



# End-to-End Proactive Talent Retention Capability using Machine Learning and Advanced Analytics

By: Jonathan S. Androvetto  
September 2023

# What this session is about

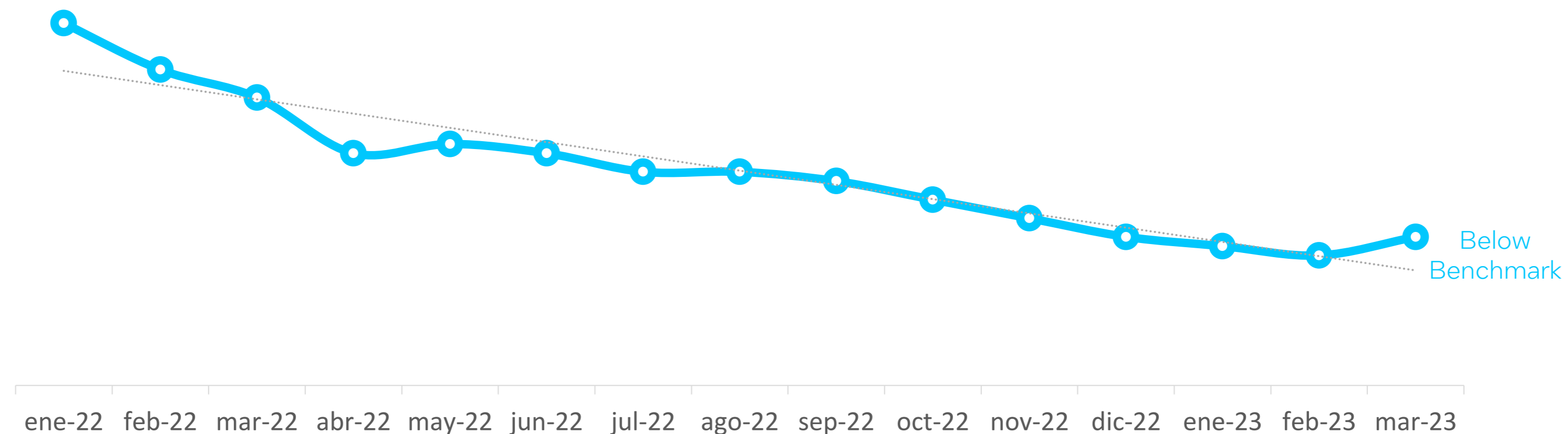
- General thoughts and observations
- Predictive Turnover Capability Architecture
- Methodologies
- Concepts
- Business Examples
- Key Takeaways

# Cost of employee replacement?

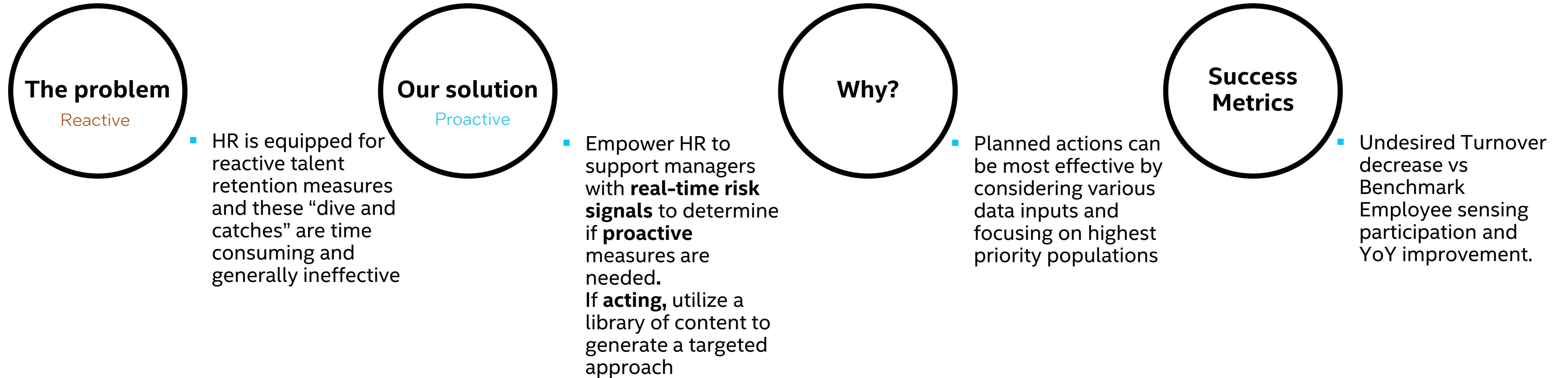
“Direct replacement can cost companies up to [50%-60% of a worker’s annual salary](#), and that is without the indirect costs associated with losing an employee  
These include missed or delayed revenue, and loss of productivity and knowledge”

Source: *Society for Human Resource Management*

The Journey: Year-over-Year **Undesired Turnover Decrease**



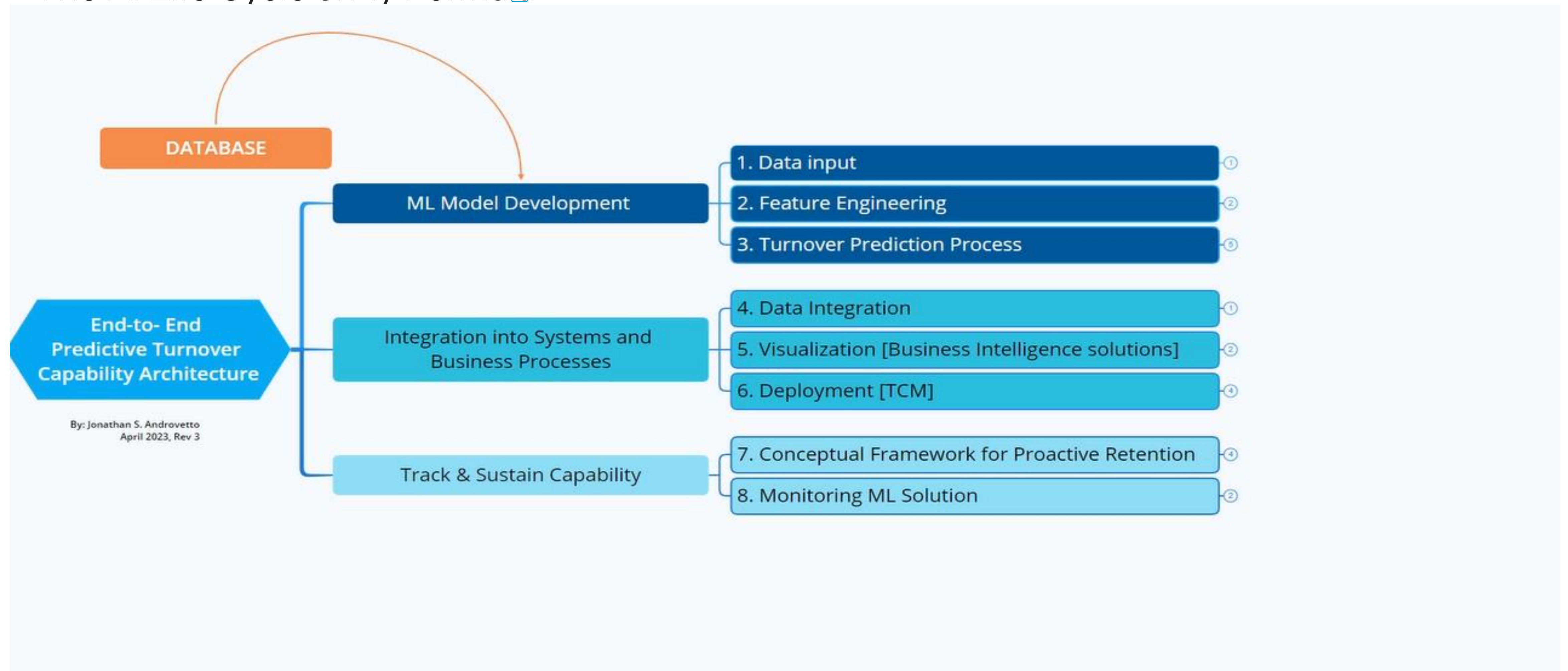
# The opportunity



Strategic Intent – Hire, Progress and **Retain** the best talent of the world

# End-to-End Predictive Capability | Model Architecture

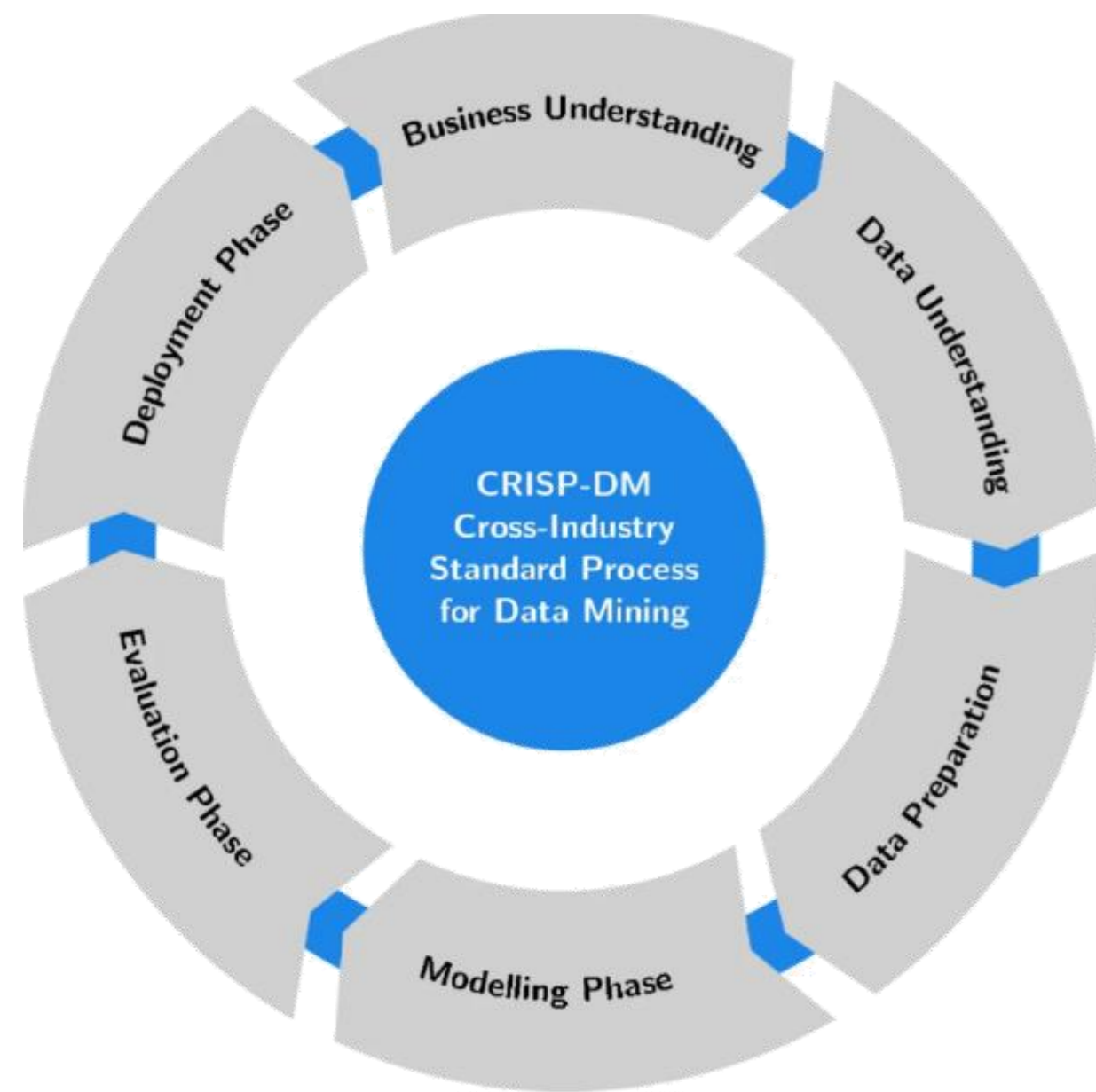
## ■ The AI Life Cycle & My Formula



\* Model architecture is based on my own experience as a consultant

# Turnover Prediction Methodology

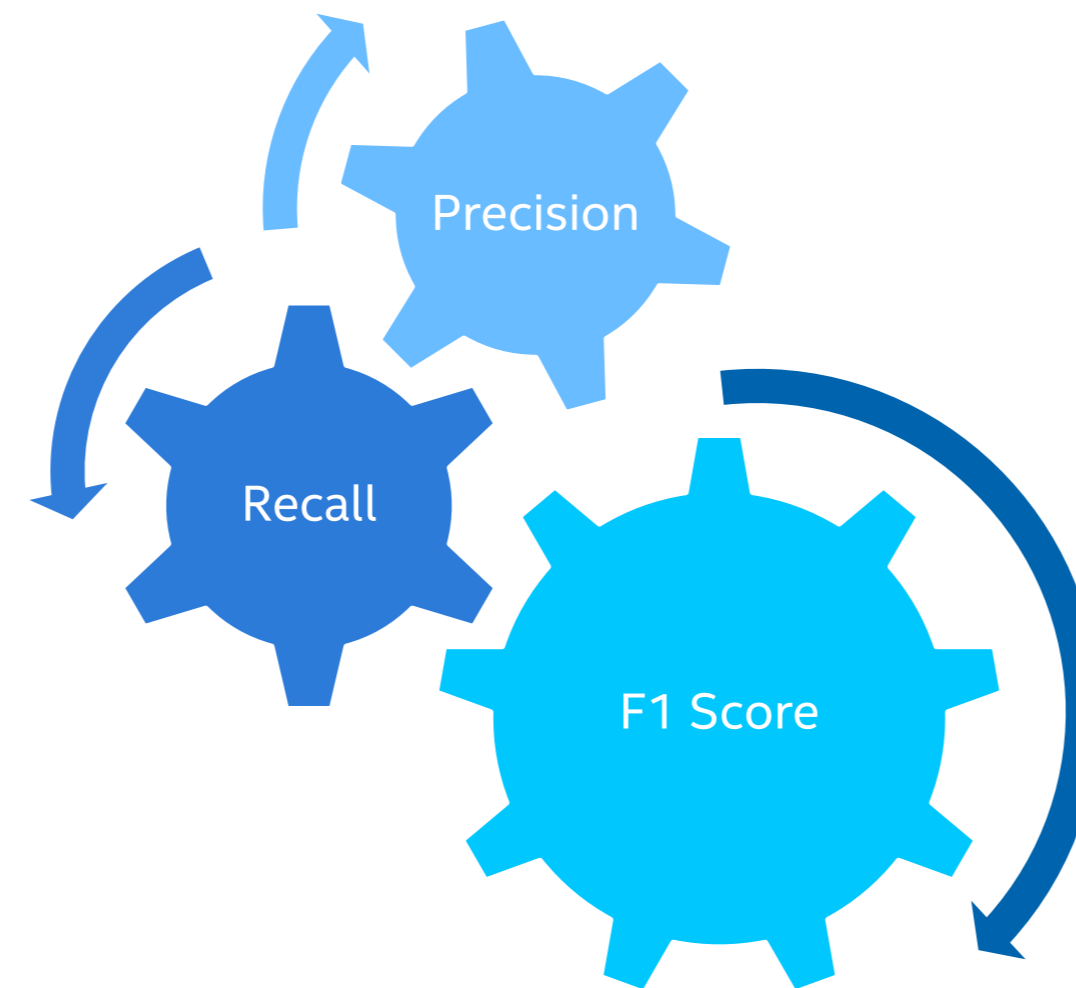
## Industry Methodology



The Cross Industry Standard Process for Data Mining (CRISP-DM) is a methodology that serves as the standard for a data science process

CRoss Industry Standard Process for Data Mining  
[CRISP-DM Guide](#)

## Performance Indicators



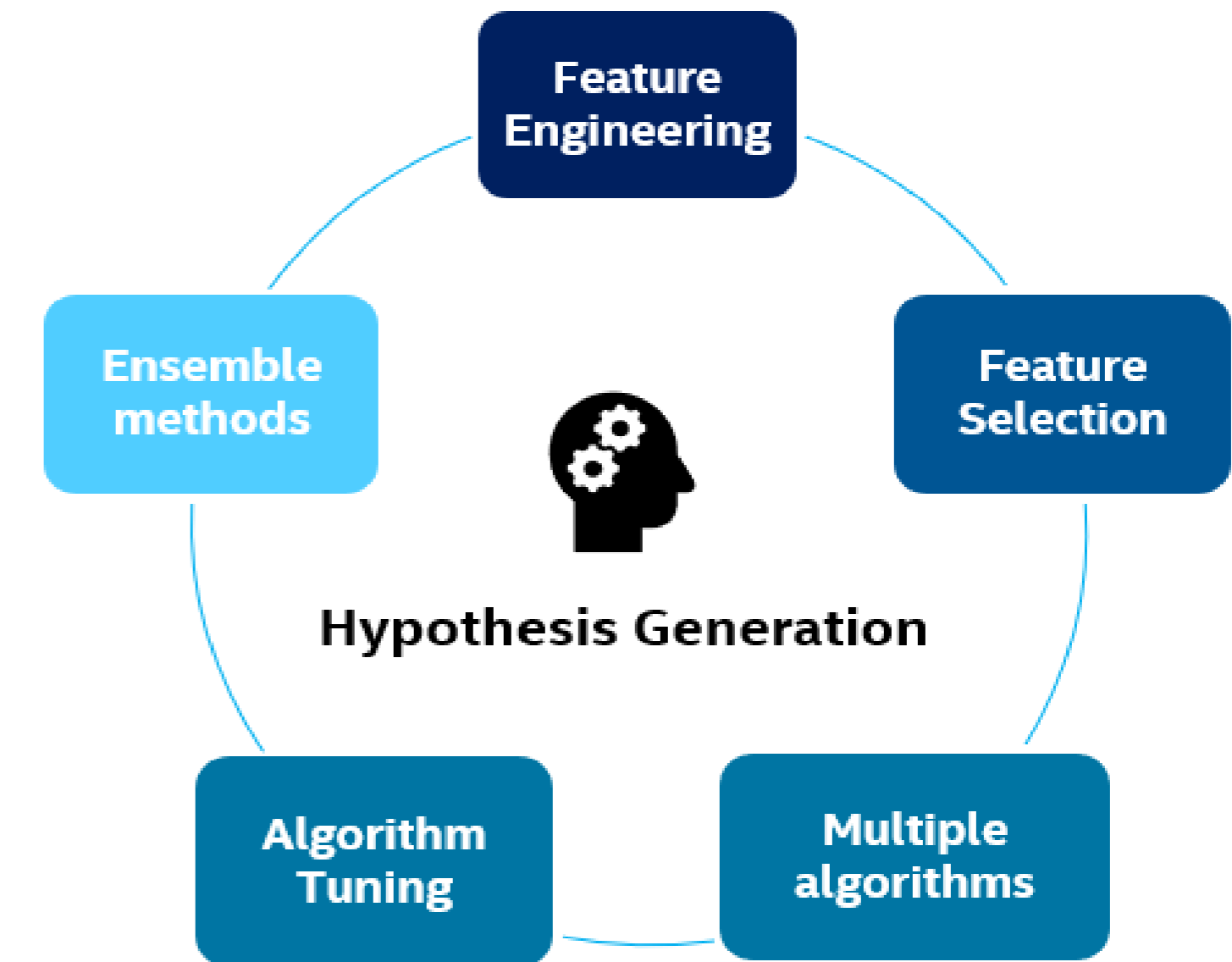
Model Performance  
**80%** F1 Score

**F1 score is a measure of a model's accuracy that combines two metrics, Recall and Precision.**

**\*Recall: ~69%** (the percentage of employees who leave that are correctly identified as a turnover risk by the model. This means the model predicted 69% of actual turnover.)

**\*Precision: ~90%** (the ratio of people identified as a turnover risk by the model who end up leaving vs. those who are identified as a turnover risk but do not leave. This means that 90% of people predicted to leave actually turned over)

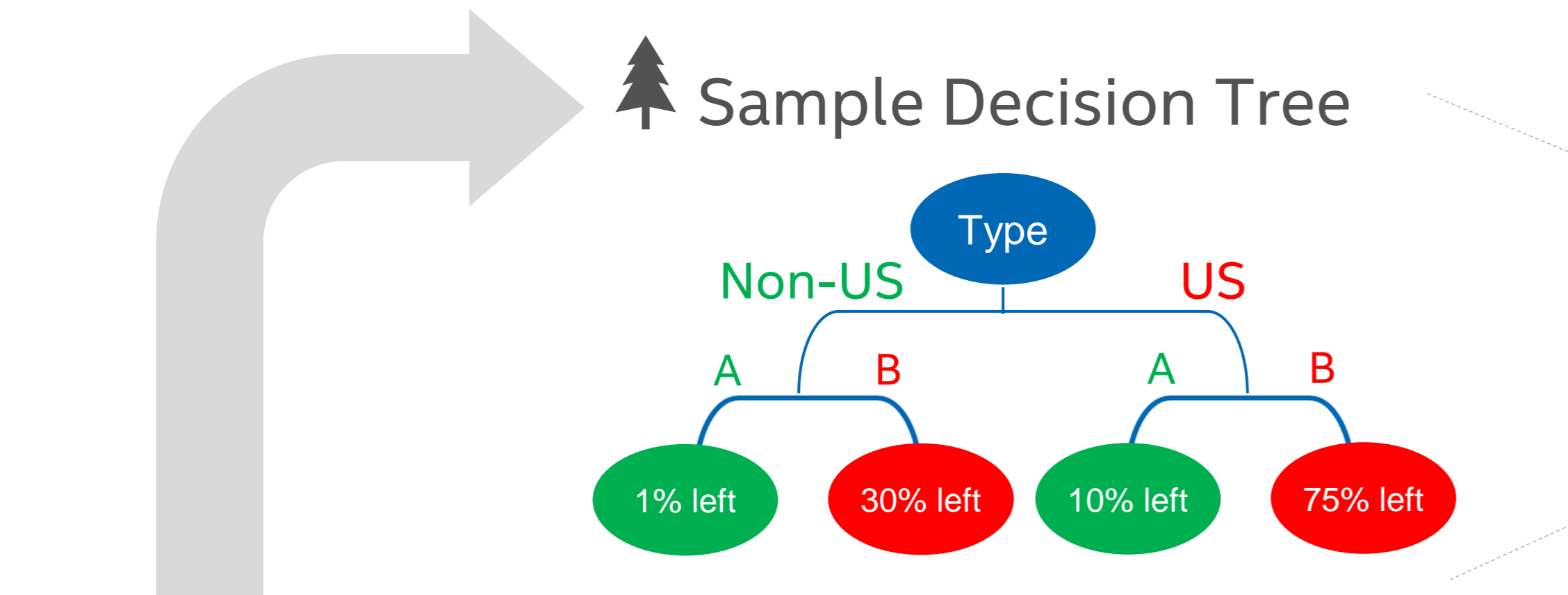
## Model Improvement Methodology



Learning Models such as the Turnover predictor model are not 100% accurate. However, a methodology was designed to support continuous improvement of the model

# How the prediction model works | Hypothetical example

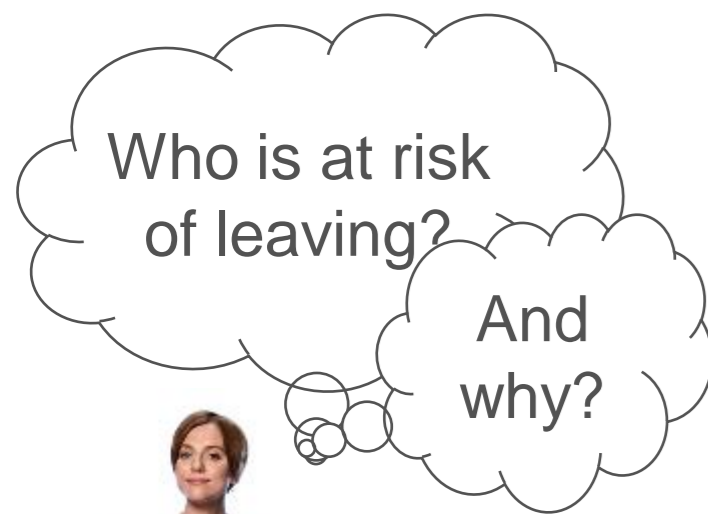
To answer these questions, the model runs 1000s of decision trees with 100s of demographic variables against data sets of terminated employees...



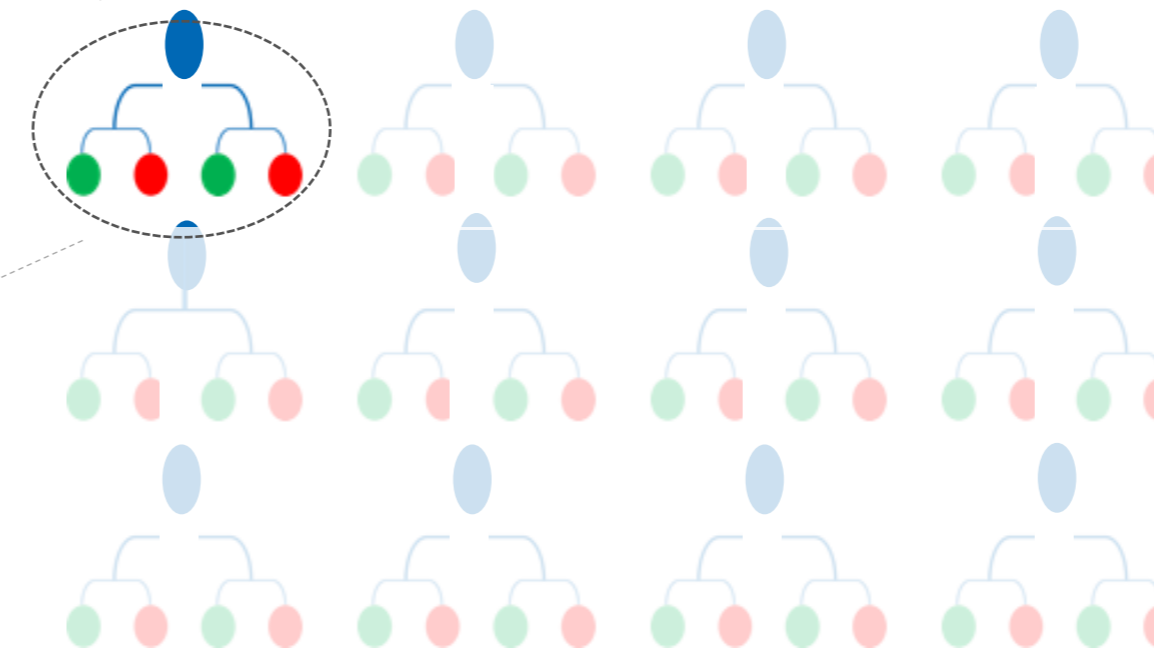
In this example, US B EEs have the highest chance of leaving.

All Active US B EEs will be flagged as "at risk"

The rest of EEs will be flagged as "not at risk"

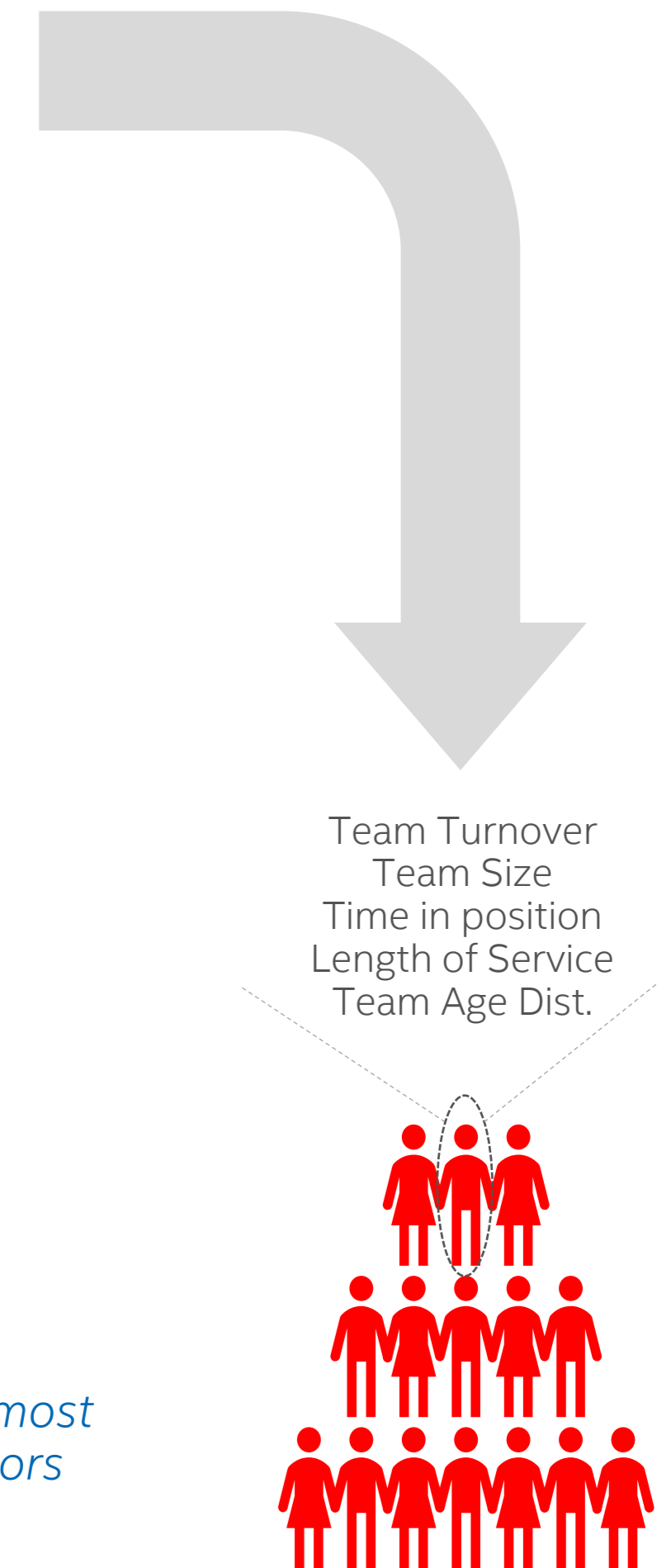


## Gradient\* Tree-based model (e.g., LightGBM, CatBoost, Xgboos)

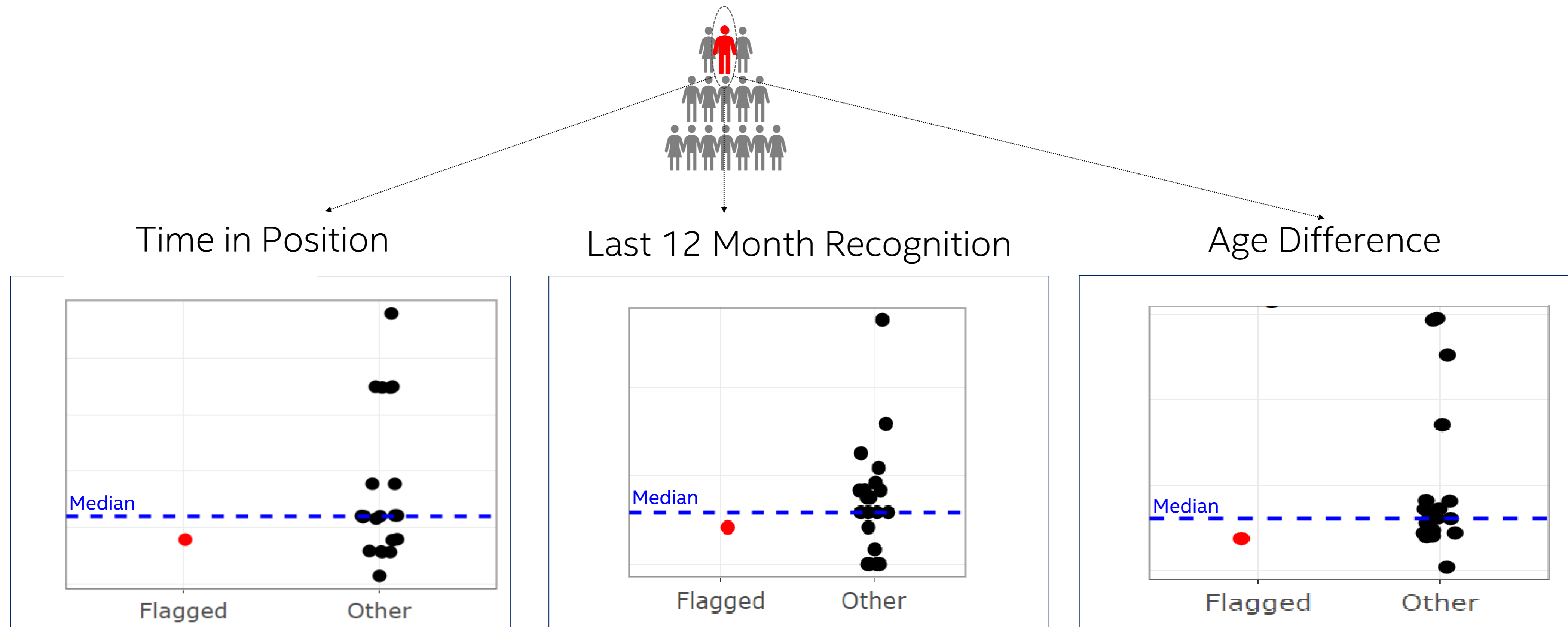


To be thorough, the model runs 1000s of decision trees of terminated EEs with 100s of demographic variables (e.g., seniority, team diversity, org changes, turnover, etc.) in different combinations. It keeps flagging as "at risk" all active EEs that have the same demographics as the terms, even if repeated

...Ultimately, the model produces a final list of Active EE most frequently flagged as "at risk" and the demographic factors associated with them



# Interpretation of signals | Symbolic Visualization



The model produces a series of signals based on demographic attributes identifying active employees that fit characteristics of the training data (employees who have left) and classified as High Risk. The employee identified as High Risk will have several different signals assigned to them (it could even be 5 or more). It is the combination of all the signals that determine the risk. It is not clear if one is more important than another to an individual (e.g., they are not weighted per se). For example, employees on the same team could have very different factors and/or combination of factors. Some factors may appear to be more actionable than others, however they cannot necessarily prescribe what actions to take on their own. The factors should be analyzed with other Turnover data points (i.e., Undesired Turnover trends, External Market Threats, Employee Survey, Ground Intelligence, reg. TO, compensation, etc.) and linked with BHR Programs for actionability.



Keep it Simple!

# Components to consider for | Customer view

## Description of Model Performance

**About this Workbook**  
The purpose of this workbook is to provide insight to the profiles of employees that are at risk of leaving. Attrition risk is based on the modeling of ~200 variables, that include demographics, data from Workday & Recognition data.

Please see [Wiki](#) for additional information and links to microlearnings that cover Privacy, Use Cases, Prediction Model overview, and Tool demo.

**Privacy Notice**  
The data in this report is classified as Intel Confidential (IC) and is required to be safeguarded under data protection laws and Intel's privacy principles. Refer to the Global Employee and Contingent Worker Privacy Notices for information on Intel's privacy practices.  
\* The names generated by the turnover predictor model should only be used for use cases that have positive intent/impact to an employee or group  
\* BHR can share insights generated from the Power BI tool, including employee names, with relevant business leaders and managers for retention purposes only. When sharing names, avoid any additional personal information unless there is a relevant business need aligned to specific retention actions  
\* When sharing personal information with the business, it is best to share your screen vs. send over email  
\* Only authorized HR personnel should have access to the Power BI tool. No data should be shared directly with the business unless first approved by BHR  
\* Personal Data must be collected and retained on an approved enterprise platform (ex: MS Teams, SharePoint) with appropriate access restriction. Reference page 2 of the [Privacy & Email Standardization documents](#).  
\* Delete the personal data when no longer needed for the purpose(s) and no later than the expiration of the retention period in [Data Retention](#).

**Cautions/Warnings**  
The dashboards are intended for analysis at an aggregate level. THE DATA IN THIS WORKBOOK SHOULD NOT BE CONSIDERED SYSTEM OF RECORD FOR INDIVIDUAL EMPLOYEE DATA. The data may not perfectly align with other sources, in part due to the day the data was last updated.

**Model Performance**

80%  
F1 Score

**F1 score is a measure of a model's accuracy that combines two metrics, Recall and Precision.**  
\* **Recall:** ~69% (the percentage of employees who leave that are correctly identified as a turnover risk by the model. This means the model predicted 69% of actual turnover.)  
\* **Precision:** ~90% (the ratio of people identified as a turnover risk by the model who end up leaving vs. those who are identified as a turnover risk but do not leave. This means that 90% of people predicted to leave actually turned over)

**Intel Model: Conceptual Example:**  
\* The predictive model correctly identified 69 of 100 employees who left Intel. That's 69% Recall. However, the model identified 77 people at risk of leaving Intel. That means included 8 employees who did not leave Intel. That means ~80% precision (i.e., 69 out of 77 were correctly predicted).  
**The F1 score is simply the harmonic mean of Precision and Recall**

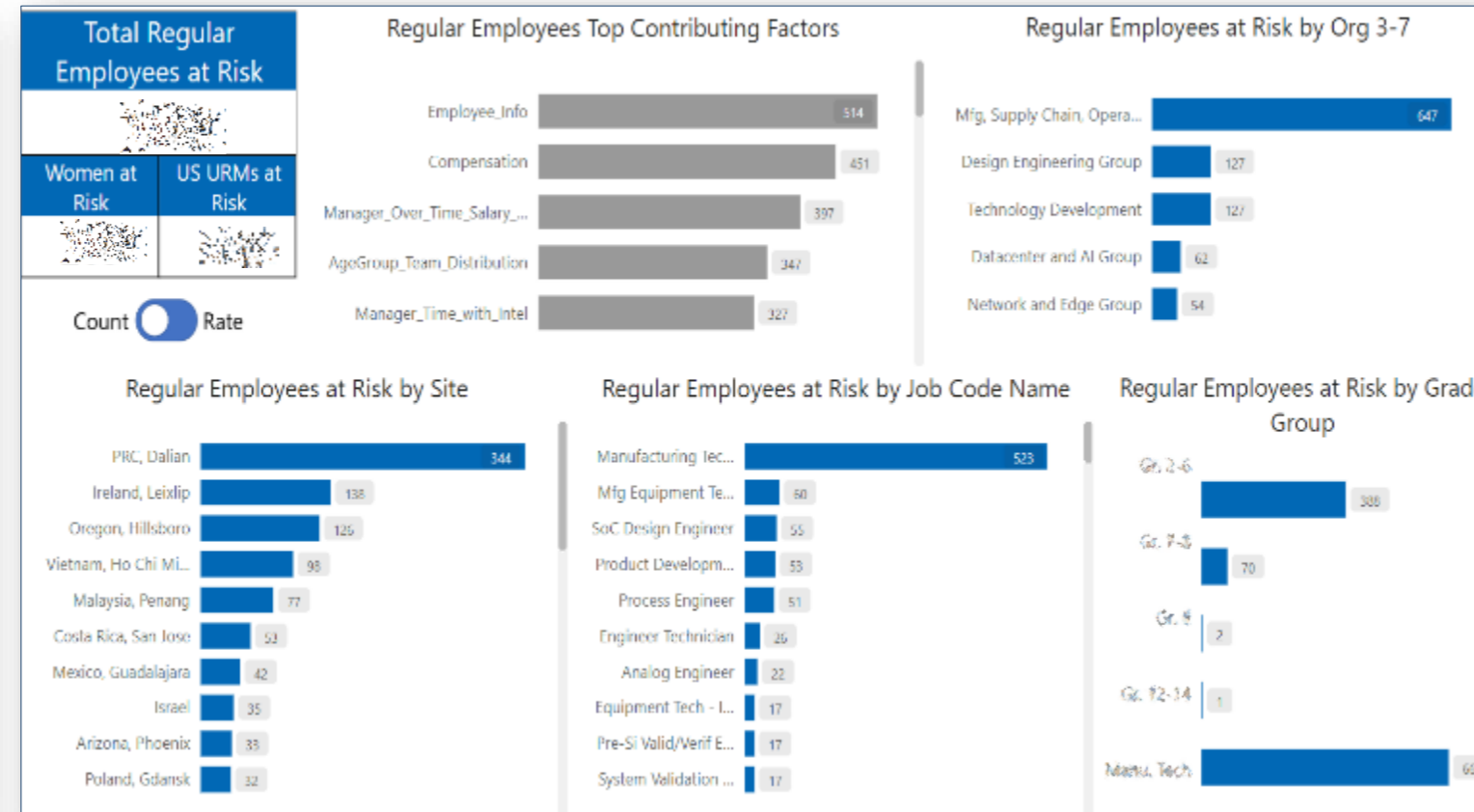
**Please note, there is variability between populations.**  
F1 scores for Organization or Location cuts are not available at this time. 80% is the F1 score for all of Intel.

\* The model output is refreshed monthly.  
\* Employees in this workbook are **active regular** employees at risk.

Select an Organization

- Executive Office (22790)
- ACOE (82358)
- CEO TA (95253)
- CLIENT COMPUTING GROUP (11374)
- Corporate Strategy & Ventures (367)
- Datacenter and AI Group (39634)
- Design Engineering Group (10294)
- FINANCE (62824)
- Global Mktg (33229)
- Intel Privacy Services (101319)
- Legal Trade & Gov (1176) (14362)
- Mfg Supply Chain Operations (409)
- Mobile (10154)
- Network and Edge Group (102204)
- OSMG (15593)
- Software & Advanced Technology (14690)
- Technology Development (14690)

## Simple Visualization



## Access to Row-Level Data

WWID	Employee Legal Name	Gender	Ethnicity	URM Indicator	Grade Code (Equivalent)	Job Code Name	Geo Code	Country

**Privacy Legal Reminder**  
URM/Ethnicity is Sensitive Employment Data, which may be subject to more restrictive processing under employment laws. Encryption is required if this information is copied or transferred in any form including screen clippings. Noncompliance may lead to disciplinary action up to and including dismissal.

## Description of Model Methodology

To answer these questions, the model runs 1000s of decision trees with 100s of demographic variables against data sets of terminated employees...

**Sample Decision Tree**

```

    Graph TD
      Root[Grade] --> Non-US
      Root --> US
      Non-US --> A
      Non-US --> B
      US --> A
      US --> B
      A --> L1[1% left]
      B --> L2[30% left]
      A --> L3[10% left]
      B --> L4[70% left]
  
```

In this example, US Gr6 EEs have the highest chance of leaving Intel.  
All Active US B EEs will be flagged as "at risk"  
The rest of EEs will be flagged as "not at risk"

**Gradient\* Tree-based model**  
(e.g., LightGBM, CatBoost, Xgboost)

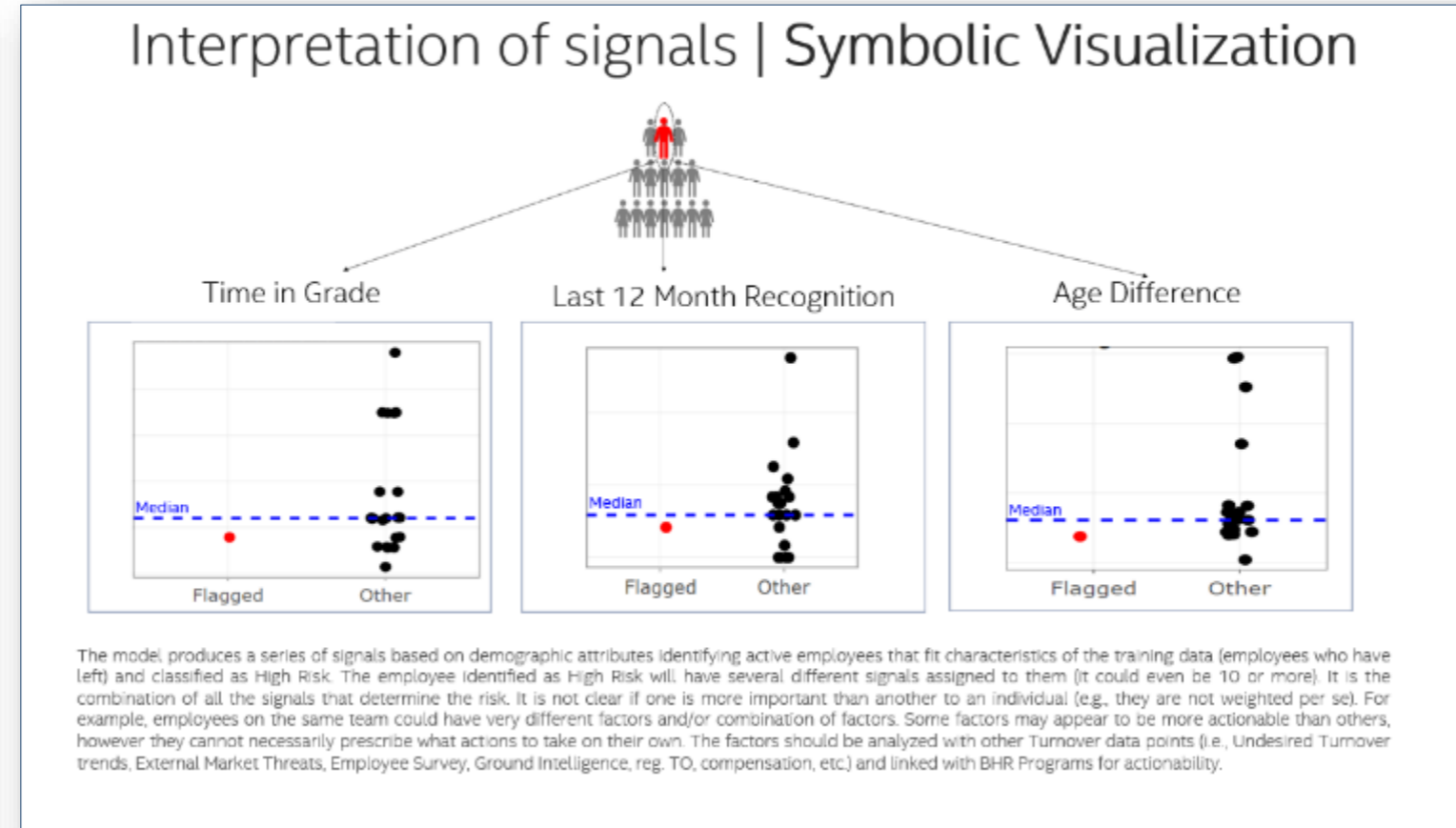
To be thorough, the model runs 1000s of decision trees of terminated EEs with 100s of demographic variables (e.g., grade, team diversity, org changes, turnover, etc.) in different combinations. It keeps flagging as "at risk" all active EEs that have the same demographics as the terms, even if repeated

...Ultimately, the model produces a final list of Active EE most frequently flagged as "at risk" and the demographic factors associated with them

**Who is at risk of leaving? And why?**

**Team Turnover**  
Team Size  
Time in Grade  
Length of Service  
Team Age Dist.

## Description of Interpretation of signals



## Description of Help and Support

**Report purpose and intended audience**  
The purpose of this workbook is to provide insight to the profiles of employees that are at risk of leaving. Attrition risk is based on the modeling of ~200 variables, that include demographics, data from Workday & Recognition data.

**How to Get Access**  
Please see access information in the follow Wiki page:  
<https://wiki.intel.com/pages/viewpage.action?pageId=2292256609>

**Owner**  
HR People Analytics

**Business Support**  
If you have questions or need help with this report, please contact:  
[TurnoverPredictor@intel.com](mailto:TurnoverPredictor@intel.com)

**For further information refer to the following Wiki page:**  
<https://wiki.intel.com/pages/viewpage.action?pageId=2292256609>

A full list of Field Definitions can be found in [Collibra](#), however some of the more commonly asked about ones are below:

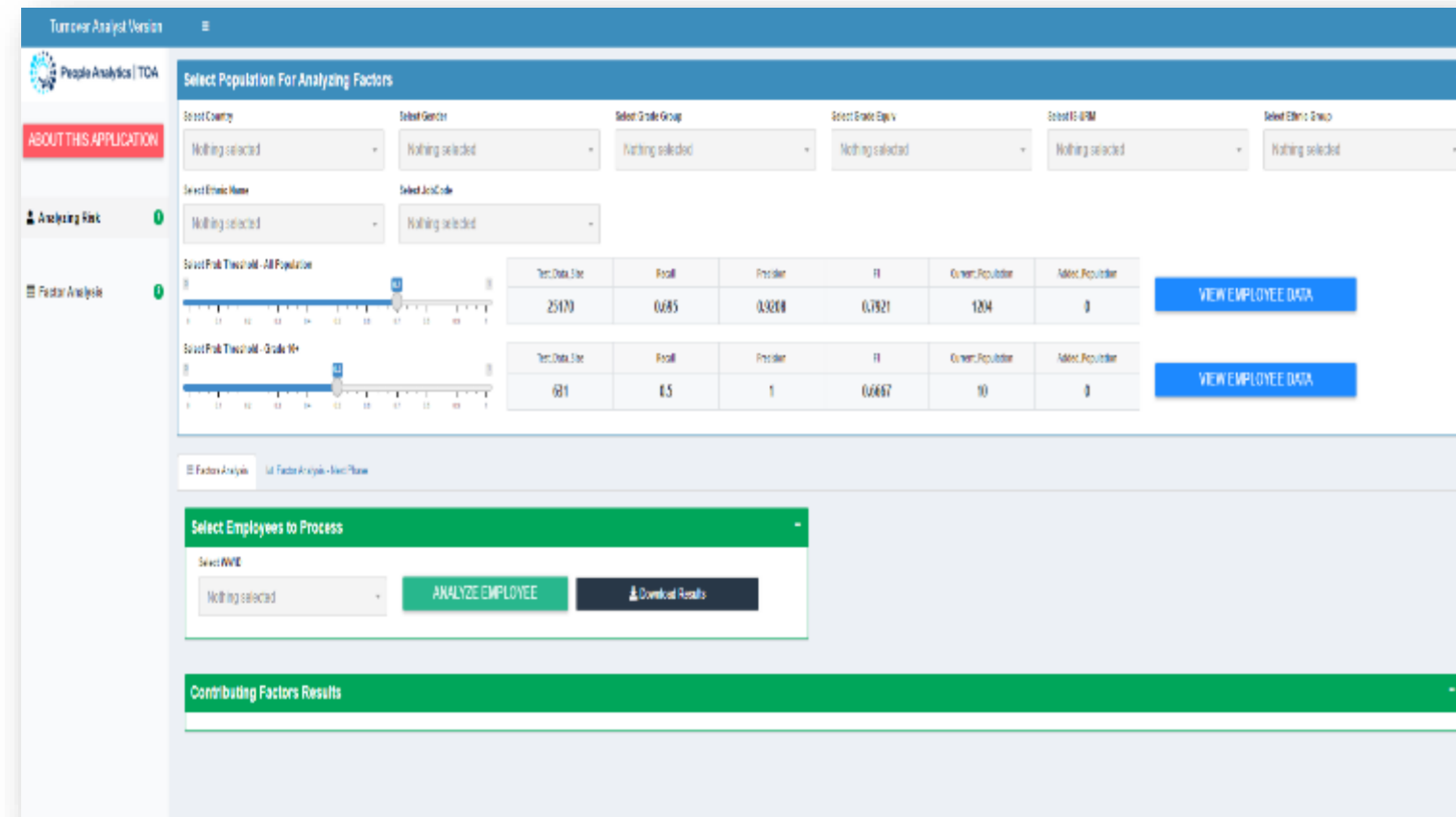
Time in Grade (years)	Length of time in years that an employee has been in their current grade, or an equivalent grade, regardless of changes to job title.
Time in Job (years)	Length of time in years that an employee has been in a particular Job Code.
Time with Manager (years)	Number of continuous years the employee has been reporting to his/her current manager.
Time in Organization (years)	Length of time in years that an employee has been in a particular Org unit.

The purpose of this **BI Solution** is to provide insight to the profiles of employees that are at risk of leaving

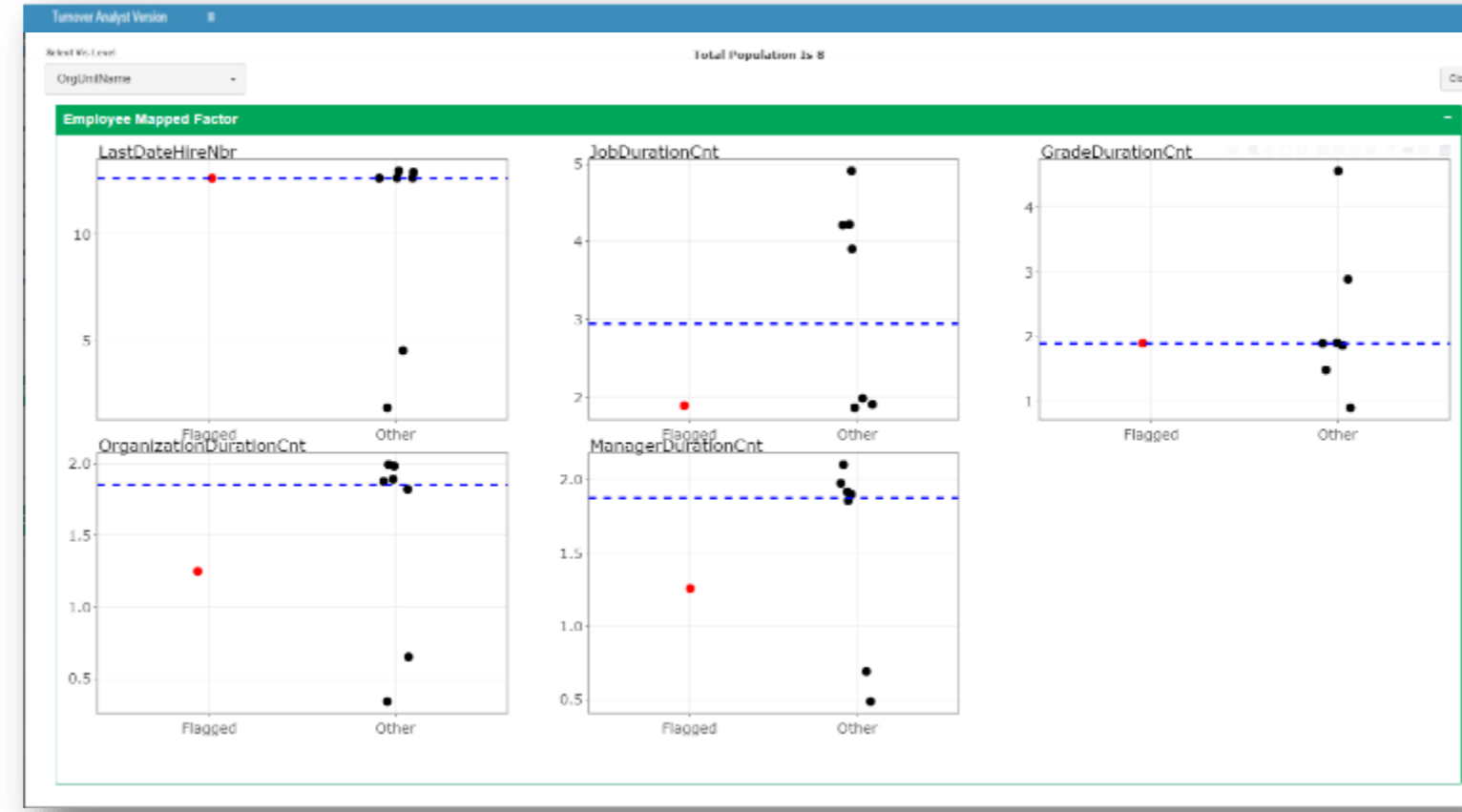
Start Small  
and Iterate!

# Components to consider for | Analyst View

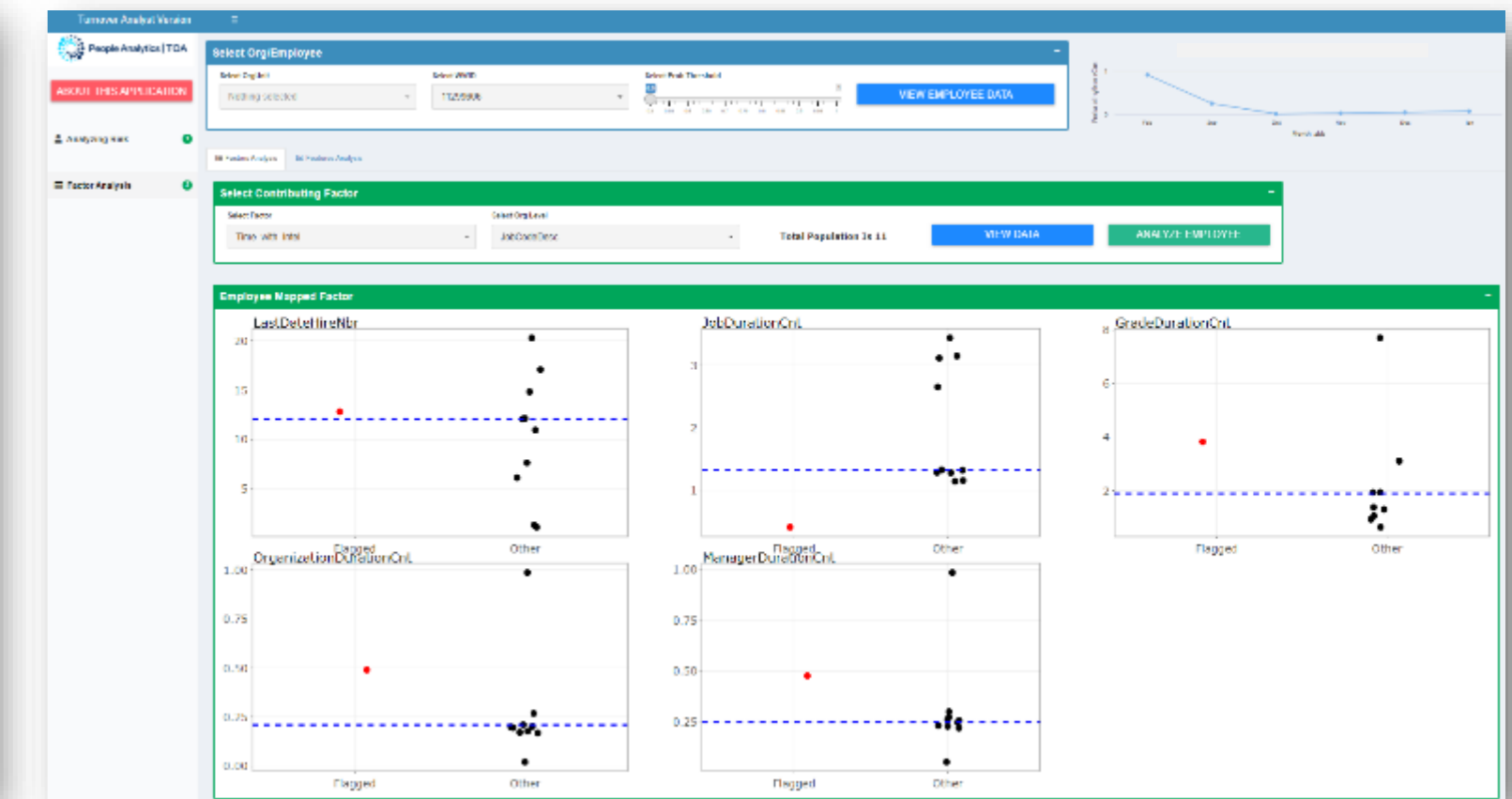
Simulation of Model Performance



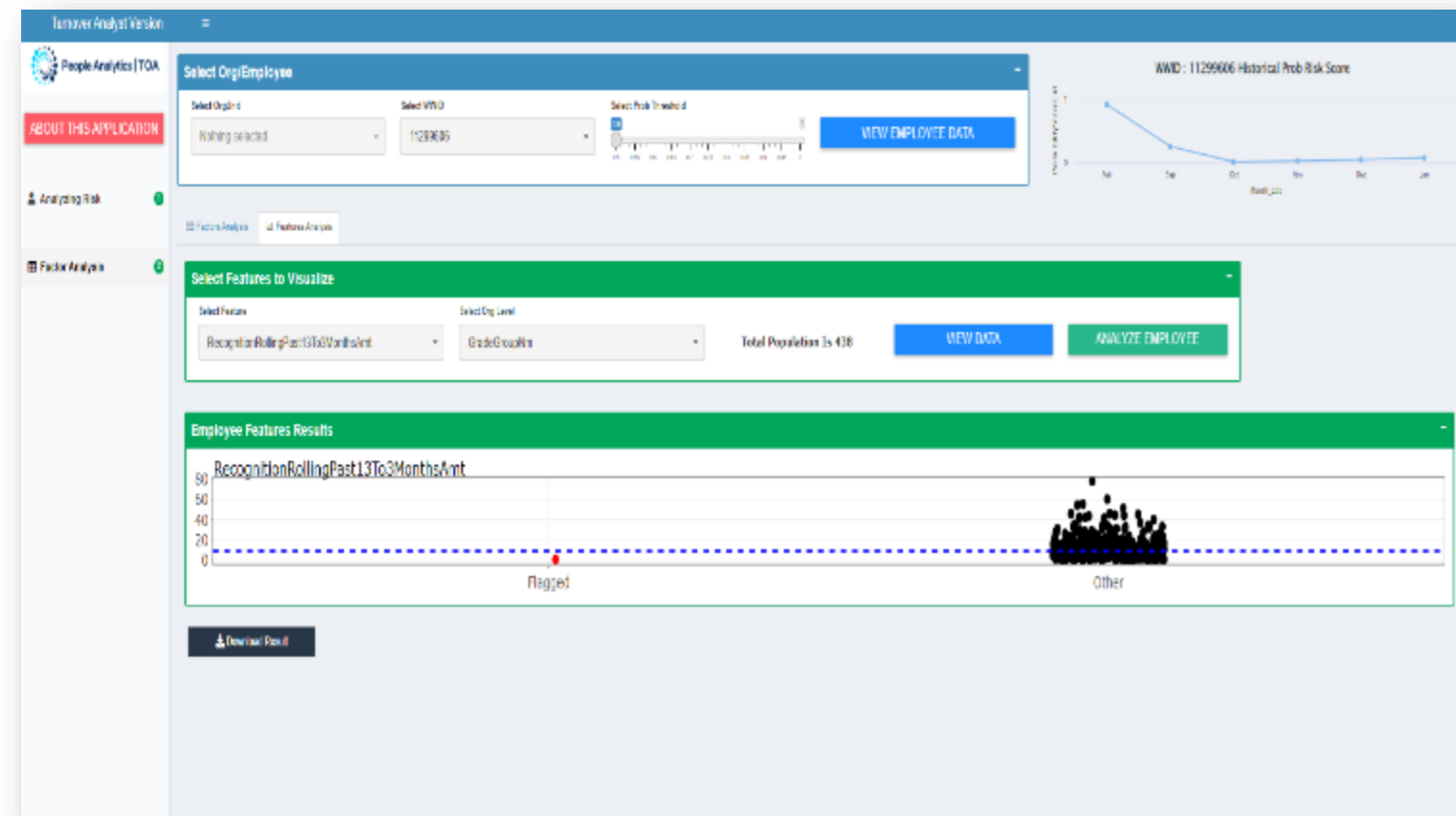
Symbolic Dynamic Visualization



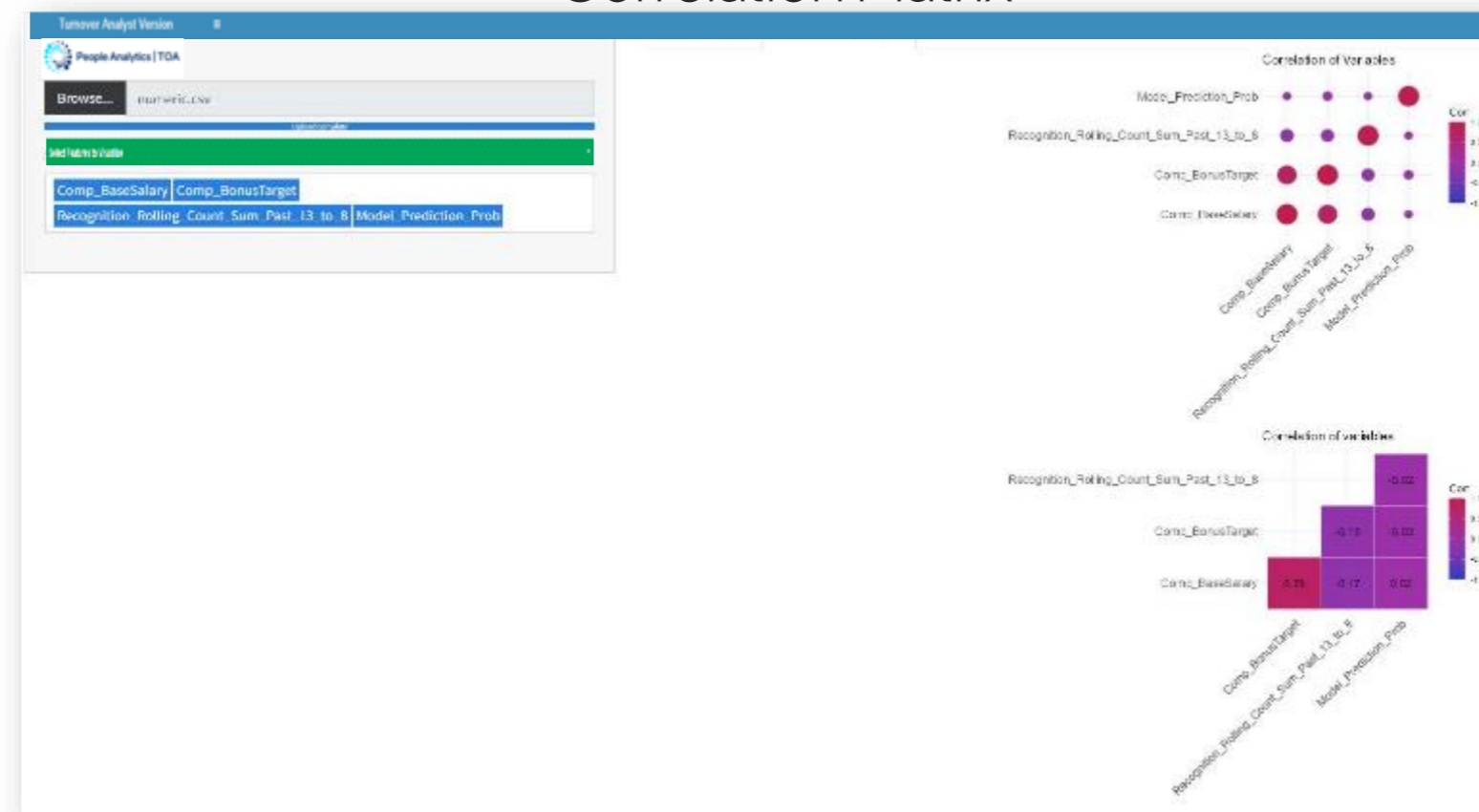
Probability of Risk Trend



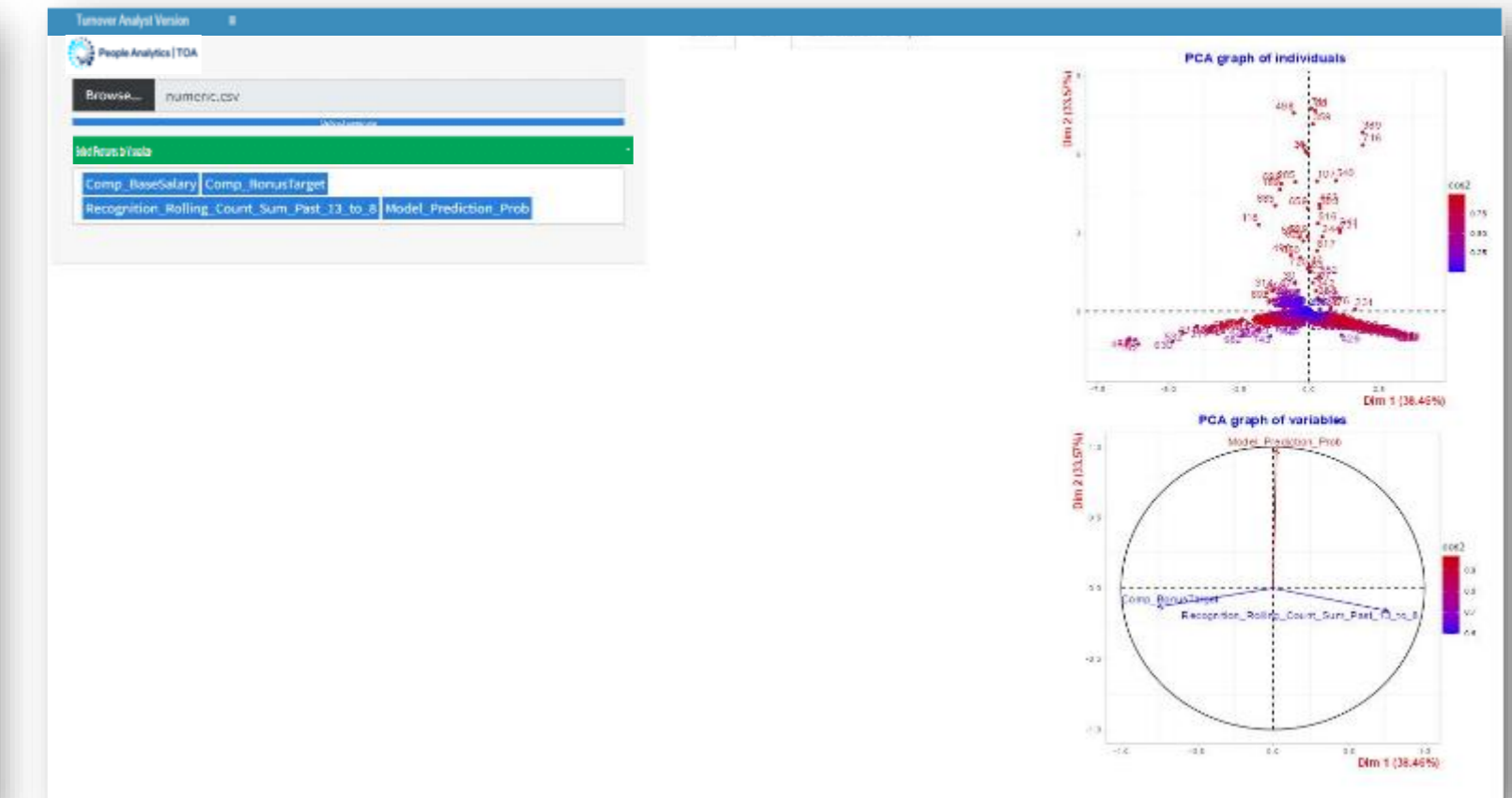
Individual Feature Visualization



Correlation Matrix



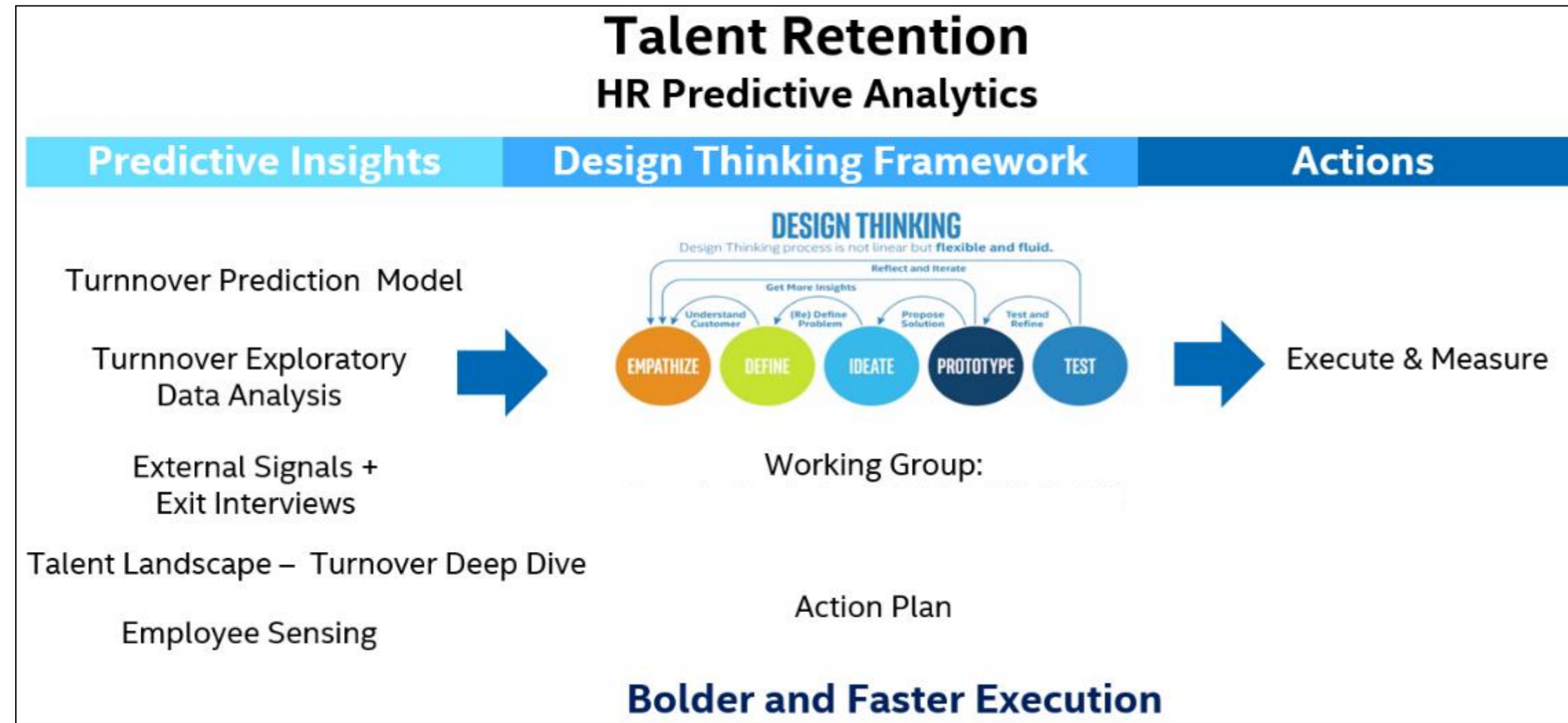
Principal Components Analysis (PCA)



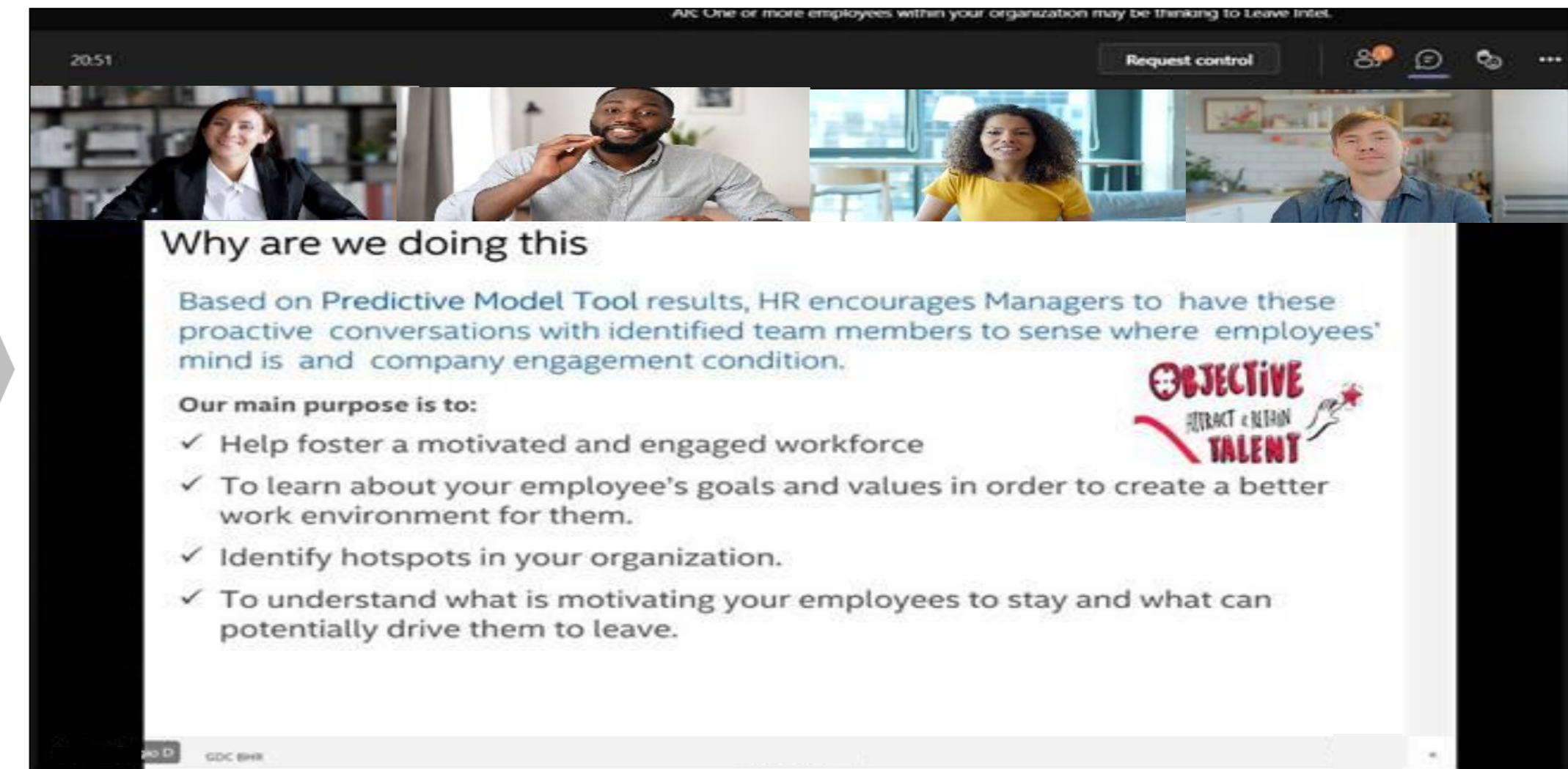
The purpose of this [BI app](#) is to provide a solution to fill gaps customer view doesn't have and facilitates characterization of additional population at risk

# Business Case | Conceptual Example

Methodology: Design Thinking

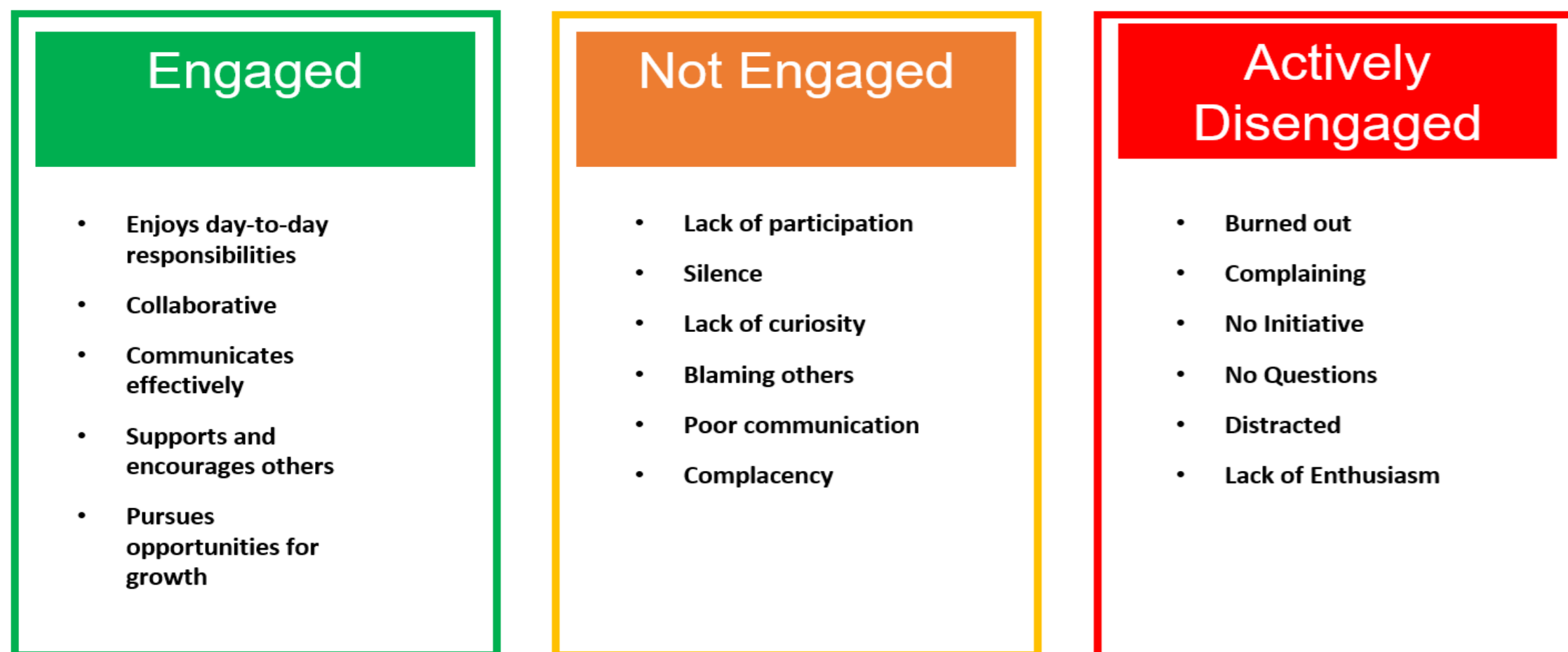


Managers: Predictive Model Intro



Employees at Risk: Manager Assessment

❖ Signs of engagement can be easily determined



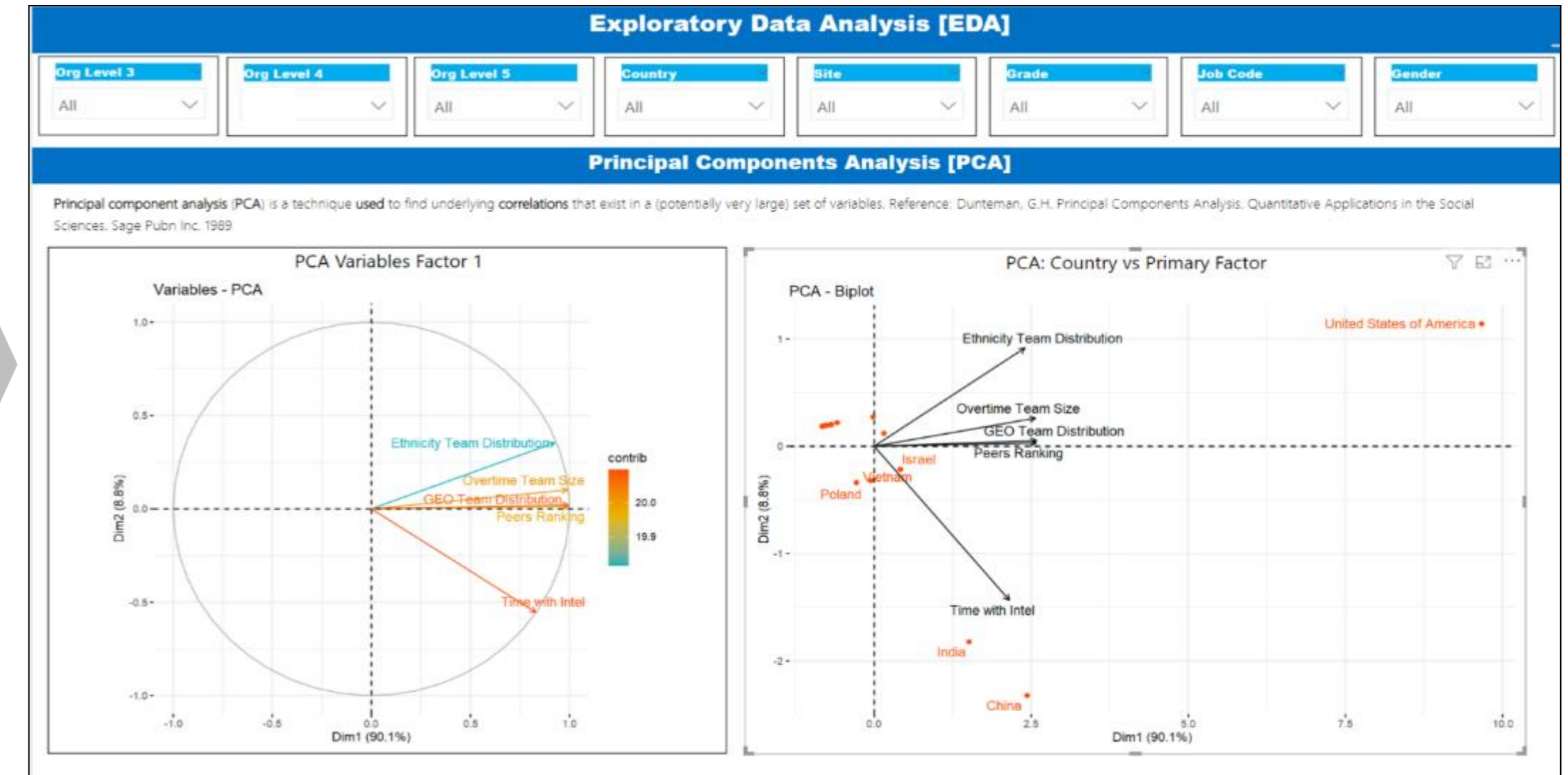
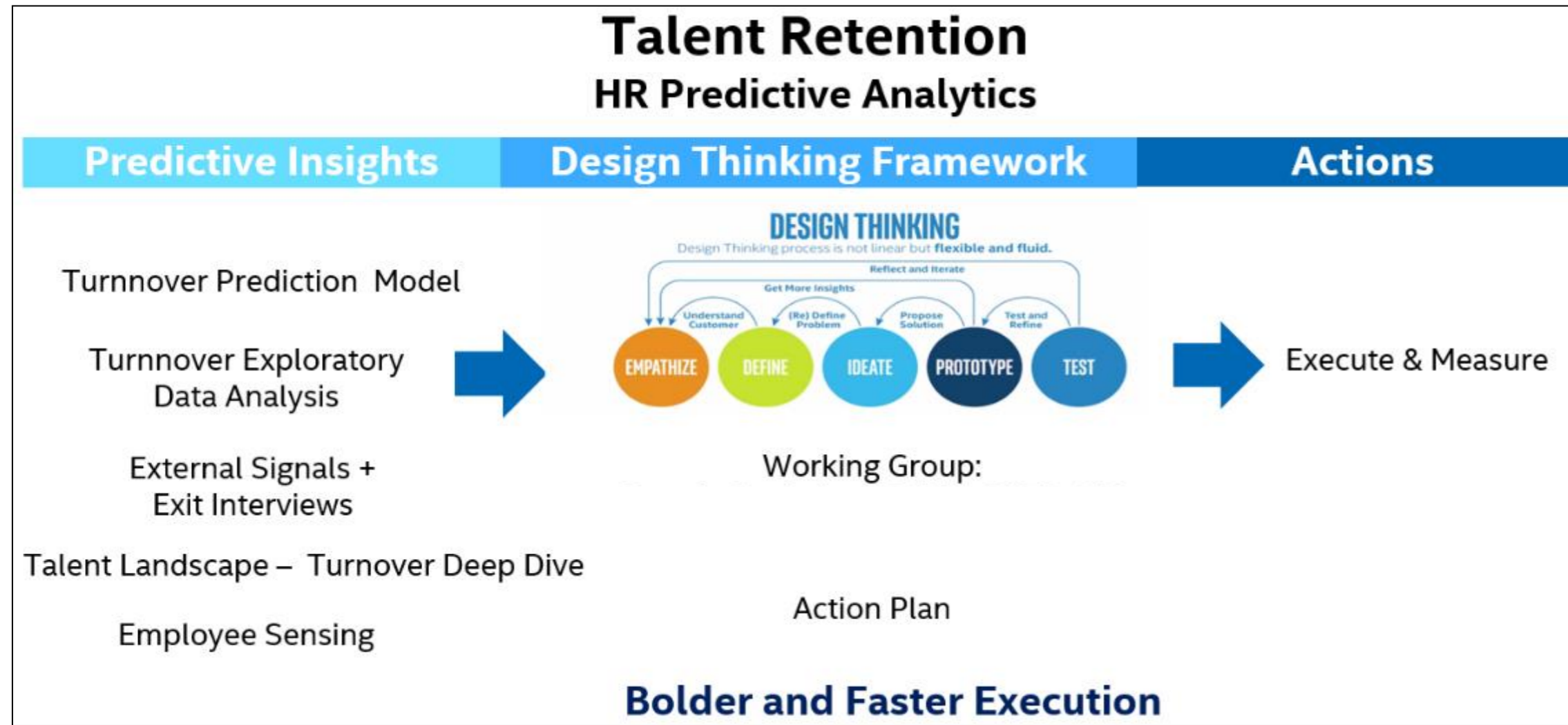
Employees at Risk: Retention Plan

Area	TIG (years)	Compa Ratio	Retention Risk Assessment Q4'2021	Retention/Engagement plans
Pre-Si Valid/Verif Engineer	7	111%	g	To keep his current engagement: Employee mentioned that this year (since I started as his manager) he felt more support and trust from his manager. So, I have to keep encourage him to show his work, providing him challenging work and listening to his ideas. I would also try to find some time in the schedule to let him propose how we could impact the project with formal work.
Pre-Si Verif	1.9	113%	g	To get him on board: Employee mentioned that at the beginning of the year, that he needed promotion so, I have been giving him challenging work and had been observing how he behaves in certain situations/situations, to make sure he has what is needed to be performing in the next grade (while analyzing his previous deliverables and collecting peers feedback). It's been a while as work and, before we had this conversation with HR, I have been showing his work to my manager and peers and, we had decided to pursue with promotion for him. I think we have all the requirements so this looks promising. Maybe this promotion comes late but, my plan is to continue following his technical grow very close so he can grow and show his skills.
PAIV/SPV	1.05	94%	g	Keep a closer communication. Help her to prioritize and plan for get noticeable progress on her tasks. Start working on people skills, so she can be ready to be a manager in the future.
Software Integration Engineer	5.5	104%	o	He will meet in a month. He is interested in a full remote position.
IP Logic Design Engineer	0.8	91%	g	Continue having challenging assignments and opportunities to lead for Samed. If ramping up new lines, owning Slice Handler component and others to explore as next Media IP program gets defined. Ensure Samed gets a mentor (first option Khalil B. but may explore others due to availability) by Q1'22 or earlier. Focus on helping to keep team's critical mass.
SW Application Engineer	1.9	107%	g	Maintaining this engagement conversations quarterly.
SW Application Engineer	0.8	91%	o	On medical leave.
IP Logic Design Engineer	0.8	105%	g	He is very motivated in his new role and particularly with recent salary bumps he is all in with Intel and Intel's strategy. I agree he is a flight risk to his previous team and that it actually happened, but in this new adventure he is committed.
Pre-Si Valid/Verif Engineer	1.7	96%	g	Upon return to Media IP provide opportunities to own cluster level verification.
Pre-Si Valid/Verif Engineer	3.3	105%	o	Upon return to Media IP provide opportunities to own cluster level verification.
Software Application Development Engineer	4.2	105%	o	1) Review options for a Career Development workshop to define career goals. 2) Explore more ways to apply automation to his work.
Cloud Application Development Engineer	1.6	93%	o	Give timely and effective feedback. Increase the frequency of engagement conversations to catch issues early. Coach Roberts to rebuild trust and credibility from the team, this piece is key to keep him engaged.

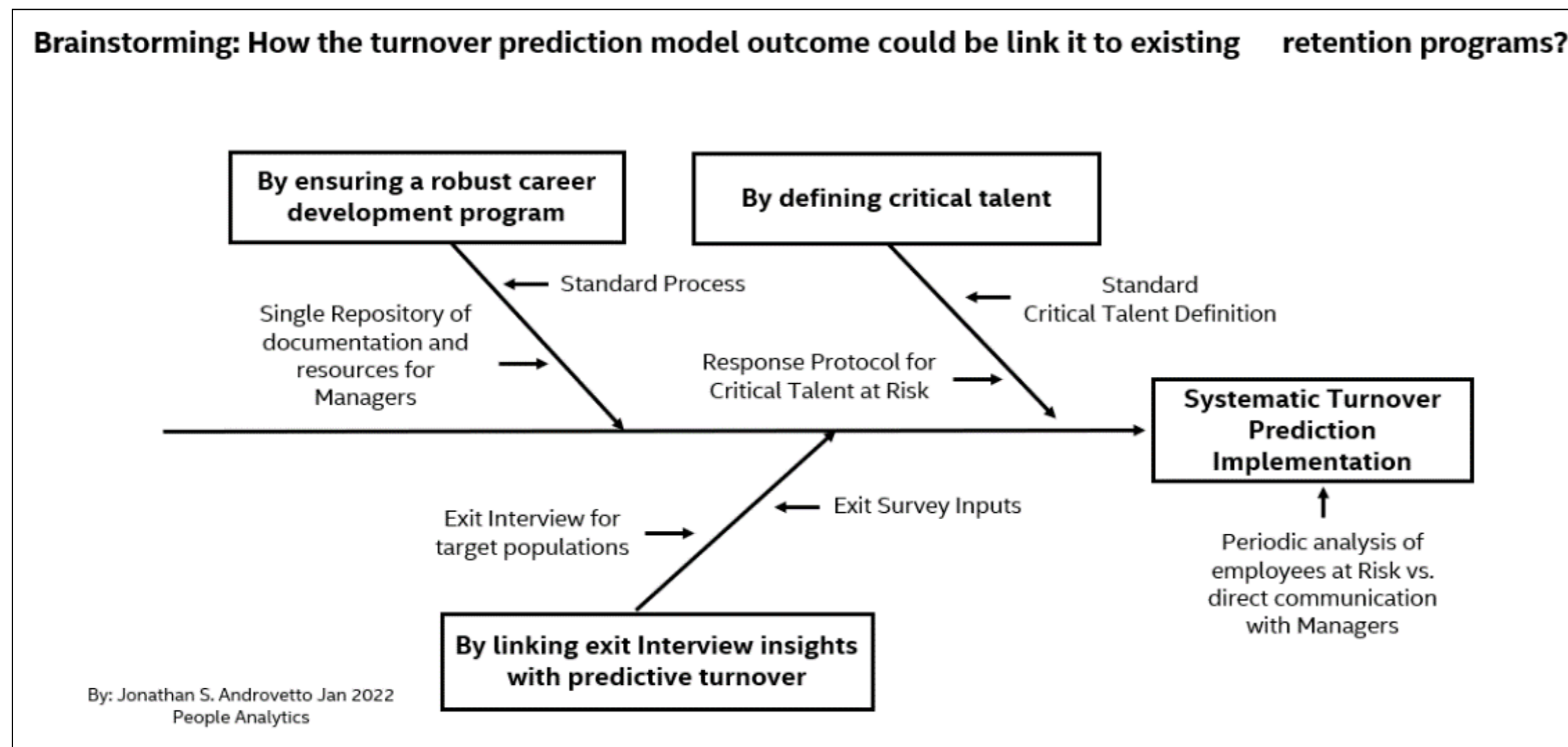
# Business Case | Conceptual Example

Methodology: Design Thinking

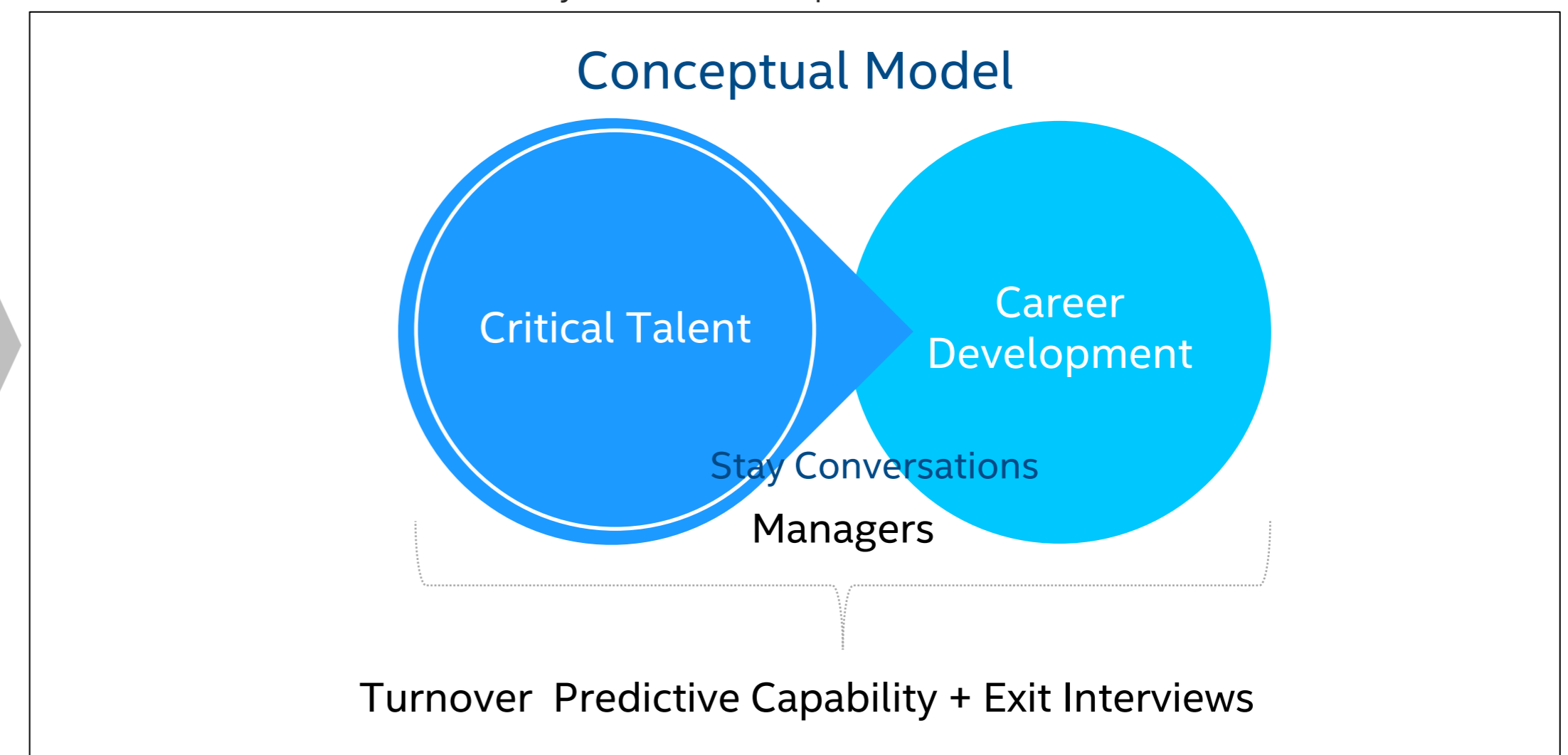
Exploratory Data Analysis



Prediction Model Results vs Retention Programs



Systematic Implementation



# Key Takeaways



## Data Protection Impact Assessment

### Required Expertise in a ML Project

**A verity of skills are required – Assemble a team with all required expertise**

Computer Science/IT

Math and Statistics

Domains/ Business Knowledge

**Prioritize delivering a working solution over a perfect one  
perfection can be a time-consuming**

**Don't aim for perfection on the first try**

**Keep it Simple!**

By adopting an interactive approach, you can quickly deliver value to the business, while continuing to improve  
Monitoring ML Solution – So you won't get surprised

**Start Small and Iterate – Remember that ML is just a small part of the end-to-end Solution**

# Thank you & iPura Vida!

By: Jonathan S. Androvetto [LinkedIn](#)

