First, let us discuss the trends with casual users versus registered users of the service. In our initial EDA phase, we quickly noticed that the majority of our users were registered, and a smaller proportion were casual users - in total, we saw some 2.67 million registered users in the database compared to only 620,017 casual users, meaning we saw over 400% more registered uses of the system than we saw unregistered, or casual, uses. Overall, it appears our registered users see much more fluctuation on a day-by-day basis, depending on factors such as weather or date, while casual users appear more influenced by the general time of year - we see a huge spike in casual users during the spring season. This trend holds pretty constant when the data is divided by season - interestingly, we see more use during fall than during the summer in both categories; I would have expected to see the most casual users in the summer - peak season for vacation and travel, meaning a pretty sizeable influx of likely unregistered users. Our peak usage by registered users was 6946, which occurred in September 2012 (a Tuesday) - we saw 787 casual users that day, making up only 10.2% of our users were casual users on this given day. For contrast, our peak date for casual users came in May 2012 (a Saturday) where we saw 3410 casual users versus 4884 registered users - even at its greatest point, the number of casual users failed to overtake that of registered users. Our minimum values did align, however - there were only 22 total users of the service on a certain day in October 2012, which was made up of 20 registered users and only 2 casual users. Unsurprisingly, this was a day with snow (or with light rain and thunderstorms) and an absurdly high humidity of .88 - the .36 wind speed was also close to double the average wind speed of .19, making this a pretty awful day to commute by bike.

weathersit <fct>	avg <dbl>	count <int>
1	4877.	463
2	4036.	247
3	1803.	21

weathersit <fct>	registered <dbl>	casual <dbl>
1	3913.	964.
2	3349.	687.
3	1618.	185.

In a convenient segue into our next point, let us discuss how the weather situation affected users of the service. On average, we saw around 4880 users per day when there was no precipitation - no rain, no snow, no mist. We saw that number drop to around 4030 users on days with some sort of mist, but no precipitation. On days with light precipitation, whether that be snow or rain, possibly with thunderstorm included, that number dropped down to 1800 - less than half the number we saw when it was simply misting, and only 37% of the amount of users we saw on days with no precipitation. When we break that down by registered versus casual users, we note that registered users seem less perturbed by the mist than our casual users - we retained 85% of our userbase going from a non-precipitous day to a day with mist, though we again see another half or so of those users vanish when there was some form of light rain or snow. Casual users appear much more bothered by mist - only 70% as many users were willing to trek out into the mist versus those who rode on non-precipitous days. That number drops even further on days with light precipitation, dropping from 687 users on misty days to a paltry 185 users on days with light precipitation. That 27% as many users on days with light precipitation as opposed to days with light mist, and a measly 19% as many users as on days with no

precipitation. That is a much harder number to stomach than the 37% retention rate we see for registered users on days with light rain/snow versus no precipitation. Likely, this can be explained in a similar way that many of the casual versus registered user trends can be - registered users are much more likely to be using the biking service as a part of their daily routine, as their main method of commuting to and from work, school, etc. For some of these users, the alternative is to either pay for a taxi or Uber, or to suck it up and walk - neither are particularly pleasant choices for someone making a living in D.C. Casual users, on the other hand, are likely first-time users who are using the service as a novelty, likely many times tourists who are using the service as a way to further enjoy their vacation. These users can simply stay in, or change their plans, or pay for an Uber - it is much easier to change your commute on a whim when your plans are much less vital to your well-being.

Let's finish the transition into the effect of weather on usage of the bike sharing service. As we noted, there was a significant drop in users on misty days, and a much more precipitous (no pun intended) drop on days with light rain or snow. However, there are other weather-related factors we had access to - namely, temperature, windspeed, and humidity. In our creation of a predictive model for the number of daily users of the service, we predicted that on average, we will see 911 fewer users on days with high humidity (defined as greater than .9) than on days with low humidity (defined as less than .3) - a trend that is intuitive, yet more stark than even I would have imagined. After all, that is a drop of 37% - quite a drastic figure. It is readily apparent that both temperature and "atemp" - our measure of how the temperature felt on that day - both had strong correlations to the number of users on a given day. Simply splitting the data down the middle generated clear insight into this trend - by finding the mean temperature and splitting the data into an above and below average group, we were able to note that days with a below average temperature saw far fewer users than days with an above average temperature. We saw 40% fewer users on days with below average temperatures than on days with above average temperatures - a trend that was strongest in the spring, where 5553 individuals used the system on days with above average temperatures, and only 2449 on days with below average temperatures - a 55% differential.

2.

a) H_0 of TV: There is no statistically significant correlation between TV advertisement budget and number of units sold. **REJECT based on p-value**
H_0 of radio: There is no statistically significant correlation between radio advertisement budget and number of units sold. **REJECT based on p-value**
H_0 of newspaper: There is no statistically significant correlation between newspaper advertisement budget and number of units sold. **FAIL TO REJECT based on p-value**

b)

i) $F(x) = 50 + 20X_1 + 0.07X_2 + 35X_3 + 0.01X_4 + -10X_5$

ii)

1) False
2) True
3) True
4) False

iii) $137,000

iv) $135,720

v) False - there is little evidence that the GPA/IQ interaction has a significant correlation to salary, however we cannot determine whether the interaction itself is significant.

c) $F = 0.001 * 0.025X_1$, where X_1 = acceleration.
The intercept appears to indicate that the starting point for mass (i.e. when acceleration and force ~ 0) is .001. No, we have not discovered new physics, this model is simply an artifact of the sample data we have. If we were able to feed in the population data of the mass of every object in existence, our model would likely show a 1-to-1 correlation between mass and acceleration.

d) Possibly. It is also possible that the friend's model simply fit the training data better than yours, and that theoretically if introduced to the entire population those R^2 would change, and possibly show that your model was indeed a better fit.