The purpose of the study was to use Divvy Bike sharing data provided by the city of Chicago to determine whether trip duration was longer in warmer temperatures. I obtained the clean dataset from Kaggle and used the following variables:

* Trip Duration of the bike ride (2014-2017 Data)
* Temperature during the time of the bike ride
* Month of the ride
* Week of the ride
* Day of the ride
* Hour of the ride

I feel that I missed the distance between the start and end of the bike ride to determine just how long it was. Was the bike returned to its original station or did it travel to another location? Were most of the bikes downtown? Where do people usually travel? Is it used for commuting, etc. I assumed that people would usually ride their bikes for fun, but by determining which was the most common hour that people rode the bikes, it makes me believe that it could also be used as another method for commuting. The challenges that I faced were not being sure of what variables to use for this analysis, even though there were so many. I wasn’t sure what categorical data would be the most useful. Also, would it have been helpful just to get the past year’s worth of data or did I need to use all of the data possible? It took some time for my jupyter notebook to complete running the analysis due to the large scope of data that it was handling.