# Applying Process Mining for Loan Approvals in a Banking Institution

Andrés Carvallo, Cristóbal Henning, Dasen Razmilic, Regina Reyes López, Jonathan Lee, Juan P. Salazar Fernández, and Michael Arias

Computer Science Department, School of Engineering
Pontificia Universidad Católica de Chile, Santiago, Chile
[afcarvallo, cahenning, dirazmilic, rreyes4, wllee, jpsalaza, m.arias]@uc.cl

**Abstract.** The BPI Challenge 2017 provides a real-life event log composed by loan applications and offers, generated by a bank to analyze the data and improve their processes. This paper analyzes the throughput times 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 for an input from the applicant. Moreover, we evaluated the influence of the frequency of incompleteness on the final outcome and if the quantity of offers requested by the customer matters. Other interesting trends are analyzed, such as efficient use of resources, business rules compliance and identification of behavioral patterns at different times of the day. Results show that a lack of customer requests for completion does not improve the credit approval rate and that this rate decreases when there are more offers. Also, the more the users engage in the case, the greater the approval rate, but the throughput times get longer as well.

**Keywords:** process mining, process analysis, loan applications, behavioral patterns, business rules

## 1 Introduction

Companies are aware that better information management translates into increased productivity. With this objective in mind, tools and solutions are constantly being developed to allow them to speed up the management of data generated everyday, allowing a better management of time and a greater use of the resources [1, 3, 6]. Process Mining can be seen as the bridge between data science and process science [4], where knowledge could be extracted from the records of events associated with a process. Although there are many available tools that support the process analysis, a detailed analysis would help to understand a prior implemented process, which could contribribute to reduce inefficiencies and improve the process performance.

The BPI Challenge 2017[1] is an opportunity to apply different tools and algorithms to a real life process. This year, the challenge is focused on a banking institution, which is particularly interested in answers to the following questions:

[1] https://www.win.tue.nl/bpi/doku.php?id=2017:challenge

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.
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 not to 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.

The aim of this paper is to address the proposed questions and give detailed feedback of the process, in order to provide another point of view of the process to the bank. This might prove useful if they want to make improvements in their internal processes, improve the allocation of resources, or even implement new communication strategies inside and outside the organization.

The paper is structured as follows. Section 2 provides a brief overview of the data description and filtering steps. Section 3 addresses the preliminary diagnosis of the process through the data. The throughput times area analyzed in Section 4. The influence on the frequency of incompleteness to the final outcome are presented in Section 5. Section 6 shows the impact of multiple offers regarding loan applications. Section 7 describes the influence of the human factor in the throughput times and application outcomes. Section 8 gives a list of recommended business rules for the process and checks their compliance. Recommendations to improve the process are given in Section 9. Finally, Section 10 presents the conclusions.

## 2 Log analysis, process familiarization and data filtering

The data is composed of offer applications and offers available in an event log comprised of a total of 1,202,267 events, with 31,509 loan applications and 42,995 offers over a period from 2016 to February 2017. Also, the contest gives additional data like the *requested loan amount*, *application type*, reason the loan was applied for, *offered amount*, *initial withdrawal amount*, number of payback *terms* agreed to, *monthly costs*, *credit score* of the customer, the employee who created the offer, whether the offer was selected and whether the offer was accepted by the customer.

After a preliminary analysis it was found that each event passes through different stages. In the case of applications and offers, they only transit through the *complete* state, while workitems transit through seven states: *schedule*, *start*, *suspend*, *resume*, *withdraw*, *complete* and *ate_abort*. Figure 1 helps to illustrate the transitions between states.

The data contains three types of events, namely *Application state changes*, *Offer state changes* and *Workflow events*. Table 1 shows the event types and the

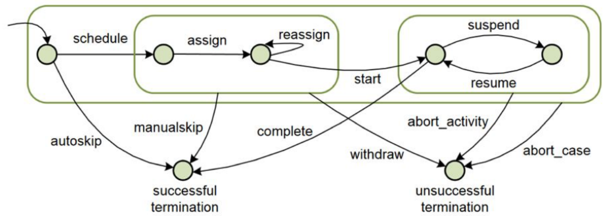



**Fig. 1:** Transition of activitys states [2].

personal of the authors, also considering the ideas provided by domain experts through the available forum about the topic of the challenge.

After analyzing the information and reviewing the clarifications made by experts domain, we elaborated a process diagram with the aim of achieving a better understanding of the process. The diagram is shown in Figure 2.

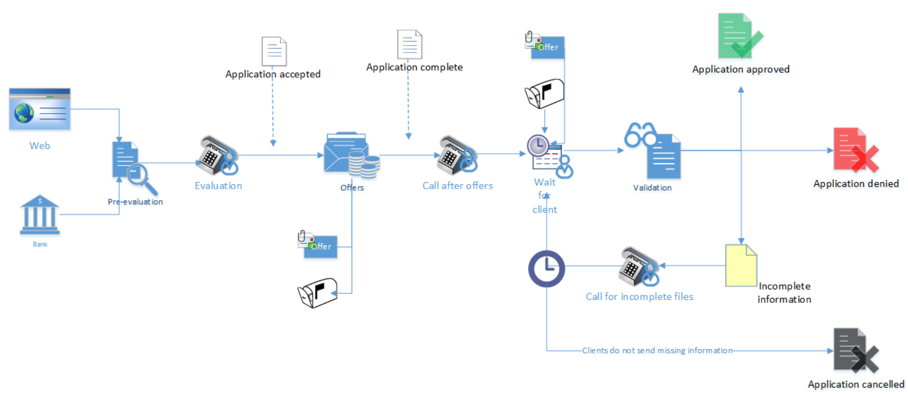



**Fig. 2:** General flow diagram of the process.

The first step afterwards was to analyze the log. Inicially, the given log was really complex and difficult to visualize. There was a lot of noise, activities that were redundant with others and even some incomplete cases. Therefore, in order to simplify the data without losing crucial information, the log was preprocessed using a custom made Python script that removed duplicated activities (activities that occurred at the same time), for example as ocurred with *O_Created* and *O_Create Offer*.

As a second step, three endpoints were identified (see Figure 2). These endpoints are: *A_Cancelled*, *A_Denied* and *A_Pending*. As a result, the Python script was also modified to remove all events following these endpoints. The removed activities were mainly associated with the automated closing of other activities, so removing them did not affect the process. Finally, when filtering the log, 98

**Table 1:** Description of events

| Event type | Description |
|---|---|
| **Application (A)** | Refers to the application itself. Indicates the application's status.<br><br>**Initial steps:**<br>**Create application:** Client or employee initiates an application.<br>**Submitted:** Client submits an application through the website.<br>**Concept:** Application gets an early evaluation.<br>**Accepted:** Application is accepted after evaluations.<br><br>**Development steps:**<br>**Complete:** Application is complete and ready to be offered.<br>**Incomplete:** Application lacks or has erroneous information.<br>**Validating:** After an offer is chosen, the application is reevaluated rigorously.<br><br>**Final steps:**<br>**Pending:** Loan is approved.<br>**Denied:** Application does not meet the acceptance criteria and is rejected.<br>**Cancelled:** The client cancels the application or does not send the needed information. |
| **Offer (O)** | Refers to the offers of an application. Indicates the offer's status.<br><br>**Initial steps:**<br>**Created offer** & **Created:** Offer is created. They are consecutive dependent events.<br><br>**Development steps:**<br>**Sent:** Offer is sent by a communication channel.<br>**Returned:** Offer is selected by the client and they have sent the information back to the institution.<br><br>**Final steps:**<br>**Accepted:** Offer is chosen by the client and the bank evaluated the approval criteria satisfactorily.<br>**Refused:** Application has been rejected by the bank, so any offer created is rejected.<br>**Cancelled:** Offer is cancelled if the application is cancelled or approved but with another offer. |
| **Work flow (W)** | Refers to the manual work that employees must perform, whether through calls, research, evaluation or others.<br><br>**Complete application:** Employee calls the customer to complete the application after it has been evaluated.<br>**Validate application:** Employee validates the information given for the application.<br>**Call:** Employee calls the customer to remind offers have been sent or to ask for missing information.<br>**Assess potential fraud:** Employee investigates an application suspicious of fraud. |

incomplete cases were identified. After analyzing their dates, they were classified as applications, that at the time of log extraction had not yet reached an endpoint, so they were also removed. After applying the aforementioned steps, a new log was generated, composed by 538,071 events, with 31,411 cases and 4,036 execution variants. All further analysis is done using the preprocessed log, unless stated otherwise.

## 3 Process analysis: general overview

Our next step was discovering the process. To do that, we used Celonis[2] tool. The obtained process graph is shown in Figure 3. It can be seen that loan applications are sent to the bank through the web page (*A_Submitted*) or directly from the bank offices. Then, there is a pre evaluation of the loan (*A_Concept*) and a call to make the final evaluation (*W_Complete application*). When the application reaches this point it is on an accepted status (*A_Accepted*). Then, the client asks for at least one offer. If it is possible, the bank makes one or more offers (*O_Create offer*). After that, the customer is called to remind him of the offers, and from this point the application reaches the complete status (*A_Complete*). Then, the process enters a second stage. In this part, the institution has to wait for the clients response, if the client sends the information needed, the application passes through different states (*A_Validating*, *A_Incomplete*, *A_Denied*), which involve some manual tasks like *W_Validate application* and *W_Call incomplete files*. After this, the bank validates the application again and asks the client for more background for the credit loan, if it is necessary. Finally, if everything goes well, the application reach the status approved (*A_Pending*). If the client does not make contact after the application is complete (*A_Complete*), the final result is a cancelled application (*A_Cancelled*).

It should be noted that some activities of the process are not shown in the process graph (Figure 3), such as: *W_Shortened completion* (occurs 74 times), *W_Asses potential fraud* (occurs 301 times) and *W_Personal Loan collection* (occurs 2 times). These activities are not analyzed here, due to their low frequency and because they are activities that are characteristic from the bank sector and every institution has its own internal rules to handle. Also denied applications (those that pass through *A_Denied* are not deeply analyzed because the bank has its own policy to deny loans, aiming to protect their business interests.

### 3.1 Preliminary diagnosis of the process

In order to obtain interesting findings, an exploratory analysis of the process was done. This analysis consisted in getting the amount of cases that ended with each of the endpoints defined in Section 2. Afterwards, a filter was applied to separate cases submitted by the client through the web application and cases submitted

[2] http://www.celonis.com/

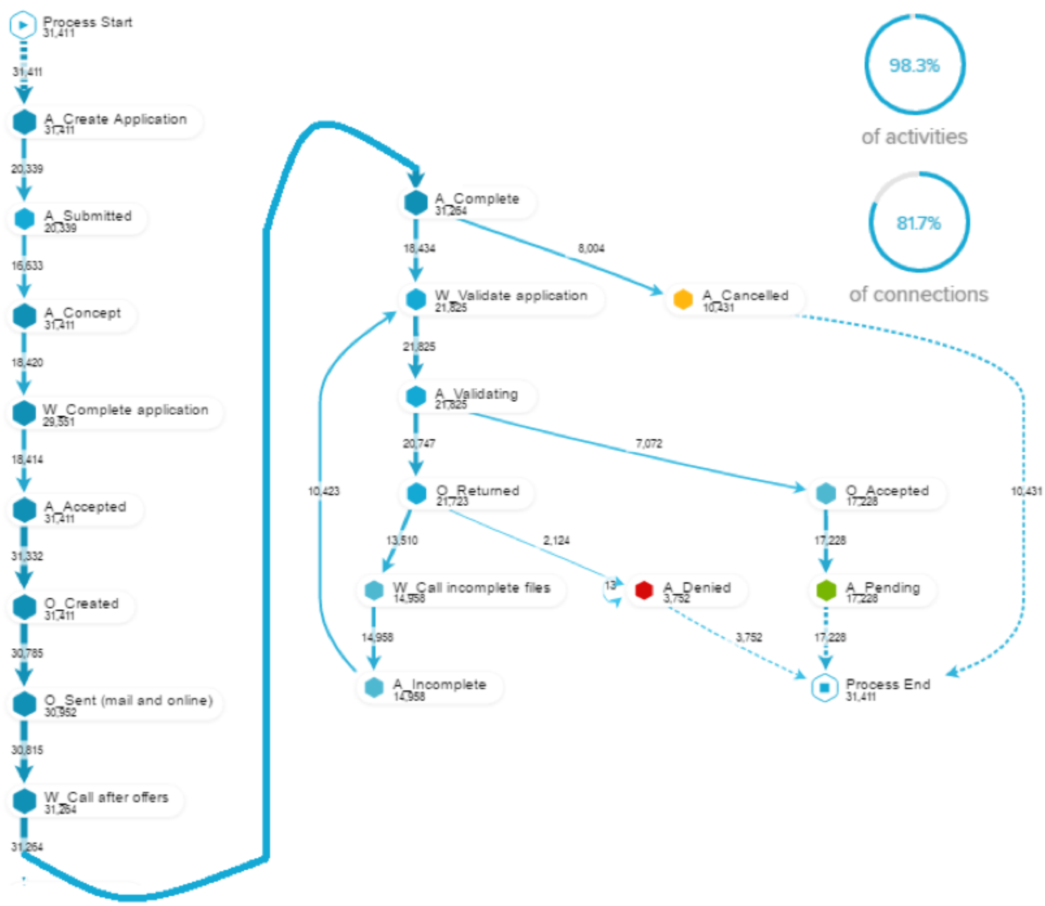



**Fig. 3:** General process graph of the offer complete process using Celonis.

by the bank, the first of these cases included the event *A_Submitted*. The results are displayed in Table 2. It can be seen that there is a high cancellation rate, 10,431 cases of 31,411 are cancelled, that is 33.2% of applications. Also, globally 54.8% of cases end up being approved, but an interesting part of the results is to compare Web applications with bank applications. Also, we can see that approval rate is greater for bank applications (64.7% vs 49.5%), whereas cancellation rate is greater in web applications (37.2% vs 25.8%).

**Table 2:** Global statistics of the process

| | **Global** | **%** | **Web** | **%** | **Bank** | **%** |
|---|---|---|---|---|---|---|
| **Approved** | 17,228 | 54.8 | 10,064 | 49.5 | 7,164 | 64.7 |
| **Cancelled** | 10,431 | 33.2 | 7,573 | 37.2 | 2,858 | 25.8 |
| **Denied** | 3,752 | 12 | 2,702 | 13.3 | 1,050 | 9.5 |
| **Total** | 31,411 | | 20,339 | | 11,072 | |

In addition, the most frequent variations are analyzed using the Trace Alignment plugin available in ProM tool[3] (version 6.6.2). Figure 4 shows that the most frequent trace alignment with 11.44% of traces and starts with *A_Create application*, *A_Submitted*, *A_Concept*, *W_Complete application* and then if the application is accepted the offer is created and the email is sent to the client. After that, the bank tries to contact the customer through a phone call. Finally, the application is completed and in most cases the process finishes cancelled.


**Fig. 4:** Trace alignment of the process using ProM 6.6.2.

## 4 Throughput times

This section addresses the first question proposed in the challenge: 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 for input from the applicant?

[3] http://www.promtools.org/

In order to analyze the applications processing times, we decided to analyze them according to their endpoints. To do that, we used the Disco[4] tool. The applications were grouped in two types using a filter: the satisfactory, all those applications that were approved (ended up with *A_Pending*), and unsatisfactory, those that have been denied or cancelled (ended up with *A_Denied* or *A_Cancelled*).

As shown in Figure 5, the average time taken for an application to be approved is 18.1 days. It was observed that one of the most demanding activities is the *analysis of fraud*. However, this happens only in 104 occasions, thus it was not considered a bottleneck. However, when analyzing the waiting times related to the client, it was observed that these are the ones that lengthen the processing time. In approved cases, it can be seen that clients take 8.7 days (connection between *A_Complete* and *W_Complete application*) on average to send the necessary information for the bank to validate the application. It takes as well 57.6 hours (connection between *A_Incomplete* and *W_Validate application*) on average for clients to send the pending information, if the application is tagged as incomplete.

After getting these results a deeper analysis was done to identify whether or not multiple offers influence processing times. To accomplish this, two filters were applied using Celonis. The first filter was used to separate cases with throughput times above the average of 18 days from those with throughput times below that average. The second filter was used to analyze the amount of offers created in each case. When comparing approved cases that are above the average of 18 days, (see Figure 6), an interesting finding was identified: applications that take the longest time (see Figure 7) tend to have more offers than those that are below the average, as shown in Figure 8.

After analyzing cancelled applications, it was observed (see Figure 9) that the average time is 29.9 days. Additionally, in 76.74% of cases it takes 27.4 days (connection between *A_Complete* and *A_Cancelled)* on average to issue the cancellation status, as noticed in Figure 9. This is mainly because customers do not send any response about the offers. Also it can be seen that there are other activities that consume a lot of time, like *fraud analysis*, but the frequency of occurrence of these is very low (104 occurrences), reason why they are not considered as critical factors. It is recommended though not to neglect these activities if the times or frequencies change.

Although it has been pointed out that the biggest proportion of processing time is due to waiting for client's responses, it was decided to analyze the activity *W_Validate application* as it was noticed that it was a bottleneck both in approved cases and cancelled ones. Figure 5 illustrates that *W_Validate application* takes on average 14.2 hours for approved applications whereas Figure 9 shows that the same activity takes 29.4 hours on average for cancelled applica-

[4] https://fluxicon.com/disco/

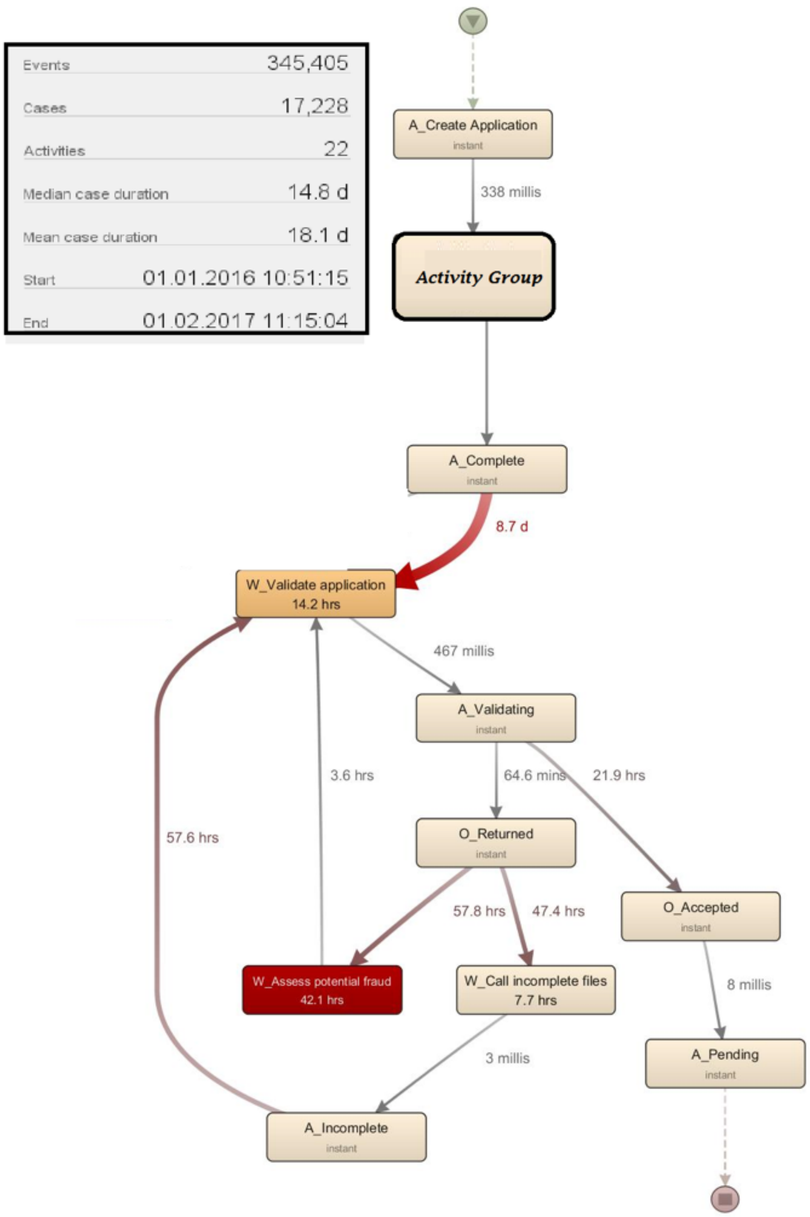



**Fig. 5:** Statistics and process flow for approved applications using Disco.

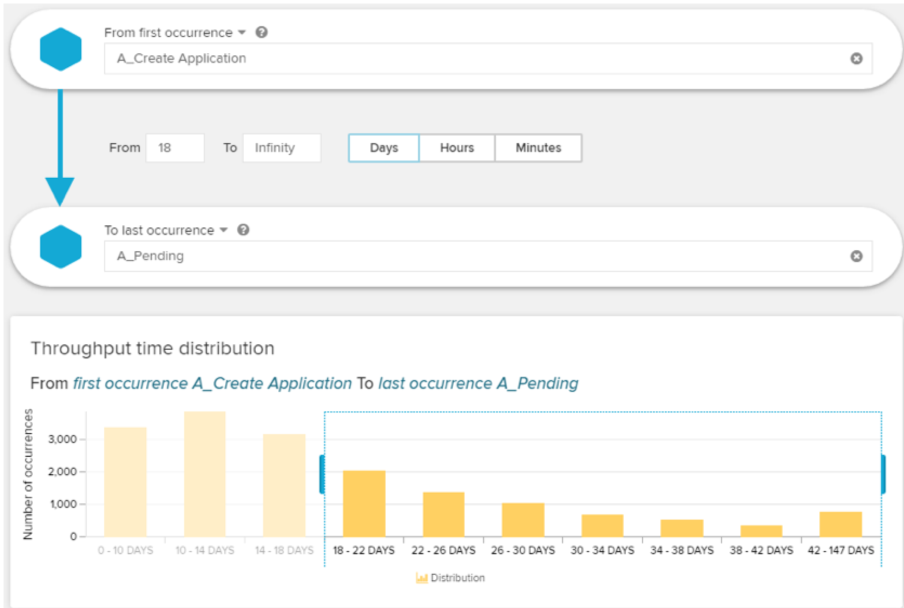



**Fig. 6:** Filter of applications that take more than 18 days using Celonis.

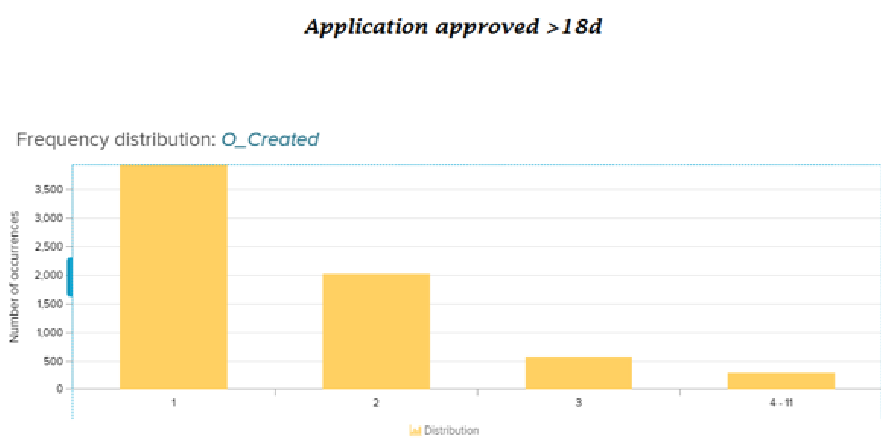



**Fig. 7:** Frequency of offers in applications over the average throughput time.

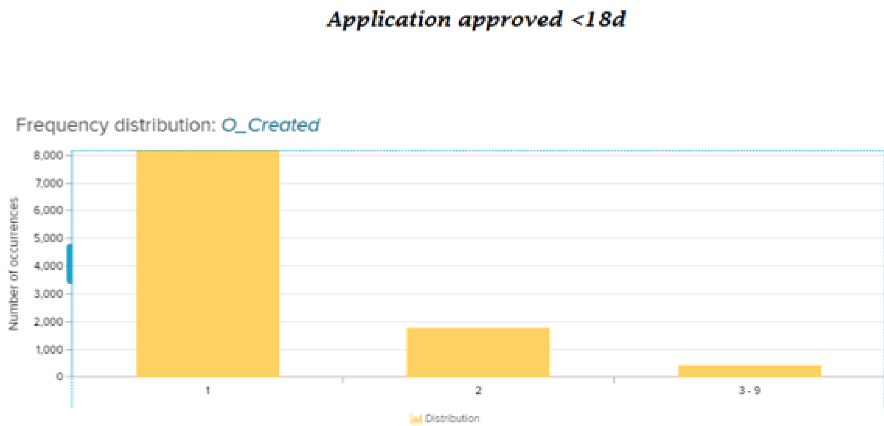



**Fig. 8:** Frequency of offers in applications below the average throughput time.

tions. Also, the activity *validation* of the application requires contact with the client and waiting for a response to obtain more background information to validate the correct delivery of the credit in process. These take an average of 8.7 days (connection between *A_Complete* and *W_Validate application)* to receive an answer from the client. This shows that this particular part of the process is definitely inefficient. In order to reduce these waiting times and improve the process efficiency, it is recommended to improve communication with the client. The background information should be requested when the credit application begins because most of the time it is returned for a new validation. According to internal processes, there are really efficient tasks, like for example *submitting loan applications*, *offer creation*, and *internal validation*.

Additional analysis on the *W_Validate application* activity presents that this activity comes from another validation in 97% of the cases and is performed mostly between 8:00 am and 2:00 pm, as shown in Figure 10. Perhaps, the schedule for requesting further background should be in the afternoon rather than between 8:00 am and 2:00 pm. This may be a good strategy to increase the customer's response rate. One alternative is adding more validation activities between 4:00 pm and 6:00 pm. As it can be seen in Figure 11, there are two loops of activities. One of them is *calling after credit offers* that needs the approval of the client to continue the process. The other is the one that was mentioned before related to the validating and return activities. Most of the time, the calls for incomplete files come and go to the same activity. This means that there is no success in asking for a deeper customer background. These activities have a similar behavior in the schedules of realization.

In summary, customers are the ones who contribute the most to the processing time of applications, mainly in activities that involve sending information and taking a decision about an offer.

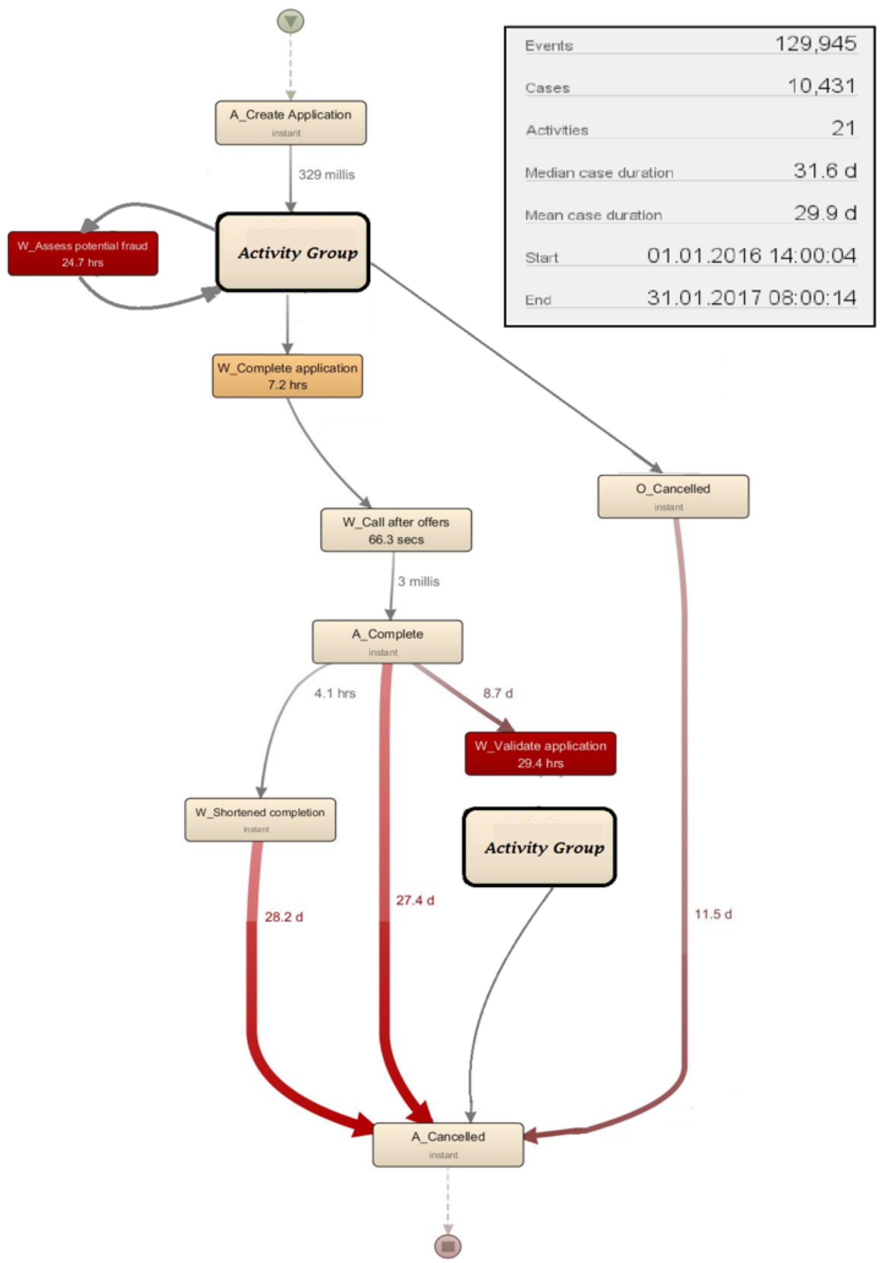



**Fig. 9:** Throughput time of cancelled applications.

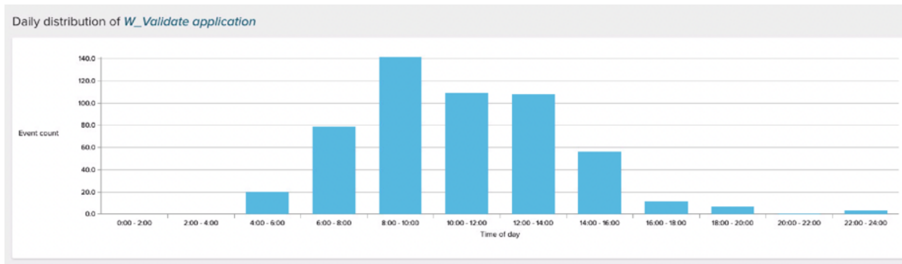



**Fig. 10:** *W_Validate application* times by hours of day.

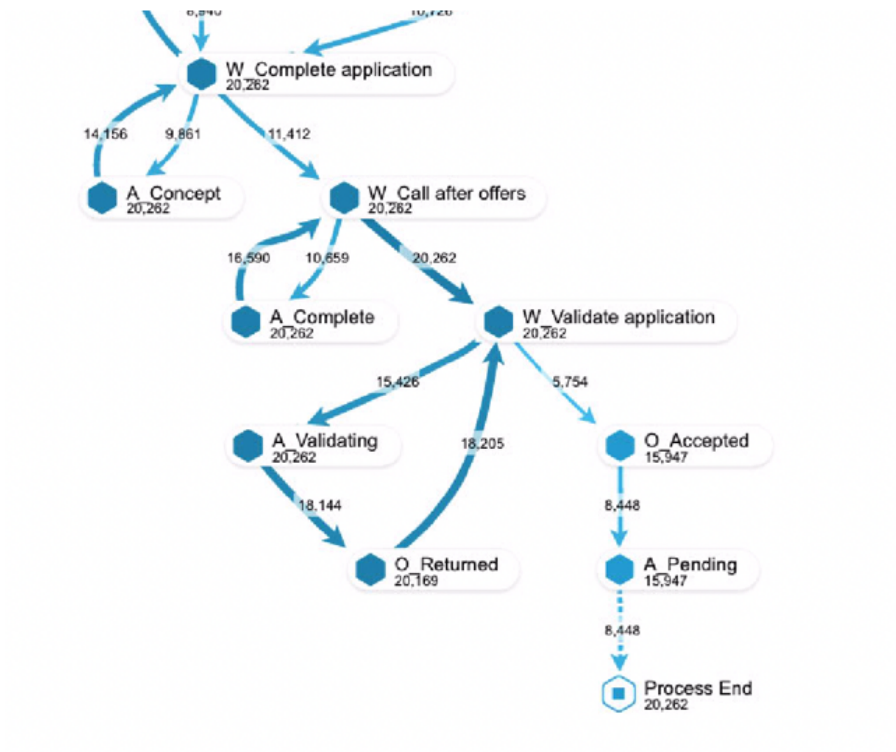



**Fig. 11:** *A_Complete* returned after *W_Call after offers* and *A_Validating* to *O_Returned* loop using Celonis.

## 5 Influence on the frequency of incompleteness to the final outcome

This section addresses the following question: What is the influence on the frequency of incompleteness to the final outcome? The hypothesis presented by the bank is that if applicants are confronted with more requests for completion, they are more likely not to accept the final offer. To answer the question, frequency of incompleteness was defined as the amount of *A_Incomplete* activities in each case. The first approach was filtering the log by the endpoints defined in Section 2, and by presence of incompleteness (at least 1). After filtering cases by the three possible outcomes and the presence or absence of incompleteness the bank's hypothesis was dismissed.

The results of the previous analysis that dismissed the bank's hypothesis are displayed in Table 3. There it can be noticed that the percentage of cases ending cancelled (dropped by the client) is greater than the percentage of cases which ended in *A_Pending* when there is no incompleteness, whereas in cases where there is presence of incompleteness, there are more cases ending with *A_Pending* than cancelled.

**Table 3:** Global statistics of complete and incomplete cases

| | **Incomplete** | % | **Complete** | % |
|---|---|---|---|---|
| **A_Pending** | 12,647 | 84.5 | 4,581 | 27.8 |
| **A_Cancelled** | 955 | 6.3 | 9,476 | 57.6 |
| **A_Denied** | 1,356 | 9.2 | 2,396 | 14.6 |
| **Total** | 14,958 | | 16,453 | |

After analyzing the influence of the presence of incompleteness on the final outcome, it was necessary to analyze the influence of incompleteness frequency. To achieve this, a second filter was applied to the log using Disco. This time, cases were filtered by the amount of *A_Incomplete* they had (multiple occurences mean the application remains incomplete after additional information was given) and their outcome.

Results of this analysis are shown in Figure 12, where it can be seen that the amount of cases which end with *A_Pending* decreases as the frequency of *A_Incomplete* increases (in cases where there is presence of incompleteness). An early conclusion to explain this might be that clients get tired of having to fetch extra documents and send them. But the reduction of approved cases can also be explained by the fact that cases with more *A_Incomplete* are less likely to be found than cases with less of this activity. Cases with four or more *A_Incomplete* are actually pretty rare.

In order to perform a deeper analysis, for each frequency of *A_Incomplete* the percentage of each outcome was calculated regarding the total amount of cases for that frequency and interesting conclusions were drawn. As can be seen in Figure 13, the percentage of *A_Pending* increases as the amount of *A_Incomplete* increases, which is quite the opposite to the hypothesis presented by the challenge. It is also posible to see that the proportion of cases which end in *A_Cancelled* has its maximum where there are no *A_Incomplete* activities, and drops down as the frequency of incompleteness increases. From this last observation it can be concluded that clients do not get tired when the bank asks them to bring in missing documents, or at least it does not affect the outcome of the case. It can also be said that clients actually get more engaged with finishing the process (not dropping the application) as the frequency of incompleteness rises, as if they had already invested too much effort on the application to drop out of the process.

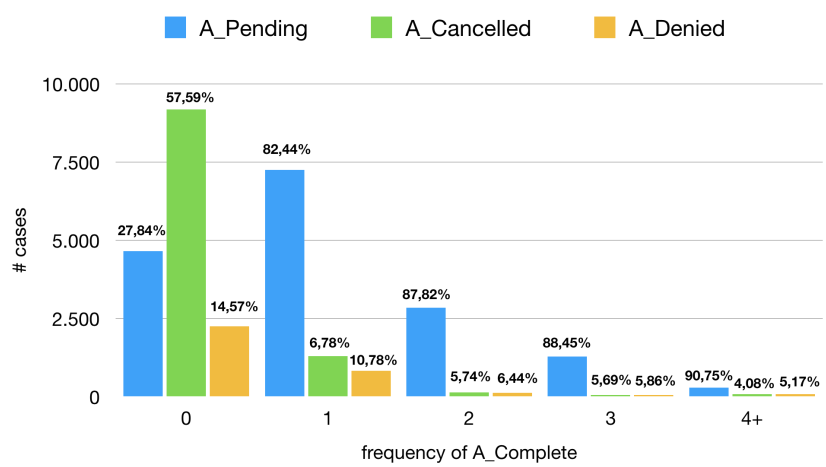



**Fig. 12:** Number of cases of each outcome for each frequency of incompleteness.

## 6 Impact of multiple offers in an application

The last questions proposed by the challenge were: 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?

To analyze the offers, the behavior of the different offers for each application was identified. The offer log was used to apply a rework filter using Celonis. Figure 14 shows how offers vary from 1 to 10 per case. In addition, the relationship between the number of offers and the final result of the process was identified. These relations can be seen in Figure 15. According to the obtained results, when there is a bigger amount of offers, clients tend to accept less offers and it takes them more time to decide. The decision time was measured from the moment the offer is created until the client returns the signed offer.

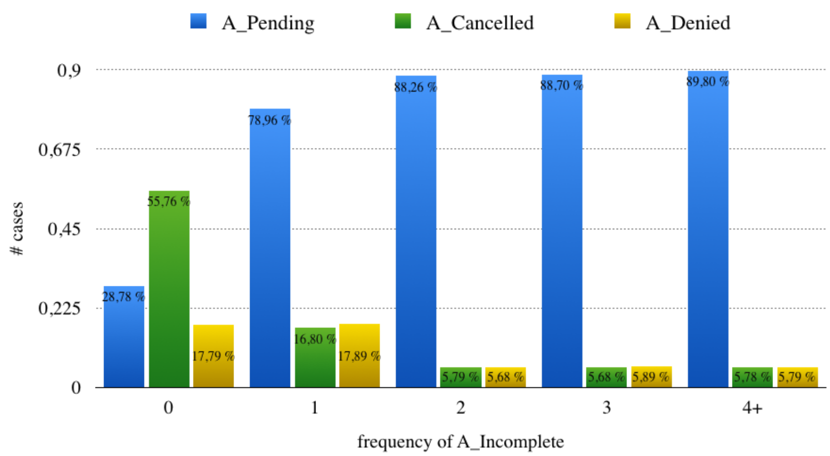



**Fig. 13:** Proportion of each outcome for each incompleteness frequency.

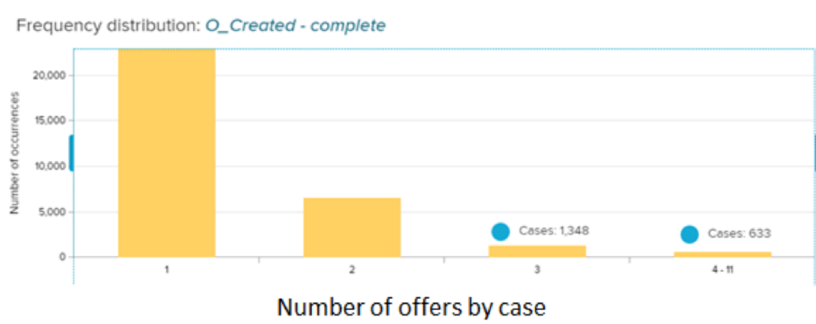



**Fig. 14:** Frequency of offers by applications.

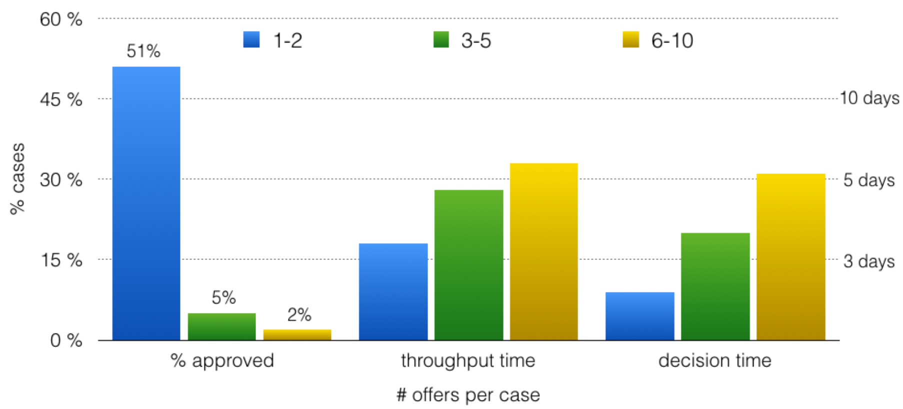



**Fig. 15:** Multiple offers and throughput time (number of offers vs approved applications).

To check if there is any effect on the offers where they are requested during a single or multiple conversations, applications of multiple offers were divided in two groups: customers who wait a week to request another offer and those who take more than a week. For this analysis, two filters using Celonis were applied, *Rework* and *Throughput time selection*. Figure 16 shows there is apparently no relationship between the acceptance of the offer and the time window when they are requested. It can be seen that applications with multiple offers made in a rapid succession have a rate of approval of 53.9% (of 8,280 total cases) while the other group has a rate of 66.7% (of 5,047 total cases). This means that raising or limiting the number of calls per client is not needed, but instead is advisable to contact them earlier. More details about this phenomenon will be explained in Section 7.

## 7 Human factor on process performance

There is an interest to characterize the influence of human factor on process performance. Human factor is associated with high variability in the execution of the process [5]. Given the characteristics of the loan application process analyzed, two human roles can be identified: clients and bank employees. Thus, checking the influence of these roles on the process throughput time and final outcome would be an interesting analysis. This section is divided in three parts. The first part addresses the influence of clients' behavior on the process. The second one, the influence of employees. Finally, the third part gives an organizational analysis of resources.

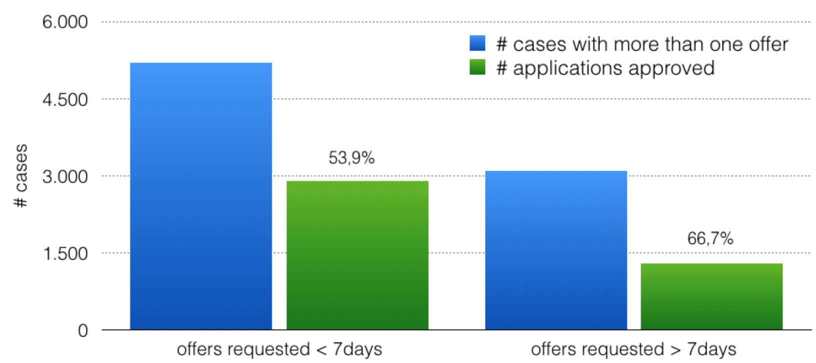



**Fig. 16:** Offers relation with calls.

### 7.1 Clients behavior influence

One of the results obtained in Section 4 shows that a major proportion of customers abandoned the application after an offer was sent to them. For this reason, behavior patterns associated with customer response were analyzed. First, cancelled applications were analyzed, and it was observed that about 90% of them were cancelled after 30 days, because once the offers have been sent and the application is completed, the client did not answer. Next, approved applications were analyzed and a simple but crucial observation was found: in these applications clients have to contact back the bank at some point. A question that surged after that was: how long does it take for clients who are really interested in finishing the application process, to contact back the bank after the offers are sent?

In order to perform this analysis, we used Celonis to apply the defined filters. The first of them was used to leave only the cases that had the *W_Validate application* activity, because the presence of this activity means that the client contacted back the bank at least once. The next filter applied was *throughput time between activities*, between the activity *A_Complete* and the first occurrence of *W_Validate application*, to determine response times of clients.

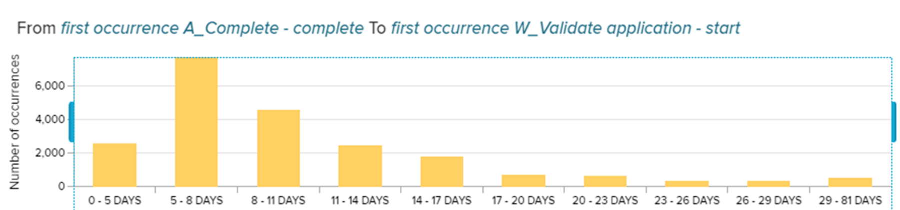



**Fig. 17:** Clients' decision time.

Results of the previous analysis are shown in Figure 17. It can be seen how in most of the cases where clients gives an answer, they do it within 15 days after the offers were sent. Summarizing, 18,761 out of 21,825 applications (85%) were answered within those 15 days, with an average response time of 7 days. It can be said that clients who do not contact back the bank within this time are likely not to finish the process and less bank resources should be allocated to those cases.

### 7.2 Resources influence

Considering that there is a large number of resources involved in the process and the human factor is decisive in approaching the client, the influence of resources is analyzed. With this goal in mind, two analysis were done. The first to determine the resources' influence in the throughput time and the second one to determine the influence in the outcome of the process.

To perform this analysis, a Celonis filter was applied to leave only applications that ended up being *approved*. Then a *throughput time* filter was applied to group cases that were above the average time and those that were below it. Afterwards, to get how many resources intervene in each case, a social analysis was done with Celonis, which has a tool fitting for this purpose. The results are shown in Figure 18, where it can be observed that slow cases tend to pass through a range between 10 and 15 users with an average of 10. While in faster cases they are manipulated by a range between 7 and 8 resources.

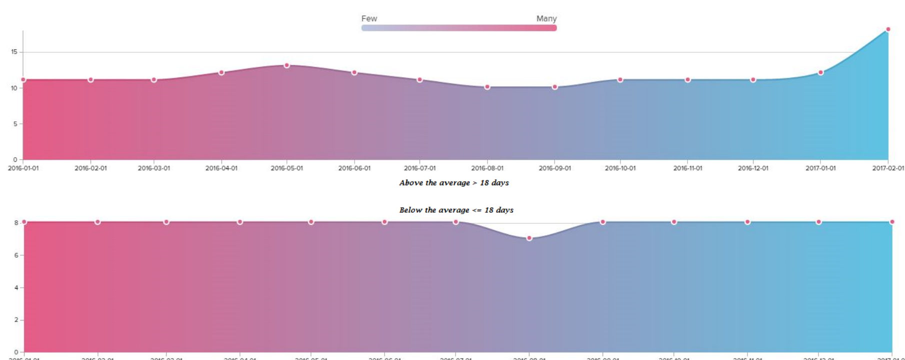



**Fig. 18:** Development users per case for approved applications

To carry out the second analysis, cases were separated in two groups with a filter: *approved* and *cancelled* applications. Then, since the majority of approved applications pass through *validation* and *incompleteness* activities and cancelled applications do not pass through these activities, we took only activities from

*A_Create application* to *A_Complete*. Then, for each group we separated the applications according to the endpoint *A_Pending* and *A_Cancelled*, and a social analysis was performed. Results are shown in Figure 19, which shows that cancelled cases tend to have a lower diversity of resources with a number of 5, while those approved have an average of 9.

After considering the results of these two analysis, it can be inferred that there is a trade-off between the processing time and the success of the cases, in terms of resource influence. Successful cases have a greater participation rate of users than those that were cancelled. Also, it is normal that the processing time increases when more users intervene. After all, the transfer of the case from one user to another implies higher response times due to work handover and queued time.

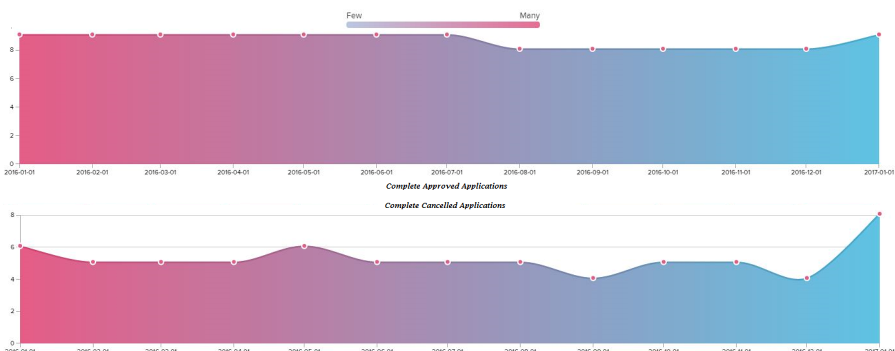



**Fig. 19:** Development users per case for complete applications

### 7.3 Organizational analysis of resources

As shown in Subsections 7.1 and 7.2, resources do influence the throughput time of the process and the final outcome. Thus, we decided to analyze the job distribution and performance analysis. To accomplish the job distribution analysis, Celonis's social analysis was used. It was sought to get the behavior of resources inside the bank and how the different activities were distributed between the resources (see Figure 20). The first conclusion reached was that *User 1* was the system, where all credit requests and loan results are entered and passed to the resources, because they had a huge proportion of the work and at unusual times, outside office hours.

In addition, the resources are grouped by the activities that were executed by them in the process. To achieve this, ProM's Mine of Handover of Social Network plugin was used. As a result (see Figure 21), five clusters were found, which are: the system, fraud cluster, application cluster, offer cluster, and the

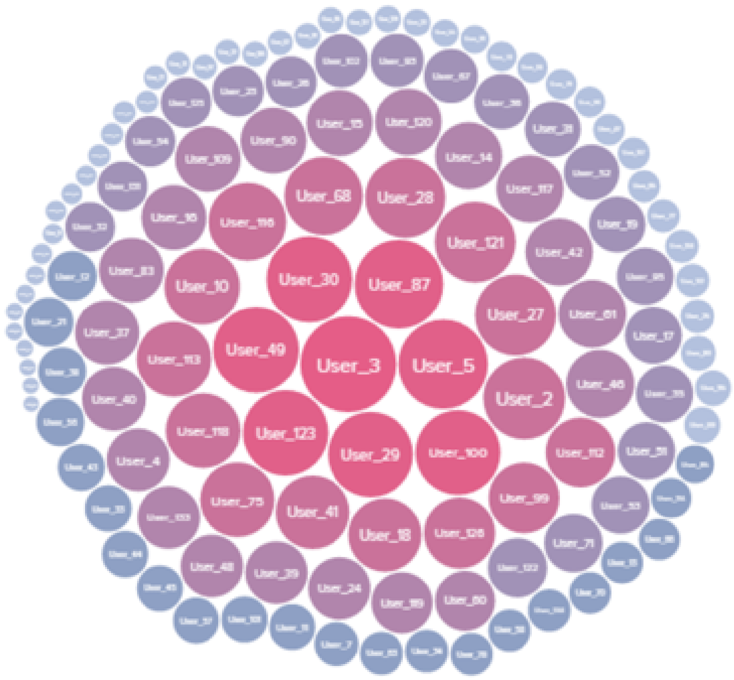



**Fig. 20:** Resource distribution of activities without considering *User 1* using Celonis.

user that connects the offer and application clusters. The labels give a general description of the type of activities they perform.

Furthermore, the performance of the resources according to the identified roles was analyzed. Taking into account activities involving manual labor, three activities that are time-consuming were identified: *W_Validate application*, *W_Call incomplete files* and *W_Assess potential frauds*, noting that the latter was not taken in consideration. Using Disco, activities were filtered in an exclusive way, creating 2 sublogs, which only contained the selected activities. Then, Celonis was used to perform the social analysis for each sublog (*W_Validate application* in Figure 22a, and *W_Call incomplete files* in Figure 22b. Three limiting resources were identified (*Users 99, 68 and 30*), since they used a greater amount of time in those activities compared to other resources. It is advisable to analyze in depth the internal steps of each activity, with the goal of being pragmatic enough to conclude that these resources must be trained to gain a better performance.

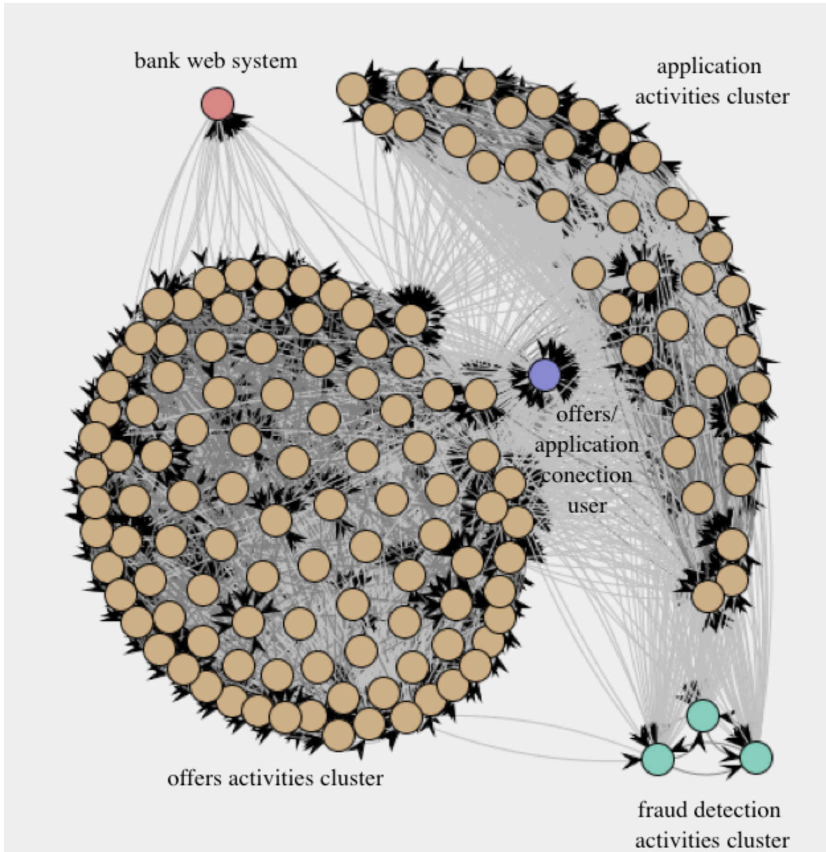
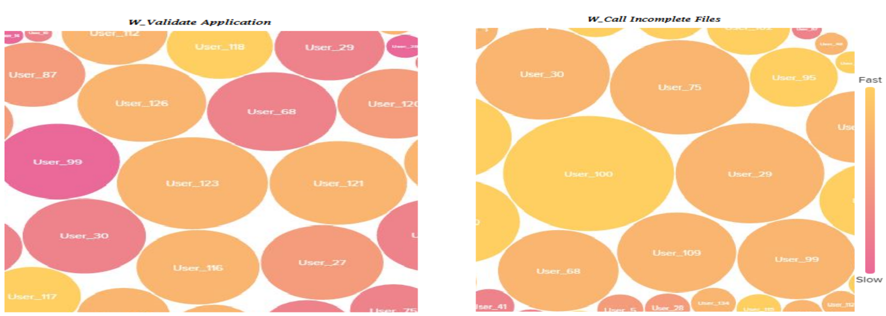



**Fig. 21:** Organization activity clusters using ProM.




**(a)** Number of cases **(b)** Number of resources

**Fig. 22:** Performance analysis

## 8 Business rules compliance

As demonstrated in Section 7, employees influence the two points analyzed (throughput time and final outcome). However, given the nature of the sector in which the analyzed process is developed, it is highly vulnerable to corruption and malpractice. We proceed in this section to analyze the compliance with possible business rules. Section 8.1 discusses business rules associated with resources, and Section 8.2 explores business rules associated with process flow.

### 8.1 Resources

An analysis was conducted to determine if there were resources who carried out activities in the same case that can be provided to infringe compliance with ethics and values. This analysis was performed over the 538,072 cases, where the activities of W_complete and W_validating were filtered, obtaining 69,390 cases, corresponding to 12.8% of the total. Additionally, a filter is applied to check if the same resource and the same case id match for the previously filtered activities and thus obtain if the same resource does the same act of completing and validating, which violates this business rule. The evaluation and validation of loan applications should be done by two different resources to avoid conflicts of interest.Similarly the presentation of the final offers and the validation of the loan applications should not be done by the same resource.

**Table 4:** Common resources for completing and validating activities.

| Common resources | Case ID | Activities |
|---|---|---|
| 75 | 927205634 | |
| 75 | 1786874274 | |
| 102 | 1191408308 | W_Complete application |
| 102 | 1582198848 | W_Validating |
| 109 | 742871702 | |
| 100 | 454790674 | |

Following with the previous analysis, W_Create and O_Created activities are filtered, obtaining 72,821 cases. After that, the cases are filtered if the same resource and the same case id do the same activity to detect if there is an infringement of these business rule.

It can be seen in Table 4, that four resources do both *assessment* and *validation* activities, which can be prone to frauds and infringes the business rule. Table 5 shows that 19 resources do both validation and final offer creation activities, violating this business rule.

### 8.2 Process flow

After analyzing the process flow (shown in Figure 2), as well as the logic of the process, the following hypothesis were proceeded to be validated. First, if all applications that have been accepted were previously evaluated. Next, if *W_Call after offers* activity has an impact on process time and if there is a significant

**Table 5:** Common resources for completing and offer creation activities.

| **Common resources** | **Activities** |
|---|---|
| 3, 5, 14, 18, 19, 24, 29, 30, 34, 35, 41, 49, 75, 87, 95, 99, 100, 102, 109 | W_Complete application O_Created |

difference between applications that pass through this activity and those who do not.

Figure 23 shows a process graph where it can be seen that 1,861 cases do not comply with the first hypothesis. Meanwhile, it was found that 86 resources intervene in those cases (resources that accept the request even though it has not been evaluated).

To address the second hypothesis the data was analyzed and it was observed that the vast majority of the applications pass through *W_Call after offers*. Nevertheless, only 341 cases reach the complete state. Assuming that completing this activity implies that it was possible to talk with the client, it was observed that regardless of the endpoint at which the process arrives, the processing time decreases to 8 days on average. As a result, it can be concluded, that completing this activity adds value to the process, since clients tend to decide faster, either to accept or cancel an offer.

Another interesting finding, is that there is a positive relation to the incompleteness of information in the validation activity, as seen in Table 6. This conclusion is reached since 80.4% of these applications pass through the *A_Incomplete* state, whereas the applications where they comply with the analyzed sequence, only 50% pass through the *A_Incomplete* state. It can be concluded that the evaluation of the application influences the completeness in information for the step validation. In other words, a case that passes through *W_Complete application* has a positive impact on the completeness of the information in the *W_Validate application* step.

**Table 6:** Cases where *W_Complete application* are followed by *A_Accepted*

| Item | Not followed by A_Accepted | % | Followed by A_Accepted | % |
|---|---|---|---|---|
| Cases where W_Complete | 1,861 | | 31,726 | |
| | Pass through A_Incomplete | | Pass through A_Incomplete | |
| Cases where W_Complete | 1,496 | 80% | 15,638 | 50% |

## 9 Recommendations

In order to improve the loan applications process presented by the BPI Challenge 2017, it would be interesting to take into account the following recommendations:

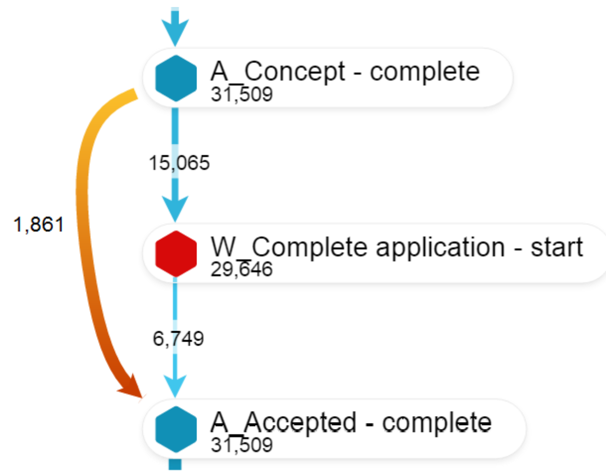



**Fig. 23:** Skipping of *W_Complete application*.

1. Conduct surveys to clients whose applications end up cancelled, only if it can be determined what caused the desertion of these. There are multiple attributes to measure the quality of a service, according to the human resources' perspective: the executive's ability to respond to customer doubts, kindness, treatment, or even related to the product itself: flexibility to certain types of customers, costs compared to competitors, flexibility of offers in terms of monthly fees and costs, availability of information online for customer inquiries through the web or in person. In other words, to gather information that complements the analysis of the desertion of customers, since the process is not designed with this approach.
2. Enrich the metadata with sociodemographic attributes such as occupation, age, net income, type of work contract, so that together with the existing data from the bank, profile of offers can be made according to the type of credit requested. This way, the number of offers can be limited to a maximum of 3, with a profile that has high acceptance in a certain segment of customers.
3. With the data received from the bank it was identified that the number of users involved in an application was a minimum of 8 and a maximum of 14. It would be interesting to know in advance which users are more capable of getting the case to an acceptable outcome. To achieve this, the creation of a pilot tests is recommended, where users are selected for each test and their performance on each activity can be measured, and compared to other users who are more experienced or have a better performance. Thus, they can focus on improving in the activities where they had a weak performance or only focus in doing the activities that suits them best according to their performance.

## 10 Conclusions

This paper analyzed the three questions proposed by the BPI Challenge 2017. There was an interest in determining whether if longer times of the process are associated with activities performed by the bank employees, or by the clients. The result obtained is that clients take a long time to answer applications and to solve this it is recommended to apply another communicational strategy. Furthermore, this paper analyzed whether if there is a relation between the completeness of the application and its approval. The results showed that loans with pending documents were accepted more often. Complete applications are more likely to be cancelled, either by clients or the bank. The last question was related to the number of clients who ask multiple offers. The greater the number of offers, the lesser the acceptance and longer the time taken. Social networks were analyzed and results showed that there are three clusters present.

After the organizational analysis of resources it was discovered that there is a trade-off between throughput time and chances of a favorable outcome, depending on the number of resources that work in the case.

Analyzing response times by customers, it was determined that the vast majority of clients with approved applications make contact within the first 15 days. Thus, a recommended way to improve the general performance of the process, is to create a communication strategy where emphasis is placed on the 15 days after the offer.

While analyzing the questions proposed by the challenge, some different questions arose. When taking a process focus and valuing the possible deviations of it, there were some activities that added value to the investigation, like *W_Complete application* and *W_Call after offers*. The first analyze activity has an influence in the incompleteness of application, and the second activity do influence throughput time of the process.

## References

1. van der Aalst, W. M.: Data scientist: The engineer of the future. Enterprise Interoperability VI 13–26 (2014)
2. van der Aalst, W. M.: Event Logs: What kind of data does process mining require? DOI 10.1007/978-3-642-19345-3_2, Springer-Verlag Heidelberg 2011.
3. van der Aalst, W. M.: Process mining in the large: a tutorial. Business Intelligence 33–76 (2014)
4. van der Aalst, W. M.: Process mining: Data Science in Action. Springer (2016)
5. Harrison-Broninski, K.: Human Interactions: The Heart and Soul of Business Process Management: How People Really Work and How They Can Be Helpful to Work Better (2005)
6. Leemans, S. J., Fahland, D., van der Aalst, W. M.: Exploring processes and deviations. International Conference on Business Process Management 304–316 (2014)