RISE BY DIGITAL BCG ACADEMY

BDA03 Group 5
RISE-ING Stars

Capstone Presentation

Executive Summary

Background

Gain **visibility** over workforce

Dashboard providing **insights**

3 Key Insights

Diverse workforce

Pay mismatch
Career mobility
Talent development

Employee performance drivers

Career trajectory

Retention risk

Location

Employee attrition drivers

Employment duration

Job responsibility

Performance / Grade

3 Recommendations

Embark on further study with **broader dataset**

Attract younger talents (20+ years old) to improve age diversity

Engage with "cruisers" in workforce to enhance potential & performance

3 Impacts

Begin objective, effective and **efficient talent conversations**

Cluster employees by profiles for **targeted** talent development

Act pre-emptively to retain high potential talents

How might we empower *Company* HR to visualize their workforce, so that insights can be generated, for better communication across employees, HR and leadership?

Achieve empowerment, not just visibility

- Develop a tool for not just timely, but pre-emptive and effective discussions between HR and workforce

Foster engagement, growth and belonging

- Unlock the value from employee data
- Uncover next steps to promote workforce diversity, performance and retention

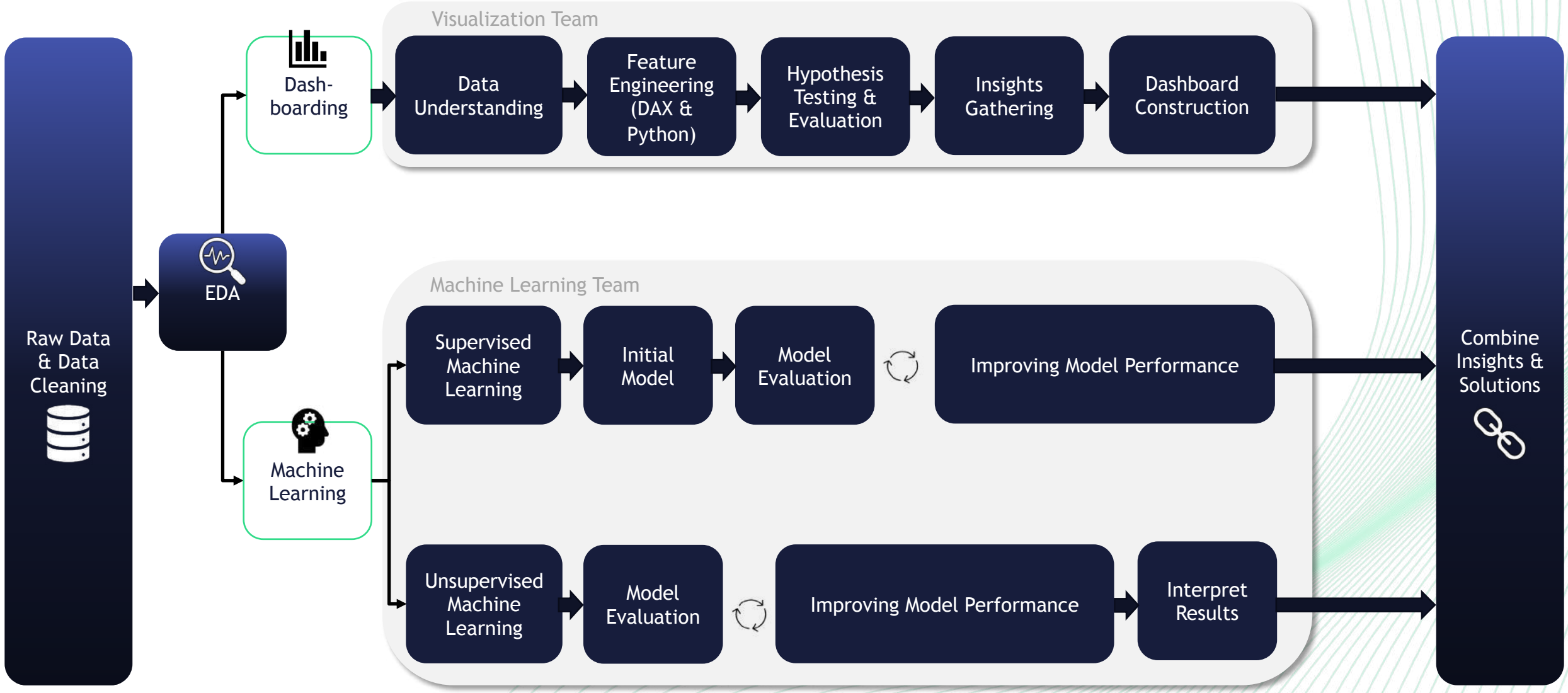
End reliance on the line managers

- Predictive analysis using employee data
- Data-driven recommendations provided

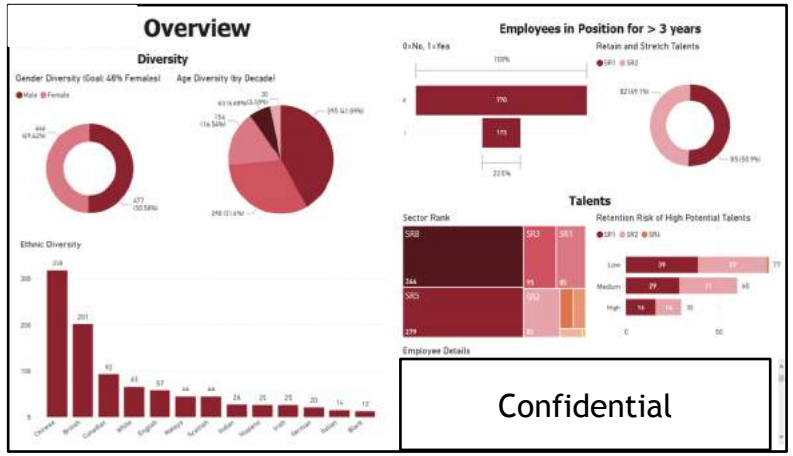
Analytical Objectives:

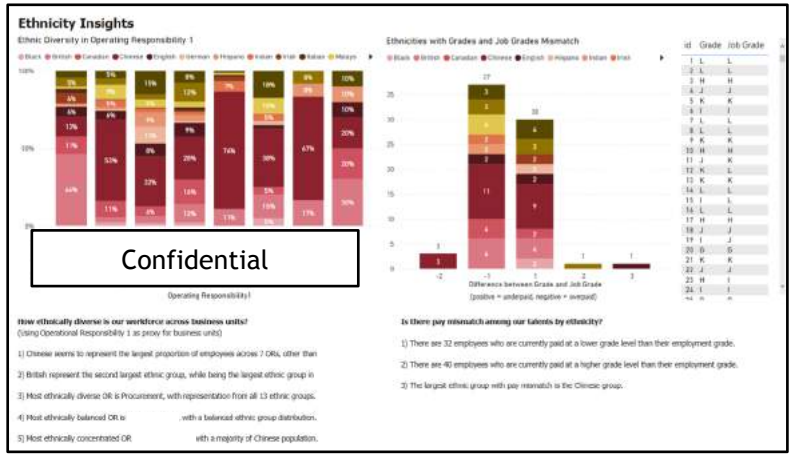
- Derive drivers for employee performance
- Derive drivers for employee retention
- Identify gaps in career mobility, workforce diversity and talent development practices

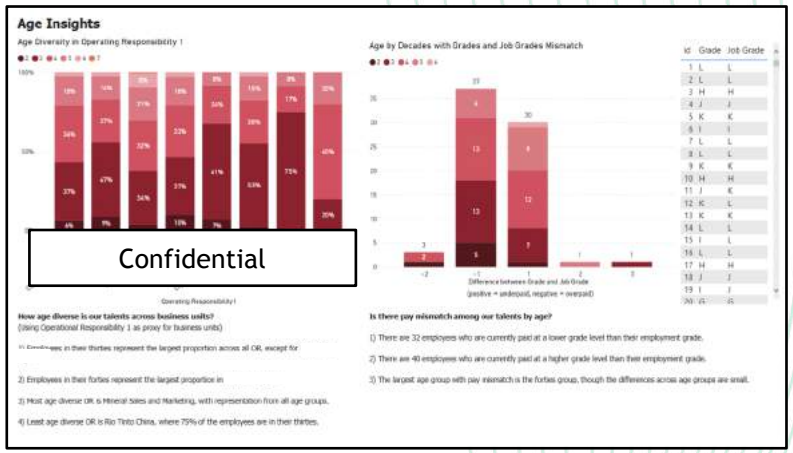
Overall Analytical Workflow

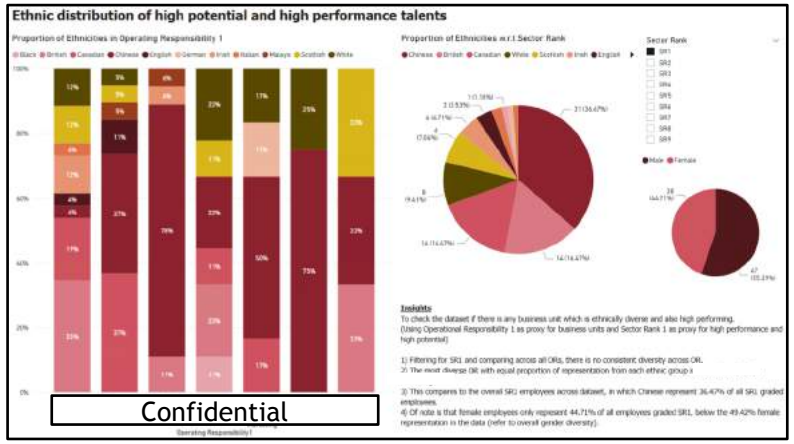


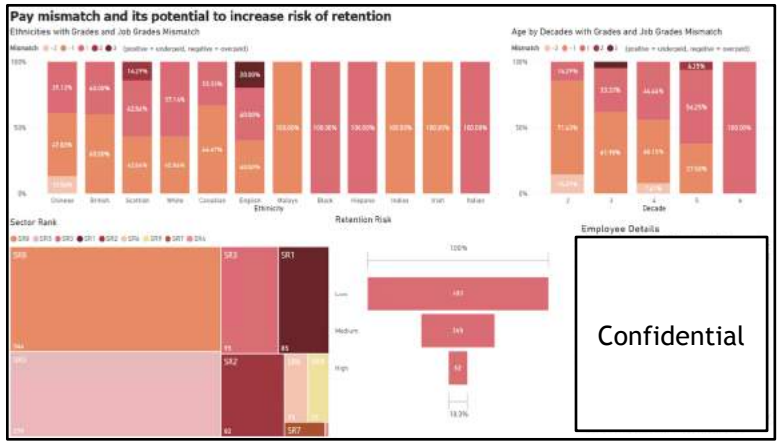
Dashboard at a glance…

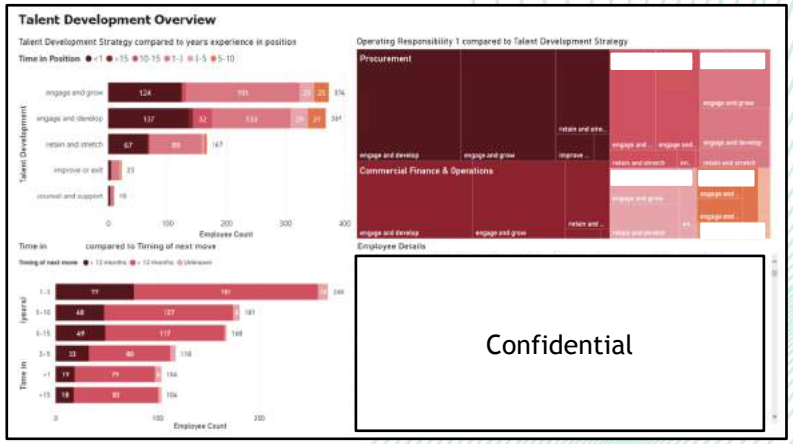


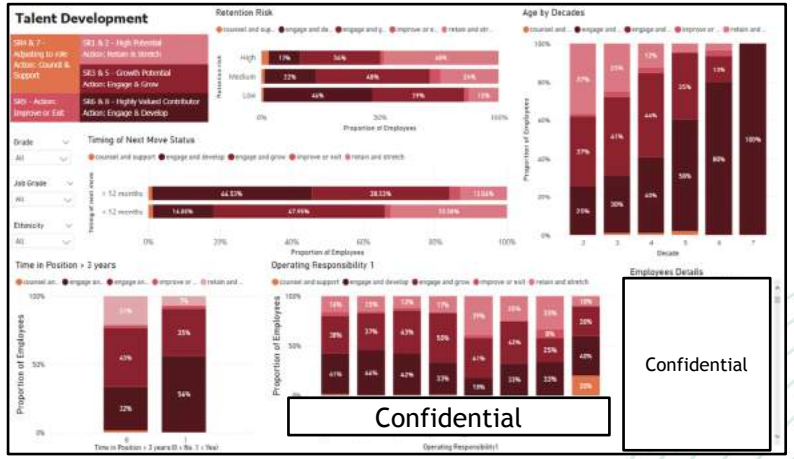


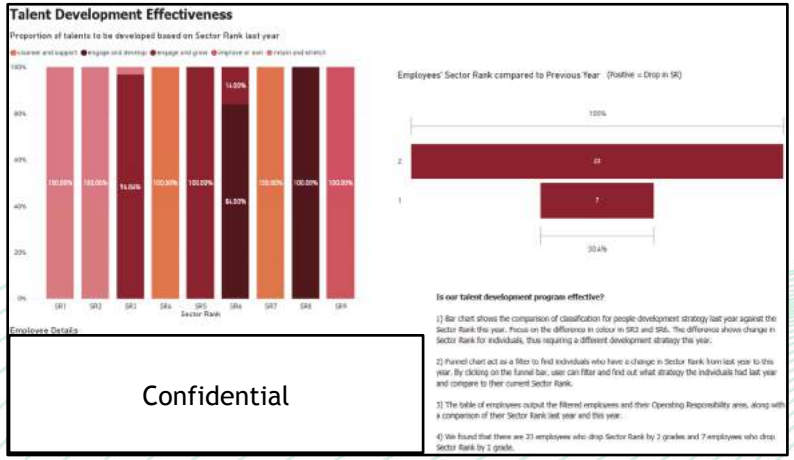


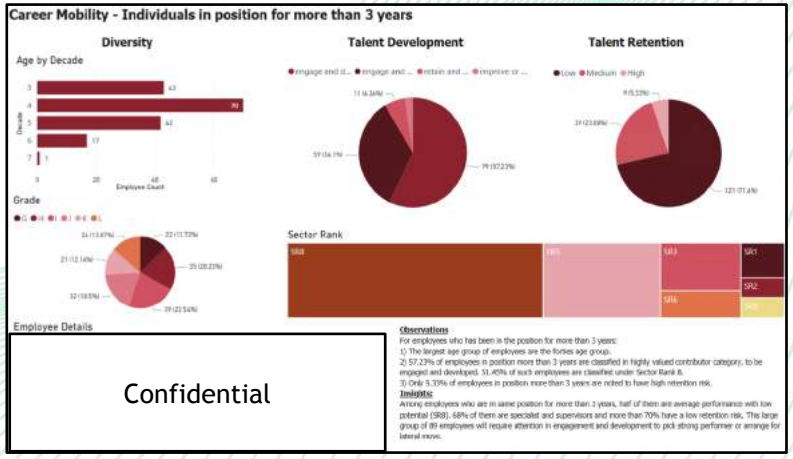












Company is fairly diverse, but more effort needed to improve diversity in high-potential and high-performance bracket.

Company

Gender

49% Females

High Potential and High Performance (SR1)

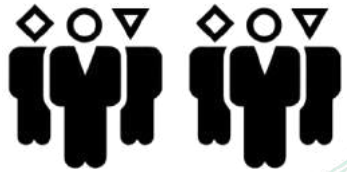
Female Representation

44% Females

Age by Decade & Ethnicities across Business Unit

Company is **ethnic diverse**
but not **age diverse**
(3 out of 8 have 50% employees in their 30s)

Ethnic Diversity in SR1



Inconsistent Ethnic Diversity

36%
Chinese

Presence of pay mismatch and its relevance to increased retention risk

No specific ethnicity and age discrimination

- Spread across all ethnicities and age

!

Underpaid	2 grades	3 grades
Ethnicity	Scottish (14%)	English (20%)
Age	50s (6%)	30s (5%)

ALL who are **underpaid**
are performers & outperformers

5% who are **overpaid**

are in SR9
(underperforming & low potential)

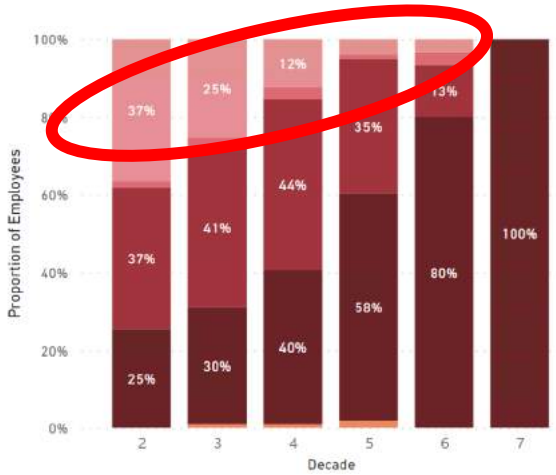
32% performers & outperformers with pay mismatch have **medium to high retention risk**

Dashboard - Diversity: Ethnicity Insights; Diversity: Age Insights; Pay Mismatch Insights

Talent Development strategies[1] could be more directed and focused

By grouping employees into *Company*'s talent development plan based on their SR, from the dashboard, we derived that:

1



'retain and stretch' more **common** among ***younger employees***

- counsel and support
- engage and develop
- engage and grow
- improve or exit
- retain and stretch

2

Timing of next move	Retain and Stretch	Engage and Grow
< 12 months	32%	47%
> 12 months	13%	38%

Higher proportion of employees who should be **groomed** in their current position, expected to move in < 12 months.

3

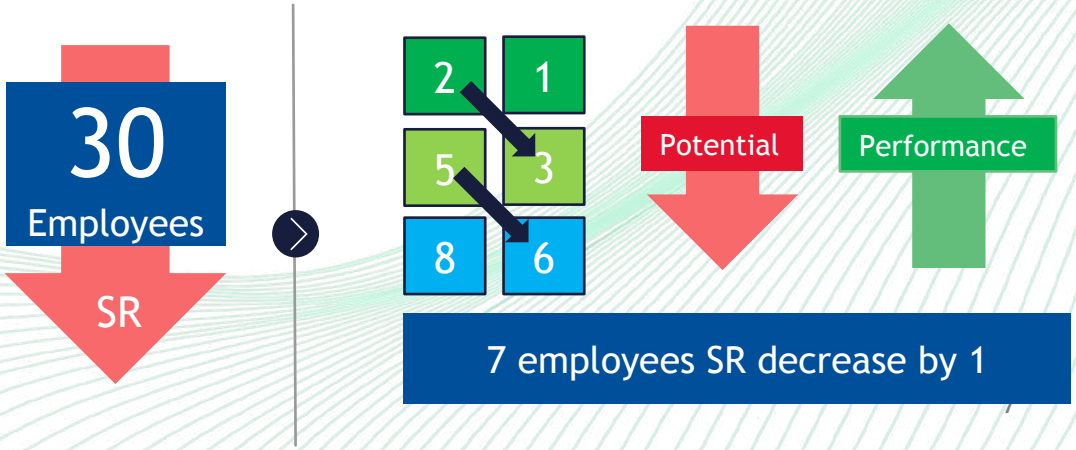
26 employees due to ***move within 6 months***

<u>However</u>

Currently planned to move in **> 12 months**

4

By comparing with last year's development plan:

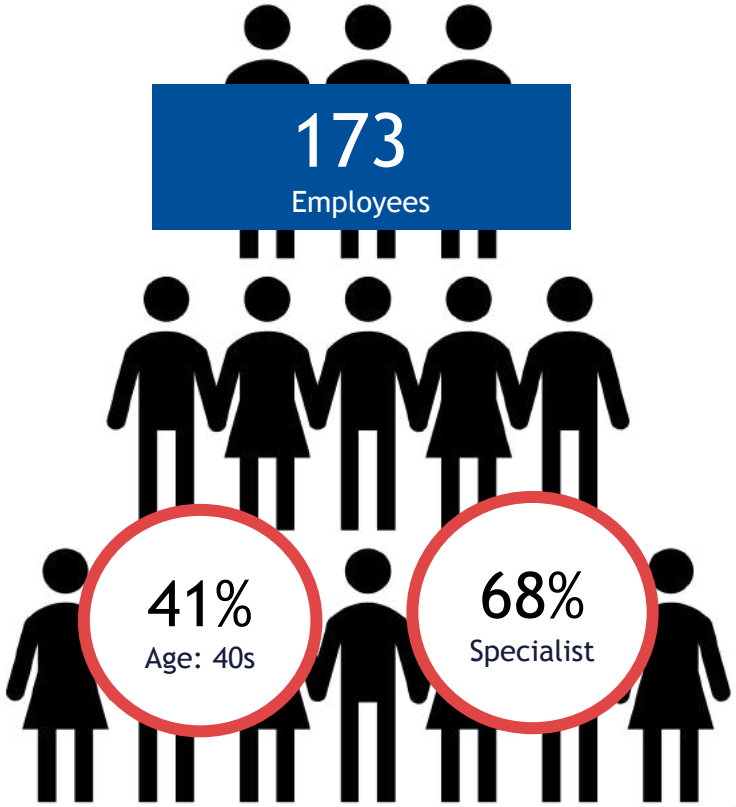


[1]Talent Development strategies: refer to SR Matrix and given development chart in the annex

Dashboard – Talent Development Overview, Talent Development

Career mobility is present but for 'stuck' employees, engagement and development should be given

For Employees who have been in their position for **more than 3 years (stuck)**:



(In position for more than 3 years)

The insights we derived were:

Sector Rank

57%
SR6 & SR8

Retention Risk

51%
Low

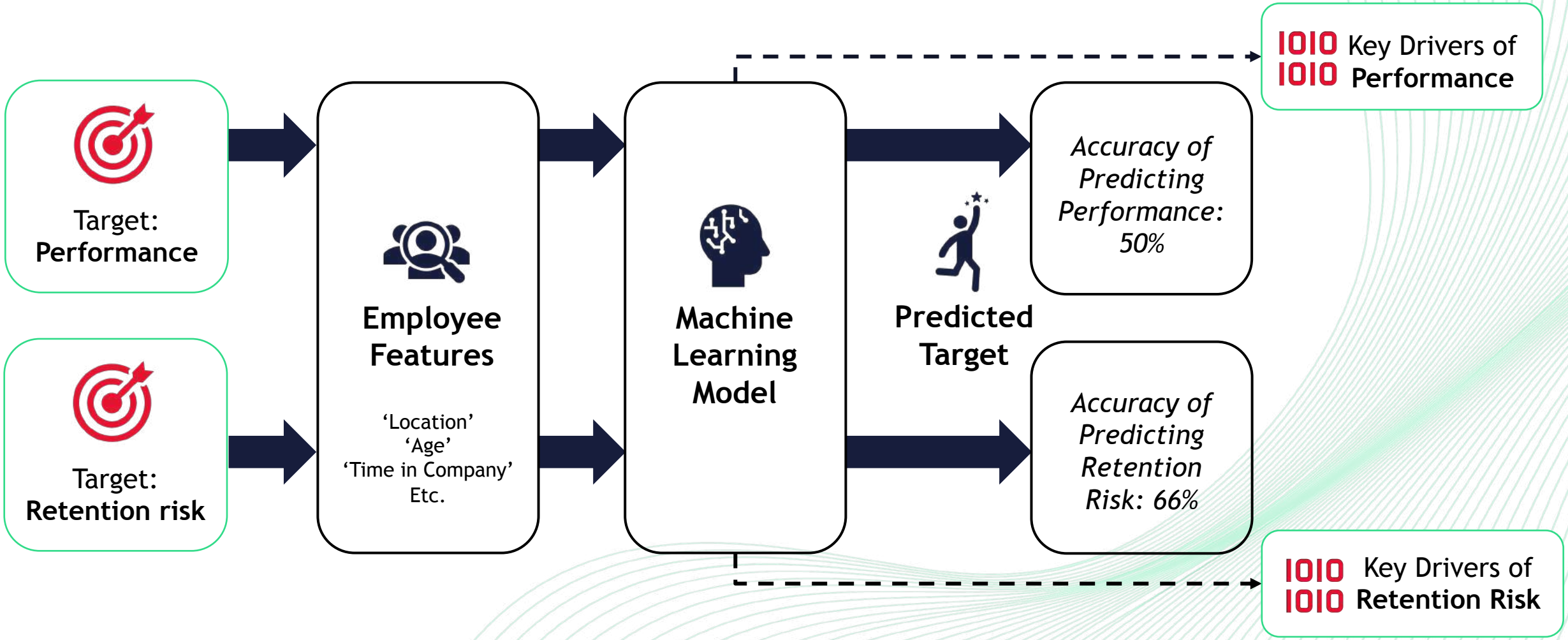
From the dashboard:

We ***identified*** the 89 employees who requires **early intervention**, **attention on engagement** and **development** to ***promote better performance*** and exposure to other positions.

Dashboard: Career Mobility – Individuals in Position for more than 3 years

Machine Learning

Utilizing Supervised Machine Learning to identify key drivers of Performance & Retention Risk



An employee's career trajectory, retention risk and location are key drivers of Performance

Career Trajectory

If an employee:
- is targeted to move in the next 12-months,
- has a suggested next job, and
- has not been in their position for >3 years,

they tend to have HIGHER performance

Employees with a promising career trajectory tend to have HIGHER performance

Retention Risk

If an employee:
- has a higher risk of attrition,

they tend to have HIGHER performance

Employees that are at a higher risk of attrition tend to have HIGHER performance

Working in HQ

If an employee:
- works in *Country1*,

they tend to have LOWER performance

There is a correlation between working in *Company*'s HQ and LOWER employee performance

Correlation of Features vs. Target Variable (Performance):

-1.0 +1.0: Target vs. Feature correlation

Feature Importance	1	2	3	4	5
Target/Feature	Ready for Promotion	Retention Risk	Works in AU	Change of Responsibilities	Years in Position > 3
Performance	(Ready for P, Higher Performance)	(Higher RR, Higher Performance)	(Works in AU, Lower Performance)	(Has CoR, Higher Performance	(>3 years, Lower Performance)

An employee's work duration, performance, grade and responsibility are key drivers of Retention Risk

Duration

If an employee:
- have been working in *Company* for a long time; and
- has been in the same position for more than 5 years,

they tend to have LOWER Retention Risk

Employees working for a long duration tend to have LOWER retention risk.

Performance/Grade

If an employee:
- has a higher performance; and
- has a higher job grade,

they tend to have HIGHER Retention Risk

Employees with a higher performance and job grade tend to have HIGHER retention risk.

Responsibility

If an employee:
- is in charge of *Business Unit 1*

they tend to have HIGHER Retention Risk

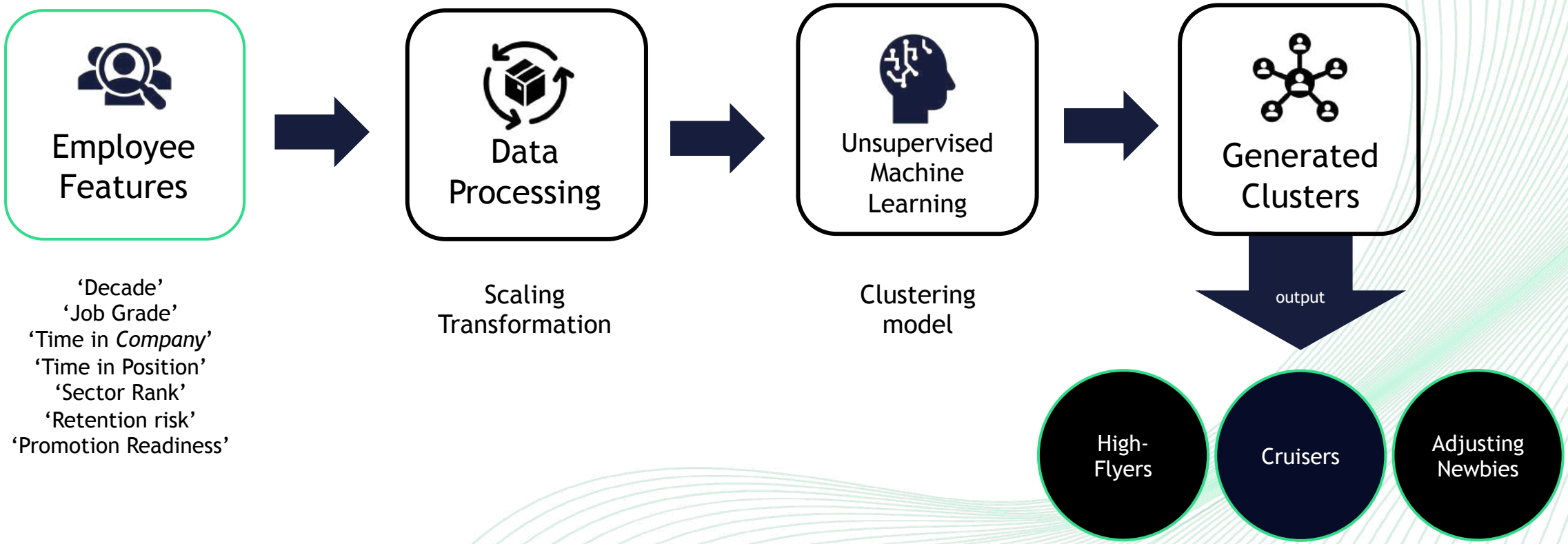
There is a correlation between working in S&M CD and HIGHER retention risk.

Correlation of Features vs. Target Variable (Retention Risk):

-1.0 +1.0: Target vs. Feature correlation

Feature Importance	1	2	3	4	5
Target/Feature	Years in *Company*	Performance	Years in Position >5	*Business Unit 1*	Job Grade
Retention Risk	(Higher Years, Lower Risk)	(Lower Performance, Higher Risk)	(>5 Years in Position, Lower Risk)	(Works in *Business Unit 1*, Higher Risk)	(Higher Job Grade, Higher Risk)

Utilizing Unsupervised Machine Learning to effectively segment employees based on profile



Understanding employee cluster profiles for more effective talent conversations

Highest retention risk

Can't wait for the next promotion. Or I might just accept the offer from another company!

High Flyer
HIGH PERFORMANCE

Medium retention risk

I've only been here for around 1 year... still adjusting to everything. Lots to learn and get used to. I wonder what's next for me in the company.

Adjusting Newbie
MEDIUM PERFORMANCE

Lowest retention risk

Been in the company for more than 10 years and in this position more than 4... I'm so comfortable.

Cruiser, Old Timer
LOW PERFORMANCE

Cluster	Profile	Age by Decade	Pay Grade	Years in *Company*	Years in Position	Sector Rank	Retention Risk	Promotion Readiness
0	High flyers	Younger	Low-Medium Pay	Short-medium service	Medium service	Highest Performance	Highest Retention risk	High readiness
1	Cruisers	Older	Low Pay	Long service	Long service	Lowest Performance	Lowest Retention risk	Low readiness
2	Adjusting Newbies	Younger	High Pay	Short service	Short service	Medium Performance	Medium Retention risk	Not ready

Low
Low-Medium
Medium
High

Action points based on insights derived from analysis

Recommendations

Further Research

- **Ethnic diversity in *Business Unit 2*;** skewed towards "Chinese"
- **Underperformance in *Country1*** (lower SR)

Engagement

- **Address the 26 employees (SR4, 7, 9) with mobility issues**
- **Review and address the pay mismatch we identified**
- **Review the employees identified as "stuck" with avg. performance (SR8)**
  - **Longer-term, leverage on the clustering model to identify these "cruisers"**

Recruitment

- **Target 20-year-olds to improve age diversity;** *Business Unit 3* has ~60% employees in their 40s

Training

- **Conduct training for line managers to align understanding on SR ranking criteria**

Area(s) for improvement

- **Larger dataset** (analysis was conducted on ~900 observations post-cleaning)
- **Increase granularity of data** (e.g. Age, Time in *Company*/Position)
- **More informative features** (e.g. KPIs met?, employee satisfaction level, managerial reporting structure, etc.)

Operationalize

- **Utilize a fuller dataset with more granular and informative employee characteristics**
- **Estimate cost/benefit of models for impact analysis**
- **Create a detailed business plan and engage with key stakeholders across the firm to be submitted for approval**

02

ANNEX

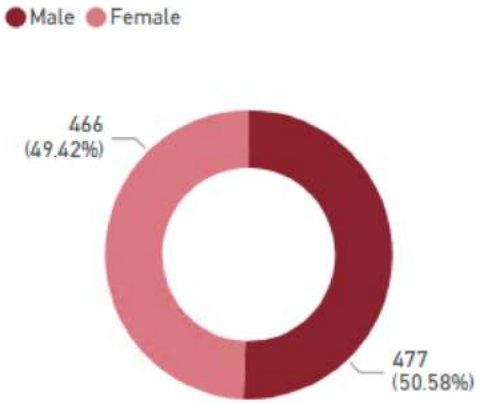
1.0 Power BI

Confidential

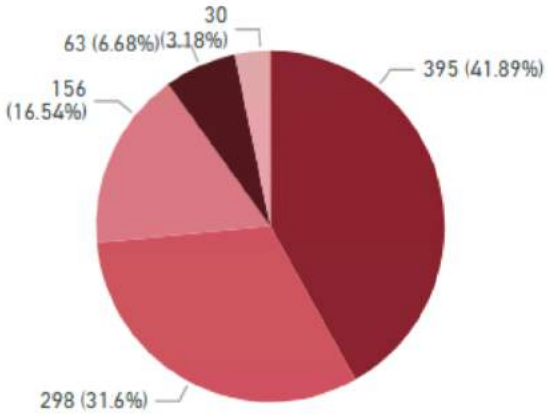
Overview

Diversity

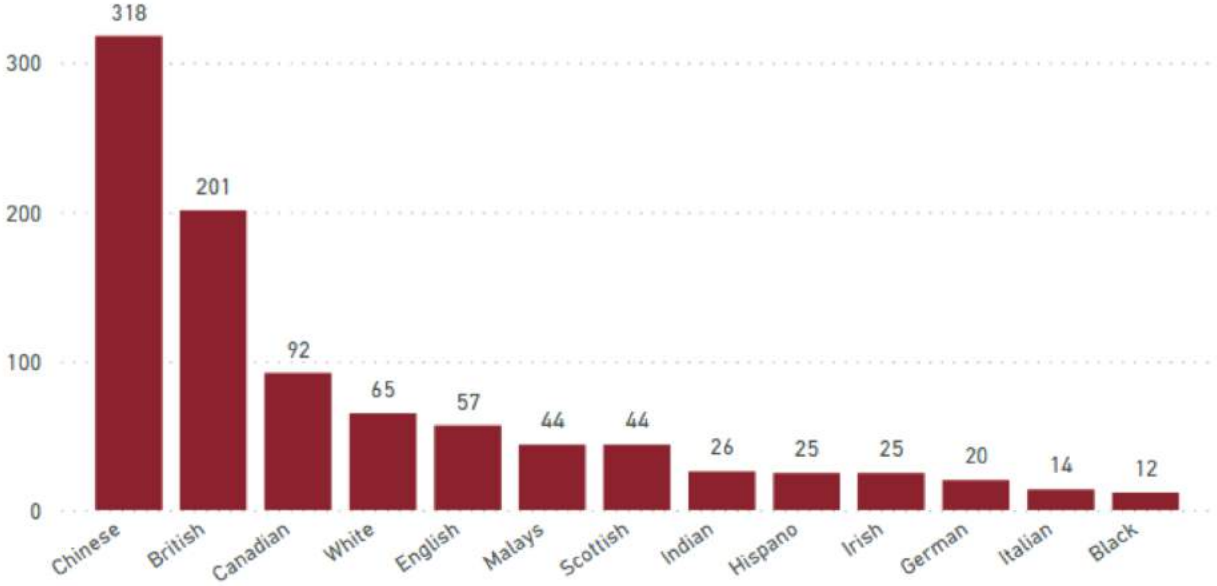
Gender Diversity (Goal: 48% Females)



Age Diversity (by Decade)

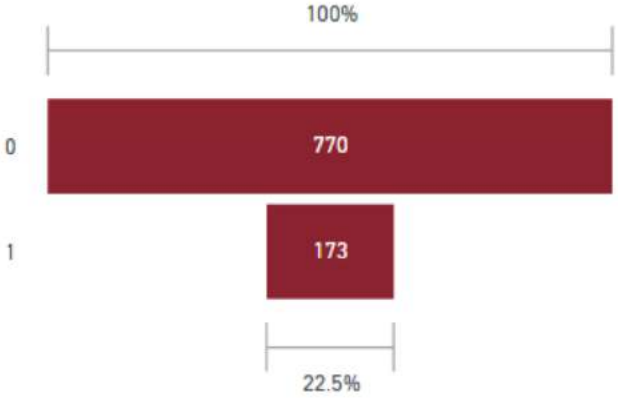


Ethnic Diversity

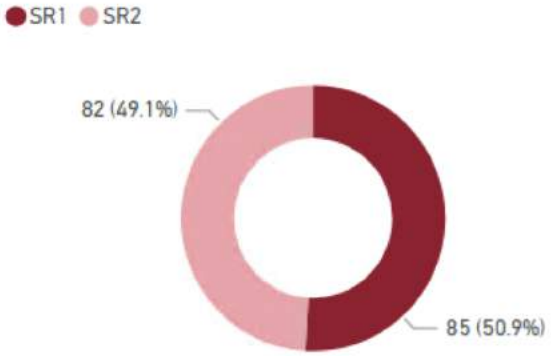


Employees in Position for > 3 years

0=No, 1=Yes

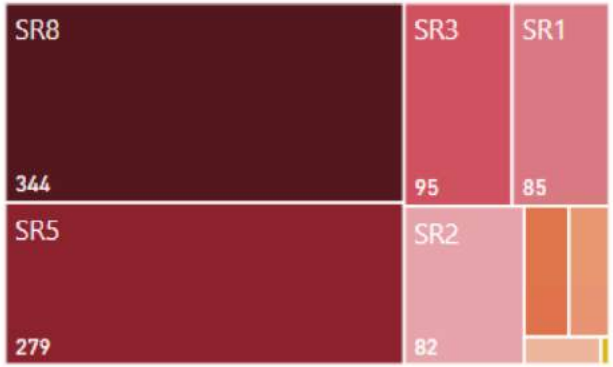


Retain and Stretch Talents

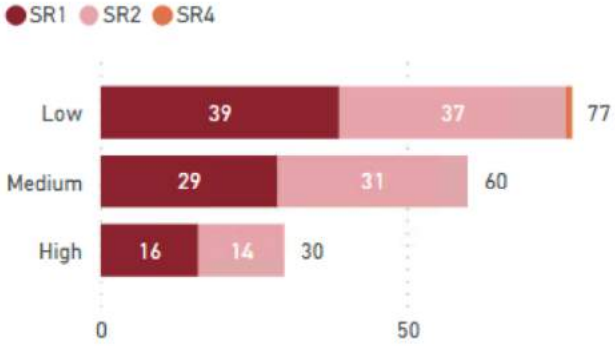


Talents

Sector Rank



Retention Risk of High Potential Talents

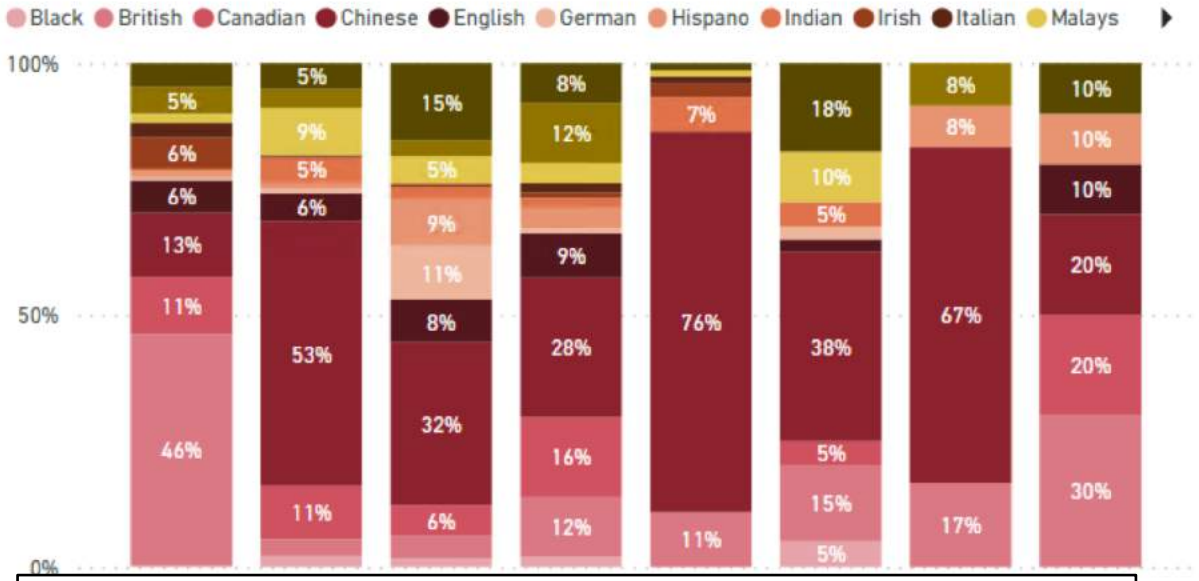


Employee Details

Confidential

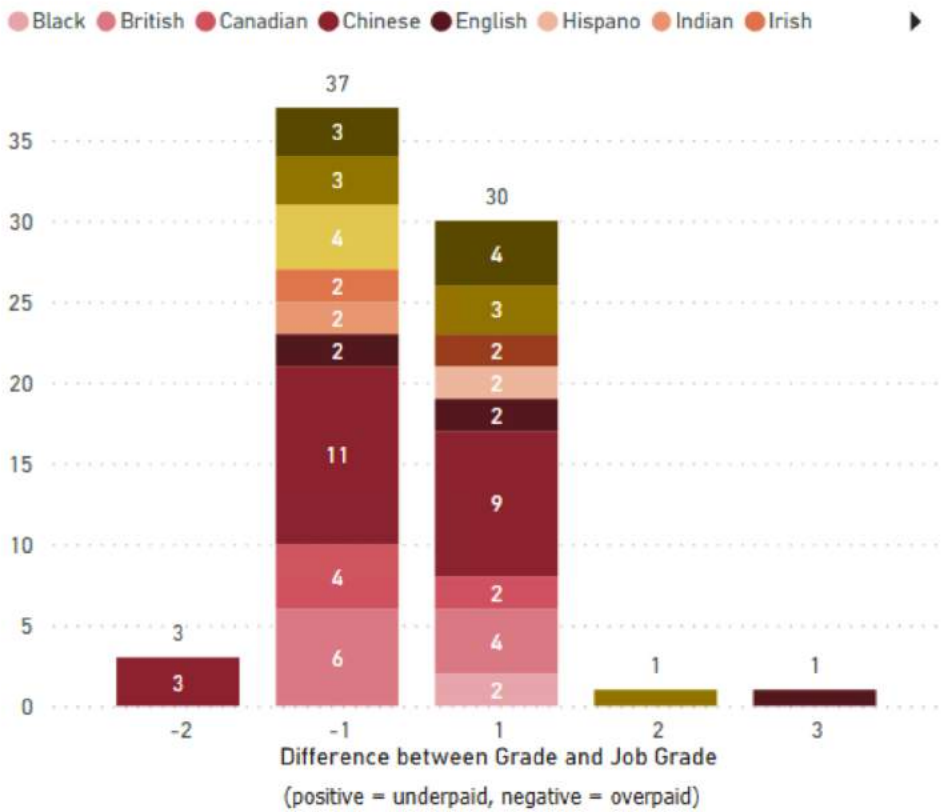
Ethnicity Insights

Ethnic Diversity in Operating Responsibility 1



Confidential

Ethnicities with Grades and Job Grades Mismatch



id	Grade	Job Grade
1	L	L
2	L	L
3	H	H
4	J	J
5	K	K
6	I	I
7	L	L
8	L	L
9	K	K
10	H	H
11	J	K
12	K	L
13	K	K
14	L	L
15	I	L
16	L	L
17	H	H
18	J	J
19	I	J
20	G	G
21	K	K
22	J	J
23	H	I
24	I	I

How ethnically diverse is our workforce across business units?
(Using Operational Responsibility 1 as proxy for business units)

1) Chinese seems to represent the largest proportion of employees across 7 ORs, other than

2) British represent the second largest ethnic group, while being the largest ethnic group in

3) Most ethnically diverse OR is with representation from all 13 ethnic groups.

4) Most ethnically balanced OR is , with a balanced ethnic group distribution.

5) Most ethnically concentrated OR is , with a majority of Chinese population.

Is there pay mismatch among our talents by ethnicity?

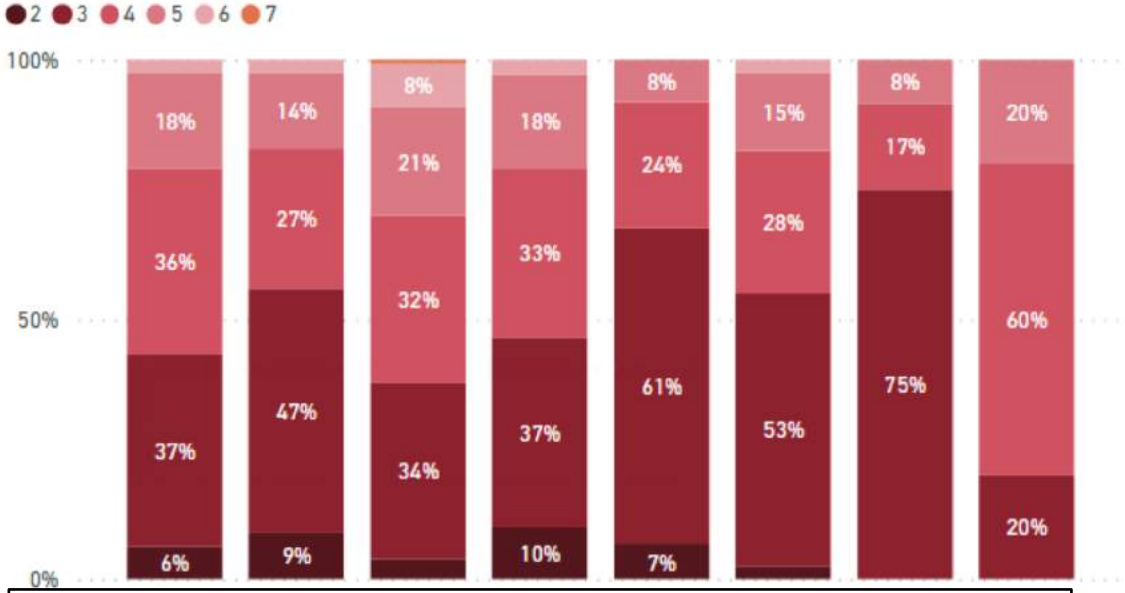
1) There are 32 employees who are currently paid at a lower grade level than their employment grade.

2) There are 40 employees who are currently paid at a higher grade level than their employment grade.

3) The largest ethnic group with pay mismatch is the Chinese group.

Age Insights

Age Diversity in Operating Responsibility 1



Confidential

Operating Responsibility1

How age diverse is our talents across business units?
(Using Operational Responsibility 1 as proxy for business units)

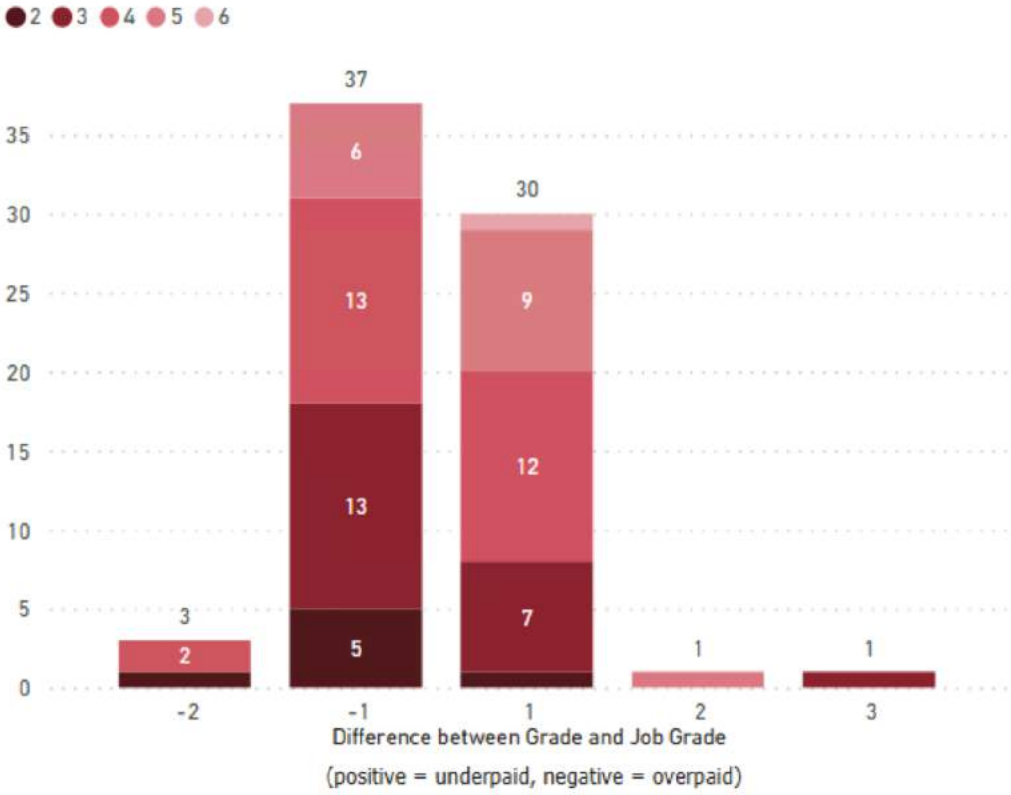
1) Employees in their thirties represent the largest proportion across all OR, except for

2) Employees in their forties represent the largest proportion in .

3) Most age diverse OR is , with representation from all age groups.

4) Least age diverse OR is . where 75% of the employees are in their thirties.

Age by Decades with Grades and Job Grades Mismatch



id	Grade	Job Grade
1	L	L
2	L	L
3	H	H
4	J	J
5	K	K
6	I	I
7	L	L
8	L	L
9	K	K
10	H	H
11	J	K
12	K	L
13	K	K
14	L	L
15	I	L
16	L	L
17	H	H
18	J	J
19	I	J
20	G	G

Is there pay mismatch among our talents by age?

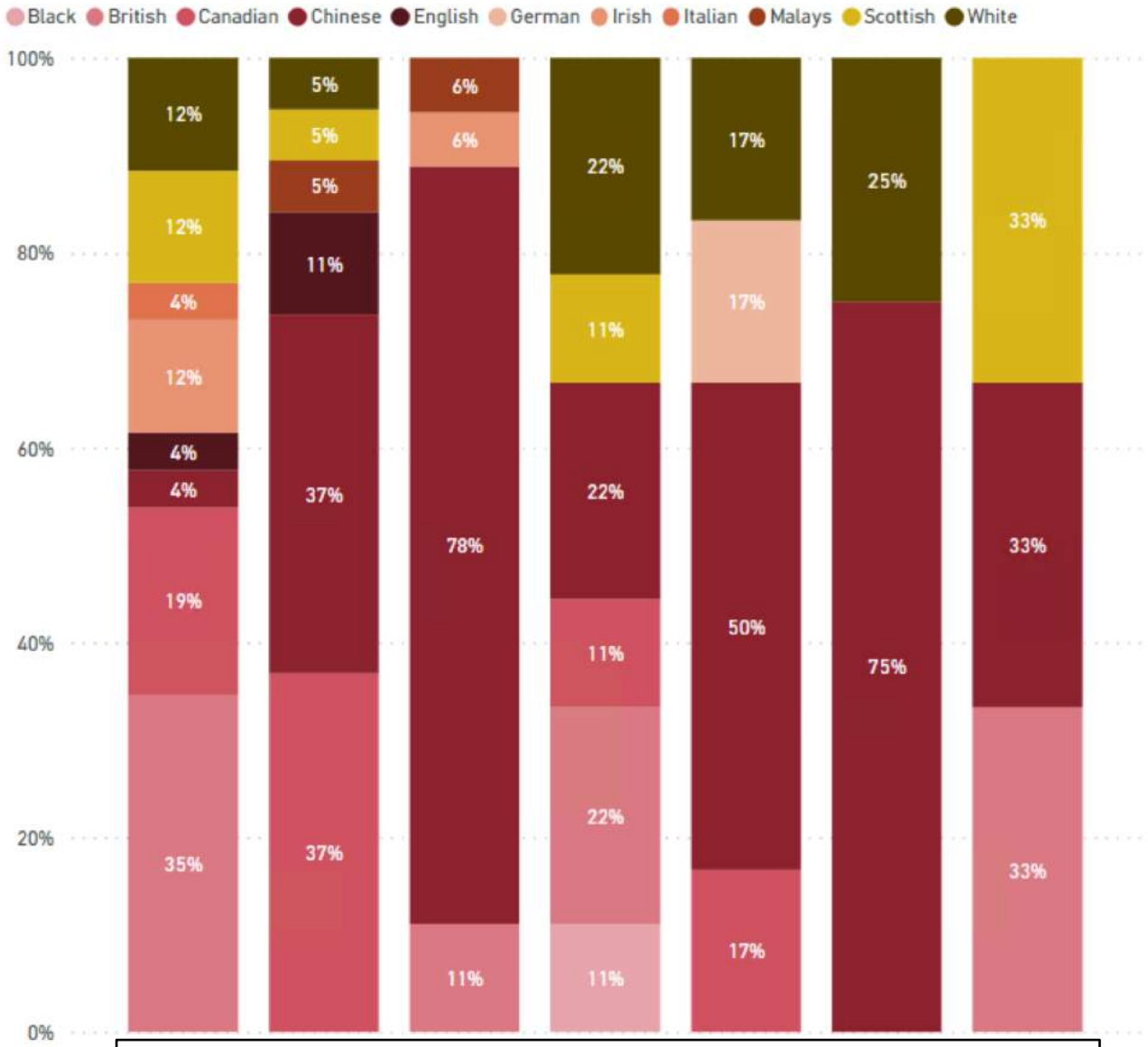
1) There are 32 employees who are currently paid at a lower grade level than their employment grade.

2) There are 40 employees who are currently paid at a higher grade level than their employment grade.

3) The largest age group with pay mismatch is the forties group, though the differences across age groups are small.

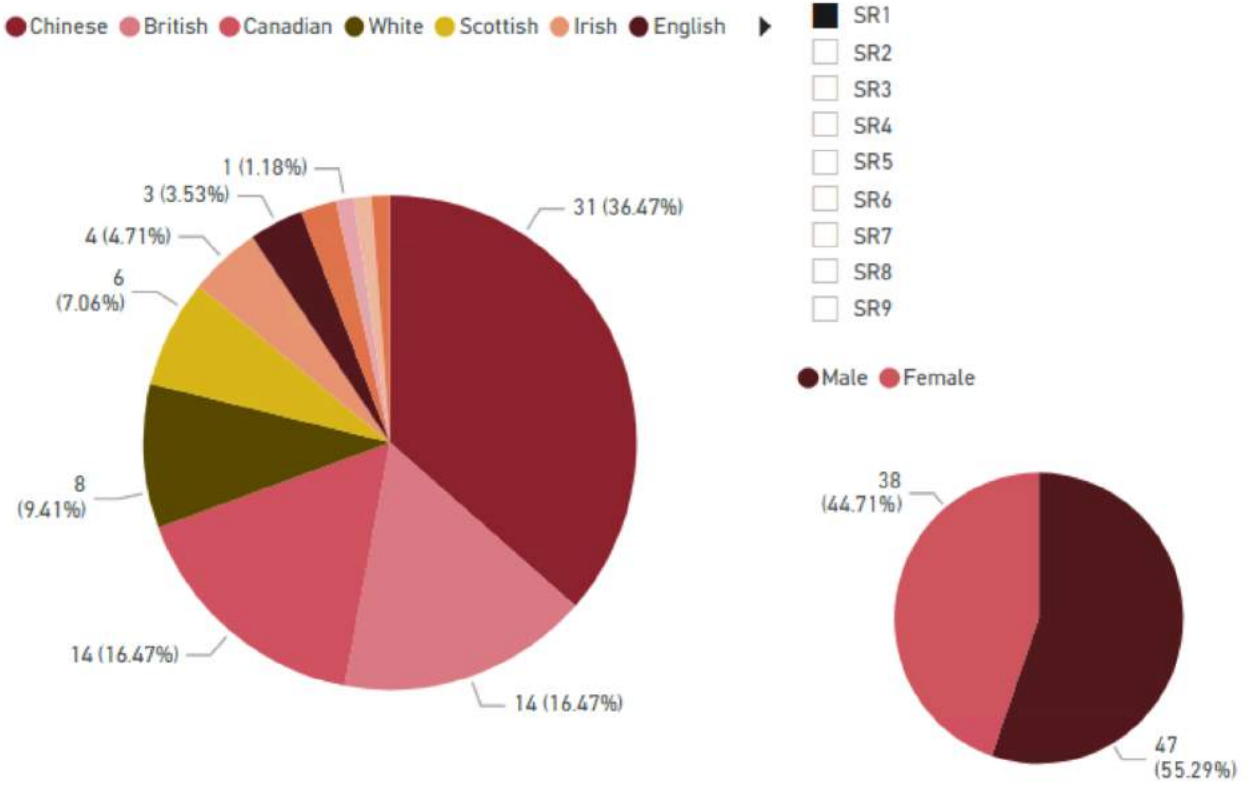
Ethnic distribution of high potential and high performance talents

Proportion of Ethnicities in Operating Responsibility 1



Confidential

Proportion of Ethnicities w.r.t Sector Rank



Insights

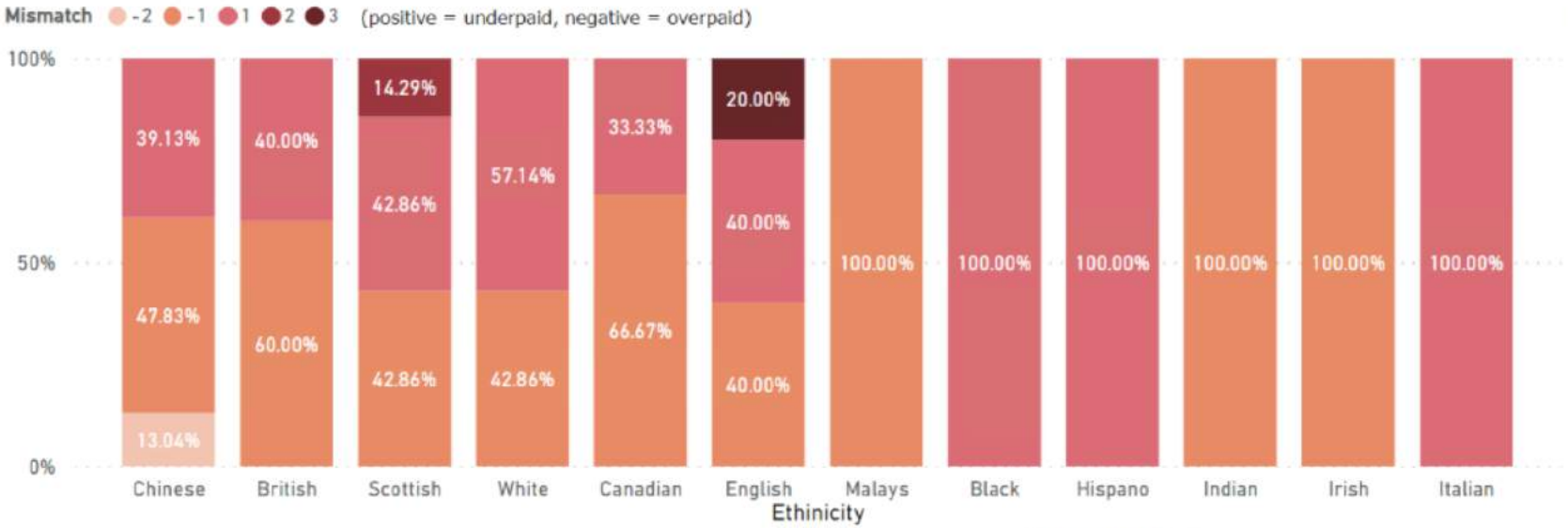
To check the dataset if there is any business unit which is ethnically diverse and also high performing.
(Using Operational Responsibility 1 as proxy for business units and Sector Rank 1 as proxy for high performance and high potential)

1) Filtering for SR1 and comparing across all ORs, there is no consistent diversity across OR.
2) The most diverse OR with equal proportion of representation from each ethnic group is

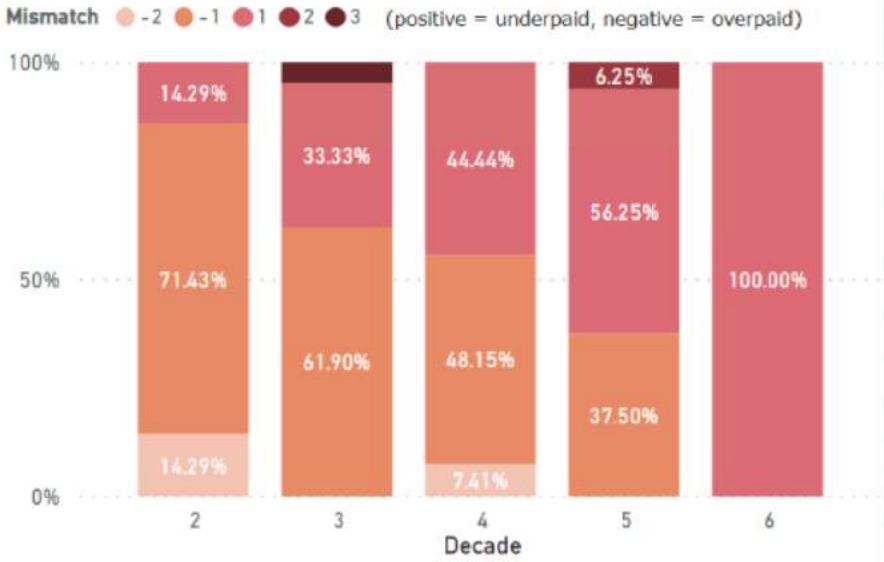
3) This compares to the overall SR1 employees across dataset, in which Chinese represent 36.47% of all SR1 graded employees.
4) Of note is that female employees only represent 44.71% of all employees graded SR1, below the 49.42% female representation in the data (refer to overall gender diversity).

Pay mismatch and its potential to increase risk of retention

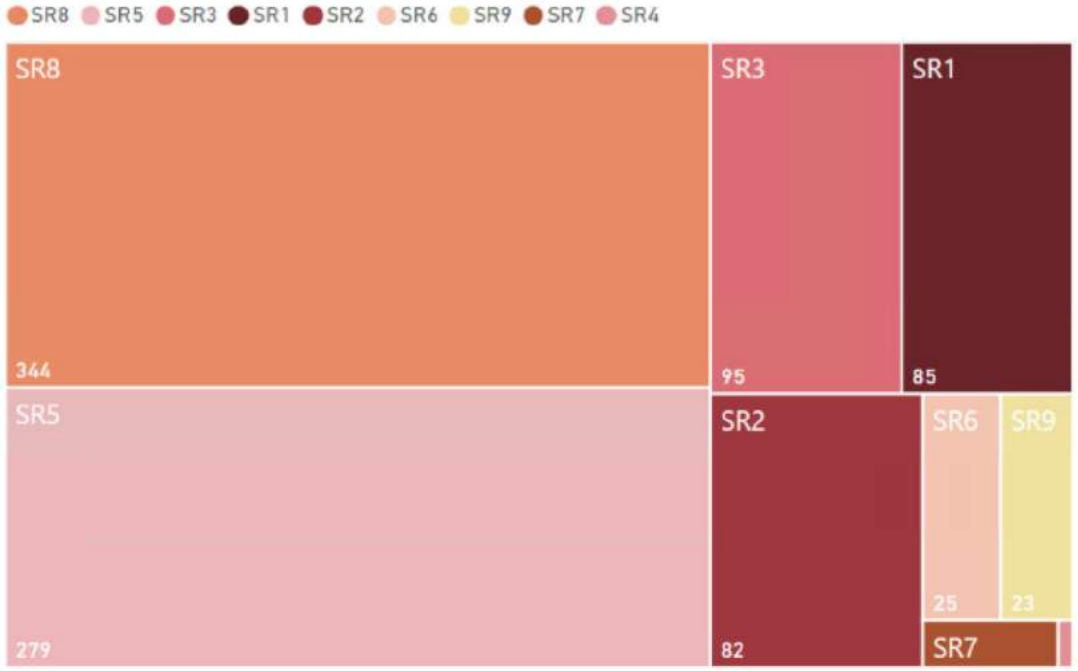
Ethnicities with Grades and Job Grades Mismatch



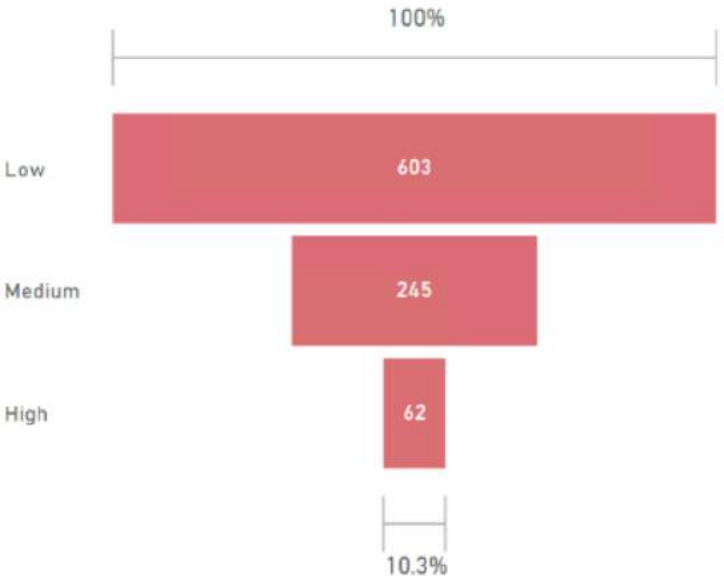
Age by Decades with Grades and Job Grades Mismatch



Sector Rank



Retention Risk

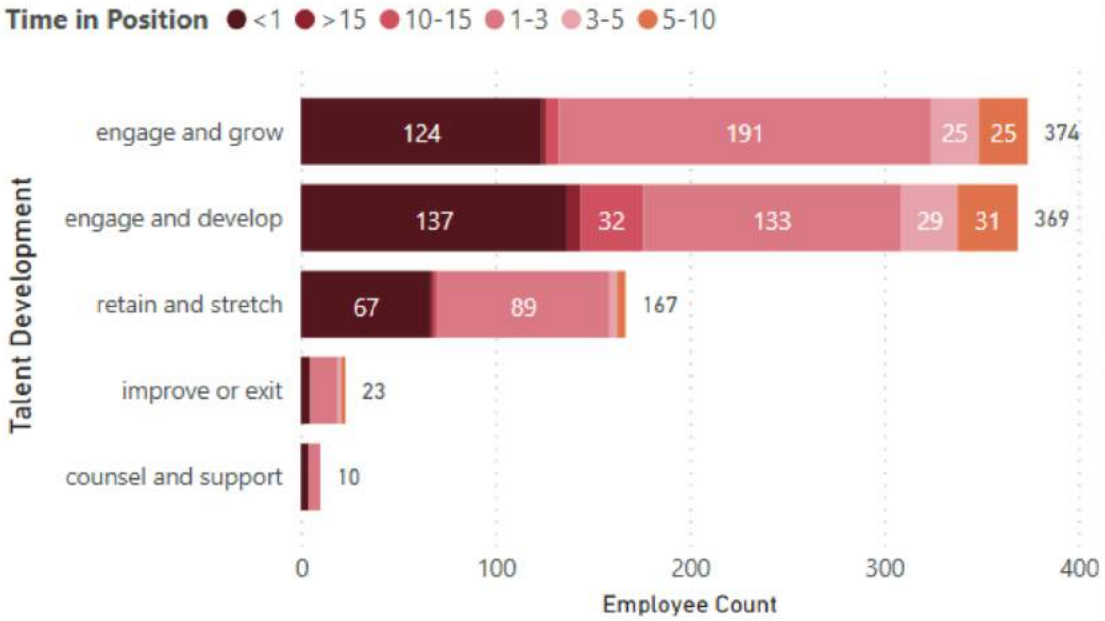


Employee Details

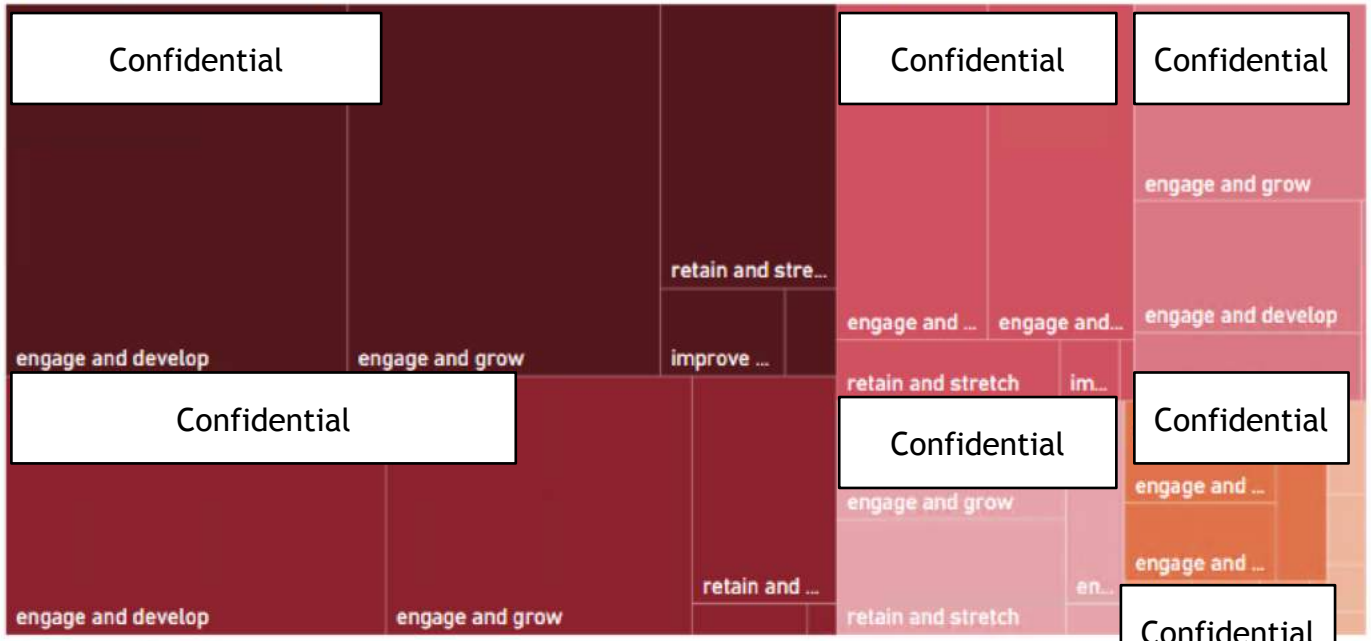
Confidential

Talent Development Overview

Talent Development Strategy compared to years experience in position



Operating Responsibility 1 compared to Talent Development Strategy



Time in compared to Timing of next move

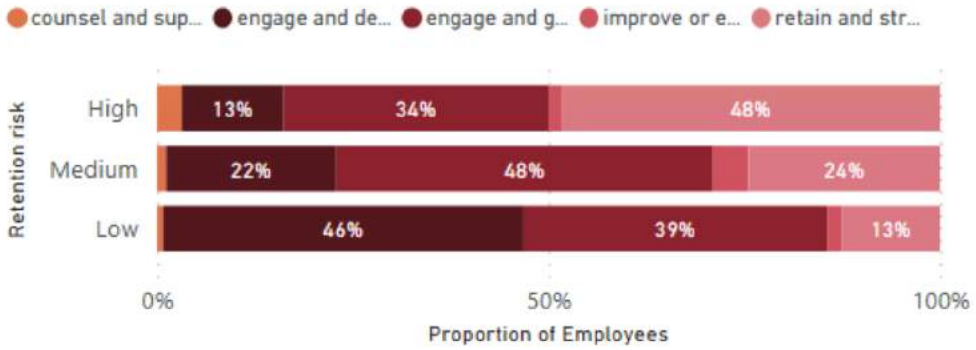
Employee Details

Confidential

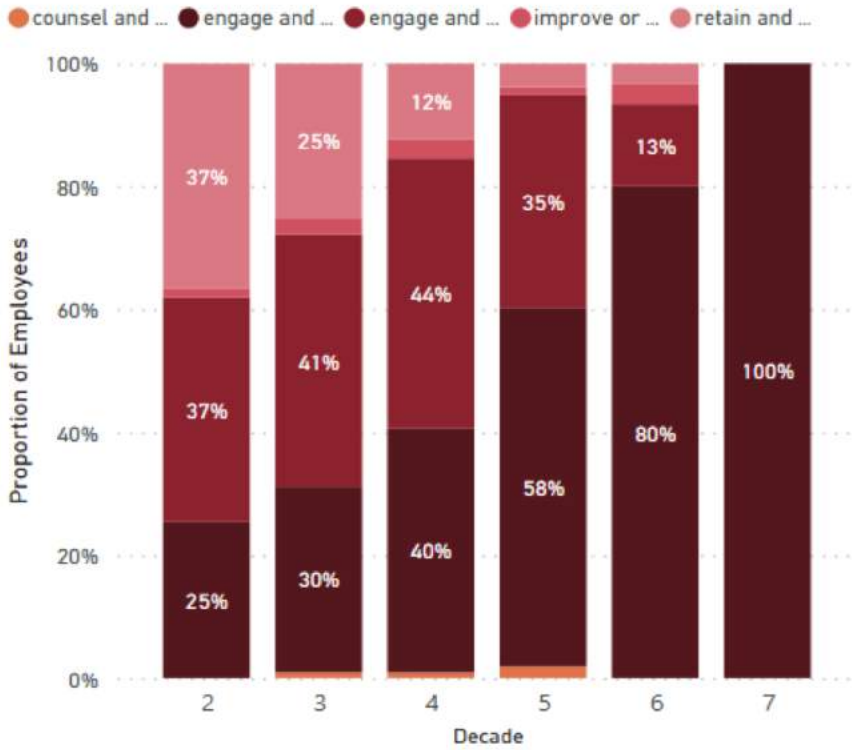
Talent Development

SR4 & 7 - Adjusting to role Action: Council & Support	SR1 & 2 - High Potential Action: Retain & Stretch
	SR3 & 5 - Growth Potential Action: Engage & Grow
SR9 - Action: Improve or Exit	SR6 & 8 - Highly Valued Contributor Action: Engage & Develop

Retention Risk



Age by Decades

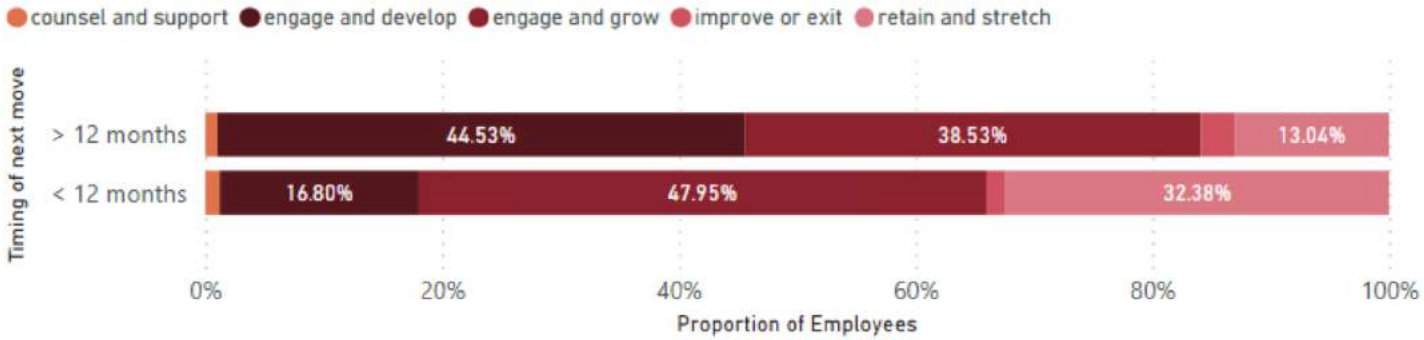


Grade
All

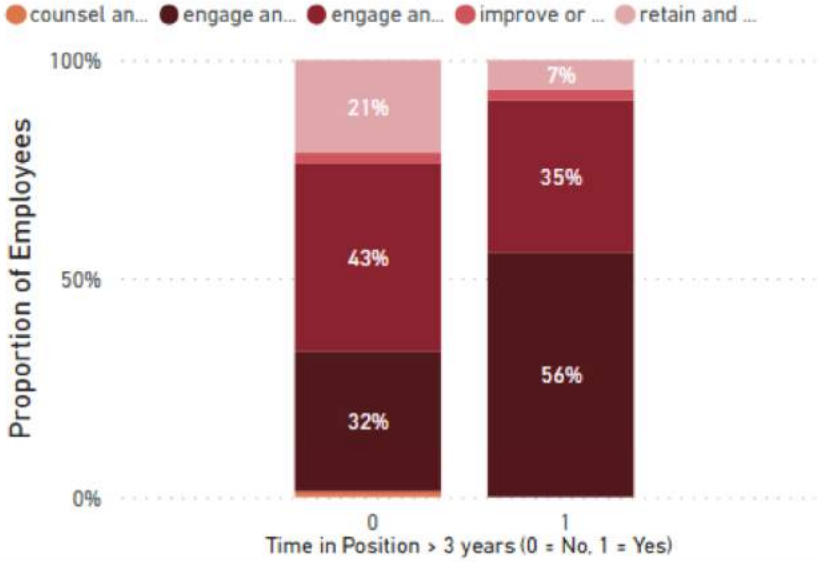
Job Grade
All

Ethinicity
All

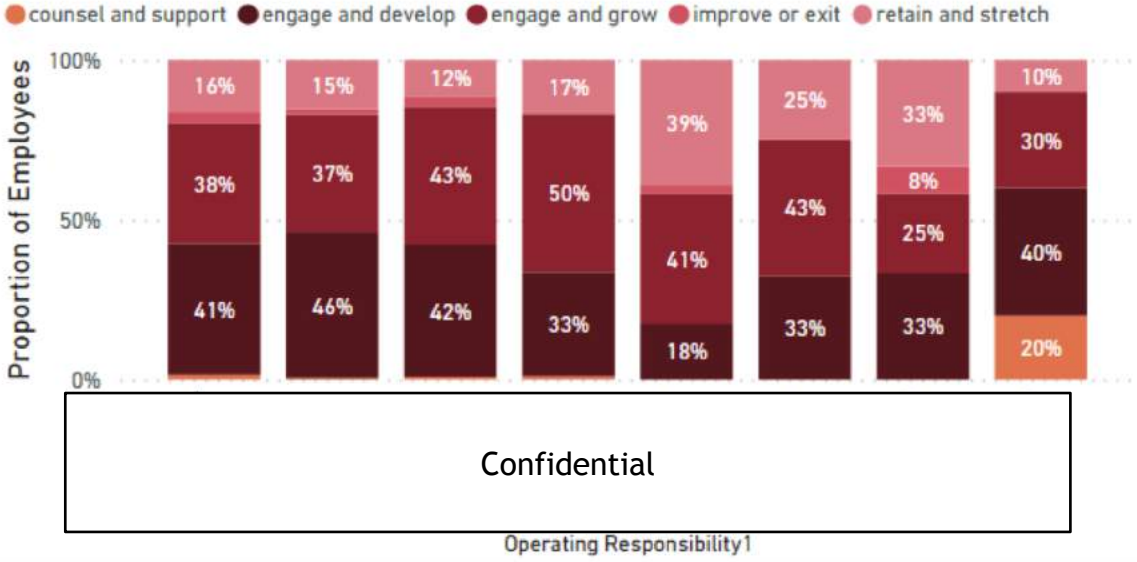
Timing of Next Move Status



Time in Position > 3 years



Operating Responsibility 1



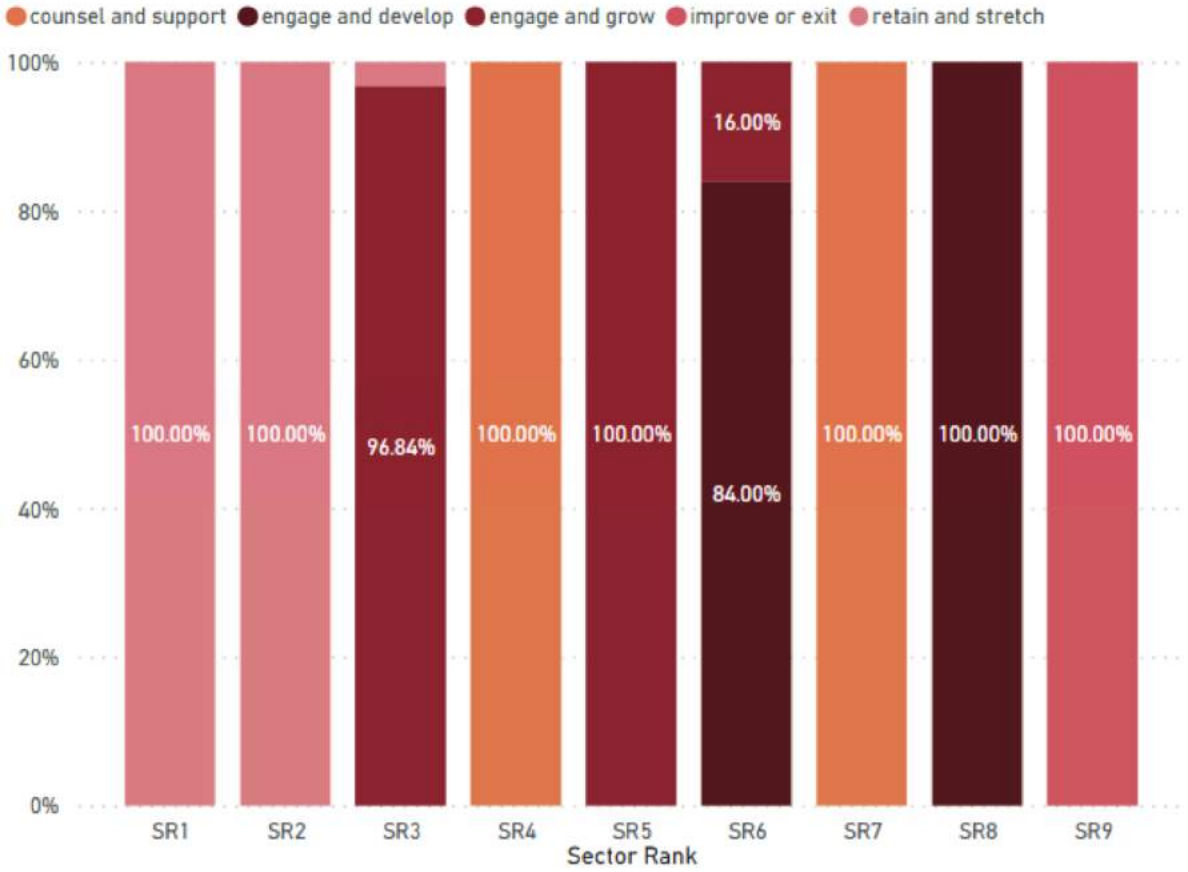
Confidential

Employees Details

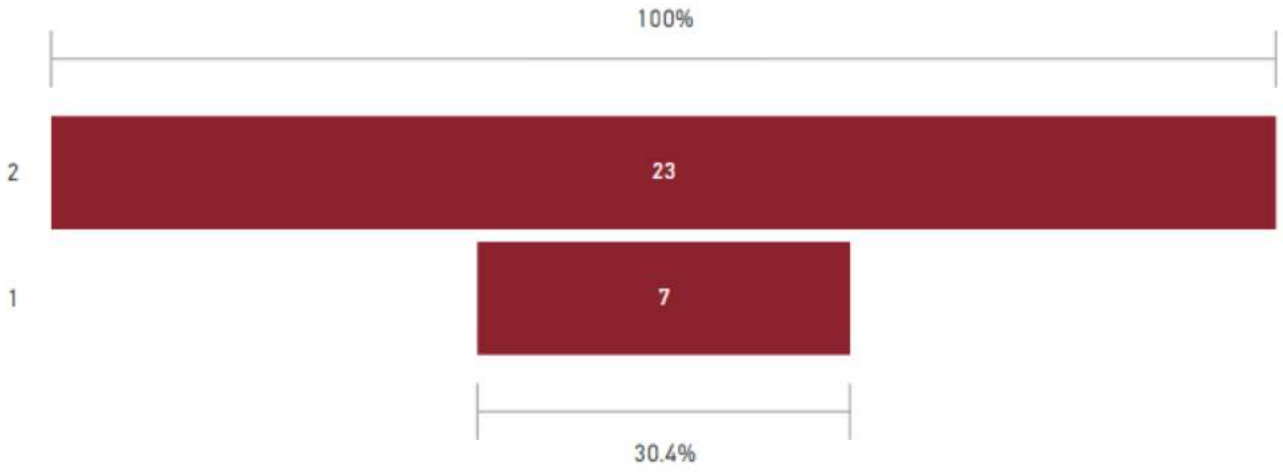
Confidential

Talent Development Effectiveness

Proportion of talents to be developed based on Sector Rank last year



Employees' Sector Rank compared to Previous Year (Positive = Drop in SR)



Employee Details

Confidential

Is our talent development program effective?

1) Bar chart shows the comparison of classification for people development strategy last year against the Sector Rank this year. Focus on the difference in colour in SR3 and SR6. The difference shows change in Sector Rank for individuals, thus requiring a different development strategy this year.

2) Funnel chart act as a filter to find individuals who have a change in Sector Rank from last year to this year. By clicking on the funnel bar, user can filter and find out what strategy the individuals had last year and compare to their current Sector Rank.

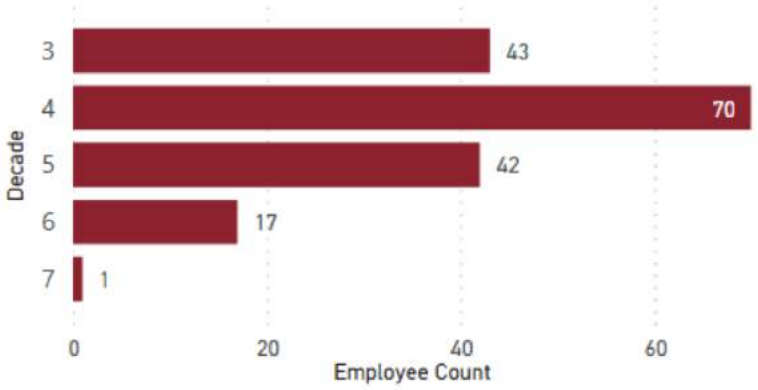
3) The table of employees output the filtered employees and their Operating Responsibility area, along with a comparison of their Sector Rank last year and this year.

4) We found that there are 23 employees who drop Sector Rank by 2 grades and 7 employees who drop Sector Rank by 1 grade.

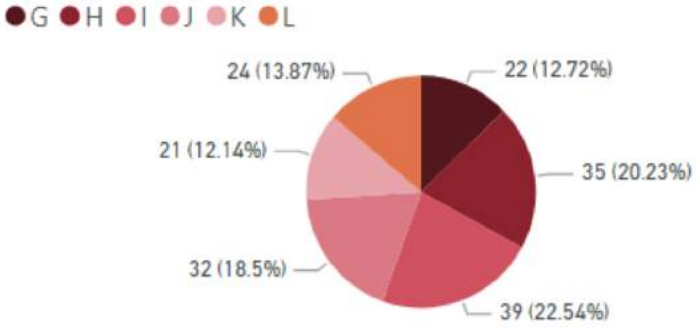
Career Mobility - Individuals in position for more than 3 years

Diversity

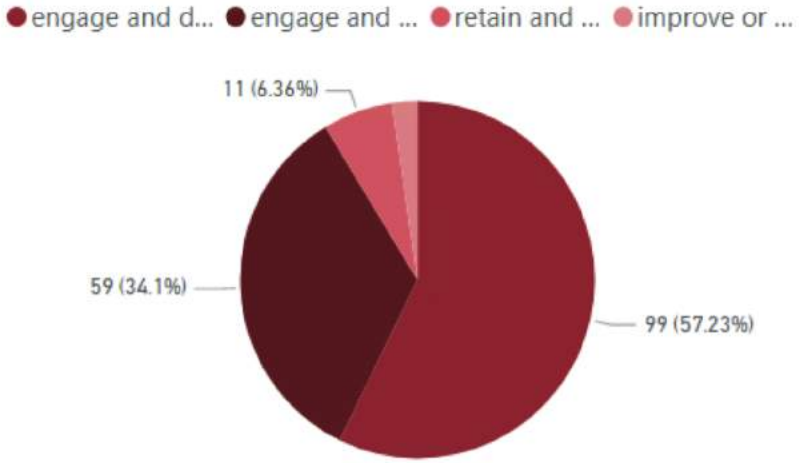
Age by Decade



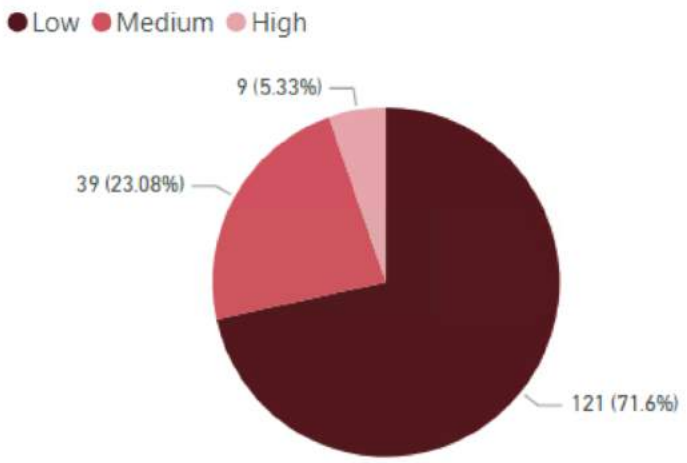
Grade



Talent Development



Talent Retention



Sector Rank

Employee Details

Confidential

Observations

For employees who has been in the position for more than 3 years:

1) The largest age group of employees are the forties age group.

2) 57.23% of employees in position more than 3 years are classified in highly valued contributor category, to be engaged and developed. 51.45% of such employees are classified under Sector Rank 8.

3) Only 5.33% of employees in position more than 3 years are noted to have high retention risk.

Insights:

Among employees who are in same position for more than 3 years, half of them are average performance with low potential (SR8). 68% of them are specialist and supervisors and more than 70% have a low retention risk. This large group of 89 employees will require attention in engagement and development to pick strong performer or arrange for lateral move.

2.0 Machine Learning

Classifier model with Performance (SR_Flag) as target

#	List of Features
1	User Country
2	Time in Confidential
3	Time_Position_gt_3
4	Retention risk
5	Decade
6	Grade num
7	Job grade num
8	Diff_Grade_Job_Grade
9	Seniority
10	Job Title
11	Operating Responsibility1
12	PromotionReadiness
13	Change of Responsibilities

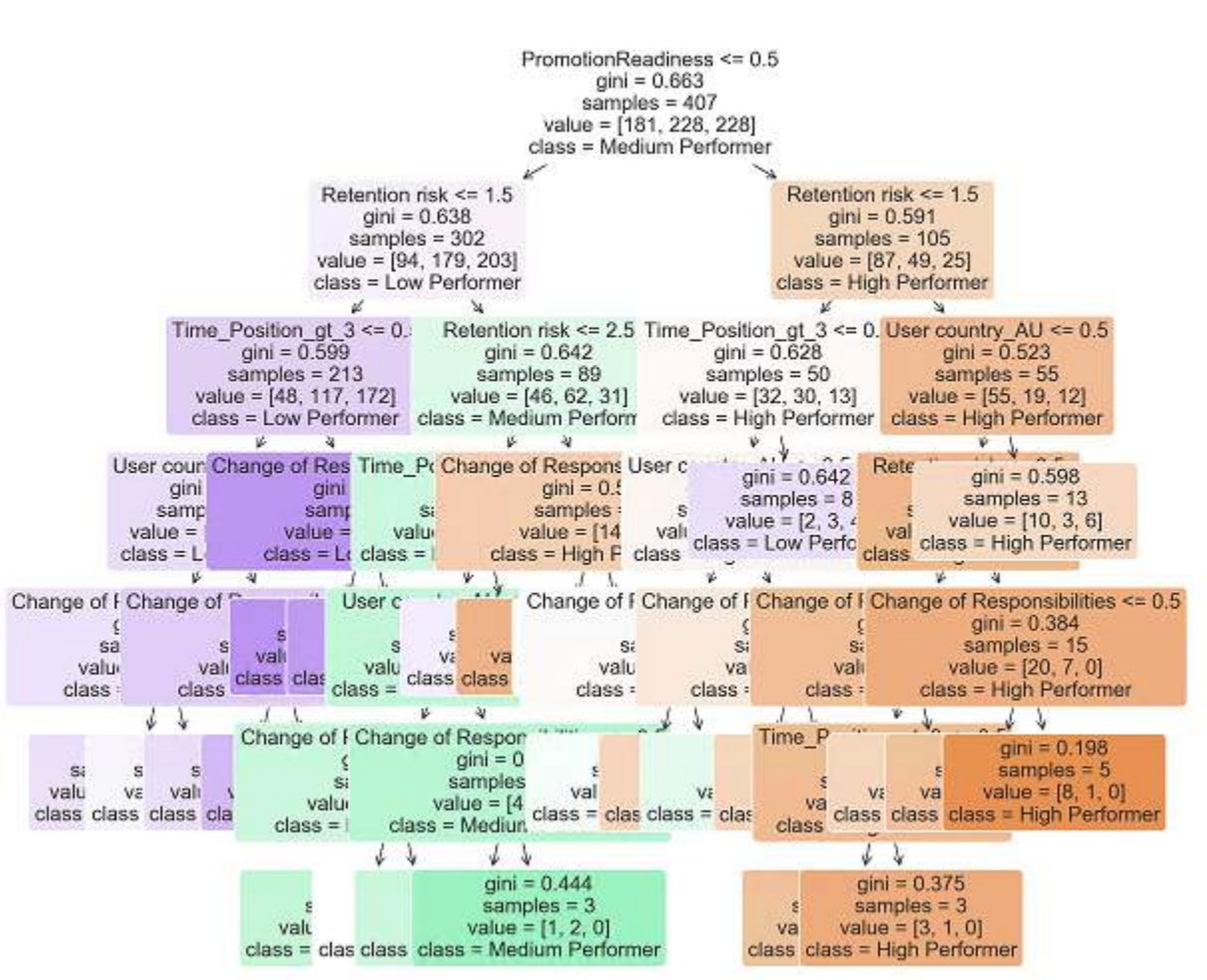
Supervised Machine Learning

Predicted SR_Flag

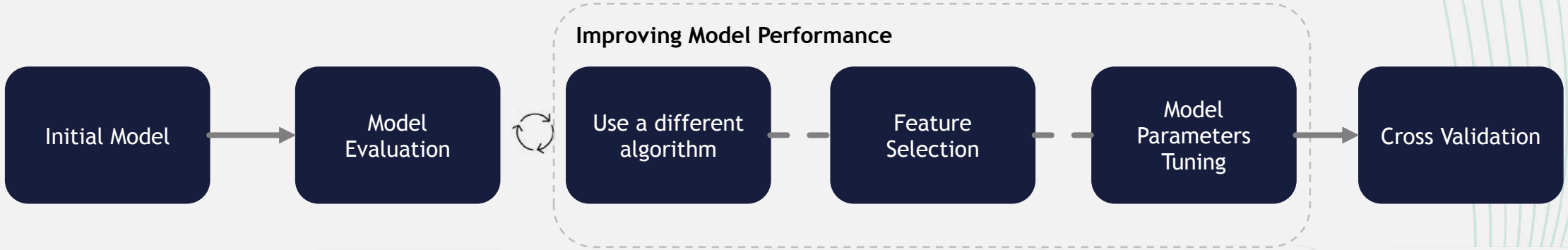
Model Evaluation:

Metric	Score
Accuracy	50%
Precision	47%
Recall	48%
F1-Score	47%

Decision Tree from SR_Flag RandomForest Classifier



Detailed workflow for Classifier for SR_Flag



Decision Tree Classifier

Metric	Score
Accuracy	41%
Precision	41%
Recall	40%
F1-Score	40%

RandomForest Classifier

Metric	Score
Accuracy	47%
Precision	47%
Recall	47%
F1-Score	47%

Permutation Importance

Metric	Score
Accuracy	50%
Precision	47%
Recall	48%
F1-Score	47%

Trimmed to 5 from 34 Features

RandomizedSearch & GridSearch

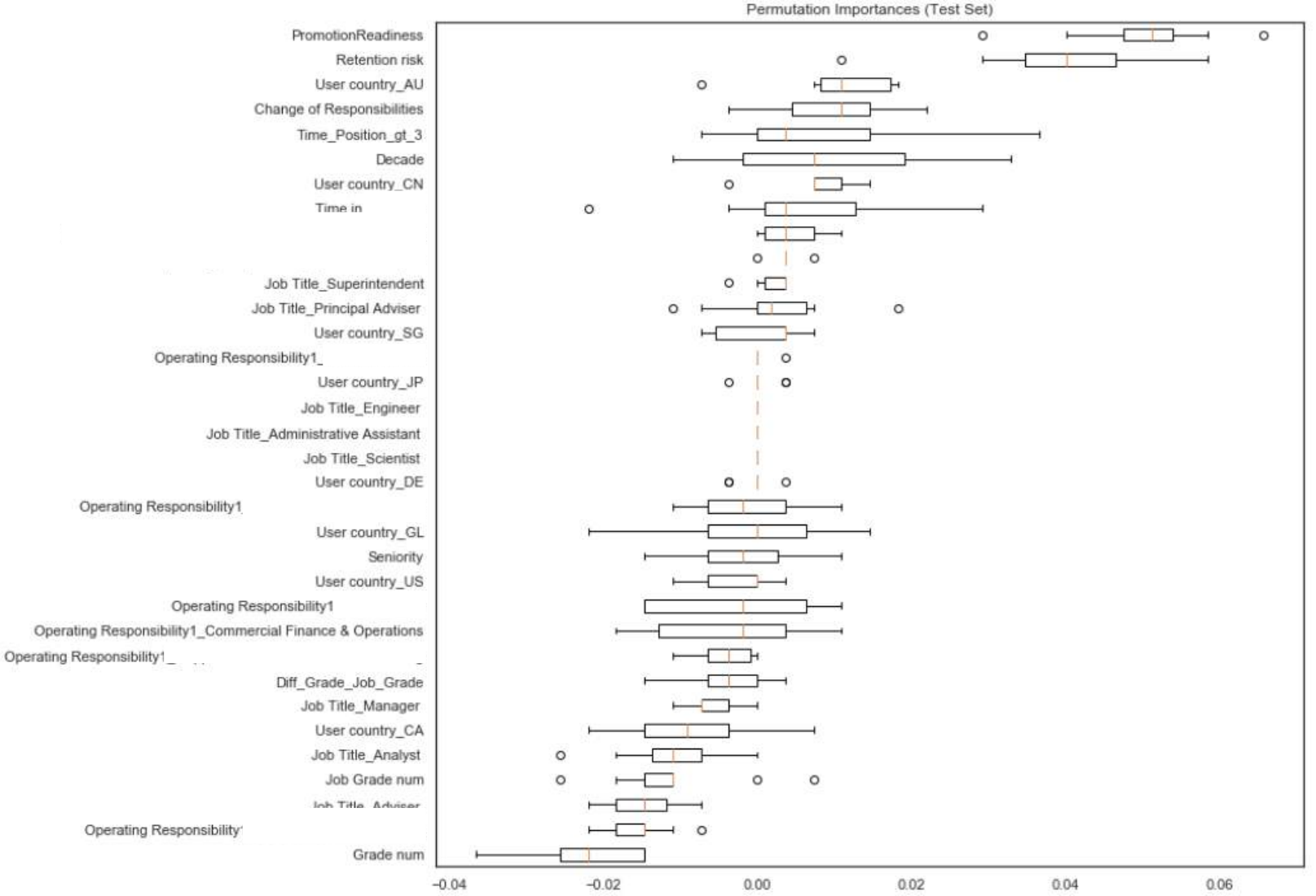
Metric	Score
Accuracy	50%
Precision	47%
Recall	48%
F1-Score	47%

- n_estimators: number of Decision Trees in RandomForest
- max_depth: max # of levels in trees
- max_features: # of features to consider at every split
- min_samples_split: min # of samples required to split a node
- min_samples_leaf: min # of samples required at each leaf node
- bootstrap: method of selecting samples for training each tree

Overall Improved Metrics

Metric	Change
Accuracy	+9%
Precision	+6%
Recall	+8%
F1-Score	+7%

Permutation Importances results for SR_Flag



Parameters:

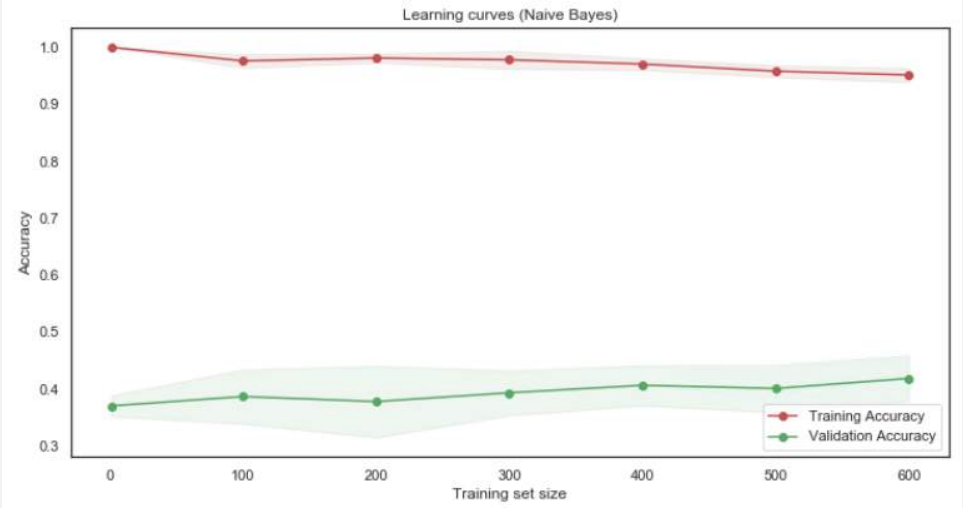
- Ran on Test-set
- Scoring: “Accuracy”
- n_repeats = 10

What is “Permutation Importance”?
It randomly shuffles a single column of Features and measures the impact to the accuracy of the model. With this, we can determine what are the most important predictors of our target variable.

Height at age 20 (cm)	Height at age 10 (cm)
182	155
175	147
...	...
156	142
153	130

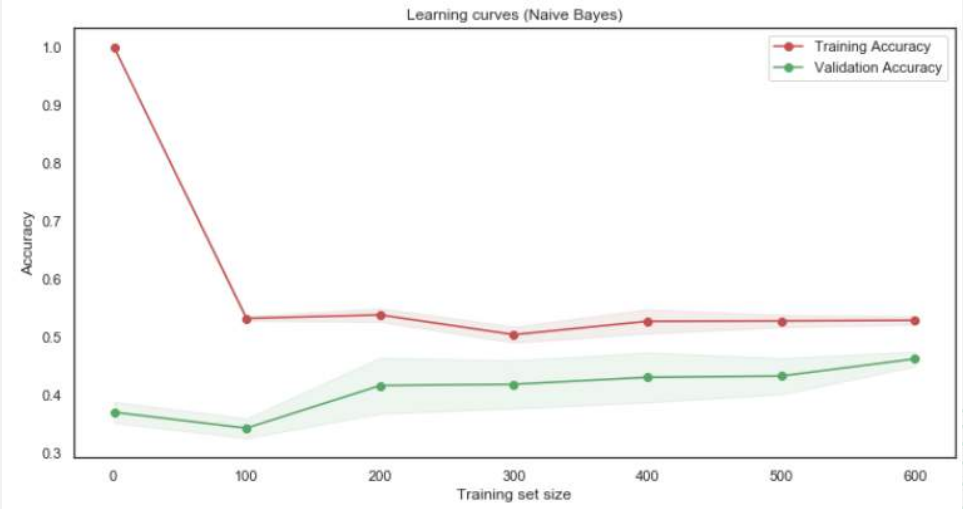
Source:
https://www.kaggle.com/dansbecker/permutation-importance

Bias-Variance tradeoff with Learning Curves (SR_Flag)



Baseline ("Out-of-bag") RandomForest:

- n_estimators: 100
- # of Features: 34



"Best" RandomForest:

- n_estimators: 1200
- max_depth: 30
- max_features: 3
- min_samples_leaf: 3
- min_samples_split: 4
- bootstrap: True
- # of Features: 5

Correlation Matrix of Features vs. Performance

-1.0 +1.0: Intra-feature correlation

-1.0 +1.0: Target vs. Feature correlation

Feature Importance	Feature	PromotionReadiness	Retention Risk	User Country_AU	Change of Responsibilities	Time_Position_gt_3
1	PromotionReadiness					
2	Retention Risk	**0.271** (Higher PR, Higher Risk)				
3	User Country_AU	**0.021** (Higher PR, Works in AU)	**-0.069** (Higher RR, Does not work in AU)			
4	Change of Responsibilities	**0.057** (Higher PR, Has CoR)	**0.025** (Higher RR, Has CoR)	**0.189** (Works in AU, Has CoR)		
5	Time_Position_gt_3	**-0.040** (Higher PR, Not >3 years in Position)	**-0.053** (Higher RR, Not >3 years in Position)	**-0.195** (Works in AU, Not >3 years in Position)	**-0.033** (Has CoR, Not >3 years in Position)	
Target Variable	SR_Flag	**-0.310** (Higher PR, Higher Performance)	**-0.269** (Higher RR, Higher Performance)	**0.055** (Works in AU, Lower Performance)	**-0.075** (Has CoR, Higher Performance	**0.155** (>3 years in Position, Lower Performance)

- PromotionReadiness: {“No”: 0, “Yes”: 1}
- Retention Risk: {“Low”: 1, “Medium”: 2, “High”: 3}
- User Country_AU: {“No”: 0, “Yes”: 1}
- Change of Responsibilities: {“No”: 0, “Yes”: 1}
- Time_Position_gt_3: {“No”: 0, “Yes”: 1}
- SR_Flag: {“High-Performer”: 1, “Mid-Performer”: 2, “Low-Performer”: 3}

Classifier model with Retention risk as target

#	List of Features
1	User Country
2	Time in
3	Time_Position_gt_5
4	SR_Flag
5	SRChange
6	Decade
7	Ethnicity
8	Grade num
9	Job Grade num
10	Diff_Grade_Job_Grade
11	Seniority
12	Operating Responsibility1
13	PromotionReadiness

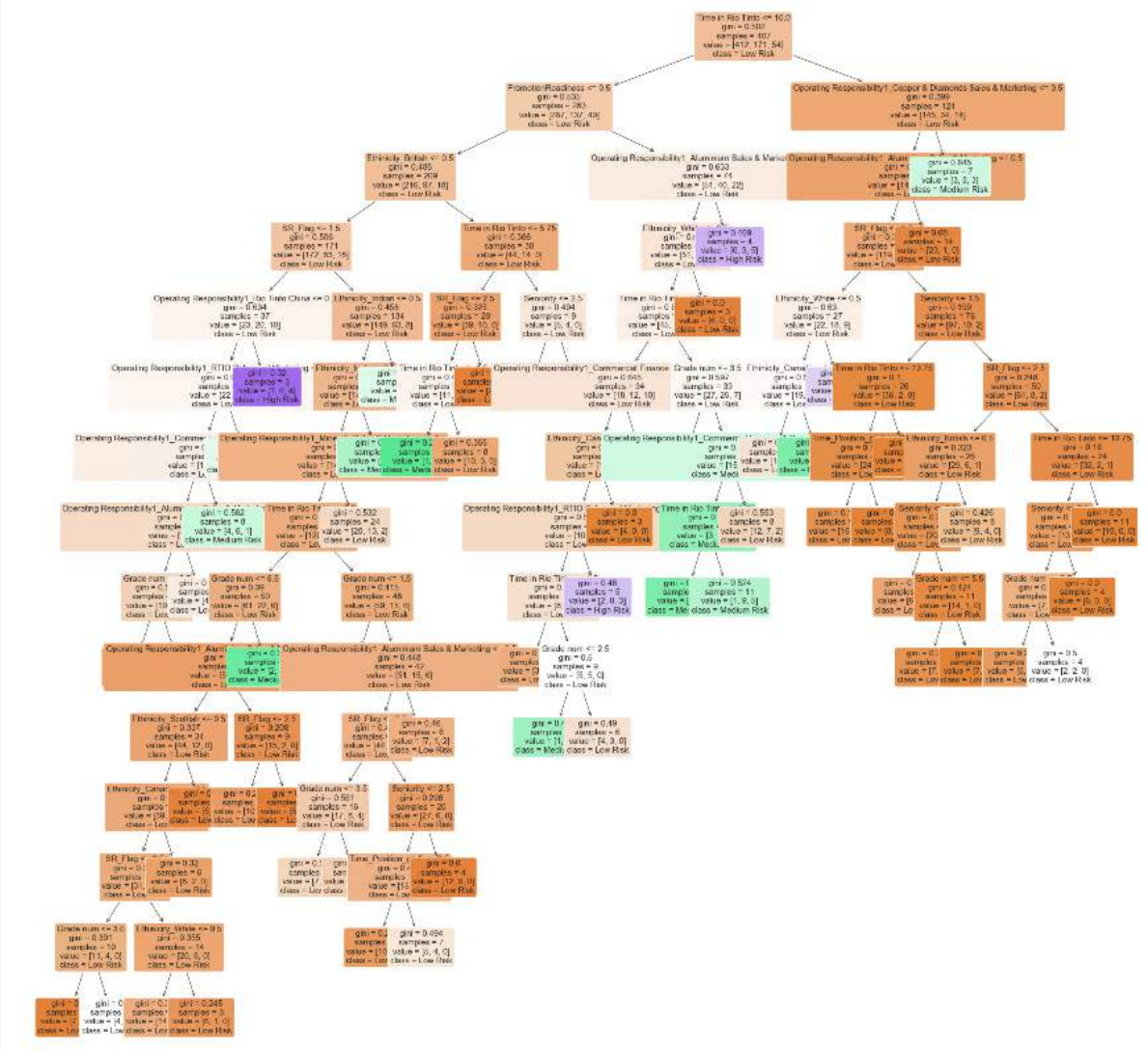
Supervised Machine Learning

Predicted Retention risk

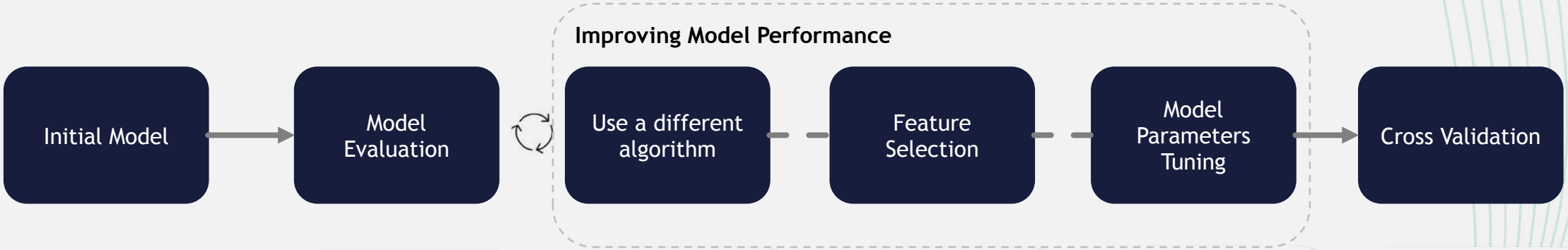
Model Evaluation:

Metric	Score
Accuracy	66%
Precision	58%
Recall	66%
F1-Score	58%

Decision Tree from Retention risk RandomForest Classifier



Detailed workflow for Classifier for Retention risk



Decision Tree Classifier

Metric	Score
Accuracy	53%
Precision	52%
Recall	53%
F1-Score	53%

RandomForest Classifier

Metric	Score
Accuracy	61%
Precision	56%
Recall	61%
F1-Score	57%

Permutation Importance

Metric	Score
Accuracy	64%
Precision	60%
Recall	64%
F1-Score	62%

Trimmed to 23 from 39 Features

RandomizedSearch & GridSearch

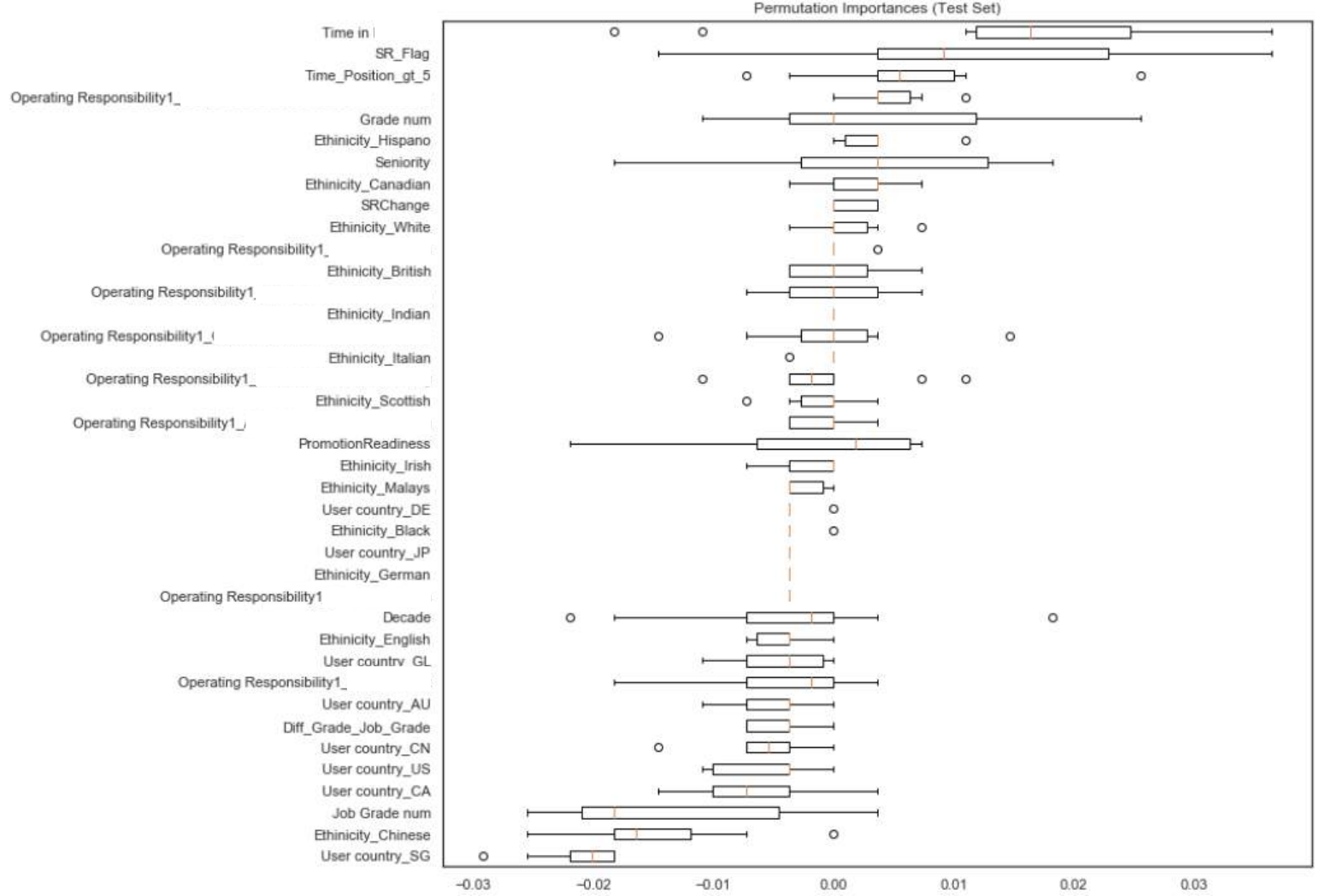
Metric	Score
Accuracy	66%
Precision	58%
Recall	66%
F1-Score	58%

- n_estimators: number of Decision Trees in RandomForest
- max_depth: max # of levels in trees
- max_features: # of features to consider at every split
- min_samples_split: min # of samples required to split a node
- min_samples_leaf: min # of samples required at each leaf node
- bootstrap: method of selecting samples for training each tree

Overall Improved Metrics

Metric	Change
Accuracy	+13%
Precision	+6%
Recall	+13%
F1-Score	+5%

Permutation Importances results for Retention risk



Parameters:

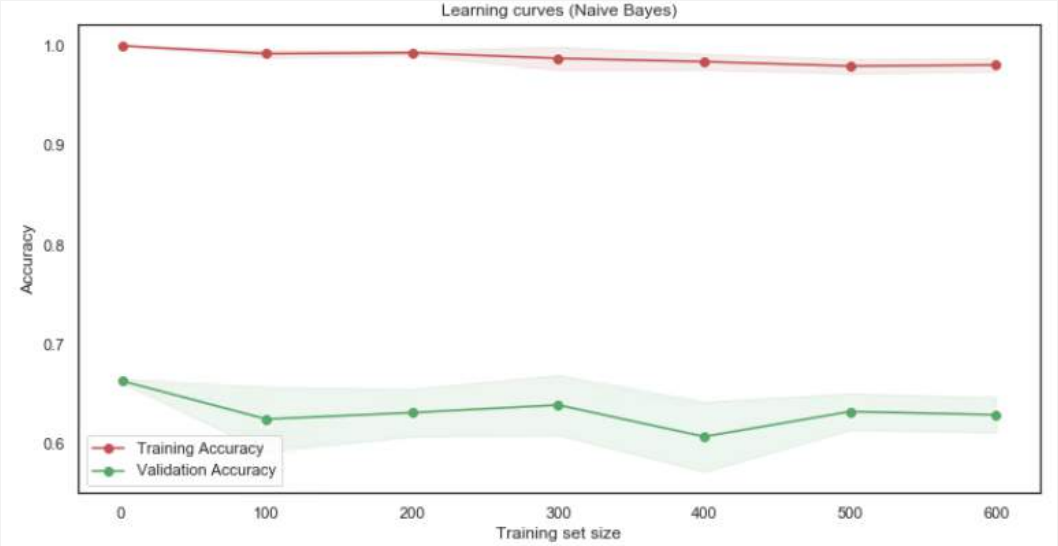
- Ran on Test-set
- Scoring: “Accuracy”
- n_repeats = 10

What is “Permutation Importance”?
It randomly shuffles a single column of Features and measures the impact to the accuracy of the model. With this, we can determine what are the most important predictors of our target variable.

Height at age 20 (cm)	Height at age 10 (cm)
182	155
175	147
...	...
156	142
153	130

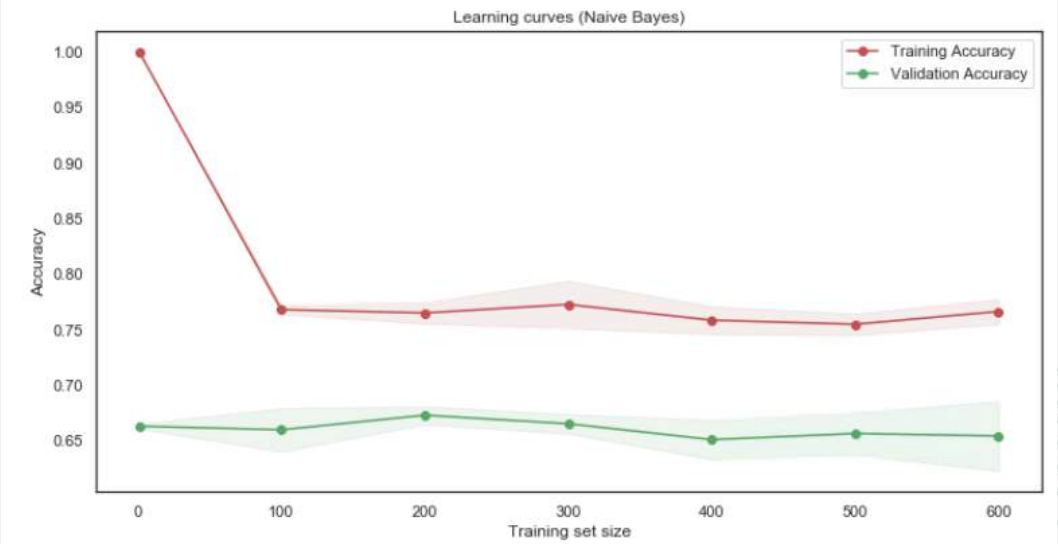
Source: https://www.kaggle.com/dansbecker/permutation-importance

Bias-Variance tradeoff - Learning Curves (Retention Risk)



Baseline ("Out-of-bag") RandomForest:

- n_estimators: 100
- # of Features: 39



"Best" RandomForest:

- n_estimators: 1200
- max_depth: 20
- max_features: 6
- min_samples_leaf: 3
- min_samples_split: 9
- bootstrap: True
- # of Features: 23

Correlation Matrix of Features vs. Retention Risk

-1.0 +1.0: Intra-feature correlation

-1.0 +1.0: Target vs. Feature correlation

Feature Importance	Feature	Years in Rio Tinto	Performance	Years in Position >5	Confidential	Job Grade
1	Years in Rio Tinto					
2	Performance	0.057 (Lower Performance, Higher Years)				
3	Years in Position >5	0.422 (Higher Years, >5 Years in Position)	0.148 (Lower Performance, >5 Years in Position)			
4	Confidential	0.083 (Higher Years, Works in	-0.011 (Higher Performance, Does Not Work in	0.100 (>5 Years in Position, Works In		
5	Job Grade	0.132 (Higher Years, Higher Job Grade)	-0.129 (Higher Performance, Higher Job Grade)	-0.001 (>5 Years in Position, Lower Job Grade)	0.029 (Works in , Higher Job Grade)	
Target Variable	Retention Risk	-0.107 (Higher Years, Lower Risk)	-0.269 (Lower Performance, Higher Risk)	-0.109 (>5 Years in Position, Lower Risk)	0.042 (Works in , Higher Risk)	0.112 (Higher Job Grade, Higher Risk)

1. Years in Rio Tinto: {1 - 15}
2. SR_Flag: {“High-Performer”: 1, “Mid-Performer”: 2, “Low-Performer”: 3}
3. Years in Position >5 {“No”: 0, “Yes”: 1}
4. Confidential {“No”: 0, “Yes”: 1}
5. Job Grade {1 (Lowest) – 6 (Highest)}
6. Retention Risk: {“Low”: 1, “Medium”: 2, “High”: 3}

Deep Dive into Clusters

	Cluster		Decade	Job Grade num	Time in	Time in Position	Sector Rank Raw num	Retention risk	PromotionReadiness	Cluster
“High-Flyers”	0	count	223.000000	223.000000	223.000000	223.000000	223.000000	223.000000	223.000000	223.0
		mean	4.542601	3.547085	12.652466	5.800448	6.623318	1.130045	0.103139	0.0
		std	0.803554	1.595681	2.764655	4.523647	1.959452	0.386883	0.304825	0.0
		min	3.000000	1.000000	2.000000	1.000000	1.000000	1.000000	0.000000	0.0
		25%	4.000000	2.000000	12.500000	2.000000	5.000000	1.000000	0.000000	0.0
		50%	4.000000	4.000000	12.500000	4.000000	8.000000	1.000000	0.000000	0.0
		75%	5.000000	5.000000	15.000000	7.500000	8.000000	1.000000	0.000000	0.0
		max	7.000000	6.000000	15.000000	15.000000	9.000000	3.000000	1.000000	0.0
“Cruisers”	1	count	466.000000	466.000000	466.000000	466.000000	466.000000	466.000000	466.000000	466.0
		mean	3.403433	3.611588	4.054721	1.663090	5.435622	1.379828	0.000000	1.0
		std	0.767841	1.435867	3.295763	0.829049	2.376665	0.563725	0.000000	0.0
		min	2.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	1.0
		25%	3.000000	3.000000	2.000000	1.000000	5.000000	1.000000	0.000000	1.0
		50%	3.000000	4.000000	2.000000	2.000000	5.000000	1.000000	0.000000	1.0
		75%	4.000000	5.000000	7.500000	2.000000	8.000000	2.000000	0.000000	1.0
		max	6.000000	7.000000	15.000000	7.500000	9.000000	3.000000	0.000000	1.0
“Adjusting Newbies”	2	count	222.000000	222.000000	222.000000	222.000000	222.000000	222.000000	222.000000	222.0
		mean	3.423423	3.554054	5.664414	1.995495	3.810811	1.734234	0.995495	2.0
		std	0.897957	1.499399	4.366288	1.407793	2.368777	0.740917	0.067116	0.0
		min	2.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	2.0
		25%	3.000000	2.000000	2.000000	1.000000	2.000000	1.000000	1.000000	2.0
		50%	3.000000	4.000000	4.000000	2.000000	3.000000	2.000000	1.000000	2.0
		75%	4.000000	5.000000	7.500000	2.000000	5.000000	2.000000	1.000000	2.0
		max	6.000000	7.000000	15.000000	12.500000	9.000000	3.000000	1.000000	2.0

Clustering detailed view: Group by (Part 1/2)

Cluster 0:

Decade		
2	28	13%
3	102	46%
4	66	30%
5	22	10%
6	4	2%
7	0	0%
	222	100%

Pay Grade		
1	27	12%
2	30	14%
3	45	20%
4	55	25%
5	44	20%
6	20	9%
7	1	0%
	222	100%

Time in [redacted]		
1	19	9%
2	77	35%
4	33	15%
7.5	46	21%
12.5	39	18%
15	8	4%
	222	100%

Time in Position		
1	72	32%
2	131	59%
4	11	5%
7.5	7	3%
12.5	1	0%
15	0	0%
	222	100%

Sector Rank		
1	54	24%
2	24	11%
3	40	18%
4	0	0%
5	67	30%
6	4	2%
7	3	1%
8	26	12%
9	4	2%
	222	100%

Retention Risk		
1	98	44%
2	85	38%
3	39	18%
	222	100%

PromotionReadiness		
0	1	0%
1	221	100%
	222	100%

Cluster 1:

Decade		
2	0	0%
3	16	7%
4	98	44%
5	84	38%
6	25	11%
7	1	0%
	224	100%

Pay Grade		
1	31	14%
2	32	14%
3	45	20%
4	43	19%
5	43	19%
6	30	13%
7	0	0%
	224	100%

Time in [redacted]		
1	0	0%
2	1	0%
4	3	1%
7.5	32	14%
12.5	96	43%
15	92	41%
	224	100%

Time in Position		
1	29	13%
2	71	32%
4	23	10%
7.5	51	23%
12.5	40	18%
15	10	4%
	224	100%

Sector Rank		
1	2	1%
2	6	3%
3	21	9%
4	0	0%
5	47	21%
6	11	5%
7	0	0%
8	132	59%
9	5	2%
	224	100%

Retention Risk		
1	199	89%
2	21	9%
3	4	2%
	224	100%

PromotionReadiness		
0	201	90%
1	23	10%
	224	100%

Cluster 2:

Decade		
2	32	7%
3	262	56%
4	124	27%
5	46	10%
6	1	0%
7	0	0%
	465	100%

Pay Grade		
1	44	9%
2	62	13%
3	106	23%
4	121	26%
5	86	18%
6	45	10%
7	1	0%
	465	100%

Time in [redacted]		
1	79	17%
2	180	39%
4	77	17%
7.5	97	21%
12.5	31	7%
15	1	0%
	465	100%

Time in Position		
1	216	46%
2	223	48%
4	24	5%
7.5	2	0%
12.5	0	0%
15	0	0%
	465	100%

Sector Rank		
1	28	6%
2	52	11%
3	34	7%
4	1	0%
5	165	35%
6	10	2%
7	6	1%
8	155	33%
9	14	3%
	465	100%

Retention Risk		
1	307	66%
2	139	30%
3	19	4%
	465	100%

PromotionReadiness		
0	465	100%
1	0	0%
	465	100%

“High-Flyers”

Cluster 0: High Performing employees with highest risk of Retention & High Promotion Readiness "High-Flyers"

Group Defined by

- Highest sector ranks (“Sector Rank Raw num”)
- Highest retention risk
- Highest promotion readiness (“Promotion Readiness”)
- Medium time in employment position (“Time in Position”)

“Cruisers”

Cluster 1: Longest serving employees with Lowest Sector Ranks & Low Promotion Readiness "Cruisers"

Group Defined by

- Longest employment time in *Company*(“Time in *Company*”)
- Longest time in employment position (“Time in Position”)
- Lowest positions in sector rank (“Sector Rank Raw num”)
- Lowest retention risk (“Retention risk”)
- Low Promotion Readiness (“Promotion Readiness”)

“Adjusting Newbies”

Cluster 2: Shortest serving, average performance employees with medium risk of Retention & no Promotion Readiness "Adjusting"

Group Defined by

- Lowest time in *Company*(“Time in *Company*”)
- Lowest time in employment position (“Time in Position”)
- Average to higher sector ranks (“Sector Rank Raw num”)
- Medium retention risk (“Retention risk”)
- Lowest promotion readiness (“Promotion Readiness”) – possibly due to short employment duration

Clustering detailed view: Group by (Part 2/2)

Cluster 0:

Decade		
2	28	13%
3	102	46%
4	66	30%
5	22	10%
6	4	2%
7	0	0%
	222	100%

Pay Grade		
1	27	12%
2	30	14%
3	45	20%
4	55	25%
5	44	20%
6	20	9%
7	1	0%
	222	100%

Time in [redacted]		
1	19	9%
2	77	35%
4	33	15%
7.5	46	21%
12.5	39	18%
15	8	4%
	222	100%

Time in Position		
1	72	32%
2	131	59%
4	11	5%
7.5	7	3%
12.5	1	0%
15	0	0%
	222	100%

Sector Rank		
1	54	24%
2	24	11%
3	40	18%
4	0	0%
5	67	30%
6	4	2%
7	3	1%
8	26	12%
9	4	2%
	222	100%

Retention Risk		
1	98	44%
2	85	38%
3	39	18%
	222	100%

PromotionReadiness		
0	1	0%
1	221	100%
	222	100%

Cluster 1:

Decade		
2	0	0%
3	16	7%
4	98	44%
5	84	38%
6	25	11%
7	1	0%
	224	100%

Pay Grade		
1	31	14%
2	32	14%
3	45	20%
4	43	19%
5	43	19%
6	30	13%
7	0	0%
	224	100%

Time in [redacted]		
1	0	0%
2	1	0%
4	3	1%
7.5	32	14%
12.5	96	43%
15	92	41%
	224	100%

Time in Position		
1	29	13%
2	71	32%
4	23	10%
7.5	51	23%
12.5	40	18%
15	10	4%
	224	100%

Sector Rank		
1	2	1%
2	6	3%
3	21	9%
4	0	0%
5	47	21%
6	11	5%
7	0	0%
8	132	59%
9	5	2%
	224	100%

Retention Risk		
1	199	89%
2	21	9%
3	4	2%
	224	100%

PromotionReadiness		
0	201	90%
1	23	10%
	224	100%

Cluster 2:

Decade		
2	32	7%
3	262	56%
4	124	27%
5	46	10%
6	1	0%
7	0	0%
	465	100%

Pay Grade		
1	44	9%
2	62	13%
3	106	23%
4	121	26%
5	86	18%
6	45	10%
7	1	0%
	465	100%

Time in [redacted]		
1	79	17%
2	180	39%
4	77	17%
7.5	97	21%
12.5	31	7%
15	1	0%
	465	100%

Time in Position		
1	216	46%
2	223	48%
4	24	5%
7.5	2	0%
12.5	0	0%
15	0	0%
	465	100%

Sector Rank		
1	28	6%
2	52	11%
3	34	7%
4	1	0%
5	165	35%
6	10	2%
7	6	1%
8	155	33%
9	14	3%
	465	100%

Retention Risk		
1	307	66%
2	139	30%
3	19	4%
	465	100%

PromotionReadiness		
0	465	100%
1	0	0%
	465	100%

Decade
- 0: Mostly 30-40s (76%)
- 1: Mostly 40-50s (82%)
- 2: Mostly 30-40s (83%)

Pay Grade
- 0: 3-5 (65%)
- 1: 3-5 (58%)
- 2: 3-5 (67%)

Time in *Company*
- 0: 2 (35%), 7.5 (21%)
- 1: 12.5-15 (84%)
- 2: 1-2 (56%)

Time in Position
- 0: 1-2 (91%)
- 1: 1-2 (45%), 7.5 (23%)
- 2: 1-2 (94%)

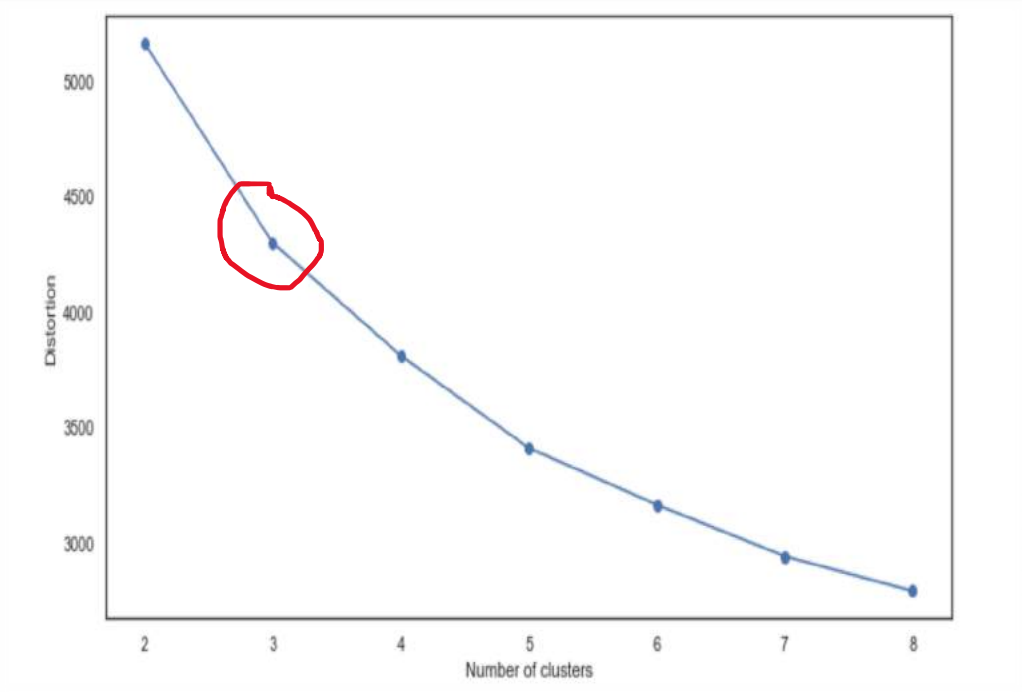
Sector Rank
- 0: 1-3 (53%)
- 1: 8 (59%)
- 2: 5 (35%)

Retention Risk:
- 0: 1-2 (82%), 3 (18%)
- 1: 1 (89%)
- 2: 1-2 (96%)

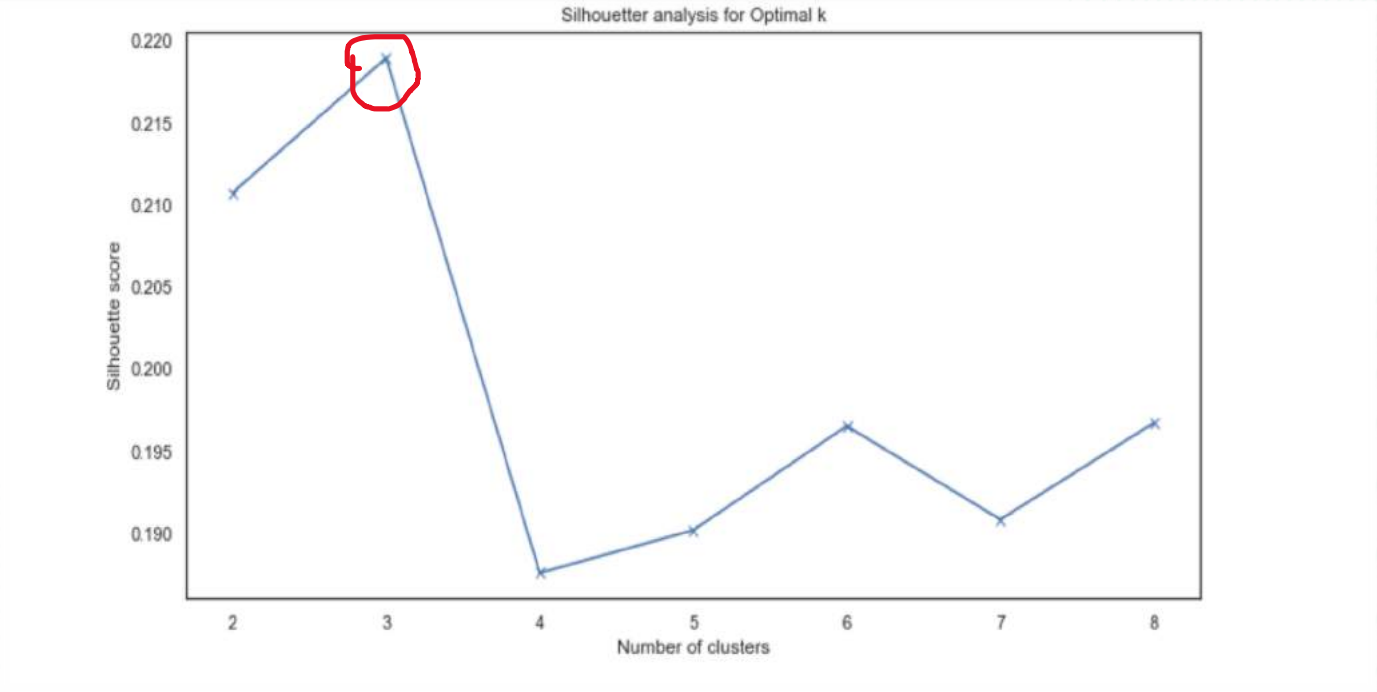
PromotionReadiness:
- 0: 1 (100%) Ready
- 1: 0 (90%) Not Ready
- 2: 0 (100%) Not Ready

How do we determine the optimal number of clusters for the Clustering model

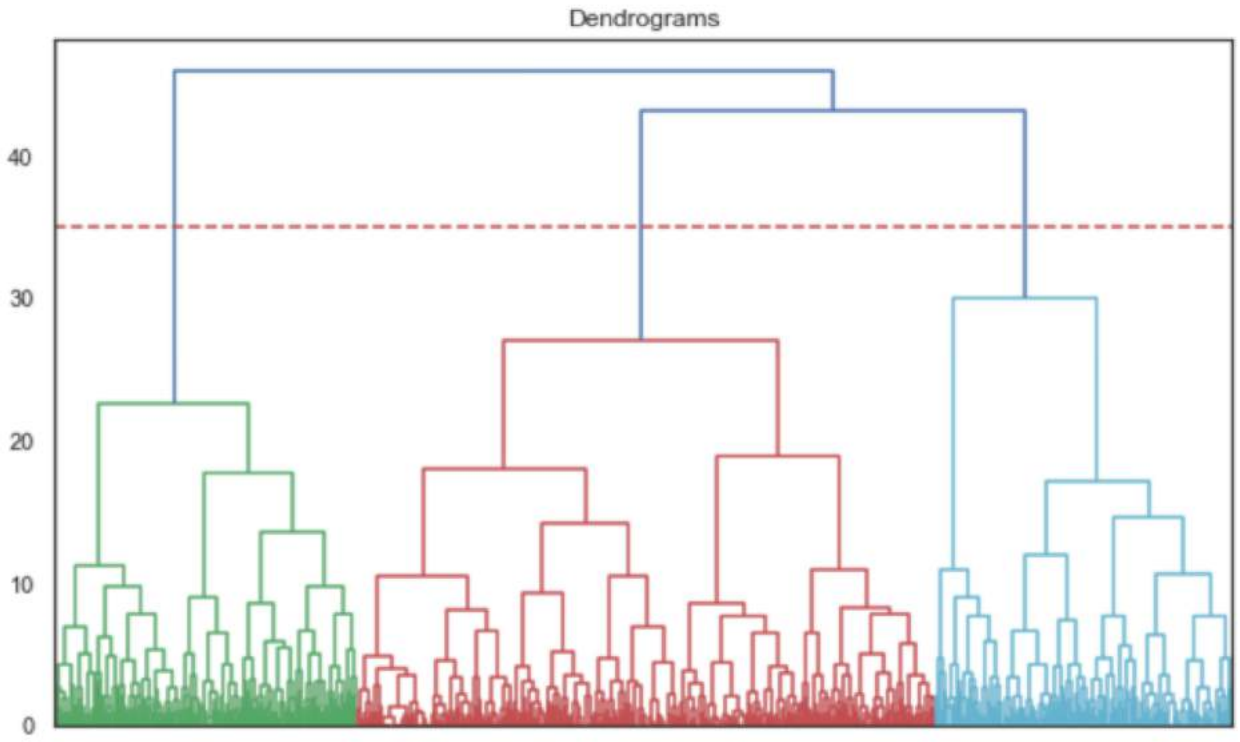
Finding optimal number of clusters with Elbow curve



Silhouette analysis for optimal number of clusters



Dendrogram: Measuring similarity across features



- Dendrogram generated to test hierarchical clustering
- Similar heights of the graph describes similarity in features
- Dissimilarity is measured by distance between features on the x axis