decision rules for best bike rental outcomes. As during optimal settings and peak times we should provide more bikes available for more bike accessibility additionally increase prices for bikes during these optimal peak times so that we gain an increase per sale on our rentals. As this is a similar tactic used by uber as they have a peak time service charge during busy hours and areas. While during worst outcomes we could offer a discount on our bike service rentals to encourage customers to utilize our services compared to other products. Such as times when we have rain or fog conditions and when it's afternoon or evening. Lastly, we could possibly create a points reward system to not only increase users but also provide discounts and coupons to bikers who don't use our system consistently and during non peak hours to encourage customer usage.

Which is a combined strategy that rewards existing and new customers to utilize our bikes in non optimal conditions.

Section 5: Logistic Regression Model

Before getting into the model, let's take a look at the data included for the logistic regression. Since our model is predicting whether DemandType is high or low and DemandType is derived directly from count, we are excluding the Count variable from the regression. Additionally, since Casual and Registered are also directly related to Cout, we are excluding those two variables from the model. Lastly, DateTime is directly related to DayOfWeek and TimeOfDay so we are excluding that variable from the model as well. Below is the screenshot of all of our variables being used in the regression. Lastly, for the regression model, we partitioned the data using a ratio of 50% training and 50% validation.

Name	Role	Level
Dayofweek	Input	Nominal
DemandType	Target	Binary
Timeofday	Input	Nominal
atemp	Input	Interval
casual	Rejected	Interval
count	Rejected	Interval
datetime	Rejected	Interval
holiday	Input	Binary
humidity	Input	Interval
registered	Rejected	Interval
season	Input	Nominal
temp	Input	Interval
weather	Input	Ordinal
windspeed	Input	Interval
workingday	Input	Binary

We used all three methods of model selection: Forward, Backward, and Stepwise. The model was identical between Forward and Stepwise, and the accuracy of the Backward selection was just 0.7% off from the Forward/Stepwise model with the only difference being that the Backward selection included WindSpeed. Going forward, we'll be looking strictly at the model that the Forward/Stepwise selection method created.

Before diving into the variables included and their effects on the model, let's get into the accuracy of the model. The image below shows the confusion matrix for validation, with the Accuracy, Specificity, Sensitivity, and Misspecification calculations below the image.

Data Role=VALIDATE Target=DemandType Target Label=' '

False Negative	True Negative	False Positive	True Positive
548	2162	556	2179

- Accuracy = (TP+TN)/Total = (2162+2179)/(548+556+2162+2179) = 79.72%
- Sensitivity = TP/(TP+FN) = 2179/(2179+548) = 79.90%
- Specificity = TN/(TN+FP) = 2162/(2162+556) = 79.54%
- Misspecification = (FP+FN)/Total = (556+548)/(548+556+2162+2179) = 20.28%

Below are the outputs from the model. One big note to keep in mind is that since our target variable (DemandType) is in the data as "Low" and "High," the SAS model automatically assigned "Low" the value of 1 and "High" the value of 0. Therefore, the odds ratios and coefficients show the odds for Low Demand. In other words, a low odds ratio estimate (below 1) would lead to a stronger association with High Demand and a high odds ratio estimate (above 1) would lead to a stronger association with Low Demand. The screenshots below are from the output in SAS.

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept		1	1.7280	0.2549	45.96	<.0001		5.629
Dayofweek	FRI	1	-0.4750	0.0944	25.32	<.0001		0.622
Dayofweek	MON	1	0.0797	0.0910	0.77	0.3811		1.083
Dayofweek	SAT	1	0.1059	0.0903	1.38	0.2405		1.112
Dayofweek	SUN	1	0.3525	0.0906	15.14	<.0001		1.423
Dayofweek	THU	1	-0.1121	0.0941	1.42	0.2337		0.894
Dayofweek	TUE	1	0.0813	0.0942	0.74	0.3885		1.085
Timeofday	Afternoon	1	-0.9589	0.0822	136.00	<.0001		0.383
Timeofday	Evening	1	-1.5538	0.0729	454.48	<.0001		0.211
Timeofday	Morning	1	0.1258	0.0603	4.35	0.0370		1.134
atemp		1	-0.1310	0.00775	285.95	<.0001	-0.6132	0.877
humidity		1	0.0260	0.00249	108.83	<.0001	0.2731	1.026
season	1	1	0.5603	0.0932	36.13	<.0001		1.751
season	2	1	-0.00193	0.0683	0.00	0.9775		0.998
season	3	1	0.3042	0.0942	10.44	0.0012		1.356
weather	1	1	-0.3225	0.0664	23.59	<.0001		0.724
weather	2	1	-0.3855	0.0666	33.49	<.0001		0.680

Odds Ratio Estimates

Effect		Point Estimate
Dayofweek	FRI vs WED	0.642
Dayofweek	MON vs WED	1.119
Dayofweek	SAT vs WED	1.148
Dayofweek	SUN vs WED	1.469
Dayofweek	THU vs WED	0.923
Dayofweek	TUE vs WED	1.120
Timeofday	Afternoon vs Night	0.035
Timeofday	Evening vs Night	0.019
Timeofday	Morning vs Night	0.104
atemp		0.877
humidity		1.026
season	1 vs 4	4.149
season	2 vs 4	2.365
season	3 vs 4	3.211
weather	1 vs 3	0.357
weather	2 vs 3	0.335

For categorical variables, SAS automatically chooses one of the categories to be the baseline for the rest of the categories to be compared to. This process was done by SAS using the last category alphabetically for each variable. Therefore, Wednesday was the baseline for DayOfWeek, Night was the baseline for TimeOfDay, Season 4 (Winter) was the baseline for

Season, and finally Weather rating 3 was the baseline for Weather. The odds ratio estimates show how each one of the categories for each variable performs against the baseline (reminder that a lower odds ratio is related to higher demand). Using these results, we can isolate each variable's effect on demand.

For the two variables that interval data, we see that aTemp has an odds ratio of 0.877. This means that the odds of Low Demand are 0.877 for every 1 degree (C) increase in "feels like" temperature. On the flip side, Humidity has an odds ratio of 1.026, so the odds of Low Demand are 1.026 for every 1% increase in humidity. This tells us that warmer and drier days lead to higher demand.

For the categorical/ordinal variables, we can create rankings in ascending order of how closely each category is related to high demand based on the odds ratios compared to the baseline. Below are the rankings for DayOfWeek, TimeOfDay, Season, and Weather, with the odds ratio next to each category (lower odds ratio leads to higher demand).

Weekdays with Most-to-Least Demand

1. Friday (0.642)
2. Thursday (0.923)
3. Wednesday (baseline)
4. Tuesday (1.120)
5. Monday (1.119)
6. Saturday (1.148)
7. Sunday (1.469)

Time of Day with Most-to-Least Demand

1. Evening (0.019)
2. Afternoon (0.035)
3. Morning (0.104)
4. Night (baseline)

Seasons with Most-to-Least Demand

1. Winter (baseline)
2. Summer (2.365)
3. Fall (2.635)
4. Spring (4.149)

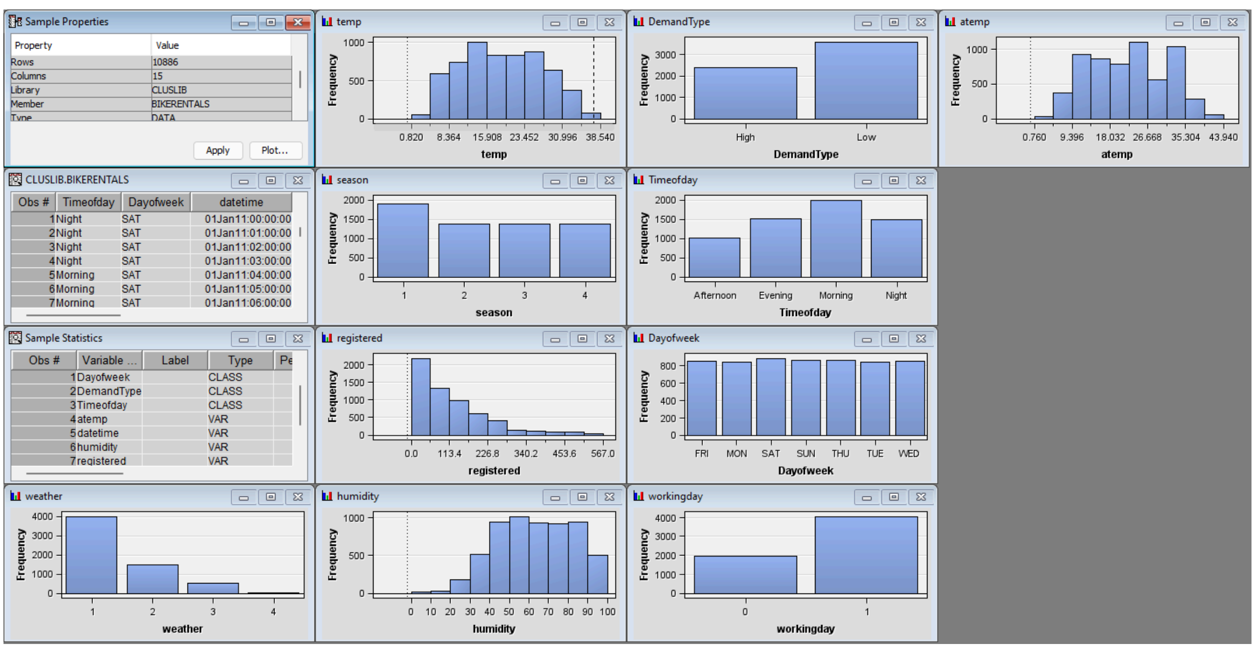
Given this information gathered from the logistic regression model, there are a couple of things that stand out. The biggest one is that Winter is the season with highest demand despite warmer temperatures being associated with High Demand. A possible explanation for this could be that people have more time available in the holiday season to visit different parts of the city. Additionally, the city could just simply be in a warmer climate where Winters are much more mild than other parts of the world. Lastly, behavioral habits may change following January 1st and new year's resolutions may include more bicycling as opposed to driving or just for additional exercise in general. This is something that we would look more into in future studies.

As for how we can use the overall data, we could offer dynamic rates that would allow discounted rates during times that would have low demand based on the model. For example, maybe during the night the rate is much cheaper, or look at pricing based on weather conditions. Additionally, we could create a mobile app that would allow users to pay/reserve bikes all on the app, and even find nearby bays where they can pick up bicycles in the future. The app could also prompt the user using a notification whenever there are good conditions for bike riding to help increase demand in peak times as well. This could result in more repeat and registered customers which is good for the long-term outlook for the business. Lastly, during the Spring, we could offer a discounted rate for the membership for a month or a couple of weeks to increase demand in the Spring.

Section 6: Clustering Analysis

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
Dayofweek	Input	Nominal	No		No	.	.
DemandType	Input	Nominal	No		No	.	.
Timeofday	Input	Nominal	No		No	.	.
atemp	Input	Interval	No		No	.	.
casual	Rejected	Interval	No		No	.	.
count	Rejected	Interval	No		No	.	.
datetime	Time ID	Interval	No		No	.	.
holiday	Rejected	Binary	No		No	.	.
humidity	Input	Interval	No		No	.	.
registered	Input	Interval	No		No	.	.
season	Input	Nominal	No		No	.	.
temp	Input	Interval	No		No	.	.
weather	Input	Nominal	No		No	.	.
windspeed	Rejected	Interval	No		No	.	.
workingday	Input	Binary	No		No	.	.

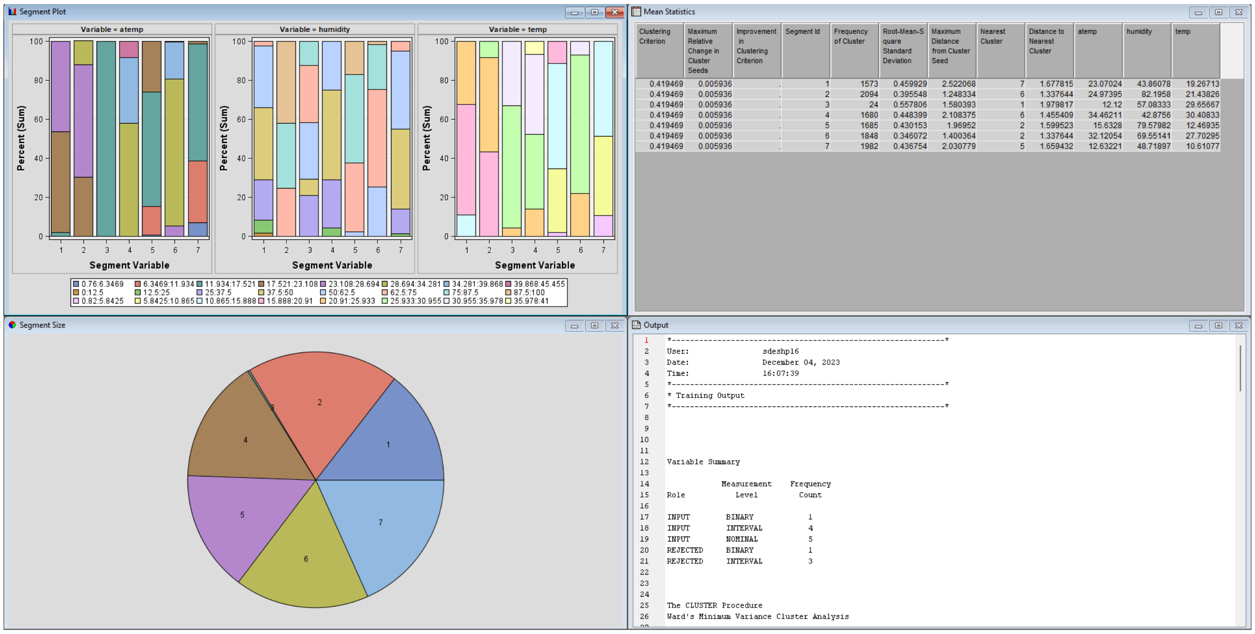
Based on the hypothesis provided above, we decided to reject the following variables: *casual, count,* and *windspeed*. In addition, we decided to reject *holiday* as well because we believed that the information gleaned from this variable was also present in the *season* variable.



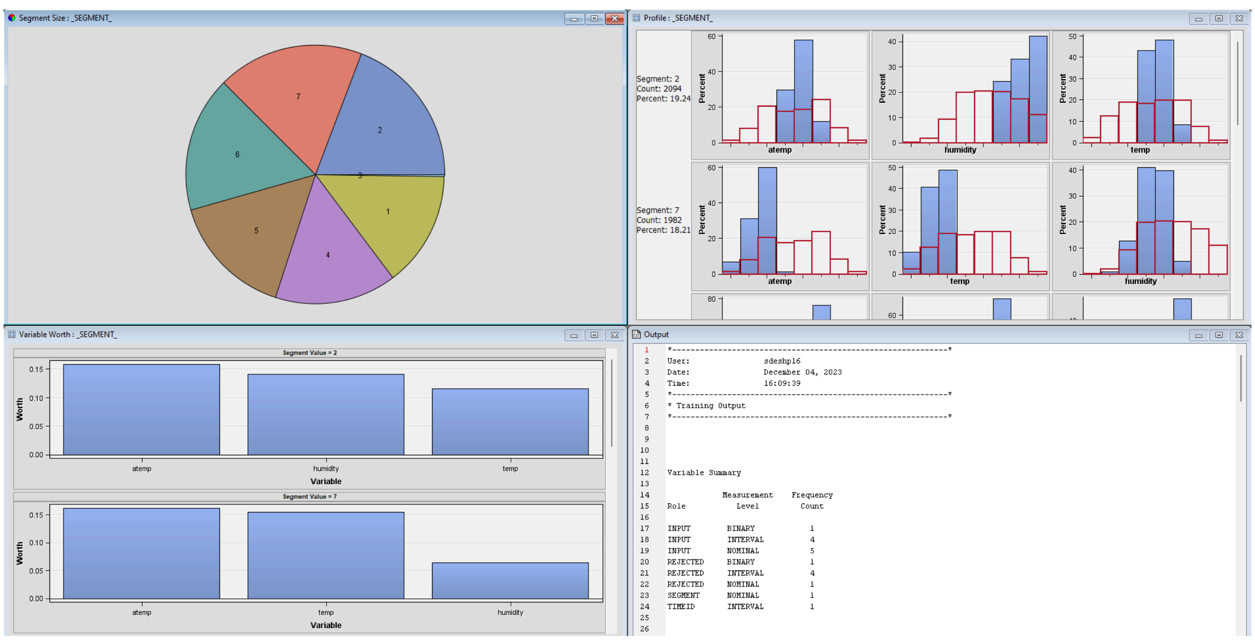
The image above provides a look at the distribution of each variable. The variables *temp, season, timeofday, dayofweek, and atemp* follow normal distributions. The variables *weather* and *registered* are skewed positively (to the right), while the variables *humidity, demandtype,* and *workingday* are skewed negatively (to the left). The chart below shows the mean for each interval variable and mode for each nominal variable:

Var_Name	**Mean/Mode**
weather	1.423
temp	18.9666
season	2.3663
registered	115.676
humidity	62.8222
DemandType	LOW
TimeOfDay	MORNING
DaysOfWeek	SAT
WorkingDay	0.6768
atemp	22.2846

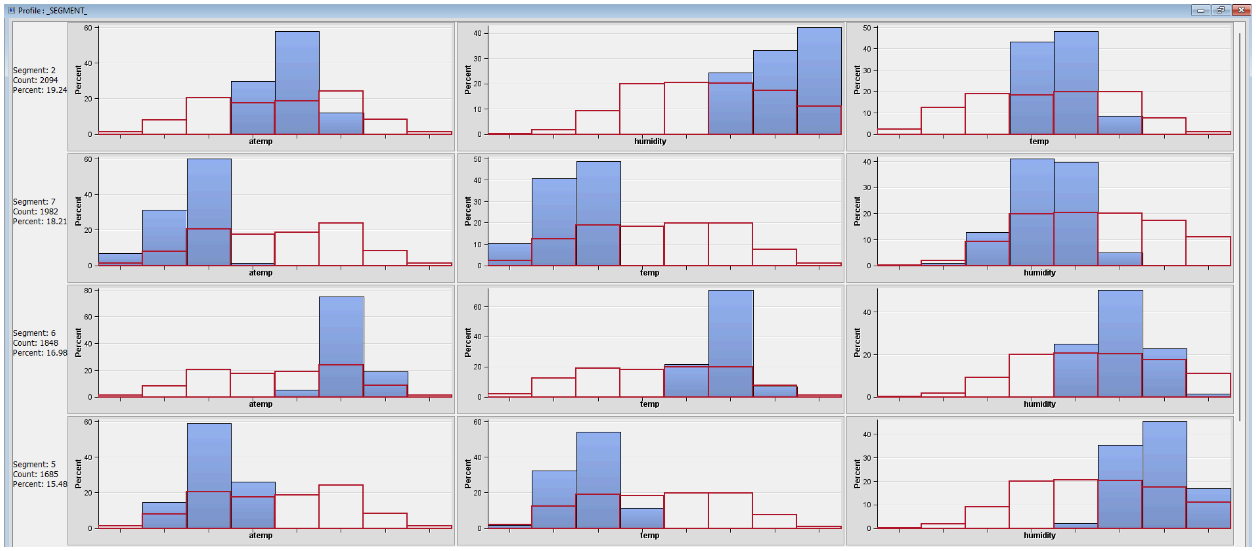
After analyzing the variables, we then ran the Clustering function in SAS Enterprise Miner. We decided to cluster on atemp, temp, and humidity. We chose these variables because we thought they would provide a good insight into how outdoor conditions can affect bike demand.

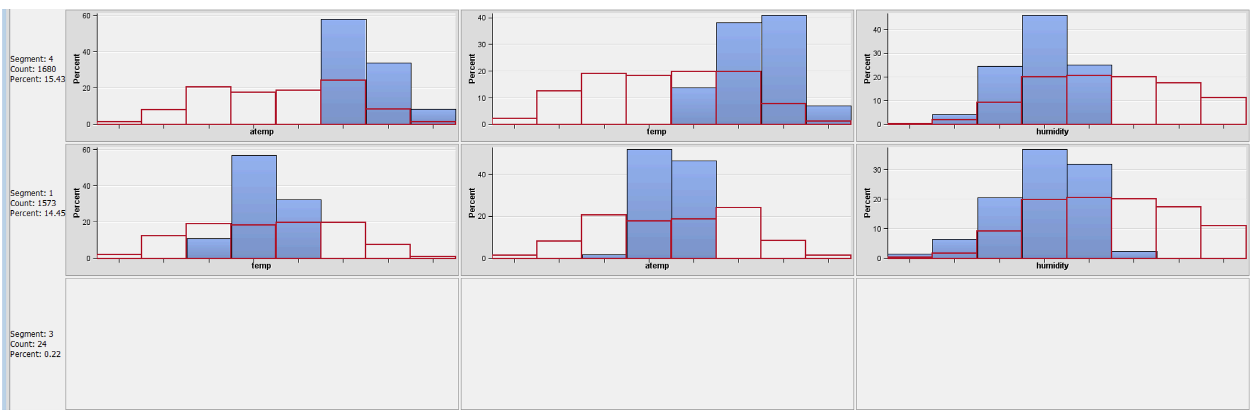


Because the data didn't have any extreme outliers, we didn't feel the need to filter the data for any cases. From this clustering, we found 7 clusters. To develop a profile on each segment, we ran the Segment Analysis command, which returned the below image:



Looking at the Segment Profile tab, we can see which variables are most important for each cluster and how the data from that cluster is distributed compared to the overall data. The breakdown for each cluster is shown and written below:





Segment 2 is the largest with 2,094 records in the cluster. The importance of variables is as follows (most important to least important):

- atemp: About normally distributed, but most of the data falls near the middle of overall data distribution
- humidity: Falls toward higher end of overall data distribution
- temp: Falls toward middle of overall data distribution

Next largest is Segment 7 with 1,982 records in the cluster. The importance of variables is as follows (most important to least important):

- atemp: Falls toward lower end of overall data distribution
- temp: Falls toward lower end of overall data distribution
- humidity: Falls toward middle of overall data distribution

Then we have Segment 6 with 1,848 records. The importance of variables is as follows (most important to least important):

- atemp: Falls toward higher end of overall data distribution
- temp: Falls toward higher end of overall data distribution
- humidity: Falls toward higher end of overall data distribution

Then we have Segment 5 with 1,685 records. The importance of variables is as follows (most important to least important):

- atemp: Falls toward lower end of overall data distribution
- humidity: Falls toward lower end of overall data distribution
- temp: Falls toward higher end of overall data distribution

Then we have Segment 4 with 1,680 records. The importance of variables is as follows (most important to least important):

- atemp: Falls toward higher end of overall data distribution
- temp:Falls toward higher end of overall data distribution
- humidity: Falls toward lower end of overall data distribution

Then we have Segment 1 with 1,573 records. The importance of variables is as follows (most important to least important):

- temp: Falls toward middle of overall data distribution

- atemp: Falls toward middle of overall data distribution
- humidity: Falls toward lower end or middle of overall data distribution

The smallest cluster is segment 3 with 24 records in the cluster. The output did not give a ranked list of variables in order of importance, and likely represents null values.

Thus, we can see that *atemp* is almost consistently the most important variable for the clusters, while humidity and temp fluctuate in importance for the cluster. From this, we can infer that *atemp* is important for the general population of data, while *humidity* and *temp* are more important for specific segments of the data population.

Part 3: Business Recommendations

Based on our findings from our constructed predictive models & clustering analysis we recommend the follow suggestion to the company in order to better manage and grow their rental business:

Rental Availability, Discounted/Surge Pricing, Loyalty Program, Mobile App

After interpreting both decision tree models we observed a significant relationship between weather, time of day, Day of the week and total bike rental demand. It can be concluded from our best outcome through the decision rule that demand rates in the afternoon or evening of nice weather weekends will tend to be higher. Therefore we suggested the following for optimal conditions:

- ***Offering additional bike rentals to allow for accessibility*** to match our high demand during peak times and days of the week for our customers.
- ***Offering higher rental rates on optimal conditions*** where the demand is higher. This is solely because of the fact that the demand is clearly present and if the conditions like that exist then the resulting demand will be present.
- ***Offering discounted services during low demand.*** For such instances like mornings or nights where there are colder temperatures, in the spring on days that are Monday-Friday individuals are less inclined to rent bicycles. Here in our worst case rental outcome we are able to make the best out of this situation by offering certain incentives.
- These suggestions utilize the ***Dynamic Pricing Strategy.*** Dynamic pricing is a strategy in which businesses are able to set flexibility for their product and/or service based on current market demand as a way to reflect the demand changes & increase profitability. Introducing dynamic pricing to adjust their prices when needed will be greatly effective

in growing their rental business. For example, when conditions for the highest potential rental demand are present, the company can charge at a higher premium and vice versa.

Create a Loyalty System incorporating all of the following above. This increases subscription and usage rates as seen with other types of services that use reward systems to increase the likelihood of customers returning again for a possibility of a free ride. Rewarding frequent customers and encouraging repeat business can entice loyal customers and increase sales.

Mobile App & Building an Online Presence through developing a user-friendly mobile app that allows customers to easily locate and rent bikes incorporating features like GPS tracking, payment integration, and bike availability status as well as leveraging social media & online marketing to increase brand awareness can increase the businesses' reach & customer satisfaction, which could lead to increased brand loyalty and bike rental demand.

All of the following suggestions can help to sustain sales and demand over the course of the days of the week that statistically have the lowest demand. While boosting sales during peak demand times and increasing customer usage.

Individual Contributions

Yvanna Tchomnou. During this project, I was responsible for completing Part 1 of the project report. I completed the research, collection of data, and the writing of Sections 1 and 2 as well as creating our initial hypothesis. Aside from that, I also created and completed the formatting and structuring of both the project report and the group presentation slides. Furthermore, I also worked with group members to go through their models and analysis to provide help when needed and to do the final checks to verify that the models were made correctly and yielded accurate results. Lastly, I wrote Part 3 of the report on "Business Recommendation" along with Majesti.

Sarang Deshpande. During this project, I was responsible for creating the Clustering Model. This included analyzing the data to determine if there was a need to filter the data in any way, deciding which variables to cluster the data with, and running the Segment Analysis and interpreting the results into actual tangible results.

Majesti Davis. During this Project, I was responsible for creating the second decision tree diagram in which I did the misclassification diagram. This includes creating the decision tree model via SAS Enterprise Miner, doing the decision matrix calculations, deciding the independent variables to include and reject within my section, and finally interpreting the best and final outcome. Aside from that I also assisted in the creation of section three of our paper which includes the data preparation. As I created the three additional variables for our project Time of day, Demand Type, and Day of week. Then I also converted our excel file to our SAS file, using SAS Enterprise Guide, for our entire group. In which I go over this in detail during section 3 of our paper. Lastly, I wrote Part 3 of the report on "Business Recommendation" along with Yvanna.

Eric Sheehan. During this project I was responsible for everything related to the Logistic Regression model. This includes the creation of the model, the partitioning of the data, and the interpretation of the model and recommendations that we can make based on the model's results. Additionally, I presented this section of the PowerPoint in class. Lastly, I contributed to the creation of the variables before creating the model along with Majesti.

Ethan Lozuke. In this project I was tasked with making the first decision tree with the data that was provided and adjusted by some of the group members. Within the decision tree I was tasked with deriving and analyzing the resulting data tree and reporting on those results. I was able to assess the nature of the optimal decision tree and find what the best and worst cases for the business were. After getting an idea of where the business was struggling or succeeding I was then able to make suggestions based off of these results that would help to increase demand for the bicycles. Once conferring with Majesti and discussing her results of the second decision tree I was then able to collaboratively make a list of final recommendations. In addition I abbreviated my results here and formatted them to be more digestible on my part of the presentation.

References

Bike Rental Market Size, Share, Competitive Landscape and Trend Analysis. (2022, August). Allied Market Research. https://www.alliedmarketresearch.com/bike-rental-market-A09610

Fishman, E., Washington, S., & Haworth, N. (2013). Bike Share: A Synthesis of the Literature. *Transport Review*, *33*(2), 148-165. Taylor & Francis Online. https://doi.org/10.1080/01441647.2013.775612

He, Y., Song, Z., Liu, Z., & Sze, N. (2022, December). Factors Influencing Electric Bike Share Ridership: Analysis of Park City, Utah. *Sage Journals*, *2673*(5). https://journals.sagepub.com/doi/full/10.1177/0361198119838981#sec-6

Normack, D., Cochoy, F., Hagberg, J., & Ducourant, H. (2018, August). Mundane intermodality: a comparative analysis of bike-renting practices. *Mobilities*, *13*(6), 791-807. Taylor & Francis Online. https://doi.org/10.1080/17450101.2018.1504651

Shaheen, S., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia. *Transportation Research Record*, *176*(1), 159-167. Sage Journals. https://doi.org/10.3141/2143-20

Sood, S. (2011, September 9). *Bike sharing around the world*. BBC. Retrieved November 30, 2023, from https://www.bbc.com/travel/article/20110909-travelwise-bike-sharing-around-the-world

Yglesias, M. (2014, October 29). *Why don't the poor use bike share systems?* Vox. Retrieved November 30, 2023, from https://www.vox.com/2014/10/29/7087331/low-income-bicycle-share