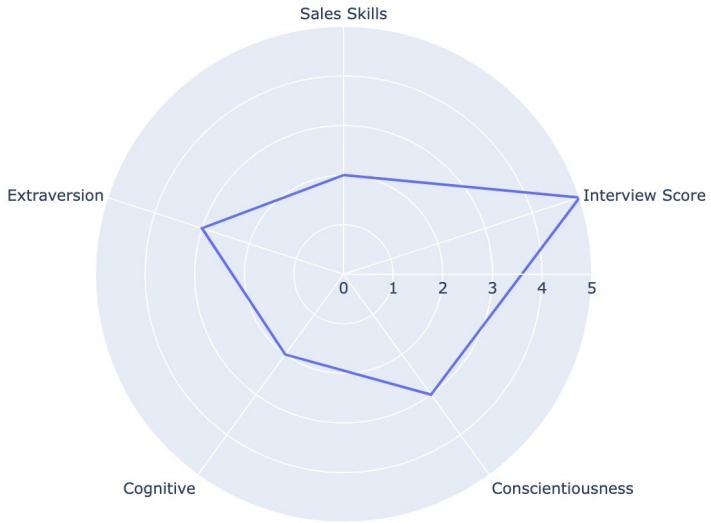
Talent Analytics Project:

Applicant Selection Analysis

Introduction

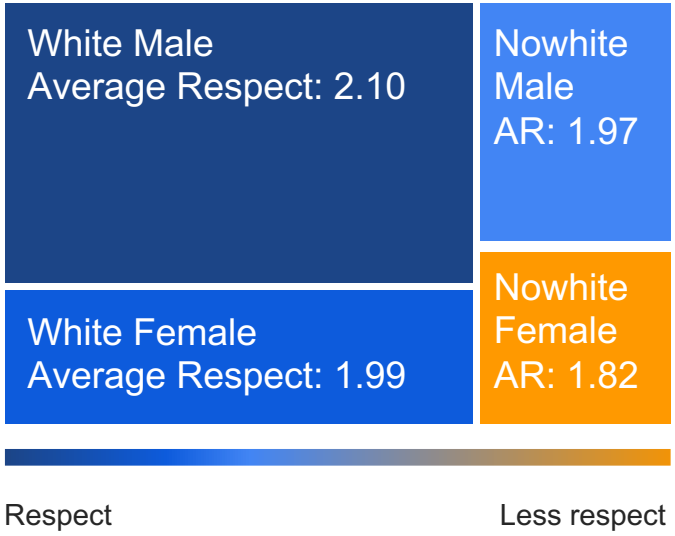
- Objective -

To hire for the position of **senior sales associate** among 54 applicants, based on the existing employees' data with several performance indicators.



- Current Dilemma -

There is a **significant disparity of gender and race** in the senior positions among all the stores.



Best Model

Data Cleaning | Exclude **Age, Gender, Marital Status, and Race** due to legislation
Keep rows **with j-level ≥ 3** only, for seniority of the target position

Model Training | Use 80% of employee data for training, 20% for validation

Model Selection | Choose the Highest R-Square (how well the model explains observed data)

Model 1

13 explanatory var.

R-Square

0.62 for Training

0.60 for Validation

- Build a model to first **predict behaviour rating** in the Applicants data
- Use the predicted values of behaviour rating as a explanatory variable when building the final model for predicting **business rating**
- Feature selection using p-values and RandomForrest

Chosen for its highest R-Square and its **inclusion of both behaviour and business performance**

$$BusinessRating = -1.16 + 0.04X_{emotionstabiliy} - 0.01X_{openness} - 0.01X_{worklife} + 0.04X_{agreeable} + 0.06X_{culturefit} + 0.05X_{congnitiveability} + 0.13X_{behavior,ating} + 0.02X_{tenure} + 0.12X_{persuasionskills} + 0.14X_{structuredinterview} + 0.38X_{salesskills} + 0.07X_{extraversion} + 0.23X_{conscientiouity} + 0.05X_{postsecondarydegree}$$ [a]

[a] The effectiveness of Conscientiousness, Extraversion and Cognitive Ability in the equation are affirmed by Kung et al. (2013).

Alternative Models [a]

Model 2

12 explanatory var.

R-Square

0.60 for Training

0.58 for Validation

Using **business rating** as the only dependent variable

$$behavior\_rating = -1.30 + 0.02X_{education} + 0.02X_{tenure} + 0.13X_{persuasion\_skills} + 0.03X_{work\_life} + 0.06X_{culture\_fit} + 0.08X_{extraversion} + 0.25X_{conscientiousness} + 0.06X_{emotion\_stability} + 0.05X_{agreeable} + 0.04X_{cognitive\_ability} + 0.15X_{structured\_interview} + 0.39X_{sales\_skills}$$

Model 3

4 explanatory var.

R-Square

0.55 for Training

0.52 for Validation

Using **business rating** as the only dependent variable

$$business\_rating = -0.36 + 0.16X_{structured\_interview} + 0.38X_{sales\_skills} + 0.27X_{conscientiousness} + 0.13X_{persuasion\_skills}$$

a. Similar assumptions applied here.

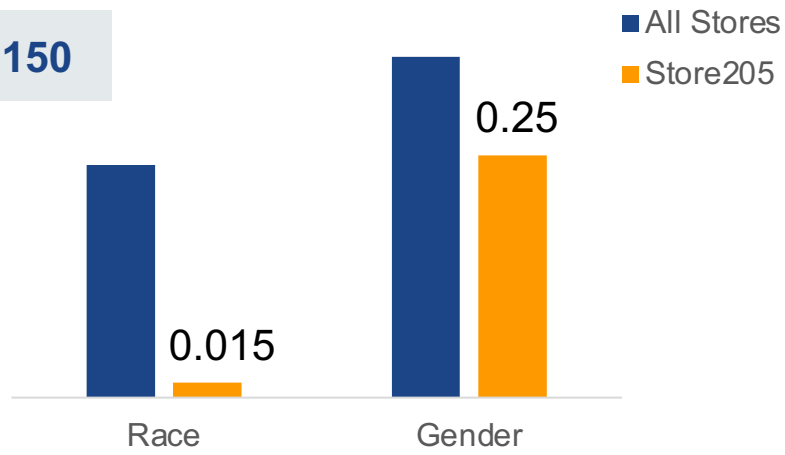
Candidate Selection – Top 3 candidates

Top 3 candidates selected by Model 1

ID	Gender	Race	Business Rating (M1)	Select by Model
150	Male	White	2.93	1, 2, 3
103	Female	White	2.88	1, 2, 3
82	**Female**	**Not White**	**2.59**	**1, 2**

Adjusted Top 3 Candidates: No.82 > No.103 > No.150

Given Store 205's more severe disparity in gender and race than stores level, it's reasonable to **prioritize candidate 82**, as it's proved by Hyun, Park, and Tian (2019) that teams with equal gender tend to perform better.



Candidate Selection - Justification for difference

Top Candidates selected by different models

ID	Select by Model
150	1, 2, 3
103	1, 2, 3
82	**1, 2**

Existence of different candidates:
Three models give the same top 2 candidates.
For the third position, model 3 choses **candidate 5** while the other two chose **candidate 82**.

ID	Select by Model
5	**3**

Justification for difference:
Candidate 5 performs much better than candidate 82 in terms of **structured interview performance**.

Recruitment - Cluster Analysis (Who to hire?)

Why is it needed?

Segmenting stores into 6 clusters, where stores within each cluster would contain similar characteristics, can help the employer conduct a **customized** and **cost-friendly** analysis for 591 stores

How to form clusters?

2 data sources to gather features for every store

- **Stores dataset** | time-series data, avg. numbers of the features for every store is used to provide an overview on store level
- **Employee dataset** | aggregate employee info based on store_id that represents which store the employee works in

How is the model evaluated?

Elbow method to find out the optimal number of clusters that would required for segmentation

How will these clusters be used?

Assign a characteristics description to every cluster, according to which we can recommend the types of applicants that would cater to the needs of every cluster

Recruitment - Cluster Analysis (Who to hire?)

Cluster 1 & 5	Cluster 2	Cluster 3 & 4	Cluster 6
221 stores (37.38%)	88 stores (14.89%)	230 stores (38.91%)	89 stores (8.79%)

Recruitment - Cluster Analysis (Who to hire?)

Cluster 1 & 5	Cluster 2	Cluster 3 & 4	Cluster 6
221 stores (37.38%)	88 stores (14.89%)	230 stores (38.91%)	89 stores (8.79%)

Characteristics Description

- Lower business ratings and sales skills
- Higher layoff and high hiring rate (fail to hire the best candidates)

Recommendations

1. Focus on recommending applicants with strong business ratings and sales skills
2. Gender/race disparity does not have to prioritized, as the cultural fit is high
3. Currently these stores are present in a low competitive market, and maybe after hiring the desired candidates with strong sales skills, we can place them in a high competitive market

Recruitment - Cluster Analysis (Who to hire?)

Cluster 1 & 5	Cluster 2	Cluster 3 & 4	Cluster 6
221 stores (37.38%)	88 stores (14.89%)	230 stores (38.91%)	89 stores (8.79%)

Characteristics Description

- High business & behavior ratings, sales skills, WLB
- Low layoff rate and a high voluntary turnover rate (which correlates with the high WLB)
- Low % of females and whites employees
- Not in competitive mkt area

Recommendations

1. Focus on recommending applicants to address the gender/race disparity
2. Place them in high competitive market areas, as the stores in this cluster show strong productivity numbers

Recruitment - Cluster Analysis (Who to hire?)

Cluster 1 & 5	Cluster 2	Cluster 3 & 4	Cluster 6
221 stores (37.38%)	88 stores (14.89%)	230 stores (38.91%)	89 stores (8.79%)

Characteristics Description

- Most ideal candidates, as they all seem to have high business and behavior ratings
- Lower than average layoffs and a good hiring rate, and are present in competitive markets area

Recommendations

1. High voluntary turnover rate, so we would recommend candidates who are least likely to turnover (favoring strong retention)
2. Address race disparity, as there seems to be a high number of whites in these clusters

Recruitment - Cluster Analysis (Who to hire?)

Cluster 1 & 5	Cluster 2	Cluster 3 & 4	Cluster 6
221 stores (37.38%)	88 stores (14.89%)	230 stores (38.91%)	89 stores (8.79%)

Characteristics Description

- Low business and behavior ratings, sales skills and work-life balance
- Low female employee ratio
- Low attrition rate (voluntary turnover and layoffs)

Recommendations

1. Focus on recommending the top level candidates to the stores in these clusters, as it contains the worst performing stores
2. Prioritize females as the ratio is poor

Turnover % - How to keep employees?

Avg Turnover
2.73 / 5

Attributes	Coeff.
Salary	-0.16
Work life balance	-0.17
Personality	
- Agreeable	-0.20
- Commitment	-0.17
- Cognitive Ability	-0.36

Turnover % - How to keep employees?

Avg Turnover

2.73 / 5

Recommendations

- Performance Evaluation

Turnover % - How to keep employees?

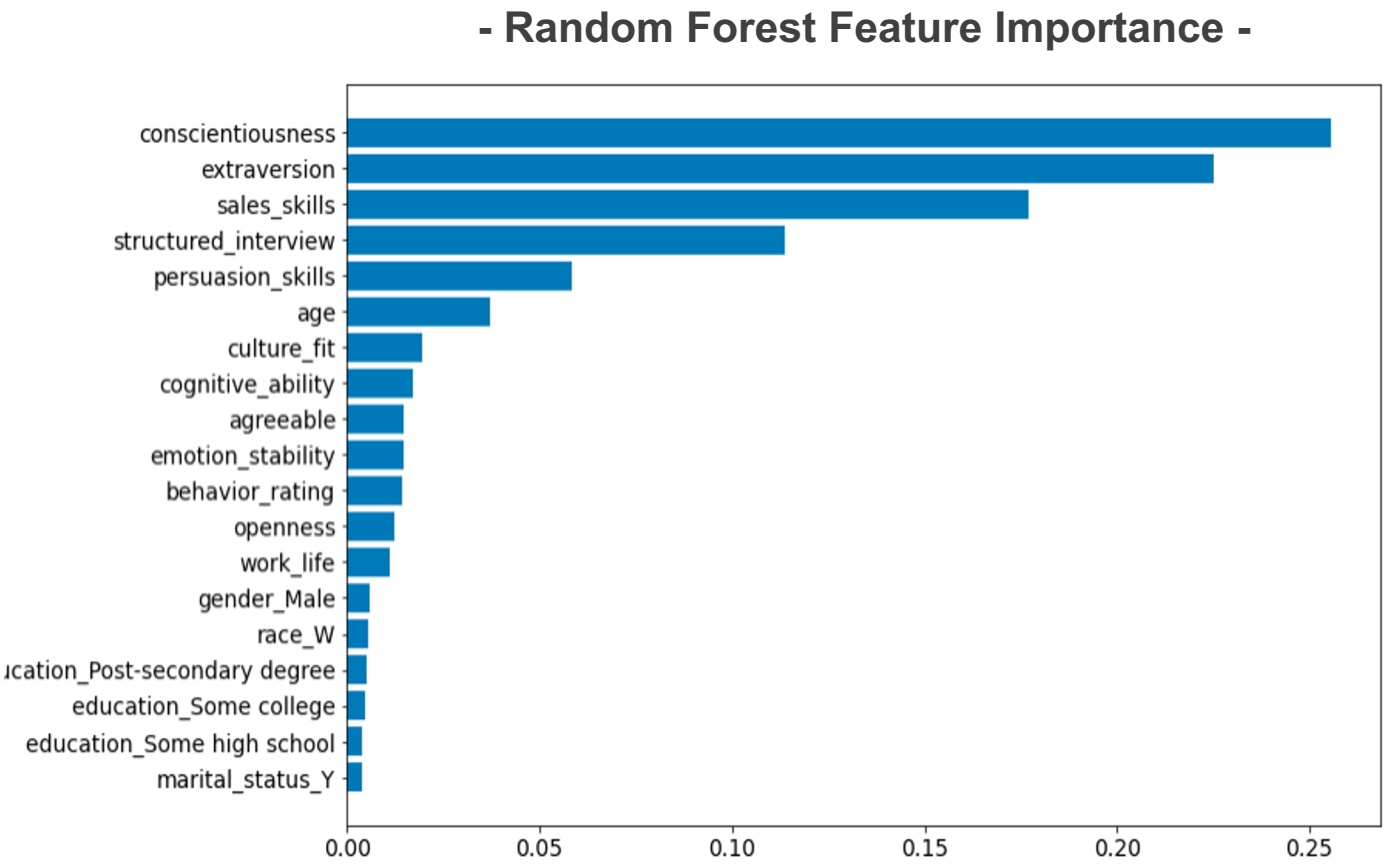
Avg Turnover
2.73 / 5

Recommendations

- Performance Evaluation
- Workshop & Events

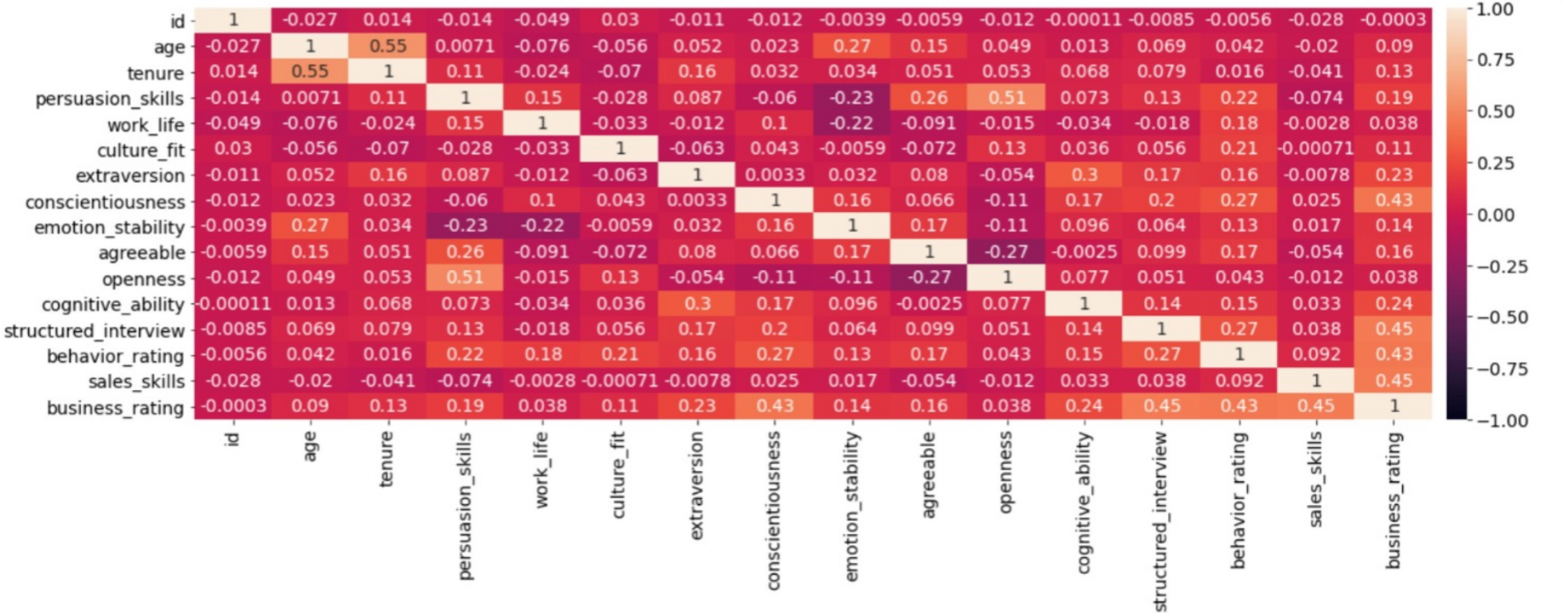
Attributes	Coeff.
Salary	-0.16
Work life balance	-0.17
Personality	
- Agreeable	-0.20
- Commitment	-0.17
- Cognitive Ability	-0.36

Appendix


- Random Forest Feature Importance -
conscientiousness
extraversion
sales_skills
structured_interview
persuasion_skills
age
culture_fit
cognitive_ability
agreeable
emotion_stability
behavior_rating
openness
work_life
gender_Male
race_W
ıcation_Post-secondary degree
education_Some college
education_Some high school
marital_status_Y
0.00
0.05
0.10
0.15
0.20
0.25

Appendix - Correlations

- Heatmap for Correlation -

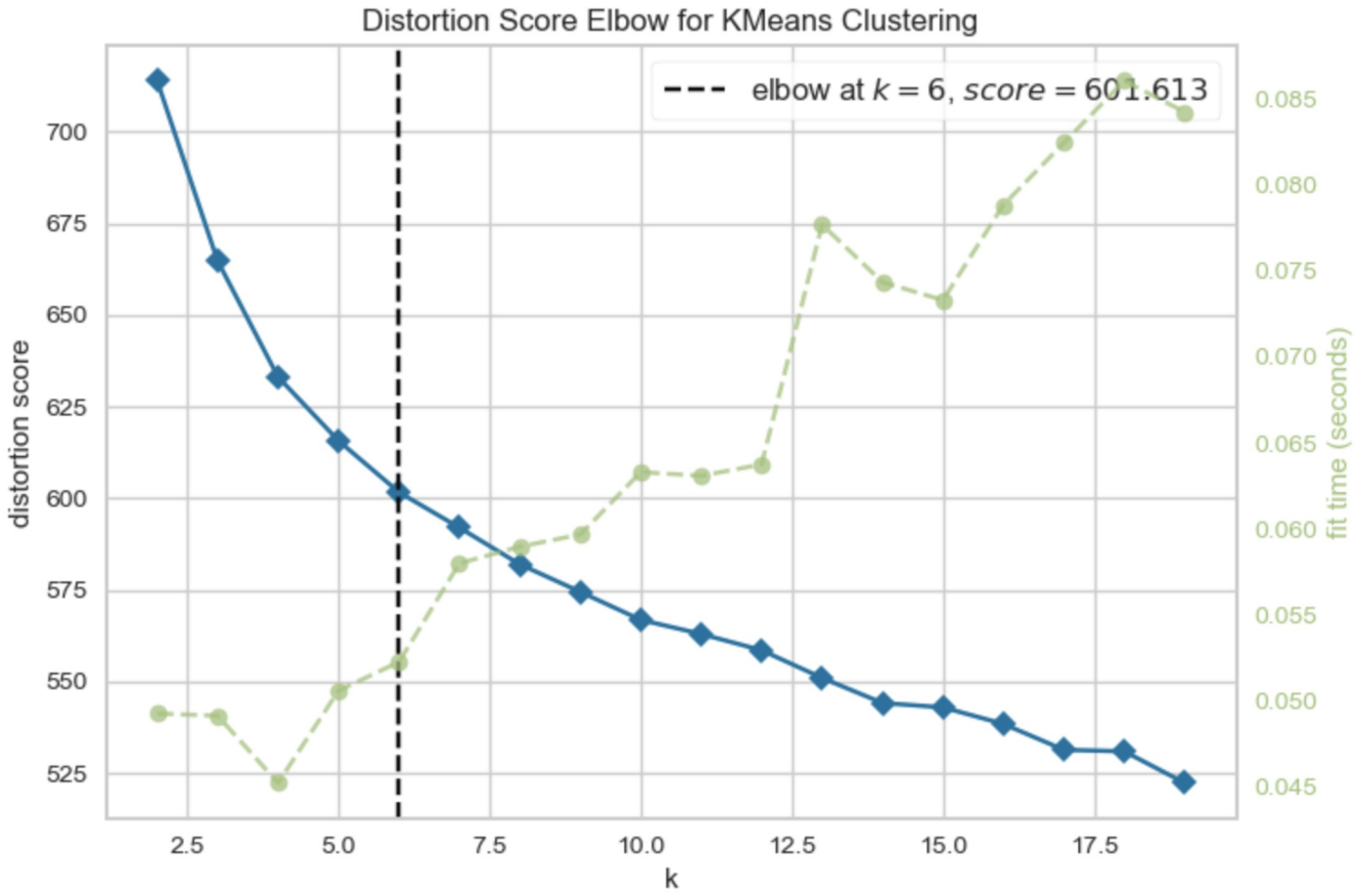


Appendix - Predicting values of three models

id	Model 1		Model 2	Model 3
	behavior	business	business	business
1	1.6334	2.0408	1.9962	2.0397
2	1.2655	1.9623	1.9266	1.8872
3	0.7514	1.6138	1.6397	1.9087
4	0.9007	1.1126	1.1432	1.2126
5	1.8840	2.5775	2.5825	2.4339
6	1.4093	1.7993	1.7885	1.7403
9	1.1790	1.7042	1.7075	1.7947
11	1.2625	1.8420	1.8521	1.9278
12	1.1907	1.8324	1.8548	1.7756
13	1.2612	1.9448	2.0446	2.1890
14	1.1733	1.2298	1.2147	1.0987
16	0.8631	1.0599	0.9655	1.0845
17	0.6110	1.0954	1.1487	1.2126
19	0.8790	1.0383	1.0437	1.0654
20	1.4222	2.1794	2.1856	2.3934
21	0.8951	1.1340	1.1510	1.4576
23	1.0031	1.1992	1.1724	1.0987
24	1.0714	0.8301	0.8372	0.7068
25	1.4298	1.7050	1.7199	1.7591
26	1.6332	1.9551	1.9774	1.9066
27	1.1215	1.6743	1.6887	1.7947
29	1.3289	1.9542	1.9736	2.1533
30	0.6384	0.7662	0.8035	0.9514
31	1.1956	1.9827	2.0393	2.0203
32	1.3575	2.2275	2.2011	2.1890
34	1.7315	2.0525	2.0195	2.0965
35	1.2741	2.0448	2.0380	2.0559

id	Model 1		Model 2	Model 3
	behavior	business	business	business
38	1.6396	2.0860	2.0722	1.9278
39	0.7195	1.2724	1.2910	1.6069
41	1.5949	2.0420	2.0053	2.1107
42	1.3319	1.9556	2.0040	1.8730
43	0.9149	1.0079	1.0491	1.0654
44	1.7678	1.9963	1.9571	1.6284
45	1.5315	2.0114	2.0083	2.0559
46	1.1230	1.8405	1.8682	1.9087
49	0.8728	1.5578	1.5545	1.6641
50	0.6912	1.0615	1.0730	1.2126
51	2.0314	1.9111	1.9537	1.7591
52	1.4850	2.2572	2.2466	2.1890
53	1.1869	1.8573	1.9066	2.2818
54	1.4158	1.9450	1.9467	2.2247
55	1.6635	1.9124	1.9914	2.0203
56	1.5236	1.7177	1.7480	1.7400
57	1.0948	1.5496	1.5613	1.7400
58	1.3670	1.1662	1.1824	1.2317
60	1.6720	2.3744	2.3590	2.3577
61	1.6423	2.0053	2.0458	1.7400
62	2.0023	1.8327	1.7916	1.4219
63	1.0062	1.0466	1.0469	1.1985
81	2.0592	2.1372	2.1229	1.6284
82	1.5769	2.5876	2.5542	2.3577
103	1.6404	2.8833	2.8564	2.8067
150	2.3134	2.9305	2.8963	2.7142
176	2.2444	1.9791	2.0712	2.0608

Appendix - Elbow method to generate cluster



Appendix - How were clusters given descriptions?

cluster_no	no.	Voluntary_turnover	Layoff	Hiring	unitmonth_female	unitmonth_white	work_life	culture_fit	business_rating	behavior_rating	sales_skills	locnearestany	respect	rmgrwhite
1	63	Low	High	Low	High	Low	High	Low	Low	High	Low	Low	Low	High
2	88	Low	Low	Low	Low	Low	High	Low	High	High	High	Low	High	Low
3	141	High	Low	High	High	High	High	High	High	High	High	High	High	High
4	89	High	Low	High	High	High	High	Low	High	High	Low	High	Low	High
5	158	Low	High	High	Low	High	High	High	Low	High	Low	Low	High	High
6	52	Low	Low	Low	High	Low	Low	High	Low	Low	Low	Low	Low	Low

- How are high and low decided : After getting clusters for all the 591 stores, we aggregated the data using avg to get the cluster level view. And as a benchmark to compare, we took avg of the features of the overall stores, and compare that with the avg of the clusters feature.

Appendix - Coefficients between 'turnover' score and other features

	term	turnover
1	persuasion_skills	-0.401843875
2	cognitive_ability	-0.363367911
3	tenure	-0.303834300
4	jlevel	-0.259779231
5	behavior_rating	-0.213660896
6	commitment	-0.204176666
7	work_life	-0.178781150
8	agreeable	-0.172408322
9	salary	-0.165304524
10	age	-0.154601770
11	perf_fair	-0.150152944
12	openness	-0.126591552

	term	turnover
13	reward_fair	-0.113117155
14	extraversion	-0.107950512
15	development	-0.096545861
16	conscientiousness	-0.089717849
17	culture_fit	-0.089035853
18	respect	-0.086325638
19	business_rating	-0.063236090
20	goal	-0.054623996
21	structured_interview	-0.026054150
22	competition	-0.018914088
23	store_id	-0.005134003
24	location_code	-0.001735437

Appendix - Reference

Hoogendoorn, S., Oosterbeek, H., & Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. Management Science, 59(7), 1514-1528.

Kung, M. C., Glatzhofer, P. E., Lawrence, A. D., & Jarrett, S. (2013). Investigating Predictors of Sales Performance: A Longitudinal Study. ResearchGate. Retrieved December 13, 2022, from https://www.researchgate.net/publication/320188393_Investigating_Predictors_of_Sales_Performance_A_Longitudinal_Study