Data Explanation

The gaps in bills and disconnects are due to service provider reading schedules and company policies. We do not issue bills on holidays, etc. The data *should* be set up so that on any given day, if there are no bills due, there are no disconnects, but you'll see we have a few unmatched single digit disconnects. We're still investigating why that is, but the purpose of this model is to identify the large spikes in disconnects, so we can essentially ignore those small single digit disconnects.

In a full sentence, the purpose of this model is to predict electricity disconnect notices on a given day, so that we can warn the call center of a potential spike in calls. You'll see there are a few days where the disconnects were upwards of 800, and those days were very bad for the call center, due to not having any expectation of a large influx of calls.

So far, the best model I could produce was an additive model, shown below in R code. With Poisson and an offset, the average difference between predicted values and actual Y is much larger. I was getting a significant p-value for including enrollments as a spline, but feel free to dispute that if you and your PhD assistance disagrees.

Thank you so much! I'm excited to be able to provide an "interesting" example.

```
my.model<-gam(Disconnects~rolling_expirations
    +s(rolling_enrollments)
    +Bills_Due+is_winter,data=data)
```

- Date: row identifier
- Rolling Expirations: A 15 day rolling average of customer plan expirations. Many customers don't pay their final bill, which can result in more disconnects
- Rolling Enrollments: A 47 day rolling average of new customer enrollments. Many new customers miss their first bill due to initial setup confusion, which can result in more disconnects. The 47 day range accounts for the amount of time that can occur between initial enrollment and actual electricity set up.
- Disconnects: Actual disconnect notices received. Notices occur one day after a bill due date has expired.
- is_summer: indicator for the summer months
- is_winter: indicator for the winter months
- is_shoulder: Not winter or summer
- Bills_Due: # of actual bill notices that were sent out (Bills_Due - Disconnects will equal the number of people who paid their bill and avoided a disconnect for that particular day)