Anlayzing IT incident and Event logs

Original Dataset (Source: Kaggle) :

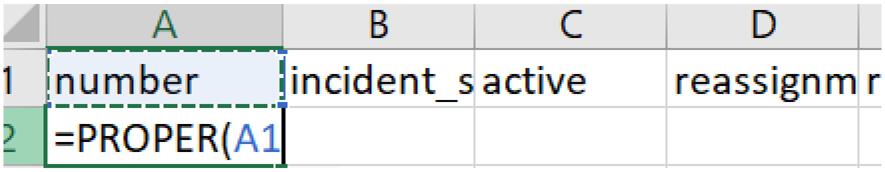
incident_event_log.csv

Data Exploration Review :

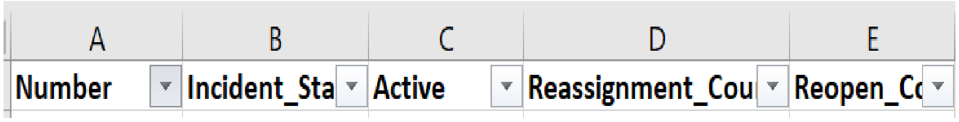
- Considering the Raw data posses same incident of all status, Analyzing only closed tickets to fetch appropriate insights.
- With the availble data, the goal is to identify tickets with Category which contributed high number of tickets inflow; Category where the incident is reported frequently and contributes more SLA Breach; Configuration Item which is causing IT incidents.

Data Cleaning :

Formating Header:



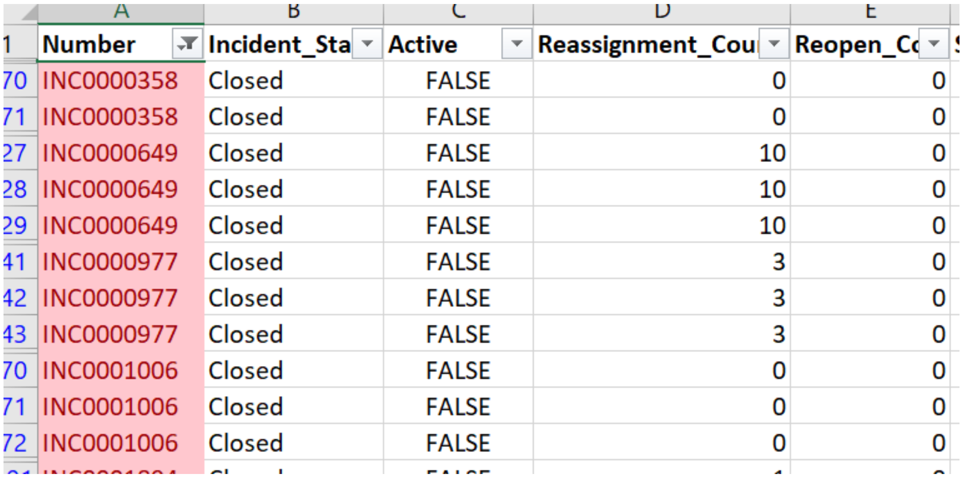
	A	B	C	D	
1	number	incident_s	active	reassignm	r
2	=PROPER(A1				



A	B	C	D	E
Number	Incident_Sta	Active	Reassignment_Cou	Reopen_Co

Removing Duplicates :

A closed cannot be re-opened and hence it cannot be closed twice. So removing duplicate Incidents.



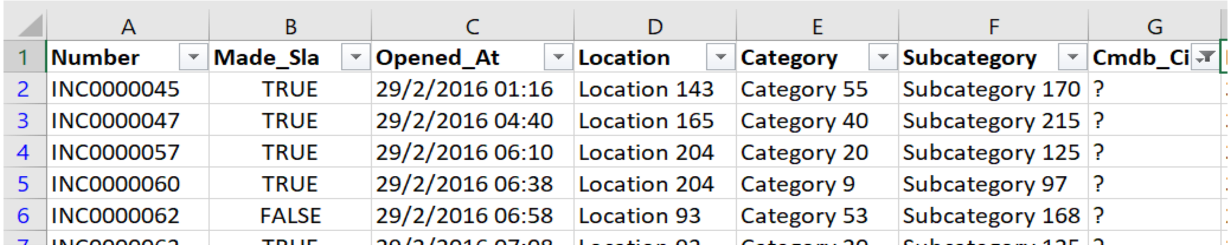
	A	B	C	D	E
1	Number	Incident_Sta	Active	Reassignment_Cou	Reopen_Cc
70	INC0000358	Closed	FALSE	0	0
71	INC0000358	Closed	FALSE	0	0
27	INC0000649	Closed	FALSE	10	0
28	INC0000649	Closed	FALSE	10	0
29	INC0000649	Closed	FALSE	10	0
41	INC0000977	Closed	FALSE	3	0
42	INC0000977	Closed	FALSE	3	0
43	INC0000977	Closed	FALSE	3	0
70	INC0001006	Closed	FALSE	0	0
71	INC0001006	Closed	FALSE	0	0
72	INC0001006	Closed	FALSE	0	0

Removing Data :

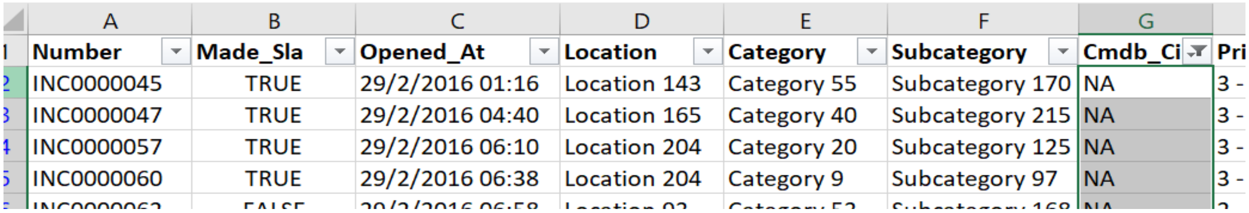
Removing unwanted data which is outside of scope, to improve performance and avoid confusions.

Data Correction :

Changing unavailable Configuration Item (Cmdb_Ci) '?' to NA to make it more precise. Technically, Application orineted tickets or user access issues may not own a CI. As the dataset lacks description and the exact issue for which the ticket raised for is unknown, it would be appropriate to formatize the unknown value than removing it.



	A	B	C	D	E	F	G
1	Number	Made_Sla	Opened_At	Location	Category	Subcategory	Cmdb_Ci
2	INC0000045	TRUE	29/2/2016 01:16	Location 143	Category 55	Subcategory 170	?
3	INC0000047	TRUE	29/2/2016 04:40	Location 165	Category 40	Subcategory 215	?
4	INC0000057	TRUE	29/2/2016 06:10	Location 204	Category 20	Subcategory 125	?
5	INC0000060	TRUE	29/2/2016 06:38	Location 204	Category 9	Subcategory 97	?
6	INC0000062	FALSE	29/2/2016 06:58	Location 93	Category 53	Subcategory 168	?



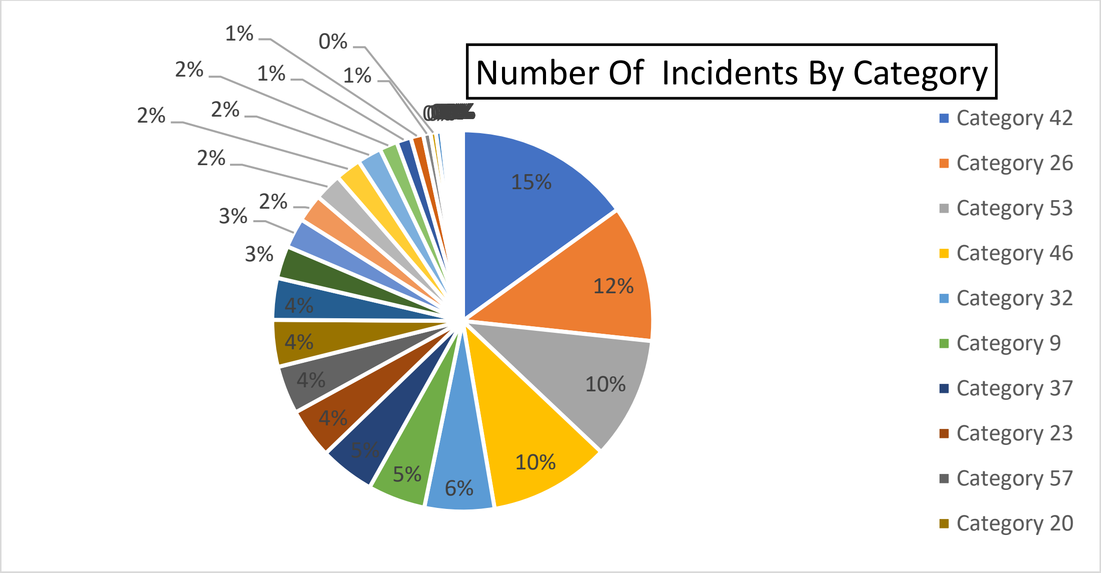
	A	B	C	D	E	F	G	
1	Number	Made_Sla	Opened_At	Location	Category	Subcategory	Cmdb_Ci	Pri
2	INC0000045	TRUE	29/2/2016 01:16	Location 143	Category 55	Subcategory 170	NA	3 -
3	INC0000047	TRUE	29/2/2016 04:40	Location 165	Category 40	Subcategory 215	NA	3 -
4	INC0000057	TRUE	29/2/2016 06:10	Location 204	Category 20	Subcategory 125	NA	3 -
5	INC0000060	TRUE	29/2/2016 06:38	Location 204	Category 9	Subcategory 97	NA	3 -

Cleaned Dataset:

incident_event_log_cleaned.csv

Category where the incident is reported frequently:

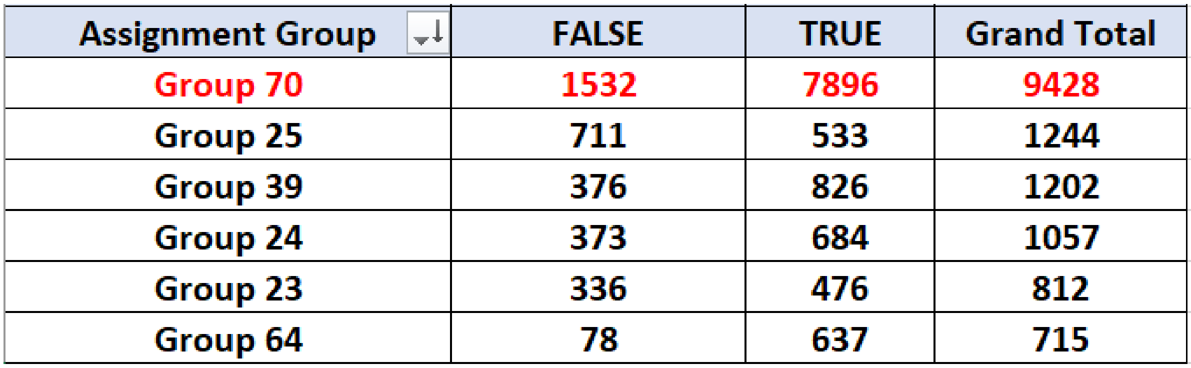
By looking at the below Pie chart, it seems that the Category 42 is the top contributor with about 15% of inflow under it, and Followed by it a high volume of incidents is obsererved under categories 26, 53, and 46. Due to limitation of data it is impossible to understand the category's exact description. However, in real-time it is worth analyzing this scenario and check for a possible automation.



Assignment Group that contributes more SLA Breach:

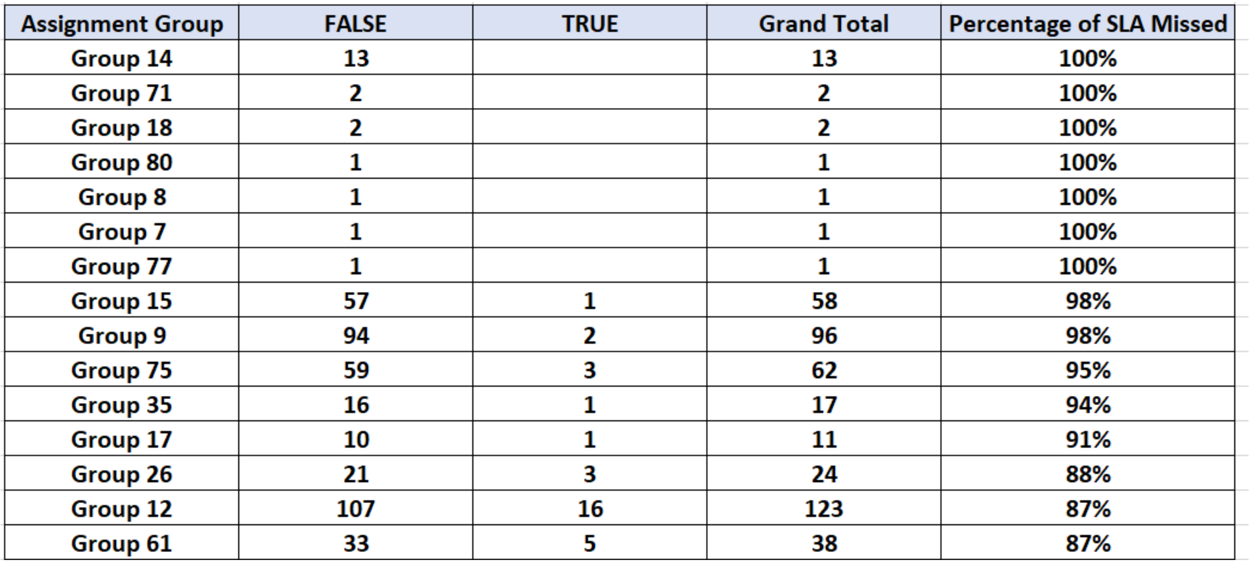
Depite the fact in summary 1 states that Group 70 is the Assignement group has the the highest number of SLA breaches, It is evident from the Summary 2 that, few groups have missed 100 % of the their SLA. Hence it is important to check with the support groups on the high SLA breach Percantage and conduct trainings to educate the importance of SLA to IT opearations. Also, the groups with high SLA breach count cannot be ignored and need to be analyzed further to understand the gap.

Summary 1 :



Assignment Group	FALSE	TRUE	Grand Total
Group 70	1532	7896	9428
Group 25	711	533	1244
Group 39	376	826	1202
Group 24	373	684	1057
Group 23	336	476	812
Group 64	78	637	715

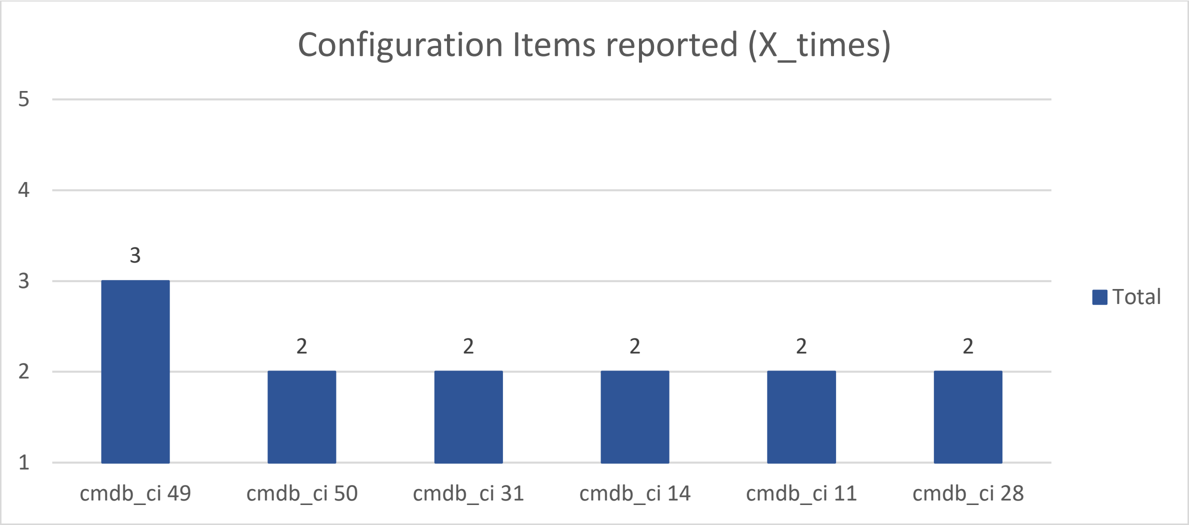
Summary 2:



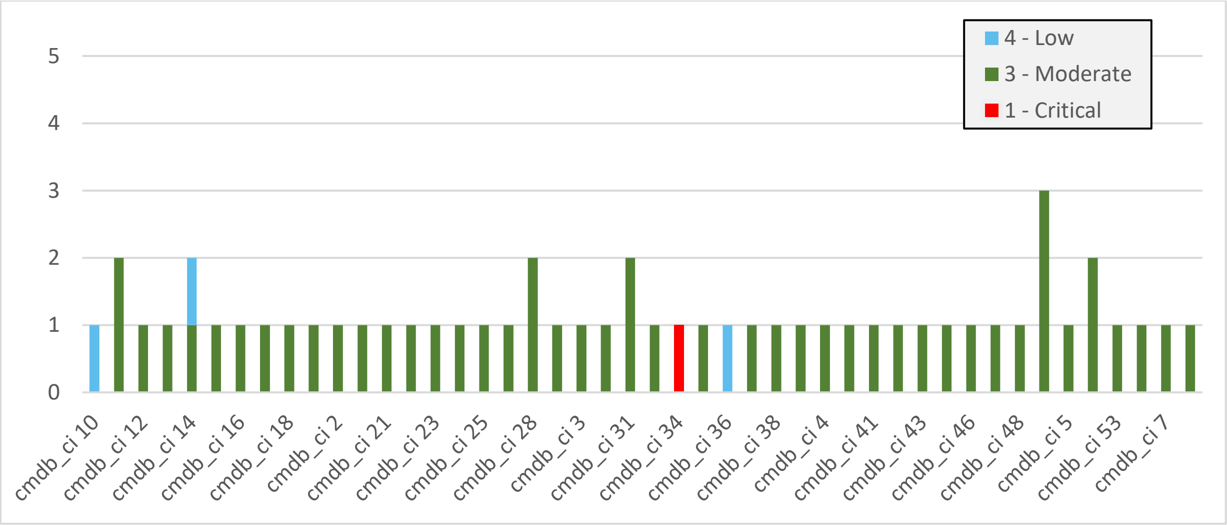
Assignment Group	FALSE	TRUE	Grand Total	Percentage of SLA Missed
Group 14	13		13	100%
Group 71	2		2	100%
Group 18	2		2	100%
Group 80	1		1	100%
Group 8	1		1	100%
Group 7	1		1	100%
Group 77	1		1	100%
Group 15	57	1	58	98%
Group 9	94	2	96	98%
Group 75	59	3	62	95%
Group 35	16	1	17	94%
Group 17	10	1	11	91%
Group 26	21	3	24	88%
Group 12	107	16	123	87%
Group 61	33	5	38	87%

Configuration Item which is causing IT incidents.

With the available data, it is identifed that Configuration item 49 in the CMBD is reported 3 times in one year, Despite the count is too low for a data spaned over a year, Let’s further analyze if there was any priority assosicated with the nodes with recursive reportings.



Configuration_X_IncidentPriroity :



It is evident from the above chart that, there is only one P1 incident among the reported Configuration items and the recursvive reports on Cis were more of P3 or low.