

**SAS EM NODES**  
TORONTO DATA MINING USER GROUP



## **AGENDA** | WHAT'S NEW IN ENTERPRISE MINER

- Survival Data Mining
- Incremental Response Modelling
- Time Series Similarity Analysis

# SURVIVAL DATA MINING



# **SURVIVAL DATA MINING ANALYSIS**

**PREDICTION OF WHEN AN EVENT WILL HAPPEN AND NOT JUST IF IT WILL HAPPEN  
VERY POPULAR FOR CUSTOMER BEHAVIOR MODELLING, SUCH AS**

- When will a customer churn?
- When will a customer cancel a service or product?

**PREDICTS EVENT PROBABILITY FOR TIME INTERVALS FOR EACH CUSTOMER**

- i.e. Customer has 50% chance to cancel next month but 75% chance to cancel the month afterward.

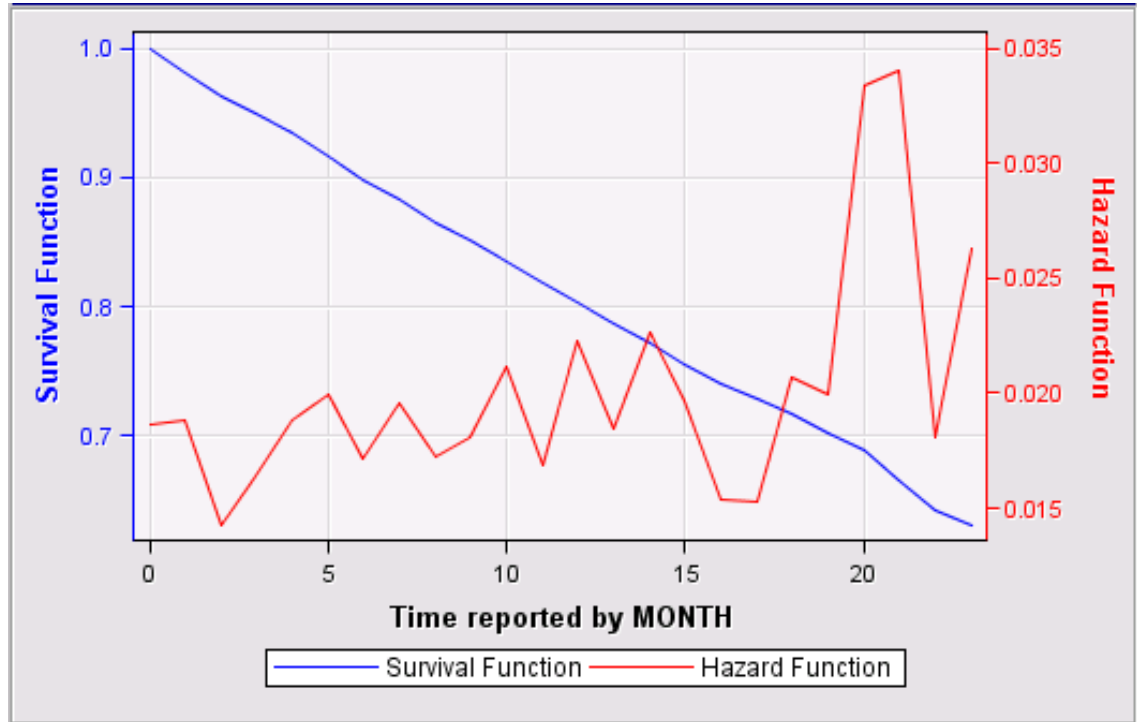
**CAN TAKE EXTERNAL FACTORS INTO ACCOUNT**

- i.e. customers with more than 2 products tend to stay longer

# SURVIVAL DATA MINING ANALYSIS

## APPROACH

- Look at probability of hazard (event) at discrete time points
- Compared to survival probability (tenure of customers overt time)



# SURVIVAL DATA MINING

## SCORING

- Mean Residual Lifetime: expected time till event occurs based on projecting hazard function into the future
  - Projection based on constant hazard function
  - Projection based on continuing trend of hazard function

Obs #	dataobs	acctno ▲	actdt	deactdt	deactrea...	goodcredit	rateplan	dealertype	Event Ty.	Mean Residual Life RMRL	Warnings	Survival ... Sur
1	1117699186...		05/24/2000			1	1A1			31.65814		0.899157
2	4117740111...		09/18/2000			1	1A1			34.00226		0.93289
3	5117748047...		07/29/1999			0	1A1			13.54889		0.640366
4	7117800373...		01/15/2001			1	2A2			34.22137		0.98088
5	1117881391...		01/25/2000	01/03/2001	NEED	1	1A1			29.17131		0.844845
6	12117889001...		01/07/2000	10/23/2000	COMP	1	1B1			27.20979U		0.850012
7	13117901984...		12/10/2000			0	1A1			23.03729		0.962881
8	14117908143...		08/29/1999			1	1A1			25.95794		0.794696
9	15117919034...		12/20/2000			1	1B1			34.03814		0.971964
10	17117949894...		09/03/2000			0	1A2			20.82965		0.894392
11	18117981146...		11/27/1999			1	2B1			25.35756		0.828127
12	19117990409...		09/01/1999			1	1A1			26.64805		0.803768
13	20117995165...		12/16/1999	06/21/2000	TECH	1	3A1			25.44135U		0.791395
14	21117995935...		09/06/2000			1	1A1			34.00226		0.93289
15	22118007089...		07/24/2000			1	1A1			32.87666		0.909987
16	23118053724...		11/15/1999			0	1A1			16.07687		0.703478
17	26118140031...		12/21/2000			1	2B1			33.54243		0.972346
18	27118145692...		04/28/1999			1	1A2			22.58443		0.727733
19	28118152852...		11/10/2000			0	1A1			22.43449		0.947059
20	30118180180...		04/16/1999			1	3A1			20.36969		0.699613
21	32118204649...		06/18/1999			0	2A1			12.27994		0.609827
22	35118253024...		08/09/1999	12/08/1999	DEBT	1	1A1			25.95794U		0.794696
23	36118257619...		05/13/2000			1	1A2			30.62251		0.873631
24	37118270251...		01/08/2000			1	1A1			29.17131		0.844845
25	38118357672...		07/21/2000			1	1A1			32.87666		0.909987
26	40118377579...		12/21/2000			1	3A1			31.73538		0.96147
27	43118435853...		02/20/1999			0	3A1			8.436454		0.441214

# INCREMENTAL RESPONSE MODELLING



## OVERVIEW WHY DO WE CARE ABOUT INCREMENTAL RESPONSE?



### Target:

customers who will buy the product if and only if it is promoted directly to them

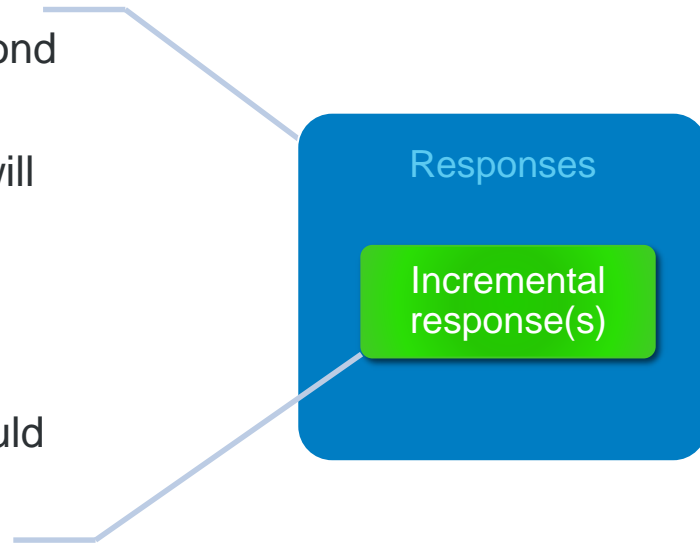
### Common Issues:

- Negative effects by annoying customers who would be less likely to respond if targeted (**Do Not Disturb**)
- **Waste on** customers who would have responded regardless of the incentive
- **Waste on** customers who would NOT respond whether they are targeted or not



## OVERVIEW RESPONSE VS. INCREMENTAL RESPONSE (IR)

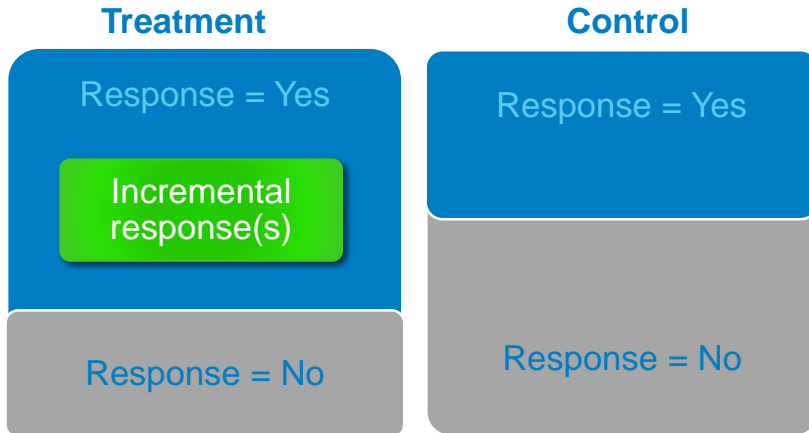
- Traditional response modeling
  - Identifies customers who are likely to respond
  - Targets all the respondents
  - Still wastes resources on customers who will respond no matter what
- Incremental response (IR) modeling
  - Identifies the additional purchases that would not have taken place without the campaign
  - Measures the true effect of a marketing campaign



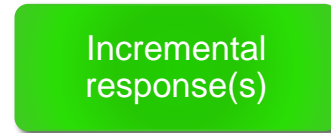
# INCREMENTAL RESPONSE MODEL

## MODELING DETAILS

- True-lift / Net-lift / Uplift Modeling
- Target only customers who produce incremental responses
- **Treatment** (promotion) group: Assume that incremental responses exist, but do not know which observations are incremental responses
- **Control** (no offer) group: No incremental response



- Build predictive model to identify



# TIME SERIES SIMILARITY



# TIME SERIES SIMILARITY

TRADITIONALLY, DATA MINING AND TIME SERIES ANALYSIS HAVE BEEN SEEN AS SEPARATE APPROACHES TO ANALYZING ENTERPRISE DATA.

HOWEVER, MUCH OF THE DATA GENERATED BY BUSINESS PROCESSES IS TIME-STAMPED.

TIME SERIES DATA MINING IS A MARRIAGE OF FORECASTING AND TRADITIONAL DATA MINING TECHNIQUES THAT USES TIME DIMENSIONS AND PREDICTIVE ANALYTICS TO MAKE BETTER BUSINESS DECISIONS.

# TIME SERIES SIMILARITY

## SIMILARITY SEARCH OR INDEXING

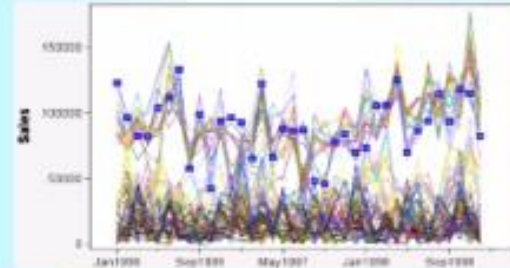
Target (Query) Series



Similarity search  
and indexing



Time Series Database



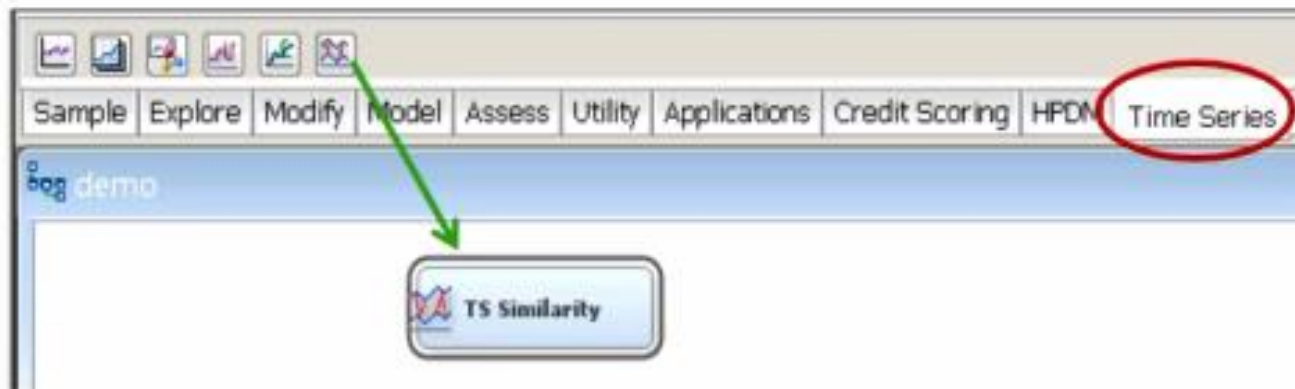
Given a measure and a target (query) time series:

- find the most similar or dissimilar time series
- index the time series database

Some real-world examples: Stock price movement, sales patterns, medical device data

# TIME SERIES SIMILARITY

## ENTERPRISE MINER NODE



- Provides similarity search and clustering
  - Similarity search: when target time series are specified
  - Clustering: no target and all time series are inputs

## USEFUL LINKS

- Videos: <http://support.sas.com/rnd/app/video/index.html#txtmine>
  - [Time Series Similarity Analysis](#)
  - [Incremental Response](#)
  - [Survival Data Mining](#)
- Papers:
  - [Time Series Data Mining with SAS® Enterprise Miner™](#)
  - [It's About Time: Discrete Time Survival Analysis Using SAS® Enterprise Miner™](#)