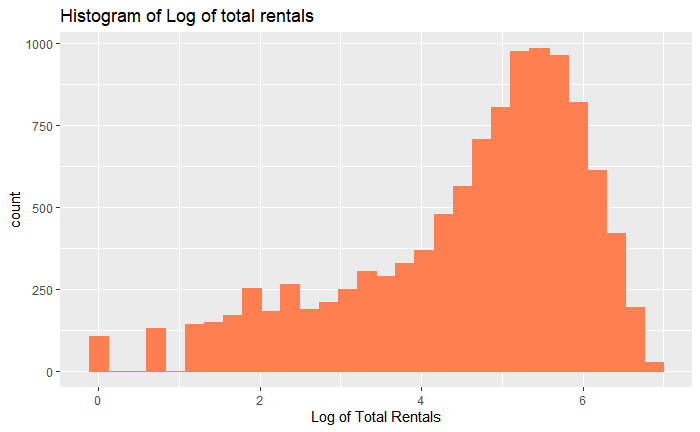
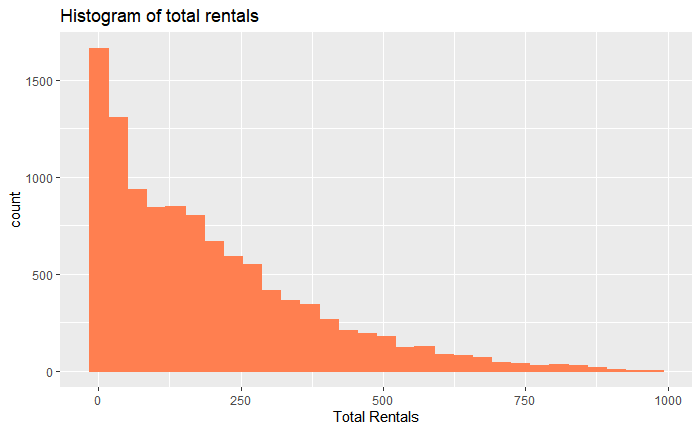
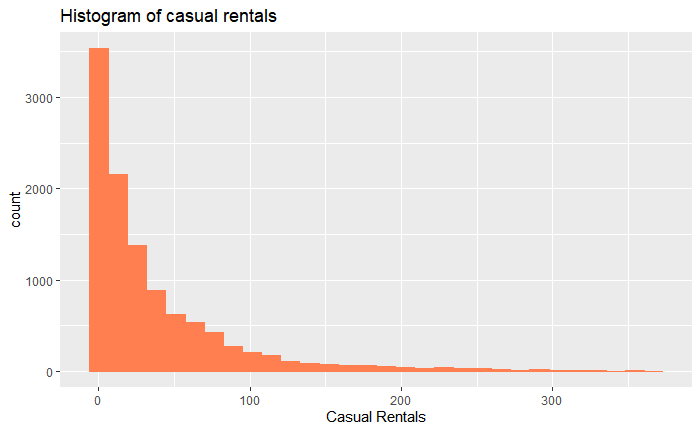
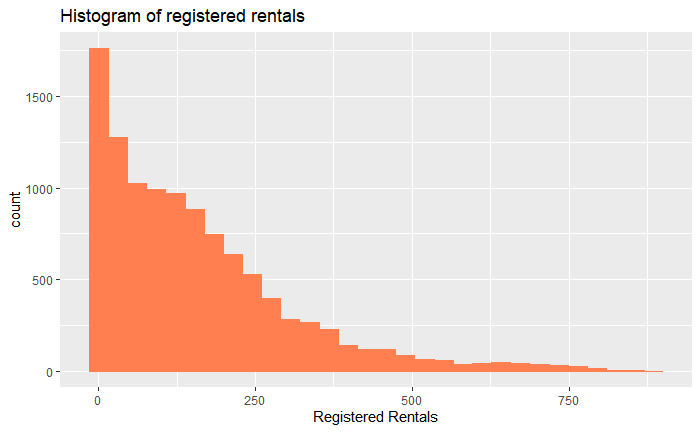
Exam 1 SDM

Submitted by: Hammad Muniem

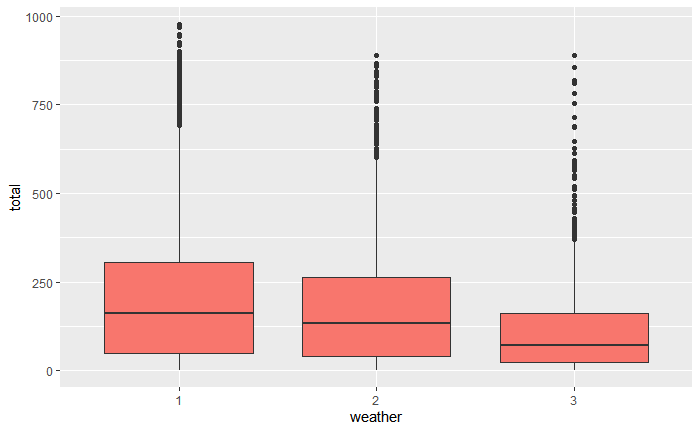
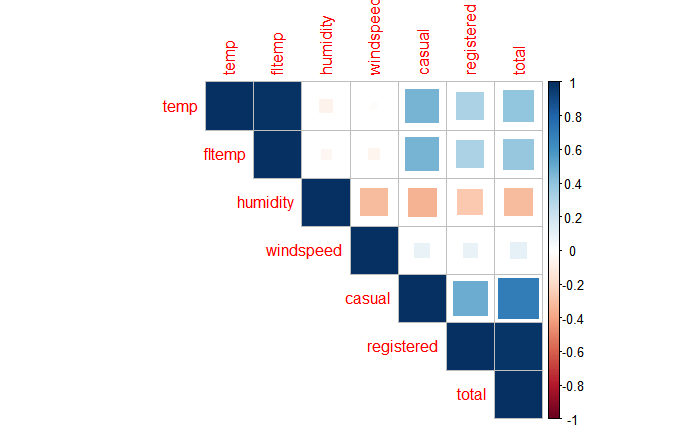
**1)** We plotted the following distributions. This was nessecary since it will help us decide the kind of model that we would have to predict the values of these dependent variables. Since the data highly skewed with lots of zero values (in casual specially) we implemented a log function on it to make it closer to a normally distributed variable. By the looks of it however, it seems relatively normal for total rentals but not for casual rentals. Therefore, running an OLS regression here would not be advisable since we will be violating OLS assumptions here. GLM models would be more appropriate here since they don’t have the OLS assumptions.

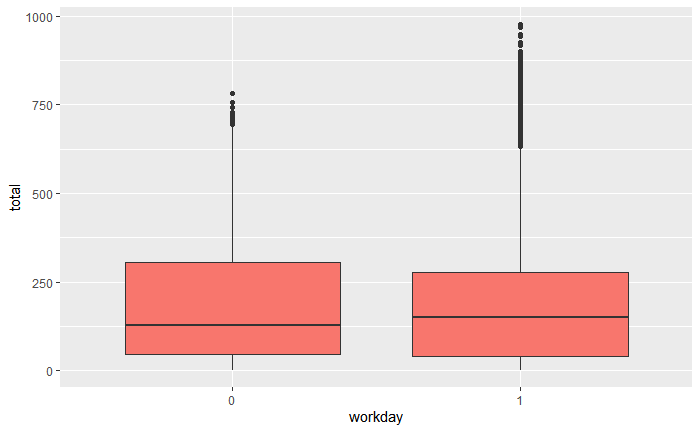
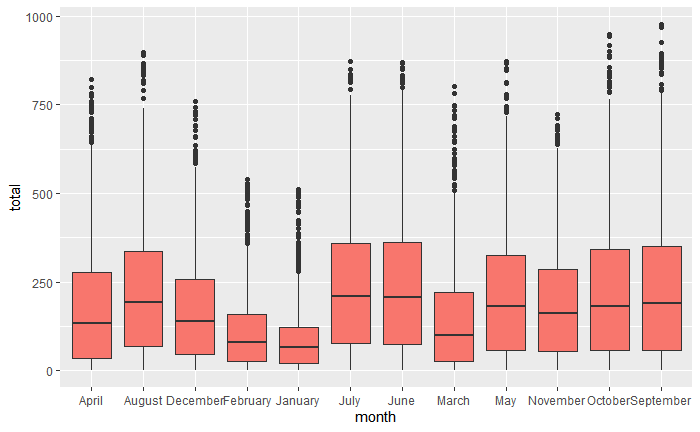
Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

**2**)



Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

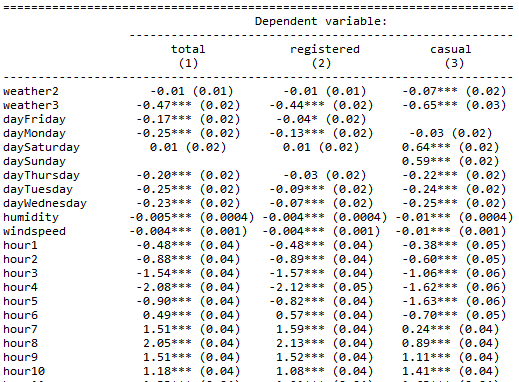
Description automatically generated

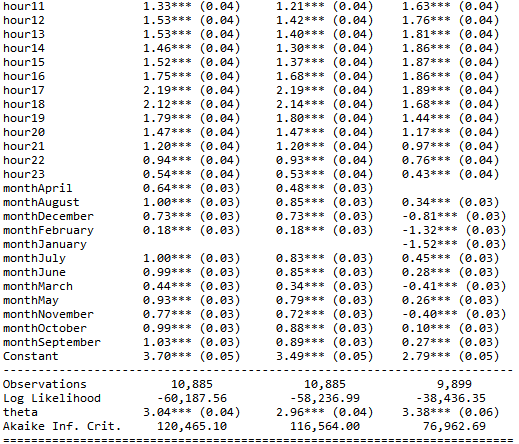
These visualizations give us an idea as to what kind of pattern to expect in the final model. For example we can infer here that on average, highest bike rentals are on days when weather equals 1. The correlations help us identify which variables might give us multicollinearity in the model.

**3)**

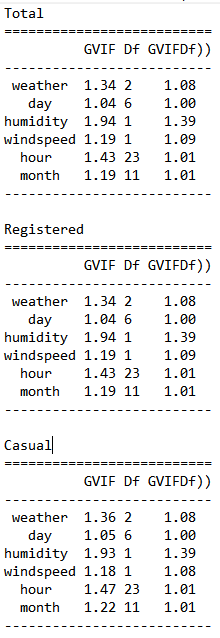
|  |  |  |
| --- | --- | --- |
| Predictor Variable | Expected sign of effect | Rationale |
| Season/Month | +- | We expect higher bike rentals when the weather is pleasant. Not too hot and not too cold. Therefore, we expect lower rentals in peak summers and winters. |
| Holiday | - | Since the data is from Europe where there is high population density, we expect people to use bikes to commute to work as well. Therefore, overall traffic should be higher on work days. |
| Workday/day | + | Since the data is from Europe where there is high population density, we expect people to use bikes to commute to work as well. Therefore, overall traffic should be higher on work days. |
| Temp/fltemp | - | Initially, rentals might go up with temperature but as temperature goes up above a certain value, we expect it to have a negative effect. |
| humidity | \_ | Higher humidity might mean rain or excessive sweat meaning that bike rentals will go down. |
| windspeed | - | As it gets more windy, handling becomes more difficult. |
| hour | + | We expect bike rentals to go up for peak working times |

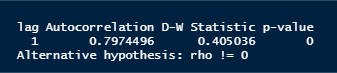
**4)**

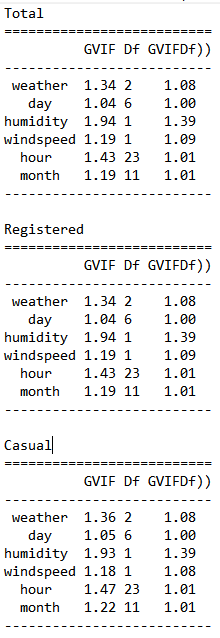


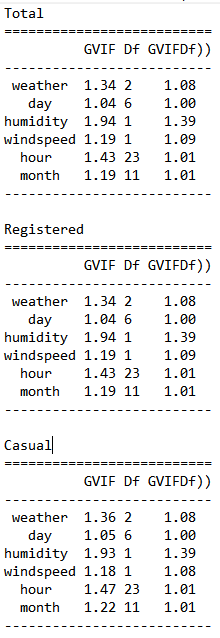
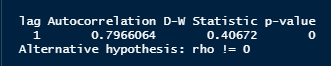


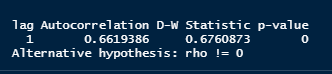
Since this distribution was skewed and was a count variable, I used poisson model (instead of OLS because the assumptions would be violated) first for each of them. Upon checking overdispersion in the data set, I ran negative binomial instead. For casual, I also removed the zero values to give it a better fit.

I removed temperature variable because my models were failing the VIF test before. But now each of them has cleared it. The selected models have also cleared the independence tests.









5)

* For weather, the highest bike rentals were in weather 1 in all categories. For weather 3, bike rentals decreased by 47%, 44% and 65% for total, registered and casual respectively. (keeping all other factors constant).  
  For weather 2, bike rentals decreased by 1%, 1% and 7% for total, registered and casual respectively. (keeping all other factors constant).
* Bike rentals were higher on the weekends as compared to the weekdays on average. Most bike rentals were on Saturday across the 3 classes.
* For total, September was the highest and January was the lowest. The difference was 103% more bike rentals in September compared to January.

For registered, September was the highest and January was the lowest. The difference was 89% more bike rentals in September compared to January.  
For casual, July was the highest and January was the lowest. The difference was 197% more bike rentals in July compared to January.

* For total, Saturday was the highest and Monday and Tuesday were the lowest. The difference was 26% more bike rentals on Saturday compared to Monday and Tuesday.

For registered, Saturday was the highest and Monday was the lowest. The difference was 14% more bike rentals on Saturday compared to Monday.  
For casual, Saturday was the highest and Wednesday was the lowest. The difference was 89% more bike rentals on Saturday compared to Wednesday.

* For total, hour 17 was the highest and hour 3 was the lowest. The difference was 373% more bike rentals on hour 17 compared to hour 3.

For registered, hour 17 was the highest and hour 4was the lowest. The difference was 431% more bike rentals on hour 17 compared to hour 4.  
For casual, hour 17 was the highest and hour 5 was the lowest. The difference was 352% more bike rentals on hour 17 compared to hour 5.