Driving Higher Operations Effectiveness in Banking: Process Mining and Data Science

Business Process Intelligence Challenge: 2017 (BPIC 2017)

Lalit Wangikar, Amit Chandra, Dikshant Yadav, Neel Biswas, Rajat Kumar, Saif Alam

Cognitio Analytics, 2 Woodland Road, Short Hills, NJ 07078, USA
{lwangikar, achandra, dyadav, nbiswas, rkumar, salam}
@cognitioanalytics.com

Abstract. This paper presents findings from our analysis of the process data for a loans application process from a financial institution (Bank) in Netherlands. The application process was segmented in to multiple stages to better understand progression of applications to decisions made by the FI and the customers applying for loans. We find that progression (timing and intensity of process events) of an application through the application process has significant impact on the outcome: acceptance or cancellation of an offer for loan. Specifically, as time and number of activities performed vary, chances of acceptance significantly change. The paper goes in to details on how this impact of timing and intensity of activities varies at different stages of the process. We then identify potential opportunities for the Bank for improving process performance.

Keywords: Process Mining, Process Improvement, Celonis, Performance Optimization, Operational Effectiveness, Loan Approval, Underwriting, Banking

1 Introduction

Exploration of innovative and better techniques to make use of Big Data is an ongoing phenomenon. Process Mining is one such technique that is becoming integral part of the tool kit used for better managing operations performance in any organization. The 2017 Business Process Intelligence Challenge (BPIC 2017) provides an opportunity to demonstrate the usefulness of process mining in a real-life business process. In this challenge, we analyze a real-world event log data for loan application process provided by a financial institution (Bank) from Netherlands. We use a combination of commercial, proprietary, and open-source tools to answer key questions posed by the

process owner and conduct further analysis to identify potential opportunities for improving process performance.

1.1 Approach and Scope

Our approach was designed to answer the key questions posed by the process owners on the BPIC 2017 challenge website: https://www.win.tue.nl/bpi/doku.php?id=2017:challenge.

Specifically, our approach included the following key steps:

- Develop a thorough understanding of the event log data
- Understand in-depth the underlying business process
- Identify key phases of the process
- Understand the relationship between activities within different process phases and eventual process outcomes
- Conduct analysis to answer the questions posed by the process owner
- Develop an analytical framework to prioritize applications based on the possible outcomes for achieving higher effectiveness in operations

The reminder of the paper presents our key findings as we conducted analysis in support of this approach.

2 Understanding the Data

2.1 Developing Thorough Understanding of the Data

The provided data capture process events and relevant attributes for 31,509 loan applications submitted during the year 2016. A total of 1,202,267 events were contained in the data for these 31,509 cases, starting with a customer applying for a loan *(or the Bank starting it)* and ending with conclusion of that application into an Approval, Cancellation or Denial. There were 98 cases which did not conclude into any of the identified end states i.e. Approval, Cancellation or Denial. We assume such cases to be unresolved and that they were in progress at the time the data was collected. We have excluded these 98 cases from most of our analyses.

The events corresponding to an application appear in the order in which they occurred and describe steps the applications go through as part of the overall process. Based on the nature of these events, they are categorized into three major groups - Application state changes, Offer state changes and Workflow events. The event names are prefixed with notations “A_”, “O_” and “W_” indicating whether they are associated with application’s state change or offer’s state change or workflow respectively.

The data field “Lifecycle:transition” indicates the lifecycle of the corresponding activities. For the workflow activities, possible lifecycle stages include – “schedule”, “start”, “resume”, “suspend”, “complete” and “ate_abort”. We identified the activity

names by the field “concept:name” in the event log. The activities along with their lifecycle at a time point, as indicated by the data field “Event-Name” are considered as events. The event log also has a resource indicator and the time of event completion.

The following table (Table 1) lists and explains the activities in the event log, corresponding to each category (Workflow/ Application/ Offer).

Table 1. Activities in the Event Log

Category of Activities	Activity	Description
"A": Activities indicating the application status	A_Create_Application	Customer applies for loan
	A_Submited	Online submission of application by the customer
	A_Concept	Completion of first assessment of an application post submission
	A_Accepted	The application is completed and is eligible for offers
	A_Complete	The offers have been sent to the applicant and the bank waits for the applicant to return a signed offer along with the required documents
	A_Validating	The applicant returns with an offer and validation of application is initiated
	A_Incomplete	The provided documents are insufficient / invalid; accordingly, the applicant needs to send additional documents
	A_Pending	Loan is final and customer is paid
	A_Denied	Denial of the application by bank due to the application being unfit for acceptance
	A_Cancelled	Cancellation of the application due to applicant being unresponsive or not requiring the loan anymore
"O": Activities indicating the Offer status	O_Create Offer	Offer created by the bank
	O_Created	
	O_Sent (mail and online)	Sending the offers to the applicant via mail and online
	O_Sent (online only)	Sending the offers to the applicant via online only
	O_Returned	The applicant returns an offer with additional documents
	O_Accepted	Acceptance of an offer by bank for loan approval
	O_Refused	Rejection of an offer by bank due to the offer being unfit for acceptance
	O_Cancelled	Cancellation of an offer due to not being returned by applicant
“W”: Workflow Activities	W_Handle leads	First assessment of an application submitted online by an applicant either automatically or manually
	W_Complete application	Calls for completing an application
	W_Call after offers	Calls for following up on the offers created
	W_Validate application	Validating documents/ information provided corresponding to the offer returned
	W_Call incomplete files	Calls for collecting additional documents/ information
	W_Assess potential fraud	Measures for detecting potential fraud
	W_Personal Loan collection	Custom workflow involving specific cases
	W_Shortened completion	Custom workflow activity involving investigation of low risk applicants

2.2 Tools Used for Analysis

Process Mining software Celonis, Programming Language R and Microsoft Excel were used for the analysis.
Celonis 4.2 was used to prepare the process map. With the help of Celonis functionalities, we could study and understand the underlying process in detail.
Microsoft Excel 2016 and Programming Language R - version 3.2.3, were used extensively to perform deep dive analysis and obtain insights. With the help of Excel, we could present our findings using charts and tables.

3 Understanding the Process in Detail

We imported the event-log data into the process mining software Celonis and obtained the corresponding process map for the loan applications.

3.1 Re-sorting the Event Log for Better Process Understanding

The process map revealed several events in a case occurring with the same exact time stamp. In Figure 1.A, the event "W_Complete application – schedule" appears before the event "A_Concept – complete". This is contrary to our understanding of the process where the activities or completing an application are scheduled after the "A_concept – complete" has occurred. The system assigns the same time stamp for both activities based on how it is programmed to work. Such activities with same time stamps can potentially appear in any sequence when one extracts events for analysis. This can lead to process diagrams where appear activities appear in unexpected patterns as above and can also create additional process variation where none should exist.

A. Event Log: Before Sort B. Event Log: After Sort

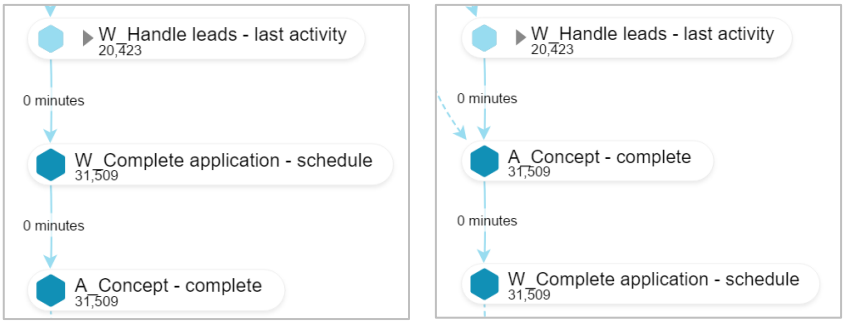

Fig. 1. Trajectory of events A_Concept – complete and W_Complete application – schedule before and after using the "Sort" capability

To overcome such challenges, Celonis allows users to specify the sorting order when events occur at the same time. This is provided through the "Sort" column in the event

log while importing the data in to Celonis. Figure 2.B shows the correct sequence of activities after applying the right sorting order.

3.2 Understanding Key Process Phases

After developing a thorough understanding of the event-log data and minutely studying the process map, we divided the overall process into eight major phases. Figure 2 shows these eight (8) process phases. The process phases are explained below, along with key observations for each of the phases.

Phase 1: Initiating the Application:

The first phase of the process - Application Initiation, involves two activities - "A_Create application" and "A_Submitted". The phase starts with the event "A_Create Application – complete".

Key observations:

- Activities in this phase help us divide the applications in to two groups: a). Created online directly by the customer and b). created by a bank employee. We do so based on the user associated with the activity "A_Create application". When this user is "User_1", the applications are created directly by the customer / applicant, otherwise by a bank employee, associated with the respective user ID. Success rates and turn around time for the two groups are different as shown in the table below.

Table 2. End State of Applications by Creator

Application Created By	Application End Status	# of Applications	% of Total Within Group
a). Customer: (Median Throughput Time: 22 days)	Accepted	10,064	49.3%
	Denied	2,702	13.2%
	Cancelled	7,573	36.8%
	Other	84	0.4%
	Total	20,423	100%
b). Bank Employee (Median Throughput Time: 16 days)	Accepted	7,122	64.7%
	Denied	1,039	9.4%
	Cancelled	2,838	25.8%
	Other	14	0.1%
	Total	11,013	100%

These differences indicate that the two groups are very different. Possible explanations are:

a. Bank employees may be helping screen out some of the applications even before being entered in to the system

b. Customers who work with bank employees to enter applications may have a better overall relationship with the bank

Further study of these two groups will potentially yield insights on improving success of applications submitted by customers.

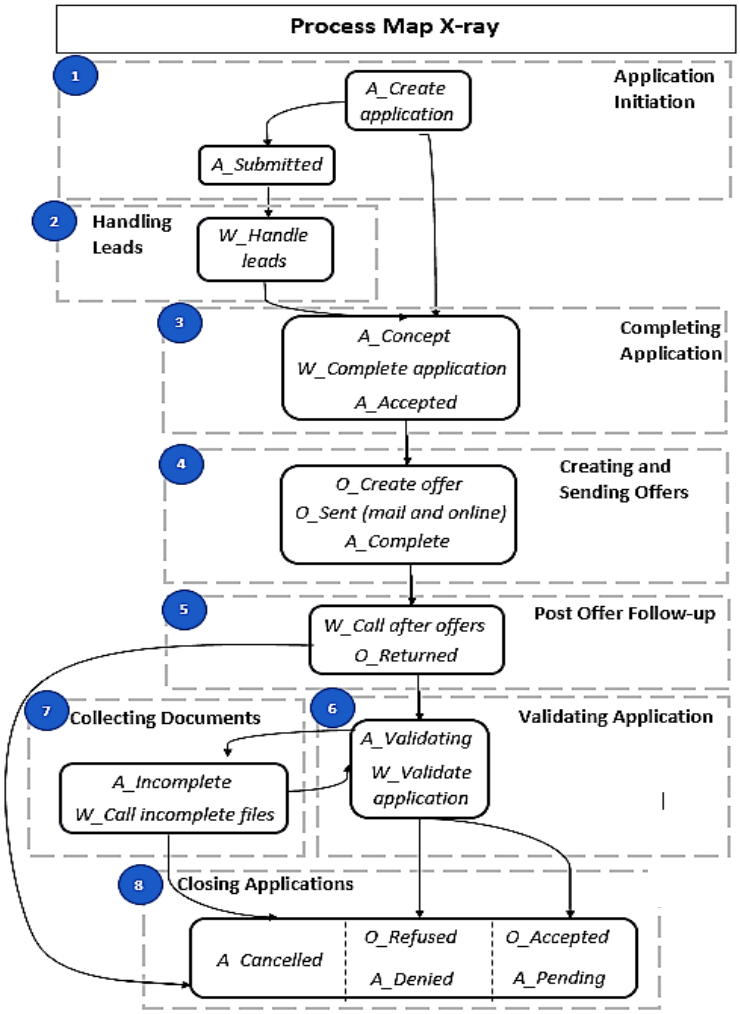

Fig. 2. Key Process Phases

Phase 2: Handling of Leads

The second phase of the process corresponds to the first assessment of the applications. It involves workflow activities related to handling of leads. Only applications created the customers (as explained in Phase 1 above) go through this process. The applications

created by a Bank employee skip this phase of the process and moved directly to Phase 3: Completing Application.

Key observations:

- Out of the 20,423 applications that go through this phase, 16,802 spend relatively little time (median: 1 minute), indicating these applications pass the initial screen without further work on part of the bank.
- The remaining 3,621 applications require additional "Handle Leads" activities, that take a median duration of 80 minutes. However, some of the cases spend significantly longer time in this phase, taking the average duration for these 3,621 applications spent in this phase to 366 minutes.

The two variants observed in this phase are shown in the figure below, along with the associated durations.

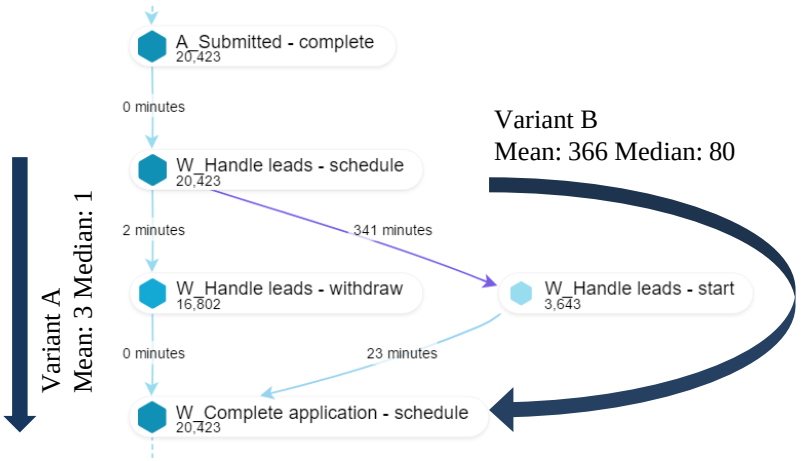

Fig. 3. Two Key Variations in the Handling of Leads Phase

The cases taking up Variant B required manual intervention for the first assessment of the application.

The variants followed by applications in this phase also had significant relationship to the end outcomes as shown in table below. Additional analysis on the impact of activities in this phase is presented in subsequent sections.

Table 3. End State of Applications by Handle Leads Activity

Handle Leads Variation	Application End Status	# of Applications	% of Total Within Group
a). Automatic Completion of Handle Leads (Median Throughput Time: 21 days)	Accepted	8,704	51.8%
	Denied	1,925	11.5%
	Cancelled	6,111	36.4%
	Other	62	0.4%
	Total	16,802	100%
b). Bank Employee Intervention (Median Throughput Time: 26 days)	Accepted	1,360	37.6%
	Denied	777	21.5%
	Cancelled	1,462	40.4%
	Other	22	0.6%
	Total	3,621	100%

3.3 Completing Application

This phase involves activities related to "W_Complete application", "A_Concept" and "A_Accepted". As per the event-log data, all the applications progress to this phase of the process.
"A_Concept - complete" i.e. the first assessment is complete and now calls can be made for completing application marks the start of this process phase. This process phase ends with the application status changed to "A_Accepted - complete".

Impact of time spent in this phase is presented as part of answers to questions posed, appearing later in this paper.

3.4 Creating and Sending Offers

The forth phase covers activities related to creating and sending offers to customers, such as "O_Create Offer", "O_Created", "A_Complete", "O_Sent (mail and online)" and "O_Sent (Online only)". The phase starts at the conclusion of the previous phase, with "A_Accepted – complete. The first event within this phase is "O_Create Offer – complete" and the phase ends with the events "O_Sent (Online only) – complete" or "O_Sent (mail and online) – complete".

Key observations:

- As per the event-log data, offers were created for all the applications, with multiple offers being created for 27% all applications.

- We noted that for some of the applications, multiple offers were created within a short span of time. We understood this to mean multiple offers being given the customer in one interaction, perhaps a live call with the customer.

Sending of the last offer for an application marked an end of this "Creating and Sending Offers" phase of the process. The application was assigned the status "A_Complete – complete". From here the applications moved to the next phase of the process i.e. Post Offer Follow up.

While we identify this as a distinct phase in the process, applications did come back to the phase after having moved down to subsequent steps in the process flow.

Additional Insight- First Offer Call

We see from the event log that several events in the phase 3 "Completing Application" and phase 4 "Creating and Sending Offers" occur in close proximity to each other. In fact, in several cases, 'O_Create offer – complete' event occurs in between the call for completing application i.e. at the time a user was handling the activity "W_Complete application". See Fig. 4 below. Detailed review of this part of the process map suggests that as soon as an application in complete, the bank attempts to make one or more offers to the customer on a call, sends out the offers to the customer and schedules follow up activities.

We observer that applications progress through first offer call via three mutually exclusive variants:

1. Variant 1: A_Accepted is directly followed by W_Complete application - start. For these cases W_Complete application - start has been taken as the beginning of the first offer call.
2. Variant 2: A_Accepted is directly followed by W_Complete application - resume. For these cases W_Complete application - resume has been taken as the beginning of the first offer call.
3. Variant 3: All the applications not moving through paths in Variant 1 and Variants 2. For these cases A_Accepted has been taken as the beginning of the first offer call.

To summarize, the first offer call made to the applicant involves three major parts of the process:

- Part of the call for completing the application (in case of variant 1 and variant 2)
- Part of the call for providing one or more offers
- Part of the call for first follow-up of offers sent

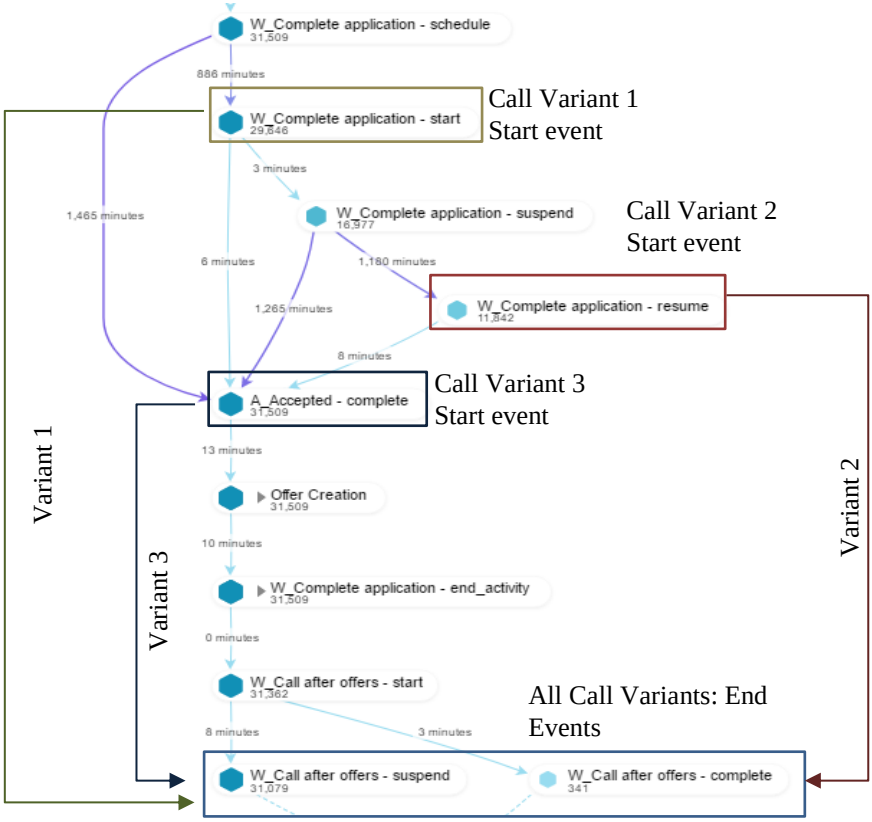

Fig. 4. Figure showing two different variants applications can take and the how the first call is defined for them (Figure shows average throughput time between activities)

Further, we see that as the time spent in the first offer call increases, the acceptance/approval rate (% applications accepted or approved) decreases, this is shown in Fig. 5 below.

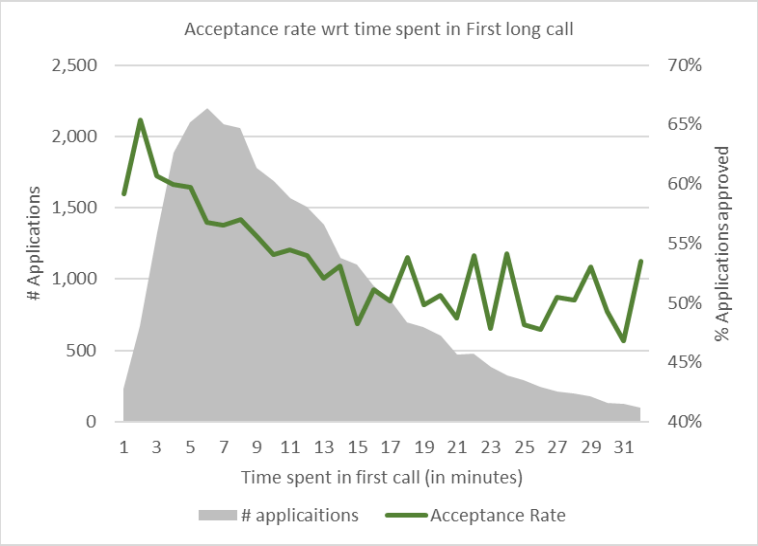

Fig. 5. Variation in application acceptance by time spent on first offer call

3.5 Post Offer Follow up

The fifth phase, "Post Offer Follow Up" involves activities performed to follow up on offers and get an offer acceptance from the customer, - "W_Call after offers", "O_Returned" . The phase starts with the event "W_Call after offers – schedule" and ends with either "W_Call after offers – complete", "W_Call after offers – ate_abort" or "W_Call after offers – withdraw".

Key observations:

- The activity "W_Call after offers" was scheduled for almost all of the applications, total of 31,362 applications, or greater than 99.5 % cases. If the applicant accepts (O_Returned) an offer, the application moves to the validation phase.

- During the Validation Phase, if the bank decides to reassess the applications because new details uncovered, then the application moves back to the "Creating and Sending Offers" phase, a new offer is created based on the new information and follow up on the offers starts again. There were 1,807 cases in which the offers were created after such changes in the validation phase.

- By the end of this phase 9,417 applications had got cancelled and 169 had got denied. Rest of the applications moved to the next phase of the process i.e. Validation phase.

3.6 Validating Application

"A_Validating" marks the application status change to "Validating" from "Accepted". This process phase involves activity - "W_Validate Application" and the events corresponding to that. The phase starts with the event "W_Validate Application – schedule" and ends with "W_Validate Application – ate_abort".

Key observations:

- A total of 21,870 applications went to the validation phase. The rest of 9,639 applications either had got denied or cancelled before progressing to the validation phase.
- Post validation, applications either got approved, or denied, or progress to the next part of the process i.e. Collecting Documents.

3.7 Collecting Documents

"A_Incomplete" marks the application status change to "Incomplete" from "Validation". This phase involves activities related to completing all the documentation for a loan, "W_Call incomplete files". It starts with the event "W_Call incomplete files – schedule" and ends with either "W_Call incomplete files – complete" or "W_Call incomplete files – ate_abort".

Key observations:

- This phase of the process focuses on obtaining additional documents from the applicant to complete the validation phase of an application. We also observed that, after/ during this process, offers were created again. From this phase, either the applications got cancelled or validation was scheduled for them post which they eventually reached any of the end states.

- For 15,003 applications reached this phase ("W_Call incomplete files" was scheduled). Out of these, 955 eventually got cancelled, 12,647 got approved after being validated again, 1,356 got denied and rest of 45 applications remained unresolved.

Time spent in this phase as well as number of activities performed in this phase had a significant impact on the success, as will be seen later.

3.8 Closing Applications

This phase simply marks the applications reaching their final status, one of Accepted, Cancelled or Denied.

Special Handling/Processing Activities:

There are a few workflow activities that indicate special handling of applications that occurs in few very applications. These involves activities such as, "W_Assess Potential fraud", "W_Shortened Application" and "W_Personal Loan Collection".

We observed that, these activities were performed for a handful of cases. "W_Assess Potential fraud", "W_Shortened Application" and "W_Personal Loan Collection" was scheduled for a total of 303, 74 and 2 applications respectively along the approvals process, ~ 1% of total applications.

4 Responses to Questions from Process Owners

4.1 Question 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,

We answer this question into two parts. In the first part we evaluate time spent in each process phase, with process phases as identified in Fig 2 and described above. The second part analyzes how the waiting time *(associated with the lifetime of an application)* is distributed between the Bank's activities – waiting in the system to be processed, and the applicant's activities – waiting for the input from the applicant.

We began with identifying the major phase of the process by observing the process map based on our understanding as was illustrated earlier in Fig.2. Thereafter, we evaluated the average time *(median and mean)* taken for a case to complete the activities corresponding to each of the identified phases.

A bottleneck analysis for some parts of the process helped us understand the difference between the waiting time associated with them.

To evaluate the throughput time per phase of the process, we computed the average time taken by an application from the first occurrence of the first event to the first occurrence of the last event of that phase of the process.

Fig. 6 shows the time spent by applications in each of the process phases.

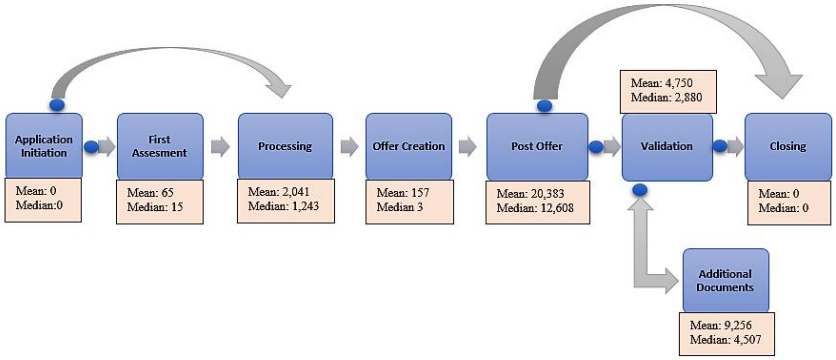

Fig. 6. Mean and Median throughput times (in minutes) per phase of the process

Our understanding of the term waiting time in the context of an applications lifetime is the time along the process when no work is done on the application.

Waiting time at Bank's end (waiting for processing) is defined as:

1. Time spent between scheduling of a workflow activity to the same being started
2. Time spent from last suspension of a workflow activity to the activity being aborted

For these sequences, we assume that the application is waiting in the Bank's system for processing.

Waiting time at Applicant's end (waiting for input from applicant) is defined as:

1. Time spent between suspension of a workflow activity which involves getting input from applicant to the same being resumed
2. Time spent from suspension of last call after offers till the application is cancelled: we assume that this time is spent waiting for the customer to respond. When no response is received by a particular time, application is cancelled.

Table 4 shows waiting times at Bank's end and at the Applicant's end, by different phases, along with specific activity sequences that are considered to contribute to the waiting time.

Table 4. Waiting Time for Applications

Process Phase	Waiting Time (Bank)	Waiting Time (Applicant)
Handling Leads	W_Handle leads - schedule to W_Handle leads – start Throughput time: Mean: 341, Median: 63 Number of Cases: 7,343	W_Handle leads - suspend to W_Handle leads - resume Throughput time: Mean: 56, Median: 7 Number of Cases: 780
Completing Applications	W_Complete application - schedule to W_Complete application - start Throughput time Mean: 892, Median: 14 Number of Cases: 29,421	W_Complete application - suspend to W_Complete application - resume Throughput time Mean: 1,180, Median: 145 Number of Cases: 11,757
Post offer Follow-up	W_Call after offer - suspend to W_Complete application - ate_abort (For applications not cancelled) Throughput time: Mean: 7,913, Median: 5,380 Number of Cases: 22,033	W_Call after offer - suspend to W_Call after offer – resume Throughput time: Mean: 5,550, Median: 5,693 Number of Cases: 25,065
		W_Call after offer - suspend to W_Call after offer - ate_abort (For applications that are cancelled after call after offers) Throughput time Mean: 31,463, Median: 38,340 Number of Cases: 9,476
Validating Applications	W_Validate application - suspend to W_Validate application - ate_abort Throughput time: Mean: 1,918, Median: 906 Number of Cases: 14,902	
Collecting Documents	W_Call incomplete files - suspend to W_Call incomplete files - ate_abort Throughput time: Mean: 7,208, Median: 3,837 Number of Cases: 3,677	W_Call incomplete files - suspend to W_Call incomplete files - resume Throughput time: Mean: 1,470, Median: 140 Number of Cases: 14,500

Impact of Waiting Times on Application Success:

As mentioned above, time spent between schedule to start of the same activity is treated as time spent by the application waiting for someone at the Bank to take a specific action. In this category of sequences, apart from schedule and start of activities "W_Handle leads" and "W_Complete application", the "start" event occurred immediately following the "schedule" event.

For these two, "W_Handle leads" and "W_Complete application" (both considered waiting time on the Bank's side), we observed that hour of the day for "schedule" event (when in the work day does someone perform the schedule activity) had an effect on the wait time of these steps. The same is illustrated in Fig 7.a and Fig.7.b below.

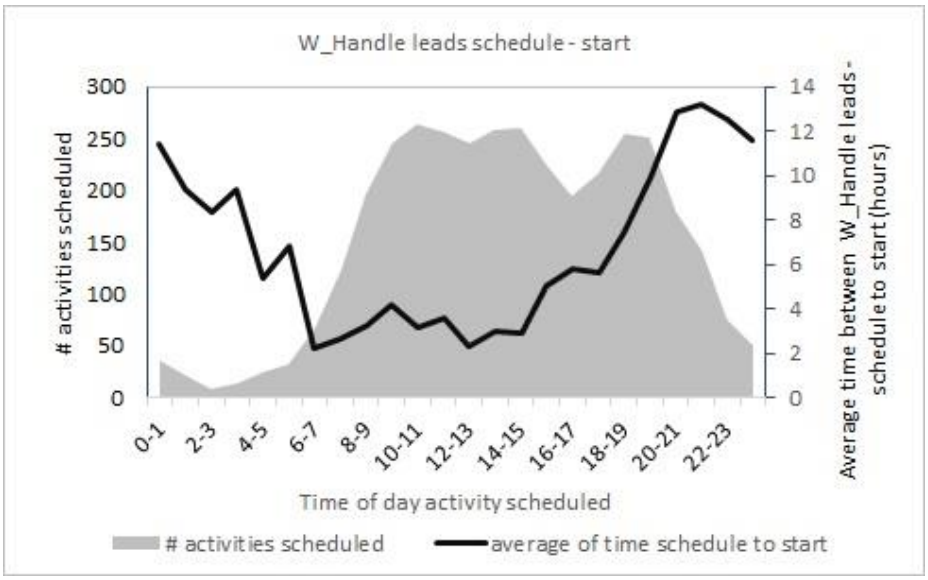

Fig. 7.a Average throughput time taken to start activty: "W_Handle leads", based on the hour of the day it was scheduled in.

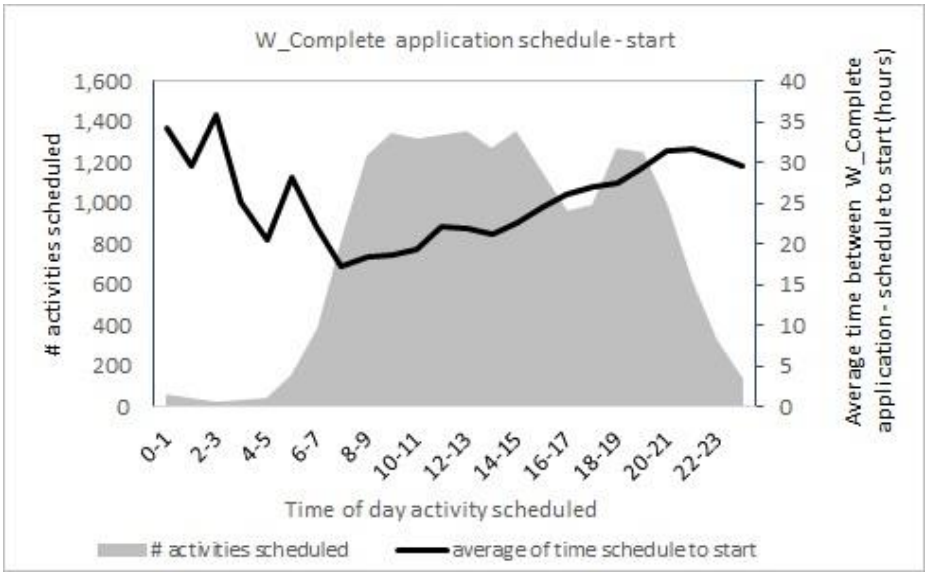

Fig. 7.b Average throughput time taken to start activty: "W_Complete Application", based on the hour of the day it was scheduled in.

As we see in Table 4, the difference between mean and median waiting times between schedule and start events for these activities is large.

- For "W_Handle leads", the mean is 341 minutes vs a median of 63 minutes

- For "W_Complete Application", the mean is 892 minutes vs a median of 14 minutes

The unusual difference in mean and median time is explained by extreme wait times for applications that had "schedule" events occur in off-hours and the corresponding "start" events occurred thereafter in the working hours. Additionally, for applications that had "schedule" events occur towards the close of the working day (Hours of 1600+), the wait time is higher as the "Start" events occur only on the following day for several such cases.

We also observed that, the waiting times (Applicant' side): W_Call after offers – suspend to resume and W_Call incomplete files – suspend to resume had a significant influence on the approvals of applications. To see the impact, we plot success rate across the waiting time taken in an activity by applications (Fig.8).

Based on our business understanding, we considered the following as "Success" for the two process phases under study:

A. Acceptance/approval of an application for the process steps of "W_Call incomplete files – suspend to resume"
B. "O_Returned" status of an offer was considered as success for the steps "W_Call after offers – suspend to resume"

In the following charts, 1 day waiting time between calls refers to a time interval of less than 24 hours between two calls (or average time b/w more than two calls), i.e. next call made within a day.

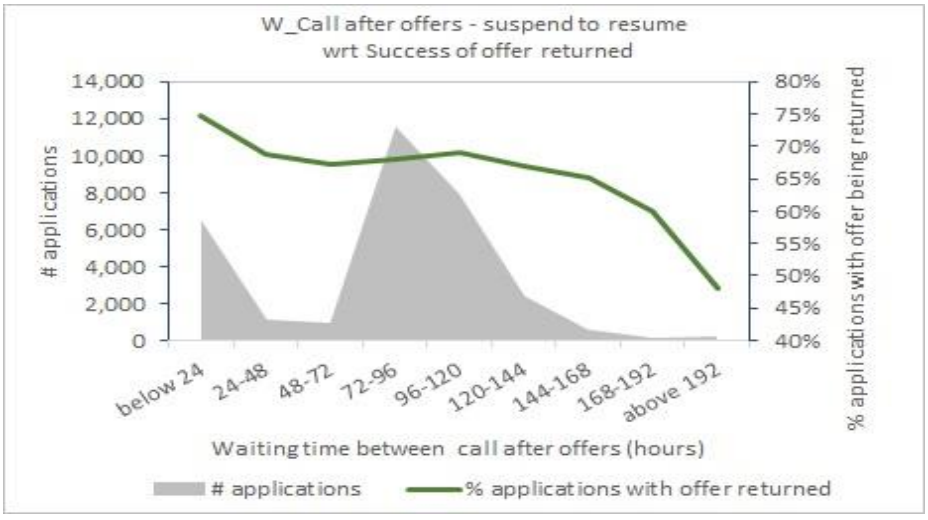

Fig. 8.a Impact of Waiting time in "W_Call after offers" on Acceptance Rate

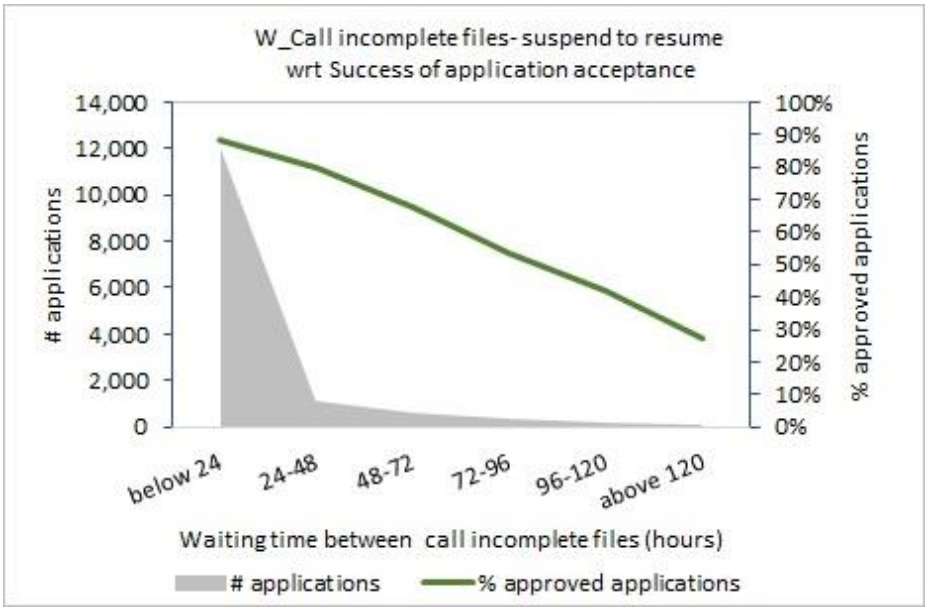

Fig. 8.b Impact of Waiting time in "W_Call incomplete files" on Acceptance Rate

We conclude from the charts in Fig.8 that the likelihood of an application to convert (Accepted) reduces as the waiting time increases.

4.2 Question 2: What is the influence of 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.

Incompleteness of an application can be related to the number of times the status of an application is changed to "A_Incomplete". The chart below shows the trend in acceptance/approval or cancellation of the final offer with the number of times the status "A_Incomplete" is assigned for an application.

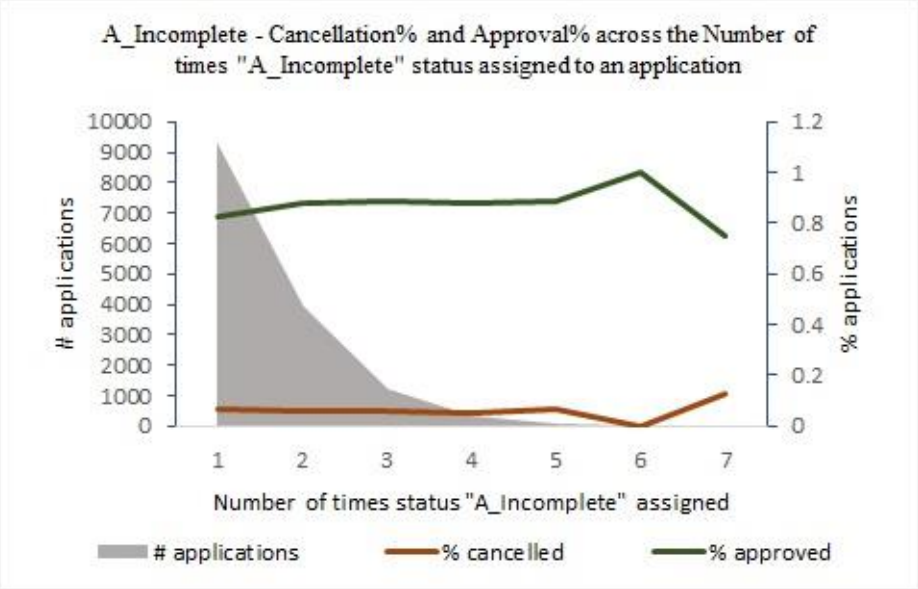

Fig. 9. Variation of Cancellation percentage and Approval percentage across the number of times an application is assigned the status "A_Incomplete – complete"

This shows that the acceptance rate or cancellation rate is not influenced by the frequency of status change to "A_Incomplete".

However, a better metric for this, based on our observation of the process map and understanding of the activities, would be the number of times activities related to "W_Call incomplete files" need to be performed for a given application and the time it takes to complete these activities. This activity is scheduled for an application post validation phase when the application status is changed to "A_Incomplete"

The lifecycle transition - "start" & "resume", corresponding to this activity, identifies the first and subsequent calls made to an applicant for the additional documents concerning his/her application. We measured the frequency of incompleteness for an application by the number of such calls made.

If the applicant provides the additional documents, the application eventually gets approved or denied post validation or, if there is no response from the applicant's side then the application gets cancelled. Hence, we identified the possible final outcomes as either the application gets approved for loan or gets denied or gets cancelled for no response from the applicant's side.

Our studied how the percentage of cancellation varies across the number of calls made for additional documents.

Analysis and Interpretation:

As per the Application Event-Log Data, the activity- "W_Call incomplete files", is scheduled for a total of 15,003 applications, which accounts for ~ 48% of all applications. We identified that out of these, 45 applications neither gets cancelled, approved or denied. We exclude such applications from our analysis, hence a total of 14,958 applications were include in this analysis.

Out of 14,958 applications, 955 get cancelled. For rest of the applications, validation is scheduled again, post which 12,647 get approved for the loan and 1,356 get denied.

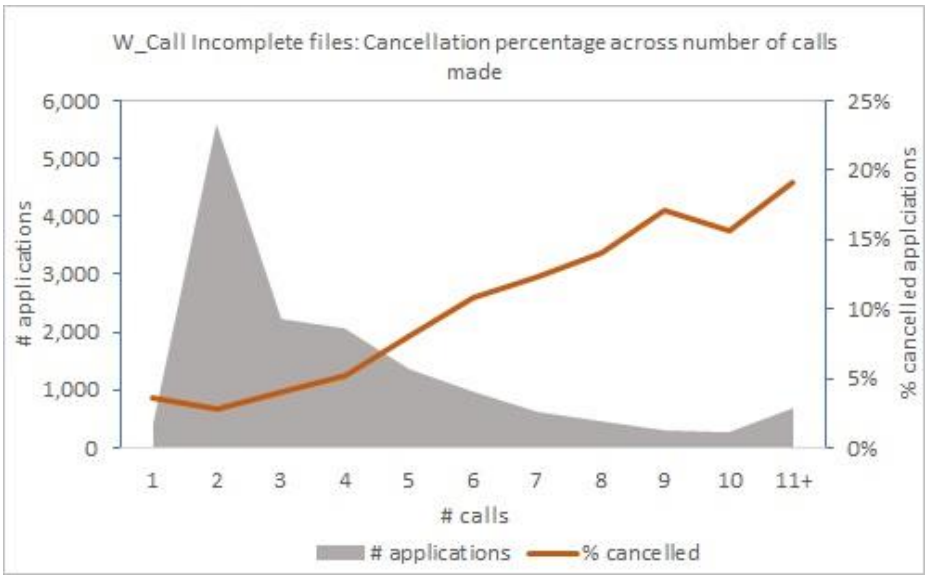

Fig. 10. Variation in Cancellation Rate by number of calls made for additional documents

In Fig.10. we see that cancellation rate increases with the number of calls made for additional documents. Additionally, the time difference between the calls (Waiting time) and the sum of call durations corresponding to an application (Total work time of the activity "W_Call incomplete files"), also impact the cancellation rate. Fig. 11 and 12 below show these relationships .

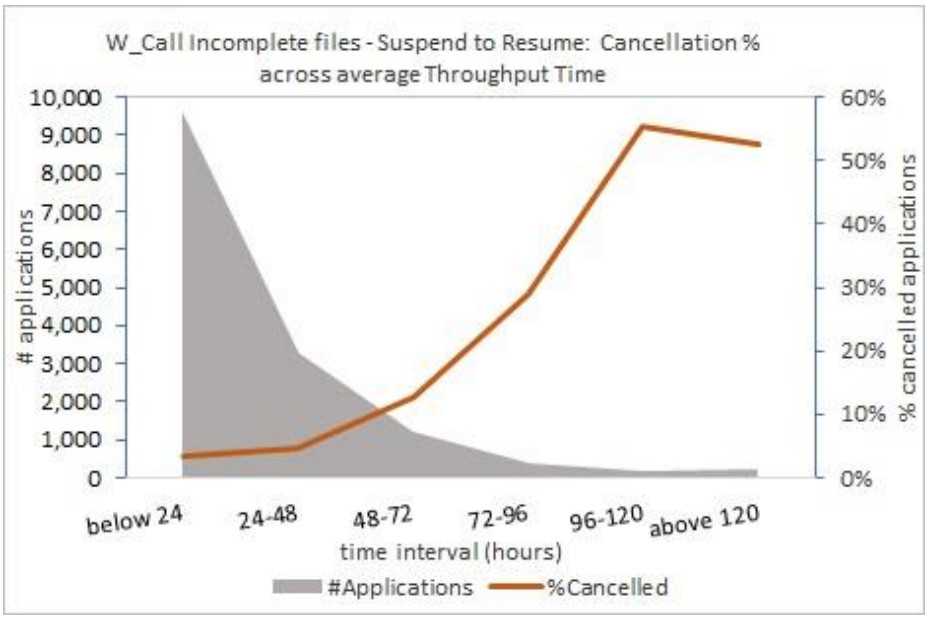

Fig. 11. Variation in Cancellation Rate by average time (in days) between calls for incomplete files

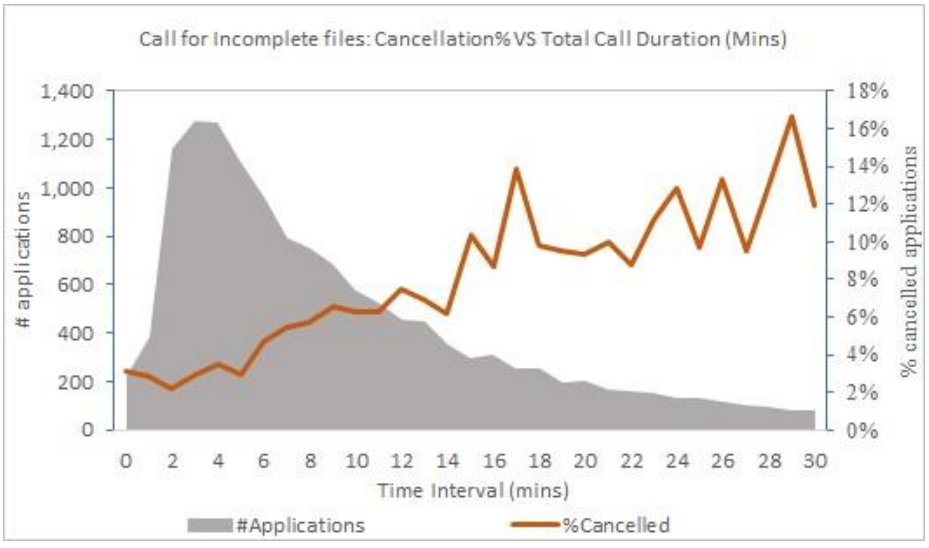

Fig. 12. Variation in Cancellation Rate by total durations of all calls incomplete files

Based on the observation from above charts we concluded that, a) higher number calls, b) longer waiting time between calls or c) longer call durations increase the likelihood of an application to get cancelled.

4.3 Question 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?

A total of 42,995 offers were created for 31,509 applications as per the Application Event-Log Data. We identified an offer creation by the Event-Name: "O_Create Offer-complete".

Based on our understanding, the Application event-log data identifies an application rather than an applicant. Hence, we based our analysis on the assumption that each application corresponds to a unique applicant and we have used the term application and applicant interchangeably for answering this question.

We understand that the Event-Name: "O_Accepted – complete", implies that any one of the offer created for an application/applicant is accepted by the Bank and the corresponding loan amount is approved for that application/applicant. We consider this as conversion of the corresponding application. For this analysis, we exclude the 98 applications that never reached either of denied, cancelled or approved state. We based our analysis on a total of 31,411 applications for whom 42,815 offers were created.

Analysis and Interpretations:

There are 8,511 applications/applicants for whom multiple offers were created. These represent approximately 27% of the total number of applications.

For an application, if the time duration between any two offers created by the same bank user is less than 301 seconds(that is up to 5 minutes), then we considered them to be created in the same conversation. Fig. 13 shows the distribution of applications based on number of offers made and total number of offer calls made for those offers.

We observe that the number of applications significantly decreases beyond the four offers and if offers are made in more than three conversations. The color intensities of the cells are in accordance with their magnitude.

Number of calls made for Offer Creation

Number of Offers Created	1	2	3	4	5	6	7	Total
1	22,900							22,900
2	2,648	3,901						6,549
3	150	622	565					1,337
4	59	142	144	93				438
5	5	28	44	29	19			125
6	1	8	5	11		4		29
7		4	3	4	1	3	1	16
8	1	1	3	4	3			12
9				1		1	1	3
10		1			1			2
Total	25,764	4,707	764	142	24	8	2	31,411

Fig. 13. Number of Applications by Number of Offers and Number of Calls for Offers

Further, we observe that if multiple offers are created in multiple conversations, then the chance of conversion gets higher, as illustrated in figure 14.

Number of calls made for Offers Creation

Number of Offers Created	1	2	3	4	5	6	7	Total
1	53%							53%
2	46%	66%						58%
3	50%	64%	72%					66%
4	42%	53%	68%	71%				60%
5	20%	75%	64%	62%	79%			66%
6	100%	63%	80%	82%	-	100%		79%
7		75%	67%	100%	100%	67%	100%	81%
8	0%	0%	67%	25%	67%			42%
9				0%		100%	100%	67%
10		0%						50%
Total	52%	65%	71%	69%	79%	88%	100%	55%

Fig. 14. Acceptance Rate by Number of Offers and Number of Calls for Offers

The color intensities for each cell is in accordance with the magnitude of approval percentage, also considering the number of applications corresponding to each combination of Number of calls & Number of Offers created in figure 13.

Out of the 27% applications for whom multiple offers were made, 59% got converted (accepted), compared to 55% for the entire population.

5 Using Event Logs for Predicting Outcomes and Optimizing Process Efforts

In addition to answering questions about process performance, the event log data can be used for predicting final state of a particular case. For the process under study for BPIC 2017, this would mean predicting likelihood of an application being accepted, denied or cancelled.

We created daily snapshots for each application based on event data available until the end of that day and used that to predict the likelihood of success: Acceptance. for that application. From the below chart, we can see that, as the information about an application increases, the ability to predict the end state of an application increases.

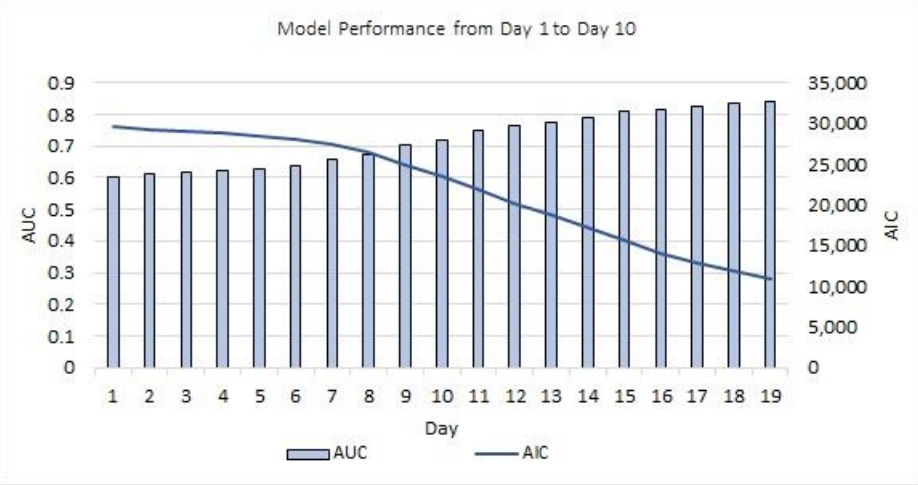

Fig.17. Model Output (AUC and AIC) from Day 1 to Day 19.

If such predictions can be made with relatively high certainty, an operations manager can redirect resources towards for achieving several critical business objectives, such as:

A. Focus efforts on applications that are highly likely to be accepted and try to close them faster
B. Change efforts on applications that are at risk of being cancelled by a customer where appropriate, such as high value customers etc.
C. Not spend efforts on applications that are not likely to convert to loans irrespective of bank efforts

The earlier we can predict the likely end state, the more we can influence both the outcomes and the efforts put in by the Bank.

6 Conclusion

Through this research paper we conclude that process events data are very valuable in not only discovering how processes are executed in a organization, but can be used to understand relationships between process execution and eventual business outcomes. Such understanding can then be used for driving a range of process improvement initiatives. Such data are also very valuable addition to any predictive analytics exercise and can support key operations decisions targeted at individual cases in a process based on how the cases are progressing. The financial institution that shared this data can benefit tremendously from such insights.

7 References

- Van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Heidelberg (2011)

- http://www.win.tue.nl/promforum/categories/-bpi-challenge-2017

- Business Process Intelligence Challenge 2012 Research Paper – Process Mining-Driven Optimization of a Consumer Loan Approvals Process (http://www.win.tue.nl/bpi/lib/exe/fetch.php?media=2012:bautista.pdf)