Deloitte.

FEBRUARY 2019, CVUT

HR Analytics

Optimization of HR Processes Using Data and Analytics

Speaker’s Introduction

Filip Trojan

Advanced Analytics

Deloitte Advisory s.r.o.

Professional data scientist with broad experience across multiple industries including banking, finance, telco, manufacturing, retail, FMCG and e-commerce.

Strong math and statistical background and always looking for application of advanced algorithms into business problems.

Skills: marketing, scorecard development, consumer lending, credit policy, bad debt provisions, supply chain, project management, six sigma, DSS, Python, R, Matlab, SAS, XML, Perl, LaTeX, AS400, Oracle, Teradata, MSSQL

Events and ways to reach out to us.

Challenge 2019

We will soon be launching our annual competition. All **students and recent graduates** are welcome to apply. You will have a chance to solve a real **business case**, benefit from **workshops**, work with data from one of our partners and **win a cash prize**. For more details follow us on Facebook.

Advanced Analytics by Deloitte

FACEBOOK

www.facebook.com/advancedanalyticsCZ

EMAIL

ceczanalytics@deloittece.com

We look forward to hearing from you.

Presentation Content

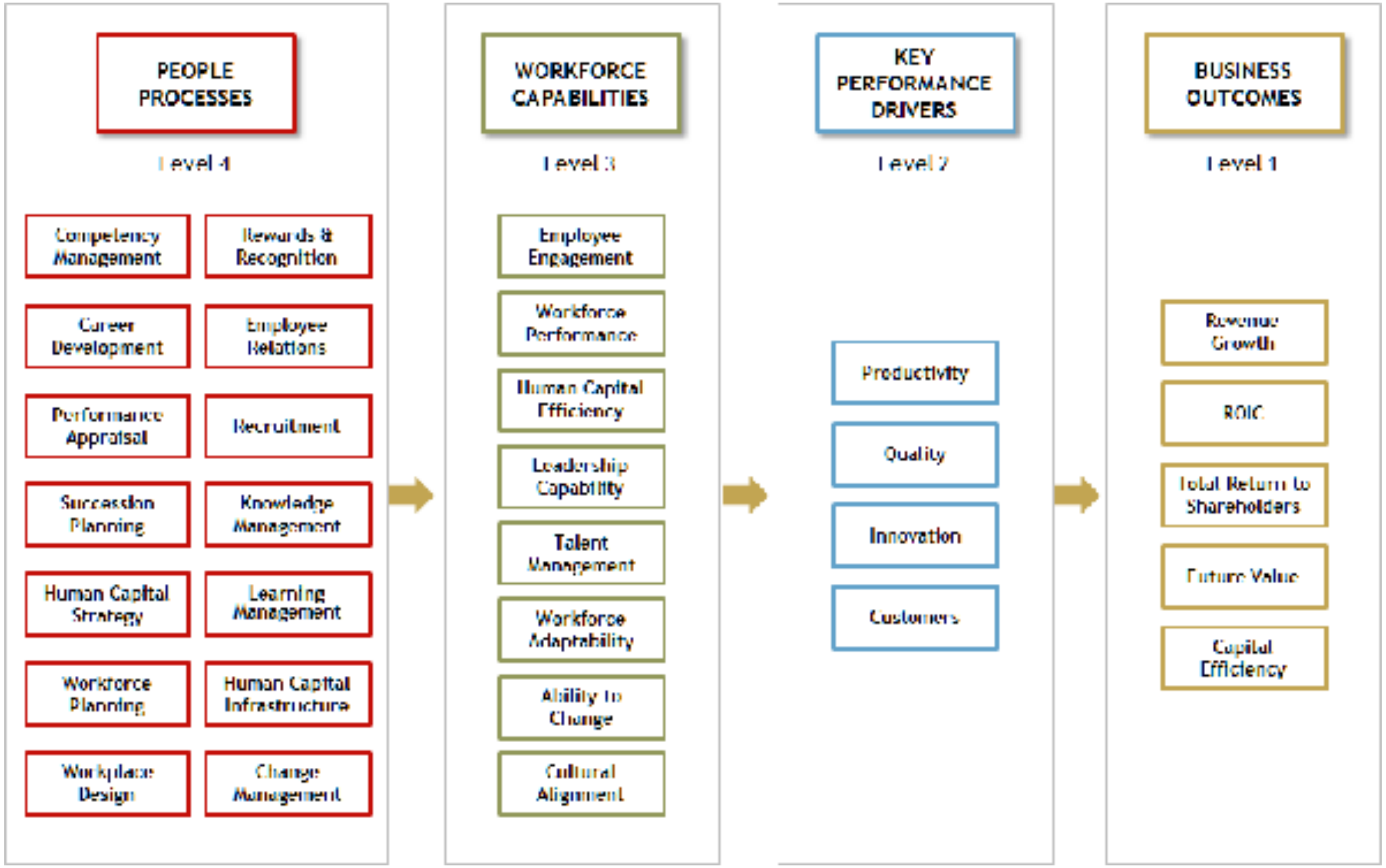
HRM’s Added Value

- **Employees can be considered an organization’s most valuable asset** – only through employees’ knowledge, skills, and abilities company can achieve its business and strategic goals (Boselie, 2014; Paauwe & Farndale, 2017).
- **Effective and/or efficient people management** (the way companies hire, deploy, develop, motivate, and retain its employees) **is thus a must** (Barney, 2001; Baron & Armstrong, 2007; Huselid & Becker, 2011; Wright et al., 1994).
- **Positive impact of the HRM function and its policies and practices** (sophisticated selection and training practices, participation programs, formal performance appraisals, contingent pay schemes among others) **on the operational and financial performance of organizations** has been supported by several studies with both cross-sectional and longitudinal research design (see Combs, Liu, Hall, & Ketchen, 2006; Crook, Todd, Combs, Woehr, & Ketchen, 2011; Huselid, 1995; Jiang, Lepak, Hu, & Baer, 2012; Subramony, 2009).

Source: van der Laken (2018)

HRM Value Chain

Model of mechanism by which HRM practices and policies do have impact on companies' financial and operational performance.

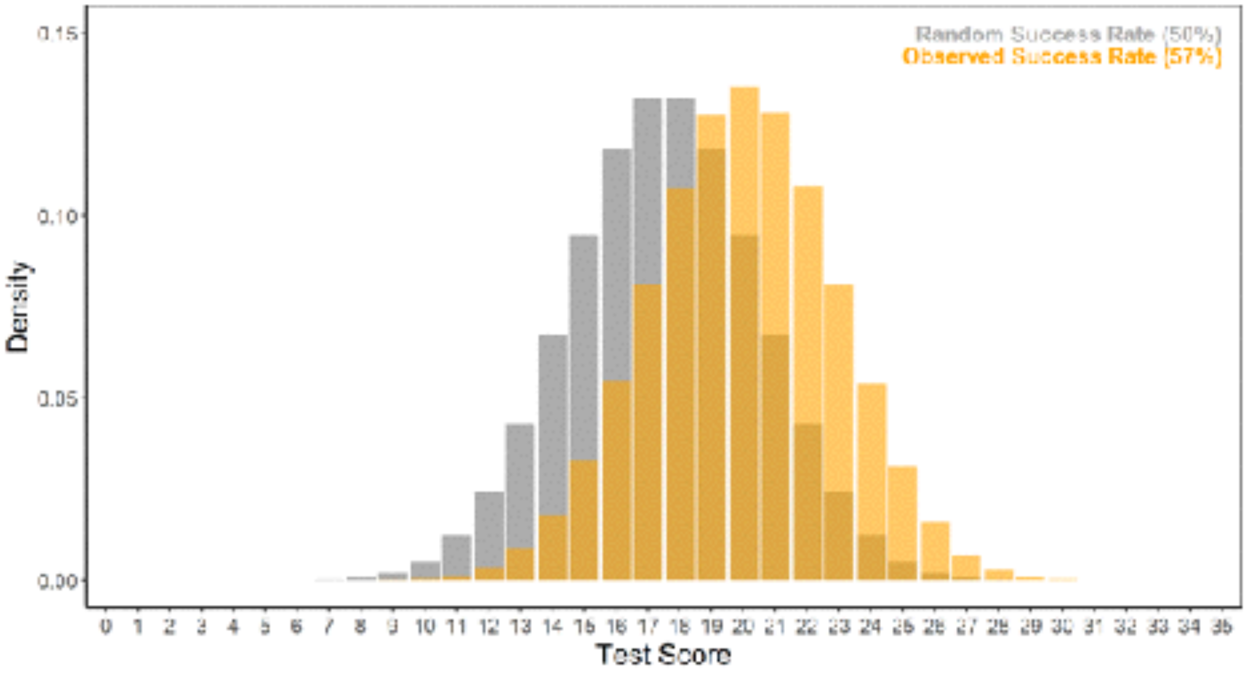


Source: Blumberg (2018), Cantrell et al. (2016)

Misleading Intuition & Believes in HRM

Unfortunately, HRM professionals' decisions are too often based on intuition, experience, and beliefs that are under influence of various fads and hypes.

Items	*Answer–% Correct (% uncertain)*	*Research Evidence*
Management Practices		
1. Leadership training is ineffective because good leaders are born, not made.	False 96% (2%)	Field study evidence that leadership behaviors and effectiveness increase following training (Barling et al., 1996). Evidence that leadership behaviors are only weakly predicted by dispositional characteristics (Judge & Bono, 2000) that are heritable (Loehlin et al., 1998; Reimann et al., 1997).
2. The most important requirement for an effective leader is to have an outgoing, enthusiastic personality.	False 82% (4.5%)	This kind of personality is, on average, an asset for leadership. A recent meta-analysis estimates a corrected validity coefficient of .31 between extraversion and leader effectiveness (Judge et al., in press). However, intelligence has an even higher correlation (.52; Lord et al., 1986). Also, some highly effective leaders are distinctly introverted (Bennis & Nanus, 1997; Collins, 2001).



Source: Rynes, Colbert, & Brown (2002)

Misleading Intuition & Believes in HRM

Test yourself!

Management Practices

On average, encouraging employees to participate in decision making is more effective for improving organizational performance than setting performance goals.

True ☐ False ☑

General Employment Practices

Most people overevaluate how well they perform on the job.

True ☑ False ☐

Training & Employee Development

Training for simple skills will be more effective if it is presented in one concentrated session than if it is presented in several sessions over time.

True ☐ False ☑

Staffing

On average, applicants who answer job advertisements are likely to have higher turnover than those referred by other employees.

True ☑ False ☐

Compensation & Benefits

Talking about salary issues during performance appraisals tends to hurt morale and future performance.

True ☐ False ☑

Source: Rynes, Colbert, & Brown (2002)

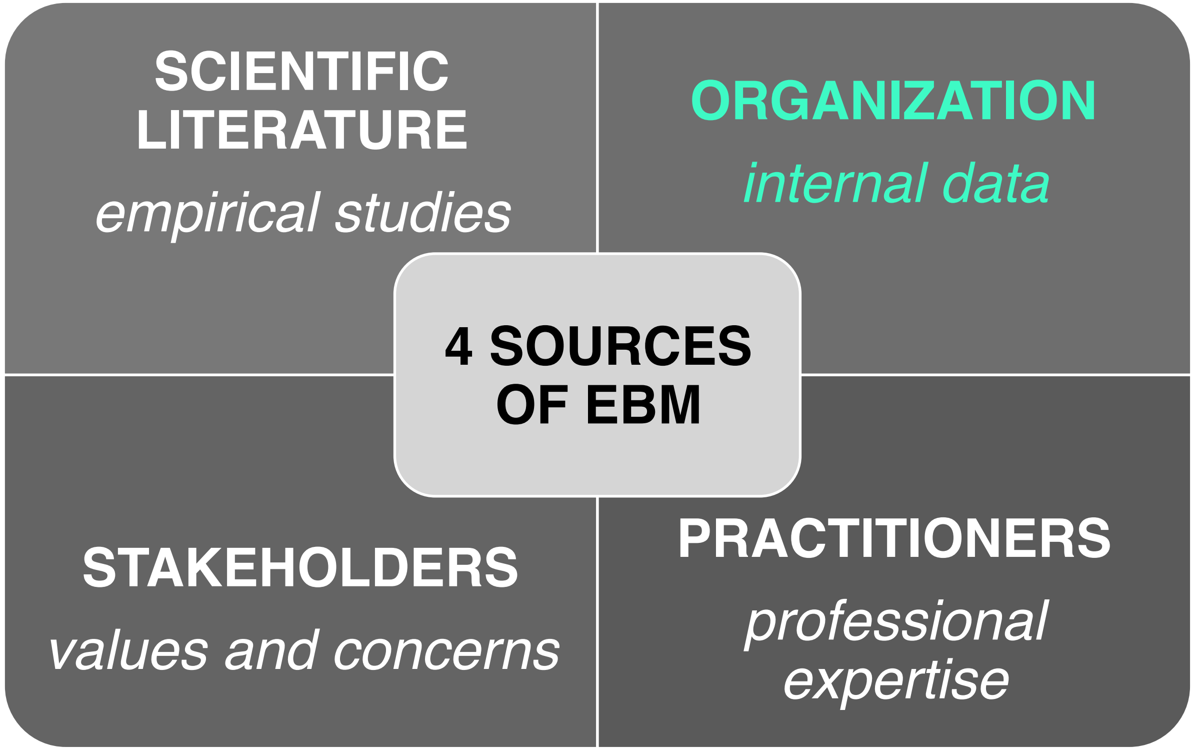
Evidence-Based HRM

- Even though **HR professionals** base their decisions on some sort of evidence (outcome of scientific research, organizational facts & data, benchmarking, best practices, collective experience, personal experience, intuition…), many of them **pay little attention to the quality of the evidence** they base their decisions on.
- Evidence-based practice is about **making decisions through, the conscientious, explicit and judicious use of the best available evidence from multiple sources**

Source: https://scienceforwork.com/blog/evidence-based-management-training/

4 Sources of Evidence-Based HRM

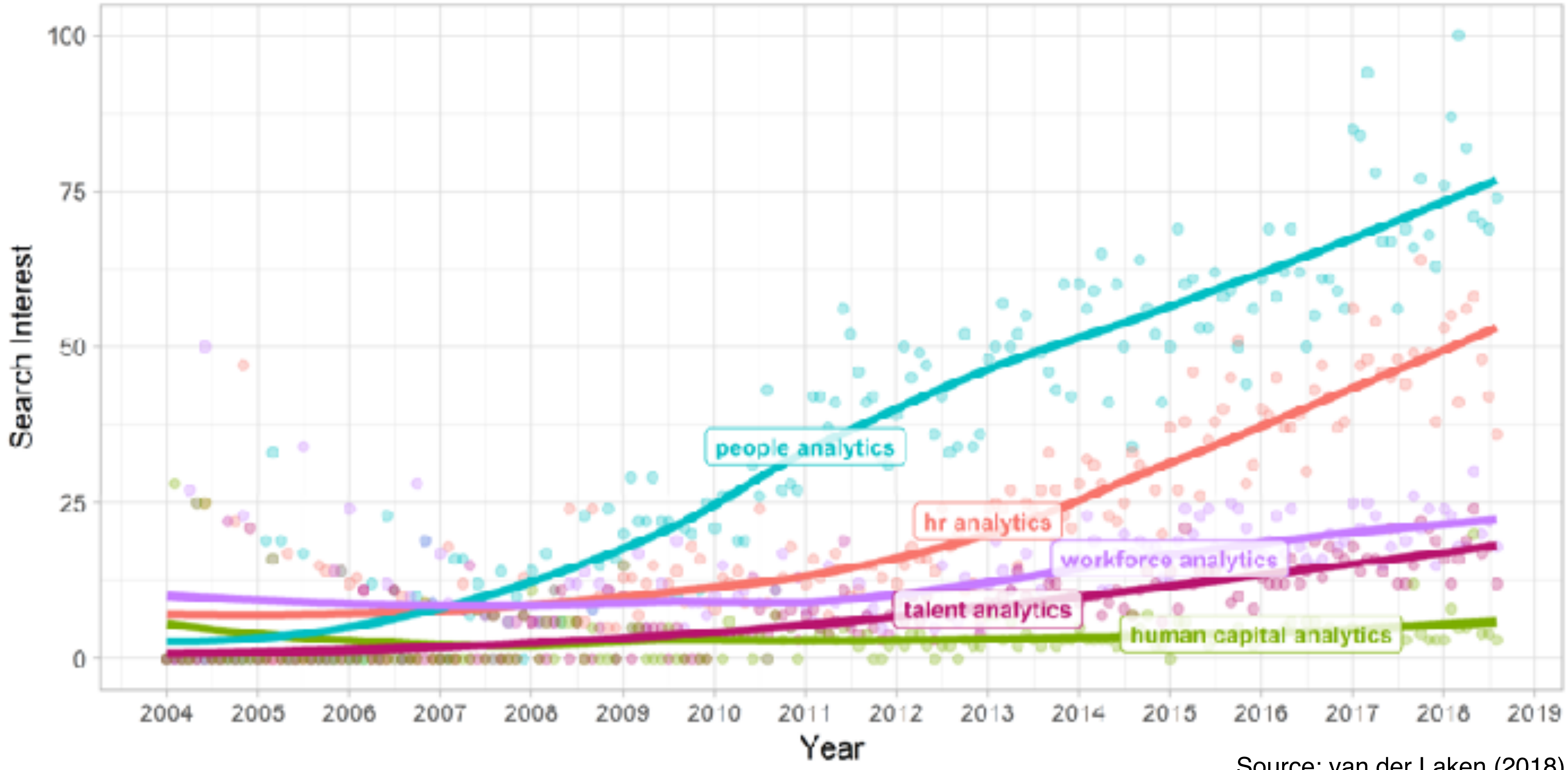
- Evidence-based HRM combines **4 sources of information:**
    1) practitioners' professional expertise,
    2) stakeholders' values and concerns,
    3) scientific evidence, and
    4) reliable and valid organizational metrics.
- The biggest blind spot of the HRM function usually lies in the fourth source of information: it **often lacks lack the capability to measure and quantify the strategic contribution of its HRM activities, its bottom-line impact, and any progress therein in its own, local organizational context…**
- … due to **missing analytical mindset among** HR professionals and **absence of reliable and relevant HR data.**



Source: Rousseau & Barends (2011), van der Laken (2018)

Public Interest in HR Analytics

Monthly Google search interest on “people analytics” and related terms over time. Values are proportional to the maximum value and fit by locally weighted regression lines (LOESS).



Source: van der Laken (2018)

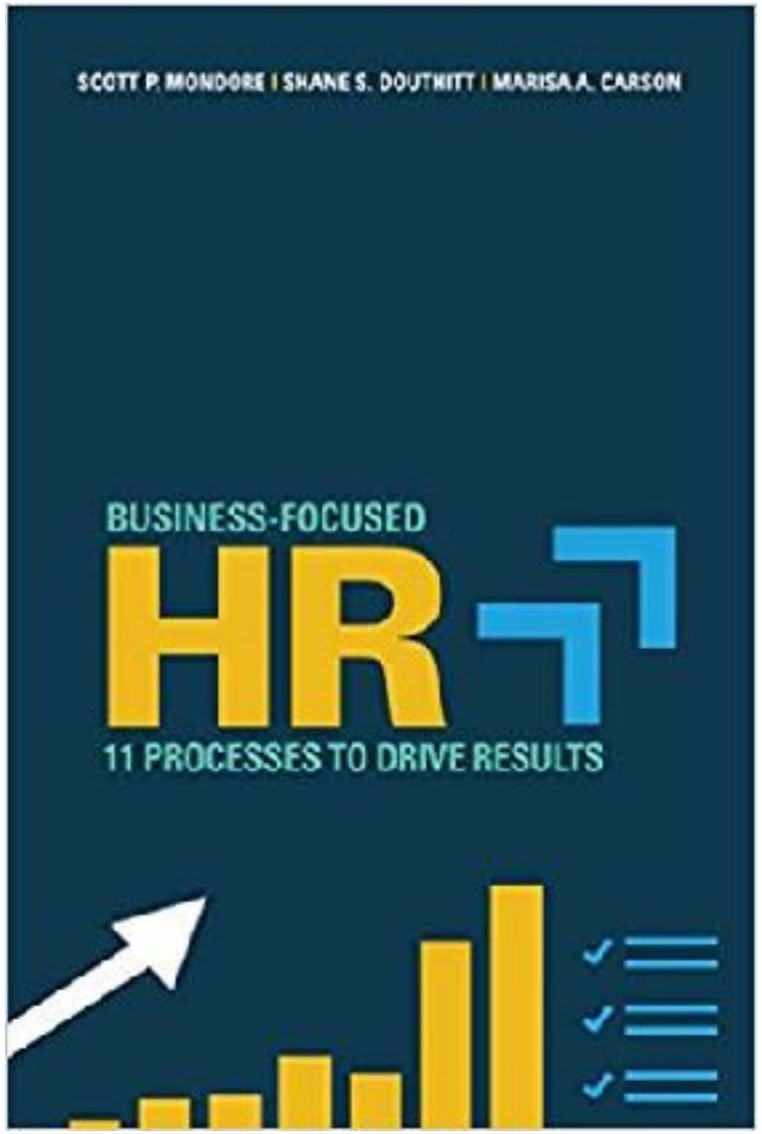
HR Analytics as Data-Driven/ Evidence-Based HRM

The key to driving business performance is understanding which competencies drive employee performance, and **then ensuring that these competencies are available in the workforce by creating people processes around these competencies**.

In practice, this means creating:

- **Recruitment processes** to hire the right competencies
- **Learning and development processes** to train the right competencies
- **Career development and compensation processes** to retain the right competencies

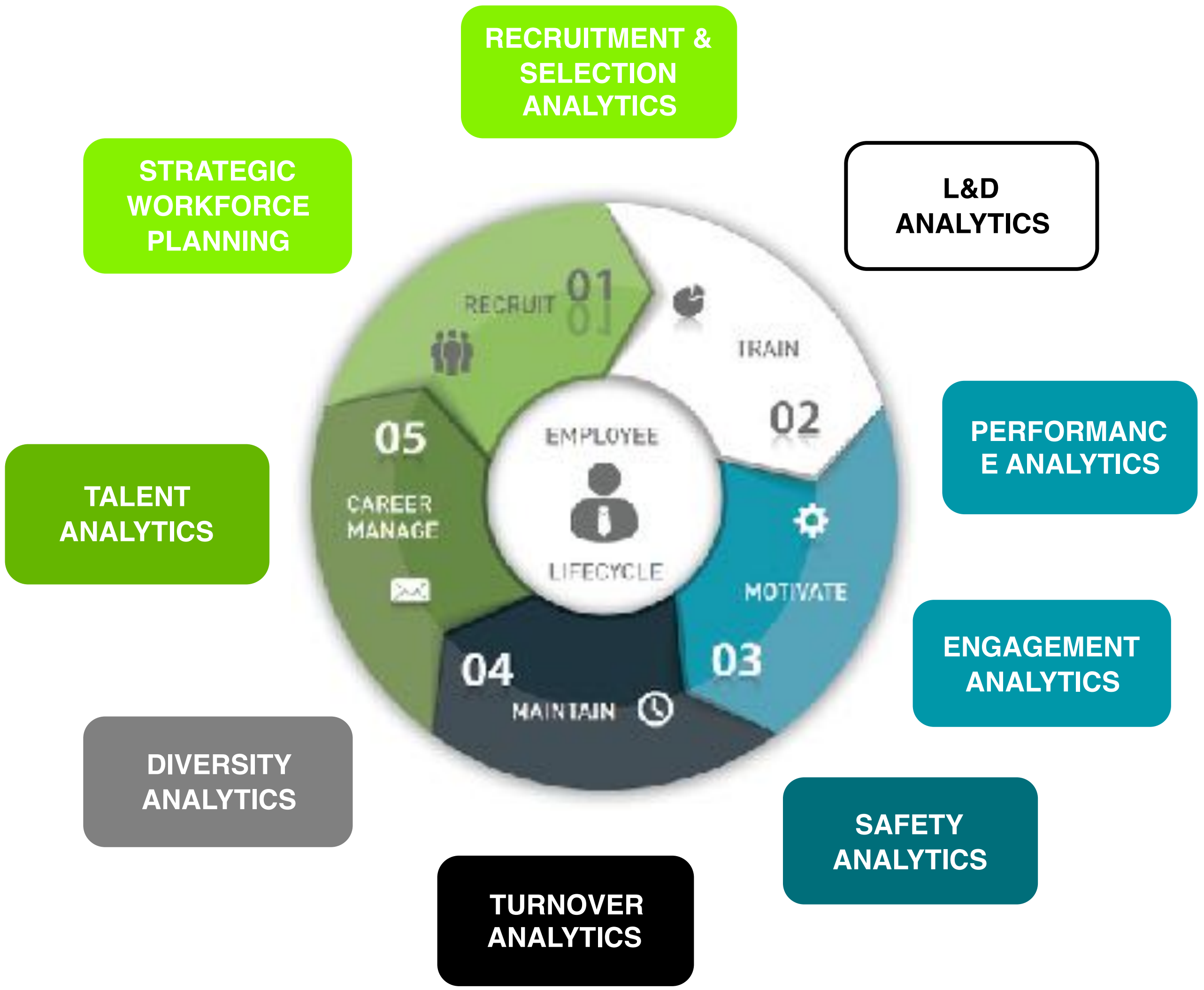
Source: Blumberg (2018)



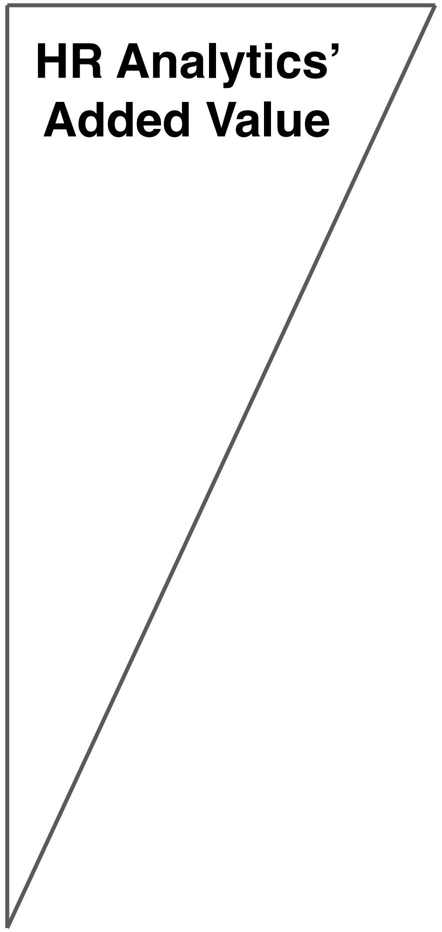
HR Analytics Agenda

HR analytics helps to **optimise mechanism behind HRM value chain** by allowing us to **find answers to certain key questions**, such as…

- Which channels bring us the best candidates?
- What characteristics differentiate successful candidates from unsuccessful ones?
- What factors contribute to successful onboarding?
- Which KPIs have the strongest link to the company's financial results?
- Which training sessions are most likely to lead to improvement of work performance?
- Which interventions have the biggest impact on well-being or work-life balance perceived by employees?
- What increases or decreases the employees' engagement level?
- Who represents hidden talent that needs to be detected and further developed?
- Where can resistance be expected with respect to planned changes in the company and who can instead be their ambassador or catalyst?
- Which factors contribute to employee turnover?
- etc.



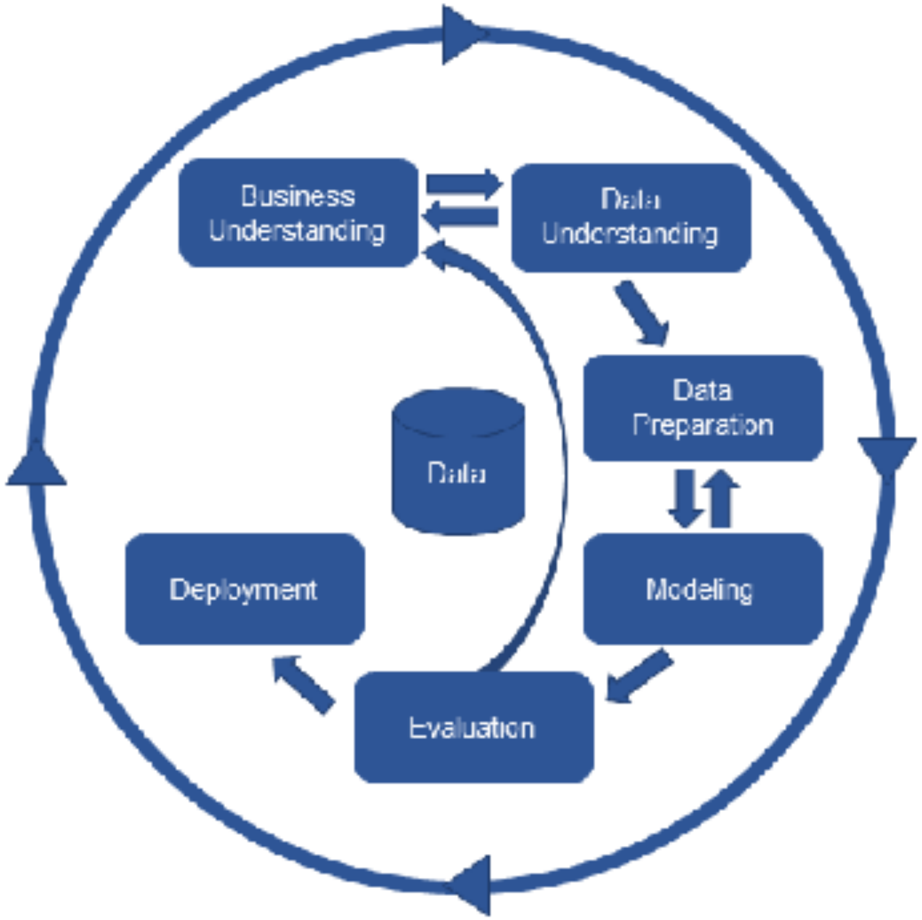
HR Analytics Maturity Model



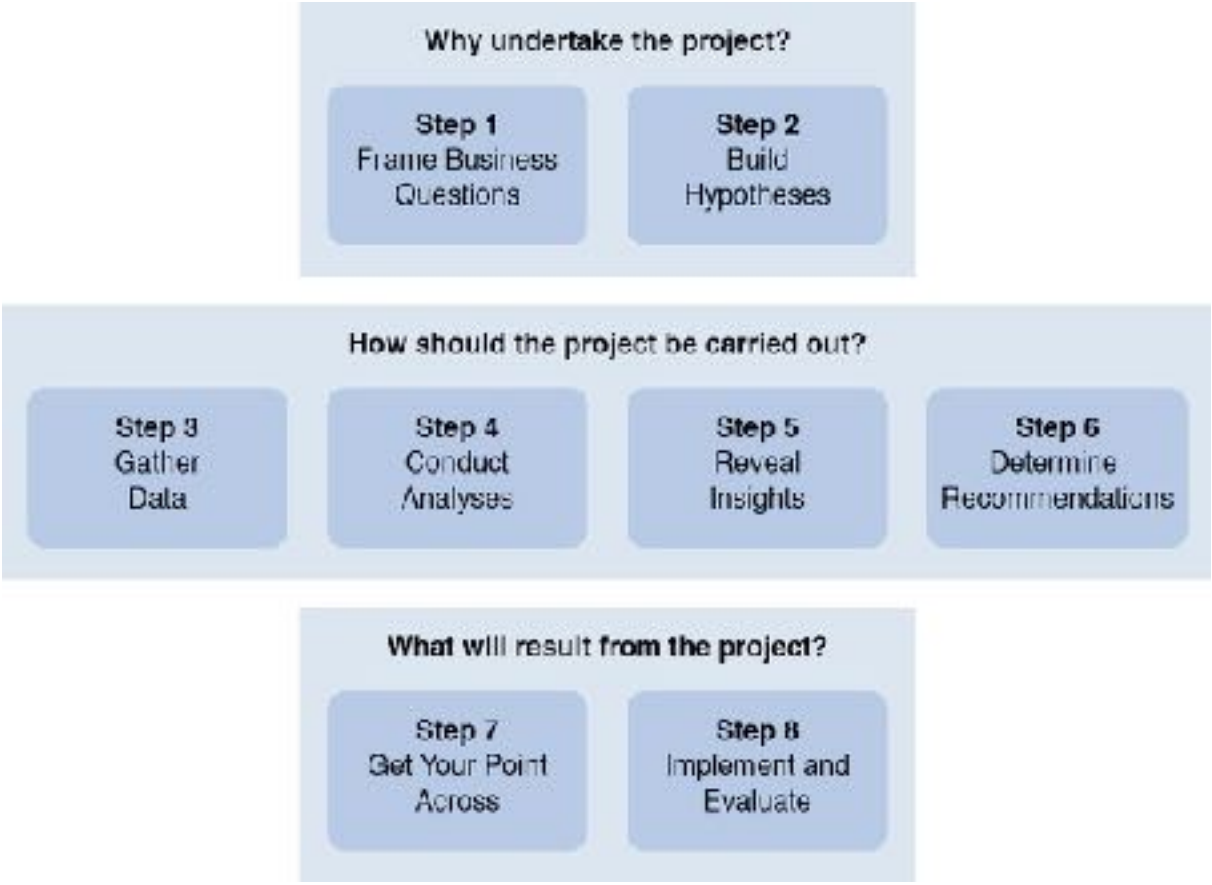
Source: Van Vulpen (2016)

HR Analytics Project Workflow

CRISP-DM



Eight Step Model for Purposeful Analytics



Source: Guenole, Ferrar & Feinzig (2017)

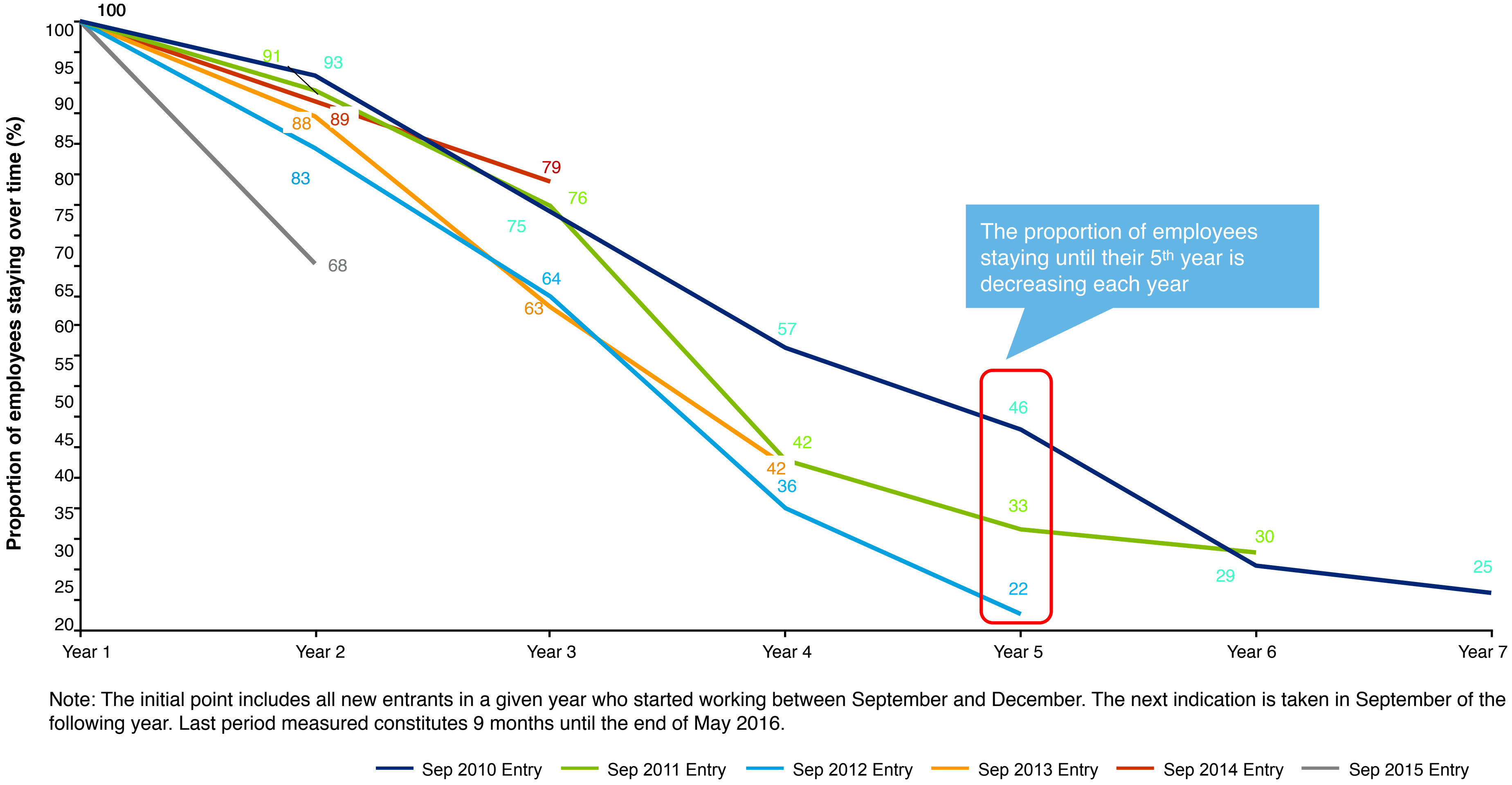
HR Analytics Skillsets

What skillsets are required for successful HR Analytics projects and what happens when one of them is missing?

Source: Van Vulpen (2016)

Case Study 1

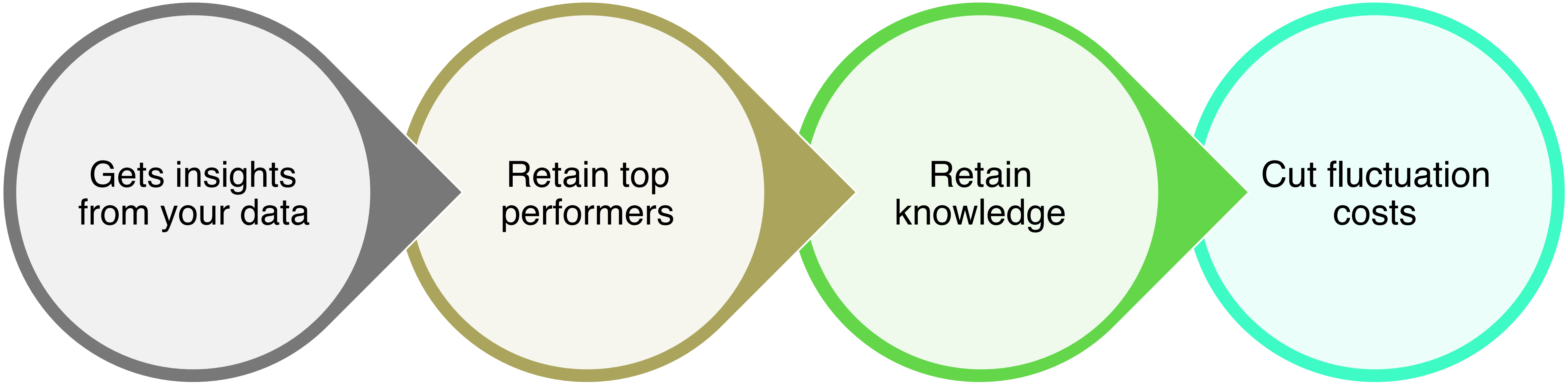
Leadership of a **professional services firm** division engaged Deloitte to address the **issue of high voluntary attrition**. The objective was to provide the leadership with data-driven insights into **why employees leave, identify which segments and individuals are at a higher risk of leaving in the near future and propose a plan to retain the key individuals.**



Note: The initial point includes all new entrants in a given year who started working between September and December. The next indication is taken in September of the following year. Last period measured constitutes 9 months until the end of May 2016.

Case Study 1

Project business case



Case Study 1

Project business case - Inputs

Some Basic Assumptions		Value
Average Monthly Gross Salary + Benefits	CZK	50 000
Average Annual Gross Salary + Benefits	CZK	600 000
Average Annual Gross Salary + Benefits + Health & Social Ins. (34,5%)	CZK	807 000
Number of Employees		230
Annual Attrition		28%

Direct Costs		Value
1. Average Separation (exit interviews, administration procedures, etc.)	CZK	5 000
2. Average Vacancy (Temporary Help + Overtime)	CZK	15 000
3. Average Acquisition (Ads, Travel, Interviews, Physicals, Bonuses, ...)	CZK	15 000
4. Average Placement (New Supplies, Onboarding, Training)	CZK	20 000
Total	CZK	55 000

Lost Productivity		Value
1. Annual Revenue (less COGS) Per Employee	CZK	1 105 820
2. Workdays Per year		240
3. Average Workdays Position Is Open		30
4. Average Onboarding / Training Period		60
5. New Hire's Effectiveness During Onboarding/Training		50%
6. Supervisor's Effectiveness During New Hire's Inboarding/Training		95%
Total (Calculation: 1. / 2. x (3. + 4. x (1 - 5.) + 4 x (1 - 6.))	CZK	290 278

Savings of Salary + Benefits		Value
1. Average Annual Salary + Benefits (Health & Social Ins. Incl.)	CZK	807 000
2. Workdays Per year		240
3. Average Workdays Position Is Open		30
Total (Calculation: 1. / 2. x 3.)	CZK	100 875

Case Study 1

Project business case - Outputs

Estimated Turnover Cost Per Employee		Value
1. Direct Costs	CZK	55 000
2. Lost Productivity	CZK	290 278
3. Savings of Salary + Benefits	CZK	100 875
Total (Calculation: 1. + 2. - 3.)	CZK	244 403

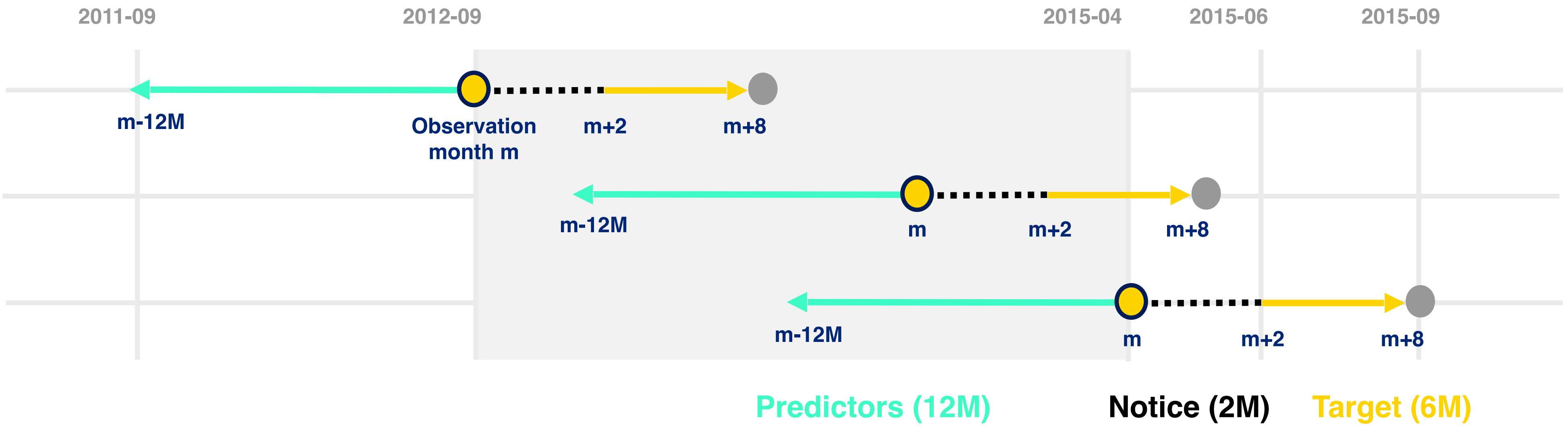
Total Cost of Employee Turnover		Value
Annual Employee Churn		78
Total Cost	CZK	19 161 178

Saved Costs With Attrition Being Reduced By...		Value
1%	CZK	684 328
2%	CZK	1 368 656
3%	CZK	2 052 983
4%	CZK	2 737 311
5%	CZK	3 421 639

Case Study 1

Target variable

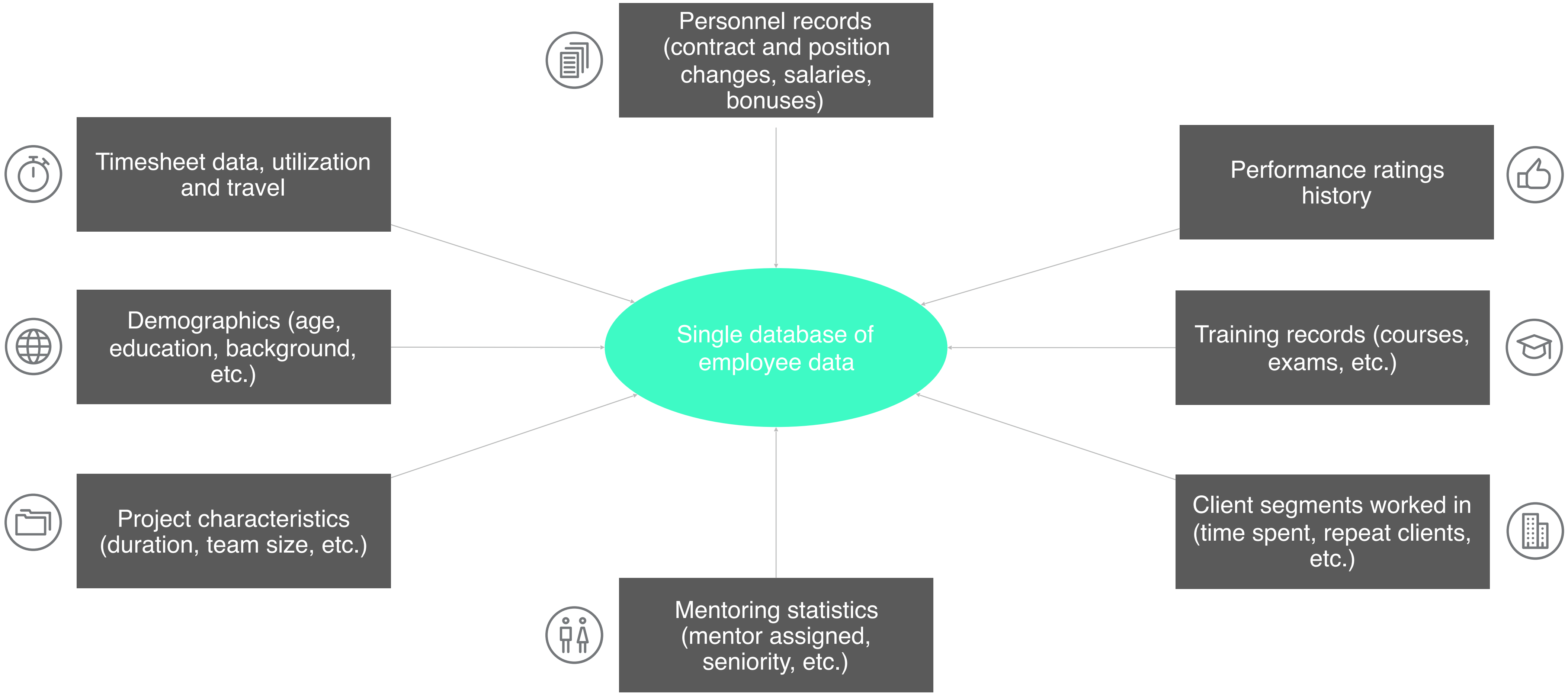
- **Target date** = date when employee left company
- **Target period** = date window when target date could occur, to define binary target
- **Observation date** = monthly snapshot of employee data
- **Blackout period** = employee notice period, when we do not observe any data about employee to avoid data leakage



Model works with each employee and his 12 months history, notice period is 2 months and projection of leave up to 6 months.

Case Study 1

Workforce Analytics relies on a **wide range of data from various sources** and combines them in a single database. Both traditional and non-traditional HR data are used.



Case Study 1

Types of predictors

Standard predictors

Monthly snapshot of employee data, E.g. Monthly average salary base

Two types (inspired by banking industry)

1. **Static data** – not very variable over time, e.g. demographics
2. **Transaction data** – changing in time, e.g. monthly performance of employee

Trend predictors

Time evolution of standard transactional predictors over several months (e.g. last month, 6months, 9months, 12months).

To indicate changes in employee behavior, e.g. trend in performance in last 6M (increasing, decreasing, same)

Types of trend predictors:

1. **Trend curve**
2. **Volatility**
3. **Difference now vs. last 1/3/6/12M**

Peer predictors

Comparison of standard predictors with peer average, i.e. group with similar working conditions

Is employee better, worse or same as peer, e.g. peer monthly salary

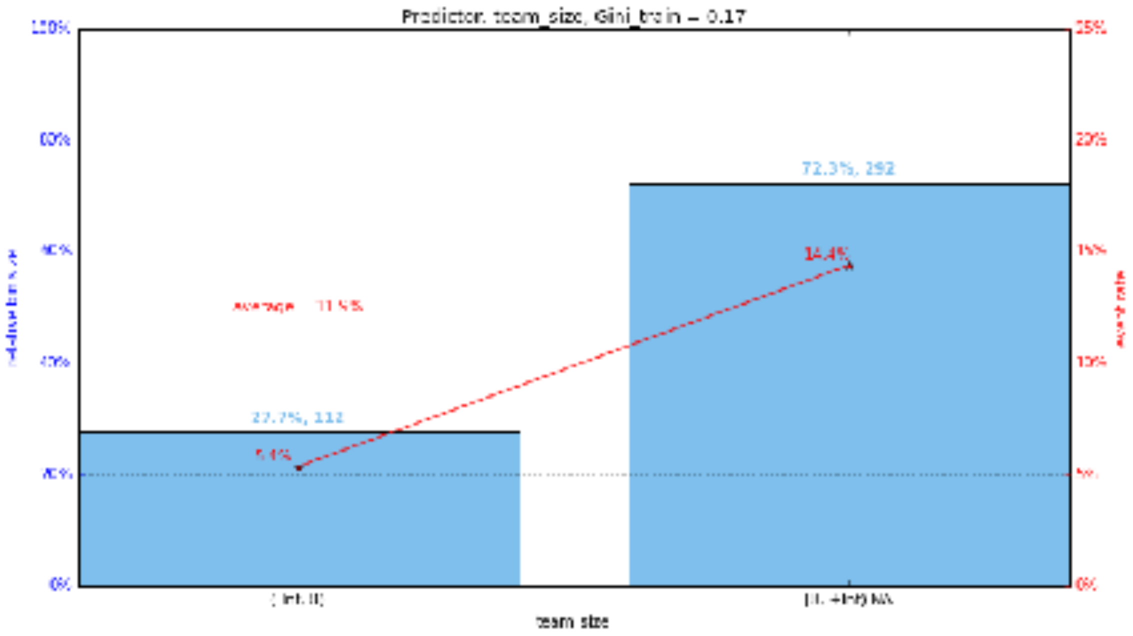
Types of peer dimension:

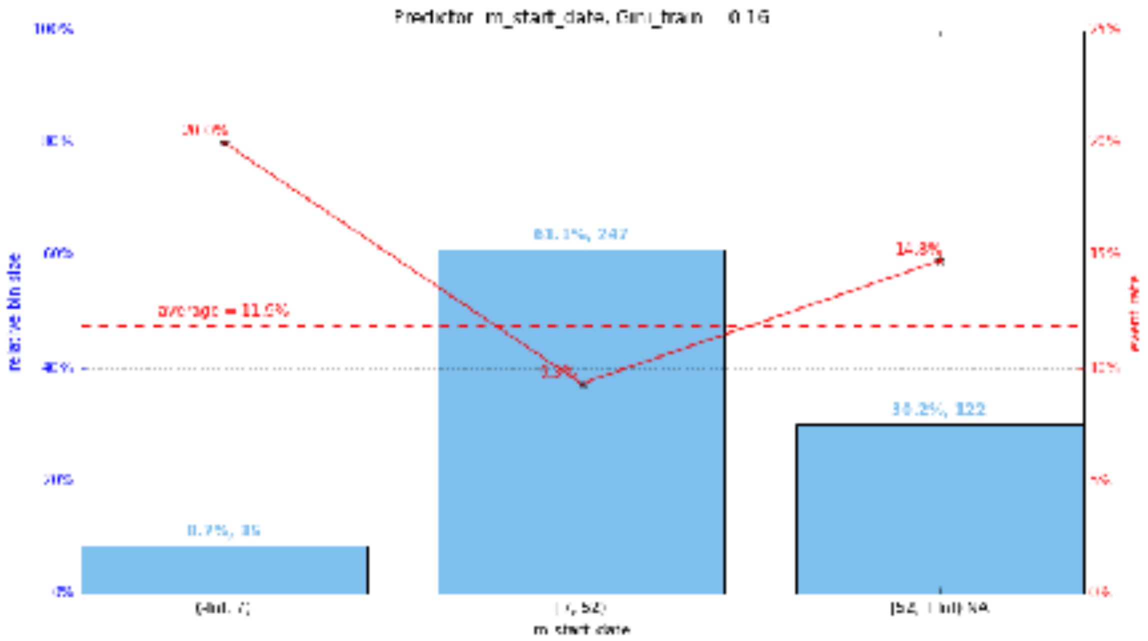
1. **Region**
2. **Team, unit**
3. **Service, client segment**
4. **Position level**
5. **Length in company**

Peer is defined based on industry specifics

Case Study 1

Binning of variables – treating missing values, outliers, and non-linearities





Case Study 1

Steps in feature selection

Predictors pool
(500-1000)

Consolidated all calculated different types of predictors from various areas

Gini and Correlation Selection
(200-300)

First selection of predictors, which have at least some relevance to target (churn)

Variable Clustering
(100)

To avoid multicollinearity among predictors in regression, group similar predictors and within each cluster select the most significant ones

Automatic and Manual Selection
(30)

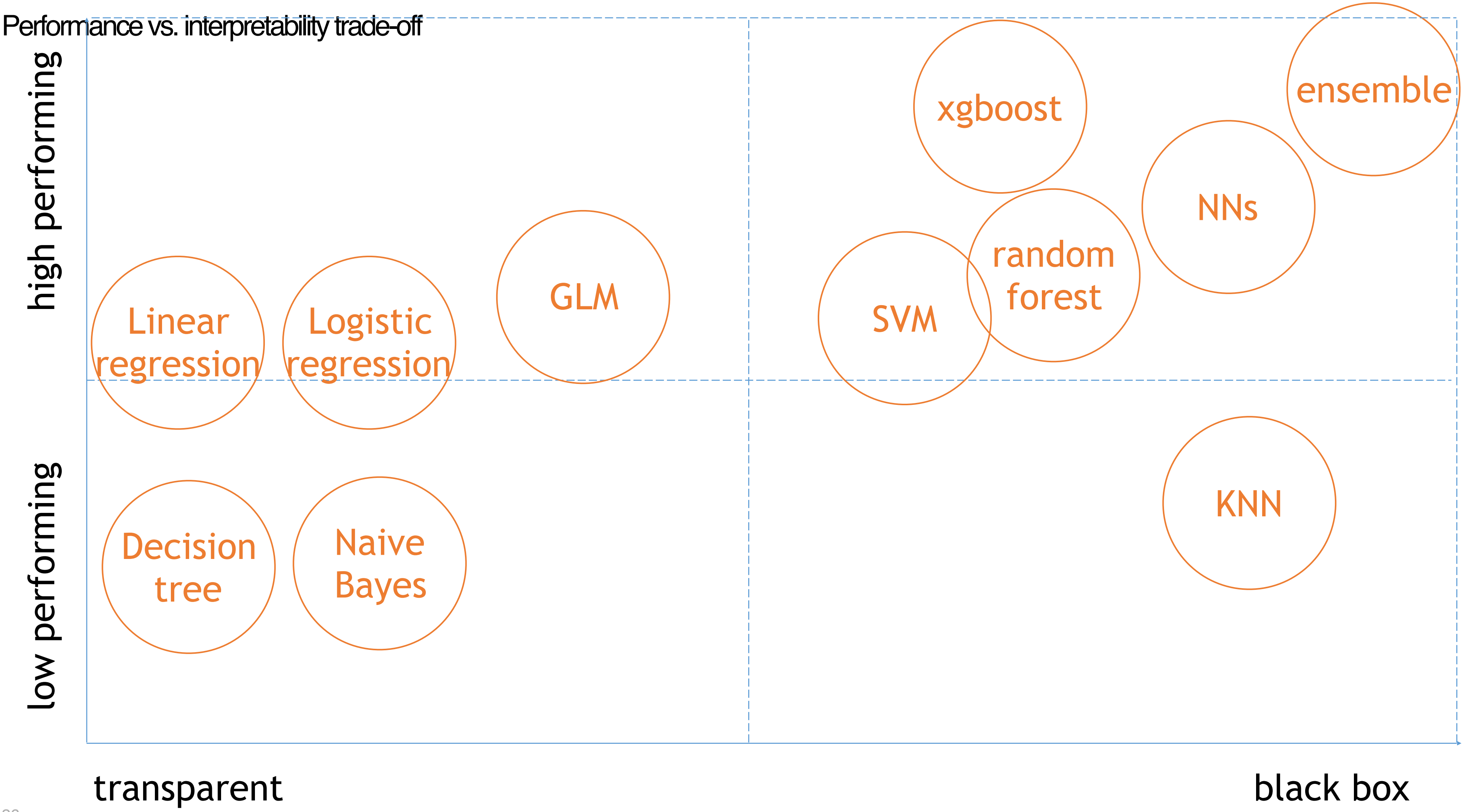
Automatic selection of best predictors (e.g., glmnet with LASSO and ridge penalties or stepwise regression) + manual selection (business interpretation, appropriate shape of event rate curve in binning)

Best predictors in final model

FINAL 5 to 15

Case Study 1

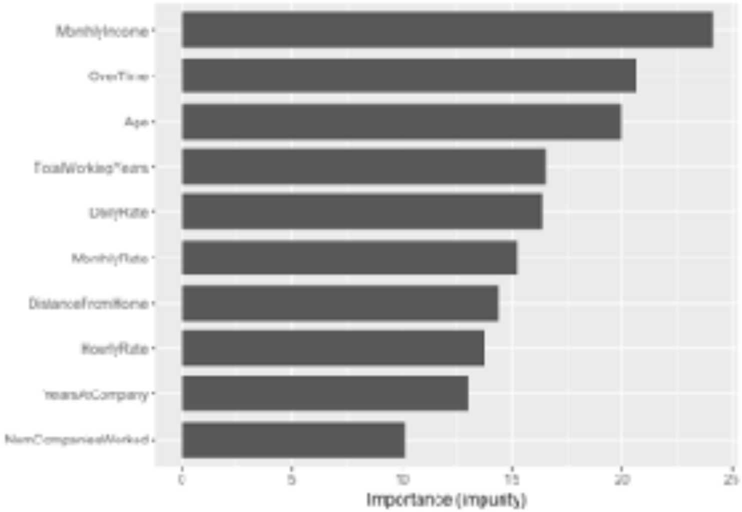
Performance vs. interpretability trade-off



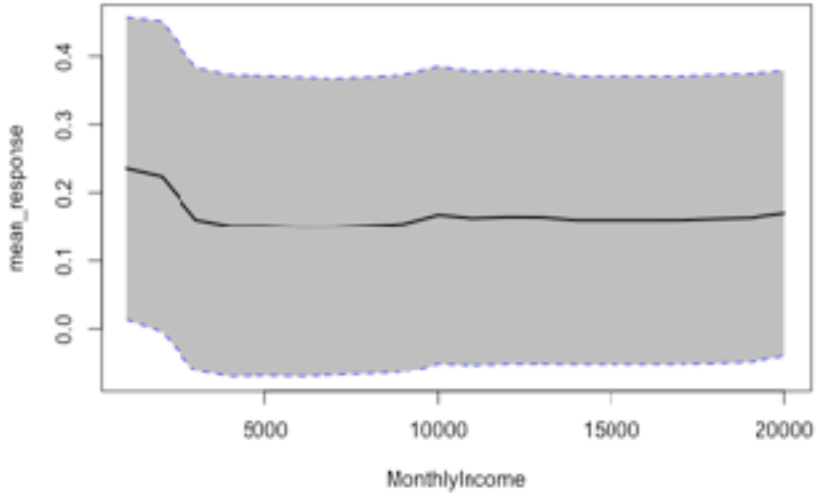
Case Study 1

Resolving performance vs. interpretability trade-off?

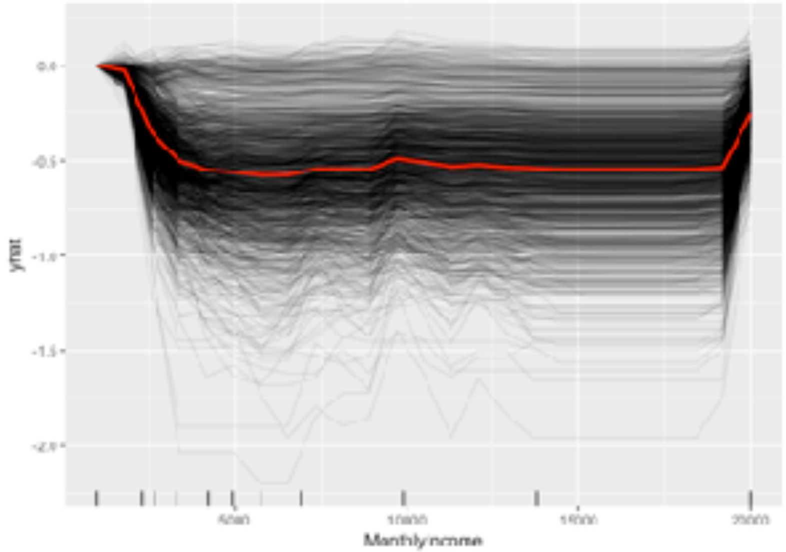
Variable importance measures



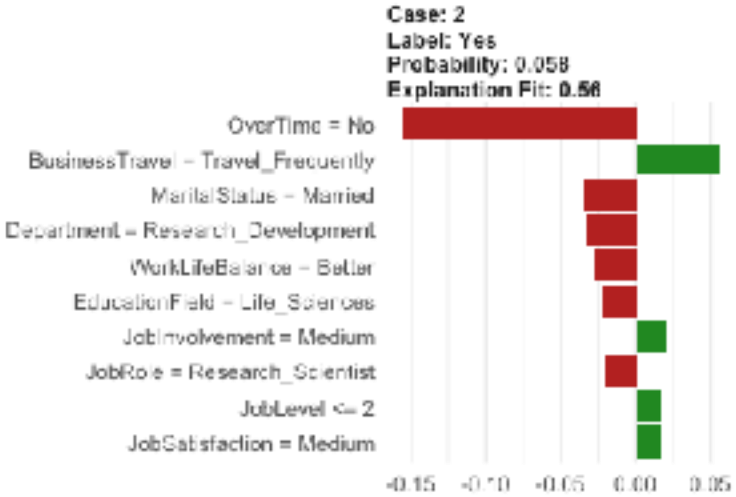
Partial dependence plots



Centralized individual conditional expectation



Local interpretation of individual predictions



Case Study 1

Why should we improve our understanding of ML models?

Source: Glander (2018)

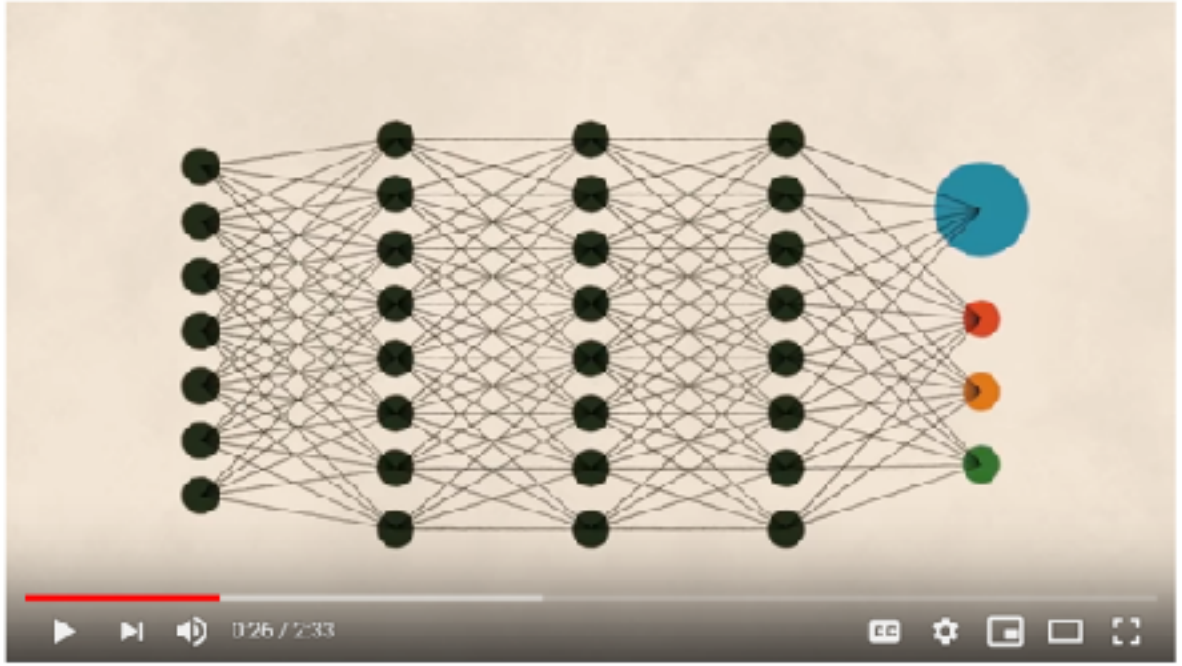
Case Study 1

Preventing biases in ML models

Amazon ditched AI recruiting tool that favored men for technical jobs

Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process

https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

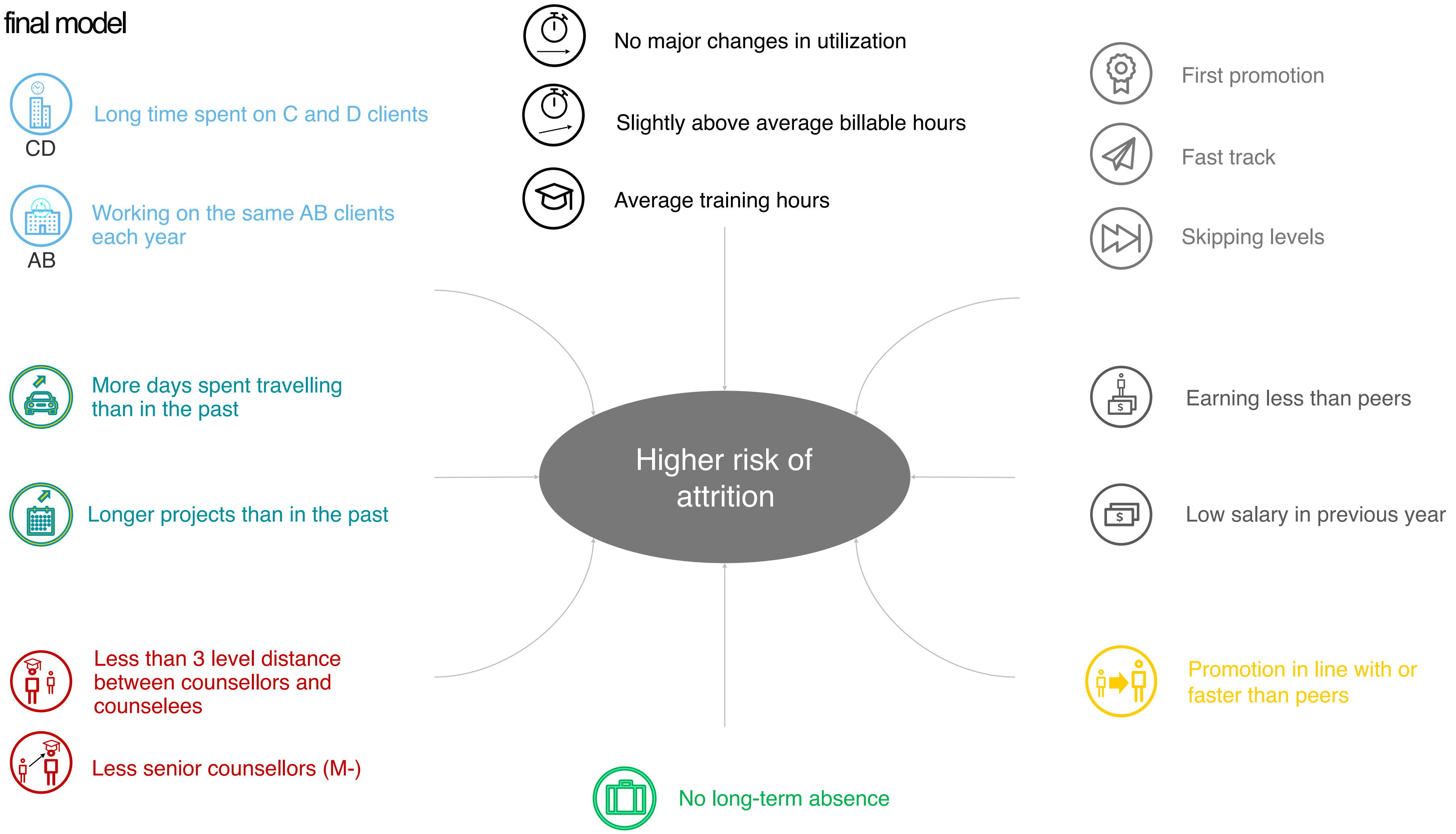


Machine Learning and Human Bias

https://www.youtube.com/watch?v=59bMh59JQDo

Case Study 1

Predictors in the final model



Case Study 1

Employees at risk of leaving with probability more than 18% to leave in next 6 months. Immediate action is required!

Employee	Prob. of leaving	Status	No. 1 Destabilizing factor	No 2. Destabilizing factor	No. 3 Destabilizing factor
	67,71%	Active	Junior counsel or (In Charge or Senior)	Growth in minimum team size from previous year	Minor loss of AB clients since previous year
	50,99%	Active	Promoted in line with peers Earning less than peers	Minor loss of AB clients since previous year	n/a
	48,74%	Active	Growth in minimum team size from previous year	Promoted in line with peers Earning less than peers	Stable utilization over time
	48,65%	Active	Higher than average billable utilization (75-79%)	Promoted in line with peers Earning less than peers	Same AB clients since previous year
	43,29%	Active	Junior counsel or (In Charge or Senior)	Promoted in line with peers Earning less than peers	Average training utilization (4-12%)
	39,73%	Active	Junior counsel or (In Charge or Senior)	Growth in minimum team size from previous year	Stable utilization over time
	36,84%	Active	Junior counsel or (In Charge or Senior)	Promoted in line with peers Earning less than peers Long time on CD projects	Average training utilization (4-12%)
	30,29%	Active	Promoted in line with peers Earning less than peers	Minor loss of AB clients since previous year	Average training utilization (4-12%)
	29,28%	Active	Minor loss of AB clients since previous year	Stable utilization over time	Junior counsellor (In Charge or Senior)
	26,56%	Active	Junior counsel or (In Charge or Senior)	Growth in minimum team size from previous year	Average training utilization (4-12%)
	26,10%	Active	Higher than average billable utilization (75-79%)	Promoted in line with peers Earning less than peers Long time on CD projects	Average training utilization (4-12%)
	20,87%	Active	Higher than average billable utilization (75-79%)	Promoted in line with peers Earning less than peers	Average training utilization (4-12%)
	19,13%	Active	Minor loss of AB clients since previous year	Higher than average billable utilization (75-79%)	n/a
	18,31%	Active	Same AB clients since previous year	Stable utilization over time	Average training utilization (4-12%)

Case Study 1

Employees with probability of leaving higher than the average churn rate 6 months later.

Employee Name	Probability of leaving	Status	No. 1 Destabilizing factor	No 2. Destabilizing factor	No. 3 Destabilizing factor
	67,71%	Active	Junior counsellor (In Charge or Senior)	Growth in minimum team size from previous year	Minor loss of AB clients since previous year
	50,99%	Left	Promoted in line with peers Earning less than peers	Minor loss of AB clients since previous year	n/a
	48,74%	Left	Growth in minimum team size from previous year	Promoted in line with peers Earning less than peers	Stable utilization over time
	48,65%	Active	Higher than average billable utilization (75-79%)	Promoted in line with peers Earning less than peers	Same AB clients since previous year
	43,29%	Left	Junior counsellor (In Charge or Senior)	Promoted in line with peers Earning less than peers	Average training utilization (4-12%)
	39,73%	Active	Junior counsellor (In Charge or Senior)	Growth in minimum team size from previous year	Stable utilization over time
	36,84%	Left	Junior counsellor (In Charge or Senior)	Promoted in line with peers Earning less than peers Long time on CD projects	Average training utilization (4-12%)
	30,29%	Left	Promoted in line with peers Earning less than peers	Minor loss of AB clients since previous year	Average training utilization (4-12%)
	29,28%	Left	Minor loss of AB clients since previous year	Stable utilization over time	Junior counsellor (In Charge or Senior)
	26,56%	Active	Junior counsellor (In Charge or Senior)	Growth in minimum team size from previous year	Average training utilization (4-12%)
	26,10%	Active	Higher than average billable utilization (75-79%)	Promoted in line with peers Earning less than peers Long time on CD projects	Average training utilization (4-12%)
	20,87%	Left	Higher than average billable utilization (75-79%)	Promoted in line with peers Earning less than peers	Average training utilization (4-12%)
	19,13%	Left	Minor loss of AB clients since previous year	Higher than average billable utilization (75-79%)	n/a
	18,31%	Active	Same AB clients since previous year	Stable utilization over time	Average training utilization (4-12%)

Case Study 1

Examples of recommendations

Case Study 2

- Professional-services firm started to **change their sales model to achieve its growth goals.**
- To **support adoption of new sales model** across the organization the firm wanted to build a **profile of high performers** around those sales professionals who were excelling in the current model.
- The goal was to use this profile for redesigning selection and training & development systems and aligning them with business goals for sales professionals to ensure…

    1) that the **current sales professionals were being trained on and dedicated attention to those key weaknesses that were leading to underperformance**, and

    2) that **new hires to the sales team had the characteristics critical to success in the organization**.

Source: Mondore, Spell, Betts & Douthitt (2018)

Case Study 2

The four-step process that is **based on research and data** but **tailored specifically to the organization**, creating a customized solution that selects employees who fit into the given context and develops the current workforce on the key drivers most **strongly linked to business success:**

1) Define greatness,
2) Assess and compare the current workforce to greatness,
3) Develop the current workforce toward greatness, and
4) Hire greatness.

Source: Mondore, Spell, Betts & Douthitt (2018)

Case Study 2

- Defining the greatness using **outputs from several interviews conducted with partners and top performing sales professionals** - gathering information about the **knowledge, skills, abilities, and behaviors that make a sales employee successful in the organization.**
- **Identification of most frequent themes/competencies** + their **behavioral examples/indicators** (e.g., Communication competency included behaviors such as "quickly builds rapport through speech and action when first meeting with decision-makers" or "uses listening as a strategy to gather information and build trust.")
- **Definition of high performance** (company tries to positively impact by implementing new selection and development systems) using combination of three common sales outcomes:
    1) sales goal attainment,
    2) average win rate, and
    3) average win size.

Source: Mondore, Spell, Betts & Douthitt (2018)

Case Study 2

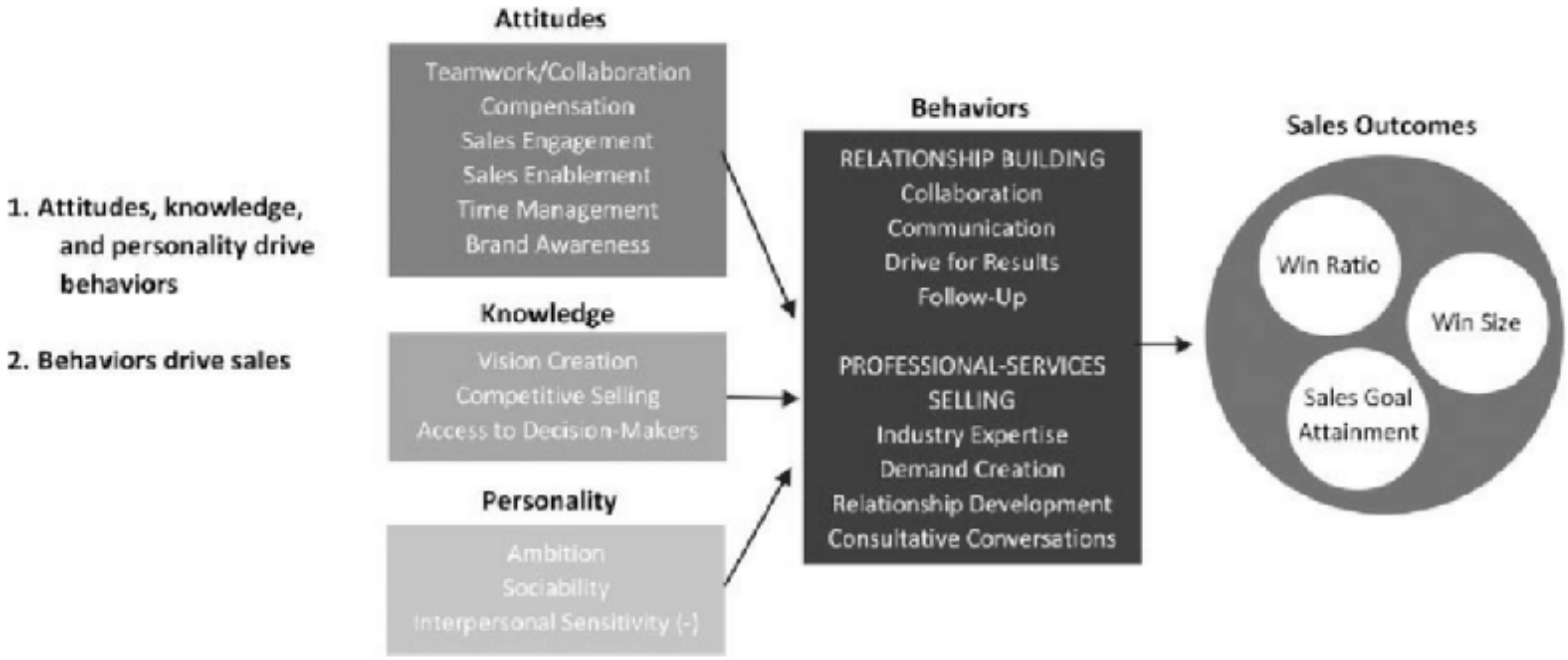
Assessing and comparing the current workforce to greatness using battery of methods by which each sales professional can be assessed on his/her **behaviors, attitudes, knowledge, and personality traits**:

- **360° feedback** to assess behavior,
- **sales-climate survey** to assess work attitudes,
- **situational-judgment test** to assess job knowledge, and
- **personality assessment** to assess personality traits.

Source: Mondore, Spell, Betts & Douthitt (2018)

Case Study 2

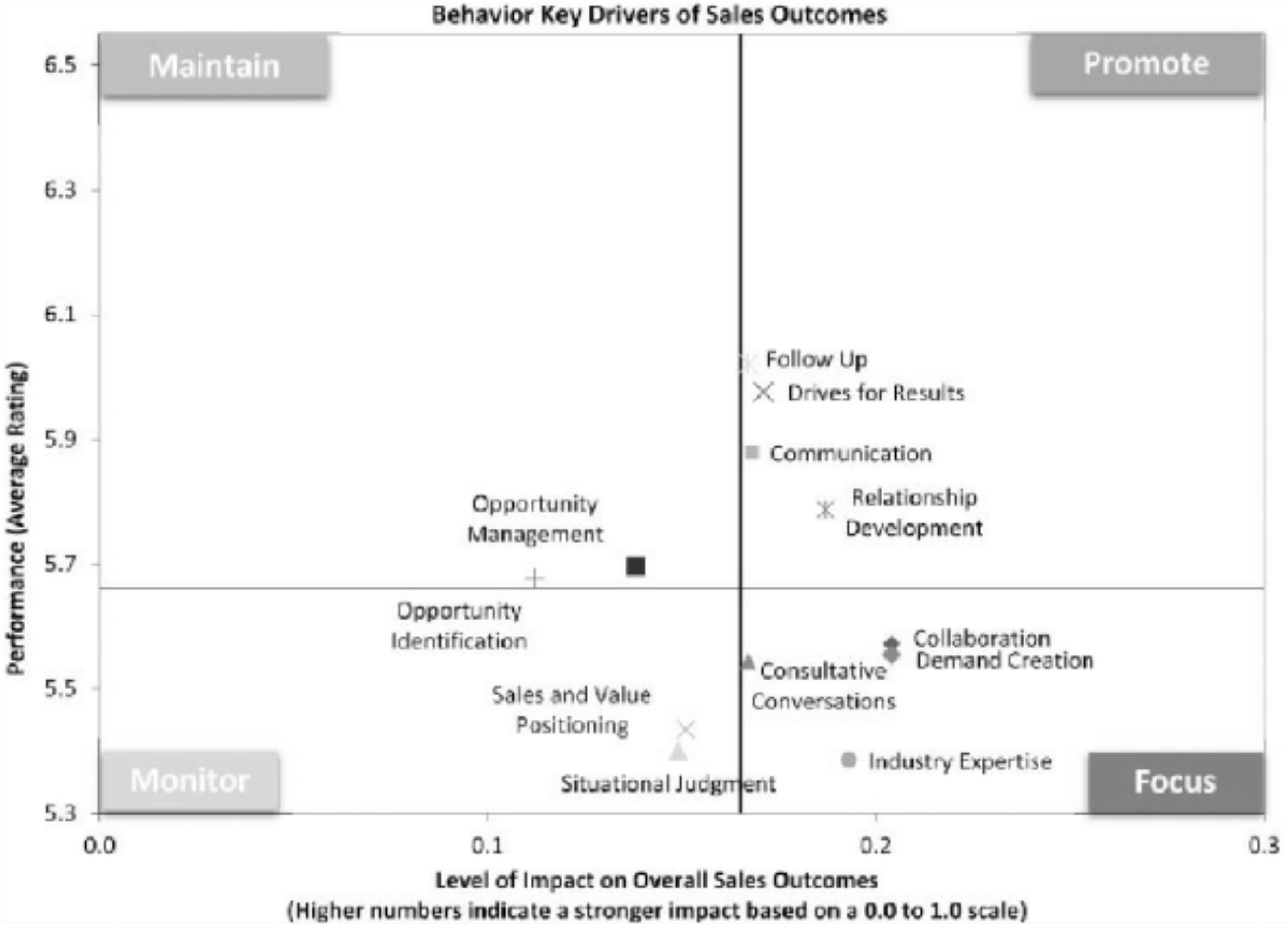
Using structural equation modeling to identify and model key drivers of sales outcomes.



Source: Mondore, Spell, Betts & Douthitt (2018)

Case Study 2

Building team and/or individual development programs around eight strongest sales drivers and weakest performance areas (Focus quadrant)



Source: Mondore, Spell, Betts & Douthitt (2018)

Case Study 2

Designing **multihurdle selection process** in combination with **predictive algorithm** weighting individual factors according to their impact on sales outcome.

Competency/Category/Dimension	Personality	Role-Playing	Interviews	SJT
Sales Engagement (A)		X		
Teamwork (A)			X	
Sales Enablement (A)			X	
Compensation & Benefits (A)			X	
Time Management (A)		X		
Brand Awareness (A)			X	
Communication (B)		X		
Drive for Results (B)			X	
Relationship Development (B)			X	
Demand Creation (B)		X		
Collaboration (B)			X	
Industry Expertise (B)		X	X	
Follow up (B)		X		
Consultative Conversations (B)		X		
Accessing Decision Makers (K)		X		X
Sales Conversations: Vision Creation (K)		X		X
Competitive Selling (K)		X		X
Interpersonal Sensitivity (P)	X			
Sociability (P)	X		X	
Ambition (P)	X		X	

A=Attitude Assessment, B=Behavioral Assessment, K=Knowledge Assessment, P=Personality Assessment

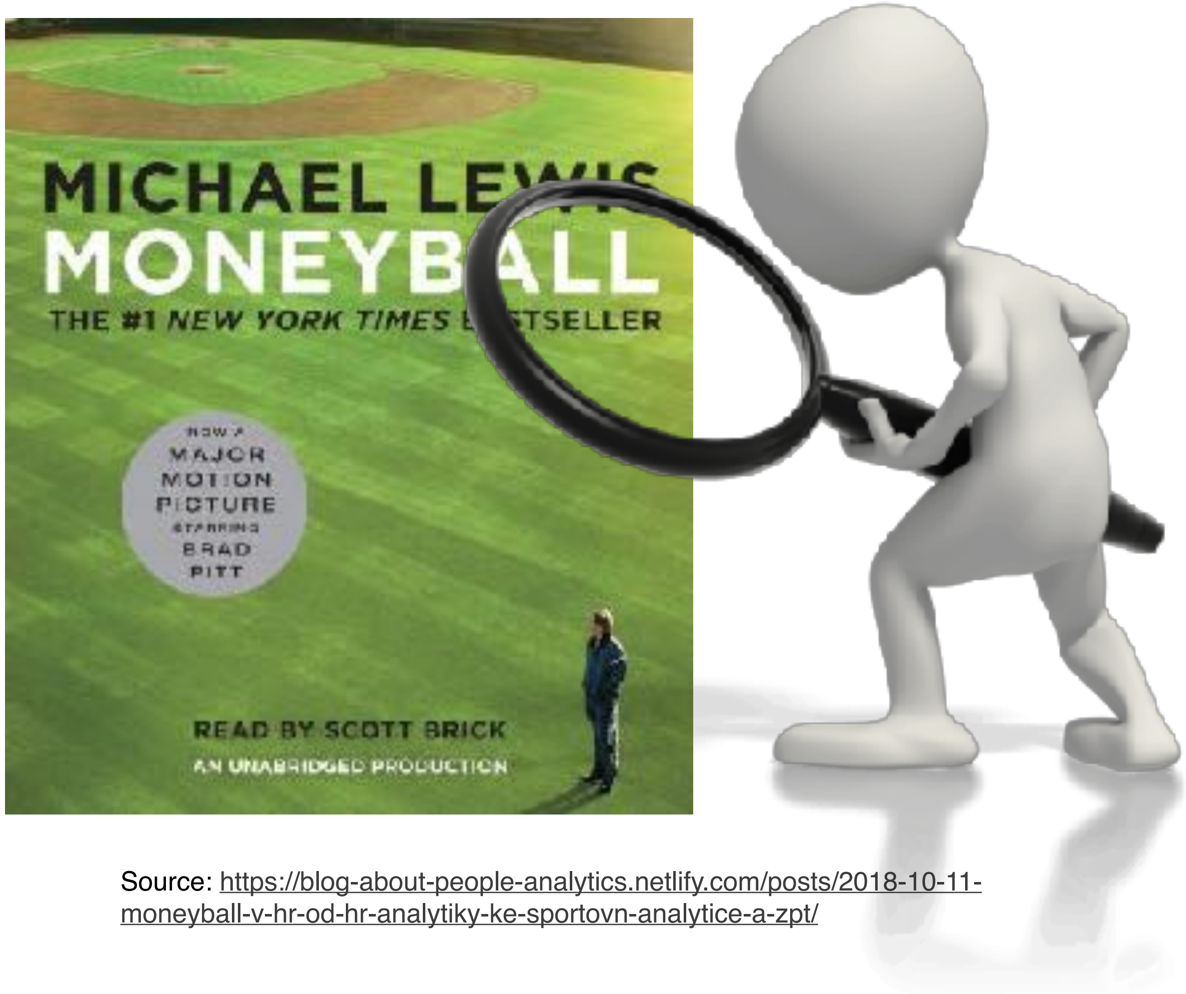
Selection process should include multiple assessments. Using a hurdled approach is recommended: hurdle online assessments (personality, situational) and on-site assessments (interviews, role-playing).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Source: Mondore, Spell, Betts & Douthitt (2018)

Case Study 3

Moneyball as a good example of HR Analytics in action



Source: https://blog-about-people-analytics.netlify.com/posts/2018-10-11-moneyball-v-hr-od-hr-analytiky-ke-sportovn-analytice-a-zpt/

Small Data Problem

HR datasets usually do not provide enough observations to train the model. We have to be careful to avoid overfitting. We are addressing the problem in two ways

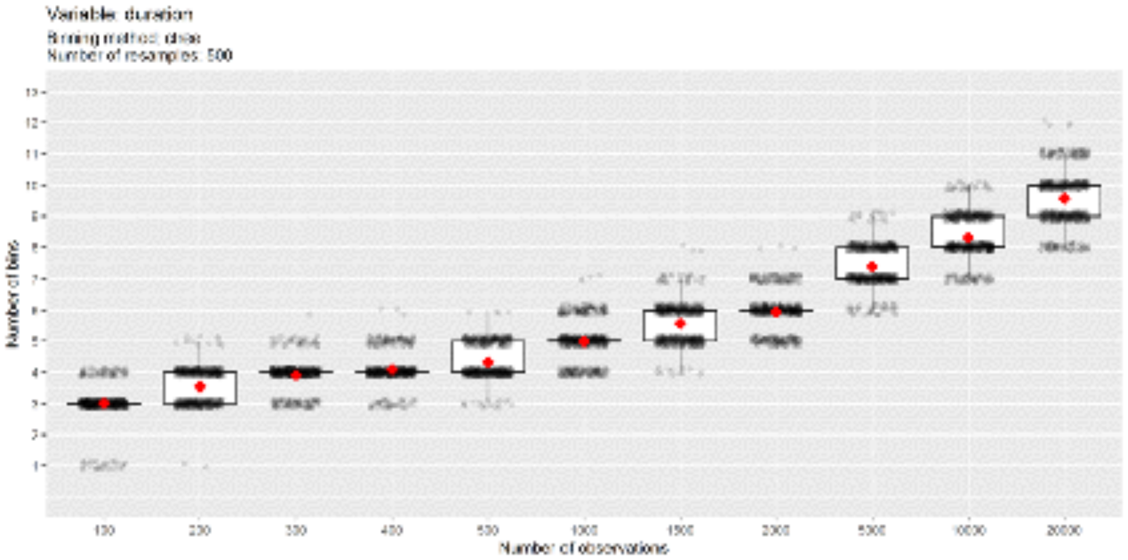
A. Use the right algorithm: logistic regression

B. Use the right binning: ctree (Conditional Inference Trees)

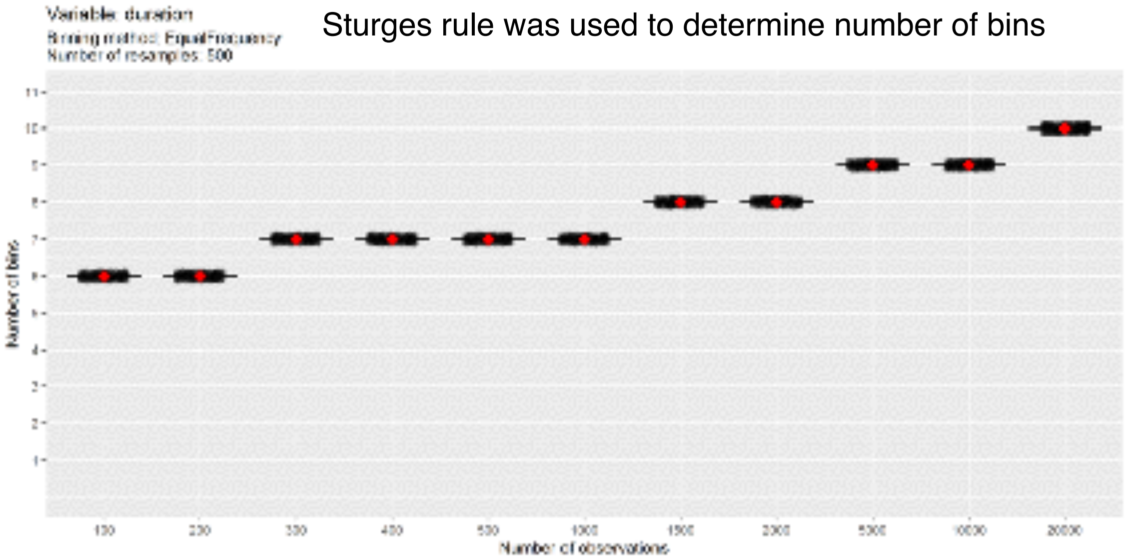
Small Data Problem

ctree versus EqualFrequency - number of bins

Duration represents a strong predictor



Sturges rule was used to determine number of bins

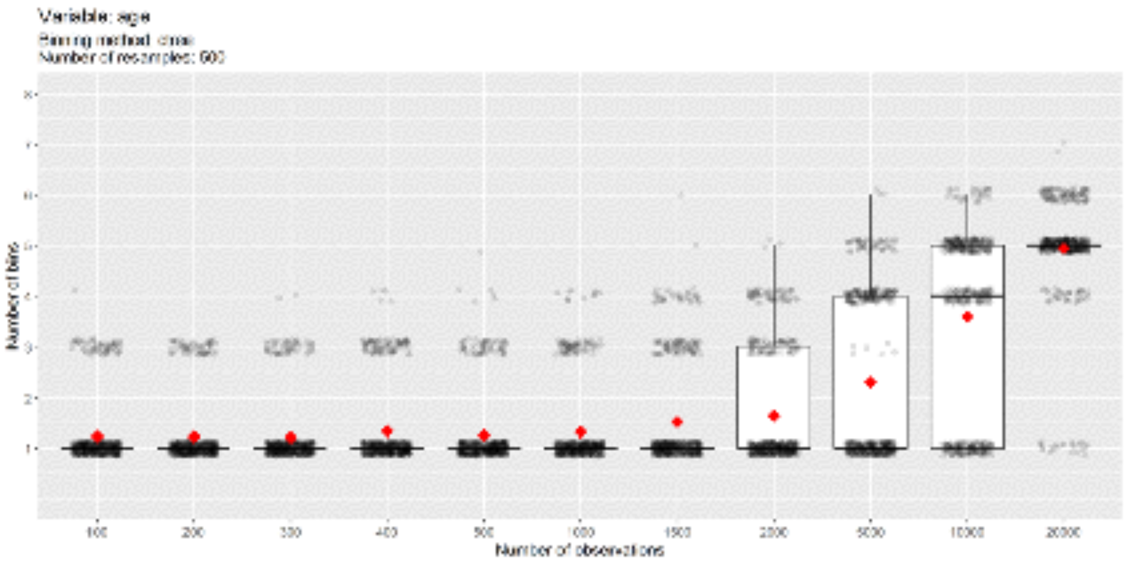


ctree suggests fewer bins compared to Sturges rule.

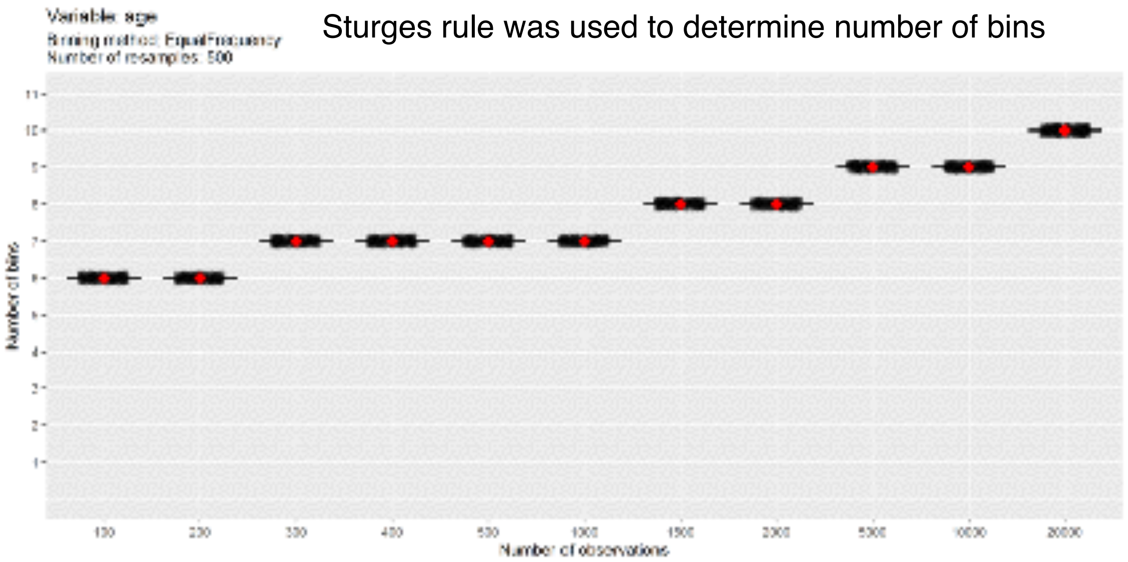
Small Data Problem

ctree versus EqualFrequency - number of bins

Age represents a weak predictor



Sturges rule was used to determine number of bins

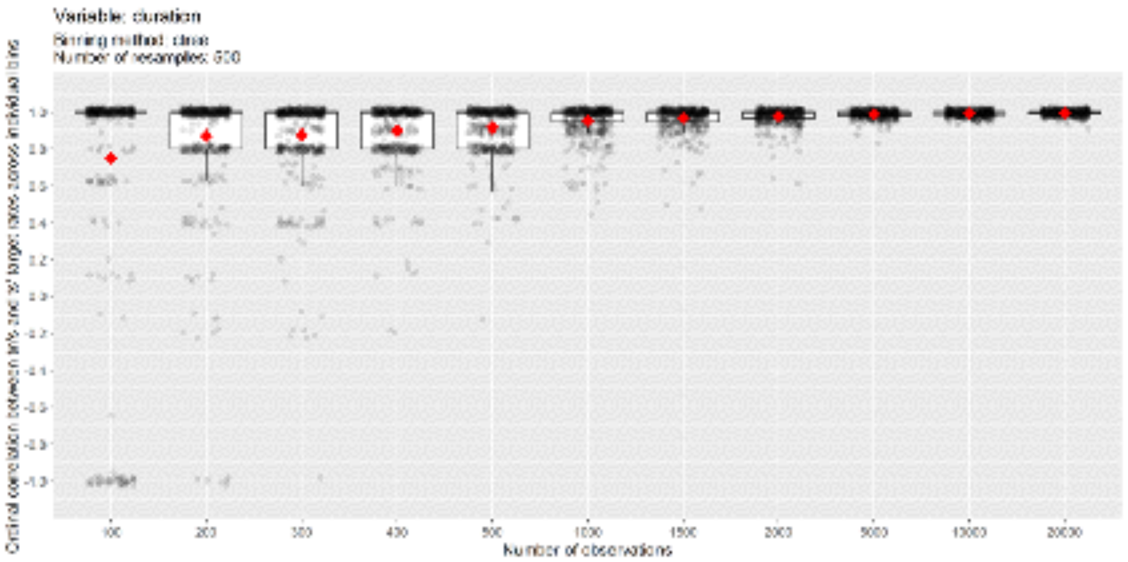


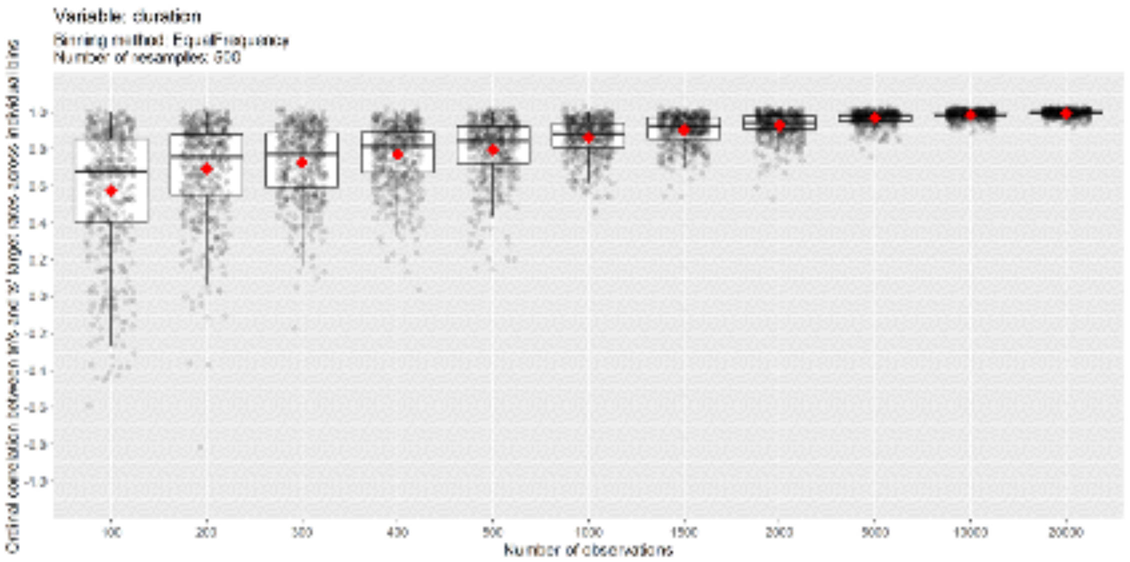
ctree suggests much fewer bins (often just one) compared to Sturges rule.

Small Data Problem

ctree versus EqualFrequency - Spearman correlation between train and test predicted probabilities

Duration represents a strong predictor



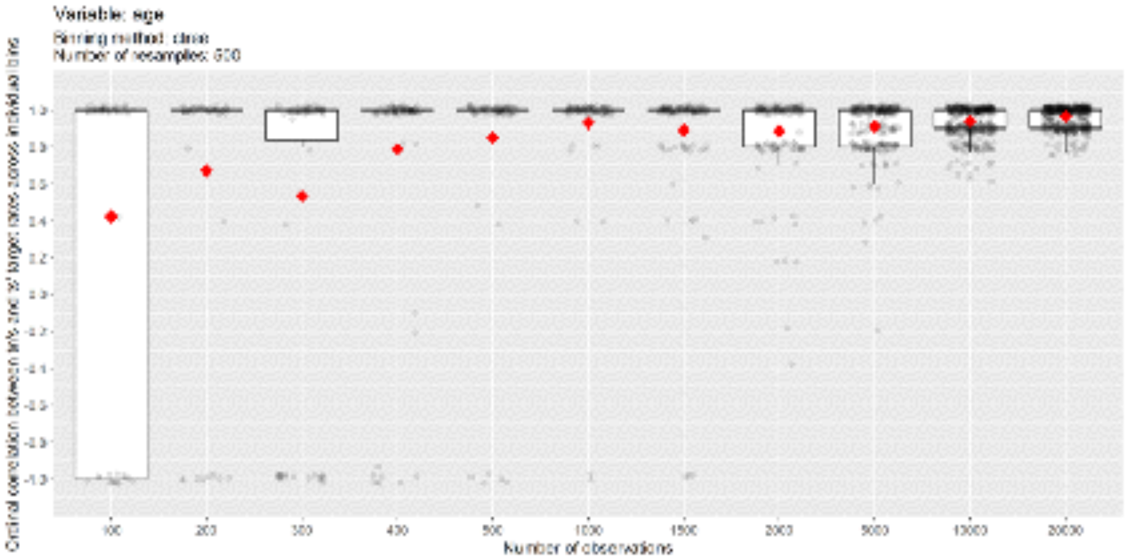


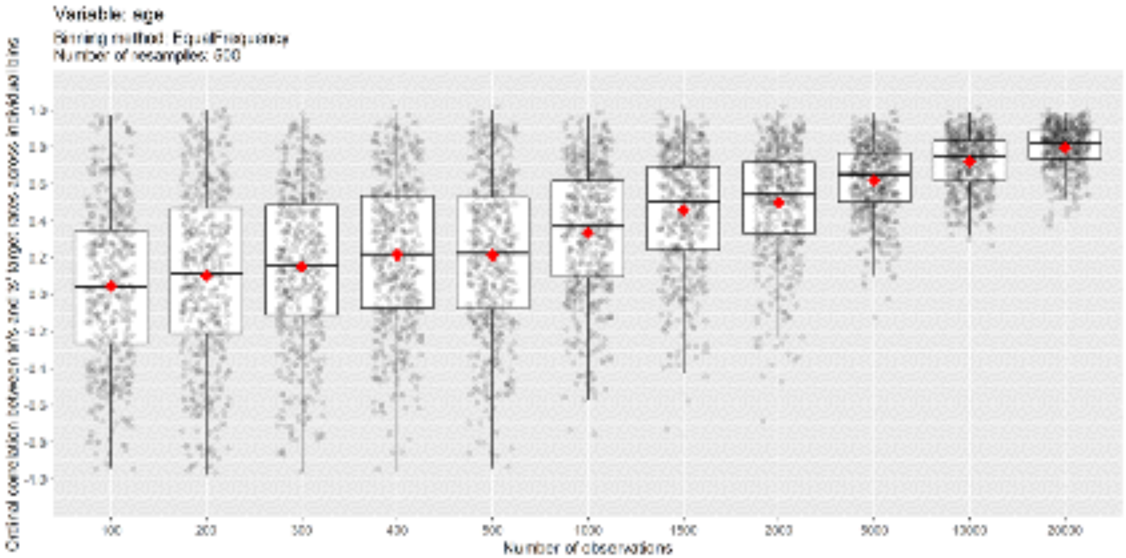
ctree gives bins which are consistent between train and test for >= 1000 observations,
EqualFrequency is prone to overfit for < 2000 observations.

Small Data Problem

ctree versus EqualFrequency - Spearman correlation between train and test predicted probabilities

Age represents a weak predictor





ctree gives bins which are consistent between train and test for >= 1000 observations,
EqualFrequency is prone to overfit even for the highest number of observations.

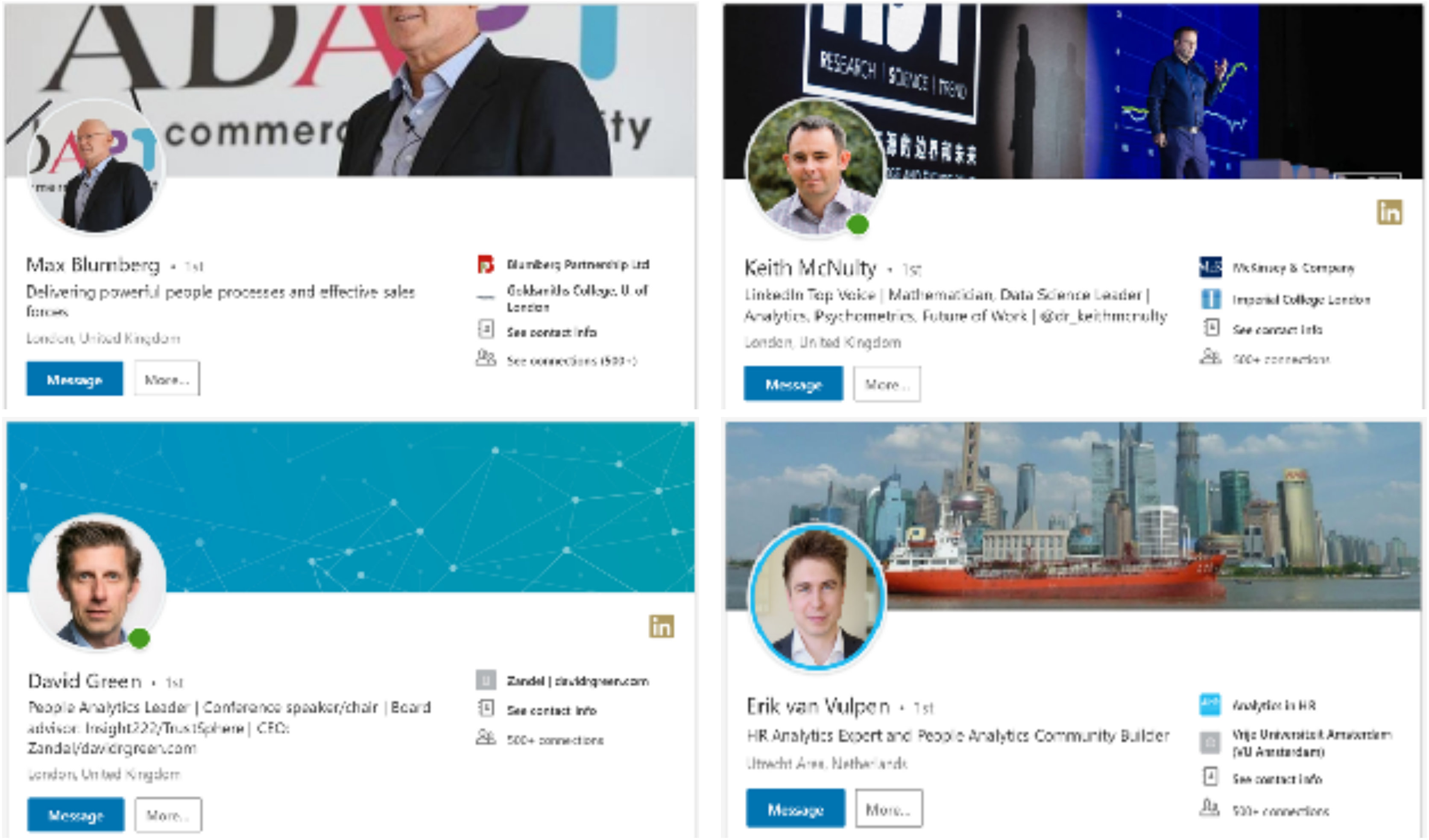
Small Data Problem

Conclusions

- Use algorithms which work well on small datasets, e.g. logistic regression
- Use coarse bins. ctree works for small samples, EqualFrequency is prone to overfitting.
- Be extremely cautious for datasets smaller than 1000 observations.

HR Analytics Resources

Follow these people…



HR Analytics Resources (cont.)

Check these websites…

https://scienceforwork.com/

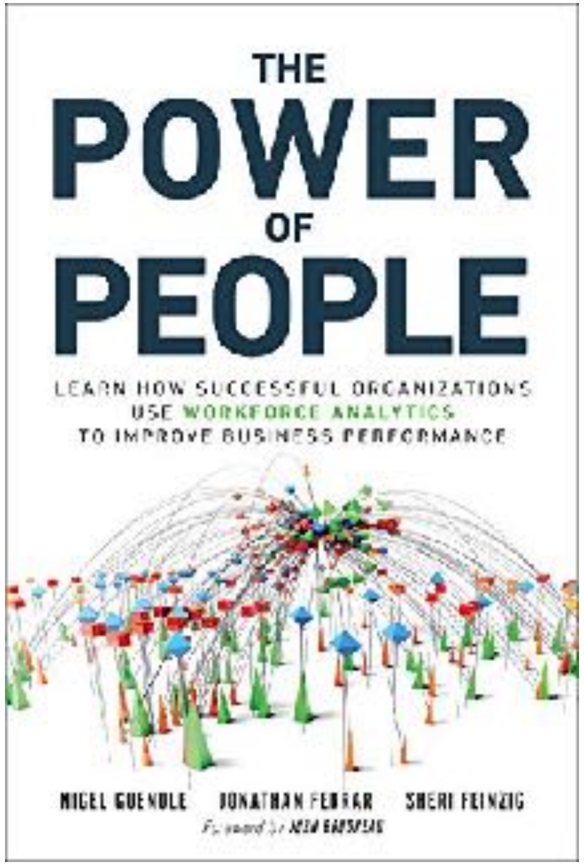
https://www.analyticsinhr.com/

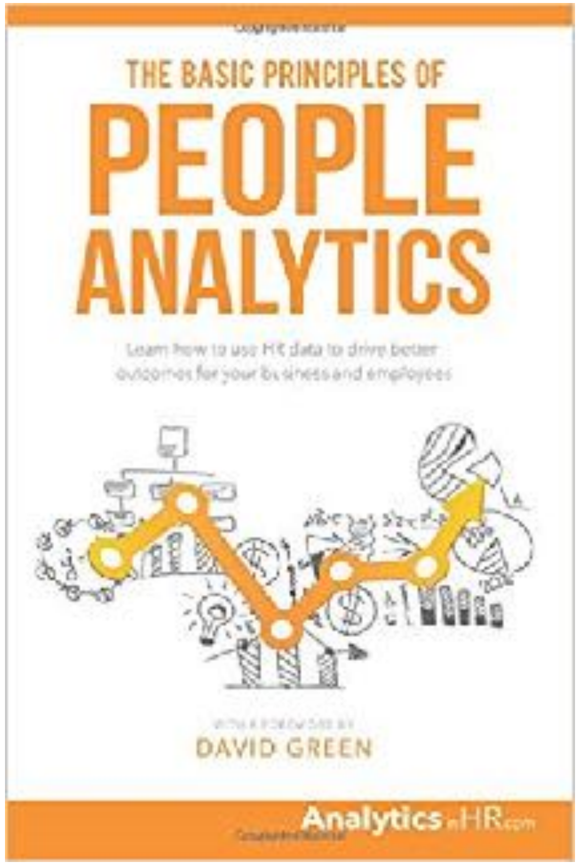
https://www.hranalytics101.com/

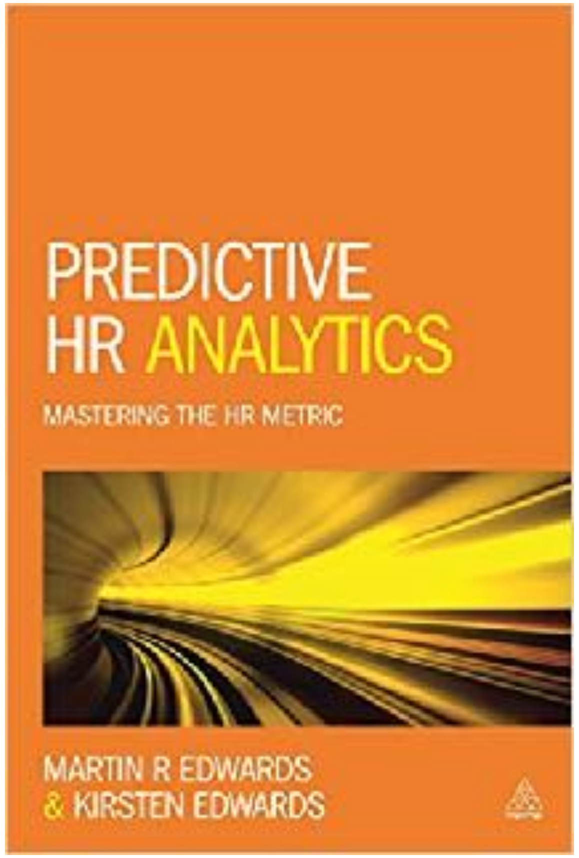
https://www2.deloitte.com/insights/us/en/focus/behavioral-economics.html

HR Analytics Resources (cont.)

Read these books…

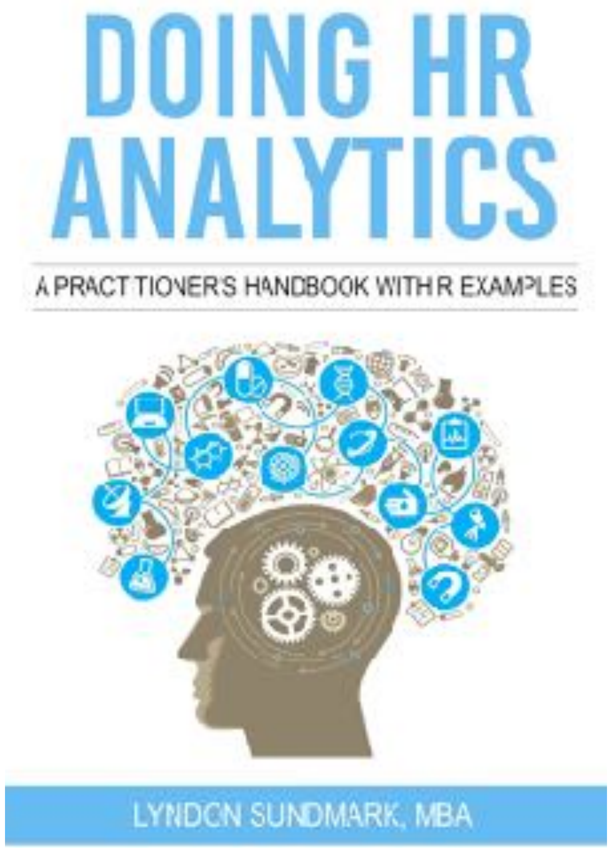


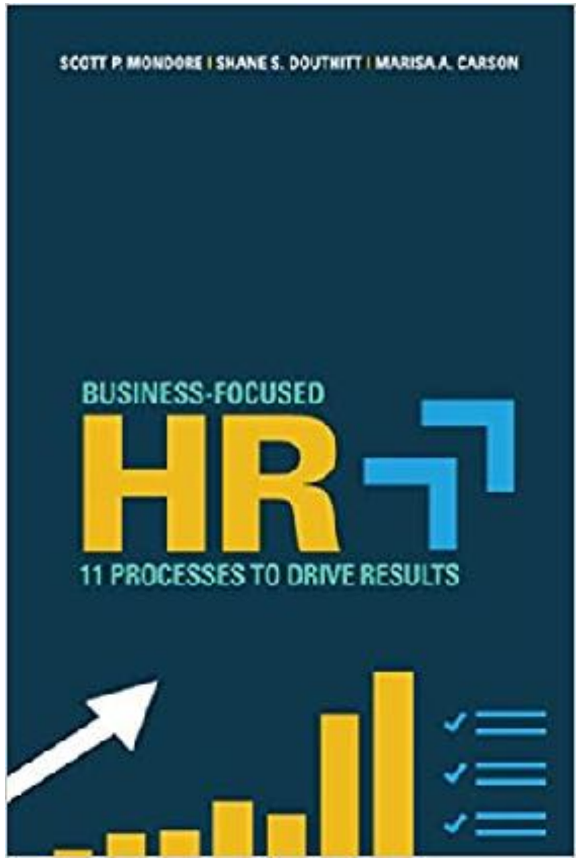


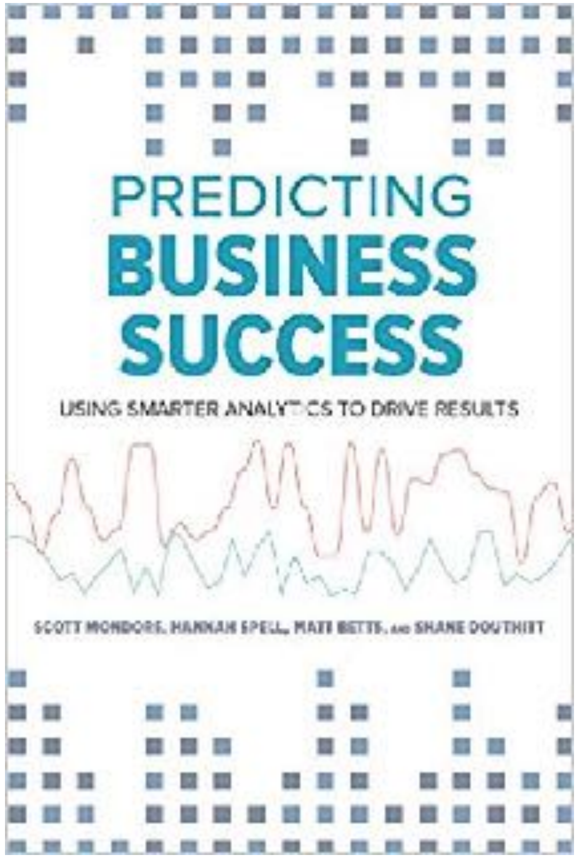


HR Analytics Resources (cont.)

Read these books…







HR Analytics Resources (cont.)

Attend following courses…

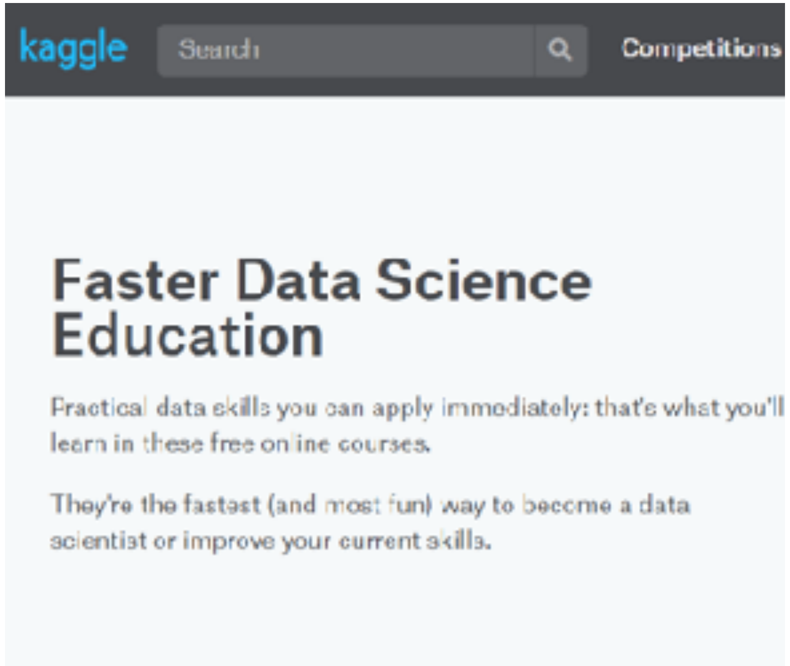
https://www.coursera.org/learn/wharton-people-analytics

https://university.business-science.io/p/hr201-using-machine-learning-h2o-lime-to-predict-employee-turnover

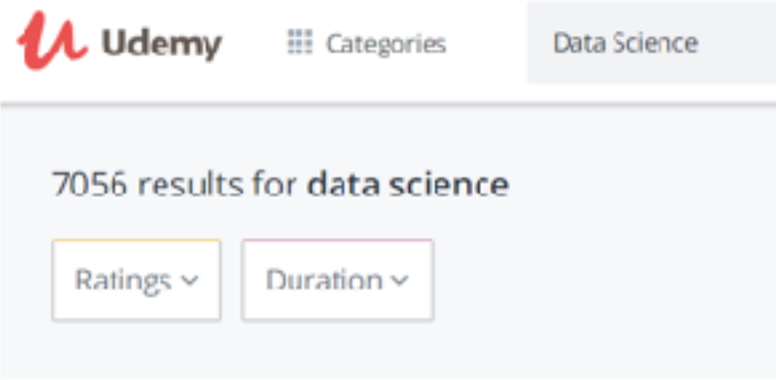
HR Analytics Resources (cont.)

Learn some Data Science basics…

https://www.datacamp.com/



https://www.kaggle.com/learn/overview



https://www.udemy.com/courses/search/?src=ukw&q=Data+Science

HR Analytics Resources (cont.)

Learn following analytical technologies…

Q&A

Thank you.

Filip Trojan
Advanced Analytics