Improving speed and effectiveness of loan processes in a bank

Ube van der Ham

Meijer & Van der Ham Management Consultants, Dijkhoeve 4, 3832 KD Leusden, Netherlands, {ube.van.der.ham}(at)meijervanderham.nl

Abstract. A Dutch bank provides loans to consumers. The bank wants to improve the consumer credit process. This paper presents an analysis of the process, based on the computer logs of the process, with the use of process mining software. The analysis provides a clear view on the process, on the throughput times of different parts of the process and on the results of interaction from employees with the customers. Several opportunities are identified that can lead to improvement of the throughput times and the percentage of offers that are converted into orders.

Keywords: process mining, data mining, big data, organizational analysis, process analysis, process improvement, bank, loans.

1 Introduction

1.1 Loan Process from a Financial Institution

A Dutch financial institution provides financial services to retail clients (consumers and small business). The institution wants to remain anonymous, so little background information is available. In this paper, we will refer to the financial institution as “the bank” since this paper targets only banking activities of an institution that may or may not be a large conglomerate. The bank wants to evaluate and possibly improve the loan process. Loans for consumers are provided for purposes like:

- purchase of a car;
- home improvement;
- boat;
- motorcycle;
- business purposes (small entrepreneurs);
- etc.

The process is supported by IT-systems that log all events.

1.2 Business Process Intelligence Challenge 2017 with Public Data from the Bank

As part of the annual international conference in the field of Business Process Management, an International Workshop on Business Process Intelligence (BPI'17) is organized. The organizers of this workshop also organize an international contest: the Business Process Intelligence Challenge 2017 [1]. For this challenge, the bank has provided a log of relevant events of the consumer loan process. This log data was made publicly available for the contest [2].

1.3 Questions from the Bank

This paper aims to identify the possibilities for improvement with regard to the loan process, that may be distilled from a process log. The company is particularly interested in answers to the following questions [1]:

1. What are the throughput times per part of the process, in particular the difference between the time spent in the company's systems waiting for processing by a user and the time spent waiting on input from the applicant as this is currently unclear,
2. What is the influence on the frequency of incompleteness to the final outcome. The hypothesis here is that if applicants are confronted with more requests for completion, they are more likely to not accept the final offer,
3. How many customers ask for more than one offer (where it matters if these offers are asked for in a single conversation or in multiple conversations)? How does the conversion compare between applicants for whom a single offer is made and applicants for whom multiple offers are made?
4. Any other interesting trends, dependencies etc.

1.4 Approach

The analysis in this paper focusses solely on the logged data about the steps taken in the loan process. The content of the information provided through the call center was not investigated. Process mining techniques should provide insight in the flow through the process. For the analyses several software tools were used: Minit, Celonis, ProM, WEKA and MS Excel.

1.5 For Contest Purposes Only

This paper serves only as a contribution to the Business Process Intelligence Challenge 2017. No formal relationship exists between the author or his company and the bank. The analysis provided in this paper may therefore not formally be regarded as an advice or consult.

The type of numeric analysis presented in this paper, can only be the starting point of further investigation and is not meant to reach final conclusions. A proper analysis of the processes does require further interpretation of the results with the members of the organization involved.

This paper is written for a target audience of management of the bank rather than academics in the field of process mining.

2 Loan Process at a Dutch Bank

This chapter describes the context of the processes that are to be analyzed.

2.1 Loan process

The loan process starts with an application by the client through the website, or through an employee. The first event in all cases is “Create application”. The application is judged several times by the bank systems and employees. If the initial tests are passed, the bank will propose an offer to the client. The client can accept and return the offer. Then the bank will ask for additional information and evidence (pay slips, bank information), in order to make sure that the client can afford to pay back the loan and interest. If all information meets the standards of the bank, then the process results in an actual loan and the money will be transferred to the client. At any time during the process, the application can be cancelled by the client or the loan can be denied by the bank.

2.2 A Highly Structured Process

The process itself is highly structured. Loan processes are typically regulated both by law and internal regulations. These regulations must protect the consumers against loans for spending that they cannot afford and protect the bank against defaults or fraud. These regulations are usually enforced by implementing a tight process that is driven by IT-systems. Since many applications for loans are received through the websites of banks, the first steps in the process are structured as well. The IT-systems do usually not allow any deviations from the obligatory process, so we may expect a process that is suitable for analysis with process mining tools.

2.3 Success Criteria Determine Perspective of the Analysis

The information that is available will be regarded from the perspective of success factors for the bank: a high conversion of offers to sales of loans, low risk obtained by a proper procedure and a good interest. In order to sell many loans, the process must be fast and easy for the customer and should take little time from employees.

Of course, there are other factors that influence whether a client will buy a loan from a bank, like friendly employees, waiting queues at the call center, etc., but these cannot be investigated from the information that is available.

3 Dataset with 1,2 Million Events

For the purpose of the challenge 3 files were made available by the bank:

1. Offerlog 2017, 193.849 rows/events from 42.995 offers with offer number as case number;
2. Eventlog 2017, 1.202.267 rows/events, from 31.509 applications/42.995 offers;
3. Eventlog 2012, 262.200 rows/events, from 13.087 applications.

The events in the offer log are also present in the eventlog 2017. The eventlog 2012 was published for the BPIC2012, and not republished, but is still available in the 4TU datacenter. The 2017 log contains cases from 01-01-2016 to 01-02-2017, the 2012 log contains cases from 10-01-2011 to 01-03-2012.

The eventlog 2012 uses similar but slightly different event names. In order to compare the processes from 2012 and 2017, the event names in the 2012 dataset were replaced with the 2017 equivalent (if available). See Table 10 in the appendix for the values that were replaced.

The Eventlog 2017 contains 31.509 cases. 98 cases are incomplete. There are 3 possible outcomes of the process. **Table 1** presents an overview of the number of cases for each outcome.

Table 1. 3 possible outcomes of the process

Outcome	# cases
cancelled (by the client)	10.431
denied (by the bank)	3.752
pending: loan provided	17.228
Subtotal	**31.411**
case incomplete	98
Total	**31.509**

The incomplete cases are all cases from the last months in the log. Because of the small number, these incomplete cases can be left in the dataset for most analyses. The cases that were cancelled include a number of cases where the client did not formally cancel the application, but simply stopped responding. The bank system automatically closes these cases.

The 2017 dataset contains for all events a timestamp and a user (employee or system) and also some additional information on the loan: purpose, amount, number of terms, etc. An overview of all dimensions is listed in appendix 1. The 2012 dataset does not have this additional information, so a comparison of the two periods is not possible on all dimensions.

It is likely that there is an issue with the timestamps in the log. Many applications are automatically cancelled at 6.00 (summer) or 7.00 (winter) o'clock. The difference is probably caused by daylight saving time in summer with transitions at 28-3-2016, and 31-10-2016. The timestamps were not corrected since it is not certain that the same clock was used for the other events and the influence on throughput times is small.

4 Process Has Become More Complex Since 2012, Because of Additional Monitoring Activities

The process of 2017 was compared with the process in 2012 using the software of Minit. For this analysis, a dataset was created that combined the data of BPIC 2017 and BPIC2012. The events in the 2012 dataset were corrected to get identical spelling and naming of similar events.

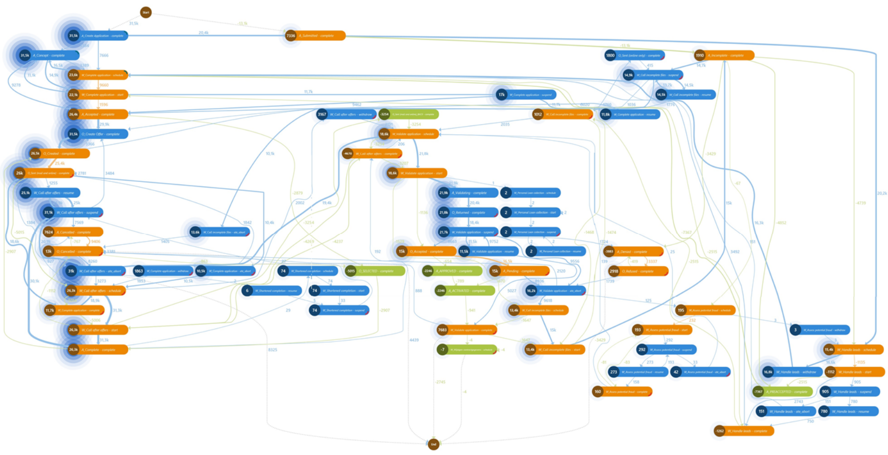

Figure 1 More activities in 2017 than in 2012 (100% activities, 1%paths).

The two processes are compared in Figure 1. The orange activities occur both in the 2012 log and the 2017 log. The blue activities are found only in the 2017 log. The green activities appear only in the 2012 log. Even without reading the labels of the activities, it is clear that there are more events in the 2017 log. The activities A_PREACCEPTED, O_SELECTED, A_APPROVED, A_ACTIVATED, O_Sent(mail and online)_BACK and W_Wijzigen Contractgegevens have disappeared. See **Figure 23** for a full-page illustration that can be inspected on screen.

New in the 2017 log are a number of suspend, resume and ate_abort events (W_Complete Application, W_Call after offers, W_assess potential fraud). These log entries indicate a new monitoring and scheduling system, while the work on the loan offers remains the same. (see Figure 2) New activities include a shortened completion procedure and a possibility for a "personal loan collection" that was hardly ever used.

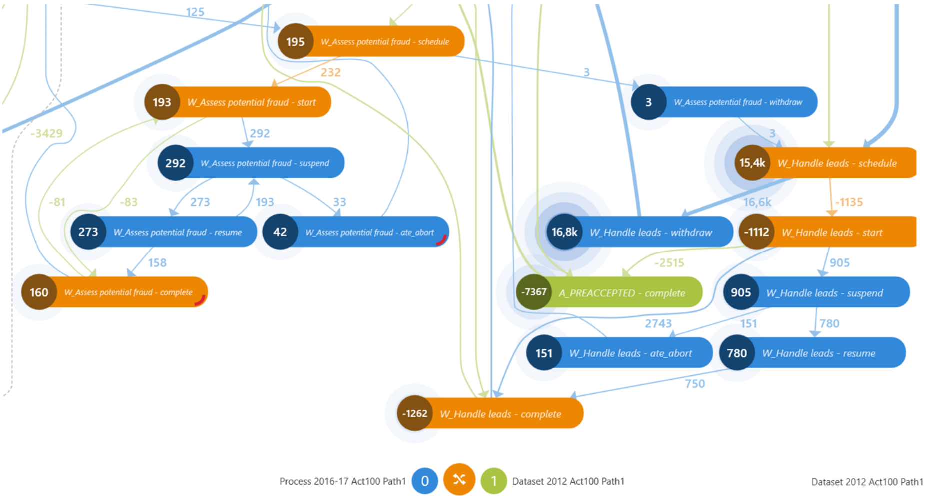

Figure 2 New suspend, resume and ate_abort events indicate a new monitoring and scheduling system.

5 Throughput Times Have Increased Since 2012

This chapter investigates the throughput times of different stages of the process and the time spent by employees. First the effect of fast service is shown. Later the causes of throughput time are investigated.

5.1 Throughput Time Has a Negative Effect on Conversion

Throughput times are hard to compare, since there are many different case characteristics. By comparing the very first stage, that all cases go though we see that speed seems to matter: if the first offer is sent within half a day, conversion from offer to a loan is (average) almost 60%, but the conversion rate declines to less than 40% when the client has to wait 5,5 days. By then, 96% of the cases have received an offer. (Figure 3)

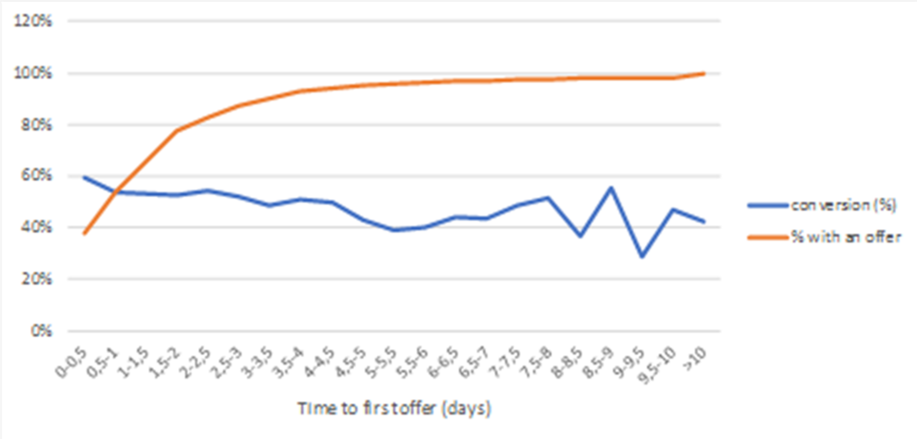

Figure 3 Conversion from offer to a loan (blue line) decreases with the time from application to the first offer.

5.2 Throughput Times Have Increased Since 2012

Since 2012 the average throughput time has increased from 8 to 22 days (all cases in both logs, see Appendix Figure 18 and Figure 19) The average should be compensated for differences in types of loans and the number of loans that are converted in order to make a fully reliable comparison, but the difference is so big, that we should look into all factors that are currently delaying the process.

5.3 Dominant Routing Has Changed Since 2012

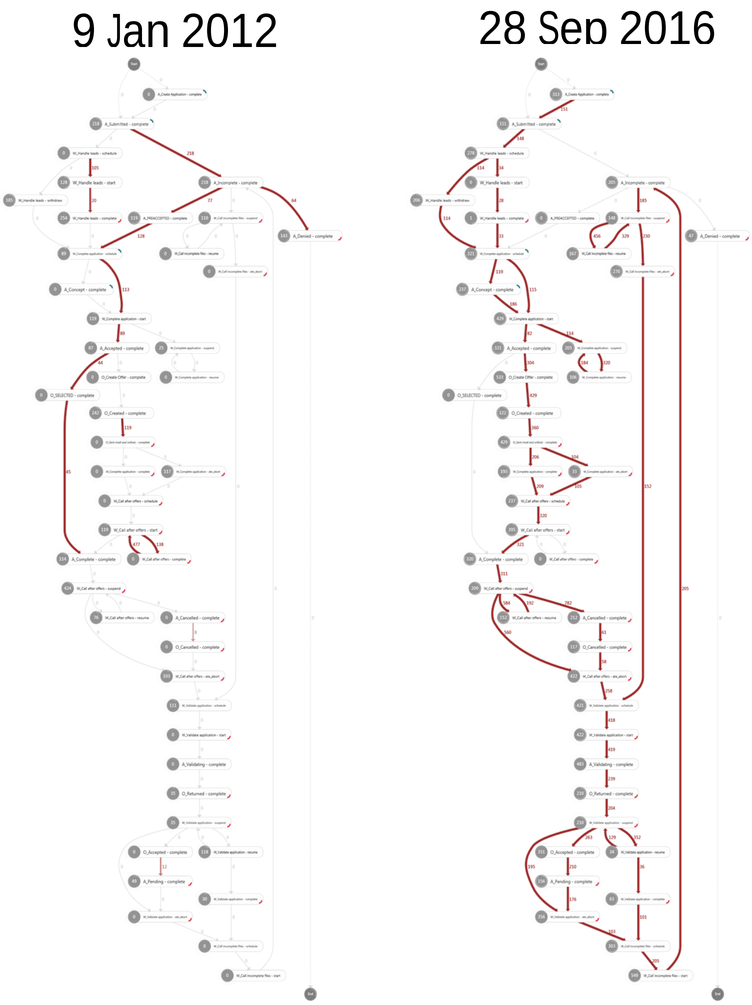

Figure 4 Increased workload in all steps of the process.

A clear way to show the differences between the two years (2012 versus 2016) is to combine the logs and freeze the animation of the process on two different moments in time. The activities in the pictures have the same layout, so differences are immediately clear. The red lines show where the activity is. We can see that the work used to pile up in the first steps of the process, but in 2016 all steps show work in process. This is partly because of additional activities, but also because of the higher workload.

5.4 Throughput Time Consists Mainly of Waiting for Client and Employee

The bank likes to know which part of the time of the process is spent idle and which part is spent working on the application. For a number of subsequent activities, it is possible to determine the start and end of activities of the employees: e.g. the time between Call after offers-start and Call after offers-suspend is probably spend by the employee working on the case. For the idle time, it is less evident if the waiting is for the client to respond or for the bank to make a phone call. There are also many activities that have no formal start, so it is impossible to calculate the exact time that is spent by employees. Only a coarse estimate can be given for the idle time and time spent working on the case.

For the purpose of this analysis, a selection of 8.464 new car loan applications was made, and the time for all transitions was counted in a directly follows matrix in Excel. Then every transition was classified as one of three categories: work, idle, consumer response. The transition matrix was then multiplied with a vector to get the average times for the three categories on each transition.

Table 2. Time allocation (average per case)

Time	# days total
work of employee	0,05
idle (bank)	3,55
consumer response time	17,56
Total	**21,16**

The estimate in **Table 2** does not take parallel processes into account, so the overall time spent will be lower. The employees time is about 78 minutes per case, which is probably a low estimate, because many activities have no start in the log. Celonis calculates an average of 20 days throughput, so the estimates are not too far off reality. In **Table 3** the time is given per stage of the process.

Most employee time is spent in the completion of applications and validation of applications. The cancelation also takes a lot of time. It must be noted that the time spent for a cancelled application is different from an application that results in a loan (which does not cancel and needs more time for validation and completion).

Table 3. Time allocation in days per stage (average per case)

Stage	Work	Idle	Consumer	Total
1_Create_Application	0,000548	0,837798	0,000003	0,838349
2_CompleteApplication	0,019586	1,327875	1,978278	3,325739
3_CallAfterOffers	0,000665	0,631087	12,414733	13,046485
4_ValidateApplication	0,011914	0,134912	1,559319	1,706145
5_CallInComplete	0,003239	0,508603	1,604088	2,115929
6_Fraud	0,002199	0,042732	-	0,044932
Cancelled	0,016172	0,064783	0,000174	0,081129
Denied	0,000000	-	-	0,000000
Total	**0,054322803**	**3,547789026**	**17,55659616**	**21,158708**

5.5 Throughput Time of the First Stage Is Worse for Sundays and Nightly Applications

The throughput time depends on the moment a client files his application. If a client applies on a Monday at 9 o'clock an offer is made in 1,08 days (see **Table 4**), while an application from Sunday 21.00 takes 2 to 3 days, while only 12 hours of idle time should be added. It seems that employees work according to Last In First Out (LIFO) The evening and Sunday applications are used as a buffer supply of work. Inspection of the log shows that employees will first work on other cases, taking new applications as well. The applications from Sunday April 3, are for instance mainly processed on Tuesday.

Table 4. Applications filed in the evening or on Sundays take longer. Average time to first offer in days plotted on a calendar with the hour and day of application. Applications between 1 and 5 AM are rare, so less significant.

Time	Mon	Tue	Wed	Thu	Fri	Sat	Sun
0	1,20	1,71	2,05	3,69	2,79	1,98	3,00
1	2,04	3,80	5,06	1,74	2,61	3,67	2,08
2	2,12	2,30	1,93	2,17	4,08	2,62	2,55
3	1,95	3,50	0,86	2,74	2,67	1,63	3,50
4	1,51	0,43	2,34	2,26	2,44	2,69	2,01
5	1,23	1,94	1,38	1,83	2,51	2,42	3,40
6	1,52	1,96	1,87	1,79	2,33	2,02	1,66
7	0,93	1,22	1,28	1,37	1,84	2,68	2,48
8	1,12	0,96	1,14	1,46	1,44	1,63	2,45
9	1,08	1,33	1,03	1,42	1,61	1,27	1,97
10	1,05	1,34	1,03	1,31	1,47	1,50	2,17
11	1,17	1,23	1,22	1,41	1,61	1,46	2,04
12	1,27	1,12	1,12	1,41	1,68	1,21	2,31
13	1,11	1,11	1,29	1,51	1,60	1,49	2,28
14	1,23	1,18	1,31	1,62	1,77	1,37	2,02
15	1,05	1,22	1,32	2,09	1,73	2,31	2,02
16	1,42	1,36	1,27	1,41	1,93	2,88	2,17
17	1,23	1,07	1,24	1,89	1,79	2,32	1,98
18	1,47	1,39	1,42	1,95	1,89	2,46	1,84
19	1,90	1,74	1,89	2,07	2,37	3,37	1,80
20	2,27	1,91	2,26	2,40	2,40	2,74	2,04
21	2,10	2,14	2,11	2,61	2,48	2,54	2,01
22	2,20	2,69	2,56	2,12	2,75	2,38	3,27
23	1,88	1,69	2,10	2,49	2,19	3,19	2,50
Total	1,29	1,35	1,37	1,68	1,78	1,87	2,12

The differences between the slow and fast cases do not result from a different process. Cases were compared in Figure 5. On the left the procedure for applications from Monday 9-16h and on the right for evening applications. The arrows indicate the major differences: Handle leads-schedule to Handle leads-start takes 12 hours extra and A_Concept-complete to Complete application-start takes 9 hours extra.

The fastest cases are those where an agent does the application for a new credit (see **Table 5**). The agent will process the first steps immediately.

A limit raise cannot be applied for through the website. Agents don't produce applications after 20:00.

Table 5. Time to first offer in days.

(days)	Monday 9-16h		Evening 20-23h	
	limit raise	**new credit**	**limit raise**	**new credit**
agent	1,16	0,18	-	-
website (user1)	-	1,81	-	2,19

Figure 5. Fast procedure on Monday 9.00-16.59h (left) and slow for evening applications (right).

5.6 Several Users May Be Bottlenecks in the Process

If we set the user as the activity in a process map, then we see the flow of work through the organization. Figure 6 shows part of the process map. After an application is received through the website (User_1) the cases are distributed over a great number of employees. In the next step a much smaller number of employees take over. They have to handle many cases and we see a long delay during the handover from one employee to the other.

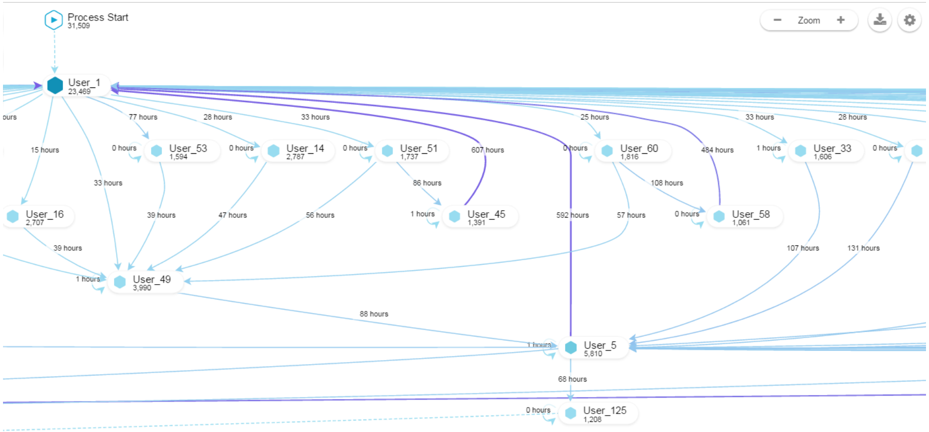

Figure 6. A layered organization of work, with several bottlenecks after handover of work to a second employee (98,7% of cases, 67% of connections)

Cases that are returned to User_1 (purple lines) are the cases that get cancelled automatically because the client does not respond to the offer or to requests for more information. Users with the highest caseload are listed in **Table 6**. Most users pass the cases on to User_3 and User_5. These 2 users each have their own cluster of users that supply a great part of their work. User_3 and User_5 have similar activity profiles. The main activities of User_5 are W_Call after offers (suspend and resume), but that is only half of his work (see Appendix Figure 22).

Table 6. Several users may be “bottlenecks” in the process.

User	#cases processed
3	6077
5	5810
49	3990
87	4407
100	3827
109	2167
123	3239

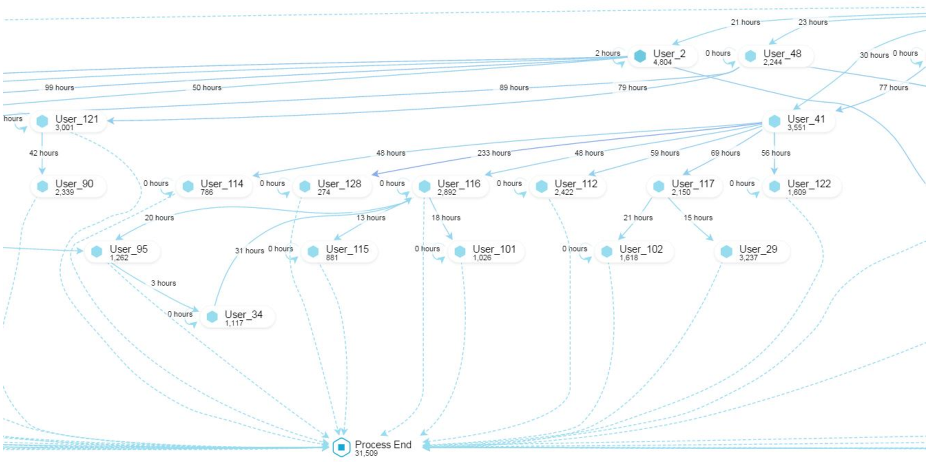

Figure 7. Last part of the work has a different structure and some bottlenecks.

User 3 passes the work mainly on to a cluster of Users (users numbers >100) that are responsible for the last steps in the process (See appendix, **Figure** 20 and Figure 21.). The cluster of users that execute the last steps, also has several bottlenecks (User 100, 109,123).

It is not clear from the log if users 3 and 5 have a unique role and must be involved in all of these cases and it also not clear if these cases are waiting for these 2 users of are waiting for the client to respond. Because of the times that cases are waiting for processing, it would be good to look into their role and see if User_3 and User_5 have an overload of work and then consider spreading the work of these resources over a larger group of employees.

6 Conversion Depends on Several Factors

The bank would like to sell as many loans as possible to the clients of their choice, at the lowest possible cost. If a consumer applies for a loan, then the bank wants to convert this consumer into a client. This requires a competitive offer and a good process. This chapter presents a number of factors that correlate with a high or low conversion.

6.1 Conversion Does Not Suffer from More Contact

If clients want a loan they must provide information and evidence about income, regular expenses, other debts etc. so the bank can judge if the client will be able to repay the loan. The bank wants to know (question 2) if applicants that are confronted with more requests for additional information/evidence, are more likely to reject the final offer.

In order to judge the loss of conversion, we must first see where the losses in the process occur. For this purpose, the process maps were converted to a simplified sales funnel as shown in Figure 8. All back loops were removed so arrows do not necessarily represent a direct transition. The main process is reduced to a few stages. The numbers represent the case count for the last stage that the case will eventually reach.

The process is divided into logical stages:

1. completing the application, creating and sending an offer
2. calling the clients after the offer
3. validation of the application
4. completing the evidence and information to finalize the loan

that are separated by the activities in Figure 8

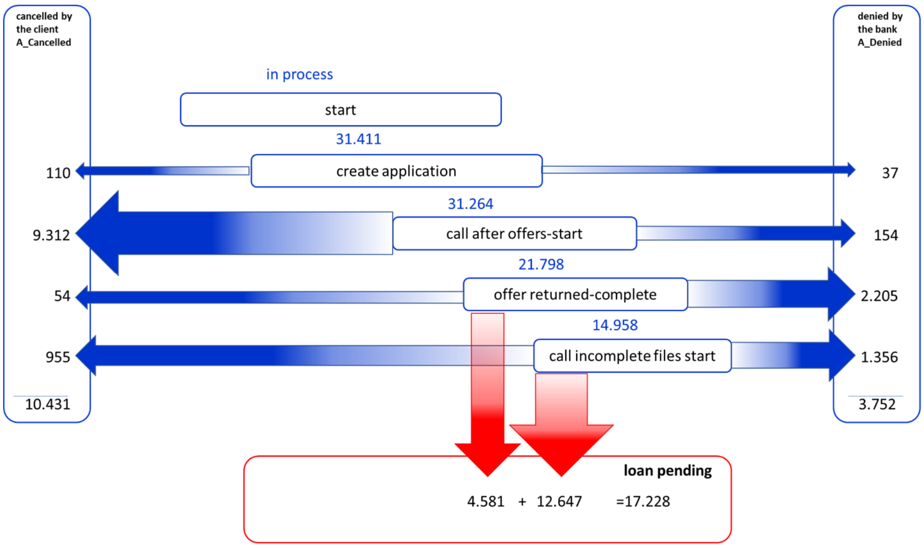

Figure 8. Most cancellations by clients during "call after offers".

From the analysis in Figure 8 we can conclude that most potential clients are lost during the call after offers stage (9.312 cancelled) and not during the call incomplete files stage (955 cancelled).

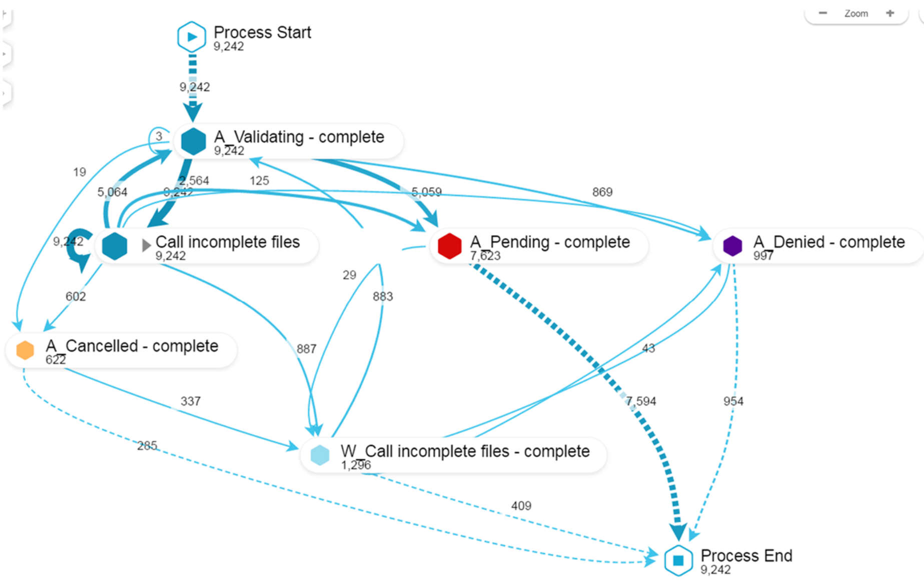

Figure 9. Example of process map with focus on Call incomplete files : 622 applications were cancelled in cases with a single "Call incomplete files-start. Analysis with Celonis.

In order to look into the effect of the number of requests to supply information, we will split up the last step in the funnel. The log contains an event "W_call incomplete files-start" that will be counted as the number of information requests. In order to judge the feared effect on conversion correctly, we must distinguish between cases that end up "pending" versus "cancelled" by the client. Cases that were "denied" by the bank should not be taken into account for this question. In order to get the right numbers, the log was filtered. Only the 31.411 cases that go through cancelled/pending/denied were used (meaning all cases that are still open were not taken into account). Then a second filter was applied to show only cases with 1 (or 2/3/4/>4) counts of W_call incomplete files-start. For every number a small process map was drawn as shown in Figure 9. This way we can also count cases that have intermediate steps between the call and the final result. Five of these process maps were summarized in **Table 7**. We can see that after each additional request, additional clients cancel the application, but the number of cancellations does not increase compared to the cases that go to "pending". Even if we calculate cancellations as a fraction of what is still in process (cancelled / (cancelled + pending + to next questions)) there is no increase in cancellations. The number of denied cases is even bigger than the number of cancellations, so skipping the additional information requests does not seem a good idea.

Table 7. An increased number of information requests, does not increasingly cause clients to cancel their applications.

after step ….	cases (eventually) go to					
#call incomplete files start	**cancelled**	**pending**	**cancelled**	**pending**	**(denied)**	**total**
1	622	7623	7,5%	92,5%	997	9242
2	232	3463	6,3%	93,7%	256	3951
3	73	1113	6,2%	93,8%	74	1260
4	19	319	5,6%	94,4%	23	361
>4	9	129	6,5%	93,5%	6	144
total	**955**	**12647**	**7,0%**	**93,0%**	**1356**	**14958**

#call incomplete files start	cancelled: % of in process
1	4,5%
2	4,2%
3	4,3%
4	3,9%
>4	6,5%

6.2 Better Conversion for Multiple Offers on Different Dates

As we have seen earlier, most cancellations are found in the call after offers stage. This stage has the biggest potential for improvements in the sales funnel. The bank wants to know (question 3) if clients appreciate getting several options to choose from and if that leads to a higher percentage of cases where a loan offer is accepted by the client and finally converted to a loan. The bank can send multiple offers at once, and/or send a new revised offer after contacting the client. It is not clear from the data if this is done at the clients request or as a special service of the bank. For this analysis, the event O_created –complete is regarded as the event that will be evaluated as the loan offer. The event that marks the conversion of the offer into a loan, is A_Pending-complete.

The great majority (22.950) of cases get only one offer. 3.491 customers (cases) get multiple offers on the first date. Then there is a group of 5.068 customers that get revised offers. A revised offer is defined here as a new offer on a different day. **Table 8** shows the conversion rates for offers. If a client gets has more offers than offer dates, then multiple offers were made on the same day. Revised offers have a better conversion, but multiple offers on a single day correlate with lower conversion. Further investigation should make clear what is cause or effect. A call center agent may give multiple offers because he feels that a client is in doubt about asking for a loan.

Table 8. Conversion rates improve for revised offers, but not for more offers on the same day.

	#offers					
#offerdates	**1**	**2**	**3**	**4**	**>4**	**total**
1	53%	48%	54%	45%	55%	52%
2		66%	66%	55%	65%	66%
3			71%	68%	63%	69%
4				78%	86%	81%
>4					69%	69%
total	53%	57%	66%	60%	67%	55%

6.3 Low Conversion for Two Special Days with a High Number of (Special) Revised Offers

A special type of revised offer is an offer with an improved (lower) interest rate. Many customers get an improved offer. For each loan, we can estimate the interest percentage. This may help us to judge the attractiveness of the loan for the customer and for the bank. For each loan offer, the dataset gives the amount of the loan, the monthly cost, and the number of months that this amount must be paid. For most loans, also the initial withdrawal is given. This amount is usually equal to the full amount of the loan. From this data, the interest percentage was estimated in Excel, assuming a 100% initial withdrawal, a constant monthly amount and a full payback of the loan. The interest rate that results from this calculation, may be slightly different from the rate that is calculated by the bank, but it can very well serve as an estimate that allows us to compare different offers.

The interest rate appears to depend on several factors, that cannot be influenced by the employees:

- purpose of loan;
- amount of the loan: higher amounts have a lower rate;
- number of months: longer running loans have lower interest rates;
- date: interest rates go down during 2016.

Despite this fact, some people get an improved offer for a lower interest. An analysis with WEKA/J48 to explain which cases get an improved offer, show that a few factors explain 97% of the cases with lower interest: several dates and cases with offers for a significantly higher amount. Higher amounts always get a lower interest rate, so that seems up to standard. The influence of the date will be investigated in this paragraph.

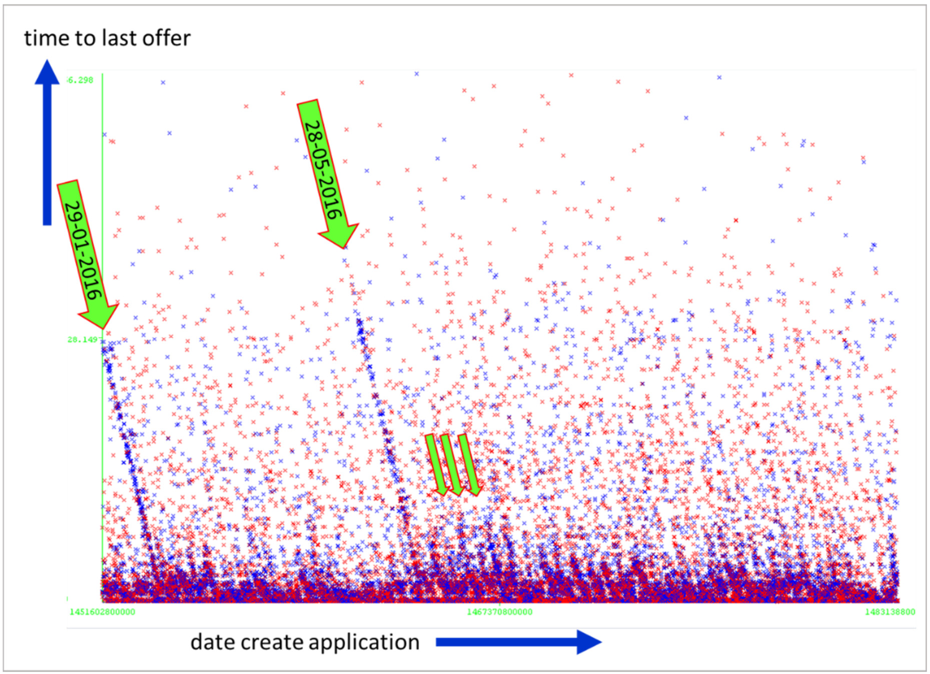

Figure 10 Two dates of last offer have a particularly low conversion rate. Blue dots represent offers that do not result in a loan. Analysis with WEKA.

Figure 10 shows the cases that are converted into a loan (red markers) and the cases that did not convert (blue markers). The horizontal axis is the date of Create_Application and the vertical axis is the time to the last offer from the date of Create_Application. Diagonal lines of dots represent cases with the same date of last offer. On two dates a large number of cases get their last offer: Friday 29-01-2016 (385 cases) and Saturday 28-05-2016 (348 cases), while an average day counts some 75 cases. In other weeks, there is a peak on Mondays (small green arrows), but that day is usually more successful in the sense that a normal conversion rate results from these cases.

It is not clear from the data what is the cause of the bad conversion. Possibly the peak was caused by a cleanup of cases that were neglected, or a poorly executed action to meet a monthly target. It looks as if it might concern a special offer, anticipating on a descending trend in interest rates. The interest rates in February and June were in general lower than before. A small number of employees participated in the action. The bad conversion should however be a reason for further investigation into these cases.

One user (User_71) was successful with a score of 49 out of 80 cases, so it might be interesting to learn from his experiences. Usually the IT-systems in a bank do not allow employees to adjust the interest rates, but employees can influence clients to boost their own performance metrics: “If you wait two weeks, I can give you a better offer.” The fact that a small group of employees have exceptionally high numbers of improved offers and that one employee has a high conversion rate on these offers, lead to the recommendation to check if the improved offers comply with company policies, though it must be stressed that User_71 often has no previous involvement with these cases. Of additional interest is that at 27-06-2016, 154 applications were cancelled by user1 (system), not in the normal batch procedure at 6:00 o'clock, but between 21:12:10 and 22:29:45 and one at a time, consistent with manual operation.

6.4 High Conversion for Limit Raise

Some clients already have a loan from the bank and want to raise the amount of that loan. These offers are usually more successful than those for new credits. Limit raise applications have a conversion of 73% versus 52% for new credits.

6.5 Low conversion for Nightly Applications and Sunday Applications

Sundays and evening applications have a lower conversion. This may be due to a slower process and influenced by the fact that limit raises are registered during office hours. Even after calculating the effect of limit raises conversion is still low during the evenings. See Figure 11.

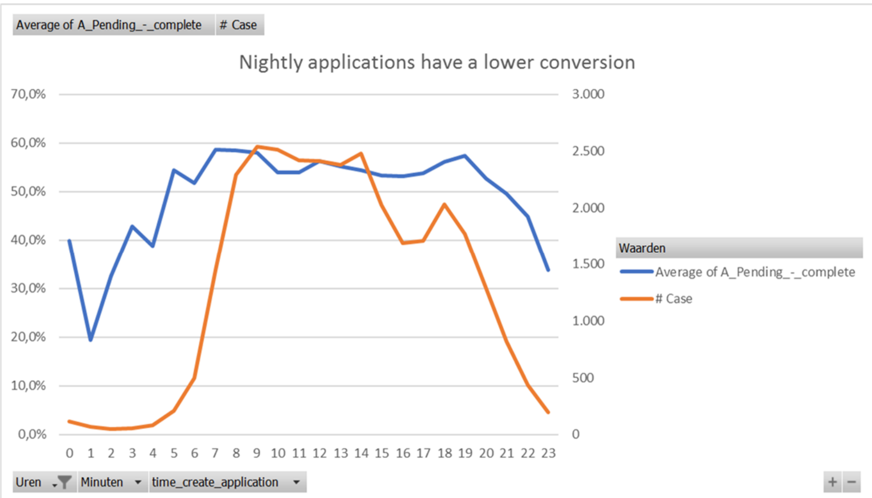

Figure 11 Conversion rate (blue line) is lower for applications during evenings and nights.

6.6 Low Conversion for Offers That Do Not Match the Requested Amount

When a client doesn't get the amount that he wants, then it is more likely that he will cancel the application. A lower amount will not allow to buy the car he likes, so these clients cannot be won with a better process. This can be seen in **Figure 12**. The diagonal lines in this graph are the result of the fact that people often ask for rounded figures (€ 1.000/5.000/10.000/15.000).

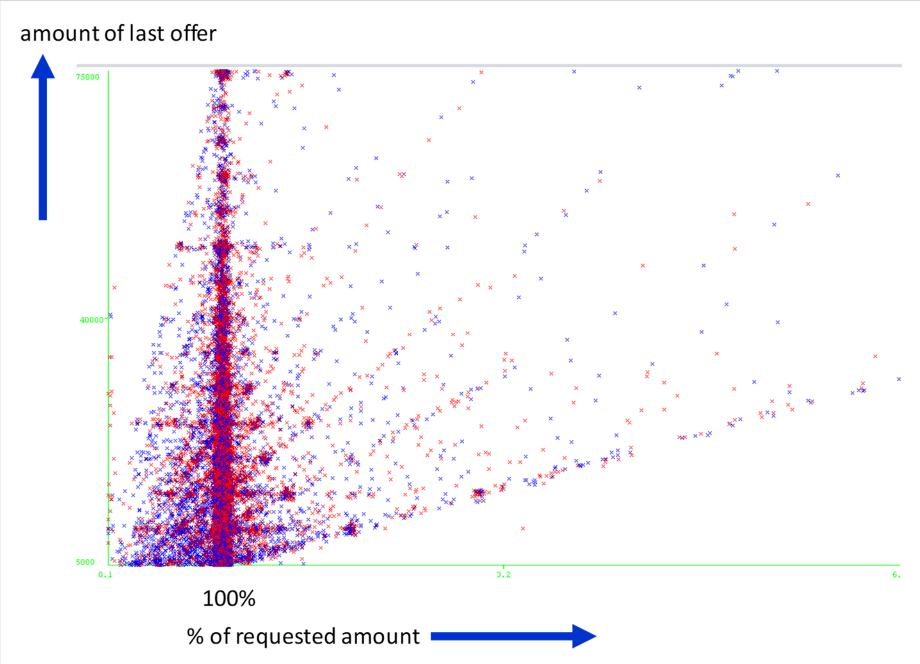

Figure 12 More blue dots (cancellations) for offers that do not match the requested amount. Analysis with WEKA.

6.7 Conversion May Be Lost, Because of Clients not Being Called Again

During a period of two months (end of October 2016 to end of December 2016), clients were scheduled to be called after offers, but 927 cases show a direct connection between W_Call after offers-scheduled and A_Cancelled-complete (almost 3% of all cases in the log) after an average 24 days. This means that customers were not called, although they were scheduled.

These cases show up when we compare in Minit the cases that go to Pending with the cases that do not go to Pending. These cases did receive a phone call earlier, but were not called again. This represents a total loan value of 16,5 mio euro.

There is no clear relation with the user that scheduled the Call, but user 69 leads the top 25 with 73 cases that are scheduled, but not called a second time.

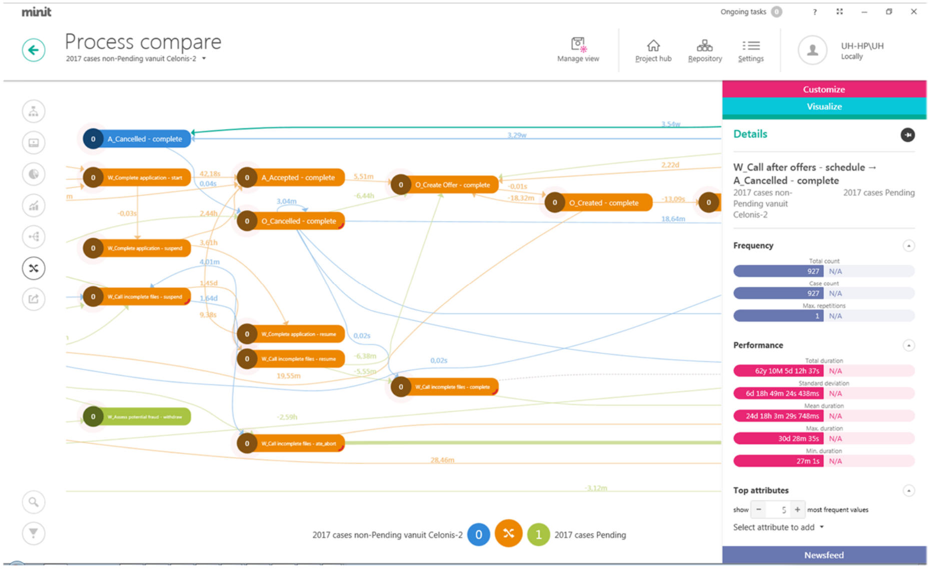

Figure 13 Clients were not called again in November and December

7 Revision of the Role of the Credit score and Accepted Status

7.1 Many Cases are Pending Though Not Accepted

An application must be accepted by the bank. According to the information provided on the ProM-Forum, the bank will not accept the customer if the loan and/or customer does not meet the criteria of the bank. Nevertheless, a significant number of cases does not have the status Accepted=true, but do go to “Pending” (meaning that the loan is provided). The cases do go through the event “O_Accepted-complete”. The information provided suggests that the loan should not be provided if the proper column is not marked as accepted.

A manual override is possible, but the frequent character of this practice, raises the question if procedures should be revised or enforced. There are 31.509 applications in the log. 17.228 of these get a loan. 3.060 offers are not marked as accepted but do nevertheless get the loan. An example is Application_1999238509. The client gets an offer on January 4, and the column “selected” is marked but the column “accepted” is not. The loan is provided (A_Pending complete) on January 27.

7.2 Not Accepted Below Credit score 700

A large part of those cases have a low credit score. A credit score is an indicator to judge if a case can be accepted. Credit scores are usually bought from an external agency and come at a price. Remarkable is that the log contains credit scores for only those cases that got a loan. Limit raise clients usually do not get a credit score. The policy may be to buy credit scores only at the end of the process to avoid unnecessary cost, but in the log the credit score is found at the event O_Create Offer-complete. This raises the question why the credit scores for other cases are missing.

Most cases in the log with a credit score lower than 700 are not accepted (see Figure 14). The horizontal axis gives the credit score, the vertical axis is the amount of the last offer divided by the amount requested. Most people get the amount they asked for (horizontal line). Jitter in the picture allows to see where most cases are.

Figure 14 Lower than 700 credit scores are usually not accepted. Analysis with WEKA.

The cases with credit score 0 are the cases that did not result in a loan, or limit raise cases, that apparently do not require a credit score.

7.3 Pending Though Credit Score Below 700

Figure 15 Lower than 700 credit scores do get a loan. Analysis with WEKA.

If we compare Figure 14 and Figure 15 it is immediately clear that cases without the accepted status go to “pending”. Cases with low credit scores are not necessarily given a lower loan than requested. Surprisingly the cases with a lower than 700 credit score do not follow a different process and do not take more time in the process. One would expect a separate check, or additional approval from a special officer, but nothing in the log indicates additional precautions. We only know from the information on the ProM Forum that a manual override is possible. Many cases with credit scores >700 are not accepted either, but do get a loan. Further investigation is needed to see if this complies with bank rules.

7.4 Revision of the Use of Credit Scores and Accepted Status

If current practices are compliant with the intentions of the bank, then management should address the following questions:

- Does the bank still want to use credit scores or are they a merely a remainder from past procedures, that are no longer enforced?
- Should the rules for the “Accepted status” be changed, since so many loans can be provided without being accepted. Manual overrides could be analyzed and translated into new rules that can be automated.

7.5 New Commercial Possibilities with Clients with High Credit Score

The credit score is currently not a factor that determines the interest rate. There may be a possibility to improve conversion for clients with a higher credit score, by offering them a lower interest rate. These loans have a lower risk of default and need a lower risk margin. This may also open opportunities to attract new reliable clients for other financial products.

8 Conclusions and Recommendations

A number of preliminary conclusions can be drawn from the datasets of the loan process. These findings need to be verified and discussed with the bank. These results can be the starting point of further research into the qualitative aspects of the processes involved. Figure 16 illustrates this point.

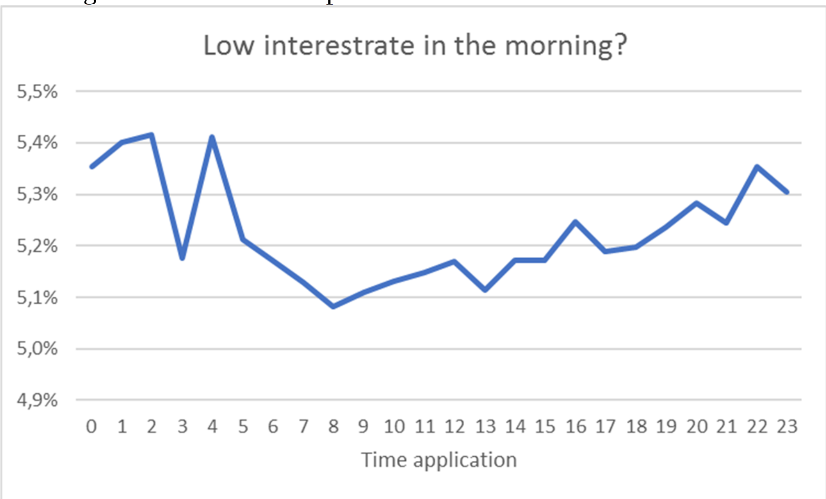

Figure 16 No final conclusions from numbers, because they can be deceptive.

An analysis based on facts and figures may seem convincing, but don't set your alarm clock for an early bird discount! The lower interest rate for morning hours is caused by a different mix in the purpose of the loans.

8.1 Summary of Conclusions

The main results from the analysis of the logs of the loan processes are:

- Since 2012 the process has become more complex and the throughput time has increased.
- The throughput time is mainly determined by waiting for the client and by idle time.
- Slow response to applications results in low conversion.
- Employees tend to work on incoming work first and leave yesterday's applications from the website as a buffer supply of work.

- Several users may be bottlenecks in the process.
- Conversion does not suffer from multiple requests for information.
- Multiple offers get a better conversion if not given at the same time.
- Special offers resulted in low conversion.
- During two months, cases were left idle for several weeks, clients not being called despite scheduling.
- An application with credit score <700 will not be accepted.
- Many applications that are not accepted, or not selected, do go to the pending stage.
- Interest rate does not depend on credit score.

8.2 Recommendations

From the analysis in this paper it is recommended to the management of the bank:

- to investigate why some clients have to wait long for an offer, and see if this time can be reduced.
- to investigate if 2 users that may be bottlenecks have special obligations and a high workload.
- to investigate if the special offers of 29-1-2016 and 28-5-2016 are compliant with bank rules.
- to investigate why 154 of these offers were not cancelled in a normal procedure, but in a separate procedure on 27-6-2016 with characteristics of a manual operation.
- to investigate why clients were not called again despite being scheduled.
- to investigate if the practice of giving loans to applications that were not marked as accepted by the bank, is compliant with bank rules.
- to investigate if manual overrides could be analyzed and translated into new rules that can be automated.
- to investigate if the credit score should still be used in the future, since it does seem not have a significant influence on the decision to provide a loan.
- to consider offering more attractive loans to clients with high credit scores.

9 References

1. TUE, http://www.win.tue.nl/bpi/doku.php?id=2017:challenge
2. Dongen, B.F. van, Dataset BPI Challenge 2017, http://dx.doi.org/10.4121/uuid:5f3067df-f10b-45da-b98b-86ae4c7a310b
3 Dongen, B.F. van, Dataset BPI Challenge 2012, http://dx.doi.org/10.4121/uuid:3926db30-f712-4394-aebc-75976070e91f

Appendix 1: Variables in the Log

The log files have a similar structure with variables that identify loan application and events.

Table 9 Variables in the log

variable	**description**
Accepted	true/false: if the bank still wants to provide the loan
Action	transition
Case-ApplicationType	new credit or limit raise
Case-Id	
Case-LoanGoal	several categories like car, home, motorcycle
Case-RequestedAmount	amount that the consumer wants for a loan
Case-concept:name	Unique casenumber
CreditScore	integer, giving indication of ability to repay the loan
Event-Id	event number
Event-Name	description of event
EventID	
EventOrigin	
FirstWithdrawalAmount	
MonthlyCost	amount to be paid by the consumer
NumberOfTerms	months that the consumer must repay
OfferID	number of offer
OfferedAmount	amounts that the bank wants to provide
Selected	true/false: if the consumer wants the loan
Sorting	
concept:name	e.g. A_Cancelled
Timestamp	date and time
lifecycle:transition	complete
org:resource	emloyee

The 2012 log has slightly different activity names and spelling. These were replaced by 2016 names. The yellow marked activities were not present in the 2016 log and were not changed.

Table 10 Replaced variables in the 2012 log

2012 Activity name	**Replaced (2016 name)**
A_ACCEPTED	A_Accepted
A_ACTIVATED	A_ACTIVATED
A_APPROVED	A_APPROVED
A_CANCELLED	A_Cancelled
A_DECLINED	A_Denied
A_FINALIZED	A_Complete
A_PARTLYSUBMITTED	A_Incomplete
A_PREACCEPTED	A_PREACCEPTED
A_REGISTERED	A_Pending
A_SUBMITTED	A_Submitted
O_ACCEPTED	O_Accepted
O_CANCELLED	O_Cancelled
O_CREATED	O_Created
O_DECLINED	O_Refused
O_SELECTED	O_SELECTED
O_SENT	O_Sent (mail and online)
O_SENT_BACK	O_Returned
W_Afhandelen leads	W_Handle leads
W_Beoordelen fraude	W_Assess potential fraud
W_Completeren aanvraag	W_Complete application
W_Nabellen incomplete dossiers	W_Call incomplete files
W_Nabellen offertes	W_Call after offers
W_Valideren aanvraag	W_Validate application
W_Wijzigen contractgegevens	W_Wijzigen contractgegevens
COMPLETE	complete
SCHEDULE	schedule
START	start

Appendix 2: Dimensions for Causal Analysis

These dimensions were used to see if it is possible to predict or explain conversion. The data was analyzed with different tools in WEKA.

Case
ID-application
amountRequested
creditScore
applicationType
purposeLoan
date_create_application
time_create_application
resource_create_application
O_Created_-_complete
#_offer_dates
#_offers
W_Call_incomplete_files_-_start
W_Call_incomplete_files_-_ate_abort
W_Call_incomplete_files_-_complete
last_resourceIncomplete
last_Date_Incomplete
last_time_Incomplete
lastEventIncomplete
date_first_offer
date_last_offer
amount_last_offer
%OfRequestedAmount
resource_last_offer
interest_last_offer
Selected_last_offer
Accepted_last_offer
simultaneous_offers_y/n
repeat_offers_y/n
improved_repeat_offer_y/n
A_Pending_-_complete
Timestamp_Pending
org:resource_Pending
TimeToFirstOffer
TimeToLastOffer
TimeToComplete
TimeToPending
Day (monday=1)

Appendix 3: Supporting Data

1 This appendix contains data that supports the text. It allows for verification, but does not provide new insights.

Process map 2016-17

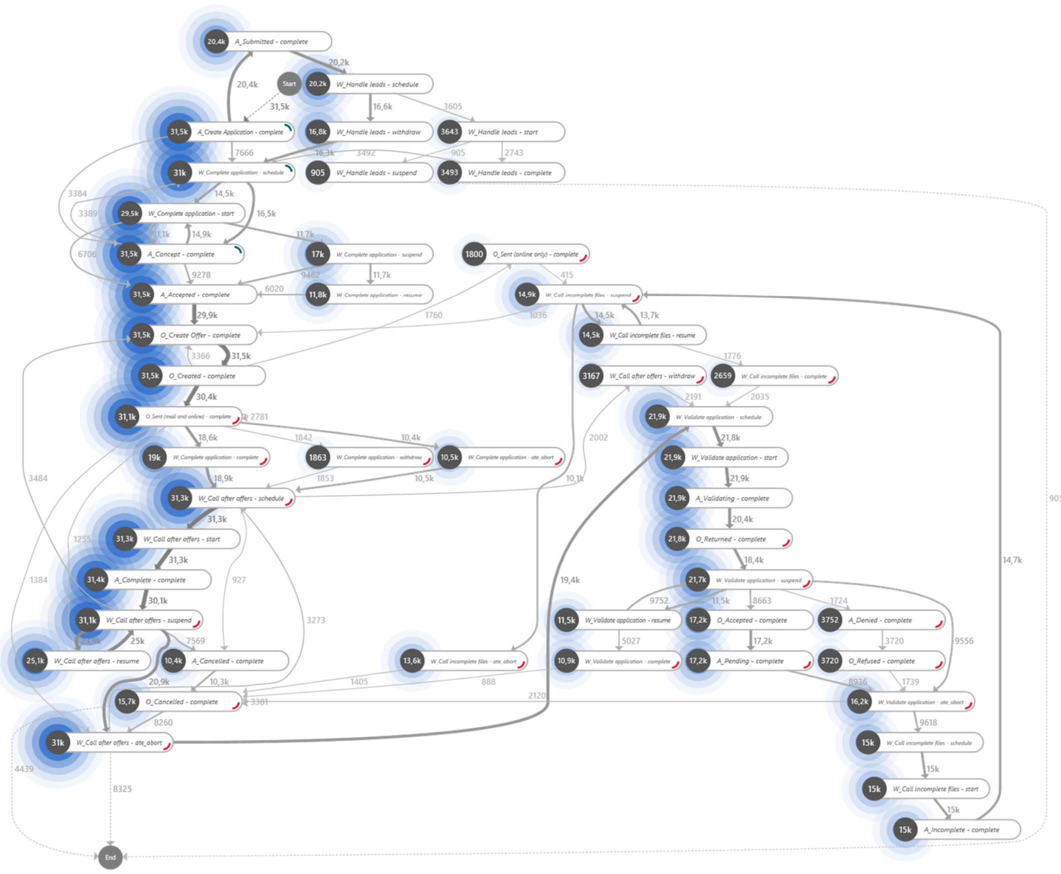

Figure 17 Process in 2016-17

Throughput Time has Increased Since 2012

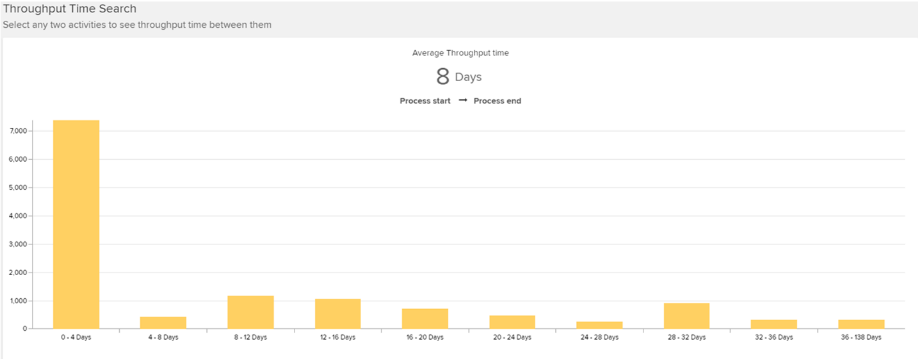

Figure 18 Average throughput time was 8 days in 2012.

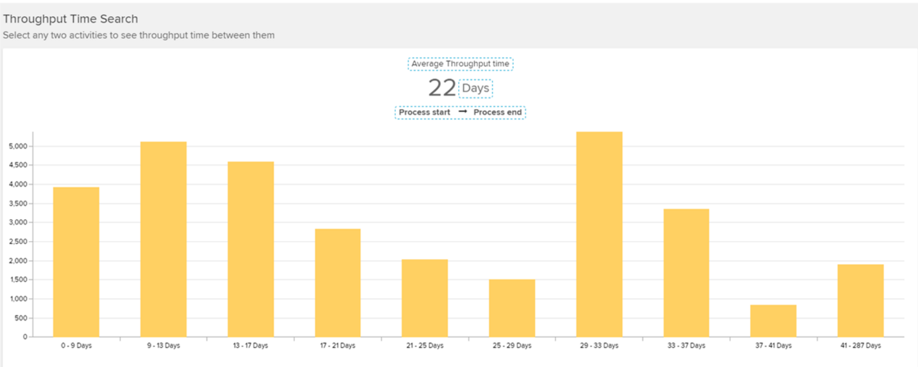

Figure 19 Average throughput time has increased to 22 days.

First stage for new credits takes longer and depends more on timeslot than for limit raises

Table 11. Time from application to first offer (in days) for new credits depends on timeslot.

applicationType	New cre t						
Average time to first offer				Day: Monday=1			
Gemiddelde van TimeToFirstOffer	Koloml						
Time	1	2	3	4	5	6	7
⊞0	1,20	1,71	2,05	3,69	2,79	1,98	3,00
⊞1	2,04	3,80	5,06	1,74	2,61	3,67	2,08
⊞2	2,12	2,30	1,93	2,17	4,08	2,62	2,55
⊞3	1,95	3,50	0,86	2,74	2,67	1,63	3,50
⊞4	1,51	0,43	2,34	2,26	2,44	2,69	2,01
⊞5	1,23	2,00	1,39	1,83	2,51	2,42	3,40
⊞6	1,28	2,21	1,90	1,85	2,31	2,02	1,66
⊞7	0,81	1,17	1,27	1,36	1,91	2,68	2,48
⊞8	0,83	0,95	1,13	1,47	1,37	1,61	2,45
⊞9	1,00	1,26	1,06	1,36	1,62	1,27	1,97
⊞10	1,05	1,34	1,05	1,22	1,45	1,55	2,17
⊞11	1,21	1,18	1,13	1,46	1,74	1,53	2,04
⊞12	1,30	1,13	1,17	1,44	1,69	1,29	2,31
⊞13	1,15	1,16	1,25	1,50	1,63	1,46	2,28
⊞14	1,22	1,18	1,31	1,58	1,79	1,38	2,02
⊞15	1,05	1,29	1,35	1,99	1,81	2,42	2,02
⊞16	1,38	1,42	1,33	1,39	1,95	2,89	2,17
⊞17	1,29	1,09	1,25	1,93	1,76	2,32	1,98
⊞18	1,48	1,44	1,41	2,01	1,85	2,46	1,84
⊞19	1,93	1,77	1,92	2,06	2,44	3,37	1,80
⊞20	2,27	1,91	2,26	2,40	2,40	2,74	2,04
⊞21	2,10	2,14	2,11	2,61	2,48	2,54	2,01
⊞22	2,20	2,69	2,56	2,12	2,75	2,38	3,27
⊞23	1,88	1,69	2,10	2,49	2,19	3,19	2,50
Total	1,29	1,38	1,39	1,69	1,81	1,94	2,12

Table 12. Time from application to first offer (in days) for limit raise is slightly shorter.

applicationType	Limit ra					
Average time to first offer				Day: Monday=1		
Gemiddelde van TimeToFirstOffer	Koloml					
Time	1	2	3	4	5	6
⊞5		1,36	1,06			
⊞6	2,56	1,32	1,60	0,77	2,68	
⊞7	1,34	1,41	1,32	1,43	1,55	
⊞8	1,75	1,00	1,22	1,35	1,76	1,73
⊞9	1,31	1,79	0,80	1,82	1,58	1,25
⊞10	1,05	1,28	0,91	1,90	1,56	1,23
⊞11	1,03	1,68	2,02	0,97	0,38	0,72
⊞12	1,07	1,01	0,74	1,18	1,58	0,83
⊞13	0,87	0,74	1,69	1,54	1,41	1,70
⊞14	1,32	1,21	1,32	1,93	1,63	1,33
⊞15	1,05	0,49	1,06	2,85	0,97	1,22
⊞16	1,74	0,81	0,76	1,65	1,73	1,90
⊞17	0,88	0,88	1,07	1,50	1,99	
⊞18	1,34	1,03	1,52	1,39	2,19	1,98
⊞19	1,53	1,22	1,43	2,40	0,91	
Total	1,31	1,15	1,21	1,61	1,52	1,25

Different Profile of Employees

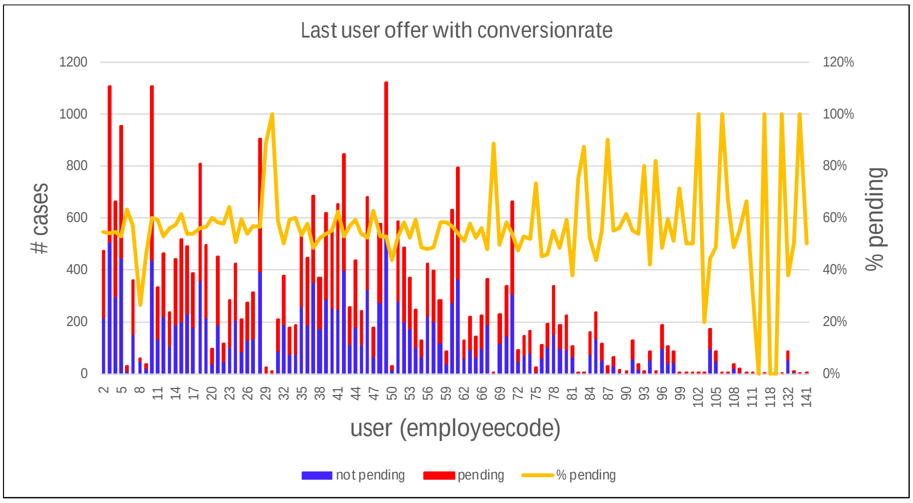

Figure 20 Offers are mainly the work of Users <100.

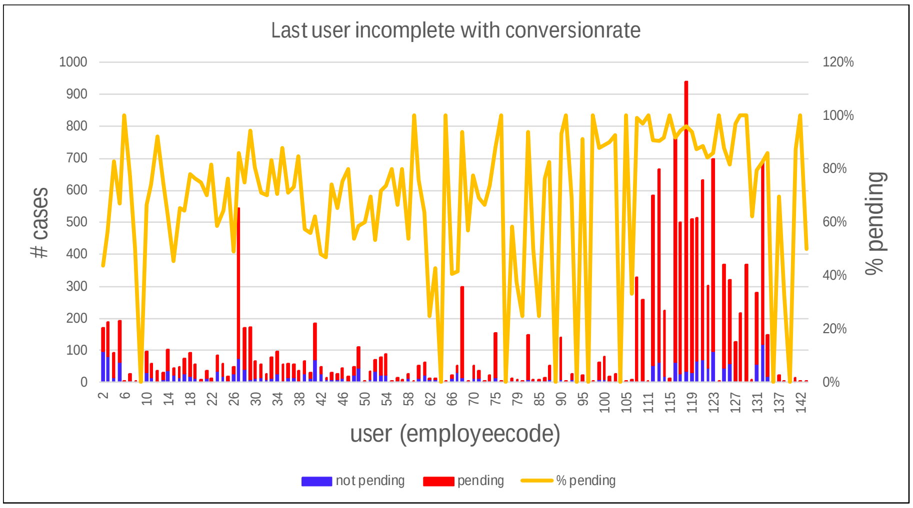

Figure 21 Completion of files is mainly the work of users >100.

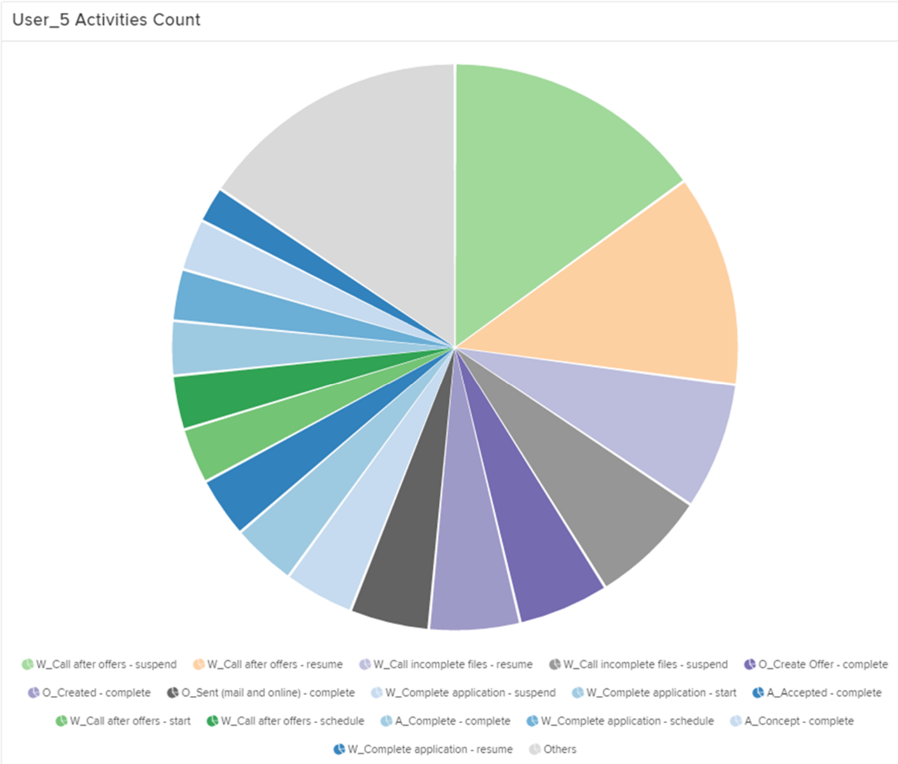

Figure 22 User_5 has many different activities, but main activities are Call after offers suspend and Call after offers resume.

Special offers with low conversion

Table 13. Low conversion for high number of offers from 2 special dates.

Date/User	#Cases last offer	# A_Pending complete	Conversion Rate
28-jan	58	29	50%
29-jan			
10	8	6	75%
11	1	1	100%
12	4	2	50%
14	5	2	40%
2	5	3	60%
27	50	25	50%
28	68	16	24%
34	7	3	43%
36	3	1	33%
38	8	5	63%
4	7	3	43%
41	5	4	80%
49	5	5	100%
5	31	11	35%
70	3	1	33%
73	60	32	53%
8	33	2	6%
85	81	17	21%
87	1	1	100%
Totaal 29-jan	385	140	36%
30-jan	37	22	59%
27-mei	40	17	43%
28-mei			
17	22	11	50%
35	34	10	29%
37	8	5	63%
39	13	9	69%
5	125	28	22%
52	9	6	67%
71	80	49	61%
76	57	23	40%
Totaal 28-mei	348	141	41%

Manual cancellations in a non-standard procedure

Table 14 Different pattern of cases that were cancelled in the normal procedure at 6:00 versus cases that were cancelled in the evening. The batch in the morning cancels 11 cases in only 18 seconds. In the evening 154 cases were cancelled in 78 minutes. Only part of the evening cancellations is listed here.

	Cancelled 27-6-2016
Timestamp.2	Case-concept:name
6:00:14	Application_796878514
6:00:15	Application_611961345
6:00:20	Application_1152955170
6:00:24	Application_1280427249
6:00:25	Application_2071699900
6:00:26	Application_769436216
6:00:27	Application_1496592829
6:00:30	Application_2001914382
6:00:31	Application_503892757
	Application_769456413
6:00:32	Application_1648426839
21:12:10	Application_1718377889
21:12:25	Application_614598233
21:12:39	Application_1977245249
21:12:54	Application_1563442669
21:13:08	Application_1455219594
21:13:21	Application_1354242264
21:13:36	Application_1265495164
21:13:50	Application_724736325
21:14:04	Application_76654009
21:16:07	Application_476277757
21:16:40	Application_1532233038
21:17:16	Application_46651536
21:17:52	Application_1470446509
21:18:15	Application_1538131961
21:18:39	Application_1985400245
21:22:11	Application_795283356

Comparing the process from 2017 vs 2012

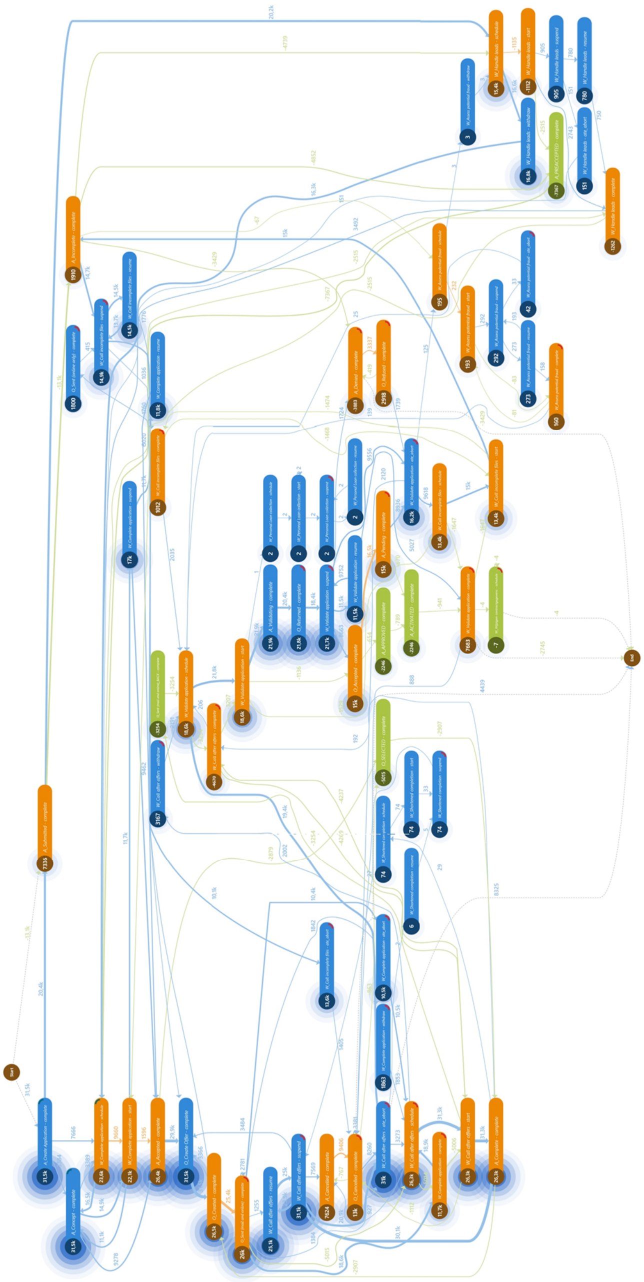

Figure 23 Comparing process from 2017 (blue) 2012 (green) and both years (orange). More steps in 2017.