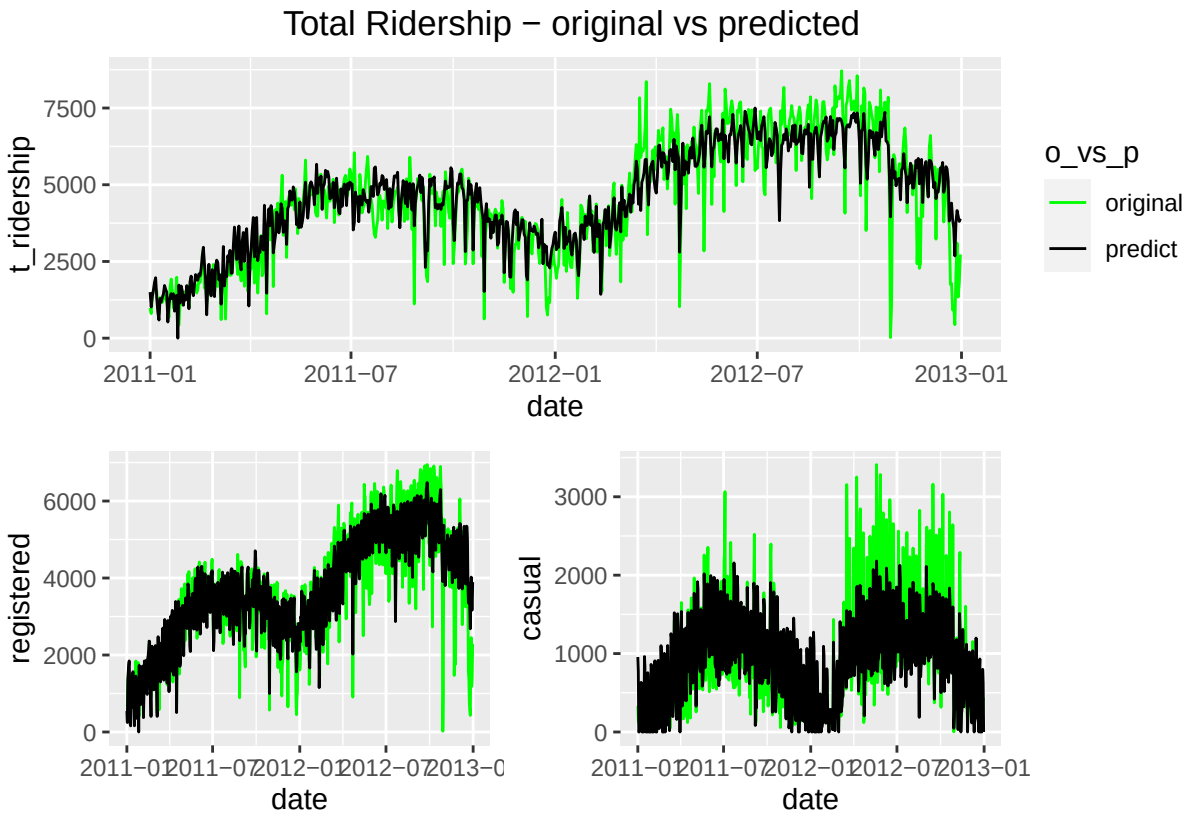
HW1 - Group 6

Caroline Searcy - 47453436, Sahar Khan - 48942652, Rachana Pokhrel - 49051554, Napat Jivakanun - 49024058, Varun Gokhale - 49045717

Objective: To predict bike ridership in year 3 onward

Executive summary

We have created two models, one for predicting casual ridership and one for predicting registered ridership with accuracies as shown in the graphs below. The three graphs depict a comparison of predicted values (from the generated multiple linear regression models) and actual values.

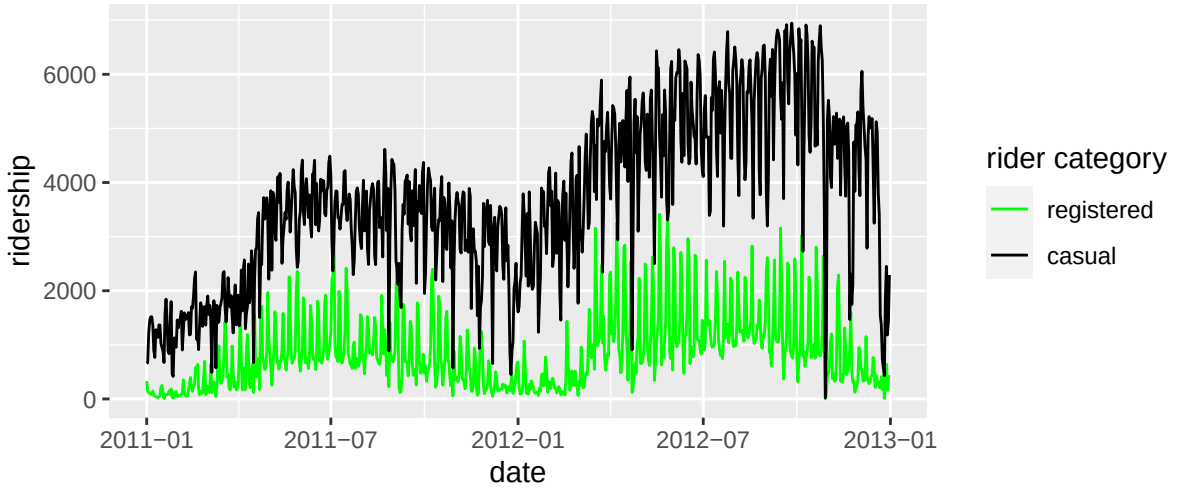



Insights taken to construct the prediction models:

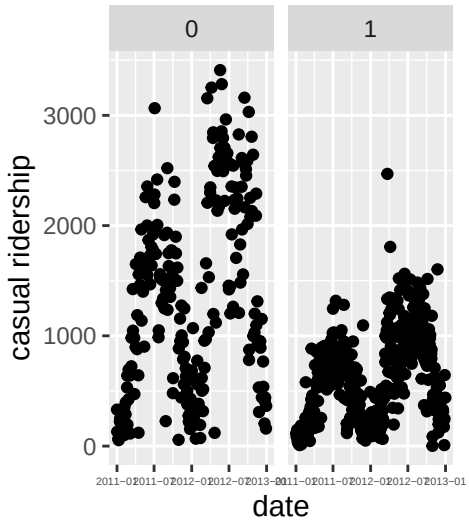
1. It was discovered that registered and casual riders behave differently. fig 3 depicts that registered riders has a higher gradient of increase over the past two years in comparison to that of casual riders. The regression gradient for registered and casual riders versus instant (instant 1 = day 1, instant 90 = day 90) are 4.87 and 0.895 respectively. fig 4 and 5 shows that casual riders has a higher propensity to ride on weekends or holidays over weekdays, and vice-versa. This indicates that different models should be used to predict registered and casual riders.

2. The increasing trend for registered riders are not explained by any of the weather related data. The coefficients of regression for temp, atemp, windspeed, and hum (humidity) versus instant are 1.305e-04, 1.178e-0.4, -0.4133e-05, and 1.104e-0.5 respectively, which are extremely small gradients compared to the 4.87 gradient of registered ridership with respect to instant. The effect of a 4.87 gradient with respect to instant was removed from the registered ridership data for regression modelling (but is reincorporated when the model was used to make prediction). Casual ridership was not adjusted because its increasing trend is negligible. fig 6 shows the adjusted values of registered ridership for registered ridership regression modelling.

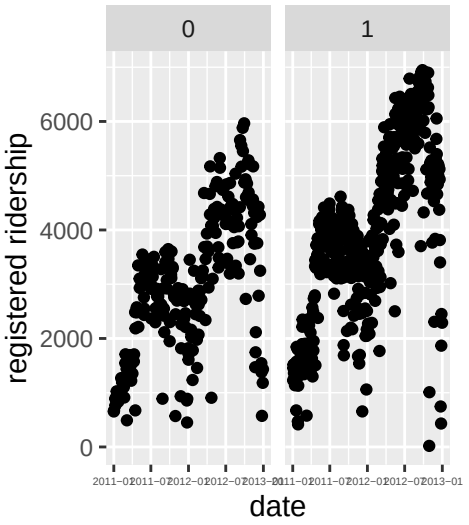
(fig 3) Ridership Trend



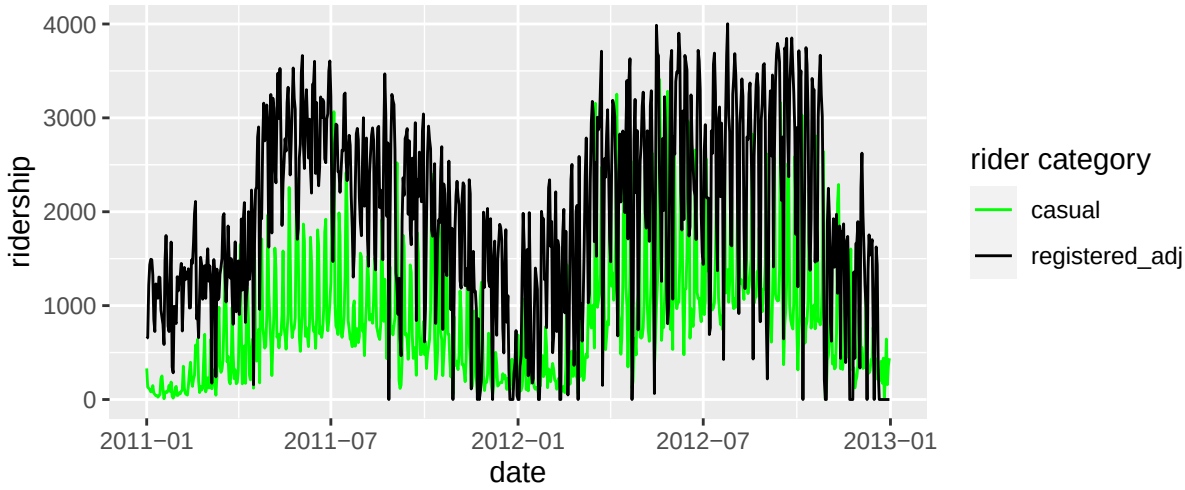

(fig 4) 0 = not working day 1 = working day




(fig 5) 0 = not_working_day 1 = working_day




(fig 6) Adjusted Ridership Trend



Key characteristics and assumptions of the formulated regression formula to predict casual and registered ridership:

1. The instant variable, which is the number of days past January 11, 2011 is required in our registered ridership prediction model to reincorporate the 4.87 gradient effect that was taken out for the purpose of model creation. If there is a projection that factors such as economic growth or increased brand awareness will cause high/low growth to registered ridership, the reincorporated 4.87 gradient cant be changed to another number.

2. Both model assumes that registered riders are growing and casual riders are constant.

Regression formula (in R) formulated to predict registered ridership and casual ridership by creating a column next to a set of data:

registered ridership:

data<- mutate(data, registered_predict = 4.8738*instant + 749.91+ ifelse(season== 1,0,0)+ ifelse(season == 2, 394.03,0)+ ifelse(season == 3, 272.14,0) + ifelse(season == 4, 744.60,0)+ifelse(mnth == 1, 0,0) + ifelse(mnth == 2, 48.20,0)+ifelse(mnth == 3, 183.76,0)+ifelse(mnth == 4, 194.67,0)+ ifelse(mnth == 5, 404.02,0)+ifelse(mnth == 6, 352.43,0)+ ifelse(mnth == 7, (-56.88),0)+ ifelse(mnth == 8, 113.32,0)+ ifelse(mnth == 9, 317.10,0)+ ifelse(mnth == 10, (-152.53),0)+ ifelse(mnth == 11, (-798.73),0)+ ifelse(mnth == 12, (-744.01),0)+ ifelse(weekday == 1, 25.86,0)+ ifelse(weekday == 2, 199.41,0)+ ifelse(weekday == 3, 246.61,0)+ ifelse(weekday == 4, 254.79,0)+ ifelse(weekday == 5, 108.62,0)+ ifelse(weekday == 6, 254.76,0)+ifelse(workingday == 1, 871.81,0)+ ifelse(weathersit == 1, 0 ,0)+ifelse(weathersit == 2, (-345.01),0)+ ifelse(weathersit == 3, (-1297.74) ,0)+ (atemp*2295.83)+(hum-905.91)+(windspeed*-1403.59))

casual ridership:

data<- mutate(data, casual_predict = 962.913+ ifelse(season== 1,0,0)+ ifelse(season == 2, 224.414,0)+ ifelse(season == 3, 49.905 ,0) + ifelse(season == 4, 35.737 ,0)+ifelse(mnth == 1, 0,0) +ifelse(mnth == 2, (-35.294 ),0)+ifelse(mnth == 3, 212.589,0)+ifelse(mnth == 4, 177.205 ,0)+ifelse(mnth == 5, 166.061,0)+ifelse(mnth == 6, (-28.007),0)+ifelse(mnth == 7, (-65.618),0)+ifelse(mnth == 8, (30.407),0)+ ifelse(mnth == 9, 246.314 ,0)+ifelse(mnth == 10, 326.270 ,0)+ifelse(mnth == 11, 141.739,0)+ifelse(mnth == 12, 1.116 ,0)+ ifelse(weekday == 1, (-741.870 ),0)+ ifelse(weekday == 2, (-802.551 ),0)+ ifelse(weekday == 3, (-805.414 ),0)+ ifelse(weekday == 4, (-803.492 ),0)+ ifelse(weekday == 5, (-628.944),0)+ ifelse(weekday == 6,144.432 ,0)+ ifelse(holiday == 1, 505.734 ,0)+ ifelse(weathersit == 1, 0 ,0)+ ifelse(weathersit == 2, (-72.077 ),0)+ ifelse(weathersit == 3, (-314.610 ) ,0)+(atemp*1990.382 )+(hum-757.427 )+(windspeed*-1197.901))

Recommendations derived from both prediction models:

1. Assuming January has 0 customers, November and December results in a loss of 799 and 744 registered customers respectively. To mitigate for loss during these winter months, an incentive should be sent out to registered customers to get them riding.

2. For casual riders, weekdays are often a loss. Assuming Sundays result in 0 casual riders, weekdays lose an average of around 756/day casual riders. To boost the customer base of casual riders on weekdays, the company should (1) partner with large employers in major cities to have bikes outside the offices and (2) position bikes next to major public transit hub.

3. With an increase of casual riders during weekdays, the company should push to convert casual riders into regular, registered users. Assuming 10% conversion rate, over the course of a year, the company could gain about 19,000 registered customers.