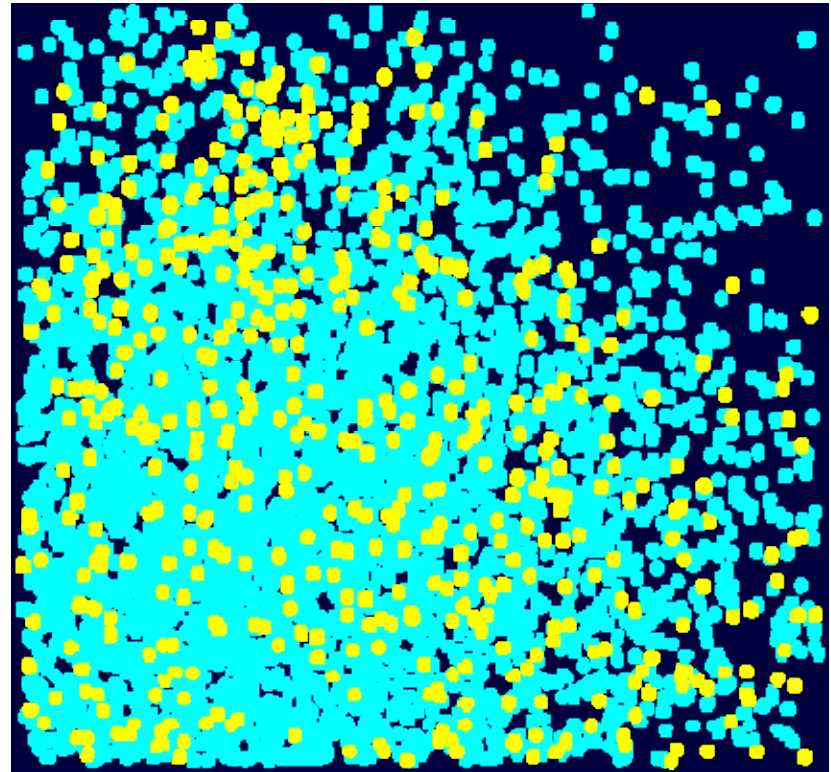


Tips & Tricks: Oversampling

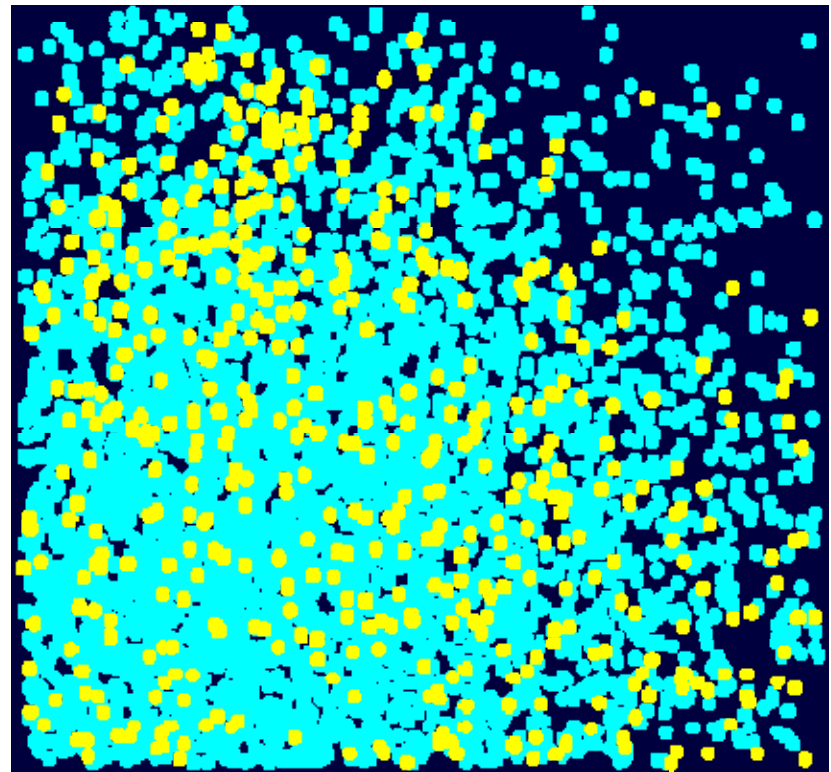
Outcome Overrepresentation

A common predictive modeling practice is to build models from a sample with a primary outcome proportion different from the original population.



Outcome Overrepresentation

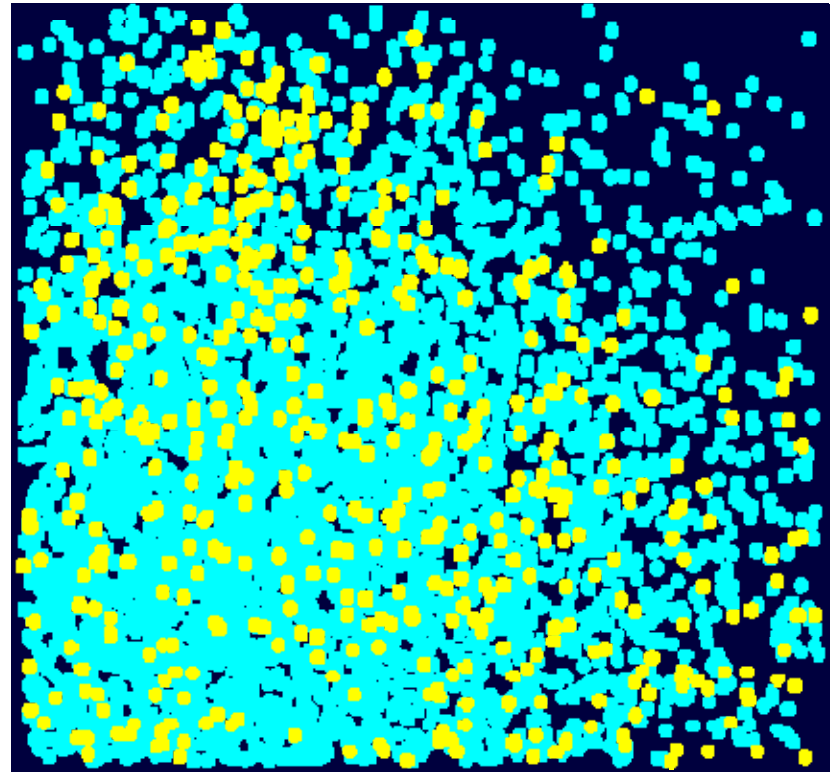
A common predictive modeling practice is to build models from a sample with a primary outcome proportion different from the original population.



Separate Sampling

secondary outcome

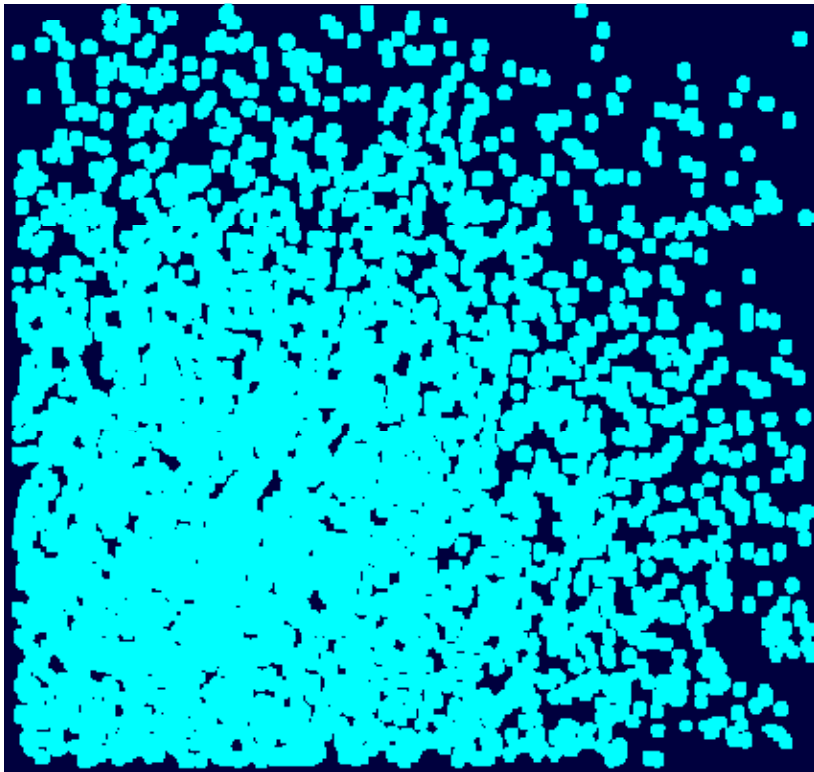
primary outcome



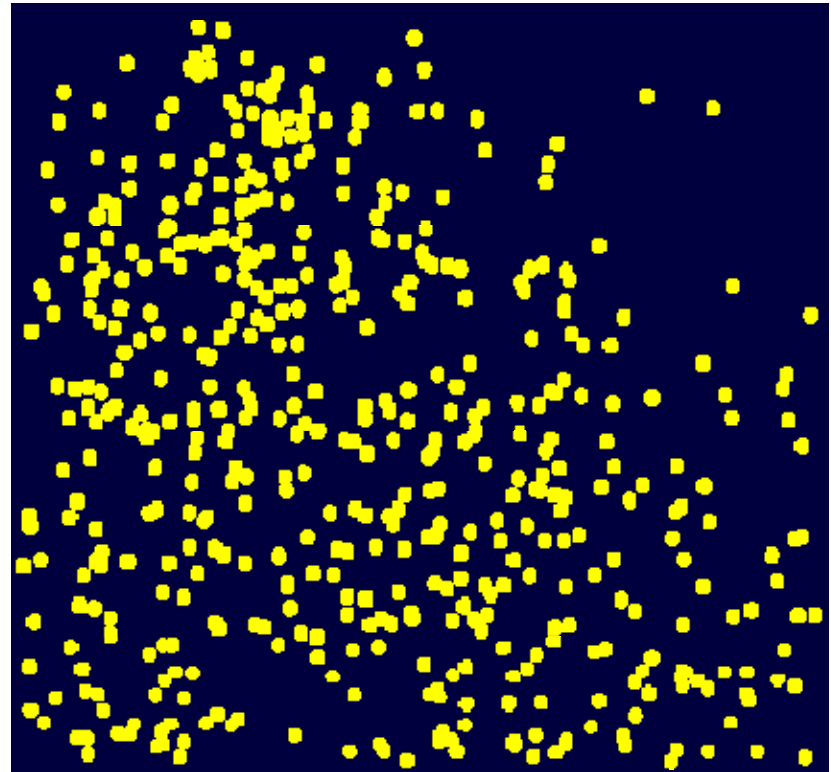
Target-based samples are created by considering the primary outcome cases separately from the secondary outcome cases.

Separate Sampling

secondary outcome



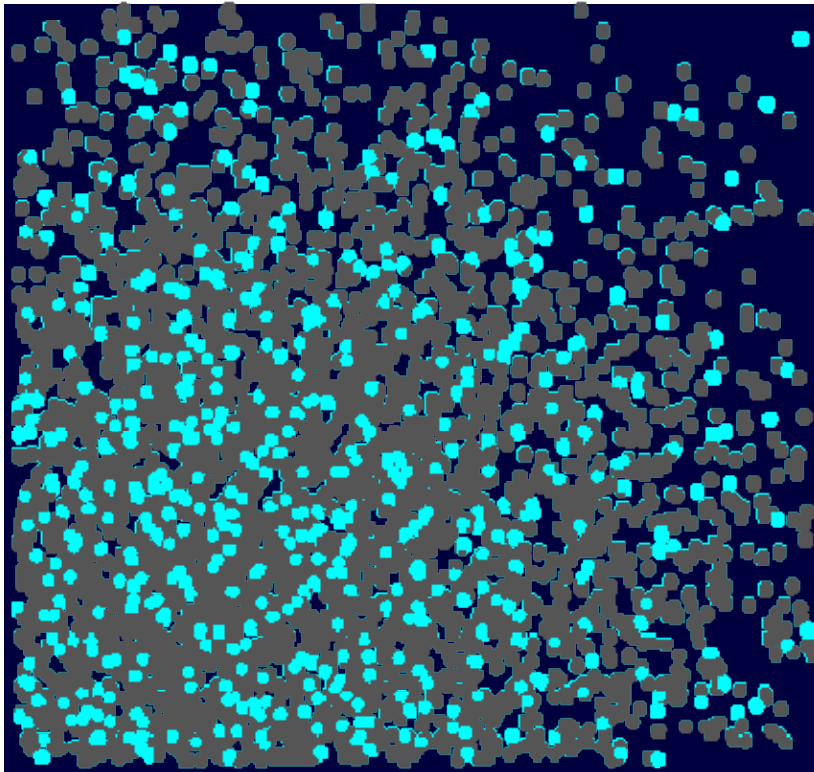
primary outcome



Target-based samples are created by considering the primary outcome cases separately from the secondary outcome cases.

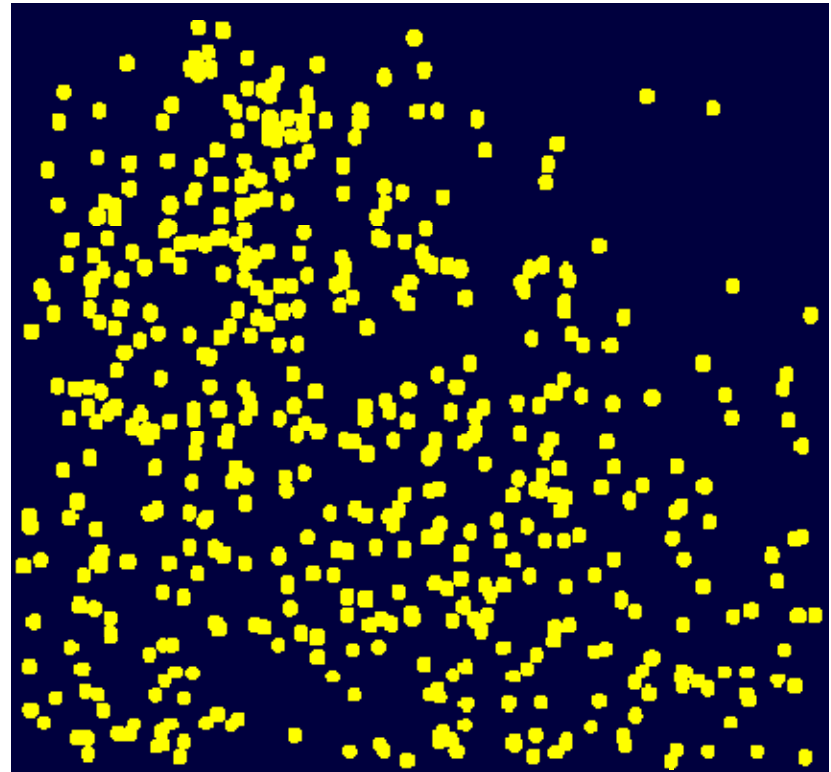
Separate Sampling

secondary outcome



Select some cases

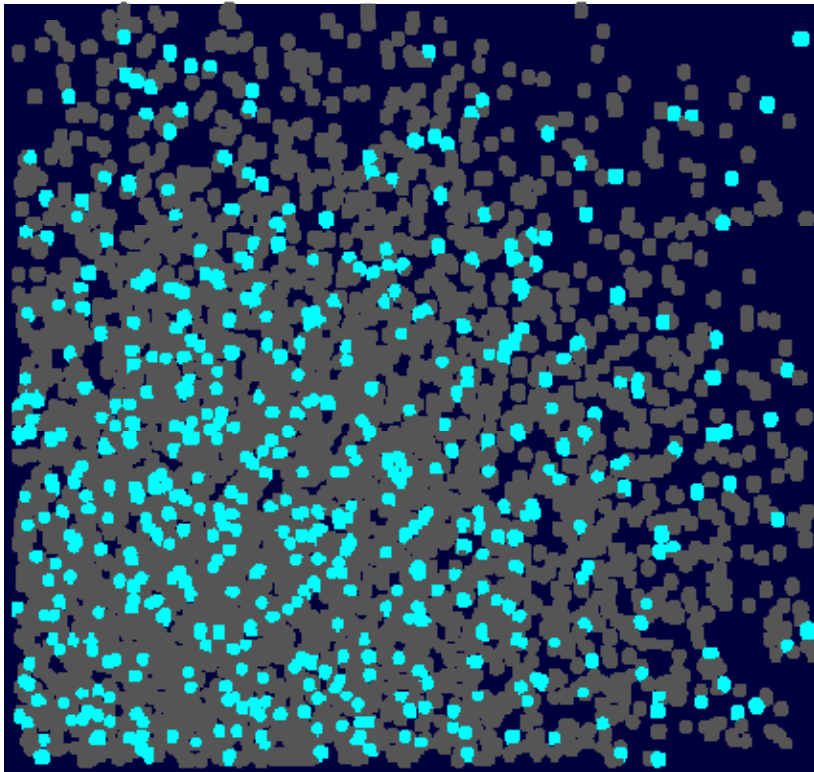
primary outcome



Select all cases

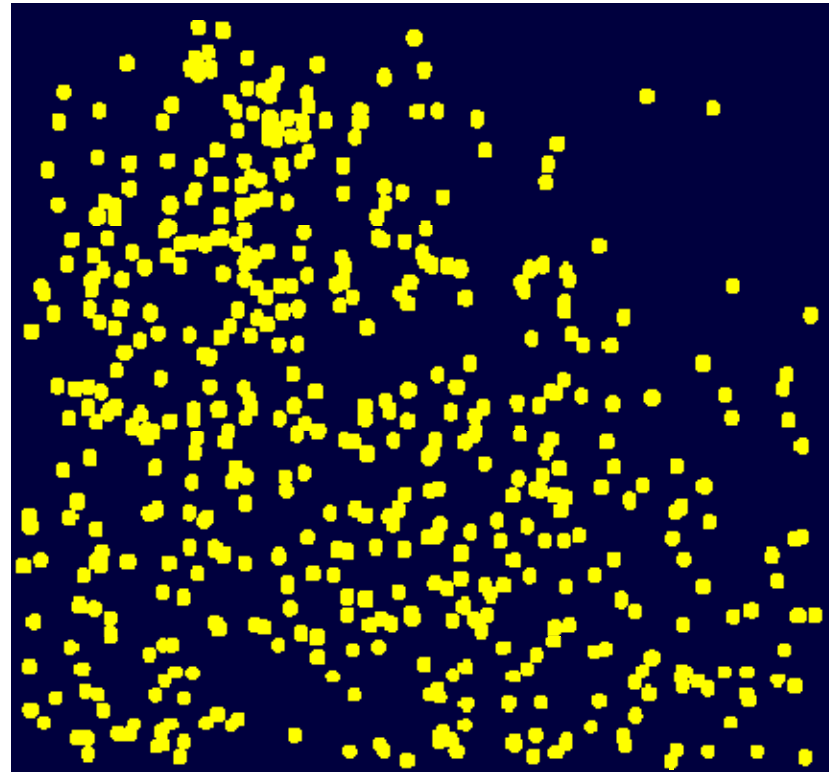
Separate Sampling

secondary outcome



Select some cases

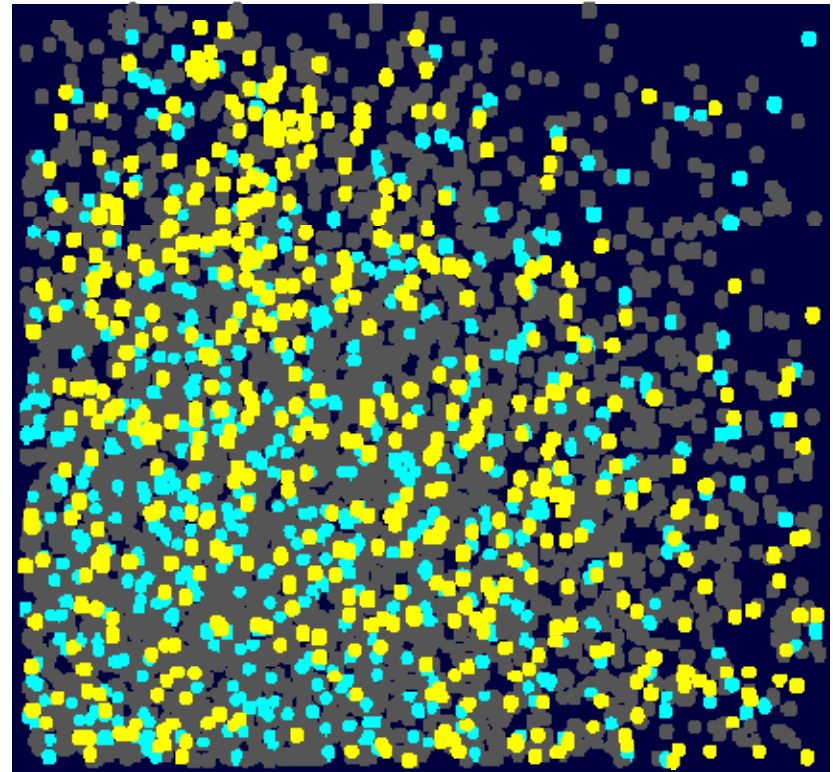
primary outcome



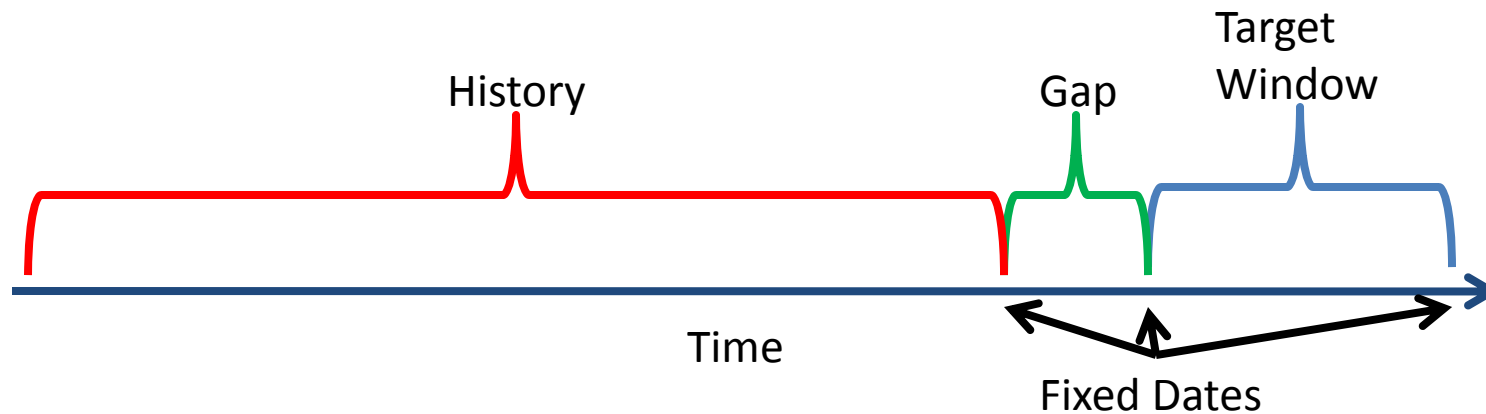
Select all cases

The Modeling Sample

- + Faster run times
- + Similar (better?) predictive power with smaller case count



Sampling For Churn/Attrition/Cross-Sell Modeling: FIXED TIME



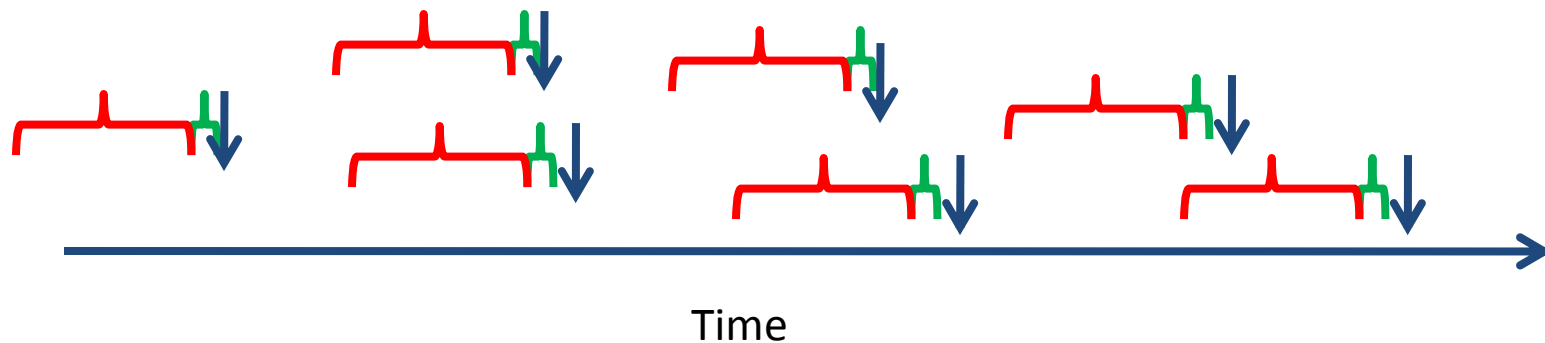
Benefits

- Easy

Drawbacks

- Model may not generalize to other times/seasons
- Insufficient number of events

Sampling For Churn/Attrition/Cross-Sell Modeling: VARIABLE TIME



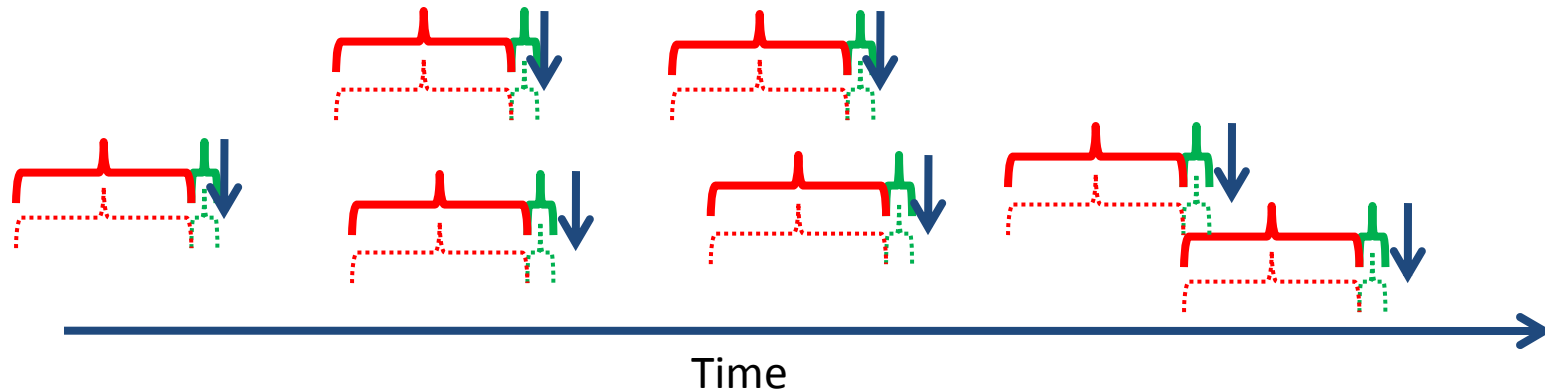
Benefits

- Better generalization to other time/seasons
- Can control for exogenous factors (e.g. economy)
- Yields larger number of events.

Drawbacks

- More date sensitive data preparation for history.
- **How do you select matching non-event cases and assign “event dates”?**

Sampling For Churn/Attrition/Cross-Sell Modeling: VARIABLE TIME



Benefits

- Better generalization to other time/seasons
- Can control for exogenous factors (e.g. economy)
- Yields larger number of events.

Drawbacks

- More date sensitive data preparation for history.
- **How do you select matching non-event cases and assign “event dates”?** **SIMPLE**