Munich RE

Actuaries and Data Scientists – Powerful Partnerships and Success Stories

Julia Druce, Manager, Integrated Analytics

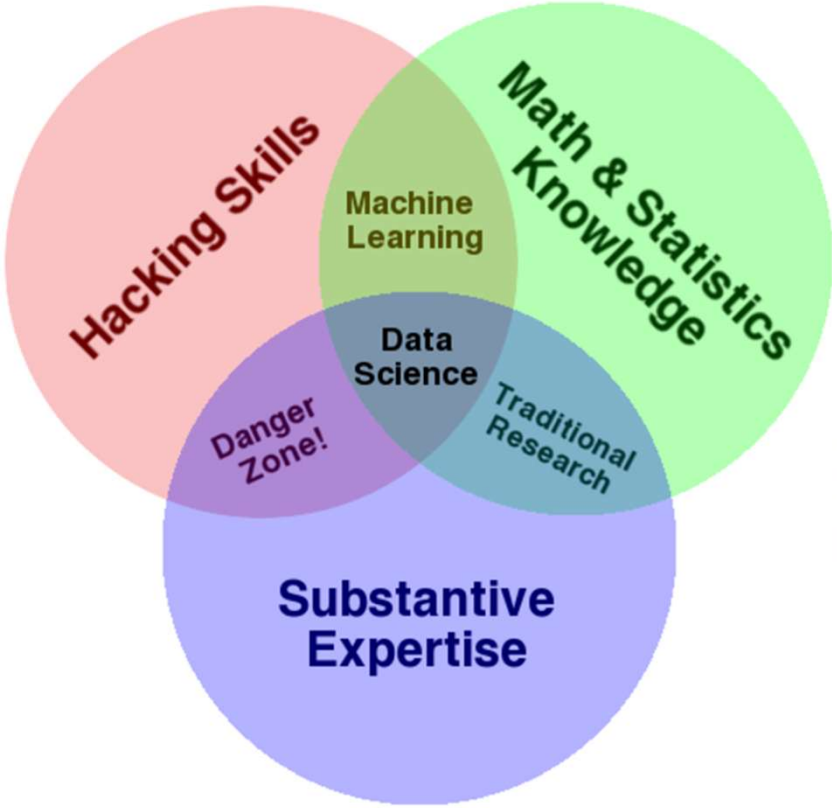
John Myslinski ASA, Sr Data Scientist, Integrated Analytics

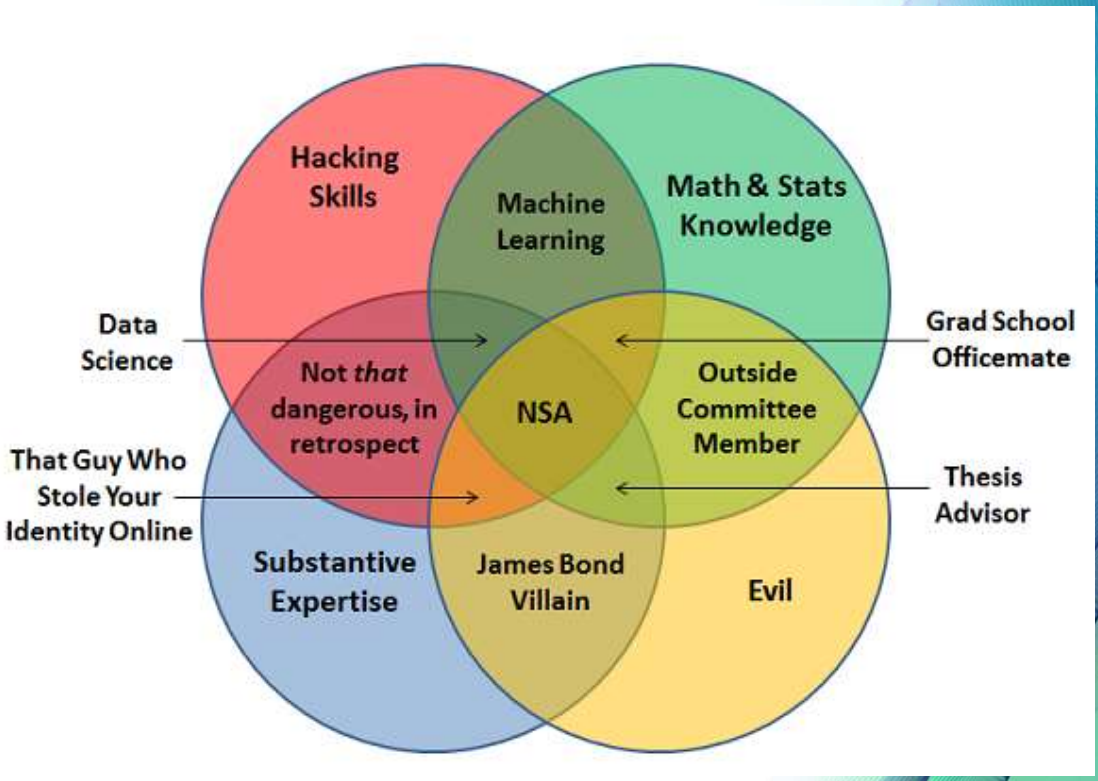
What's the difference?

	Actuarial Science	Data Science
Definition	"Relating to calculations of risk for insurance companies and pension funds."	"The use of scientific methods to obtain useful information from computer data."
Education	mathematics, statistics, finance, economics, or other quantitative subjects	mathematics, statistics, computer science, data science (newer MS degrees)
Credentialing	CAS and/or SOA	None required

Definitions taken from the Cambridge English Dictionary

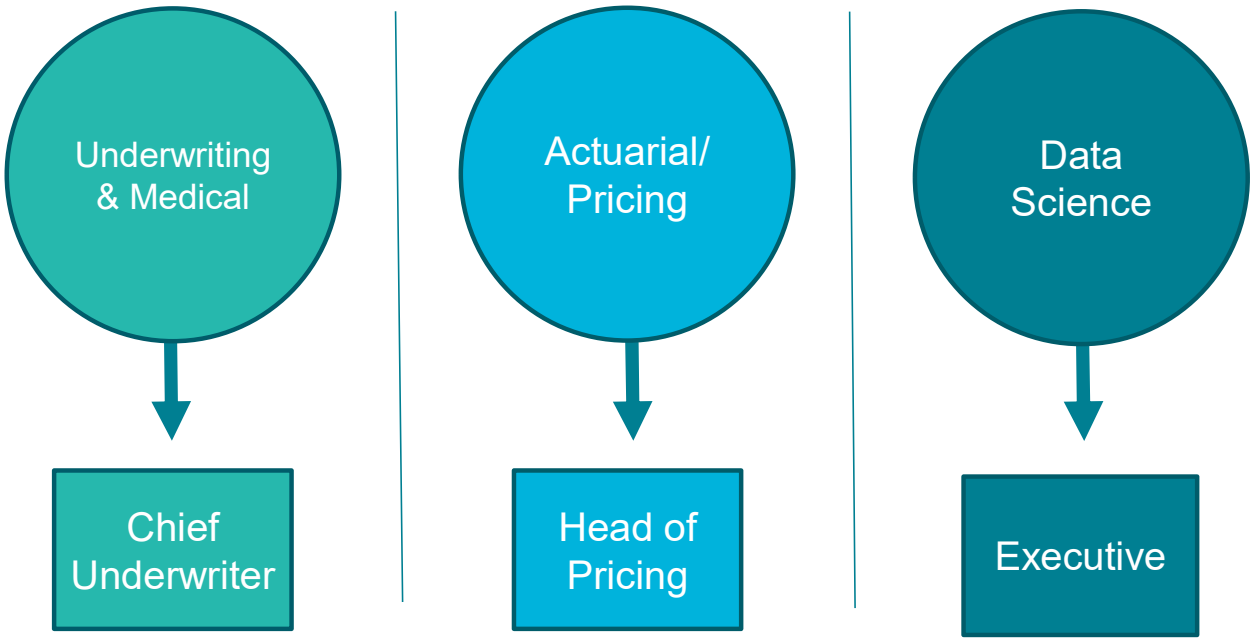
Data Science Skillsets





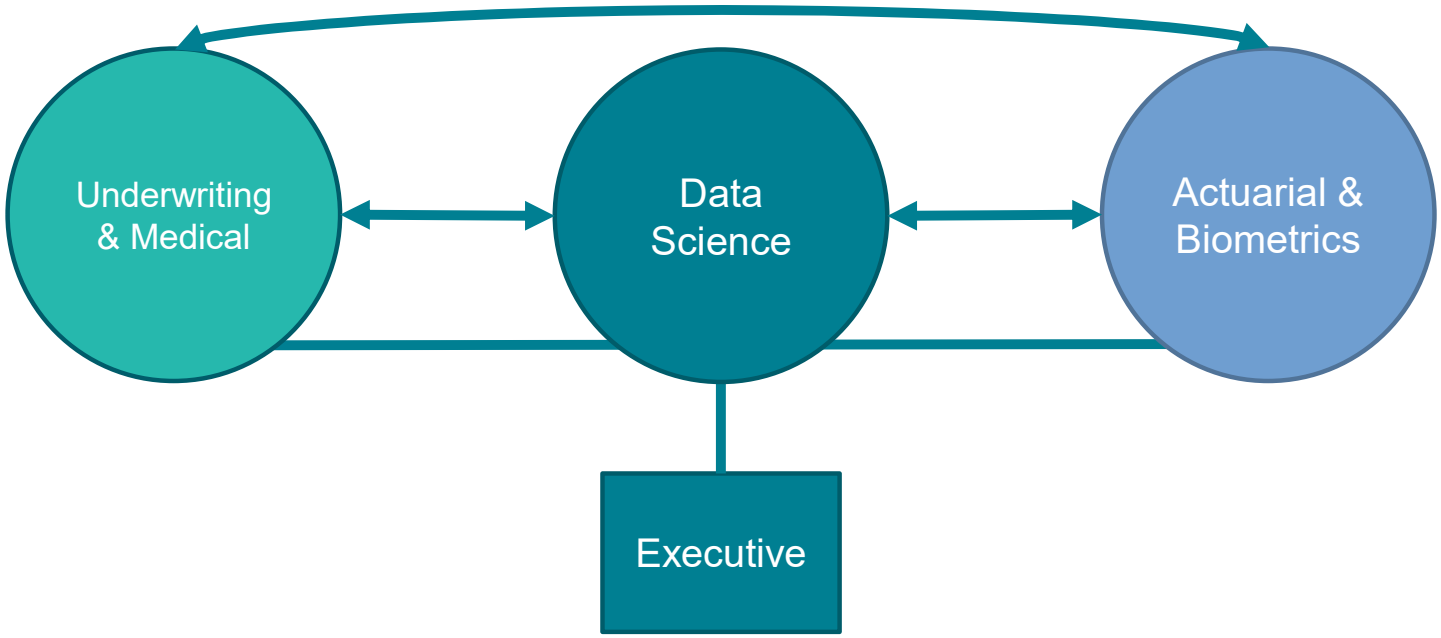
How do actuaries and data scientists partner in a life insurance company?

Early Organizational Structures with a Data Strategy



- Data science strategy separate from the business units

Underwriting, Actuaries & Data Science as Strategic Partners



- Each practice area has autonomy to work directly with strategic partners to find solutions, with shared executive oversight.

Data Science Organization Structure

Standalone

Data scientists exist as a standalone team.

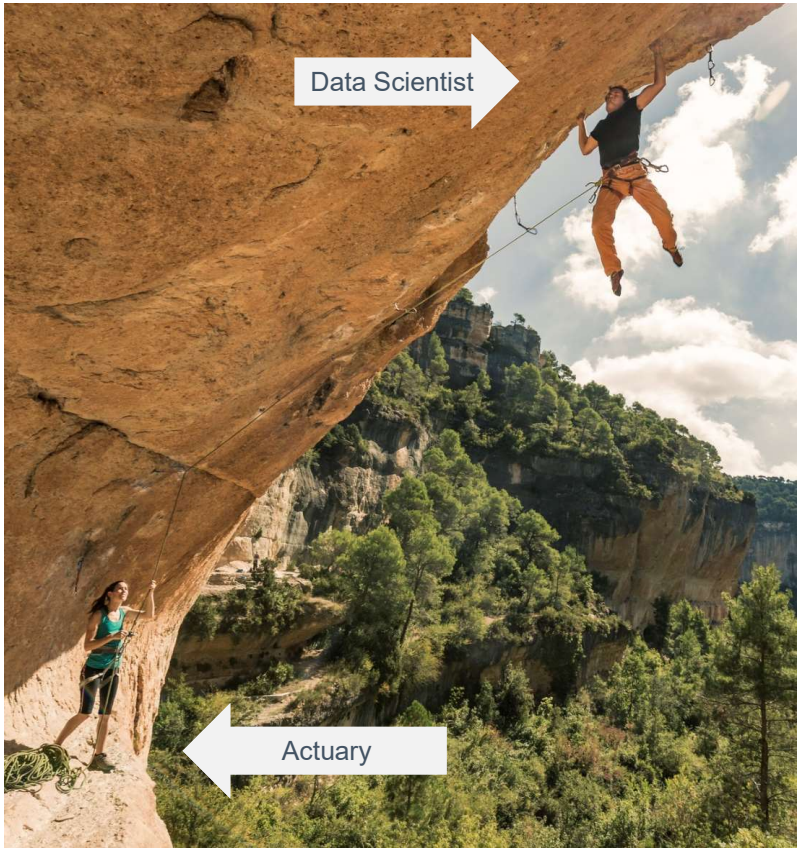
- Creates knowledge sharing and sense of community with data science/engineering.
- Generally, more autonomy for developing own tech stack.
- Can be difficult to integrate data science throughout the organization.

Integrated

Data science team exist within business units.

- Easy cross-pollination of skills and business knowledge
- Non-standard definition of data science role (may be more difficult to recruit)
- Challenges in managing career paths

Collaboration



Munich RE

Speaking the Same Language

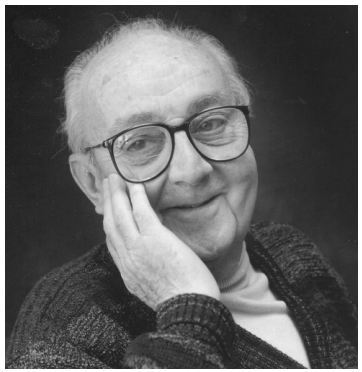
Data scientist says: “Our model **target** was lapse (yes/no).”

Actuary says: “We predicted lapse rates.”

Data scientist says: “4 **features** were important.”

Actuary says: “4 variables were found to differentiate risk.”

Common Ground



"All models are wrong,
but some are useful."

George E. P. Box
Statistician

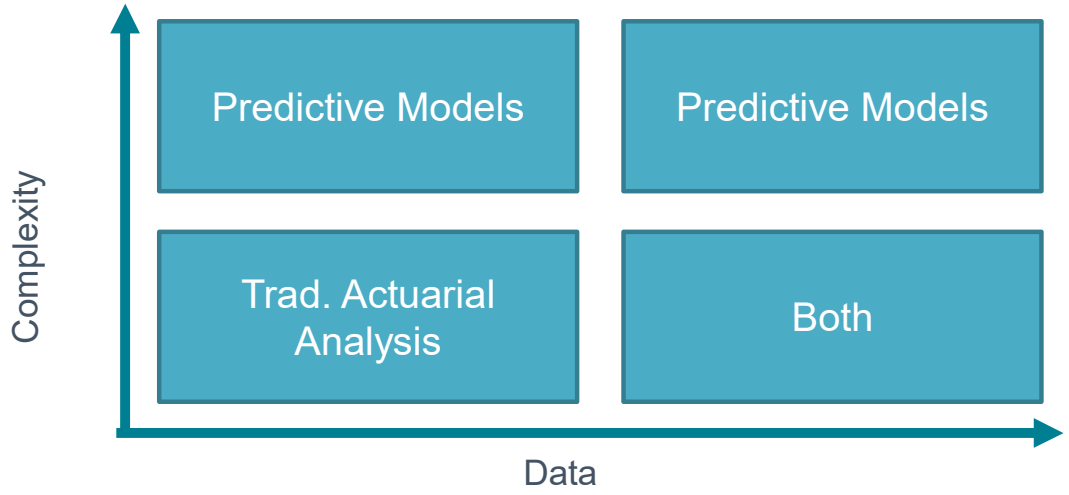
Traditional Actuarial Analysis vs. Predictive Models

Traditional Actuarial Analysis

- Works well when relationships in data are well understood and not overly complex.
- Inefficient for understanding large, new, or complex datasets.

Predictive Models

- Automatically discovers rules and relationships within data.
- With the appropriate amount and type of data, should outperform traditional actuarial analysis.

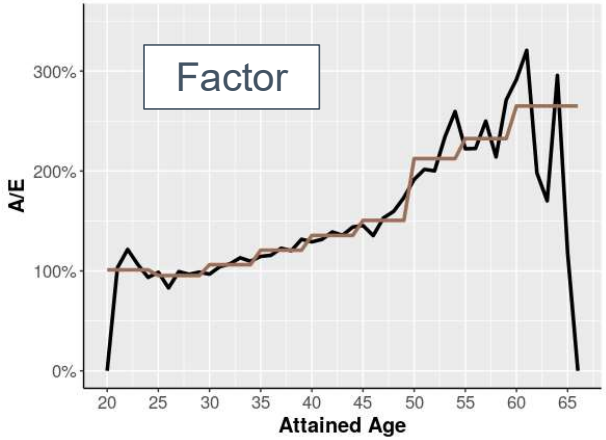


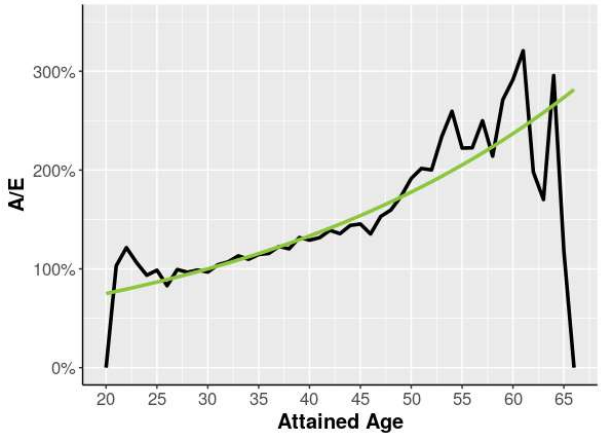
Success Stories from Life Insurance

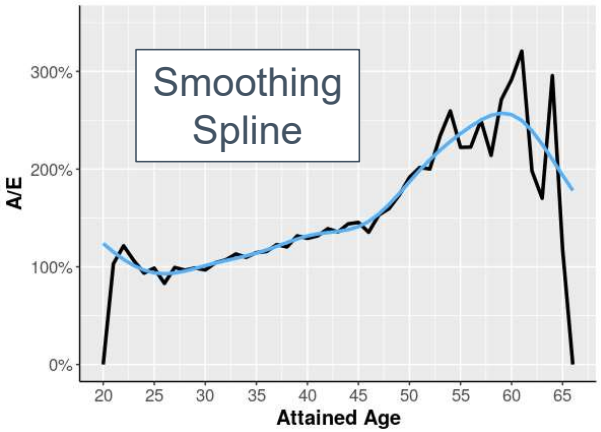
Problems and How We Solve Them

Predict mortality rates & incorporate new predictors: geodemographic, financial, behavioral	• Survival models - Cox proportional hazard, exponential (parametric) • Other regression models – GLM, GAM, survival trees
Summarize unstructured medical data	• Natural Language Processing
Predict applicant misrepresentation (smoker status)	• Classification models – logistic regression, support vector machines, boosted / bagged decision trees, etc.

Predict Mortality Rates - Mortality Modeling using GAMs



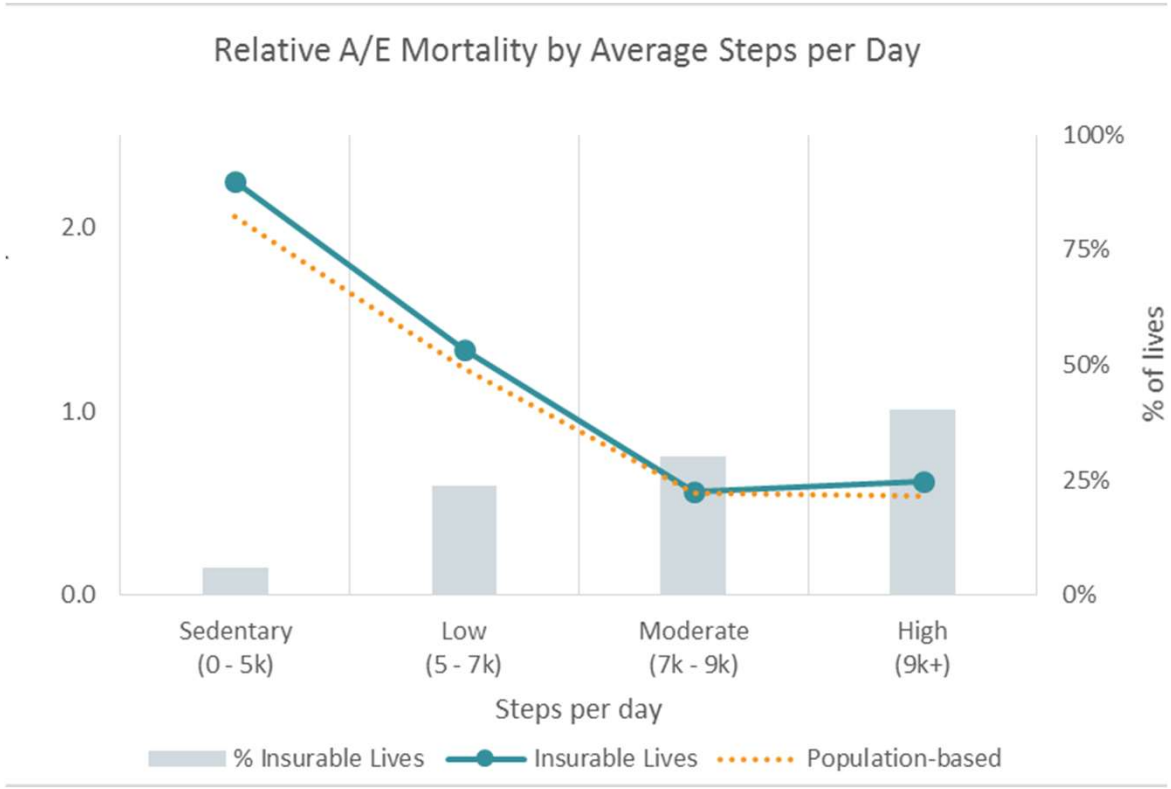




1. $\text{Claims} \sim \beta * \text{AttainedAge Bucket}$
2. $\text{Claims} \sim \beta * \text{AttainedAge}$
3. $\text{Claims} \sim f(\text{Attained Age})$

- Generalized Additive Models (GAMs) extend GLMs by incorporating smoothing splines into our model specification.
- GAM allows us much greater modeling flexibility

Predict Mortality Rates - Physical Activity and Mortality



Model

- Cox PH

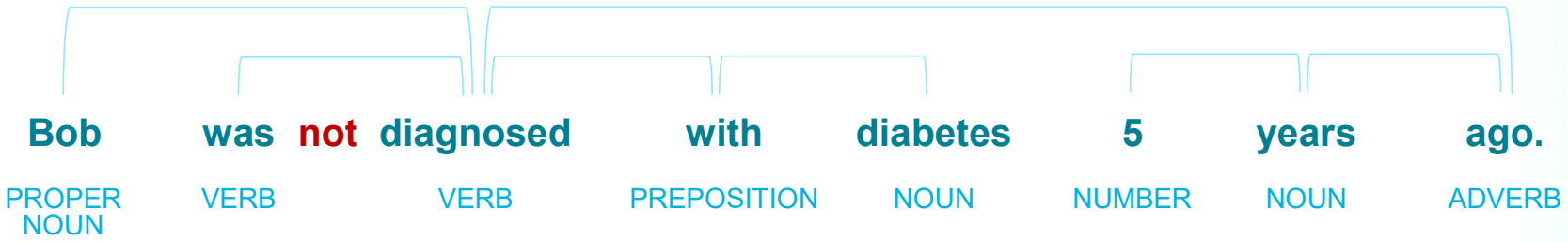
Results

Daily step count stratifies mortality risk

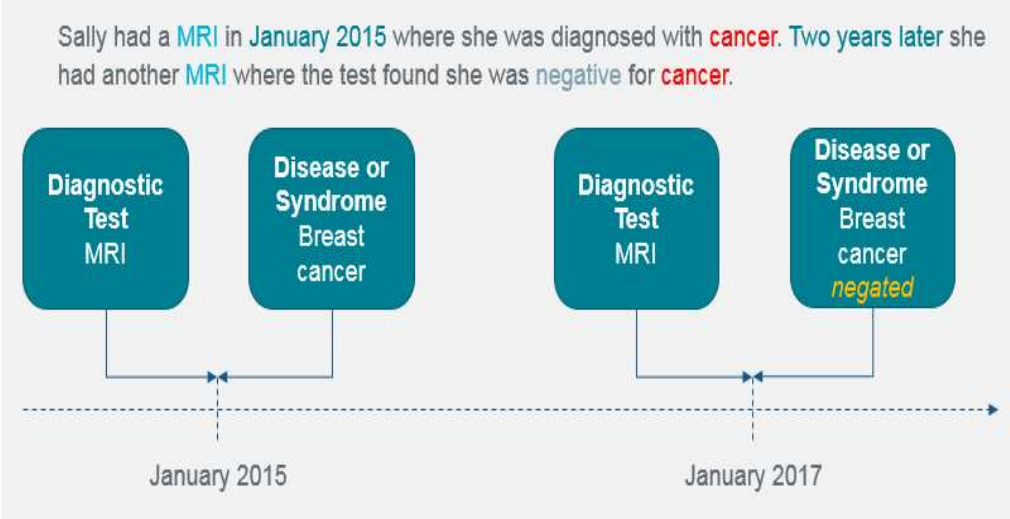
- People with sedentary/low steps per day have a higher mortality risk, while those with moderate/high steps per day have lower mortality risk.
- People with sedentary behavior have 3x relative mortality of active individuals.

Summarize Unstructured Medical Data - Natural Language Processing

Speech Tagging and Dependency Parsing

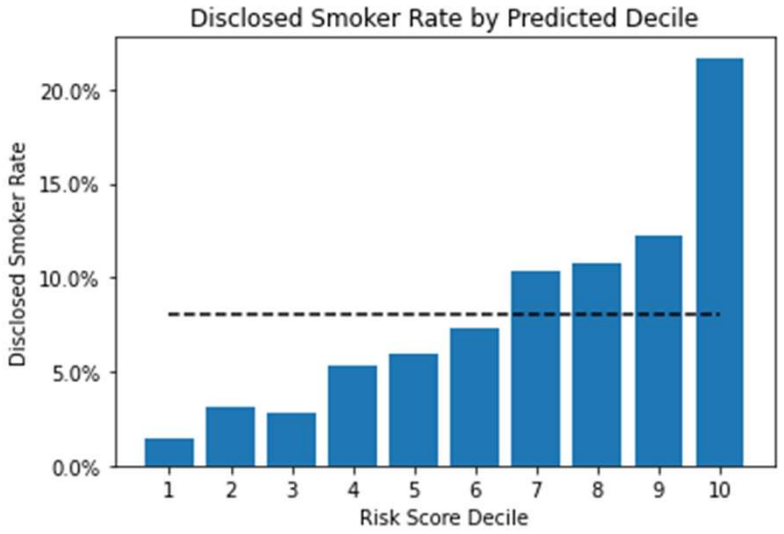


Summarize Unstructured Medical Data - Timeline creation



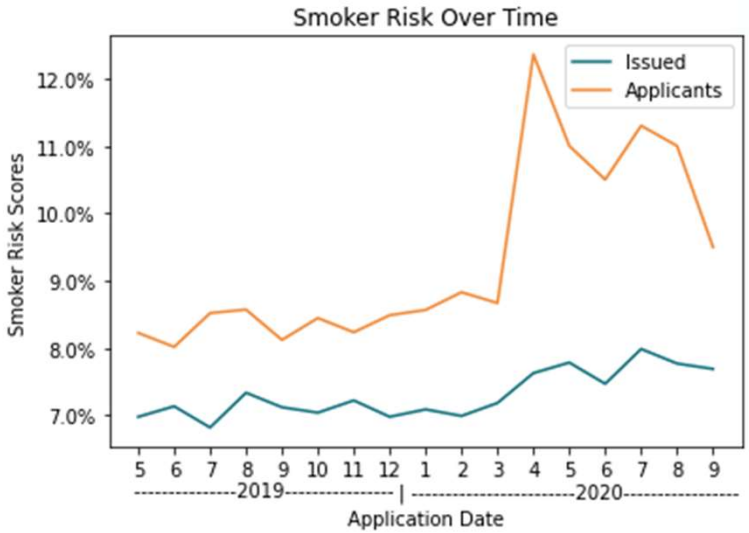
- Extract all dates within an Attending Physician Statement (APS) and associate them with medical terms
- Construct a timeline of a person's medical history

Predict Applicant Misrepresentation - Smoker Status



A machine learning model was trained to predict if a life insurance applicant was likely to be a smoker (classification model).

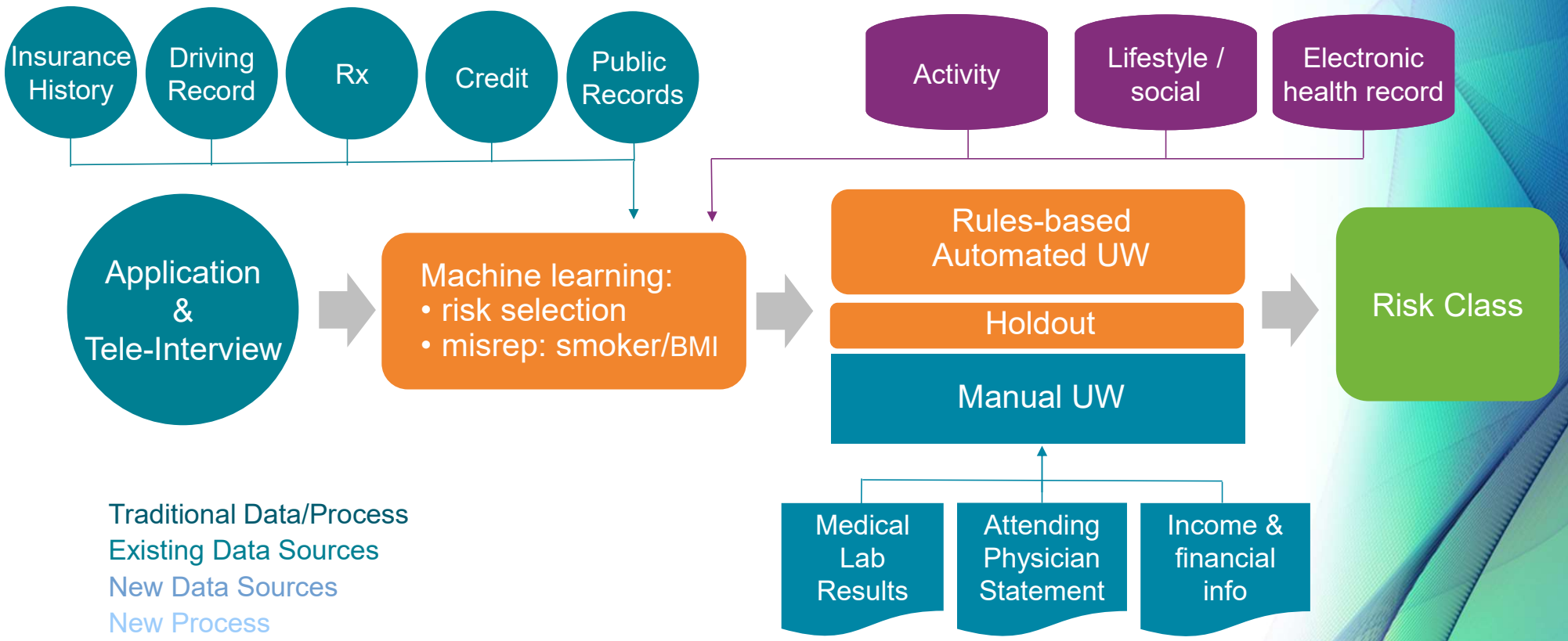
The model segments the population and can be used to flag applicants who are most likely to smoke and require additional screening.



Monitoring is a crucial part of successful model deployment.

Here, we see a concerning divergence between the applicant and issued populations.

Putting it all together in AUW



Conclusions

- Actuaries and data scientists complement each other, and both want to improve understanding and management of risk.
- Currently, data scientists are generally more comfortable with manipulating large datasets, implementing new model types, or integrating new data sources than the average actuary, but there are always exceptions.
- The difference in roles and skills may shrink as more actuaries learn more programming and data science through school/exams and as data scientists gets embedded in actuarial departments.
- Practical recommendation for actuaries – get programming experience
  - Automate repetitive Excel tasks using VBA
  - Recode an Excel spreadsheet in R/Python – compare cell output to code output
  - Pre-process large datasets outside of Excel (you may already know SQL, perfect)
  - Train your first GLM in R or Python
- Practical recommendation for data scientists – learn more about the business
  - Attend industry presentations or internal lunch and learns
  - Learn the language of the business
  - Check your loss metrics – are you using a business relevant measure of model performance?

How it started, how it's going