Predictive Analytics:
Case Studies in Life and Annuities

So you've got data, now what?

High level tips and tricks for working with data

November 30, 2018
Craig W. Reynolds

Milliman

© Randy Glasbergen for RapidBI.com

"If we learn from our mistakes, shouldn't I try to make as many mistakes as possible?"

What's in store today

Tools & Tricks

- Build a team
- Planning for success
- Data cleaning
- Data exploration
- Predictive modeling & machine learning

Building a team

Who do you need on your team?

Actuaries

Subject matter experts – people who understand the product and industry you're studying

Data scientists

Data manipulators/modelers - people who clean and manipulate data and build models to answer questions

Technologists

Data storage/computation experts – make sure data is stored effectively and you have the tools needed to work with it and present it

Stages of a Project

- Problem definition
- Data gathering
- Data cleaning
- Data enrichment
- Data exploration
- Model building
- Model validation
- Sense making
- Implementation
- Refinement

Problem definition

- Give your analysis direction
- What is your objective?

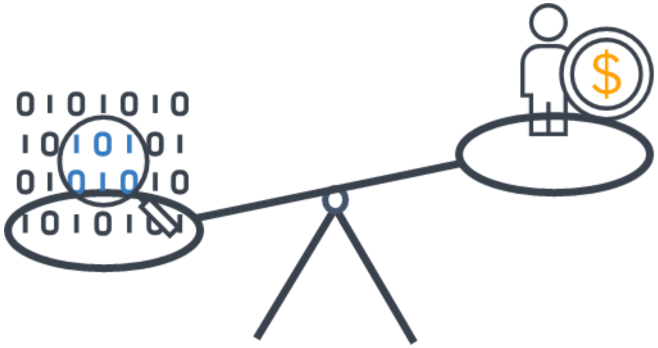
What drives variable annuity benefit utilization?

How will policy lapses vary with economic conditions?

How much universal life premium will policyholders pay?

What data do you need?

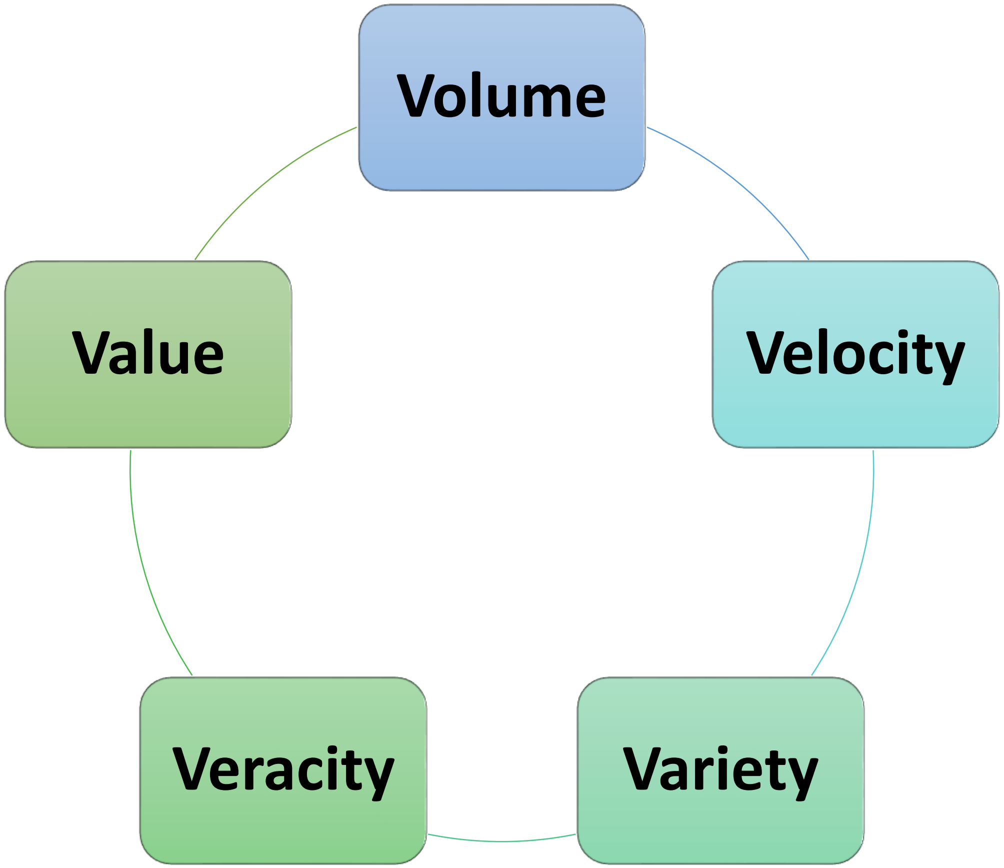
- Data should be specific to the question posed
- May be limited by data you have available
- Longitudinal vs cross-sectional
- Should you consider 3rd party data?



What data should you save?

- Demographic data
- Transaction data
- Underwriting data
- Applications
- Economic data
- Policy features and charges
- Distribution system info

Five V's of big data



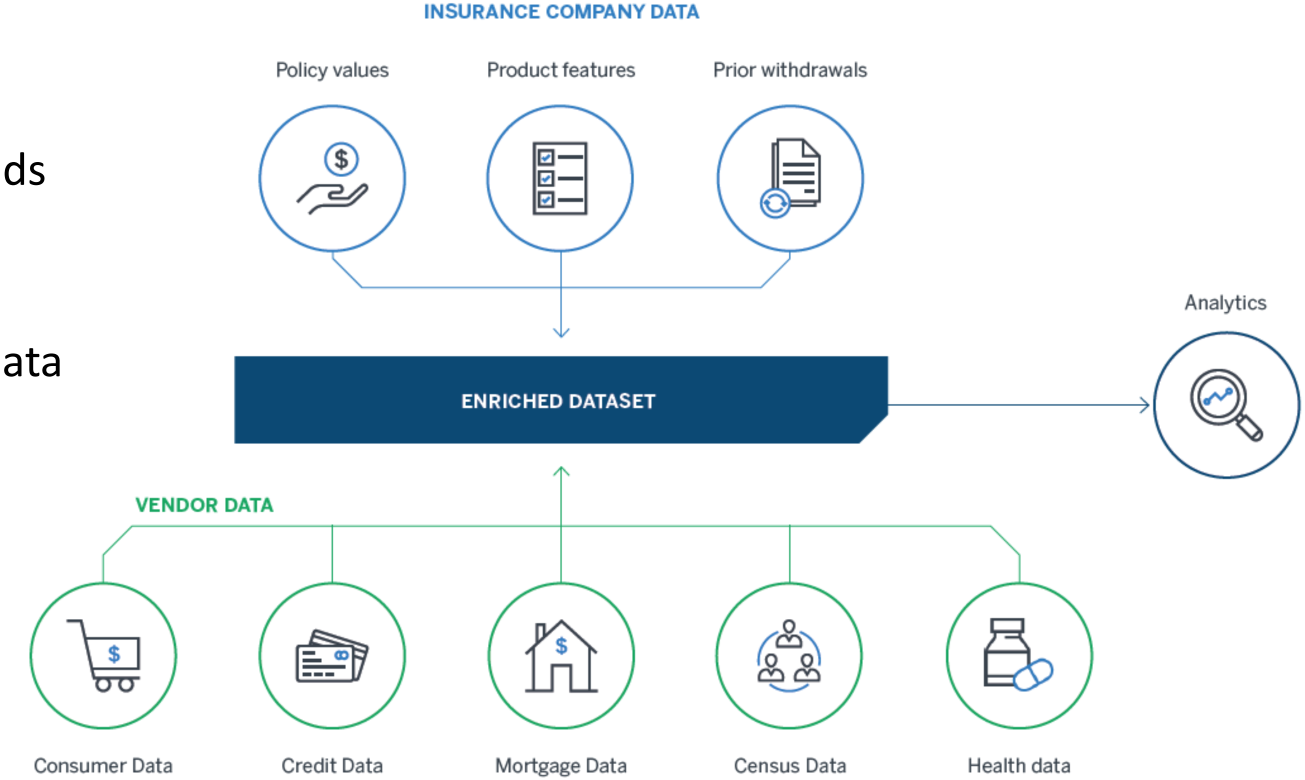
What you want vs. what you have

Want:

- Liquidity needs

Have:

- Policy values
- Third party data



Considerations when using third party data

Legal restrictions

- Consider whether there are legal restrictions that govern the use of particular data sources
- US: Fair credit reporting act (FCRA)
- Europe: General data protection regulation (GDPR)

Privacy

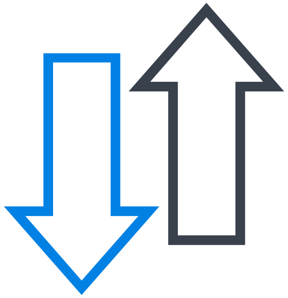
- Will customers feel that their privacy has been violated from the use of specific data
- Are you comfortable?

Security

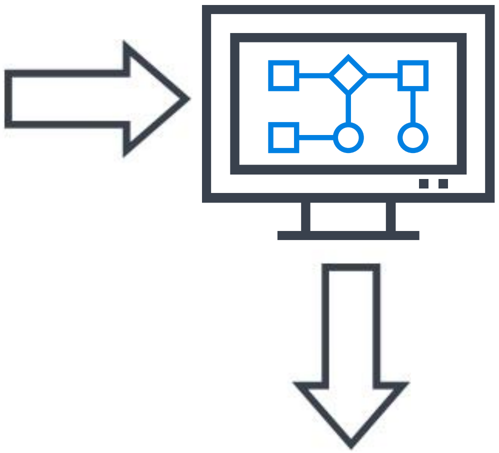
- If data requires PII you must be prepared to ensure the security of your data storage

What’s your end goal?

- Implementation
- Presentation of findings

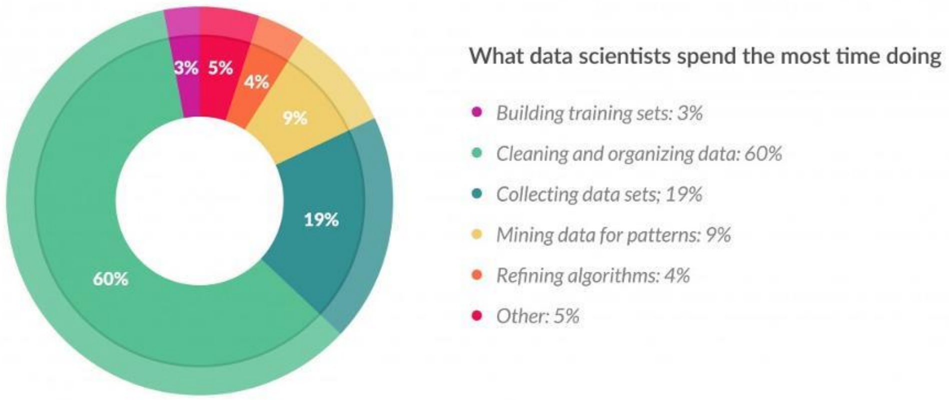


- Traditional experience analysis
- Predictive model
- Machine learning model



Data cleaning

***Data preparation** accounts for about 80% of the work of data scientists*



*Source: Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says, **Forbes***

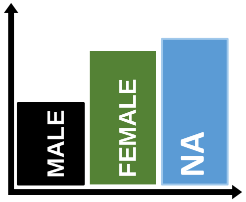
The bad news:

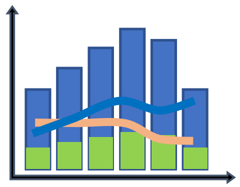
- There’s no way around it
- It’s time consuming
- If you don’t get it right your analysis will suffer

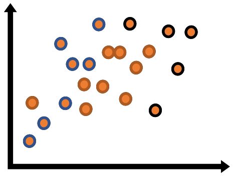
The good news:

- …
- There are many tools available to make it more approachable

Data exploration







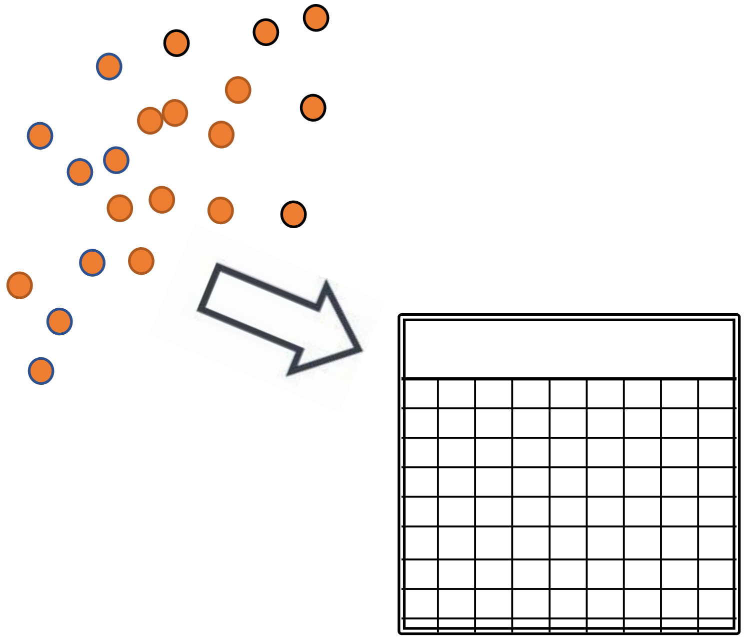
- What's in the data
- Does anything look weird?
- Handle missing values
- Filter where needed
- Check exposure and distribution of variables
  - Consider transformations where necessary

Tools for data cleaning and exploration

- Are you prepared to write your own code?
- Is a drag and drop interface more accessible?
- Do you have the computing power necessary to handle your dataset in-house?

Prep model data

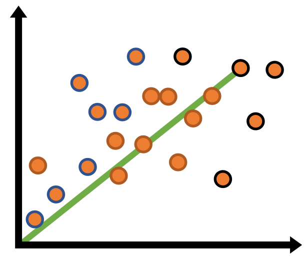
- Create target variable
- Derive additional fields
- Map data from separate tables
- Set aside holdout dataset

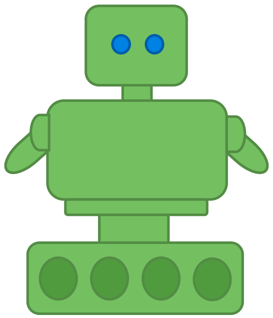


Tips we learned the hard way

- Create a data spec, that shows how to get from raw to clean data
- Establish a robust data quality assurance process
- If you’re writing your own code use version control
- Never tell yourself “oh we’ll never need that…”

Modeling





Experience analysis

Predictive modeling

- **Regression:** OLS, Generalized linear models (GLM), ridge, lasso
- **Tree-models:** Classification and regression trees (CART)
- **Survival models:** Cox proportional hazard, accelerated failure time
- **Others:** Generalized additive model (GAM)

Machine learning

- **Deep-learning:** Neural networks
- **Ensemble models:** Random forest, Gradient boosted machines (GBM)
- **Clustering:** k-means, hierarchical clustering
- …

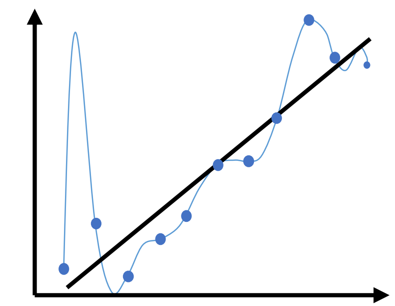
Why should I even consider machine learning?

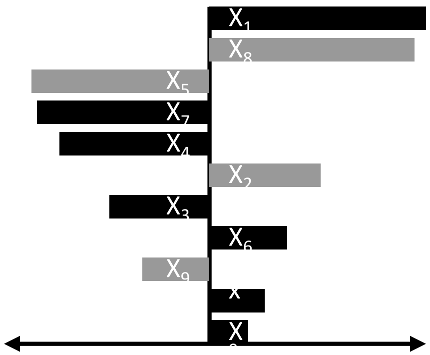
- Data continue to grow
- Powerful
- Flexible
- Computational enhancements
  - Cheaper
  - More available
- It’s sexy

Challenges of machine learning

- Limited transparency
- Implementing the resulting model
- Difficult to extend / extrapolate
- If you get it wrong……

Are we done yet?





Validation

- Actuarial judgement

Implementation

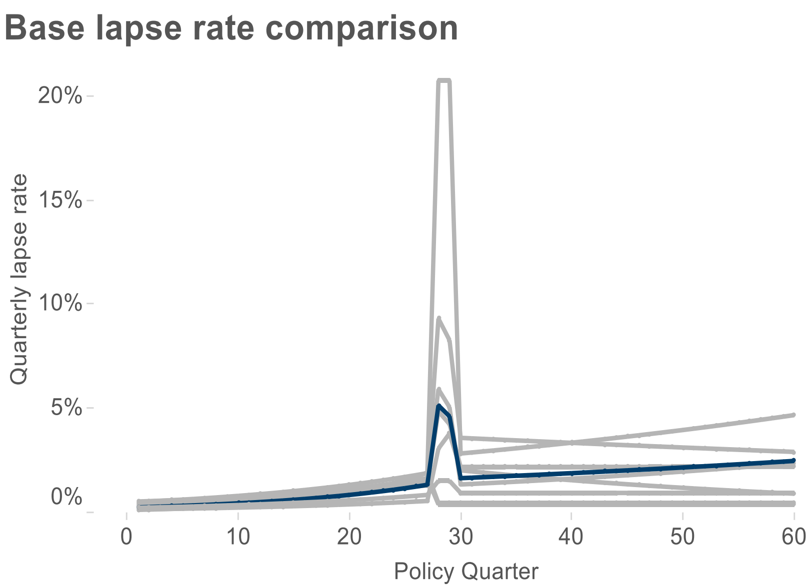
- Structure
- Extrapolation

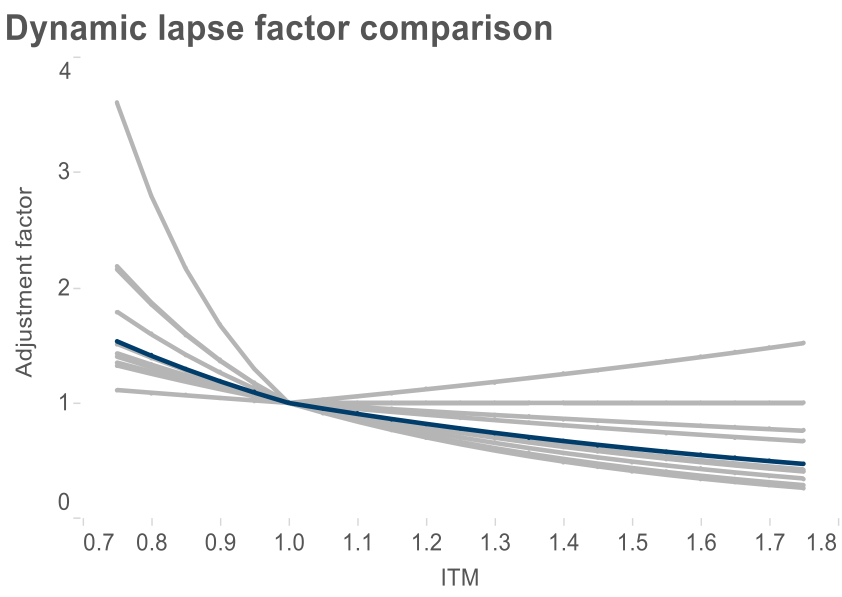
Presentation

What can this look like in practice?

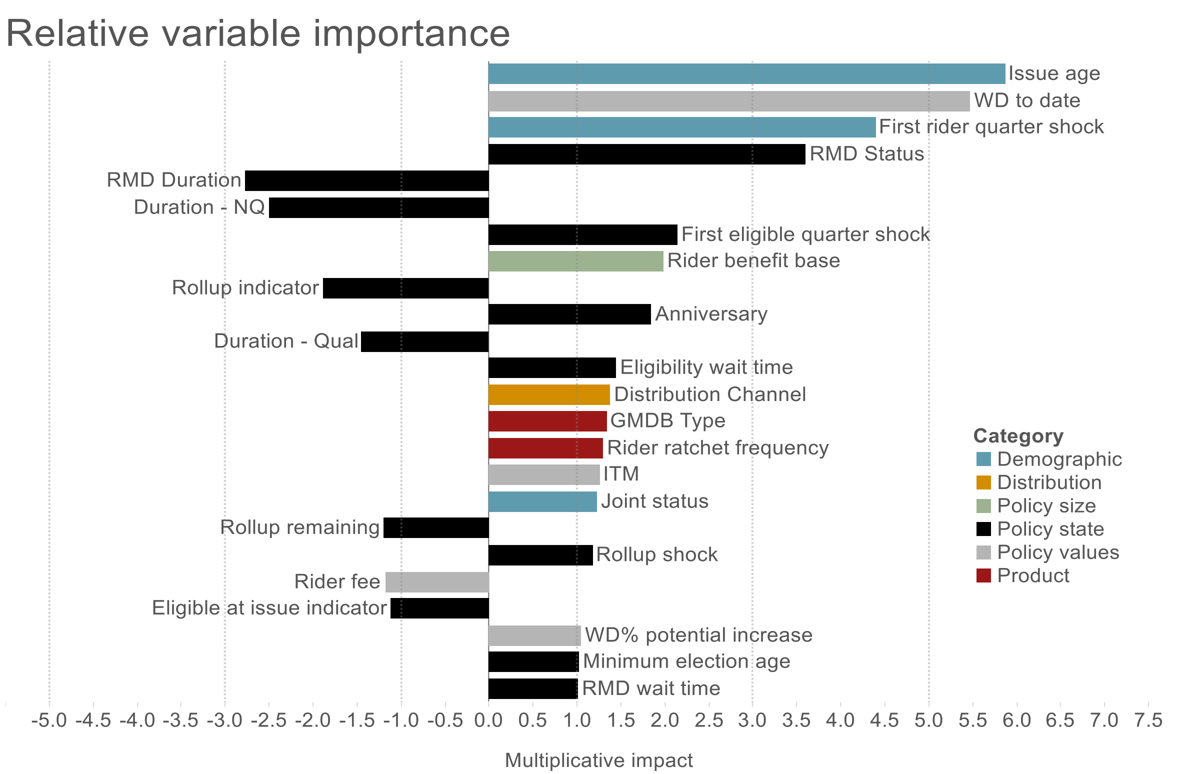
VA lapse model: industry study

Simple predictive model





VA with GLWB withdrawals: Industry study



Life Insurance Applications

Life insurance is a bit behind P&C in leveraging Predictive Analytics, but we have observed companies use Predictive Analytics for the following:

- Predictive Underwriting
- Sales/Marketing
  - Customer segmentation
  - Cross and up-selling
  - Propensity to buy
  - Lead generation
- Retention/Proactive Lapse Management
- Fraud Detection
- Distribution Management
- Assumption Setting
- Customer Value Analysis

Predictive Underwriting Case Study

Use of additional data sources to develop a less expensive and intrusive proxy for traditional medical based underwriting

Traditional Data Used
Age
Gender
BMI
MIB
MVR
Rx histories

Examples of Consumer Data Used	
Brand Name Medicine Propensity	Economic Stability Indicator Financial
Rx Online Search Propensity	Household Income
Advertised Medicine Inquirer Propensity	Home Assessed Value
Prescriptions by Mail Propensity	Home Lot Square Footage
Number of Sources	Media Channel Usage
Credit Card Indicator	Consumer Prominence Indicator
Under Banked Indicator	Consumer Hits: Exercise items
Casino Gambling Propensity	Total Consumer Hits

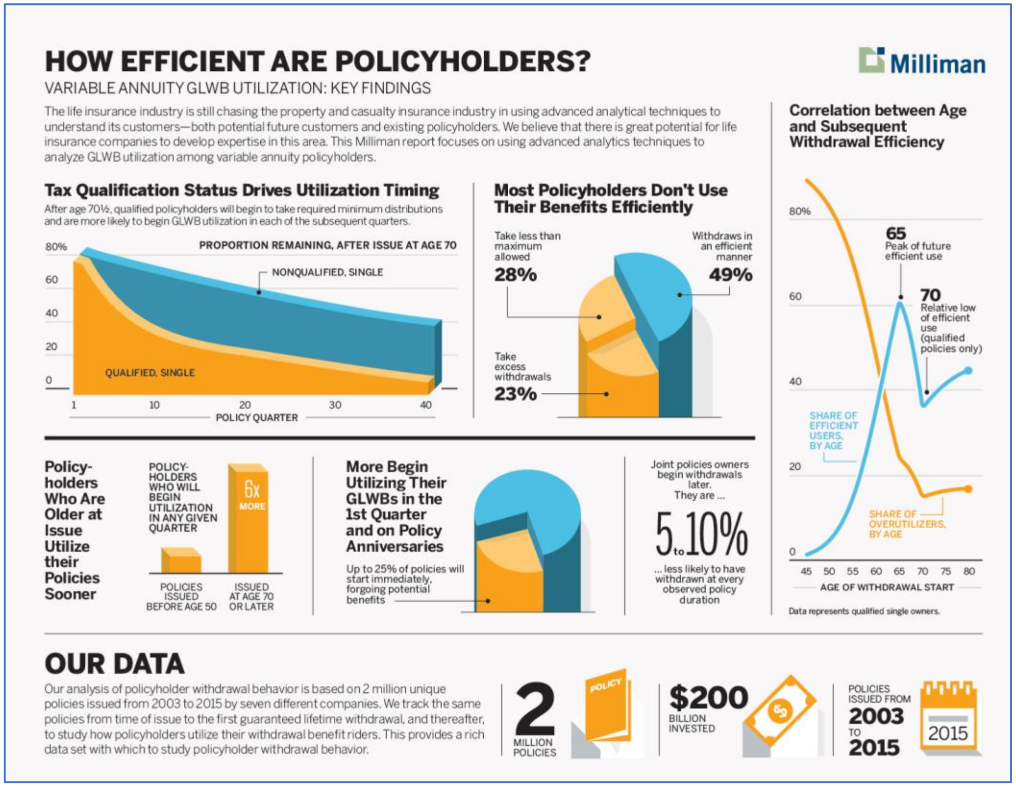
Findings

- A small percentage of policies would be issued under this program who would have otherwise been declined
- However, this should be more than offset by the underwriting savings from the policies that are rapid issued
- Traditional factors were stronger predictors for determining the best preferred class
- Consumer and financial factors were more influential in determining whether or not to decline

CARIBBEAN ACTUARIAL ASSOCIATION

Assumption Setting Example: VALUES Industry Utilization Study 2016

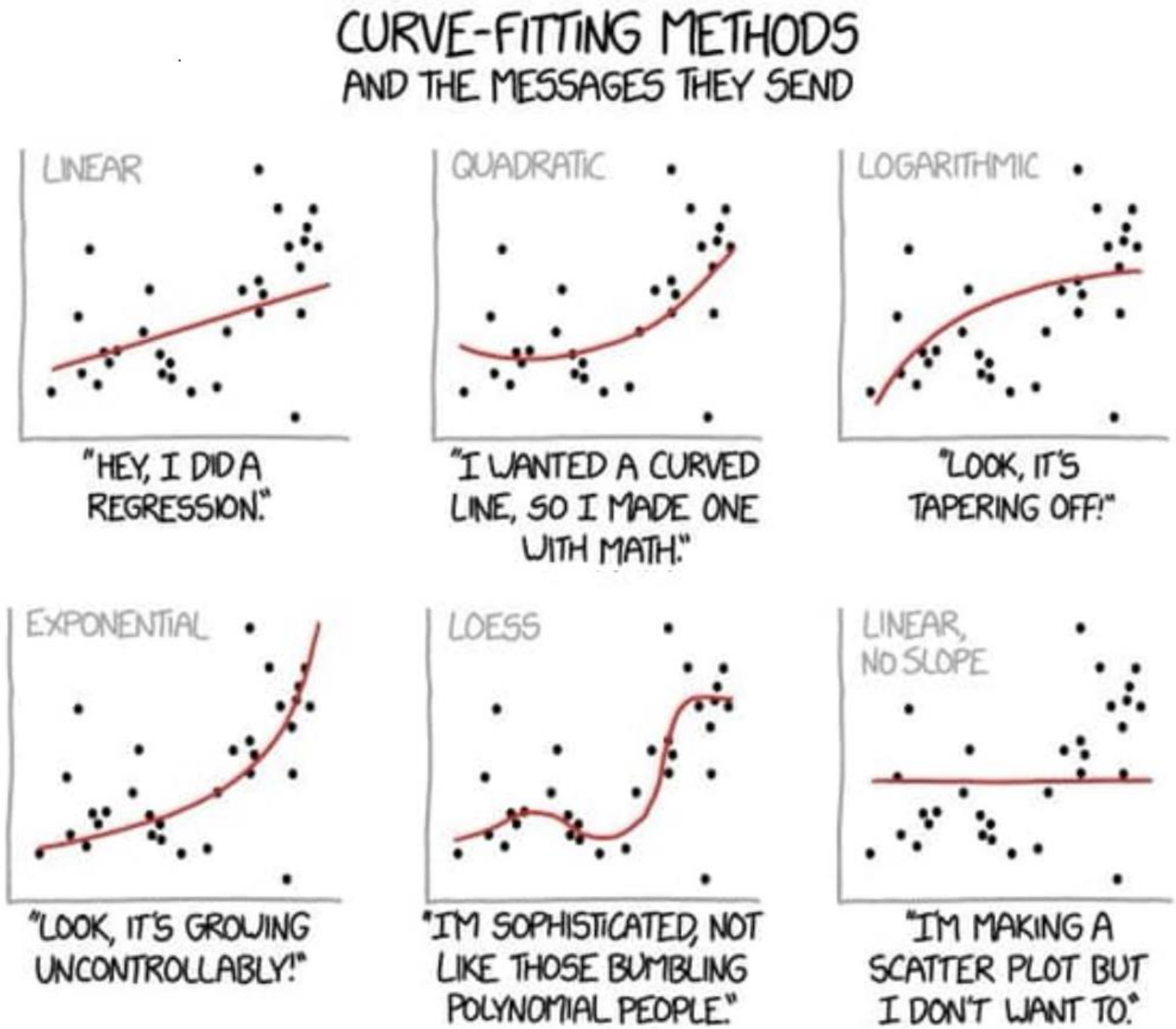
- Study covered 7 companies, 2 million policies, $200bn AV
- Studied both timing of first WB withdrawal and amount of withdrawals relative to MAWA using experience data from 2007 to 2015
- Impact of drivers and predicted behavior are analyzed by applying advanced statistical modeling.
- Study showed that policyholders who are older at issue tend to utilize their policies sooner
- Policyholders with rollup feature wait longer to utilize the GLWB.
- Less than half of all policyholders currently taking GLWB withdrawals utilize their GLWB benefit with 100% efficiency.

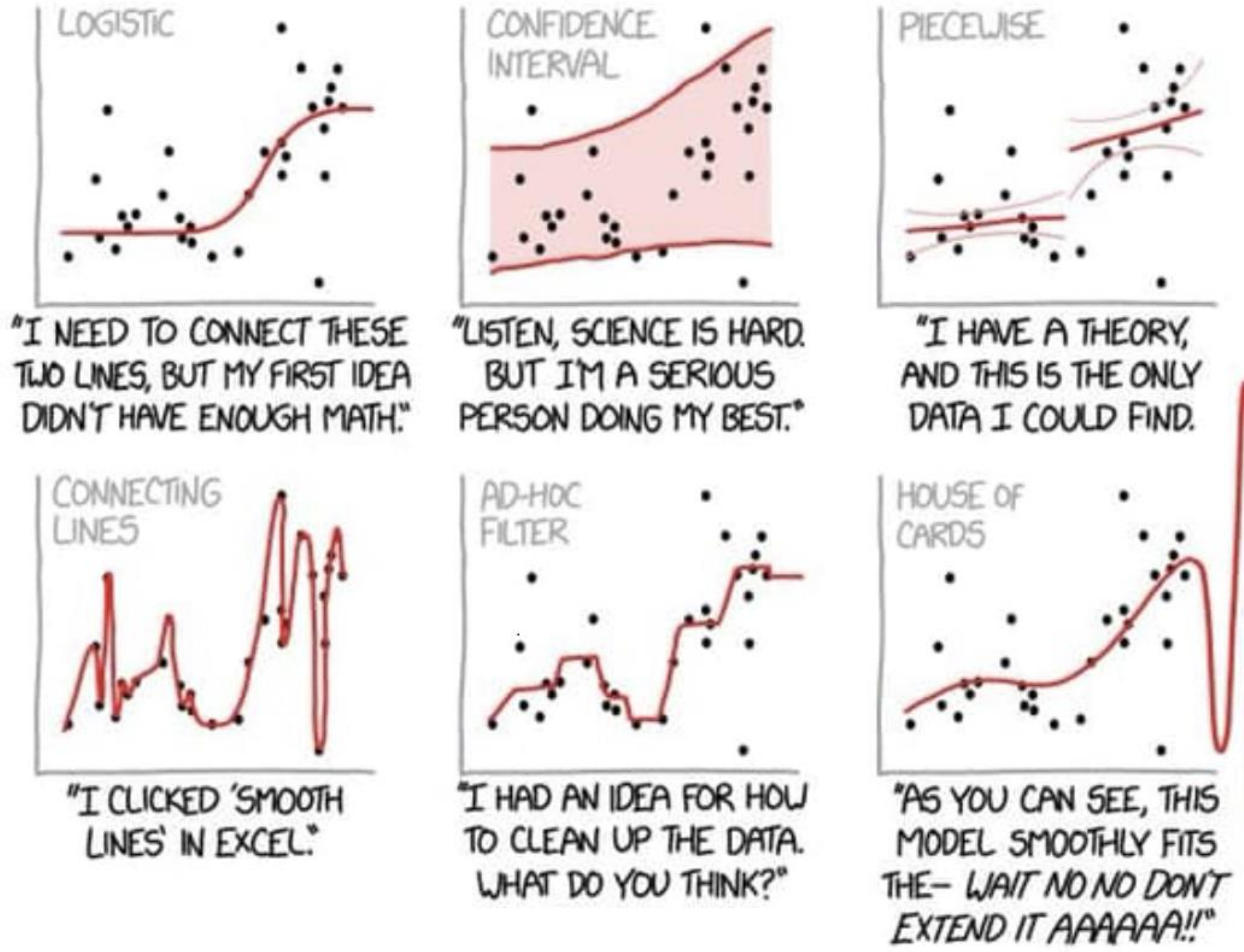


Proprietary and Confidential

Product Development Example: Vitality

- Vitality is a leading company in integrating wellness benefits in life insurance products, and has partnered to launch life insurance products in different countries
- John Hancock has launched a UL product in partnership with Vitality
- Customers accumulate points and rewards for maintaining a healthy lifestyle (diet, exercise, health screenings)
- Points status is used to determine discounts for each year's premium
- The product proposition is empowered by Predictive Analytics and new data
  - Steady stream of data is captured from customer
  - Historical dataset used to analyze impact of various lifestyle indicators on mortality rates
  - Presented as a win-win proposition to customer
  - Data from customer can be used for other purposes (cross-sell/up-sell)


CURVE-FITTING METHODS
AND THE MESSAGES THEY SEND
LINEAR
"HEY, I DID A REGRESSION."
QUADRATIC
"I WANTED A CURVED LINE, SO I MADE ONE WITH MATH."
LOGARITHMIC
"LOOK, IT'S TAPERING OFF!"
EXPONENTIAL
"LOOK, IT'S GROWING UNCONTROLLABLY!"
LOESS
"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE."
LINEAR, NO SLOPE
"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO."


LOGISTIC
"I NEED TO CONNECT THESE TWO LINES, BUT MY FIRST IDEA DIDN'T HAVE ENOUGH MATH."
CONFIDENCE INTERVAL
"LISTEN, SCIENCE IS HARD. BUT I'M A SERIOUS PERSON DOING MY BEST."
PIECEWISE
"I HAVE A THEORY, AND THIS IS THE ONLY DATA I COULD FIND.
CONNECTING LINES
"I CLICKED 'SMOOTH LINES' IN EXCEL."
AD-HOC FILTER
"I HAD AN IDEA FOR HOW TO CLEAN UP THE DATA. WHAT DO YOU THINK?"
HOUSE OF CARDS
"AS YOU CAN SEE, THIS MODEL SMOOTHLY FITS THE— *WAIT NO NO DON'T EXTEND IT AAAAAA!!*"

Questions?

Full Member of

International Actuarial Association
Association Actuarielle Internationale