



SMM750

Digital Technologies and Value Creation

*Final coursework*

# Integration planning at SFB

***Name: Linh Nguyen***

***ID: 170037344***

***MSc Business Analytics, cohort 2022/23***

**Question 1:**

*Using Python, complete the prediction task outlined in Part B of the case. In particular, develop a model that allows you to predict the probability of employees at the Lyon facility accepting an RCC if it is offered to them. Make sure that your code prints out the predicted probability for each employee (e.g., as a list). What are the most important factors in determining whether an employee will accept an RCC?*

The list of predicted probability for each employee can be found in the folder “output”, file name “lyon\_acceptance.csv”.

Based on the logistic model’s results, with an RCC offered, the most important factors affecting the chance of an employee leaving the company are observed below.

**Table1.** Logistic regression Model results: significant negative factors

(At 0.05 significance level):

Variable	Coefficient	p-value
YearsInCurrentRole	-0.785603	3.474443e-02
EnvironmentSatisfaction	-0.517245	1.076597e-03
JobSatisfaction	-0.337484	2.806172e-02
JobInvolvement	-0.300468	2.944488e-02

The employees who had a higher total number of years in their current roles, higher environment satisfaction, higher job satisfaction and higher job involvement would have lower probabilities of leaving the company which means higher probabilities of accepting the RCC. These are the indicators that should be concentrated on to assess the possibility of staying of each employee.

**Table 2.** Logistic regression Model results: significant positive factors

(At 0.05 significance level):

Variable	Coefficient	p-value
NumCompaniesWorked	0.587660	6.226938e-04
BusinessTravel_Travel_Frequently	0.712686	9.946865e-03
MaritalStatus_Single	0.744941	1.504396e-03
OverTime_Yes	0.917913	4.265849e-10

The employees who had a higher total number of companies that the employee had worked in their careers would tend to have a higher probability of leaving. This can be explained by the nature of the employees that tend to frequently switch jobs or change careers, or those employees might have a long time of work experience, so these numbers were higher, and it was easier for them to change jobs, etc.

Compared with employees who did not have to travel, people who had a frequency of employee business travel had a higher probability of leaving.

Compared with employees who divorced, single people had a higher probability of leaving.

Compared with people who did not have to work overtime, employees who worked overtime had a higher probability of leaving. This is the factor that had the biggest impact on attrition. For future implications, there should be appropriate actions to manage workload, manpower and time to avoid overtime working issues.

## Questions 2:

*Formulate the problem outlined in Part C of the case as an optimization problem. Be sure to clarify what are the key elements of the problem as learned in class.*

In this problem, the **decision variables** are  $x_1, x_2, \dots, x_j$ , which are binary variables representing whether an RCC is offered to each group, with  $j$  being the number of groups.

The **objective** is:

$$\text{minimize } (x_1 * r_1 + x_2 * r_2 + \dots + x_j * r_j)$$

where  $r_j$  being the sum of the severance pay for all employees in group  $j$ .

The **constraints** are

- The minimum average number of employees to offer an RCC is 40. This can be represented as:

$$E_{\text{offer}} = x_1 * p_1 + x_2 * p_2 + \dots + x_j * p_j \geq 40$$

where  $p_j$  being the sum of probabilities of taking an RCC of the employees in group  $j$

- Yearly saving from salary is greater than 3 million Euros. This can be represented as

$$x_1 * p_1 * s_1 + x_2 * p_2 * s_2 + \dots + x_j * p_j * s_j > 3000000$$

where  $s_j$  being the sum of the annual salary rate of employees in group  $j$ , calculated by multiplying the MonthlyRate of each employee by 12. In this case, we cannot use MonthlyIncome because this figure represents what the employee receives, but we are actually interested in what the company actually has to pay out, which is represented by the MonthlyRate. This also needs to be normalised by the probability that an employee would leave.

- The percentage of job roles held by the remaining employee remain similar to before, subject to some tolerance.

$$M_{r,\text{before}} * (1 - \text{tolerance}) \leq M_{r,\text{after}} \leq M_{r,\text{before}} * (1 + \text{tolerance})$$

where:

$M_{r,\text{before}}$  being number of employees in a job role  $r$  divided by the total employees before the RCC ( $E_{\text{before}}$ )

$M_{r,after}$  being the number of employees in a job role  $r$  divided by the total employees after the RCC. The number of employees after the RCC is calculated by

$$E_{after} = E_{before} - E_{offer}$$

- At least 80% of employees need to stay in each department. This is given by

$$E_{d,after} \geq 0.8 * E_{d,before}$$

Where  $E_{d,after}$  being the number of employees in each department after the RCC and  $E_{d,before}$  being the number of employees in each department before the RCC.

### **Questions 3:**

*Open Starting\_model.xlsx. Each employee is given with a probability of accepting an RCC (please use this probability, not the one you predicted). Determine employee categories which may or may not be opened up to RCCs. In the xlsx file, employees have been assigned at random to placeholder categories - make sure to overwrite the group assignments based on your choices and modify the optimization problem accordingly. Moreover, make sure to justify your choice of employee categories, especially in light of possible discrimination issues.*

In this project, the categories are determined so that:

- They are not specific to a certain number of employees. In this case, 5 was chosen to be the minimum number of employees that a category needs to have.
- The results balance some sensitive attributes. In this project, the attributes were determined to be gender, age, and marital status.
- They allow efficient optimisation of the severance pay. To achieve this, multiple categorisation strategies were tested. The strategy that minimises the severance pay while balancing the sensitive attributes will be selected as final.

The following 4 tested categorisation strategies along with their reasonings are as follows:

- Severance pay: Because the objective is to minimise the severance pay, categorising based on this is a logical starting point and it lets the optimiser optimise directly the objective.
- Monthly Income: Because the severance pay was calculated directly based on this factor, optimising based on MonthlyIncome would allow indirectly optimise severance pay.
- Monthly Income x Job satisfaction and Monthly Income x Environment satisfaction: As the RCC can also be considered as a way to offer the unsatisfying employees to leave, categorising based on Job satisfaction and environment satisfaction lets the optimiser minimise the severance payment by offering the RCC to employees with high attrition probabilities.

**Question 4:**

*In your Solver model, define the severance cost for each employee (the current values are randomly generated), based on the case description.*

The severance pay for each employee was defined according to Maurice's email, which is:

- $\frac{1}{4}$  of monthly salary per work year for up to 10 years at the current company
- $\frac{1}{3}$  of monthly salary per work year from 10 years onward
- 3 times the monthly salary

Since the predicted probabilities were obtained using a machine learning model trained on the attrition data from the last acquisition, it is important to mimic the context of the last acquisition as much as possible. Therefore, because in the last acquisition, 3 months' salary on top of the base severance compensation were offered to all employees, it makes sense to apply the same logic when calculating the severance pay for the current acquisition so the probabilities can be comparable.

**Questions 5:**

*Solve the optimization problem using Excel Solver. Be sure to note which groups are opened up to RCC, as well as the cost you achieve.*

The groups opened up to RCC for each categorisation strategy can be found in the folder "output", the sub-folder named "severance\_\_offer".

**Question 6:**

*The optimization problem is incomplete. Add the following two constraints: (i) The average yearly amount gained from salary cuts must be greater than the savings expected by management and found in part A; (ii) In each of the three departments, at least 80% of the employees have to stay on. With the additional constraints, solve the problem again. How do the decision variables and the objective change?*

The difference in the decision variables is summarised as follows. In this case, a very minimal difference can be noted, especially in the middle severance groups, which corresponds to groups with average monthly income.

**Table 3.** Differences in decision variables and objective with 2 add constraints

Severance group	Incomplete Case (Income)	Complete Case (Income)	Difference
1	1	1	
2	1	1	
3	0	1	TRUE
4	1	1	
5	0	0	
6	0	0	
7	0	0	
8	0	0	
9	0	0	
10	0	0	
11	0	0	
12	0	0	
13	1	1	
14	1	0	TRUE
15	1	1	
16	1	1	
17	1	0	TRUE
18	0	1	TRUE
19	1	1	
20	0	0	

By adding the 2 constraints, the severance pay increased from €2.99 million to €3.37 million, which makes sense as the more constrained an optimisation problem is, the harder it is to achieve the same objective.

The achieved objective and constraints of the tested categorisation strategies are summarised as follows:

**Table 4.** Summary of achieved objective and constraints

Strategy	Categorise based on	Severance Pay (€million)	Employees leaving (average)	Yearly saving (€million)
1	MonthlyIncome	3.37	40.05	7.4
2	MonthlyIncome x Job Satisfaction	4.81	40.04	6.97
3	MonthlyIncome x Environment Satisfaction	4.23	40.07	7.48
4	Severance Pay	3.08	40.01	7.08

It can be seen that although categorising based on severance pay leads to the lowest severance pay of 3.08, the yearly saving is also much lower compared to categorising based on MonthlyIncome and Monthly Income x Environment satisfaction. Categorising based on Monthly Income x Job Satisfaction appears to be the least optimal as it yields the highest severance pay and lowest yearly saving.

The estimated yearly saving for each strategy is summarised below. The table suggests that although categorising based on severance pay leads to high saving in year 1 (albeit only slightly higher compared to Monthly Income), categorising based on Monthly Income leads to the highest saving in the long term.

Therefore, categorising based on Monthly Income can be considered the most financially optimal strategy.

**Table 5.** Summary of estimated yearly saving for each strategy

Strategy	Categorise based on	Severance Pay (€million)	Yearly saving (€million)	Year 1	Year 2	Year 3	Year 4
1	MonthlyIncome	3.37	7.4	4.03	11.43	18.83	26.23



2	MonthlyIncome x Job Satisfaction	4.81	6.97	2.16	9.13	16.10	23.07
3	MonthlyIncome x Env Satisfaction	4.23	7.48	3.25	10.73	18.21	25.69
4	Severance Pay	3.08	7.08	4.00	11.08	18.16	25.24

The tables below summarise the satisfaction of the remaining two constraints. It can be seen that all categorisation strategies have satisfied the constraints at a similar level.

**Table 6 & 7.** Summaries of the satisfaction of the remaining two constraints of all categorisation strategies

Job Role	Income	Income x JobSatisfaction	Income x EnvSatisfaction	Severance Pay
Healthcare Representative	8.18	7.32	8.87	7.88
Human Resources	-3.44	-4.21	-2.54	-2.51
Laboratory Technician	-9.71	-8.20	-8.68	-9.95
Manager	9.99	9.99	9.99	9.98
Manufacturing Director	5.14	3.05	5.48	5.83
Research Director	9.99	9.99	9.99	9.98
Research Scientist	-4.26	0.96	-1.65	-3.56
Sales Executive	2.42	-0.99	-0.94	1.60
Sales Representative	-9.59	-9.64	-8.57	-8.67
<b>Sum of absolute changes</b>	<b>62.71</b>	<b>54.34</b>	<b>56.71</b>	<b>59.96</b>

Department	Income	Income x JobSatisfaction	Income x EnvSatisfaction	Severance Pay
Human Resources	89.53	88.94	90.23	90.27
Research & Development	90.41	91.65	91.47	90.60
Sales	92.20	89.69	89.85	91.71
<b>Average remaining pct</b>	<b>90.72</b>	<b>90.09</b>	<b>90.52</b>	<b>90.86</b>

**Question 7:**

*Ensure that, given the optimal solution, whether an employee is offered an RCC is not implicitly linked to sensitive variables such as gender. Make sure you describe your verification process.*

This section analyses the RCC offer rate by each sensitive attribute.

In terms of Gender, categorising based on Income and based on Severance Pay alone led to the lowest disparity, with Female receiving less offer than Male by 2.29 and 1.34 percentage point, respectively. The remaining two strategies led to the highest disparities.

**Table 8.** RCC offer rates by gender

Gender	Income	Income x JobSatisfaction	Income x EnvSatisfaction	Severance Pay
Female	43.75	56.82	52.84	44.32
Male	46.04	49.43	44.15	45.66
Difference	-2.29	7.38	8.69	-1.34

The similar pattern can be seen when analysing by age groups, with categorising by Income and Severance Pay led to significantly lower in RCC offer rate. However, all strategies led to junior employees receiving RCC offer at a higher rate compared to more senior employees. Interestingly, categorising by Income x Environment Satisfaction sees more offers extending to the 34-39 age group. Age groups were defined as 5 bins, each containing the same number of employees.

**Table 9.** RCC offer rates by age group

Age group	Income	Income x JobSatisfaction	Income x EnvSatisfaction	Severance Pay
(17.999, 31.0]	51.58	54.74	48.42	52.63
(31.0, 34.0]	51.56	64.06	45.31	48.44
(34.0, 39.0]	42.11	53.95	61.84	40.79
(39.0, 46.0]	33.73	40.96	43.37	42.17
(46.0, 60.0]	29.41	33.82	26.47	29.41
Variance	81.89	115.58	128.28	62.41

Interestingly, in terms of marital status, categorising by Income leads to the highest disparity, at an extremely high level compared to the other strategies, with single employees receiving

more offers than divorced staffs by approximately 20 percentage points. Categorising by Severance pay led to the lowest variance.

**Table 10.** RCC offer rates by marital status

MaritalStatus	Income	Income JobSatisfaction	xIncome EnvSatisfaction	xSeverance Pay
Divorced	36.84	50.53	44.21	44.21
Married	40.93	50.26	46.63	43.01
Single	56.00	56.67	51.33	49.33
Variance	67.86	8.76	8.74	7.53

In conclusion, although categorising by Income yields the highest financial advantage in the long run, categorising by Severance Pay appears to balance the sensitive attributes. Therefore, it depends on the acquiring company to decide which factors they would like to emphasise more – financial by choosing categorising by Income, or publicity by choosing categorising by Severance Pay.

### **Question 8:**

*Discuss possible pros and cons of the optimization approach. What are the assumptions you are making implicitly, and how likely are they to be fulfilled?*

In terms of advantages, the optimization approach followed a quantitative and data-driven approach to optimising severance payment. It has identified multiple categorising strategies and has identified one optimal strategy for each business requirement (financial or publicity) that satisfies all constraints. Additionally, optimising with Python allows more flexibility compared to Excel. It allows us to quickly test out new categorisation strategies by modifying a few lines of code.

Regarding the downsides of the optimization approach, it assumes that this acquisition would be the same as the last when calculating leaving probabilities. An assumption was made that all employees are independent (assumption of the logistic regression model). This is not true because employees are given several weeks to think over the RCC offer, which means they could talk to each other and that could affect the predicted probabilities. For example, a colleague with a strong intention to stay but did not receive an RCC offer might become dissatisfied when talking to someone who had received the offer. Furthermore, this method only optimised 1 objective at a time, which is not ideal. This is because long-term saving is also important to the business, and we could have implemented a multi-objective optimisation method to minimize the severance payment and maximise yearly saving at the

same time. Additionally, since balancing the sensitive attributes is also a strong requirement, it is better to include them in the optimisation problem as constraints.