

# Incident completion time prediction

[TRANSLATED VERSION - This document was originally developed in Portuguese. In images containing graph titles and axes titles in the original language, the translation will be provided alongside with the graph.]

This questionnaire is part of a research on process mining. Therefore, your answers will help us in the development of this research. We appreciate your collaboration and remain at your disposal to talk more about process mining if you are interested.

Best regards,

Thais Rodrigues Neubauer, Profa. Dr. Sarajane Marques Peres and Prof. Dr. Marcelo Fantinato.

\* Required

1. What is your full name? \*

---

Consider a "incident management system" which stores information that can be used to help designing a "incident completion time predictor system". An incident completion time predictor system is capable of estimating, at any time point, how much time is left for a incident to be resolved.

Also consider that:

- 1) the incident manager system is used for a company that support services based on hardware and software structures - for instance, a company that offers telecommunication services;
- 2) the incident management process followed by this compnay is supported by the ITIL framework;
- 3) the completion time predictor can be triggered at any point of the life cycle of an incident;
- 4) the life cycle of an incident goes through states as: new, active, waiting for information, resolved, closed etc.

Our goal:

Create a "incident completion time predictor system" that make predictions based on knowledge derived from the management history of incidents already handled by the company.

Our problem:

Using the entire history of incidents handled by the company will result in a predictor system of low precision since there is a great variability in the incidents' features and its resolution process.

## Context

Our solution:

To use an algorithm that clusters similar incidents based on their life cycles and other incidents' information. This information do not include the incidents' completion time as this information would not be available in a real world environment since it is exactly the information we want to predict. We suppose that we can build "one completion time predictor per group of incidents", increasing the precision of the prediction since each predictor will work on in a group of similar incidents, with also similar resolution process.

Your help:

Our algorithm produces cluster of similar incidents that might me enhanced by an expert's analysis. Hence, we ask you to analyse the clusters resulting from the application of the clustering algorithm and indicate possible enhancements.

To orient your analysis, we elaborated four tasks for you to complete. We would like for you to feel involved in the tasks in order to solve them in the best possible manner. To this regard, chose a moment in which you feel motivated to participate in this research. We estimate you will spend no more than 40 minutes to solve all four tasks.

To assist you in some tasks, an example of the expected resolution will be presented using an illustrative context.

Before we start....

- For us to be able to estimate the time it took you to carry out the proposed tasks, <sup>\*</sup> enter below the time when you are starting the tasks:

---

*Example: 8:30 AM*

### Task 1: Enhancing the characterization of the clusters

#### Illustrative context

Suppose that our algorithm created the following groups of dishes served in a restaurant based in information of who prepare them and of the equipments used in the preparation. Also consider that the goal of this analysis is to reorganize the prices of the dishes served at this restaurant.

Group 1	
Involved Professionals	Required equipments
Chef	Stove
Cooker	Oven
Assistant 1	Processor
Assistant 2	Fridge

Group 2	
Involved Professionals	Required equipments
Chef	Stove
Barbecuer	Oven
Assistant 1	Freezer
Assistant 2	Blowtorch

Group 3	
Involved Professionals	Required equipments
Cooker	Oven
Assistant 1	Blowtorch
Assistant 2	Freezer
	Mixer

Group 4	
Involved Professionals	Required equipments
Cooker	Oven
Barbecuer	Grill
Assistant 1	Processor
	Freezer

Now consider that when we showed this four groups of dishes to the chef of this restaurant we had received the following observations from:

- Observation 1: all the dishes in which I am involved have a higher cost than the ones in which I am not. Therefore, all the dishes that need my work as a chef must be priced higher than the others.
- Observation 2: the dishes that the barbecuer is involved in and that the grill is used have a different value than the frozen meals (freezer involved). In general, the frozen meals are much cheaper.

Note that the observations made by the chef involved the same characteristic of more than one group or different characteristics of the same group, as highlighted below:

Group 1	
Involved Professionals	Required equipments
<b>Chef</b>	Stove
Cooker	Oven
Assistant 1	Processor
Assistant 2	Fridge

Group 2	
Involved Professionals	Required equipments
<b>Chef</b>	Stove
Barbecuer	Oven
Assistant 1	Freezer
Assistant 2	Blowtorch

Group 3	
Involved Professionals	Required equipments
Cooker	Oven
Assistant 1	Blowtorch
Assistant 2	Freezer
	Mixer

Group 4	
Involved Professionals	Required equipments
Cooker	Oven
Barbecuer	<b>Grill</b>
Assistant 1	Processor
	<b>Freezer</b>

The observations made by the chef tell us that we need to "teach" the algorithm to do two things to improve the quality of the groups:

- 1) assure that all the dishes prepared by the chef are assigned to the same group;
- 2) assure that dishes involving the grill and dishes involving the freezer are assigned to different groups.

Now transfer this analysis strategy to the case of designing an incident completion time predictor system and click "NEXT" to solve this task.

Task 1:  
Improving the  
characterization  
of the groups

The algorithm created five groups of incidents using information about state, priority and category. The groups are presented bellow. To interpret them, consider that:

- The "state" that the incidents assumed during its lifecycle are: new, active, resolved, closed, awaiting evidence, awaiting problem, awaiting user info e awaiting vendor. The presentation order of these states below are not necessarily correspondent to their occurrence order in the lifecycle of the incidents. A state is presented in a group whenever at least one of its incidents passed through this state. Also, the states are also indicated only once even if more than one incident of the group passed through it.

- The "priority" values that the incidents were registered in the incident management system are: 1 - Critical, 2 - High, 3 - Moderate e 4 – Low. The same order is used below to present the priority information of each group.

- There are 26 values for "category". We presented below only the first ten most frequent categories in each group. The category information is presented in alphabetical order.

All the other information on the incidents were not considered by the algorithm (as subcategory, location, symptom, impact, urgency).

Grupo 1		
Incident_state	Priority	Category
New	1 - Critical	Desktop
Active	2 - High	Internet
Resolved	3 - Moderate	Notebook
Closed	4 - Low	Passwords and access
Awaiting Evidence		Printer
Awaiting Problem		SAP ECC
Awaiting User Info		SAP ECC EN
Awaiting Vendor		Server
		Software - industry
		Telephony

Grupo 2		
Incident_state	Priority	Category
New	1 - Critical	Others
Active	2 - High	Passwords and access
Resolved	3 - Moderate	Remote access
Closed	4 - Low	SAP
Awaiting Evidence		SAP CRM
Awaiting Problem		SAP ECC
Awaiting User Info		SAP ECC EN
Awaiting Vendor		SAP GRC
		Software
		Software - industry

Grupo 3		
Incident_state	Priority	Category
New	1 - Critical	Peripheral
Active	2 - High	Radio frequency antenna
Resolved	3 - Moderate	Software - industry
Closed	4 - Low	SAP
Awaiting Evidence		Telephony
Awaiting Problem		Passwords and access
Awaiting User Info		Internet
Awaiting Vendor		Remote access
		Infrastructure
		Network

Grupo 4		
Incident_state	Priority	Category
New	1 - Critical	Infrastructure
Active	2 - High	Internet
Resolved	3 - Moderate	Network
Closed	4 - Low	Passwords and access
		Peripheral
Awaiting Problem		Radio frequency antenna
Awaiting User Info		Remote access
Awaiting Vendor		SAP
		Software - industry
		Telephony

Grupo 5		
Incident_state	Priority	Category
New	1 - Critical	Desktop
Active	2 - High	Internet
Resolved	3 - Moderate	Office pack
Closed	4 - Low	Passwords and access
		Remote access
Awaiting Problem		SAP
Awaiting User Info		SAP CRM
Awaiting Vendor		SAP ECC EN
		Software
		Software - industry

### Your task:

Considering the information describing the incidents assigned to this groups, do as the chef and point out at least three observations to better organize the incidents and provide a more precise database to build our incident completion time predictors.

Remember that to improve our prediction system, the groups should be compound by incidents with similar profiles in terms of completion time.

You can submit more than three observations if you feel it is relevant. Order your observations from the most to the least important.

3. Observation 1: \*

---

---

---

---

---

4. Observation 2: \*

---

---

---

---

---

5. Observation 3: \*

---

---

---

---

---

6. Other observations:

---

---

---

---

---

Task 2:  
Improving  
the  
organization  
of the  
incidents in  
the groups

Consider those same five groups of incidents, but now with more detailed information about the "states" by which the incidents went through during their lifecycles.

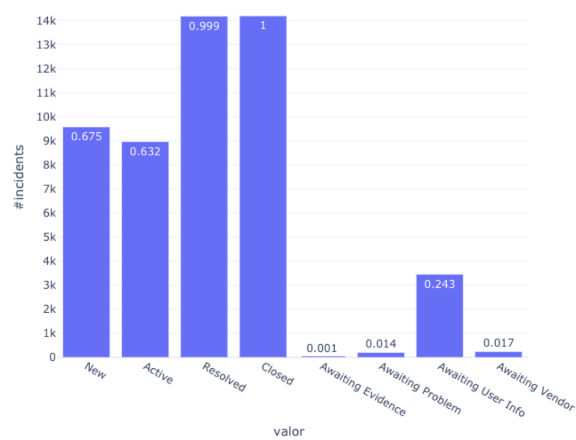
To help you to interpret these groups, observe the following characteristics in the graph below for the forth group (Grupo 4):

- none of the incidents went through the states "Resolved" e "Closed";
- lots of incidents (but not all of them) went through the states "New" e "Active" - even though it is not expected the existence of incidents that do not go through, for example, the state "New", this kind of situation is possible in the system registering these incidents;
- some of the incidents went through the state "Awaiting User Info";
- very few incidents went through the states "Awaiting Problem" and "Awaiting Vendor".

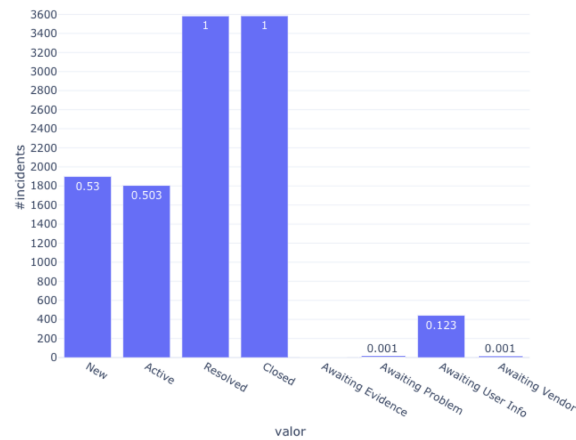
\* "Grupo" = "Group" and "Valor" = "Value"



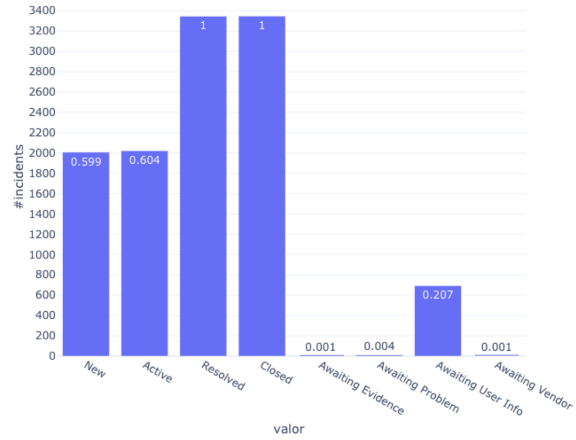
Grupo 1



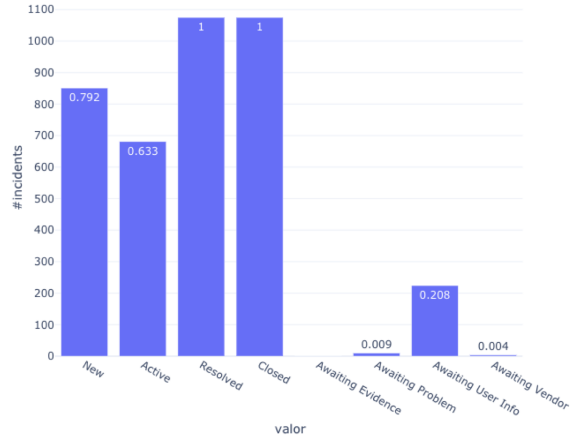
Grupo 2



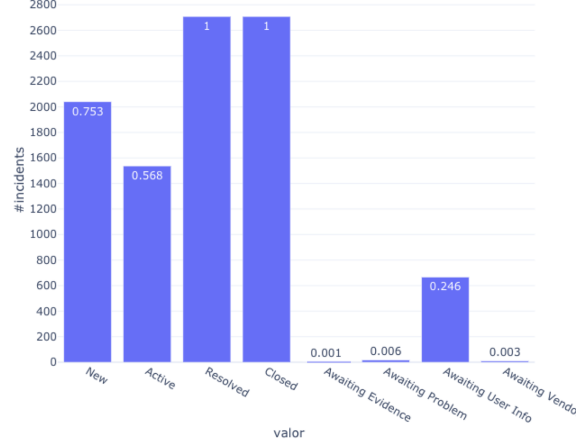
Grupo 3



Grupo 4



Grupo 5



For a better visualization of these graphs, use the links bellow:

- Group 1 (Grupo 1) -

[https://temporesolucaoincidentes.netlify.com/grupo%201/incident\\_state.html](https://temporesolucaoincidentes.netlify.com/grupo%201/incident_state.html)

- Group 2 (Grupo 2) -

[https://temporesolucaoincidentes.netlify.com/grupo%202/incident\\_state.html](https://temporesolucaoincidentes.netlify.com/grupo%202/incident_state.html)

- Group 3 (Grupo 3) -

[https://temporesolucaoincidentes.netlify.com/grupo%203/incident\\_state.html](https://temporesolucaoincidentes.netlify.com/grupo%203/incident_state.html)

- Group 4 (Grupo 4) -

[https://temporesolucaoincidentes.netlify.com/grupo%204/incident\\_state.html](https://temporesolucaoincidentes.netlify.com/grupo%204/incident_state.html)

- Group 5 (Grupo 5) -

[https://temporesolucaoincidentes.netlify.com/grupo%205/incident\\_state.html](https://temporesolucaoincidentes.netlify.com/grupo%205/incident_state.html)

Remember that the groups should gather incidents with similar profile in terms of completion time.

7. About the quantity of groups, you believe that:

*Mark only one oval.*

- ☐ more than five groups are necessary to appropriately separate the incidents in terms of resolution time similarity;
- ☐ five groups is a lot; it would be better to have fewer groups, as in fact the incidents are quite similar;
- ☐ five groups is the ideal number to accommodate incidents.

8. Explain your motivations for the choice above.

---

---

---

---

---

9. Evaluate the observations below in terms of how relevant they might be to improve the groups of incidents considering completion time similarity. Assign "0" to irrelevant observations, and rank those you think are relevant ("1" for low relevance; "2" for medium relevance; "3" for high relevance): \*

Mark only one oval per row.

	0	1	2	3
<b>(A) All incidents that passed through any of the "Awaiting..." states should be grouped together in a single group</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>(B) Incidents that did not go through the "Active" state and did not go through any "Awaiting..." state should be grouped together in a single group and separated from other incidents.</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>(C) Incidents that have passed the "Awaiting User Info" state should be in an isolated group from other incidents.</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. Explain your motivations for the choice above. In this space, you can also add other rules that you think are necessary to improve the distribution of incidents in groups. \*

---



---



---



---



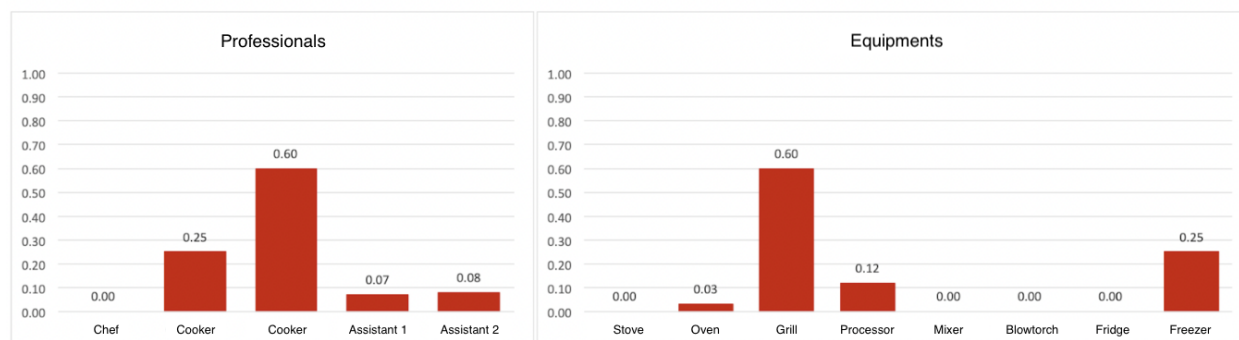
---

### Task 3: Individually analyzing the quality of each group

## Illustrative context

Back to our illustrative example of the served dishes at a restaurant. The graphs informing how the four groups created by the algorithm are presented bellow, describing each group by the features of the dishes (involved professionals and required equipments). A dish preparation may or may not include more than one professional or equipment.

In the group presented bellow, no dish required the chef and no dish required the use of stove, mixer, blowtorch and fridge. We can also note that, for example, 60% of the dishes assigned to this group required the barbecuer and 12% required the use of the processor.

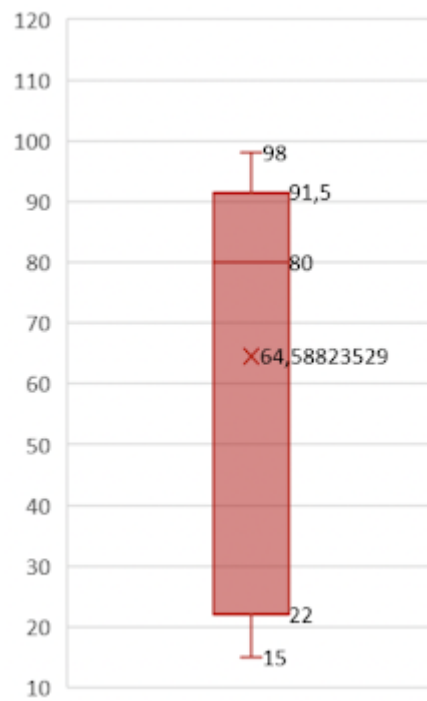


We showed this graphs to the chef and asked him which of the characteristics presented in this group (create by the algorithm) would work better if they were separated (in different groups instead of in the same group). Remember that the goal in this context is to group together dishes with similar prices.

He pointed out the following suggestions to better form the groups:

- 1) The dishes that involve both Assistant 1 and the Grill are more complex to be prepared and, hence, should be priced higher than the one that do not involve this resources. Therefore, it would be reasonable to separate this dishes from the other dishes assigned for this group.
- 2) The Assistant 2 is a professional allocated to prepare dishes of low complexity and low price. Therefore, the dishes involving the Assistant 1 or the Grill should not be assigned to the same group as the ones involving Assistant 2.

Then, we showed the chef the variation of the prices of the dishes assigned to this group (in the figure bellow, we can see the variation from \$15 to \$98) and we emphasize that the price range in this restaurant is from \$10 to \$120. We asked them to analyze again the situation and check if he could provide any other suggestion.



The chef analyzed the features of the groups and reaffirmed that the dishes involving Assistant 1 or the Grill have higher costs than the other and, hence, should be assigned to a particular group (exclusively formed by this dishes).

Now transfer this analysis strategy to the

Now transfer this analysis strategy to the case of designing a incident completion time predictor system and click "NEXT" to solve this task.

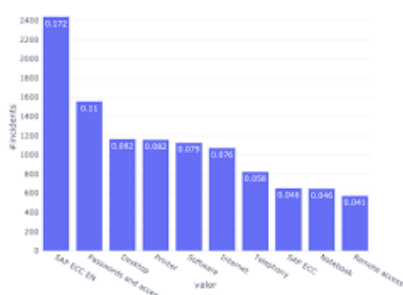
**Task 3:**  
Individually  
analyzing the  
quality of  
each group

Analyze the information about the incidents assigned to Group 1 (Grupo 1), described in terms of the frequency of the values of the features "category", "state", "priority".

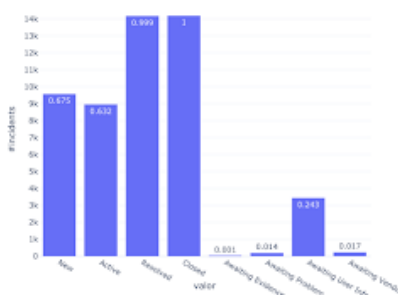
\* "Grupo" = "Group" and "Valor" = "Value"

## Group 1 (Grupo 1)

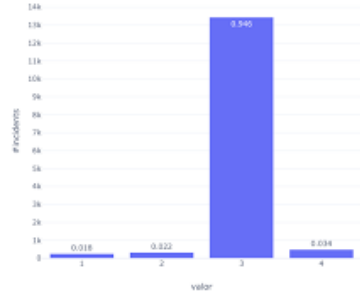
Grupo 1 - category



Grupo 1 - incident\_state



Grupo 1 - priority



For a better visualization of these graphs, use the links bellow:

- Category - <https://temporesolucaoincidentes.netlify.com/grupo%201/category.html>
- State - [https://temporesolucaoincidentes.netlify.com/grupo%201/incident\\_state.html](https://temporesolucaoincidentes.netlify.com/grupo%201/incident_state.html)
- Priority - <https://temporesolucaoincidentes.netlify.com/grupo%201/priority.html>

Which of the characteristics presented in this group (created by the algorithm) do you believe that would work better if they were separated (in different groups instead of in the same group)? Always remember that the goal in this context is to group together incidents that are similar in terms of completion time.

You can assemble suggestions similar to these formats:

- 1) priority "X" + estado "Y" should be always separated from category "Z".
- 2) state "A" should be always separated from category "B".

Another way of reading this suggestions is:

- 1) priority "X" + estado "Y" should be in a group that category "Z" is not present.
- 2) state "A" should be in a group that category "B" does not appear.

11. Suggestion 1 (the most important): \*

---

---

---

---

---

12. Suggestion 2:

---

---

---

---

---

13. Other suggestions (one per line, enumerate):

---

---

---

---

---

Now consider the information about the completion time of the incidents. Observe in the graph bellow that the completion time of the incidents assigned to Group 1 varies from 0 to 341 days (from a overall interval of 0 to 341 considering all the incidents in the event log):

Incident completion time (in days):



For a better visualization of these graphs, use the links bellow:

- <https://temporesolucaoincidentes.netlify.com/grupo%201/tempo%20de%20resolucao>

Remember that these incidents were grouped considering only "priority", "state" and "category", expecting that the incidents were indirectly grouped with potentially similar "completion time".

14. Observing this distribution of the completion time of the incidents assigned to this group, do you identify any issue regarding the assignment of these incidents to the same group?

*Mark only one oval.*

- ☐ No, I do not identify any issue about the way this group was formed.
- ☐ Yes, I believe that the variability of the completion time in this group is too high and I will point out bellow suggestions to decrease this variability, considering that the information that the algorithm that creates the groups is capable of using: "state", "priority" and "category".

15. Suggestion 1 (the most important):

---

---

---

---

---

16. Suggestion 2:

---

---

---

---

---

17. Other suggestions (one per line, enumerate):

---

---

---

---

---



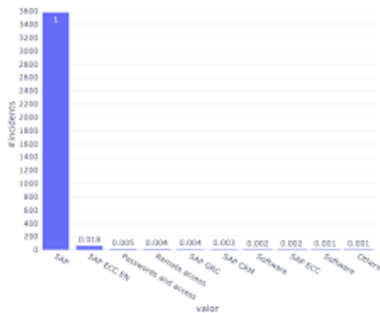
**Task 3:**  
Individually  
analyzing the  
quality of  
each group

Analyze the information about the incidents assigned to Group 2 (Grupo 2), described in terms of the frequency of the values of the features "category", "state", "priority".

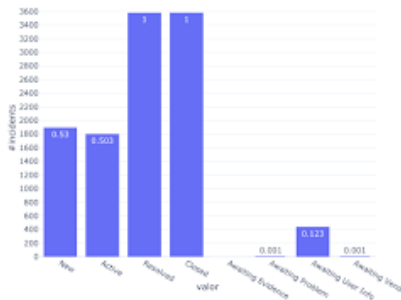
\* "Grupo" = "Group" and "Valor" = "Value"

**Group 2 (Grupo 2)**

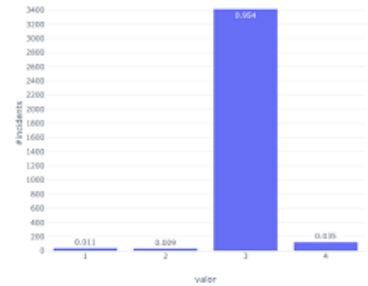
Grupo 2 - category



Grupo 2 - incident\_state



Grupo 2 - priority



For a better visualization of these graphs, use the links below:

- Category - <https://temporesolucaoincidentes.netlify.com/grupo%202/category.html>
- State - [https://temporesolucaoincidentes.netlify.com/grupo%202/incident\\_state.html](https://temporesolucaoincidentes.netlify.com/grupo%202/incident_state.html)
- Priority - <https://temporesolucaoincidentes.netlify.com/grupo%202/priority.html>

Which of the characteristics presented in this group (created by the algorithm) do you believe that would work better if they were separated (in different groups instead of in the same group)? Always remember that the goal in this context is to group together incidents that are similar in terms of completion time.

**18. Suggestion 1 (the most important): \***

---



---



---



---



---

19. Suggestion 2:

20. Other suggestions (one per line, enumerate):

Now consider the information about the completion time of the incidents. Observe in the graph bellow that the completion time of the incidents assigned to Group 2 varies from 0 to 330 days (from a overall interval of 0 to 341 considering all the incidents in the event log):

Incident completion time (in days):



For a better visualization of these graphs, use the links bellow:

- <https://temporesolucaoincidentes.netlify.com/grupo%202/tempo%20de%20resolucao>

Remember that these incidents were grouped considering only "priority", "state" and "category", expecting that the incidents were indirectly grouped with potentially similar "completion time".

21. Observing this distribution of the completion time of the incidents assigned to this group, do you identify any issue regarding the assignment of these incidents to the same group?

*Mark only one oval.*

- ☐ No, I do not identify any issue about the way this group was formed.
- ☐ Yes, I believe that the variability of the completion time in this group is too high and I will point out bellow suggestions to decrease this variability, considering that the information that the algorithm that creates the groups is capable of using: "state", "priority" and "category".

22. Suggestion 1 (the most important):

---

---

---

---

---

23. Suggestion 2:

---

---

---

---

---

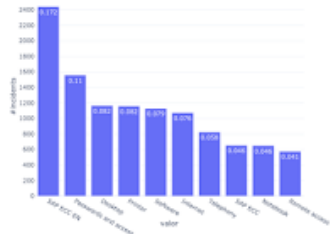
24. Other suggestions (one per line, enumerate):

Task 4:  
Analyzing the  
general  
clustering

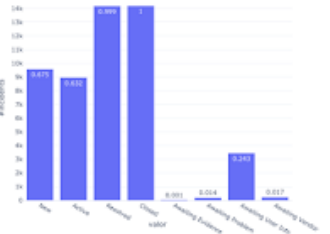
Now you can analyze the all the information on all the five groups at once:

\* "Grupo" = "Group", "Valor" = "Value", "Tempo de resolução de incidente (em dias)" = "Incident completion time (in days)".

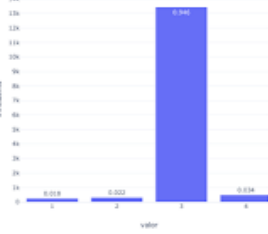
Grupo 1 - category



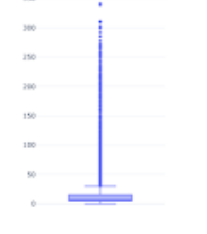
Grupo 1 - incident\_state



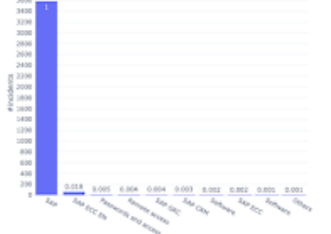
Grupo 1 - priority



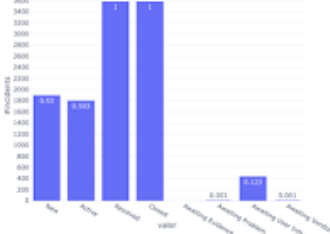
Tempo de resolução de incidente (em dias)



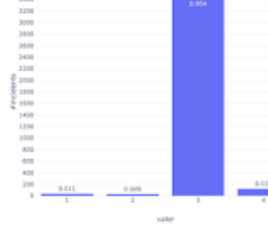
Grupo 2 - category



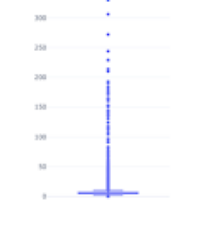
Grupo 2 - incident\_state



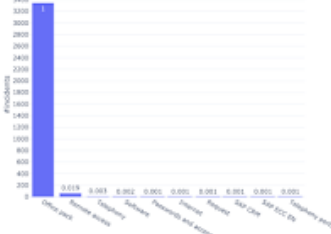
Grupo 2 - priority



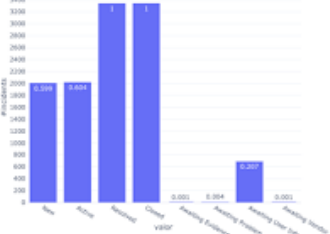
Tempo de resolução de incidente (em dias)



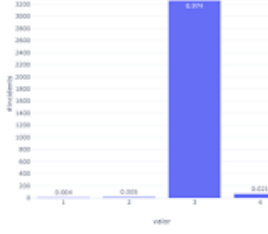
Grupo 3 - category



Grupo 3 - incident\_state



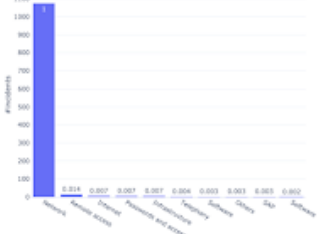
Grupo 3 - priority



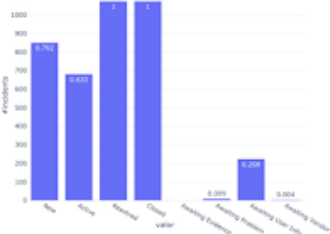
Tempo de resolução de incidente (em dias)



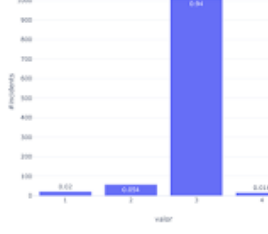
Grupo 4 - category



Grupo 4 - incident\_state



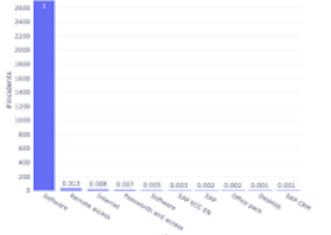
Grupo 4 - priority



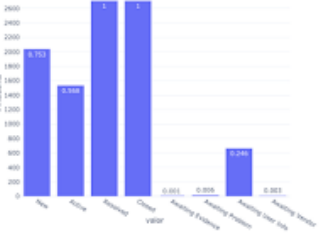
Tempo de resolução de incidente (em dias)



Grupo 5 - category



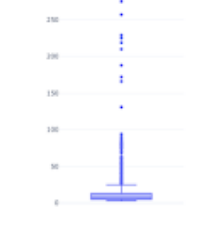
Grupo 5 - incident\_state



Grupo 5 - priority



Tempo de resolução de incidente (em dias)



25. Considering this overview of incidents in each group and the variations in incident completion time for each group, would you add or change any observations or suggestions about the way incidents were grouped? \*

---

---

---

---

---

Almost finished...

26. So that we can estimate the time it took you to carry out the proposed tasks, inform below the time when you are finishing the tasks: \*

---

*Example: 8:30 AM*

Thanks for your collaboration!

27. Use this space for suggestions, comments or adicional information.

---

---

---

---

---

In case you want to know more about this research, contact us:

- Thais Rodrigues Neubauer: [thais.neubauer@usp.br](mailto:thais.neubauer@usp.br)
- Sarajane Marques Peres: [sarajane@usp.br](mailto:sarajane@usp.br)

And do not forget to submit your answers!!! :-)

# Google Forms