

Employee Survey Sentiment Analysis

Nikhil Gupta

Background

- 1. Company XYZ conducted a survey of 18,999 employees
- 2. 60 questions (rating from 1 to 6)
- 3. Employee sentiment across 14 dimensions
- 4. **Goal:** Gain insights from survey.

Manager (8)	Development Opportunities (6)	Engagement (6)	Teamwork (6)
My Work (5)	Inclusion (4)	Total Rewards (4)	Action Planning (3)
Customer Focus (3)	Diversity (3)	Ethics and Integrity (3)	Processes & Resources (3)
	Senior Leadership (3)	Workplace (3)	

Agenda

Overview: Data and Analysis

Part 1: Diversity and Inclusion

1. Does response to “Diversity” and “Inclusion” vary by **Gender**?
2. Does response to “Diversity” and “Inclusion” vary by **Race/Ethnicity**?
3. Next Steps to address findings

Part 2: Factors Influencing Attrition

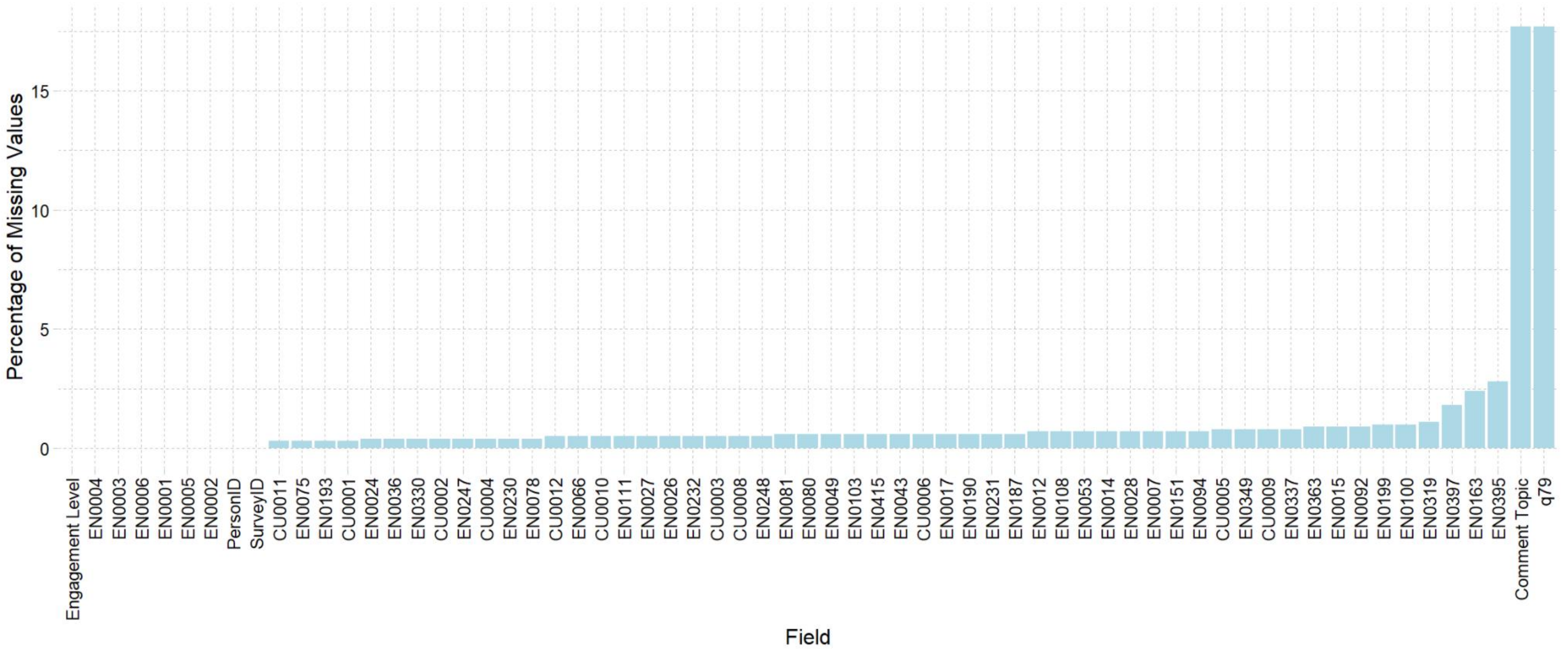
1. Identifying **Factors** that might be **Correlated** with **Attrition**

Part 3: Reducing Attrition

1. Model to **Predict Attrition**
2. **Improve** Attrition **Prediction**
3. Action Plan to **Mitigate Attrition**

Overview | Data and Analysis

Missing Data Analysis



Missing Data Imputation

Employee ID	Q1	Q2	Q3
1	●	●	●
2	●	●	●
3	●	?	●
4	●	●	●
5	●	●	●
6	●	●	●
7	●	●	●

Original Responses

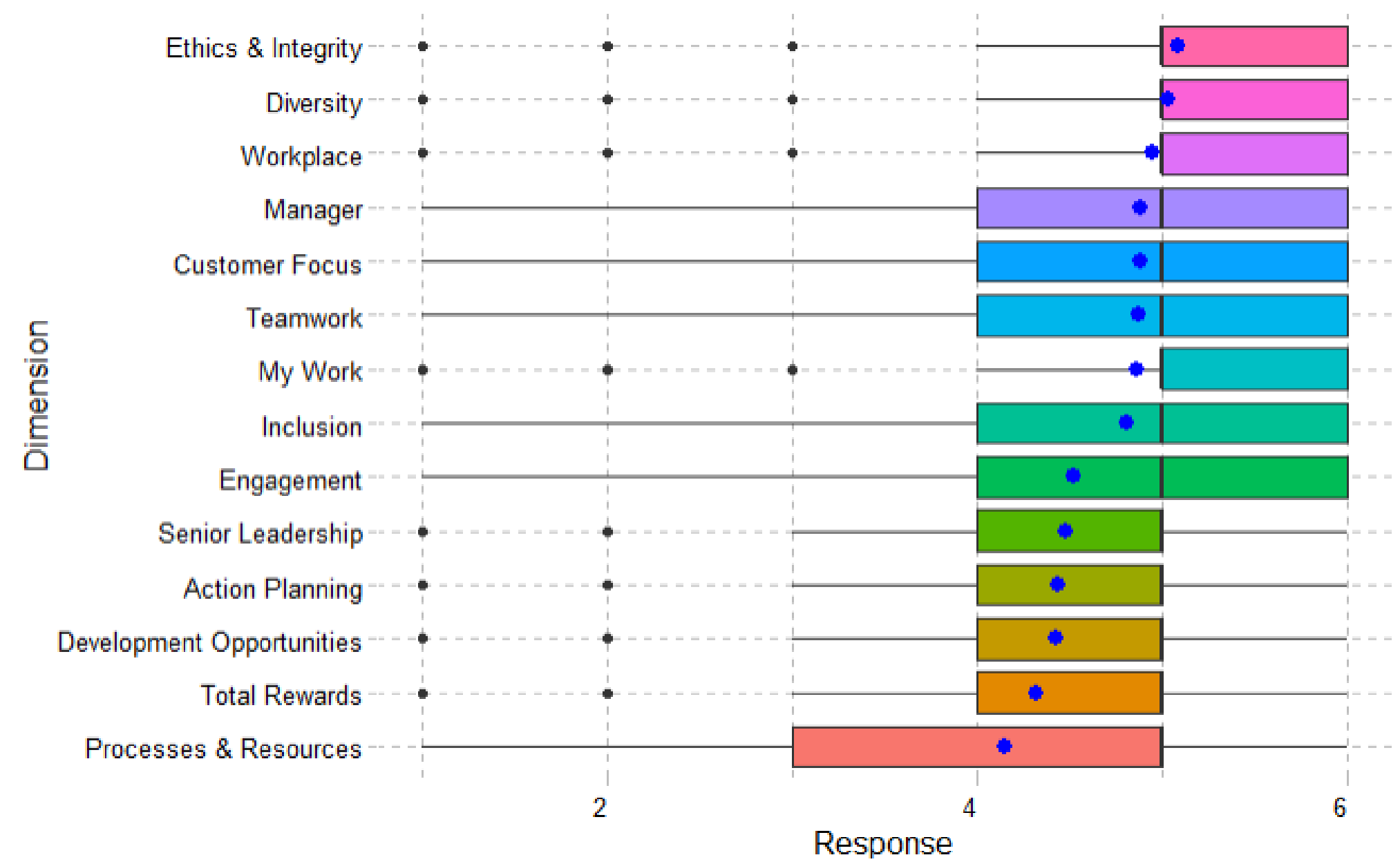
Employee ID	Q1	Q2	Q3
1	●	●	●
2	●	●	●
3	●	●	●
4	●	●	●
5	●	●	●
6	●	●	●
7	●	●	●

Majority Voting
(Not Recommended)

Employee ID	Q1	Q2	Q3
1	●	●	●
2	●	●	●
3	●	●	●
4	●	●	●
5	●	●	●
6	●	●	●
7	●	●	●

Nearest Neighbor
Imputation

Overall Sentiment across Dimensions



Ethics, Diversity,
Workplace, Manager

Engagement, Development
Opportunities, Rewards, Processes

Top 10 Questions | Bottom 10 Questions

Top 10 Question Words

manager
working goals
supports team
work
environment
differences
individual

Bottom 10 Question Words

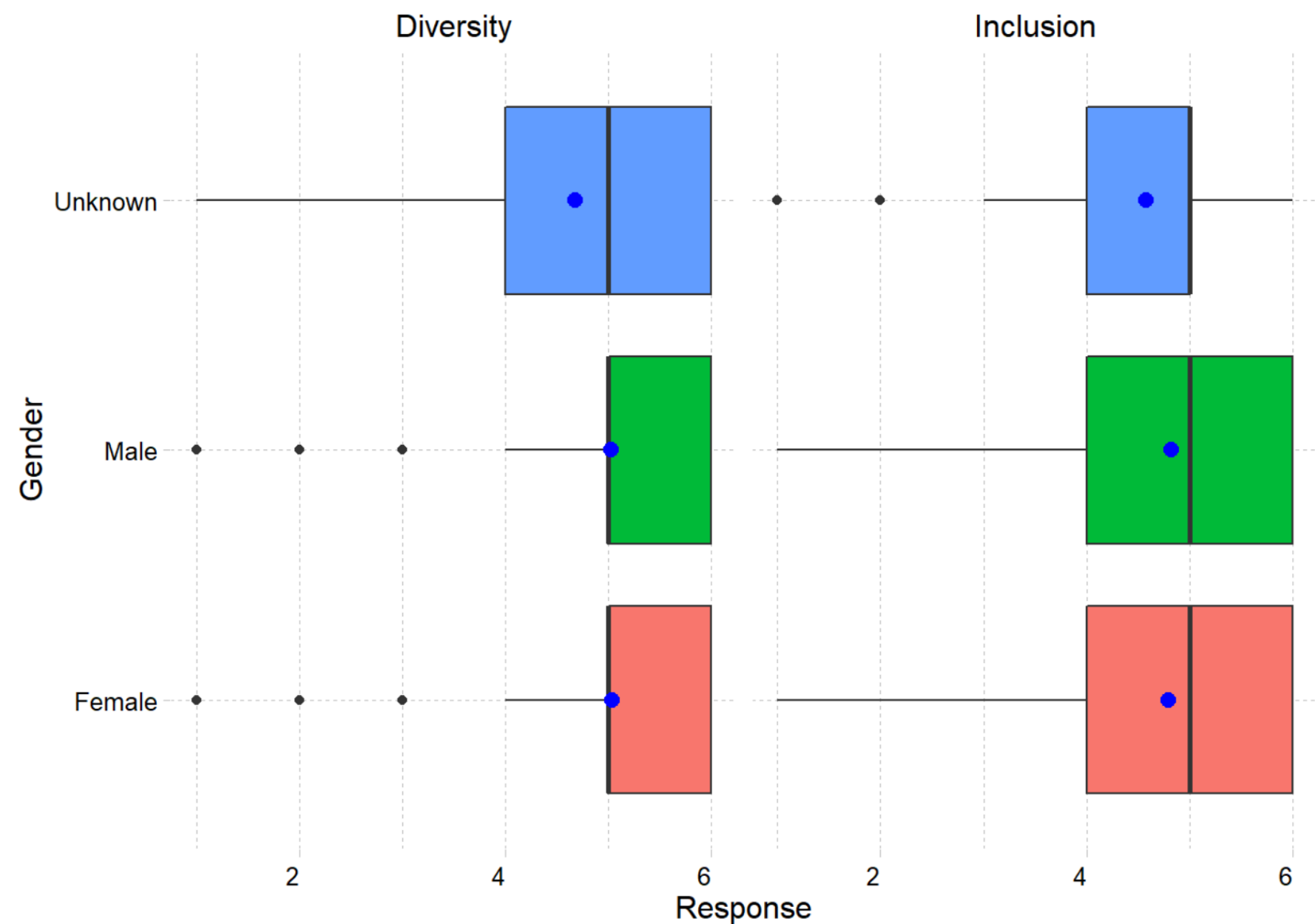
contributions
thinkpay
effective
company
work
survey
performance

Part 1: Diversity and Inclusion

By Gender

By Race/Ethnicity

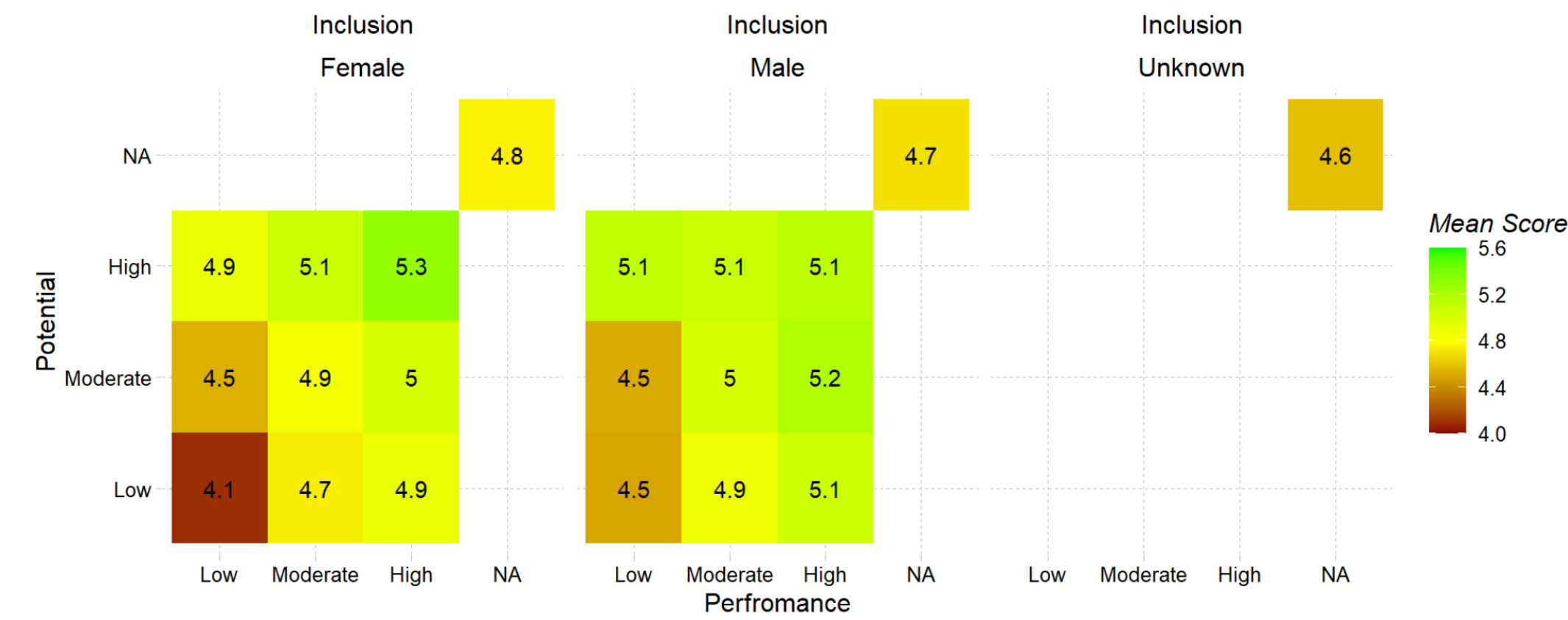
Diversity & Inclusion | By Gender



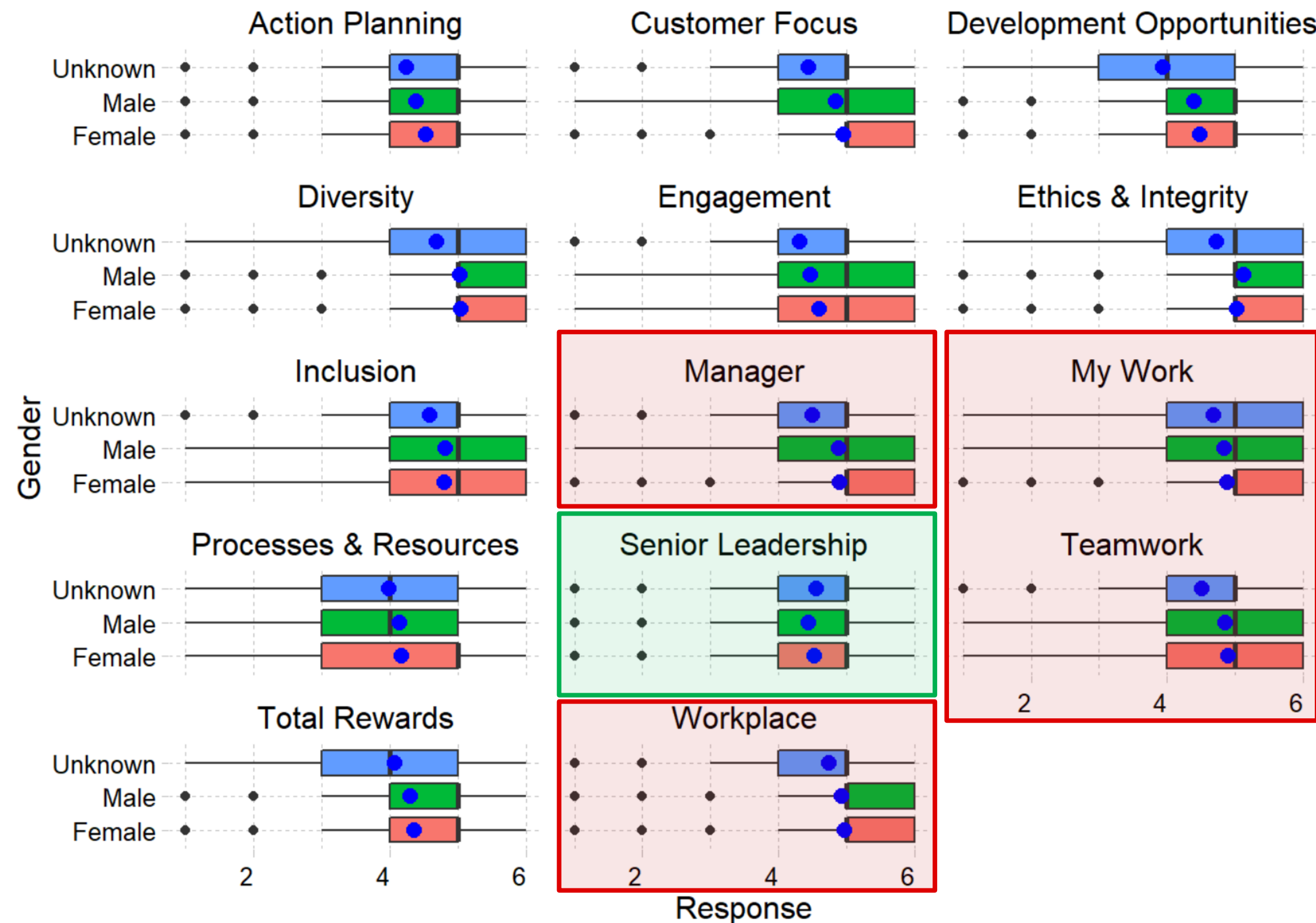
Diversity	Female	Unknown
Male	-0.01 (p-value: 0.34)	-0.35 (p-value: 3e-5)
Female		-0.36 (p-value: 7e-5)

Inclusion	Female	Unknown
Male	0.02 (p-value: 0.06)	-0.22 (p-value: 0.01)
Female		-0.238 (p-value: 0.01)

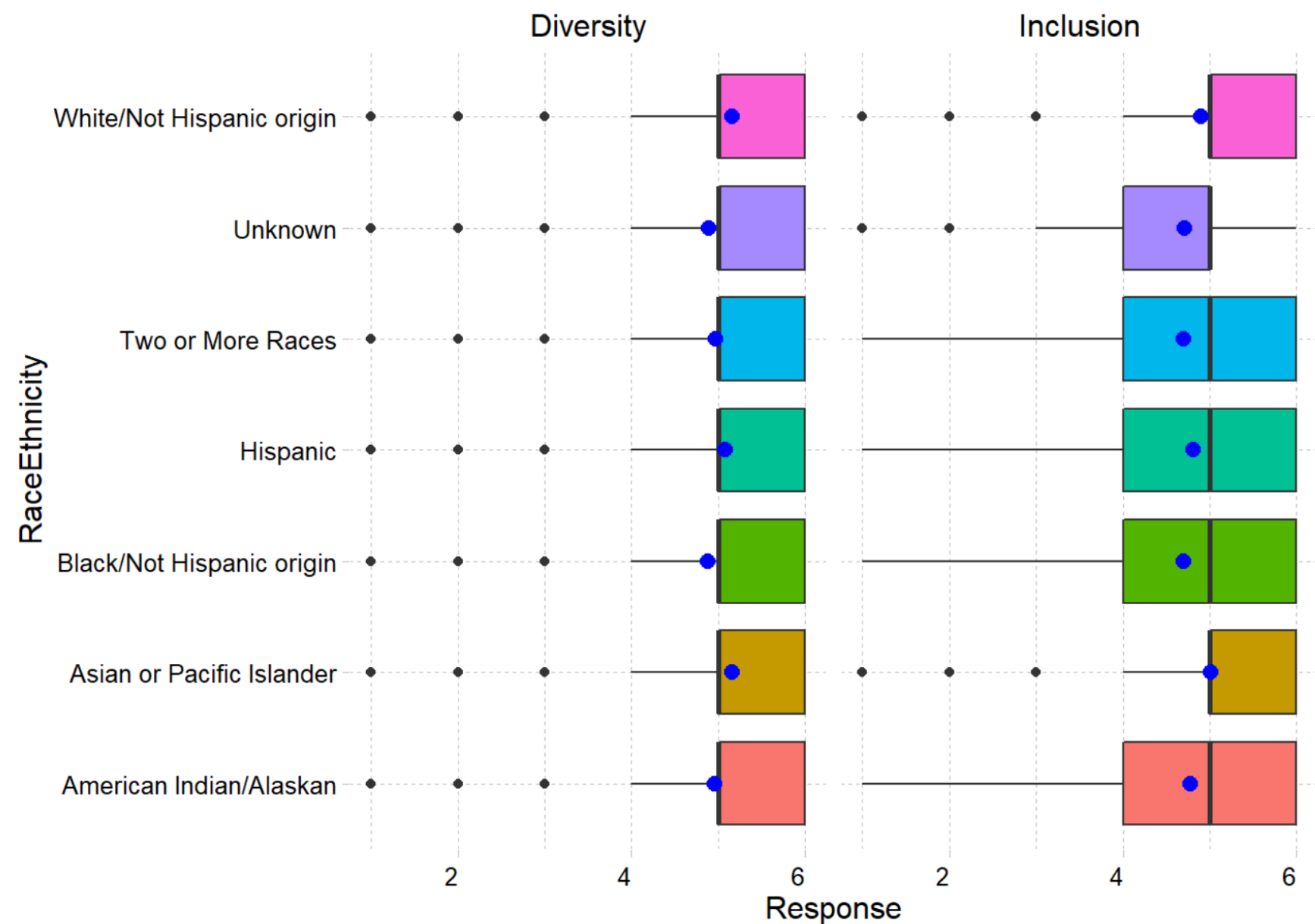
Diversity & Inclusion | By Gender & 9 Box



All Dimensions | By Gender



Diversity & Inclusion | By Race and Ethnicity



Diversity

All groups poll similarly
(some statistical difference found)

Inclusion

White/Non Hispanic Origin
Asian or Pacific Islander

Hispanic
Black/Non Hispanic Origin American
Indian/Alaskan



**"DIVERSITY IS BEING INVITED TO THE PARTY;
INCLUSION IS BEING ASKED TO DANCE."**

VERNA MYERS, DIVERSITY AND INCLUSION EXPERT

Diversity and Inclusion | **Action Plan**

Gender

- Differences between Male and Female groups is not significant
- Group with “unknown” gender generally scores lower
- Regards “Senior Leadership” highly. Issue seems to be coming from day to day interactions
- Further investigation warranted without violating employee privacy

Race/Ethnicity

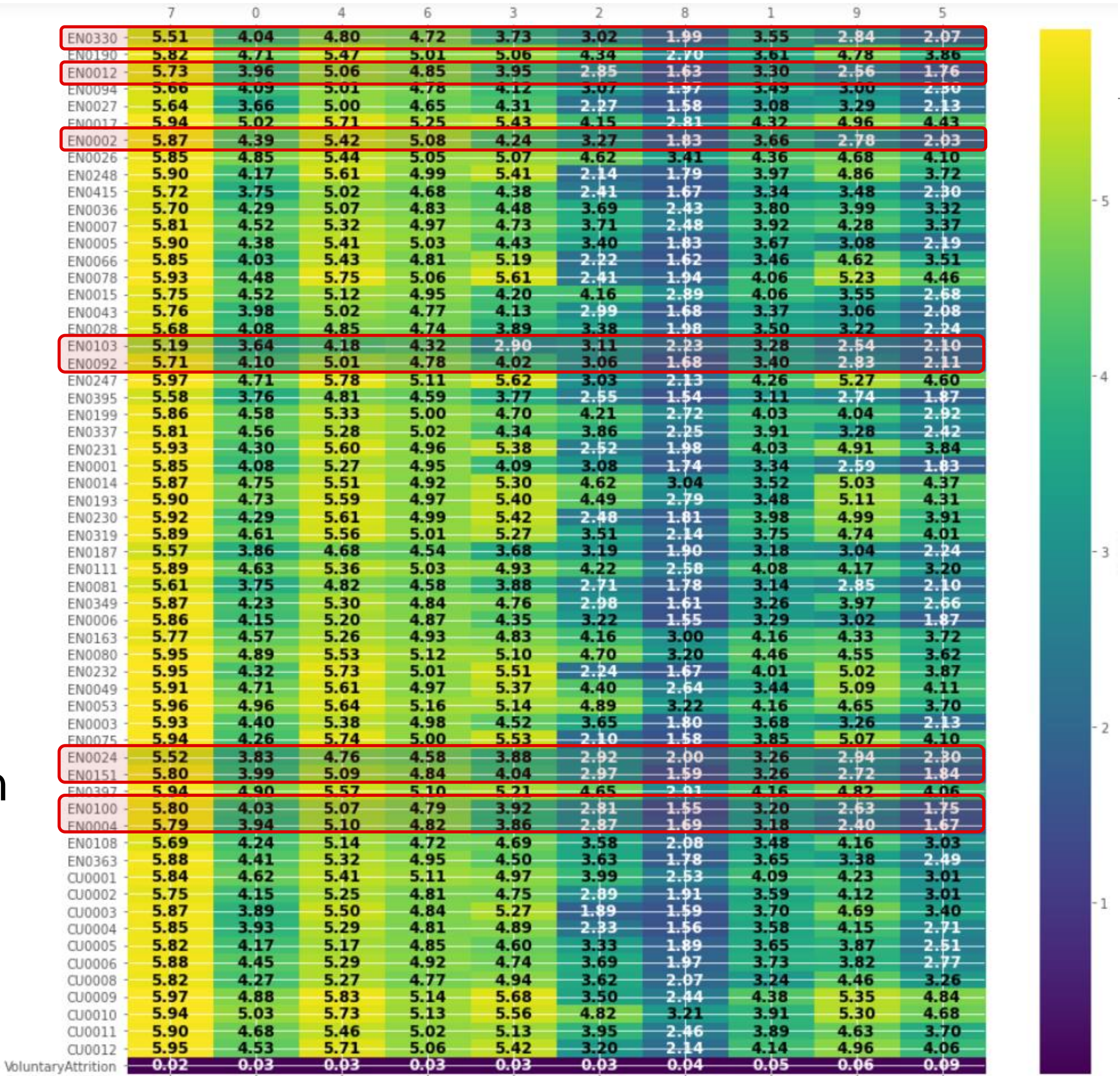
- Not much difference in “Diversity” scores
- However, some groups not feeling as “included” (Hispanic, Black/Non Hispanic Origin, American Indian/Alaskan)
- This may warrant further investigation and action planning to address underlying causes (e.g. **“Implicit Bias”** training at **regular intervals**)

Part 2: Factors Influencing Attrition

Important Predictors of Attrition

Which questions (out of 60) are important for predicting risk of attrition?

- 1. Divide employees into 10 groups (based on survey responses).
- 2. Evaluate attrition in each group.
- 3. Evaluate responses of high attrition groups and how they differ from responses of low attrition groups.



Most Important Questions

Dimension

Question Words

Senior Leadership

Engagement

Total Rewards

Action Planning

Processes Resources

[illegible]

Related to Non Monetary Topics such as Engagement and Career Opportunities

Part 3: Reducing Attrition

Reducing Attrition | Overview and Game Plan

Overview

1. Cost of replacing voluntary attrition
 - A. $0.5x - 2x \text{ Base Pay}^* = \$100,000$
2. Cost to retain top talent likely to leave voluntarily
 - A. Stock Incentives, Career Opportunity: **\$40,000**
 - B. Cost Savings: **\$60,000** per employee prevented from leaving
3. Percentage Retained: 50% (out of 600)
→ Cost Savings: **\$18M**

Game Plan

1. Employees who left within 6 months of taking the survey are known.
2. Build a model to identify employees at high risk of attrition.
3. Build a retention package for key identified employees (**High Performance/Potential**) – **Monetary or Non Monetary**
4. **Challenge:** Underestimating or Overestimating Attrition

* [Reference](#)

Model | Random Guessing vs. Informed Model



😊 18999 employees surveyed

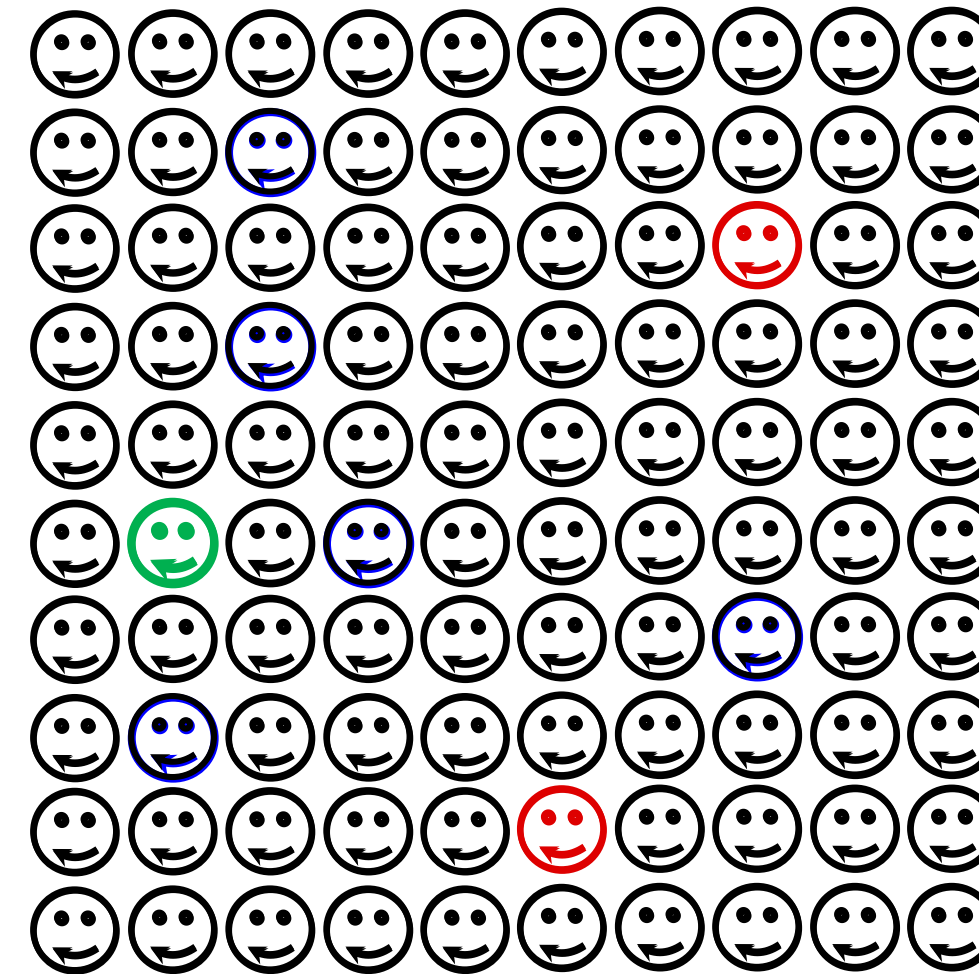
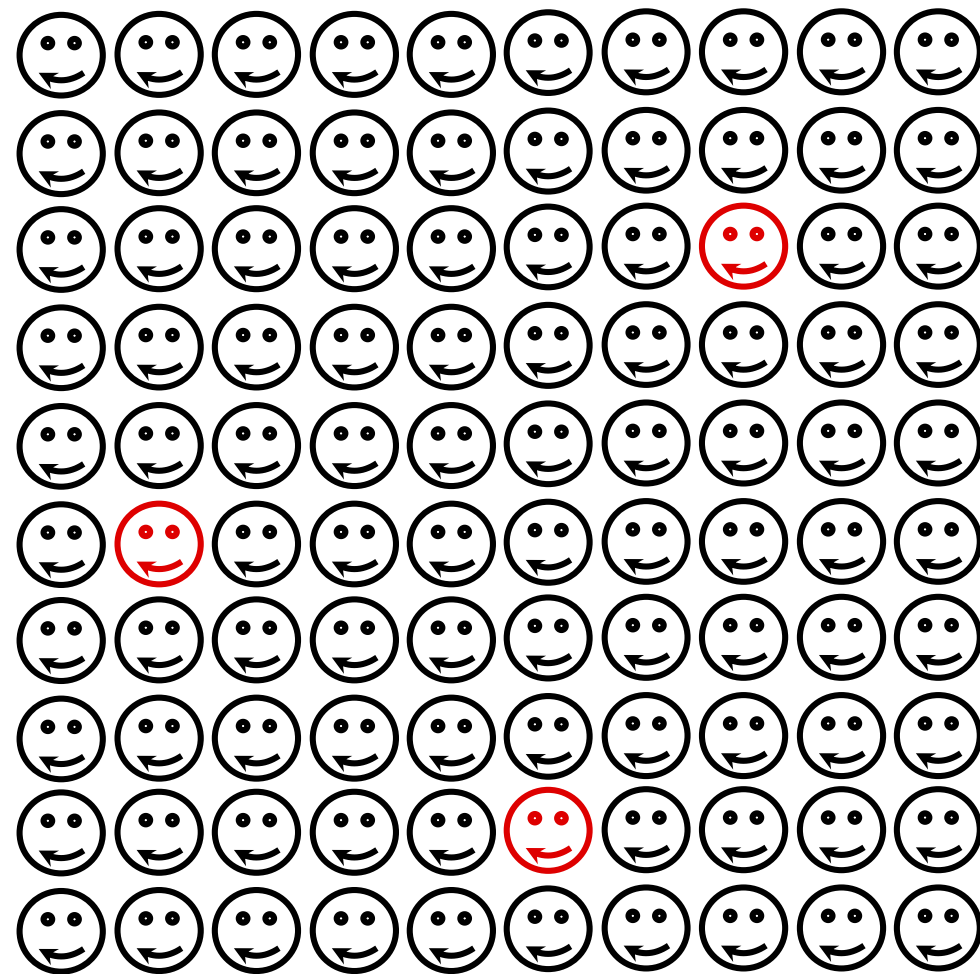
😊 600 left voluntarily

Probability of correctly identifying attrition by **random guessing ~ 3%, however overestimating.**

Model developed to predicted employees at risk of attrition

Probability of predicting attrition correctly with **model ~ 15%**

Model Performance



- 👤 Left Voluntarily
- 👤 Model correctly identifies **1 in 3** employees that are likely to leave.
Would have identified **200 (out of 600)** employees who left.
- 👤 For every 1 employee that needed retention incentives, we would be providing additional retention incentives to 5 more employees that did not likely need it.
For the 200 employees correctly identified, 1000 employees would have been incorrectly identified as well.

Improve Prediction

1. Data did not contain important variables that could impact attrition
 - A. **Salary placement vs. market**
 - B. **Existing long-term incentives** (stock options, RSU's, etc.)
 - C. **Job Grade**
2. Addition of these variables could improve the prediction power of the model.
3. There may be other data points that could be included in the model as well after talking to HR domain experts.

Action Plan to **Mitigate Attrition**

1. Create a retention package for employees that are highly likely to leave (could be monetary or non-monetary).
2. Model should not be the sole point of reference.
 - A. Triangulate based on multiple reference points.
 - B. The model incorrectly predicts that 5 employees will leave for every 1 employee that actually leaves.
 - C. However, some of these 5 employees could be eliminated from the list of high attrition employees if they have high pending stock incentives.
 - D. Other sources of data reference points could be added as well after talking to the HR team.

Questions?



**"DIVERSITY IS BEING INVITED TO THE PARTY;
INCLUSION IS BEING ASKED TO DANCE."**

VERNA MYERS, DIVERSITY AND INCLUSION EXPERT