# Analysis Overview

Analysis began by searching for what factors played the biggest role in determining the number of bike rentals I could expect in any given situation. I expanded beyond just dates and times to look at many factors. Planning and initial testing lead me to use the following factors:

* Season - I accounted for Fall, Winter, Summer and Spring
* Hour of day - I accounted for all 24 hours in a day.
* Holidays - True or false. Was that day a holiday?
* Working days - True or false. Was that day a working day?
* Weather situation - I accounted for four generalized weather situations
  + Ideal weather - Clear, Few clouds, Partly cloudy
  + OK Weather - Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds
  + Not Ideal Weather - Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
  + Bad Weather - Heavy Rain, Ice Pellets, Thunderstorm + Mist, Snow, Fog
* Humidity - The measure of moisture in the air as a percentage.
* Wind Speed - The speed of the wind in “miles per hour”
* Temperature and “feels like” temperature in Celsius.

Once factors were defined I built a neural network that took these in as inputs and ran them through various scenarios to make predictions. After some tweaking, optimizing, and exploring with the data I found it rather difficult to be very accurate with the data. Predictions led me to be able to successfully predict how many bikes would be needed on a given day, give or take 35 bikes. We will expand upon this finding below.

# Bike Rental Prediction

When analyzing the data I found there were many cases that were very similar (i.e. the input factors were fairly similar) but the resulting number of riders was very different. This caused some issues because the model would then correct for a given number of riders and that scenario would be slightly thrown off.

One of the things that really hurt calculations was not initially including hours in the pool of input factors. This caused the model to get even more confused because the problem mentioned above. The scenarios were all too similar with wildly varying outputs. As I trained the model more and experimented with other features, I found that having the hour available for the model to analyze very significantly improved the accuracy of the model.

As mentioned in the analysis overview led me to the conclusion that I can predict, within 35 bikes, how many bikes will be needed in any given scenario that arises. 35 bikes is within 1 standard deviation of how many bikes would normally be rented in a given circumstance, and the model was exactly right as to how many bikes would be required 450 times in the test data, while being within less than a single standard deviation over or under the majority of the time.

As far as a factor analysis goes, the weather seems to have the largest impact over time on bike usage. The standouts in this category are the temperature and the humidity. The warmer days tend to have higher usage while the humid days tend to have lower usage.

# Conclusions and Caveats

In conclusion, the model is accurate and will likely be beneficial in determining how many bikes to put out on a given hour of the day. It will give accurate results up to plus or minus 35 bikes, which is within a single standard deviation of how many bikes would be rented in any given hour. Weather-related factors tend to have the greatest predictive value.

One caveat that I would like to put extra emphasis on, is that due to the feature Holiday occurring so few times a year, the model tends to perform poorly on those days. One noticeable improvement that could be made would be to give extra weight towards Holidays as a feature, since it is likely that, even given the same weather situation and temperature, a day that is a Holiday is more likely to see an increase in riders.

Another issue I ran into that would help to optimize model was the overfitting of the data. Model was able to predict based on the training data much more accurately than on the testing data. This “overfitting” can happen for a lot of reasons but I ran out of time to dive any further into fixing this past initial attempts (see [Overfitting Correction](https://colab.research.google.com/gist/ipetty14/361bfb0b4be6a07f4e2a240c8631f025/biking-neural-network.ipynb#scrollTo%3DsZfPeMJUwl-0)) . There were simulations where training data was getting within a difference of only 20 to 25 of the true number of riders. If given more time to dive into this, \model could be optimized to yield more accurate results with the test data and allow me to get an even more accurate number.