The features would be represented up to Feb 28th -(that is what I usually call a cut off date.

The target would be March 1-March 30 – that is what I usually call a target time period.

In addition you want a start date of which the data is available. In this data that is Jan 1 -2015.

So Some of your features would look like this. Let’s say you use 2

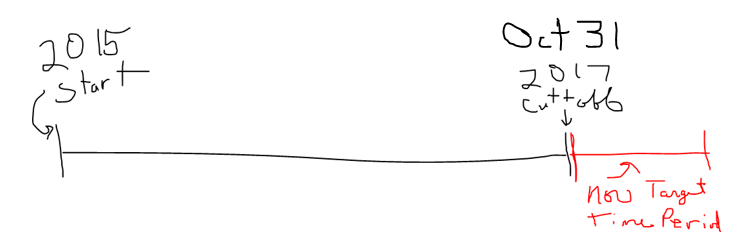
(1) Days\_since\_last\_donation

(2) Total\_Donations

The transactional will look something like this.

|  |  |
| --- | --- |
| ID | Registration Date |
| Person A | 1/1/2016 |
| Person B | 9/20/2015 |
| Person A | 6/19/2017 |
| Person B | 10/19/2016 |
| Person C | 10/1/2015 |
| Person B | 1/15/2017 |
| Person B | 8/2/2017 |
| Person D | 11/5/2017 |
| Person B | 11/15/207 |

Let’s have the cutoff date be October 31 - 2017 and the target time period be from Nov 1 – Nov 30 -2017.



You have to calculate Total\_Donations and Days\_Since\_Last\_Donation relative to the cut off date Oct 31 -2017.

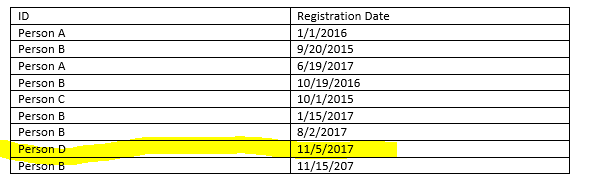
|  |  |  |
| --- | --- | --- |
| **ID** | **Total Donations** | **Registration Recency (Days)** |
| Person A | 2 | 136 |
| Person B | 4 | 91 |
| Person C | 1 | 762 |

Total donations is the count of donations from 2015 – October 31 – 2017

Registration recency is the date diff from the last registration date and the cut off date.

That’s the covariate query.

I put a person D in there that is not used because they were new in the target time frame and did not have any covariates to calculate. So we can not use them.



Now for the Target Query.

Who registered in the target time period (Nov 1 – Nov 30 – 2017)

|  |  |
| --- | --- |
| **ID** | **Target** |
| Person A | 0 |
| Person B | 1 |
| Person C | 0 |

The final training data will look like this

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Total Donations** | **Registration Recency (Days)** | **Target** |
| Person A | 2 | 136 | 0 |
| Person B | 4 | 91 | 1 |
| Person C | 1 | 762 | 0 |