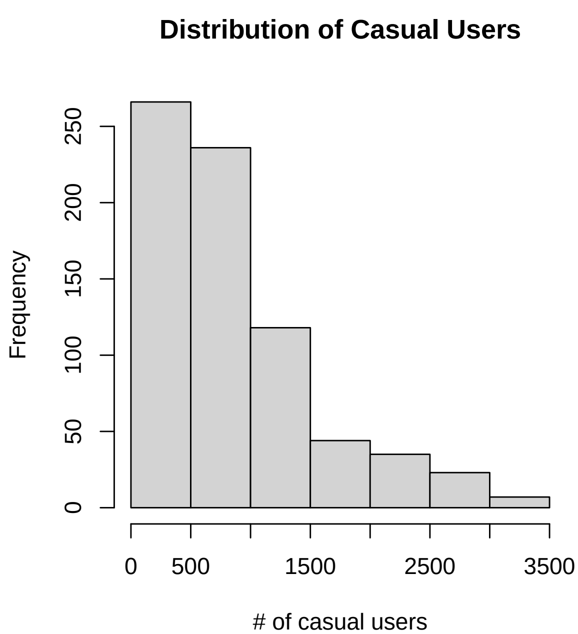
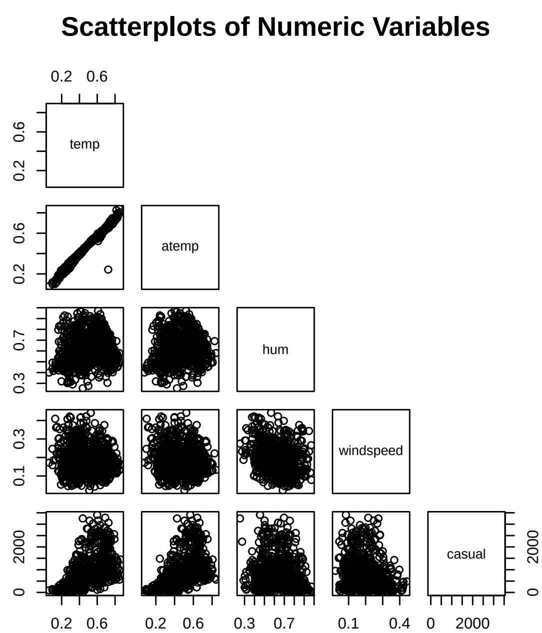
MAST 6251: HW1

Group 4: Abiba Bamba (48953884), Kareem Baber (48892108),
Gangadhar Sallagundla Chowdary (49031287), Sumita Nair (48992260),
Nicole Parmelee (47653402)

This report discusses our findings from a regression analysis of the company's bike share data from 2011 and 2012. The data primarily contains environmental and seasonal variables that affect the demand for bike sharing services. Our analysis solely focuses on "casual" usage and doesn't consider "registered" usage. We inferred that registered users primarily use bike sharing services out of necessity, to commute to work. However, casual users use the service on a more occasional basis as a form of entertainment and transportation to get around in downtown areas. Because of this, we believed casual users provided more of an opportunity for the company to take actions (i.e. marketing efforts, promotions, discounts, etc) that will truly influence demand and increase revenue.

Before running the regression analysis, we needed to observe and correct any extraneous values in the data. We checked for outliers in all of the numeric predictor variables (temperature, feels-like temperature, humidity, windspeed, and number of casual users) and removed a few of the most extreme points in the humidity and windspeed variables. This ensures that the insights gained from the regression will be as accurate as possible.





As shown in the graph above on the left, the number of casual users is strongly right-skewed. This means that for a majority of the observed days, there are relatively low numbers of casual users, but that are still some days with much larger amounts. This variation is, in part, due to the fact that casual users tend to ride bikes more for pleasure than out of necessity. Our analysis attempts to understand and explain which situational factors lead to days with higher numbers of casual users.

We found relationships between multiple variables in the data by observing descriptive plots and statistics (see graph above on the right). Intuitively, the data shows that temperature and feels-like temperature (labeled as "atemp") are extremely correlated. Because of this, we had to choose only one of these variables to include in our model so that it could represent an accurate relationship with the number of casual users. We decided to include feels-like temperature because it considers the other weather-related factors that could influence a user's decision to ride a bike. We also found evidence that the number of casual users is related to feels-like temperature, humidity, windspeed, month, and day of the week. From the data, we observed that the peak usage days for casual users tend to be on weekends and in the Spring, Summer, and Fall months.

term	**estimate**	**std.error**	**p.value**
(Intercept)	730	112	0.0000
yr	288	27	0.0000
mnth2	-14	66	0.8378
mnth3	350	70	0.0000
mnth4	482	77	0.0000
mnth5	530	88	0.0000
mnth6	368	101	0.0003
mnth7	233	110	0.0345
mnth8	330	101	0.0012
mnth9	451	91	0.0000
mnth10	445	77	0.0000
mnth11	215	68	0.0017
mnth12	38	66	0.5673
holiday	-247	214	0.2478
atemp	1618	198	0.0000
windspeed	-986	187	0.0000
hum	-571	142	0.0001
weathersit2	-94	36	0.0091
weathersit3	-281	96	0.0034
weekday1	-741	50	0.0000
weekday2	-790	49	0.0000
weekday3	-798	49	0.0000
weekday4	-781	49	0.0000
weekday5	-608	49	0.0000
weekday6	153	49	0.0018
holiday:atemp	1741	446	0.0001

The table above shows our final regression model used to understand the effects of the predictor variables on the number of casual users. The model does a good job of explaining the number of casual users (adjusted $R^2 = .74$). As shown, year, month, holiday, feels-like temperature, windspeed, humidity, weather situation, and day of the week all significantly help explain casual usage. For the categorical variables (year, month, holiday, weather situation, and weekday), a positive coefficient estimate suggests that when an observation falls into the given category, the casual usage will increase by the amount of that category's coefficient. For example, on Saturdays (weekday6), holding all other variables constant, the number of casual users is expected to increase by 153, on average. The opposite is true for categorical variables with negative coefficients. For the numeric variables (feels-like temperature, windspeed, humidity), a one unit increase* will result in an increase/decrease of casual users by the corresponding coefficient, holding all other variables constant (*note: these variables were converted to a 0 to 1 scale, so the "one unit increase" is on this adjusted scale as well).

The model shows that holidays significantly amplify the effect of the feels-like temperature on the number of casual users. For example, on holidays, the temperature has a larger effect on the number of casual users than it does on non-holidays. This means that if the temperature is 5°C higher on a holiday, this will increase casual demand more than this same temperature change would on a non-holiday. Similarly, a 5°C drop in temperature on a holiday would decrease casual usage by a larger amount than it would on a non-holiday. Because of this interaction, we think that promotional offers on holidays in the Spring, Summer, and Fall months would help increase casual usage, which will drive the company's revenue. While these situations already have a positive relationship with casual usage, this can be further improved with specific promotion efforts. We suggest a "rent two bikes for the day, get one free" promotion. This will hopefully appeal to users who are interested in spending the holidays with larger groups of their friends and family. Additionally, the company should release marketing material that advertises this promotion, and also emphasizes the non-financial benefits of ride-sharing on holidays. For example, for those who want to attend downtown holiday events, renting bikes would allow them to avoid the hassle of driving and parking in crowded areas. The cost of this promotion itself would be about $30 per free bike, which is the price users typically pay for a full-day bike rental. To cover the fixed cost of the marketing campaign, the company should find a sponsor. This sponsoring company's logo will be on the advertisement, and the money they pay us for this ad will offset our marketing costs. We anticipate that both the promotion and marketing campaign will attract more casual users on these days, which will increase revenue enough to where it covers the variable promotion costs.

The regression implies that the humidity, the feels-like temperature, and the windspeed all have large effects on the number of casual users. As the humidity or windspeed increases, the number of casual users decreases on average. Similarly, as the feels-like temperature decreases, the number of casual users decreases. To encourage increased usage on these days with unfavorable weather, we suggest that the company implement a punch card system. This system would reward users who ride on windy, cold, and/or humid days. Once users accumulate 10 "punches", they can trade in their card for a free full-day bike rental. For this recommendation to be feasible, we're assuming that the company currently has bike docking stations where users pay for rides and that these stations are connected to a central database, which has up-to-date weather information. When the stations determine that the windspeed and/or humidity is above a certain threshold or that temperature is below a certain level, it will punch a user's card accordingly. This aims to increase usage on otherwise low-usage days. Even though the company will incur a cost associated with each free rental ($30), this system will attract customers on days that would otherwise have very minimal usage, and is therefore expected to increase revenue overall.

Similarly, we suggest stimulating casual usage in the typically slow Winter months (November, December, January, February) by offering a 20% discount. These months are currently associated with lower numbers of casual users, meaning the company has a large amount of excess bikes that aren't in use or generating any revenue. Because of this, the increased traffic that the seasonal promotion attracts is very likely to offset the 20% discount cost. The rides sold at 80% of their normal price during the winter are rides that wouldn't be sold if the discount wasn't given, meaning the 20% loss isn't a significant detriment to the company's revenue. To advertise this discount, the company should print large banners to display at the bike docking stations. This way, the only associated cost would be the cost to design and print the ad (real estate would be free since the company owns the property). From initial research, banners would cost around $40 each. While buying a banner for each docking station would be a somewhat significant investment, this only has to be done once; for every subsequent Winter, the banners could be reused. Again, we anticipate that the additional users obtained from this effort will offset the overall cost.

All of these actionable suggestions made based on our regression model will encourage not only an increase in the number of casual users, but also an increase in company revenues. These actions will (1) stimulate even more usage on popular days, such as holidays and those with favorable weather, and (2) generate new usage during off-peak days, including during the winter and in extreme weather, which would otherwise have very low revenue. In addition, these promotions and discounts will ideally attract brand new users, which will help the company further expand its customer base.