BA 550
Group 2

CANADARETAIL EMPLOYEE SELECTION MODEL

PRESENTATION

Emery (Yuju) Chang
Tanmay Jain
Junda Liu
Nurzhan Uapov
Cynthia Wang

OUR PEOPLE

Attract, retain and develop diverse talent

CanadaRetail

TABLE OF CONTENT

1 Objective

2 Predictor Selection

3 Model Comparison

4 Final Model

5 Candidate Evaluation

6 Other Considerations

7 Appendix

Objective

To develop a predictive model that will aid the selection process for the top candidates for the Senior Sales Associate position in our largest flagship store:

Position: Senior Sales Associate

Store: StoreID 205

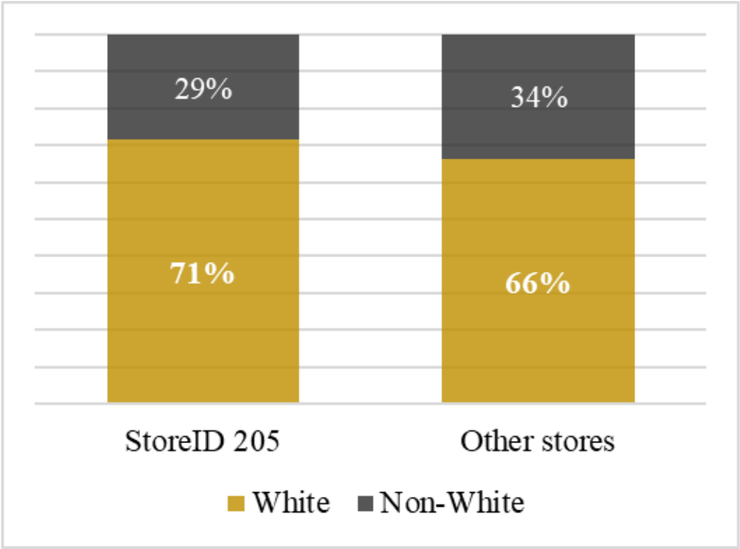
Location: Vancouver, BC

Objective

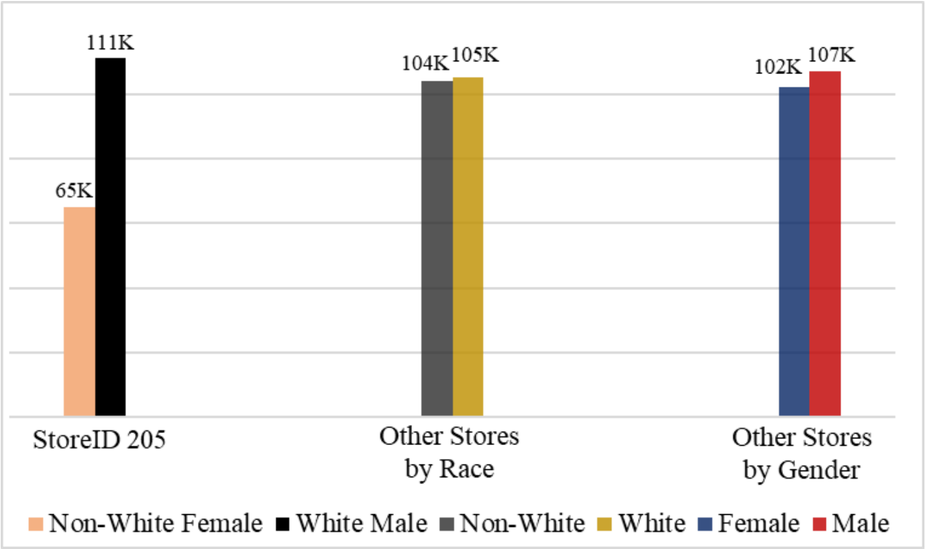
Addressing gender and racial disparities

Store 205 has one of the highest levels of gender and racial inequalities within the organization:

% of Employees by Gender



% of Employees by Race



Jlevel 4 Salary (CAD) by Race & Gender

Predictor Selection

1. Feasibility

Selected only variables available in both Employee and Applicant datasets

2. Legality

Excluded variables that can result in discrimination:

- age
- gender
- race
- marital status

3. Validity

Removed variables with high VIF to mitigate multicollinearity:

- openness ***(dropped)***
- persuasion skills ***(kept)***

4. Reliability

Psychological climate as a predictor of sales performance*:

- extraversion
- conscientious-ness
- agreeable
- emotion stability

*Refer to Appendix: References

Model Selection

Model 1

- Linear Regression
- Predicting Business rating
- All data from Employee dataset*

Model 2

- Linear Regression
- Predicting Business rating
- Filtered on Job Level = 3 or 4*

Model 3

- Linear Regression
- Predicting Business rating
- All data from Employee dataset*
- Exclude behavioral predictors to isolate performance measure

*Refer to Appendix: Model 1

Final Model: Model 1

- **Model Fitting**

Model 1 has the highest r-squared out of the 3.

- **v.s. Model 2: Number of Observations**

Model 2 restricts our data to specific job levels and decreases our data to 20% of the original dataset, whereas Model 1 is more generalized utilizes all data.

- **v.s. Model 3: Behavior Predictors**

We want to include behavioral/phsycological metrics as they are good indicators of job responsibilities, i.e. customer service.

Potential Candidates

Predicted Business Rating	**2.86**	**2.85**	**2.60**	**2.55**	**2.34**
Applicant ID	**103**	**5** ✓	**150**	**82** ✓	**60** ✓
Age	32	34	40	31	26
Gender	Female	Female	Male	Female	Male
Race	White	Non-White	White	Non-White	Non-White
Marital Status	Married	Married	Married	Married	Married
Education	Post-secondary	Post-secondary	College Diploma	High School Diploma	High School Diploma
Years of Experience	11.6	6.8	12.5	8	4.5

Selection of Final 3:

- The median **sales experience** for the employee data is 6.5 years. Therefore, applicant ID 150 & 103 seem to be **overqualified**.
- Considering the **racial** and **gender disparity** in store 205, the firm should hire from other groups to create **equity**.

Potential Candidates Skills

Applicant ID 5

Applicant ID 82

Applicant ID 60

To better aid CanadaRetail's final selection among the top 3 candidates for the Senior Sales Associate position, we visualized the hard and soft skills of each candidate via spider charts.

OTHER CONSIDERATIONS

MODEL LIMITATIONS

- When creating our model, we did not include behavior rating as part of the predicting metric, which may decrease the weight of soft skills when evaluating applicants.
- The model cannot predict the likelihood of retention rate given limited predictors available in the applicant dataset.

POTENTIAL SOLUTIONS

- Update the business rating to incorporate behavior rating based on the level desired from the People Operations Team.
-
  - Create supportive and inclusive environments
  - Provide proper training and create career development plan
  - Monitor absent days and timely communicate to offer help

THANK YOU

PRESENTATION

Appendix

Dictionary of terminologies in regression results

- **Adjusted R-Square:** a measurement of the goodness of fit of our model, higher value means better goodness of fit.
- **Coe f:** coefficient of each predictor, a measurement of the influence of a specific predictor on the business rating.
- **Beta:** standardized coefficient, a measurement of the ranking of influence of each predictor, higher value means higher weight in the influence of business rating.
- **Std.Err:** the standard errors for each variable.
- **t:** value of t statistics, used to test the significance of each predictor.
- **P>|t|:** the p value of each variable, a measurement of degree of significance for each. predictor, the less p value is, the more significant the variable is.
- **VIF**: variance inflation factor for each candidate variable, a measurement of multicolinearity of our model. Higher VIF means more serious multicolinearity.
- **Pairwise-Correlation:** A correlation matrix among all the candidates variables. Higher value means higher correlation between the variables.
- **Source of regression results:** output from STATA, a well-known statistical software for data science

Appendix

References

Riley Dugan, Maria Rouziou, Willy Bolander, The case for hiring neurotic salespeople: A longitudinal growth modeling analysis, Journal of Business Research, Volume 116, 2020, Pages 123-136, ISSN 0148-2963, https://doi.org/10.1016/j.jbusres.2020.04.021.

Martin, C. A., & Bush, A. J. (2020). Do Gender and Length of Employment Impact the Relationship Between Psychological Climate and Subjective Salesperson Performance? Journal of Management Policy & Practice, 21(5), 28–38. https://doi.org/10.33423/jmpp.v21i5.3872

Nye, C. D., Ma, J., & Wee, S. (2022). Cognitive Ability and Job Performance: Meta-analytic Evidence for the Validity of Narrow Cognitive Abilities. Journal of Business & Psychology, 37(6), 1119–1139. https://doi.org/10.1007/s10869-022-09796-1

Appendix

Pairwise Correlation & VIF

	persuasion_ski	extraversion	structured_interview	tenure	agreeable	cognitive_ability	conscientiousness	emotion_stability	work_life	sales_skills
extraversion	0.13									
structured_interview	0.16	0.16								
tenure	0.2	0.14	0.09							
agreeable	0.26	0.06	0.09	0.1						
cognitive_ability	0.06	0.3	0.1	0.06	0.02					
conscientiousness	0.05	0.01	0.19	0.14	0.13	0.11				
emotion_stability	-0.16	0.03	0.04	0.07	0.17	0.08	0.12			
work_life	0.14	0	0.02	-0.01	-0.07	-0.01	0.06	-0.21		
sales_skills	0	0.01	0.06	0	-0.01	0	0.01	0	0.01	
culture_fit	-0.07	-0.08	0.03	-0.06	-0.07	0.06	0	-0.02	-0.04	0.01

Variable	VIF	1/VIF
absent	5.21	0.192032
culture_fit	3.94	0.253712
emotion_st~y	3.04	0.328969
salary	2.94	0.340086
jlevel	2.88	0.347431
persuasion~s	2.77	0.361135
marital_st~l	2.71	0.368455
turnover	2.49	0.401272
agreeable	2.36	0.424421
openness	2.26	0.442869
training_s~l	2.07	0.482266
tenure	1.95	0.511812
cognitive_~y	1.95	0.511947
age	1.84	0.542489
genderl	1.61	0.622645
conscienti~s	1.42	0.705781
work_life	1.36	0.734934
reward_fair	1.31	0.761612
extraversion	1.31	0.762528
development	1.28	0.783543
commitment	1.27	0.787846
racel	1.20	0.831312
competition	1.17	0.854164
respect	1.16	0.864572
structured~w	1.11	0.897724
perf_feedb~k	1.11	0.898474
perf_fair	1.11	0.904210
sales_skills	1.02	0.977502
goal	1.02	0.982012
Mean VIF	1.96	

Appendix

Model 1

Source	SS	df	MS
Model	2289.41614	11	208.12874
Residual	1489.39972	6,988	.213136766
Total	3778.81586	6,999	.539907966

Number of obs	=	7,000
F(11, 6988)	=	976.50
Prob > F	=	0.0000
R-squared	=	0.6059
Adj R-squared	=	0.6052
Root MSE	=	.46167

business_rating	Coef.	Std. Err.	t	P>|t|	Beta
tenure	.0173238	.0028889	6.00	0.000	.047304
education_rank	.0151431	.0056662	2.67	0.008	.0212099
persuasion_skills	.145418	.0060063	24.21	0.000	.199758
culture_fit	.0500737	.0048953	10.23	0.000	.0781351
extraversion	.0834005	.0052285	15.95	0.000	.1304668
emotion_stability	.0371511	.0056051	6.63	0.000	.0524491
conscientiousness	.238352	.0062517	38.13	0.000	.2994884
agreeable	.051908	.0059424	8.74	0.000	.0702776
cognitive_ability	.0323647	.0056963	5.68	0.000	.0476256
structured_interview	.1558345	.0041223	37.80	0.000	.2967746
sales_skills	.4183334	.0068343	61.21	0.000	.4608169
_cons	-1.247497	.0411151	-30.34	0.000	.

Validation:

- **RMSE: 0.47**
- **MAE: 0.35**

Explanation:

- **The top 3 predictors for employee performance in Model 1 are sales skills, conscientiousness and structured interview according to beta**
- **Goodness of fit: our model explains 60.52% of the changes in business rating**

* To create our model, we've split our employee data into 80% training and 20% validation.

Appendix

Model 2

Source	SS	df	MS
Model	502.231526	11	45.6574115
Residual	344.600786	1,783	.193270211
Total	846.832312	1,794	.472035848

Number of obs	=	1,795
F(11, 1783)	=	236.24
Prob > F	=	0.0000
R-squared	=	0.5931
Adj R-squared	=	0.5906
Root MSE	=	.43963

business_rating	Coef.	Std. Err.	t	P>|t|	Beta
tenure	.0245661	.0055282	4.44	0.000	.0692359
education_rank	.02379	.0107281	2.22	0.027	.0350412
persuasion_skills	.1248644	.0116162	10.75	0.000	.179529
culture_fit	.063492	.0091734	6.92	0.000	.1059142
extraversion	.0849328	.0098949	8.58	0.000	.1411915
emotion_stability	.0543764	.010445	5.21	0.000	.0848849
conscientiousness	.2549105	.012084	21.09	0.000	.3359787
agreeable	.0483134	.0109102	4.43	0.000	.0721906
cognitive_ability	.0383743	.0106478	3.60	0.000	.0611509
structured_interview	.1419334	.0080211	17.70	0.000	.280548
sales_skills	.391923	.0132686	29.54	0.000	.4501325
_cons	-1.298027	.0873822	-14.85	0.000	.

Validation:

- **RMSE: 0.45**
- **MAE: 0.33**

Explanation:

- **The top 3 predictors for employee performance in Model 1 are sales skills, conscientiousness and structured interview according to beta**
- **Goodness of fit: our model explains 59.06% of the changes in business rating**

Appendix

Model 3

Source	SS	df	MS
Model	1850.33768	6	308.389613
Residual	1928.47818	6,993	.275772656
Total	3778.81586	6,999	.539907966

Number of obs	=	7,000
F(6, 6993)	=	1118.27
Prob > F	=	0.0000
R-squared	=	0.4897
Adj R-squared	=	0.4892
Root MSE	=	.52514

business_rating	Coef.	Std. Err.	t	P>|t|	Beta
tenure	.0368841	.0032271	11.43	0.000	.1007151
education_rank	.0167944	.0061667	2.72	0.006	.0235228
persuasion_skills	.1356464	.0066071	20.53	0.000	.1863349
agreeable	.0810377	.0065608	12.35	0.000	.109716
structured_interview	.1977071	.0045633	43.33	0.000	.3765176
sales_skills	.4210104	.0077734	54.16	0.000	.4637658
_cons	-.33715	.0362486	-9.30	0.000	.

Validation:

- **RMSE: 0.55**
- **MAE: 0.42**

Explanation:

- **The top 3 predictors for employee performance in Model 1 are sales skills, conscientiousness and persuasion skill according to beta**
- **Goodness of fit: our model explains 48.92% of the changes in business rating**

Appendix

Ordinal Logistic Model

```
Iteration 0:   log likelihood = -7445.0594
Iteration 1:   log likelihood = -5140.8919
Iteration 2:   log likelihood = -5031.7897
Iteration 3:   log likelihood = -5030.2492
Iteration 4:   log likelihood = -5030.2478
Iteration 5:   log likelihood = -5030.2478

Ordered logistic regression                     Number of obs     =      7,000
                                                LR chi2(6)        =    4829.62
                                                Prob > chi2       =     0.0000
Log likelihood = -5030.2478                     Pseudo R2         =     0.3244
```

business_rating	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
tenure	.1524622	.0135742	11.23	0.000	.1258572	.1790672
education_rank	.0658656	.0257976	2.55	0.011	.0153033	.1164279
persuasion_skills	.5674179	.0286596	19.80	0.000	.5112461	.6235897
agreeable	.3410809	.0279032	12.22	0.000	.2863916	.3957702
structured_interview	.8589684	.0223683	38.40	0.000	.8151275	.9028094
sales_skills	1.746064	.0401221	43.52	0.000	1.667426	1.824702
/cut1	8.146065	.1949999			7.763872	8.528257
/cut2	11.51882	.2242376			11.07932	11.95832

- We build this model to further test the robustness of the resuts of our main models (model 1,2 and 3)
- The response variable in this model is the logit of the probability of the each rank for business rating.
- We backtransformed the predicted value of response variable in the applicants dataset, and the recommended top 3 applicants of this model is consistent with our previous models

Appendix

Turnover Model-Overall Model

Source	SS	df	MS
Model	3271.23644	11	297.385131
Residual	6550.06627	6,988	.937330605
Total	9821.30271	6,999	1.40324371

Number of obs	=	7,000
F(11, 6988)	=	317.27
Prob > F	=	0.0000
R-squared	=	0.3331
Adj R-squared	=	0.3320
Root MSE	=	.96816

turnover	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
tenure	-.1302682	.0062961	-20.69	0.000	-.1426105	-.117926
work_life	-.2010858	.0114604	-17.55	0.000	-.2235516	-.1786199
culture_fit	-.1344639	.010421	-12.90	0.000	-.1548922	-.1140356
extraversion	.0818232	.0108527	7.54	0.000	.0605486	.1030977
ognitive_ability	-.496045	.0122184	-40.60	0.000	-.5199967	-.4720934
perf_fair	-.0863052	.013109	-6.58	0.000	-.1120029	-.0606075
reward_fair	-.1300747	.0107211	-12.13	0.000	-.1510913	-.1090581
respect	-.1492582	.0128829	-11.59	0.000	-.1745126	-.1240039
development	-.2891326	.0162734	-17.77	0.000	-.3210334	-.2572318
goal	-.079503	.0219578	-3.62	0.000	-.122547	-.036459
salary	-5.31e-06	5.42e-07	-9.79	0.000	-6.37e-06	-4.25e-06
_cons	7.07547	.0866815	81.63	0.000	6.905548	7.245392

- Predictors like work life, culture fit, salary, reward fairness, performance fairness and higher employee satisfaction - respect, development and goal - are shown to reduce turnover rate of employees.
- Employees with better cognitive ability and less extraverted personality have less turnover rate compared with other employees

Appendix

Turnover Model-Gender difference in turnover

Source	SS	df	MS
Model	3461.04554	12	288.420461
Residual	6360.25718	6,987	.910298723
Total	9821.30271	6,999	1.40324371

Number of obs	=	7,000
F(12, 6987)	=	316.84
Prob > F	=	0.0000
R-squared	=	0.3524
Adj R-squared	=	0.3513
Root MSE	=	.9541

turnover	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
tenure	-.1380834	.0062282	-22.17	0.000	-.1502926	-.1258743
work_life	-.1452116	.0119384	-12.16	0.000	-.1686145	-.1218087
culture_fit	-.1646327	.01048	-15.71	0.000	-.1851767	-.1440887
extraversion	.0618229	.0107844	5.73	0.000	.0406823	.0829636
cognitive_ability	-.4886368	.0120518	-40.54	0.000	-.512262	-.4650116
perf_fair	-.0769026	.012935	-5.95	0.000	-.1022592	-.0515461
reward_fair	-.1186373	.010595	-11.20	0.000	-.1394068	-.0978679
respect	-.1431526	.0127028	-11.27	0.000	-.1680539	-.1182513
development	-.2767647	.0160599	-17.23	0.000	-.3082469	-.2452824
goal	-.0812746	.0216392	-3.76	0.000	-.123694	-.0388551
salary	-4.46e-06	5.38e-07	-8.28	0.000	-5.51e-06	-3.40e-06
l.genderl	-.3708271	.0256806	-14.44	0.000	-.4211689	-.3204854
_cons	7.25119	.0862848	84.04	0.000	7.082045	7.420334

Male employees have lower turnover ratio than Female employees

Appendix

Turnover Model-Race difference in turnover

Source	SS	df	MS
Model	3532.79292	12	294.39941
Residual	6288.5098	6,987	.900030027
Total	9821.30271	6,999	1.40324371

Number of obs	=	7,000
F(12, 6987)	=	327.10
Prob > F	=	0.0000
R-squared	=	0.3597
Adj R-squared	=	0.3586
Root MSE	=	.9487

turnover	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
tenure	-.1257083	.0061753	-20.36	0.000	-.1378138	-.1136028
work_life	-.1923496	.0112417	-17.11	0.000	-.2143868	-.1703124
culture_fit	-.1364649	.0102122	-13.36	0.000	-.156484	-.1164459
extraversion	.0652185	.0106791	6.11	0.000	.0442842	.0861527
cognitive_ability	-.4848473	.0119908	-40.44	0.000	-.5083529	-.4613418
perf_fair	-.0701861	.0128803	-5.45	0.000	-.0954354	-.0449369
reward_fair	-.1086307	.0105806	-10.27	0.000	-.1293719	-.0878894
respect	-.1308492	.01267	-10.33	0.000	-.1556863	-.1060121
development	-.2737658	.0159718	-17.14	0.000	-.3050754	-.2424563
goal	-.0804604	.0215166	-3.74	0.000	-.1226394	-.0382814
salary	-5.19e-06	5.32e-07	-9.77	0.000	-6.23e-06	-4.15e-06
l.race1	-.4175721	.024495	-17.05	0.000	-.4655897	-.3695545
_cons	7.185774	.0851853	84.35	0.000	7.018785	7.352763

White employees have lower turnover ratio than Non-White employees