Process Mining: The first successful Peruvian case

Guillermo Calderón-Ruiz, PhD in engineering[1], Danny Fernández Luque, Systems Engineer[1]
[1]Universidad Católica de Santa María, Perú, gcalderon@ucsm.edu.pe, dfernandez@ucsm.edu.pe

***Abstract*– Process mining is a research discipline that allows to analyze and optimize business processes, it has been used in different types of companies (government, health, among others). In this paper we want to reinforce that process mining is a powerful tool in organizations, in addition we want to demonstrate to the Peruvian entrepreneur that its application brings substantial benefits to the company and the people involved.**

In 2016, different process mining algorithms were applied to the sales process of a Peruvian company and its optimization was achieved, five years later it was verified that the process was still operating correctly.

Keywords—process mining, business process optimization, application of process mining.

I. Introduction

Process mining is a research discipline that is between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand. Process mining aims to discover, monitor and improve real processes by extracting knowledge from event logs available in information systems [1]. In its early years, Process mining considered three main activities, (i) *process discovery* (i.e., extracting process models from event logs), (ii) *conformance checking* (i.e., comparison of real data, included in event logs, versus an "ideal" model; in order to find deviations), and (iii) *enhancement* (i.e., extract different types of information, e.g., performance of business processes, organizational mining, among others). [1,2]. "Modern" or online Process Mining considers additional activities as detect, predict and recommend [3].

Process mining has been used in various types of companies: governmental, health, private [4,5,6], among others, and has proven to be an excellent tool to improve processes and also the work of the people involved in these processes. This paper aims to reinforce to the international community, but especially to the Peruvian businessman, that process mining does improve processes; this demonstration was done by applying process mining in a local company in 2016 and then doing a follow-up in the year 2021.

This paper explains and shows the application of process mining in the sales process of a Peruvian food company. It explains how the event log was built for the analysis because the company's information system did not generate this data, this activity was the most costly in time and effort. It also details all the types of analysis that were applied to the process, firstly, process discovery to identify the flow and verify how the process was being executed; this allowed obtaining the ideal flow. Secondly, a performance analysis was performed, which allowed identifying bottlenecks and proposing alternative solutions. Finally, an analysis was made of the people involved in the process and it was possible to identify that they were doing things correctly, but there was generality in the execution of some activities. Five years later, a new analysis of the process was made, and it was possible to verify that the process was being executed correctly. The objective is to reinforce to the international community that process mining is a powerful tool to improve processes and also to improve the work of the people involved

This paper is structured as follows: section II explains the selected sales process, section III details the application of the different process mining techniques, and finally section IV lists the conclusions of the paper.

Digital Object Identifier (DOI):
http://dx.doi.org/10.18687/LACCEI2022.1.1.527
ISBN: 978-628-95207-0-5 **ISSN:** 2414-6390

II. Selected process

This work was carried out in a Peruvian food company, for confidentiality reasons we cannot mention its name. The work was developed in 2016 [7].

The sales process was selected for this work due to its critical importance, since it involves a direct relationship with the customer and also the main business objective depends on it (i.e., to satisfy in the best possible way a customer need in order to create a loyalty relationship). Currently the company has 5 channels for the sale of the product, which are: *pre-sale*, *office sales*, *intermediaries*, *supermarkets* and *own stores*, the difference between these is the way of contacting the client and the classification of the client. (to which the sale price, credit and distribution are associated), the sales process in the first 4 channels has the same basic flow, only in the case of sales in own stores is billing immediate, due to direct interaction with the client.

In the own stores channel, there is a direct interaction with the customer who arrives at the store, the attention and payment are immediate; these are classified as store customers. In the pre-sales channel, a salesperson contacts the customer to offer the product and take the order, this customer is classified according to the type of business they manage, who have a special sales price and variable credit according to the evaluation of each one, then goes to registration, assembly and billing, performed by different roles, the order is delivered to the customer's home and payment can be credit to 7 or 15 days according to evaluation and agreement with the customer.

For the office sales channel, the customer is the one who contacts the company to place the order, which is registered, assembled and invoiced; their classification is office customers, who have the same price as pre-sales, but the payment is in cash and they pick up their order at the office; in

the case of intermediaries, they send their orders by e-mail, they are registered, assembled, invoiced and finally sent for delivery; their classification is according to their region (north, center or south) and they are given a credit of up to 30 days according to their evaluation. In the self-service channel, upon receiving the order request from the customer, it is registered in the system, fully assembled, invoiced and sent for delivery, and a credit of up to 60 days is given.

As can be seen in the existing sales channels, there is communication with the customer at the beginning and end of the process, delivering the invoice and the order, regardless of whether it was complete or not. There are problems such as the lack of a clear definition of roles and responsibilities of users, confusing registration and parameterization of data, time deficiency and incomplete cases. After analyzing the problems encountered in the sales process, it was decided to apply process mining to optimize the process.

III. Applying Process Mining

The methodology used in 2016 was Bozkaya's process diagnostics, which consists of five phases: (i) *log preparation*, where the extraction of the system event log is performed, (ii) *log inspection*, with which a first view of the process is obtained, (iii) *control flow analysis*, an analysis of the information obtained from how the process is carried out, (iv) *performance analysis* and (v) *role analysis*, i.e. people and resources that execute the activities in the process [8]. Therefore, this section is divided according to these phases.

We used de ProM 5.2 [9] framework to perform these analyses.

A. Log preparation

In this phase, the files generated by the system managed by the company were inspected, concluding that the system does not generate a log of the activities carried out by the users, so the log of the process to be analyzed was assembled based on the transactional database. The data considered in the log were: *Case ID*, *Task ID*, *User*, *Timestamp*, which are the fundamental fields required by a log, additionally the *Category* attribute was considered as an attribute of each event, this describes the category of the client to which the case is associated. Fig. 1 shows a view of the log obtained.

The activities defined for the process were: (i) *Receive Order*: The store saleswoman or biller receives the order directly from the customer; (ii) *Take Order*: The horizontal channel salespeople visit the customer and take the order they wish to place through an app on their phone; (iii) *Register Order*: The Biller registers the customer's order in the information system, whether it was received directly or taken by a salesperson; (iv) *Evaluate Order*: The information system verifies whether the order meets the sales condition defined for the salesperson; (v) *Approve Order*: The Biller approves the order if it passed the evaluation performed by the information system; (vi) *Delete Order*: The Biller deletes the order if it does not pass the information system evaluation or contains incorrect data; (vii) *Review Order*: the warehouse area receives the approved orders, reviews them and assigns them to be placed; (viii) *Place Complete Order*: The complete order is placed when the total quantities of the required items are available; (ix) *Place Incomplete Order*: The incomplete order is placed when one of the required products is missing; (x) *Issue bill of landing*: The bill of landing is issued indicating the products covered by the order; (xi) *Invoice Order*: The order is invoiced according to the bill of landing issued by the warehouse area; when the order is invoiced, it is closed, whether it has been fully or partially filled; (xii) *Cancel Invoice*: If the order and the invoice are returned, the invoice must be cancelled; and (xiii) *Cancel bill of landing*: Once the invoice is cancelled, the bill of landing is cancelled so that the product can be returned to the warehouse area.

caseID	taskID	user	timestamp	categoria
361868	RecibirPedido	VendedoraT7	14/03/2016 17:07	Tienda
361868	RegistrarPedido	VendedoraT7	14/03/2016 17:11	Tienda
361868	FacturarPedido	VendedoraT7	14/03/2016 17:12	Tienda
361868	AceptarFactura	VendedoraT7	14/03/2016 17:14	Tienda
361869	RecibirPedido	VendedoraT7	14/03/2016 17:12	Tienda
361869	RegistrarPedido	VendedoraT7	14/03/2016 17:12	Tienda
361869	FacturarPedido	VendedoraT7	14/03/2016 17:13	Tienda
361869	AceptarFactura	VendedoraT7	14/03/2016 17:13	Tienda
1227019	TomarPedido	Vendedor1	09/03/2016 12:08	Oficina
1227019	RegistrarPedido	Facturadora	09/03/2016 12:13	Oficina
1227019	EvaluarPedido	Sistema	09/03/2016 12:14	Oficina
1227019	AprobarPedido	Facturadora	09/03/2016 12:14	Oficina
1227019	RevisarPedido	EncargadoAlmacen3	09/03/2016 12:17	Oficina
1227019	ArmarPedidoCompleto	EncargadoAlmacen3	09/03/2016 12:25	Oficina
1227019	EmitirGuiadeRemision	EncargadoAlmacen3	09/03/2016 12:29	Oficina
1227019	FacturarPedido	Facturadora	09/03/2016 12:41	Oficina
1227019	AceptarFactura	Vendedor1	09/03/2016 13:30	Oficina
1227020	RecibirPedido	Facturadora	09/03/2016 12:26	Oficina
1227020	RegistrarPedido	Facturadora	09/03/2016 12:31	Oficina
1227020	EvaluarPedido	Sistema	09/03/2016 12:32	Oficina
1227020	AprobarPedido	Facturadora	09/03/2016 12:32	Oficina
1227020	RevisarPedido	EncargadoAlmacen1	10/03/2016 13:48	Oficina
1227020	ArmarPedidoIncompleto	EncargadoAlmacen1	10/03/2016 16:24	Oficina
1227020	EmitirGuiadeRemision	EncargadoAlmacen1	10/03/2016 17:41	Oficina
1227020	FacturarPedido	Facturadora	10/03/2016 17:53	Oficina
1227020	AceptarFactura	Facturadora	10/03/2016 18:42	Oficina

Fig 1. A view of the created event log, containing the 4 log attributes: Case ID, Task ID, User and Timestamp. Also containing the case attribute Category.

B. Log inspection

As can be seen in Fig. 2, it was possible to prepare a log with 1994 cases, 21 roles (originators), and in a period of 1 month (i.e., March 9 to April 9).

Fig 2. Dashboard of the Sales Process, showing 1994 cases, 21 roles, start and end dates, among other information.

After reviewing the log, several start activities and several end activities were identified (due to cases that had started before March 9 and ended after April 9). Therefore, these

incomplete cases were filtered out and 1888 cases remained, which were used in the following phases.

C. Control flow analysis

In this phase, the process control flow was analyzed. In order to know how the real process is. For this purpose, *Alpha*, *Heuristic* and *Fuzzy* discovering algorithms were applied, then a conformance analysis was also performed.

Fig. 3 shows the result of applying the *Alpha* algorithm, we can appreciate a semi-spaghetti process, therefore we proceeded to apply other algorithms (i.e., *Heuristic* and *Fuzzy*) to identify the ideal process and on which improvements could be proposed.

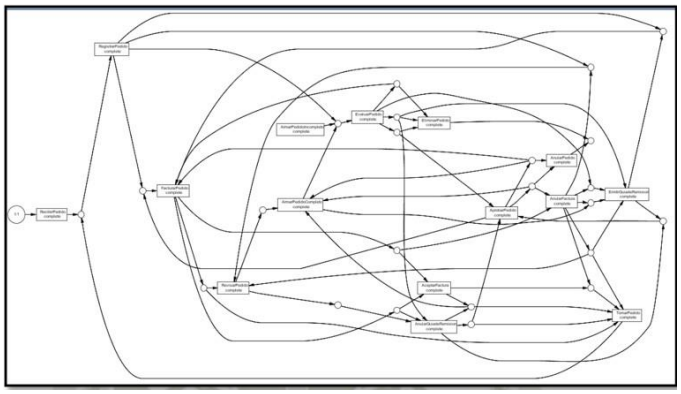

Fig 3. Resulting petri net of the Sales process, after applying the Alpha algorithm.

One of the algorithms that helped a lot in identifying the ideal process was the *Performance Sequence Diagram Analysis* algorithm, which allowed us to identify the most frequent flow patterns. Table I shows the five most frequent patterns in the process, which account for almost 94% of the event log execution. Taking these patterns into consideration, it was possible to filter the log to obtain a more structured process, which is shown in Fig 4.

In addition, the *Conformance Checker* algorithm was applied, and three inconsistencies were found in the activities *Review order*, *Invoice order* and *Cancel bill of landing* (see Fig. 5), so the company was recommended to regulate and control the execution of these activities.

D. Performance analysis

In this phase we applied the performance analysis algorithm with petri nets (see Fig. 6). Two bottlenecks (fuchsia-colored dots) and four moments to observe (yellow-colored dots) were found. In this article we will explain only one of the bottlenecks found.

Between the activities *Evaluate order* and *Approve order* we found a bottleneck (see Fig. 7), the average it gives us is 61.64 minutes, and a maximum of 1440 minutes, which indicates that after being evaluated, the order is put on hold until it is approved. Therefore, we analyzed the causes and we found three reasons: (i) that there is no authorized user for approval, at the time the request was passed (ii) that the request does not meet the requested conditions but is waiting to consult a supervisor to see if it can be given the special pass. (iii) delay of attention to the order by the biller because she is working on other orders.

TABLE I
MOST FREQUENT FLOW PATTERNS IN THE SALES PROCESS.

Pattern	Frequency (#)	Frecuency (%)
Receive Order - Register Order - Invoice Order - Approve Order	971	51.43%
Take Order - Register Order - Evaluate Order - Approve Order - Review Order - Place Complete Order - Issue bill of landing - Invoice Order - Approve Order	585	30.99%
Take Order - Register Order - Evaluate Order - Approve Order - Review Order - Place Incomplete Order - Issue bill of landing - Invoice Order - Approve Order	127	6.73%
Receive Order - Register Order - Evaluate Order - Approve Order - Delete Order	57	3.02%
Receive Order - Register Order - Evaluate Order - Approve Order - Review Order - Place Complete Order - Issue bill of landing - Invoice Order - Approve Order	33	1.75%
Total	1773	93.92%

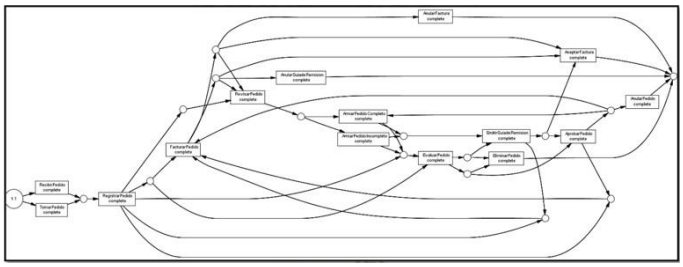

Fig 4. A refined petri net of the Sales process, after considering the most frequent flow patterns.

E. Role analysis

Several process mining algorithms were applied to perform the role analysis, in this section we highlight two. The first one is the application of the *Organizational Miner* algorithm (Fig. 8) which allows identifying a group of workers, group 3, who perform many tasks, which allows identifying a high degree of generality.

The second algorithm to be highlighted is *Handover of work*, which was analyzed with the *Analyzer Social Network* algorithm (Fig. 9), this algorithm shows that the interaction of the workers is correct. It can be seen that there is a high degree of interaction with the invoicing role (in the corresponding channels) and a high degree of independence of the saleswomen working in the own stores.

F. Transfer of results and 2021 verification

All findings were discussed with the company and those involved in the process, among the most relevant we have:

- Apply controls to the process flow, to ensure the correct execution of critical activities.
- Correctly define the roles of the users, since there are cases in which unauthorized roles execute activities that do not correspond to them (e.g., the activities of canceling invoices and approving orders are executed by warehouse personnel).
- Control the activities where bottlenecks were found, as well as place suitable personnel so that the execution of these activities does not generate delays in the process.

To verify if the improvements are maintained over time and to see the impact of this study on the company, a new analysis was performed in 2021. This paper shows only the process flow after applying the Alpha algorithm, but this time using ProM Lite 1.3 [10]. As can be seen in Fig. 10, the process is still structured and maintains the proper flows. Another important change made in the 2021 analysis was the change of methodology, this time the PM² methodology [11] was used.

The detail of the 2021 analysis and the new opportunities for improvement are not part of this paper and will be shown in a future work.

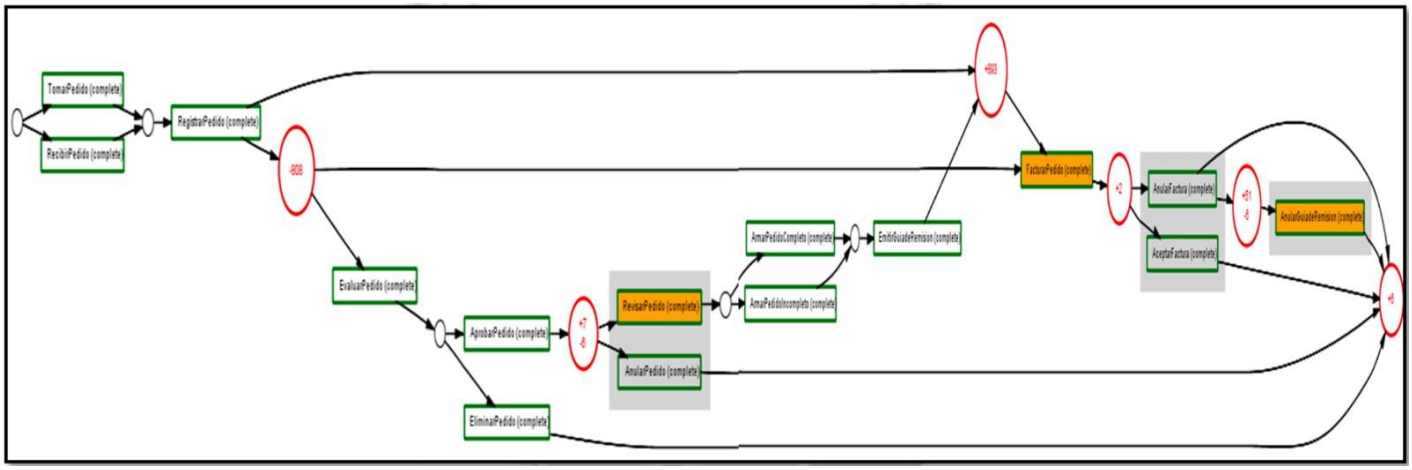

Fig 5. Inconsistencies found after applying the Conformance Checker algorithm, the three orange-colored activities (*Review order*, *Invoice order* and *Cancel bill of landing*) are executed without control.

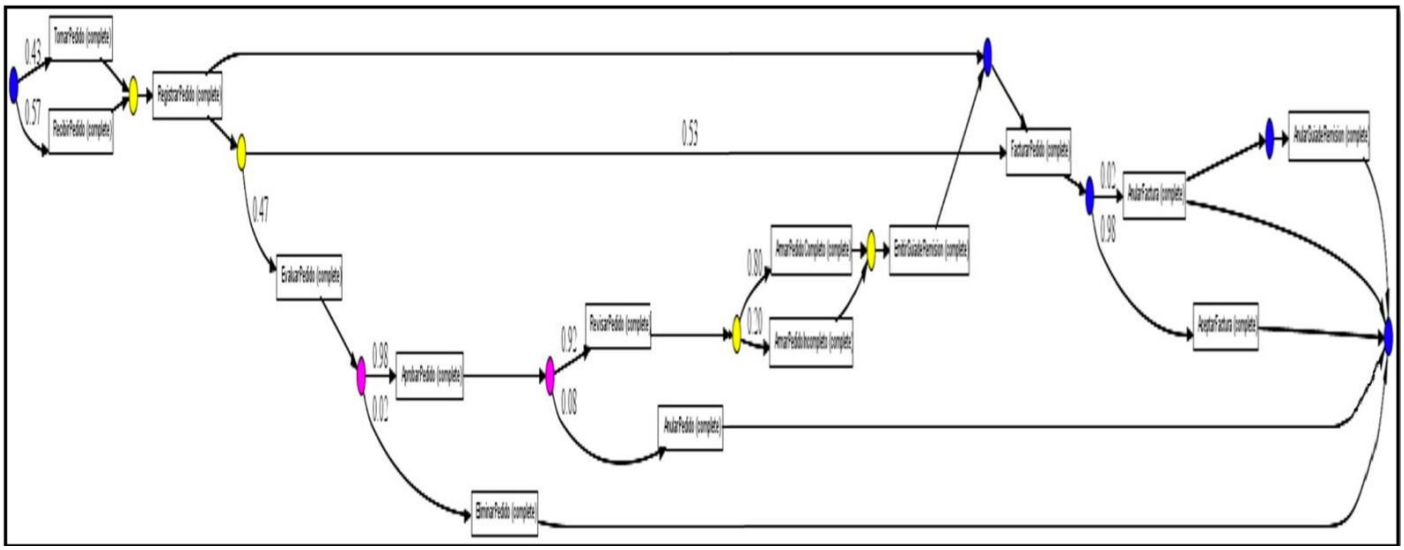

Fig 6. Petri net for analyzing performance of the Sales process. Blue dots indicate that the execution times are within the established limits, yellow dots indicate that the execution times are at the limit and may become bottlenecks. Fuchsia dots show bottlenecks. There are two bottlenecks in this process.

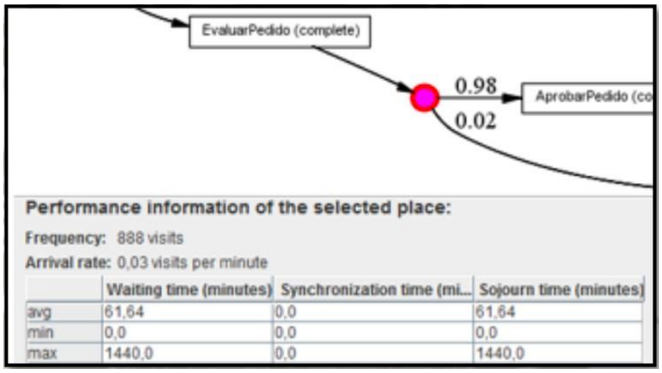

Fig 7. Identified bottleneck between the activities *Evaluate order* and *Approve order*. The average waiting time is 61.64 minutes and a maximum of 1440 minutes.

III. CONCLUSIONS

The objective was to reinforce to the international community that process mining is a powerful tool to improve processes and also to improve the work of the people involved, and we believe we have achieved that goal. In addition, our intention was also to demonstrate and motivate the Peruvian entrepreneur that the use of process mining will greatly improve their performance, and we are sure that now process mining will be applied in our country.

It should also be noted that the analysis time of a process is greatly reduced when applying process mining, what used to take weeks, even months, is now done in hours or days. In addition, the analysis (monitoring) can be done more frequently.

Another important aspect related to process mining is the existence of free and commercial tools that allow an exhaustive analysis of the processes. In this work we used ProM, a free and academic tool; this tool shows its results through graphical models, which allows a faster and clearer analysis, which was demonstrated in this paper. It was not necessary to go into much detail to understand the results after applying the different process mining algorithms.

The complete results of the analysis carried out in 2021 will be presented as future work.

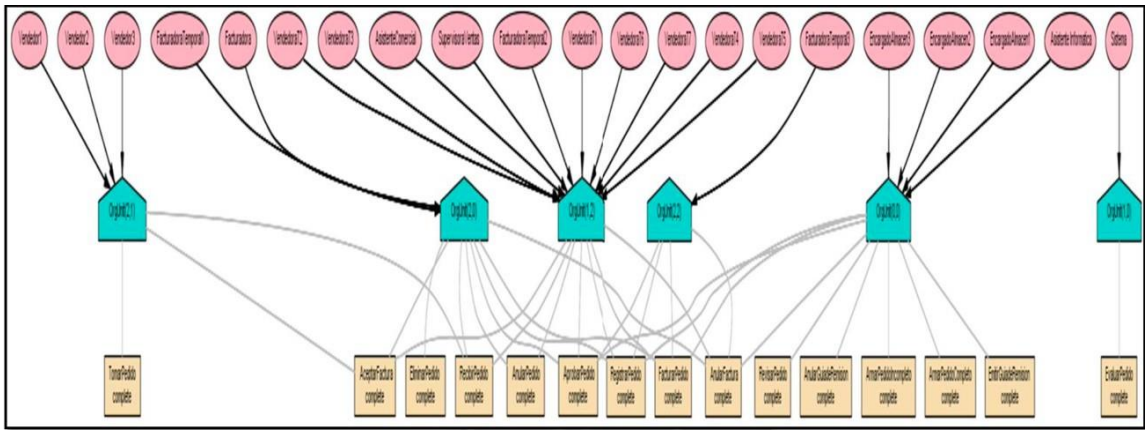

Fig 8. Result of applying Organizational Miner to the Sales process event log. We can identify that group three has several roles associated (pink nodes) and also several activities (orange squares), it means that there is a high degree of generality.

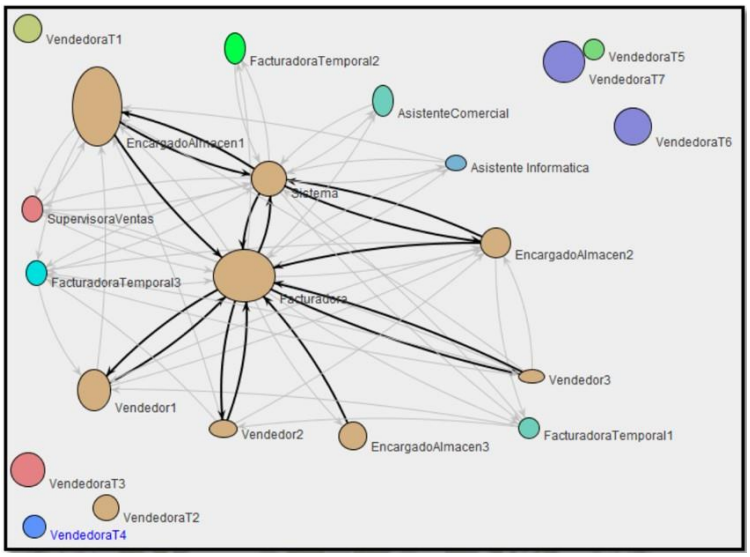

Fig 9. Results after applying the Handover of work algorithm and analyzed with the plugin Analyzer Social Network. The correct interaction with the invoiced role can be seen and also the correct independence of the saleswomen in the own stores.

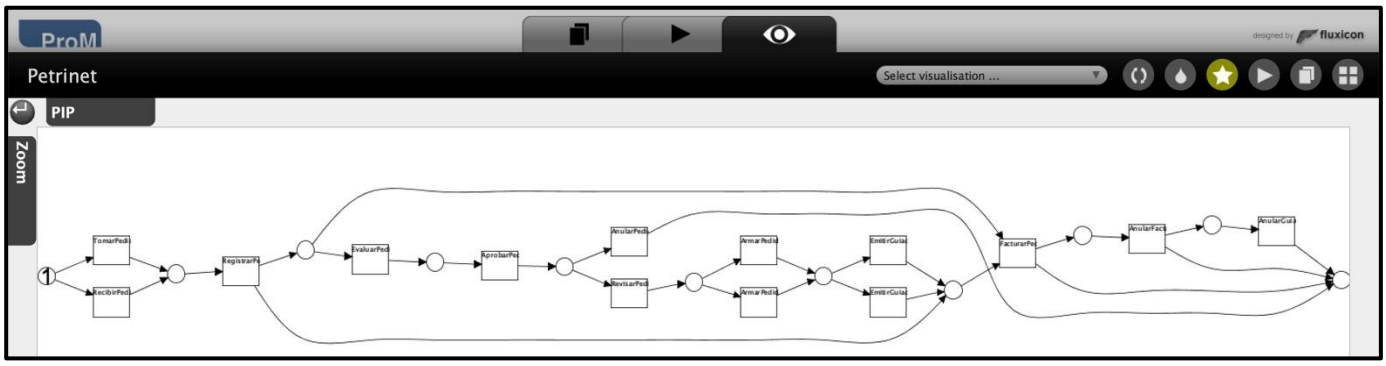

Fig 10. Resulting petri net 5 years later, improvements are sustained, a structured process is evident.

ACKNOWLEDGMENT

We would like to thank our great friend and professor Héctor Velarde Bedregal, for his motivation and support in this work, he collaborated a lot in the development of the analysis tasks, but unfortunately, he passed away last year. Rest in peace, dear and remembered friend.

REFERENCES

[1] W.M.P. van der Aalst, et al, "Process mining manifesto", *Lecture Notes in Business Information Processing*, vol. 99, pp. 169-194, 2012.

[2] W.M.P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer, 2011.

[3] W.M.P. van der Aalst, *Process Mining: Data science in action*, Springer, 2016.

[4] D. Dakic, et al, "Business process mining application: a literature review", *Annals of DAAAM*, vol 29, 2018.

[5] C.C. Osman, A.M. Ghiran, "When industry 4.0 meets process mining", *Procedia Computer Science*, vol. 159, pp. 2130-2136, 2019.

[6] E. Rojas, et al, "Process mining in healthcare: A literature review." *Journal of biomedical informatic*, vol 61, pp. 224-236, 2016.

[7] D. Fernández Luque, "Minería de procesos para el análisis y mejora del proceso de ventas de una empresa de insdustria alimentaria", thesis, Universidad Católica de Santa María, Arequipa, Perú, 2016.

[8] M. Bozkaya, J. Gabriels, J. van der Werf, "Process Diagnostics: a Method Based on Process Mining.", *Information, Process, and Knowledge Management eKNOW'09*, pp. 22-27, 2009.

[9] ProM tools (2022), ProM 5.2, available: http://www.promtools.org/doku.php?id=prom52.

[10] ProM tools (2022), ProM Lite 1.3, available: http://www.promtools.org/doku.php?id=promlite13.

[11] M. van Eck, et al, "PM²: a process mining project methodology", *Lecture Notes in Computer Science CAiSE 2015*, vol. 9097, 2015.