INSEAD The Business School for the World®

Pre-Release Version

Case Study

Integration Planning at SFB (B)

09/2021-6684

This case study was written by Anton Ovchinnikov, Visiting Professor of Operations, Technology and Decision Sciences and a Distinguished Professor of Management Analytics and Scotiabank Scholar of Customer Analytics at Smith School of Business at Queen's University in Canada, Georgina Hall, Assistant Professor of Decision Sciences and Piyush Gulati, PhD student. It is intended to be used as a basis for class discussion rather than to illustrate either effective or ineffective handling of an administrative situation.

To access INSEAD teaching materials, go to https://publishing.insead.edu/

"We first have to predict the probability of each employee leaving the firm and taking up the RCC option of," said John, as his initial reaction when Kusha explained the need to come up with a headcount reduction plan for the Lyon office. John had already reviewed the employee census files (Exhibit B1) that Kusha had shared with him.

"Yes, that makes sense. But how do we do that?" she asked, trying to work out how this fit with the analysis toolkit she had used as a consultant.

John Williams, the strategy analyst

Like Kusha, John was a recent addition to the BAP corporate strategy team. After completing a Masters in Management with a specialization in analytics two years earlier, he first worked in the Strategy and Artificial Intelligence (AI) team of a consumer products company, which focused on enabling strategy decisions with the extensive use of data to predict consumer behaviour.

At BAP, Nikitha planned to inculcate data-driven decision-making within her team and was hoping to recruit someone who could bring an understanding of predictive analytics to organizational decision making. Luckily for her, John was keen on the challenge and agreed to take up the role after being pursued by a recruiter for several months.

"It is great that we have data from a previous acquisition," he responded. "We could use that as a 'training dataset' to develop a supervised learning model – a model that predicts the probability of leaving based on the input (features of the employees) and output (decisions made by these employees).".

"But how can we be sure that the decision made by employees in Lyon will be similar to the ones made in the previous transaction?" asked Kusha, trying to understand the process of developing predictions.

"Great question! The short answer is – we cannot. The perfect model would be one based on similar past decisions by the same employees or a randomly selected subgroup in the same situation. That's not possible – hence we rely on the next best option. In my opinion, data from a previous transaction that was similar in deal value, similar in the size of shut-down, and similar in the distance of movement from the shutdown office – offers a wonderful next best alternative."

"That makes sense. I have another question. As per French regulations, we cannot target specific employees with an offer to leave – this can only be done at the employee-category level. In this case, at what level do we build our prediction model - employee or employee category?"

"I read through the RCC overview from Maurice that you forwarded. I think the best way would be to build a prediction model at the employee level that calculates the probability of leaving. Subsequently, we select employee categories that have, on average, the highest probability of leaving, and satisfy other requirements such as cost and headcount targets – perhaps

using an optimization model. This way our final output will be at the *employee-category level* and not *employee-level.*"

"Okay, sounds good."

"Now I need help from you," said John. "Could you please help me interpret each of the features in the census files? While column names are mostly intuitive, I want to make sure I don't make any mistake in how I use these features in the prediction model."

"Absolutely. I did go through the exercise of reviewing data from the Lyon office and previous transaction census files. I've compiled a data dictionary with the interpretation of each column and a few additional comments. Let me share that with you." (Exhibit B2)

"Thanks. Once you share, I will start the process of developing the prediction model – first data pre-processing, then tuning and testing different machine learning models, and finally developing predictions for the Lyon employees."

"Great – thanks, John. I feel so much better about this task now."

Exhibit B-1

Lyon Employee Census Snapshot

	A	B	C	D	E	F	G	…	AB	AC	AD	AE	AF	AG	AH
1	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField		TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
2		Travel	866	Sales		3	Medic		6	4	3	5	4	1	3
3	53	Travel	1084	Resea	13	2	Medic		5	3	3	4	2	1	3
4	24	Travel	240	Huma	22	1	Huma		1	2	3	1	0	0	0
5	45	Travel	1339	Resea	7	3	Life Sc		25	2	3	1	0	0	0
6	36	Travel	1396	Research & D		2	Life Sc		16	3	4	13	11	3	7
7	34	Travel	204	Sales	14	3	Techn		8	3	3	8	2	0	6
8		Travel	144	Resea	22	3	Life Sc		6	2	3	5	4	4	3
9	39	Travel	1431	Resea	1	4	Medic		7	1	3	3	2	1	2
10	45	Non-T	1052	Sales	6	3	Medic		23	2	3	19	7	12	8
11		Travel	1229	Resea	8	1	Life Sc		1	2	3	1	0	0	1
12	47	Travel	1454	Sales	2	4	Life Sc		3	3	2	3	2	1	2
13	43	Travel	531	Sales	4	4	Marke		23	3	4	21	7	15	17
14	44	Travel	625	Resea	4	3	Medic		10	2	2	5	2	2	3
15	40	Travel	1171	Resea	10	4	Life Sc		10	3	3	7	7	1	7
16	22	Travel	581	Resea	1	2	Life Sc		4	2	4	3	2	1	2
…															
428	29	Travel	442	Sales	2	2	Life Sc		10	3	2	10	7	0	9
429	36	Travel	566	Resea	18	4	Life Sc		4	2	3	1	0	0	0
430	33	Travel	213	Resea	7	3	Medic		14	3	4	13	9	3	7
431	36	Travel	506	Research & D		3	Techn		10	2	3	8	0	7	7
432	38	Travel	1495	Resea	10	3	Medic		18	4	3	1	0	0	0
433	54	Travel	376	Resea	19	4	Medic		9	4	3	5	3	1	4
434	37	Non-T	1413	Research & D		2	Techn		7	2	1	6	5	1	3
435	33	Non-Travel		Sales	8	1	Life Sc		3	2	2	2	2	2	2
436	50	Travel	410	Sales	28	3	Marke		20	3	3	3	2	2	0
437	49	Travel	470	Resea	20	4	Medic		16	2	2	15	11	5	11
438	33	Travel	430	Sales	7	3	Medic		5	2	3	4	3	0	3
439	27	Non-T	210	Sales	1	1	Marke		6	0	3	5	4	1	4
440		Travel	1157	Resea	2	4	Medic		5	4	2	3	2	2	2
441	39	Travel	1383	Huma	2	3	Life Sc		13	2	3	5	4	0	4
442		Travel	104	Resea	2	3	Life Sc		1	2	3	1	0	0	0

Source: "Credit goes to Prashant Patel (https://www.kaggle.com/patelprashant) from where the dataset has been collected.".

Exhibit B-2

Employee Census Data Dictionary

#	Column	Description	Value (Lyon employee census)
1	Age	Age of the employee in years as of 31st Dec 2018	Min: 18, Max: 60 (Not available: 55)
2	Attrition	Dummy variable, "Yes" if the employee took the offer to leave, "No" otherwise [Kusha: only available in previous transaction census]	'Yes', 'No'
3	BusinessTravel	Frequency of employee business travel	'Non-Travel', 'Travel_Rarely', 'Travel_Frequently' (Not available: 5)
4	DailyRate	Daily prevailing wage rate for the Job Role + Job Level (in Euros) [Kusha: Appears to be legacy system value, not correlated with Monthly Income]	Min: €104, Max: €1,499 (Not available: 14)
5	Department	Department in which the employee works	'Human Resources', 'Research & Development', 'Sales'
6	DistanceFromHome	Distance of the new office location from home in kms [Kusha: Note that this is from the new office, not the old office]	Min: 1 km, Max: 29 kms (Not available: 54)
7	Education	Highest level of education received by the employee	Min: 1, Max: 5
8	EducationField	Subject in which the employee completed their education	'Medical', 'Human Resources', 'Life Sciences', 'Technical Degree','Marketing', 'Other'
9	EmployeeCount	Column that verifies if the employee is available in the company ERP system	Min: 1, Max: 1
10	EmployeeNumber	A unique id assigned to each employee	Unique 2-4 digit id
11	EnvironmentSatisfaction	Self-reported satisfaction with the work environment in the Yearly Employee Survey 2018 on a 1-4 scale	Min: 1, Max: 4
12	Gender	Gender of the employee	'Male', 'Female'
13	HourlyRate	Coded hourly rate for the job role in the ERP system (in Euros) [Kusha: Appears to be legacy system value, not correlated with Monthly Income]	Min: €30, Max: €100
14	JobInvolvement	Self-reported involvement (identification) with their work in the Yearly Employee Survey 2018 on a 1-4 scale	Min: 1, Max: 4

Exhibit B-2 (Contd.)

#	Column	Description	Value (Lyon employee census)
15	JobLevel	Current job level of the employee in the company	Min: 1, Max: 5
16	JobRole	Role of the employee in the company	'Sales Executive', 'Manufacturing Director', 'Human Resources', 'Research Scientist', 'Laboratory Technician', 'Sales Representative', 'Healthcare Representative', 'Manager', 'Research Director'
17	JobSatisfaction	Self-reported satisfaction with their specific job role in the Yearly Employee Survey 2018 on a 1-4 scale	Min: 1, Max: 4
18	MaritalStatus	Marital status of the employee	'Single', 'Divorced', 'Married' (Not available: 3)
19	MonthlyIncome	Actual monthly income received by the employee in Euros	Min: €1,051, Max: €19,740
20	MonthlyRate	Coded hourly rate for the job role in the ERP system (in Euros) [Kusha: Appears to be legacy system value, not correlated with Monthly Income]	Min: €2,104, Max: €26,959
21	NumCompaniesWorked	Total number of companies that the employee has worked in his/her career	Min: 0, Max: 9
22	Over18	A system flag variable to highlight any underage employees recruited, "Y" if over 18 years of age and "N" otherwise	'Y'
23	OverTime	Dummy variable, "Yes" if the employee completed overtime hours in the last year, "No" otherwise	'Yes', 'No'
24	PercentSalaryHike	Percentage salary increase for the employee last year	Min: 11%, Max: 25%
25	PerformanceRating	Rating received by the employee in the previous annual review, on a 1-4 scale [Kusha: Strange distribution, 379 of 441 employees are rated 3 and rest are rated 4]	Min: 3, Max: 4
26	RelationshipSatisfaction	Self-reported satisfaction with their manager in the Yearly Employee Survey 2018 on a 1-4 scale	Min: 1, Max: 4

Exhibit B-2 (Contd.)

#	Column	Description	Value (Lyon employee census)
27	StandardHours	Standard number of hours that the employee is expected to work in a biweekly cycle	Min: 80, Max: 80
28	StockOptionLevel	A scale of 0-3 that indicates the stock option ownership of the employee as part of their total compensation	Min: 0, Max: 3
29	TotalWorkingYears	Total number of years that the employee has been engaged in work, across companies	Min: 0, Max: 40
30	TrainingTimesLastYear	Number of trainings attended by the employee last year	Min: 0, Max: 6
31	WorkLifeBalance	Self-reported work life balance score for their job role in the Yearly Employee Survey 2018 on a 1-4 scale	Min: 1, Max: 4
32	YearsAtCompany	Total number of years that the employee has been working for SFB	Min: 0, Max: 40
33	YearsInCurrentRole	Total number of years in the current role	Min: 0, Max: 18
34	YearsSinceLastPromotion	Total number of years at the same job level	Min: 0, Max: 15
35	YearsWithCurrManager	Total number of years working for the same reporting manager	Min: 0, Max: 17

Source: "Credit goes to Prashant Patel (https://www.kaggle.com/patelprashant) from where the dataset has been collected."