

코멘토 직무부트캠프_4주차

No. 1 데이터 소개

| Part |

✔ '4th_data'의 앞부분과 뒷부분의 일부를 확인해보면, 한 관측치에 대해 단계별 수행 기록을 나타낸다는 것을 볼 수 있으며, 'taskID' 변수의 값을 확인해 본 결과, 현재 데이터는 제품에 대한 A/S 수행 기록이 담긴 데이터를 의미한다.

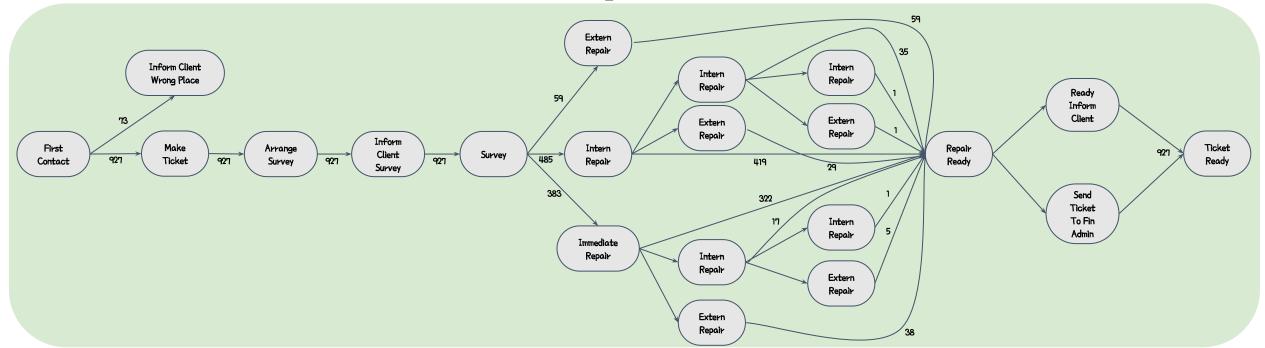
✔ 'originator' 변수를 통해 각 단계에 대한 담당자를 알 수 있으며, 사람이 담당하는 경우와 시스템이 담당하는 경우로 나뉘어져있다.

✔ 또한, 'eventtype'와 'timestamp' 변수를 통해, 각 단계가 언제 시작되고 언제 마무리되었는지 확인할 수 있다.

✔ 그 외, 나머지 변수들은 후에 데이터 탐색을 통해 자세히 알아볼 필요가 있다.

✔ 아래의 Process를 통해 단계의 진행 방식을 알 수 있으며, 처음에는 1000개의 관측치가 접속했지만, 접속 후 바로 73개의 관측치가 손실되었음을 확인할 수 있다.

[4th_data의 Process]



No. 2 데이터 문제점

[4th_data의 앞부분 일부]

						_	_				
caseID	taskID	originator	eventtype	timestamp	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK
1	FirstContact	Dian	complete	1970-01- 02 08:08	Phone	NaN	NaN	NaN	NaN	NaN	NaN
1	MakeTicket	Dian	start	1970-01- 02 08:08	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	MakeTicket	Dian	complete	1970-01- 02 08:11	NaN	Е	1340.0	NaN	NaN	NaN	NaN
1	ArrangeSurvey	Dian	start	1970-01- 02 08:11	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	ArrangeSurvey	Dian	complete	1970-01- 02 08:16	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	InformClientSurvey	System	complete	1970-01- 02 08:16	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Survey	Cindy	start	1970-01- 11 21:33	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Survey	Cindy	complete	1970-01- 11 21:56	NaN	NaN	NaN	True	240.0	1.0	NaN
1	InternRepair	John	start	1970-01- 17 04:36	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	InternRepair	John	complete	1970-01- 17 08:12	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	RepairReady	System	complete	1970-01- 17 08:12	NaN	NaN	NaN	NaN	NaN	NaN	True
1	SendTicketToFinAdmin	System	complete	1970-01- 17 14:03	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	ReadyInformClient	System	complete	1970-01- 17 15:44	NaN	NaN	NaN	NaN	NaN	NaN	True
1	TicketReady	System	complete	1970-01- 17 15:44	NaN	NaN	NaN	NaN	NaN	NaN	NaN

✔ 왼쪽의 그림을 통해, 'taskID' 변수의 단계마다 특정 변수에 대한 정보를 얻을 수 있음을 확인할 수 있다.

FirstContact -> contact MakeTicket -> RepairType, objectKey

 $\textbf{Survey} \rightarrow \texttt{RepairInternally, EstimatedRepairTime, RepairCode} \quad \textbf{RepairReady} \rightarrow \texttt{RepairOK}$

ReadyInformClient -> RepairOK

✔ 'eventtype'와 'timestamp' 변수를 통해, 각 단계가 언제 시작되고 언제 마무리되었는지 확인할 수 있으며, start와 complete로 나뉘어 있는 경우와 complete만 있는 경우가 있음을 확인할 수 있다.

✔ 각 단계에 대한 담당자가 지정되면, 해당 단계가 완료되기 전까지는 담당자가 바뀌지 않음을 확인할 수 있다.

위의 정보를 통해 현재 데이터에 다음과 같은 문제점이 있음을 확인할 수 있다.

- 1. 한 개의 case에 대해서 모든 단계와 변수에 대한 정보가 포함되어 있다. 이로 인해, 특정 단계에서 필요없는 변수임에도 값을 가져야 하는 현상이 발생해, case가 추가될 때 마다 Null 값이 저장된다는 문제점이 있다.
 - →> 각 단계마다 데이터를 따로 저장하여, Null 값이 저장되는 것을 방지하도록 하자.
- 현재 데이터에서 단계별 수행 기록을 한 개의 case에 대해서는 쉽게 확인할 수 있다는 장점이 있지만, 전체적인 관점에서 각 단계마다 case의 유입과 유출을 확인하기 어렵다는 문제점이 있다.
 - -> 1번과 동일
- 3. 각 단계에 start와 complete로 나뉘어져 있는 경우, 1번을 수행했을 때 중복된 값이 존재하고, 여전히 Null 값이 존재한다는 문제점이 있다.
 - -> 하나의 row에 StartTimeStamp와 CompleteTimeStamp를 저장함으로써, 중복과 Null을 방지하도록 하자.

No. 3 데이터 재구조화

1. FirstContact

	caseID	taskID	originator	contact	StartTime	CompleteTime
0	1	FirstContact	Dian	Phone	1970-01-02 08:08:00	1970-01-02 08:08:00
1	2	FirstContact	Dian	Personal	1970-01-08 05:17:00	1970-01-08 05:17:00
2	3	FirstContact	Dian	Phone	1970-01-03 01:03:00	1970-01-03 01:03:00
3	4	FirstContact	Monica	Personal	1970-01-03 08:23:00	1970-01-03 08:23:00
4	5	FirstContact	Monica	Letter	1970-01-07 20:41:00	1970-01-07 20:41:00

2-1. InformClientWrongPlace

	caseID	taskID	originator	StartTime	CompleteTime
0	6	InformClientWrongPlace	Monica	1970-01-03 02:29:00	1970-01-03 02:29:00
1	7	InformClientWrongPlace	Monica	1970-01-07 09:21:00	1970-01-07 09:21:00
2	13	InformClientWrongPlace	Monica	1970-01-02 12:22:00	1970-01-02 12:22:00
3	40	InformClientWrongPlace	Monica	1970-01-06 05:07:00	1970-01-06 05:07:00
4	61	InformClientWrongPlace	Monica	1970-01-06 12:40:00	1970-01-06 12:40:00

2-2, MakeTicket

	caseID	taskID	originator	RepairType	objectKey	StartTime	CompleteTime
0	1	MakeTicket	Dian	E	1340.0	1970-01-02 08:08:00	1970-01-02 08:11:00
1	2	MakeTicket	Dian	Р	8161.0	1970-01-08 05:17:00	1970-01-08 05:21:00
2	3	MakeTicket	Dian	Е	3844.0	1970-01-03 01:03:00	1970-01-03 01:10:00
3	4	MakeTicket	Monica	Р	6582.0	1970-01-03 08:23:00	1970-01-03 08:27:00
4	5	MakeTicket	Monica	Р	9857.0	1970-01-07 20:41:00	1970-01-07 20:47:00

3. ArrangeSurvey

	caseID	taskID	originator	StartTime	CompleteTime
0	1	ArrangeSurvey	Dian	1970-01-02 08:11:00	1970-01-02 08:16:00
1	2	ArrangeSurvey	Dian	1970-01-08 05:21:00	1970-01-08 05:25:00
2	3	ArrangeSurvey	Dian	1970-01-03 01:10:00	1970-01-03 01:14:00
3	4	ArrangeSurvey	Monica	1970-01-03 08:27:00	1970-01-03 08:32:00
4	5	ArrangeSurvey	Monica	1970-01-07 20:47:00	1970-01-07 20:52:00

4. InformClientSurvey

	caseID	taskID	originator	StartTime	CompleteTime
0	1	InformClientSurvey	System	1970-01-02 08:16:00	1970-01-02 08:16:00
1	2	InformClientSurvey	System	1970-01-08 05:25:00	1970-01-08 05:25:00
2	3	InformClientSurvey	System	1970-01-03 01:14:00	1970-01-03 01:14:00
3	4	InformClientSurvey	System	1970-01-03 08:32:00	1970-01-03 08:32:00
4	5	InformClientSurvey	System	1970-01-07 20:52:00	1970-01-07 20:52:00

5. Survey

	caseID	taskID	originator	RepairInternally	EstimatedRepairTime	RepairCode	StartTime	CompleteTime
0	1	Survey	Cindy	True	240.0	1.0	1970-01-11 21:33:00	1970-01-11 21:56:00
1	2	Survey	Anne	True	120.0	1.0	1970-01-12 04:19:00	1970-01-12 04:57:00
2	3	Survey	Paul	True	220.0	2.0	1970-01-06 06:31:00	1970-01-06 06:47:00
3	4	Survey	Anne	True	120.0	1.0	1970-01-04 17:03:00	1970-01-04 17:42:00
4	5	Survey	Anne	True	120.0	1.0	1970-01-09 22:35:00	1970-01-09 23:22:00

6. Intern & Immediate & Extern

	caseID	taskID	originator	StartTime	CompleteTime
0	1	InternRepair	John	1970-01-17 04:36:00	1970-01-17 08:12:00
1	2	ImmediateRepair	Anne	1970-01-12 04:57:00	1970-01-12 06:57:00
2	2	ExternRepair	Dolt	1970-01-12 06:57:00	1970-01-12 06:57:00
3	3	InternRepair	Paul	1970-01-06 19:50:00	1970-01-06 23:30:00
4	4	ImmediateRepair	Anne	1970-01-04 17:42:00	1970-01-04 19:42:00

7. RepairReady

	caseID	taskID	originator	RepairOK	StartTime	CompleteTime
0	1	RepairReady	System	True	1970-01-17 08:12:00	1970-01-17 08:12:00
1	2	RepairReady	System	True	1970-01-12 08:57:00	1970-01-12 08:57:00
2	3	RepairReady	System	True	1970-01-06 23:30:00	1970-01-06 23:30:00
3	4	RepairReady	System	True	1970-01-04 19:42:00	1970-01-04 19:42:00
4	5	RepairReady	System	True	1970-01-10 01:10:00	1970-01-10 01:10:00

8-1. SendTicketToFinAdmin

	caseID	taskID	originator	StartTime	CompleteTime
0	1	SendTicketToFinAdmin	System	1970-01-17 14:03:00	1970-01-17 14:03:00
1	2	SendTicketToFinAdmin	System	1970-01-12 15:14:00	1970-01-12 15:14:00
2	3	SendTicketToFinAdmin	System	1970-01-07 06:26:00	1970-01-07 06:26:00
3	4	SendTicketToFinAdmin	System	1970-01-04 23:35:00	1970-01-04 23:35:00
4	5	SendTicketToFinAdmin	System	1970-01-10 06:05:00	1970-01-10 06:05:00

8-2. ReadyInformClient

	caseID	taskID	originator	RepairOK	StartTime	CompleteTime
0	1	ReadyInformClient	System	True	1970-01-17 15:44:00	1970-01-17 15:44:00
1	2	ReadyInformClient	System	True	1970-01-12 11:17:00	1970-01-12 11:17:00
2	3	ReadyInformClient	System	True	1970-01-07 07:04:00	1970-01-07 07:04:00
3	4	ReadyInformClient	System	True	1970-01-04 20:56:00	1970-01-04 20:56:00
4	5	ReadyInformClient	System	True	1970-01-10 01:14:00	1970-01-10 01:14:00

9. TicketReady

	caseID	taskID	originator	StartTime	CompleteTime
0	1	TicketReady	System	1970-01-17 15:44:00	1970-01-17 15:44:00
1	2	TicketReady	System	1970-01-12 15:14:00	1970-01-12 15:14:00
2	3	TicketReady	System	1970-01-07 07:04:00	1970-01-07 07:04:00
3	4	TicketReady	System	1970-01-04 23:35:00	1970-01-04 23:35:00
4	5	TicketReady	System	1970-01-10 06:05:00	1970-01-10 06:05:00

- ✔ 두 번째 슬라이드의 Process를 참고하여, 크게 9가지의 단계로 구분하여 각 단계에 대한 데이터를 추출하였다.
- ✓ 각 단계마다 공통으로 들어가 있는 변수는 'caseID', 'taskID', 'originator', 'StartTime', 'CompleteTime' 이며, 특정 단계에서 다른 변수에 대한 정보를 얻을 수 있는 경우 해당 변수를 추가해주었다.
- ✔ 각 단계에 'eventtype'과 'timestamp' 변수를 삭제하고, 'StartTime', 'CompleteTime' 두 변수를 생성해줌으로써, 시작 시간과 마무리 시간에 대한 정보를 얻을 수 있도록 하였다.
- ✔ 6번의 경우 InternRepair, ImmediateRepair, ExternRepair 세 단계가 포함되어있지만, 세 단계 모두 Repair을 의미하므로 하나로 묶어주었다.

No. 4 데이터 재구조화

[6번 Repair 단계의 일부분]

	caseID	taskID	originator	StartTime	CompleteTime
170	153	InternRepair	Anne	1970-01-07 05:55:00	1970-01-07 07:43:00
171	154	InternRepair	Paul	1970-01-08 00:25:00	1970-01-08 04:49:00
172	155	ImmediateRepair	Jacky	1970-01-02 11:38:00	1970-01-02 13:26:00
173	155	InternRepair	Jacky	1970-01-02 16:23:00	1970-01-02 18:11:00
174	157	InternRepair	John	1970-01-04 14:42:00	1970-01-04 19:06:00
175	158	InternRepair	Edd	1970-01-16 15:12:00	1970-01-16 19:14:00
176	160	InternRepair	Paul	1970-01-09 13:58:00	1970-01-09 17:34:00
177	160	InternRepair	John	1970-01-11 06:43:00	1970-01-11 10:43:00
178	161	InternRepair	Edd	1970-01-01 18:56:00	1970-01-01 22:36:00
179	162	ImmediateRepair	Anne	1970-01-09 03:44:00	1970-01-09 05:32:00
180	163	InternRepair	Cindy	1970-01-13 07:08:00	1970-01-13 11:32:00
181	163	ExternRepair	Dolt	1970-01-13 11:32:00	1970-01-13 11:32:00
182	164	InternRepair	Jacky	1970-01-12 19:28:00	1970-01-12 23:28:00
183	165	ImmediateRepair	Anne	1970-01-10 21:38:00	1970-01-10 23:26:00
184	166	InternRepair	Eric	1970-01-04 13:18:00	1970-01-04 18:06:00
185	167	InternRepair	Eric	1970-01-04 05:25:00	1970-01-04 09:25:00

1. Intern

		caseID	taskID	originator	StartTime	CompleteTime
	0	1	InternRepair	John	1970-01-17 04:36:00	1970-01-17 08:12:00
	3	3	InternRepair	Paul	1970-01-06 19:50:00	1970-01-06 23:30:00
1	20	20	InternRepair	Anne	1970-01-12 06:48:00	1970-01-12 10:48:00
	24	24	InternRepair	Cindy	1970-01-15 11:16:00	1970-01-15 14:52:00
	26	26	InternRepair	John	1970-01-06 23:44:00	1970-01-07 03:20:00

2. Immediate

		caseID	taskID	originator	StartTime	CompleteTime
	4	4	ImmediateRepair	Anne	1970-01-04 17:42:00	1970-01-04 19:42:00
	5	5	ImmediateRepair	Anne	1970-01-09 23:22:00	1970-01-10 01:10:00
•	6	8	ImmediateRepair	Jacky	1970-01-14 02:14:00	1970-01-14 04:14:00
	7	9	ImmediateRepair	Anne	1970-01-03 21:25:00	1970-01-03 23:37:00
	8	10	ImmediateRepair	Jacky	1970-01-11 05:09:00	1970-01-11 07:09:00

3. Extern

		caseID	taskID	originator	StartTime	CompleteTime
	16	17	ExternRepair	Dolt	1970-01-03 23:04:00	1970-01-03 23:04:00
/	22	22	ExternRepair	FixIt	1970-01-07 18:57:00	1970-01-07 18:57:00
	31	30	ExternRepair	Dolt	1970-01-07 18:24:00	1970-01-07 18:24:00
	33	32	ExternRepair	FixIt	1970-01-14 10:58:00	1970-01-14 10:58:00
	109	100	ExternRepair	FixIt	1970-01-01 23:43:00	1970-01-01 23:43:00

4. Intern -> Intern

	caseID	taskID	originator	StartTime	CompleteTime
10	12	InternRepair	Paul	1970-01-08 04:49:00	1970-01-08 09:13:00
11	12	InternRepair	Cindy	1970-01-08 18:10:00	1970-01-08 22:10:00
42	39	InternRepair	Paul	1970-01-18 21:37:00	1970-01-19 02:01:00
43	39	InternRepair	Edd	1970-01-19 13:23:00	1970-01-19 16:59:00
59	52	InternRepair	Cindy	1970-01-02 02:41:00	1970-01-02 02:41:00

5. Intern -> Extern

	caseID	taskID	originator	StartTime	CompleteTime
17	18	InternRepair	Cindy	1970-01-13 23:36:00	1970-01-14 03:12:00
18	18	ExternRepair	Dolt	1970-01-14 03:12:00	1970-01-14 03:12:00
29	29	InternRepair	Anne	1970-01-06 23:17:00	1970-01-07 02:53:00
30	29	ExternRepair	FixIt	1970-01-07 02:53:00	1970-01-07 02:53:00
57	51	InternRepair	Paul	1970-01-11 08:03:00	1970-01-11 11:43:00

6. Immediate -> Intern

	caseID	taskID	originator	StartTime	CompleteTime
36	35	ImmediateRepair	Barbara	1970-01-04 12:05:00	1970-01-04 13:53:00
37	35	InternRepair	Jacky	1970-01-04 19:53:00	1970-01-04 22:05:00
38	36	ImmediateRepair	Anne	1970-01-03 20:32:00	1970-01-03 22:20:00
39	36	InternRepair	Jacky	1970-01-03 23:50:00	1970-01-04 02:02:00
169	153	ImmediateRepair	Anne	1970-01-07 03:43:00	1970-01-07 05:55:00

7. Immediate -> Extern

	caseID	taskID	originator	StartTime	CompleteTime
1	2	ImmediateRepair	Anne	1970-01-12 04:57:00	1970-01-12 06:57:00
2	2	ExternRepair	Dolt	1970-01-12 06:57:00	1970-01-12 06:57:00
13	15	ImmediateRepair	Barbara	1970-01-06 06:24:00	1970-01-06 08:12:00
14	15	ExternRepair	Dolt	1970-01-06 08:12:00	1970-01-06 08:12:00
53	48	ImmediateRepair	Jacky	1970-01-12 09:49:00	1970-01-12 12:01:00

8. Intern -> Intern -> Intern

	caseID	taskID	originator	StartTime	CompleteTime
913	865	InternRepair	Paul	1970-01-06 15:50:00	1970-01-06 19:50:00
914	865	InternRepair	Cindy	1970-01-06 20:00:00	1970-01-06 00:57:00
915	865	InternRepair	Cindy	1970-01-06 20:00:00	1970-01-06 23:36:00

9. Intern -> Intern -> Extern

	caseID	taskID	originator	StartTime	CompleteTime
827	776	InternRepair	Eric	1970-01-03 16:54:00	1970-01-03 16:54:00
828	776	InternRepair	Eric	1970-01-03 16:54:00	1970-01-03 19:42:00
829	776	ExternRepair	FixIt	1970-01-03 19:42:00	1970-01-03 19:42:00

10. Immediate -> Intern -> Intern

	caseID	taskID	originator	StartTime	CompleteTime
545	512	ImmediateRepair	Anne	1970-01-10 21:08:00	1970-01-10 23:20:00
546	512	InternRepair	Barbara	1970-01-10 23:21:00	1970-01-11 01:33:00
547	512	InternRepair	Jacky	1970-01-11 02:26:00	1970-01-11 04:14:00

11. Immediate -> Intern -> Extern

	caseID	taskID	originator	StartTime	CompleteTime
44	41	ImmediateRepair	Anne	1970-01-01 17:35:00	1970-01-01 19:35:00
45	41	InternRepair	Jacky	1970-01-01 20:47:00	1970-01-01 22:47:00
46	41	ExternRepair	FixIt	1970-01-01 22:47:00	1970-01-01 22:47:00
83	456	ImmediateRepair	Anne	1970-01-08 00:41:00	1970-01-08 02:29:00
84	456	InternRepair	Anne	1970-01-08 02:56:00	1970-01-08 04:56:00

✔ 6번 Repair 단계의 경우 Intern, Immediate, Extern 3개의 유형이 존재했으며, 왼쪽의 그림을 통해 한 개의 case에 여러 유형이 포함되어 있음을 확인할 수 있다.

✔️ 6번 단계에 대해 자세히 살펴보면, 6번 단계는 또 다시 11 단계로 나뉘어짐을 확인할 수 있으며, 오른쪽 그림을 통해 11단계의 각 내용을 확인할 수 있다.

✔ 새롭게 나뉜 11단계에서 8번, 9번, 10번 경우는 한 개의 case만이 해당함을 확인할 수 있다.