Behaviour Analysis MDX

Andrew Baker

- Content - A SSAS MDX Deep Dive session covering the implementation of a business requirement to identify rogue behaviour over time and within SCD organisation groups.
- Credits
  - Rich Carr & Dave Tolladay (Alerts4FM) – Requirements, Code & Testing
  - Chris Webb – a Code suggestion, and
  - Andrew Baker (Guest & Baker) – 30 years developing MIS

Copyright Alerts4FM/Guest & Baker 2015

DETECT, ANALYSE & RANK

Good or Bad Behaviour

“CRADBURDS”

- CONCEPT (clarify the initial ideas via a PoC)
- REQUIREMENTS (Functional & Non-Functional)
- ANALYSIS (Both Business & Technical, Solution Architecture)
- DESIGN (Both User perspective & Technical Implementation)
- BUILD (Regression Tests, Code, Code Review, SIT)
- USER ACCEPTANCE TEST (verify it is what they had in mind)
- RECONCILIATION (intra & inter system data reconciliation)
- DEPLOYMENT (Re-Baseline, Revisit Capacity Planning, Rollback)
- SUPPORT (*Pro-active* monitoring against a Baseline)

CONCEPT

- A senior Business Sponsor drives the Concept
- In this case, the Basic Concept is to detect rogue behaviour
- Why? Rogue Behaviour can lead to a HUGE cost
- The Trader who brought down Barings Bank in 1990s
- But, behaviour analysis applies to *both good and bad behaviour*
- Need a meaningful PoC to clarify the concept
- Establish the Boundaries at an early stage
- SCOPE – what is IN and what is OUT

REQUIREMENTS

- First – DEFINE what you are trying to measure
- Next – you need to DETECT it
  - Need to capture the required data and validate its correctness
- Next - You need to ANALYSE it
  - Which you can do in (at least) 3 ways
  - Measure against an Absolute Threshold value
  - Measure against the Historical Data for the person
  - Measure against the person's Peer Group
  - Measure using "statistical analysis" e.g. Standard Deviation
- Finally – RANK people, using a POINTS system

ANALYSIS & DESIGN

- An Example – Non-Core Hours
- Definition – Activity outside a person's official working hours
- How important is this Activity? Depends on the context
- Good or Bad – Why is it happening? Regularly?
- Data Capture – Need the person's core hours of work
- and the time of their transactions
- Design – User Perspective – How to "report" it?
- Design – Technical – Platform – SQL v MDX
- PoC – Explore preferred alternative (technical design)

MDX? …

LAG, ANCESTORS, AVG, STDEV(P)

But – Slowly Changing Dimensions?

BUILD - PoC

- I have a simple example (for this presentation) …
- Sales by Persons to Customers of Products over Time
- The Customers & Products are not strictly relevant
- I want to analyse each Person's non-core hour activity
- I want to ***leverage*** existing data structures i.e.
- The sales data already exists in a relational database
- and also in a OLAP (SSAS) Cube
- I just need to build on this platform … (show SSMS & SSDT)

CODE DEMONSTRATION

- Use SQL to DETECT the Non-Core Hour: COUNT per Person
- The example SQL is simple – can get much more complex
- Use SSAS cube technology to **aggregate** the counts
  - For all Customers and Products, for a defined Period e.g. Week
- Only interested in the Total NCH for each Person per Week
- Use MDX to ANALYSE these aggregated counts
  - Use MDX LAG function for “Historical” comparison
  - Use MDX ANCESTORS function for “Peer Group” comparison
  - Use STDEV function for “sample” population group analysis
  - Use STDEVP function for “precise” peer group analysis

GOTCHAs

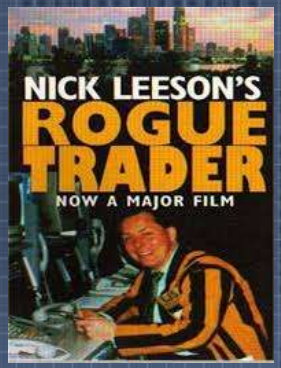
- The Devil is always in the Detail … every time … ☺
- First – Reporting Requirement
- Combine Look-Up and Look-Back Calcs in *same* query
- Separate Look-Up and Look-Back calcs will *not* work
- Need to combine the two – start at the lowest "DimId" level
- Next – People move between Peer Groups
- aka Slowly Changing Dimensions ("SCD")
- Need to associate "DimPersonId" with the Date dimension
- Achieved by using a PersonActiveDate Measure Group

RANKING

- Calculate POINTS for each Person
- Use the MDX RANK function to rank the Persons dynamically
- And look-out for (beware of) Anomalies
- Especially when it comes to Statistics!
- “There are Lies, Damned Lies, and then there are Statistics”
- Statistics are sometimes used to bolster weak arguments
- But, in this case, they can give the “wrong” impression

SUMMARY

- Rogue behaviour can be immensely costly
- These people cost their organisations:-





$1.3bn → $2.3bn → $2.6bn → $6.9bn

- Rich Carr and Alerts4FM have shown one way to *try* to STOP it!
- Finally, if rogues know they are being monitored, it might just deter them.