Analysis in Corporate Transaction Environments

Performance

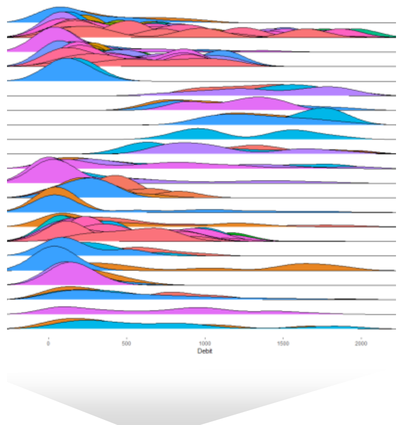
- *Corporate*
- *Subsidiary*
- *Division*
- *Channel*

…

Data Levels & Dimensions

- *Source Systems*
- *Accounts & Attributes*

…



Drivers

Economic

- *Industry*
- *Competition*
- *Channels*
- *Technology*

…

Transactional

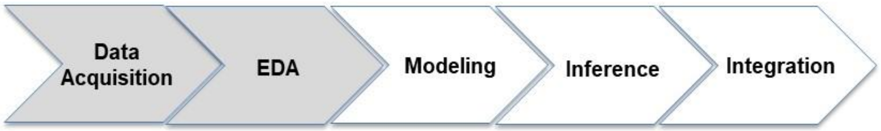
- *Suppliers*
- *Customers*
- *Transactional Attributes*

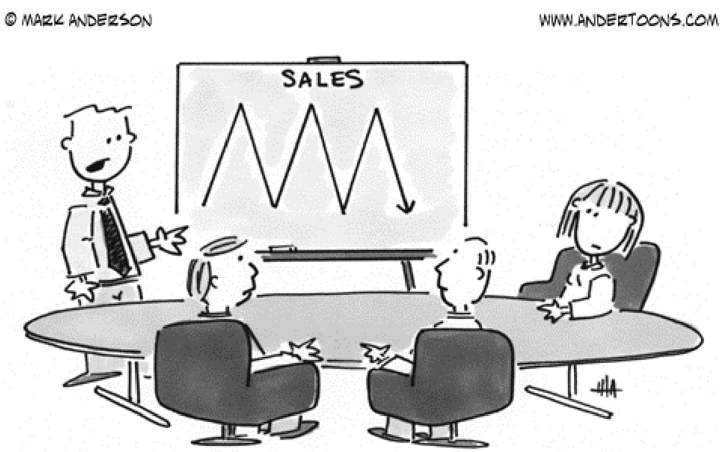
…

Corporate systems can be extensive in scope and scale. ERP systems generally manage the accounting processes, and some portion of operations, but rarely manage all of the end-to-end processes. *[Companies choose to manage operations outside ERPs for many reasons, but mostly strategic: "Very rarely will individual companies gain durable advantages through the deployment of packaged applications"]* [1]

While analysts *(and auditors)* often begin work by reviewing top-level dashboards, data in these business intelligence *(BI)* warehouses are highly aggregated and dimensions are pre-defined. So, the detailed transactional data needed to associate many operational drivers to financial performance is often not available.

[1]*Michael Porter – Harvard Business School*



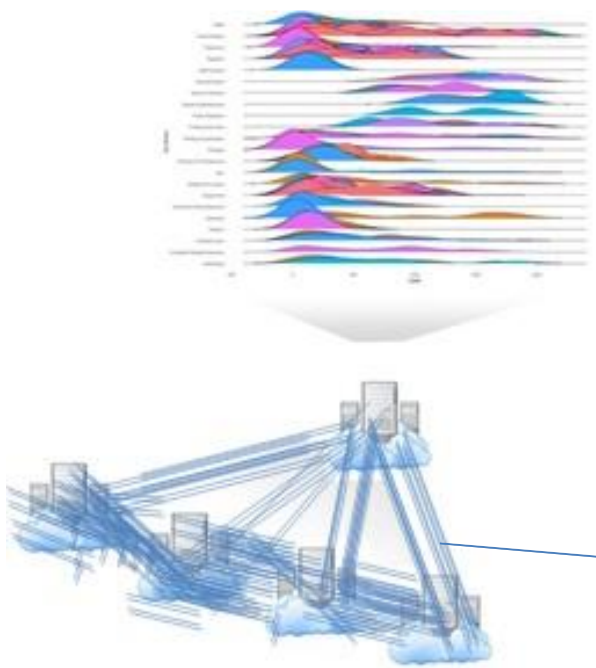


"..and then another drop this month. But, I have a really good feeling about next month."

Visualization dashboards can build intuition and understanding. And sometimes, that's the goal: Recall how we used visualization to understand data in *Exploratory Data Analysis* **(EDA).**

But other times, we need prediction and inference[1], in which case we move on to *Modeling* processes. Recall from level 1, that we defined 3 types of models and parameter production:

1. ***Non-Parametric*** *(e.g., visualization)* does not provide prediction or inference.
2. ***Noninterpretable Parametric*** *(e.g., black box machine learning),* provides prediction but inference becomes intractable because of the dimensions and complexity of parameters.
3. ***Interpretable Parametric*** *(e.g., regression),* provides prediction and inference, but can becomes intractable with large numbers of dimensions.



An economic organization is a living breathing organism and ***all*** of its parts are ***interrelated*** through ***levels*** and ***dimensions***. For example:

- ***Levels***. Margins of product A are *increased* by complimentary channel demand, as well as brand improvement. Margins are *decreased* by competitor capacity, political friction and import clearances, and supplier disruptions.
- ***Dimensions***. Margins of product A are *increased* by *decreased* machine time in factory 1, and decreased material costs with supplier B. Margins are decreased by discounts to customer segment 4 and increased material costs with supplier 3.

How much do increases and decreases impact a level and variable of interest? That's the ***Effects.*** Effects compliment and conflict with each other within and across levels and dimensions. ***Understanding these effects is key to understanding the business and constructing inferences in a complex, high volume, dynamic world. Effects give us a way to quantify business motion.***

Bayesian Modeling in Transaction Environments

Bayesian modeling is especially strong in **business transaction environments:**

- ***Multi-Level.*** Transaction environments are inherently multi-level *(i.e., hierarchical)*, and Bayesian models produce ***parameters that measure the effects within and between levels.*** This adds a great deal of flexibility and power to inference *(i.e., we can state how much a channel and region will influence product introduction)* and we can also make statements about the credibility of those parameters [1]

- ***Distributions.*** Variables in transaction environments rarely fit a normal distribution, and neither do the parameters *(e.g., I assume CLT/normal to determine sample size, drawing from a skewed distribution. Now I need to predict / project – on what parameter do I base my prediction?).* Bayesian models fit each variable and parameter to the appropriate distribution, which then provides a basis for credible prediction.

- ***Judgment.*** Judgment *(beliefs)* be quantified, integrated into models, and validated with data. Sometimes, we don't have enough data *(this happens more often than you think - new processes are always popping up in enterprises and data will be sparse to begin –we can use priors and effects to pool the data)*.

- ***Generalization.*** Bayesian models have fine control over generalization. In enterprise data, simple models tend to underfit and complex models tend to overfit (especially with L1/L2 algorithms). Bayesian models give us a means to "fine-tune" generalization to the level / parameter.

- ***Operationalization***. Bayesian models are easier to deploy to operations *(e.g., continuous audit, securities trading)* and still remain relevant as models can be trained in parallel with deployed models and simply pass parameters between.

[1] *it should be said here that multilevel frenquenist modeling is possible, but Bayesian process is far more extensible*

BAUER COLLEGE OF BUSINESS UNIVERSITY of HOUSTON

Review and Look-forward

- Visualization *(and dashboards)* can be helpful in describing data, finding errors, and understanding trends. These are valuable insights and can set direction for further analysis. But visualization can also be misleading, and it is NOT a basis for inference (including prediction, projection, and association), it is a basis for a belief.

- Beliefs *(i.e., judgment, prior data)* can be important inputs for inference, provided that parameters can be assigned (and sometimes that begins with significant uncertainty). But over time, as data is acquired, these beliefs become more and more certain, and the credible range for inference narrows. Keep in mind that data often exists in total, but not downlevel or in a timebox.

- When modeling organizational transactions, holistic models are important for defining overall trends *(i.e., parameter pools)*. Projection becomes an exercise in credibility control across levels - usually the more downlevel you go, the more credible the projection. To this point, beliefs and pooling control the influence of trends downlevel and provide a basis for models to arrive at parameters without sufficient data inlevel.

So to review, any projection based on visualization should be challenged and transferred to a belief to be parameterized and validated across organizational levels.

Section Outline

1. Distributions, Parameters and Likelihood
2. Introduction to Bayesian Theory
    1. Conditional Probability and Bayes Theorem
    2. Single Parameter Models
        1. Closed form solutions: Normal, Beta Binomial, Skew Normal
        2. Stan and MCMC sampling
        3. Simple Bayesian Regression
        4. Multivariate Bayesian Regression
3. Multilevel Modeling
    1. Multilevel Architectures *(hierarchical / nested vs. crossed effects)*
    2. Pooling (full vs partial pooling)
4. Application
    1. Algorithm Assurance Project
    2. Transaction Divergence Project
5. Student Projects
6. Final