

# 4주차 과제

김유진

2021.08.22

# 1. 데이터 기본 구조 및 현황 파악

수리업체의 수리업무 진행 관련 데이터

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	E	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
...	...	...	...	...	...	...	...	...	...	...	...
1000	InternRepair	John	complete	1970-01-19 18:34	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1000	RepairReady	System	complete	1970-01-19 18:34	NaN	NaN	NaN	NaN	NaN	NaN	True
1000	SendTicketToFinAdmin	System	complete	1970-01-20 01:11	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1000	ReadyInformClient	System	complete	1970-01-20 02:25	NaN	NaN	NaN	NaN	NaN	NaN	True
1000	TicketReady	System	complete	1970-01-20 02:25	NaN	NaN	NaN	NaN	NaN	NaN	NaN

총 13,262 rows x 12 columns

## 12개 열 정보 정리

caseID	수리 의뢰 ID – 총 1000개
TaskID	업무명
Originator	담당자 이름
Eventtype	업무 상태
Timestamp	시간
contact	연락 수단
RepairType	수리타입(P/E/B)
ObjectKey	??
RepairInternally	내부수리 여부(T/F)
EstimateRepairTime	수리 시간
RepairCode	수리 코드 (1,2,3,4)
RepairOK	수리 완료(T/nan)

# 1. 데이터 기본 구조 및 현황 파악

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	E	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
...	...	...	...	...	...	...	...	...	...	...	...
1000	InternRepair	John	complete	1970-01-19 18:34	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1000	RepairReady	System	complete	1970-01-19 18:34	NaN	NaN	NaN	NaN	NaN	NaN	True
1000	SendTicketToFinAdmin	System	complete	1970-01-20 01:11	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1000	ReadyInformClient	System	complete	1970-01-20 02:25	NaN	NaN	NaN	NaN	NaN	NaN	True
1000	TicketReady	System	complete	1970-01-20 02:25	NaN	NaN	NaN	NaN	NaN	NaN	NaN

총 13,262 rows x 12 columns

## TaskID 종류

FirstContact	최초 연락
MakeTicket	청구서생성
ArrangeSurvey	상담내용정의
InformClientSurvey	고객문의전산처리
Survey	상담진행
InternRepair	단순 수리
RepairReady	수리종료
SendTicketToFinAdmin	청구서발송 to 관리자
ReadyInformClient	고객정보준비
TicketReady	청구서송달
ImmediateRepair	당업체 직접 수리
ExternRepair	외부업체 수리
InformClientWrongPlace	고객의 잘못된 정보

## 2. 해당 업체 문제 파악

예) Case 하나씩 살펴보기

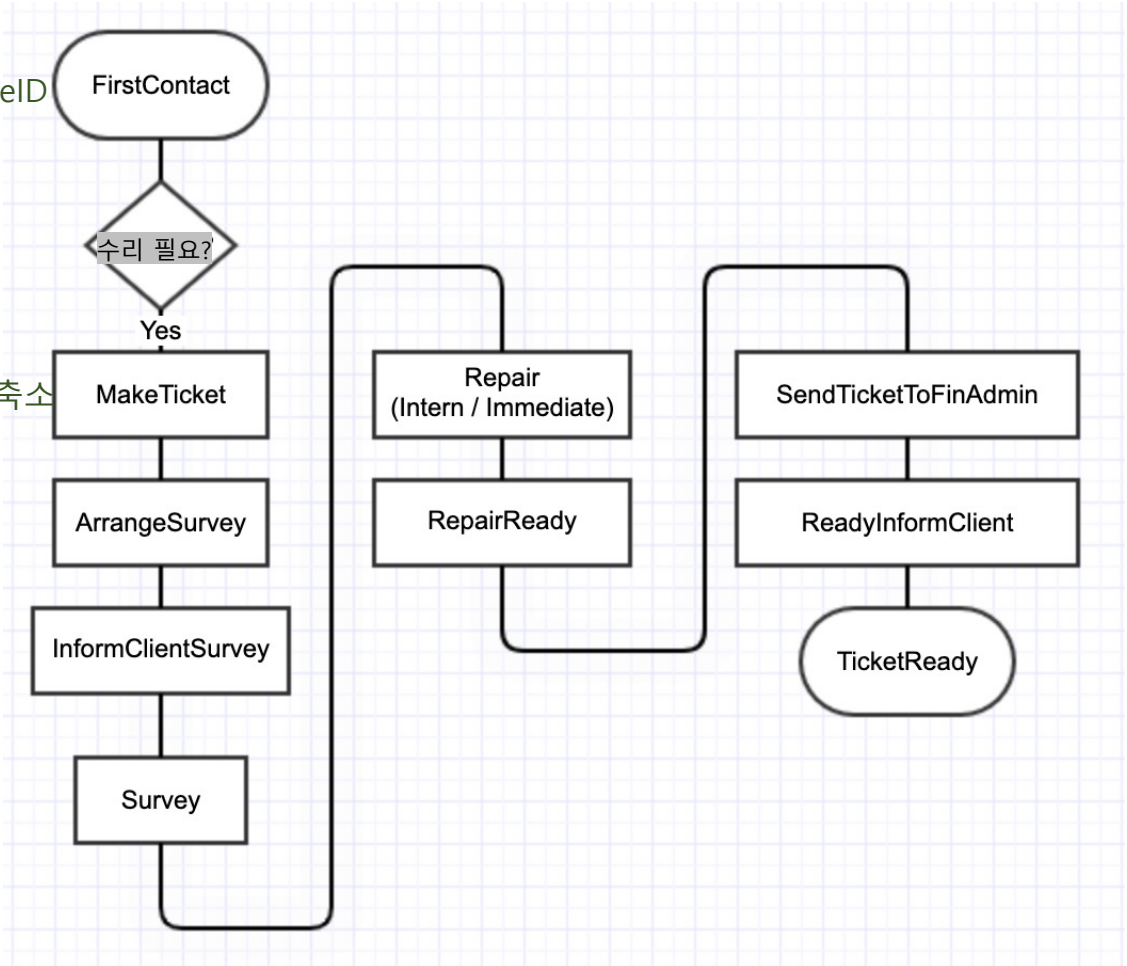
Case 1)

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	E	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

<taskID열을 이용한 수리업무 Flow>

총 1000개의 caseID

927개로 축소

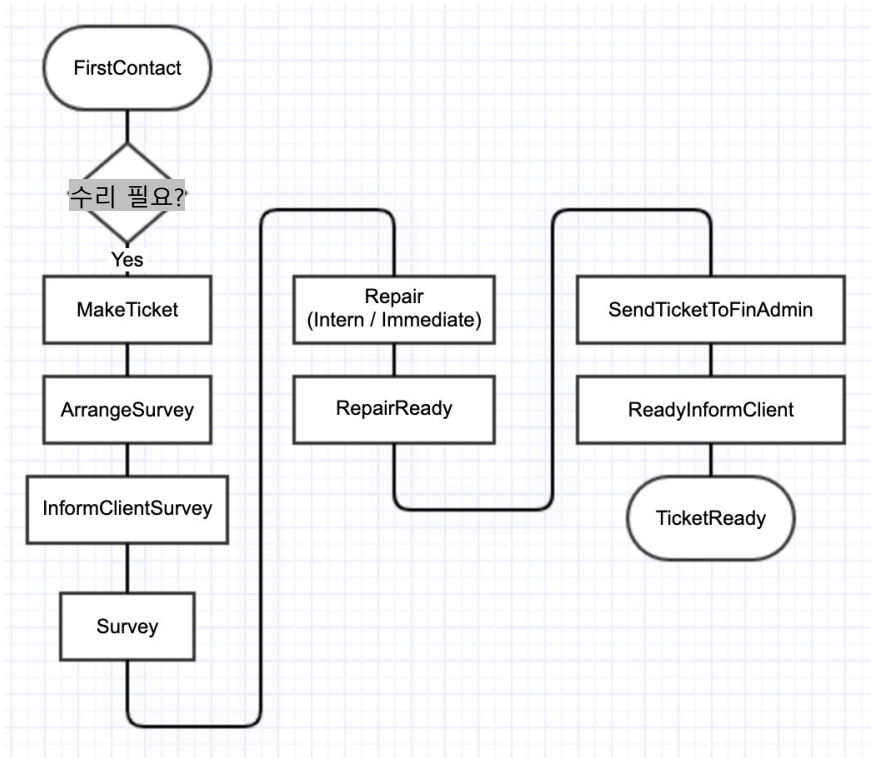


<수리 전 CS 관련 업무>

<수리 업무>

<수리 후 CS 관련 업무>

## 2. 해당 업체 문제 파악



### 문제점 1) 모든 업무에 대한 column을 동일하게 통일

: flow chart를 통해 업무의 종류를 구분할 수 있었으나, 데이터에서는 모든 업무가 한 데이터 프레임에 혼재

- > 업무에 따라 불필요한 column이 생김
- > 데이터에 null 값이 대부분을 차지해 공간 낭비
- > 업무별 비교가 어려움

예시) 전체 case는 1000가지이지만 모든 column을 공유하고 있어, MakeTicket과 같이 업무 소요 시간이 큰 비중을 차지 하지 않는 column에도 complete, start를 구분하기 위해 1000개보다 많은 row가 요구됨

eventtype	data[ 'taskID' ].value_counts()	
complete	ArrangeSurvey	1854
	MakeTicket	1854
start	Survey	1854
	InternRepair	1094
complete	FirstContact	1000
	InformClientSurvey	927
start	RepairReady	927
	SendTicketToFinAdmin	927
complete	ReadyInformClient	927
	TicketReady	927
...	ImmediateRepair	766
	ExternRepair	132
complete	InformClientWrongPlace	73
	Name: taskID, dtype: int64	
complete		
complete		
complete		
complete		

## 2. 해당 업체 문제 파악

caseID	taskID	originator	eventtype	timestamp	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK
2	FirstContact	Dian	complete	1970-01-08 05:17	Personal	NaN	NaN	NaN	NaN	NaN	NaN
2	MakeTicket	Dian	start	1970-01-08 05:17	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	MakeTicket	Dian	complete	1970-01-08 05:21	NaN	P	8161.0	NaN	NaN	NaN	NaN
2	ArrangeSurvey	Dian	start	1970-01-08 05:21	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	ArrangeSurvey	Dian	complete	1970-01-08 05:25	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	InformClientSurvey	System	complete	1970-01-08 05:25	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Survey	Anne	start	1970-01-12 04:19	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Survey	Anne	complete	1970-01-12 04:57	NaN	NaN	NaN	True	120.0	1.0	NaN
2	ImmediateRepair	Anne	start	1970-01-12 04:57	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	ImmediateRepair	Anne	complete	1970-01-12 06:57	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	ExternRepair	Dolt	start	1970-01-12 06:57	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	RepairReady	System	complete	1970-01-12 08:57	NaN	NaN	NaN	NaN	NaN	NaN	True
2	ReadyInformClient	System	complete	1970-01-12 11:17	NaN	NaN	NaN	NaN	NaN	NaN	True
2	SendTicketToFinAdmin	System	complete	1970-01-12 15:14	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	TicketReady	System	complete	1970-01-12 15:14	NaN	NaN	NaN	NaN	NaN	NaN	NaN

### 문제점 2) 불명확한 taskID

예) RepairReady는 수리 종료의 의미임에도, ready란 단어를 사용해서 혼란 (+Ticket Ready)

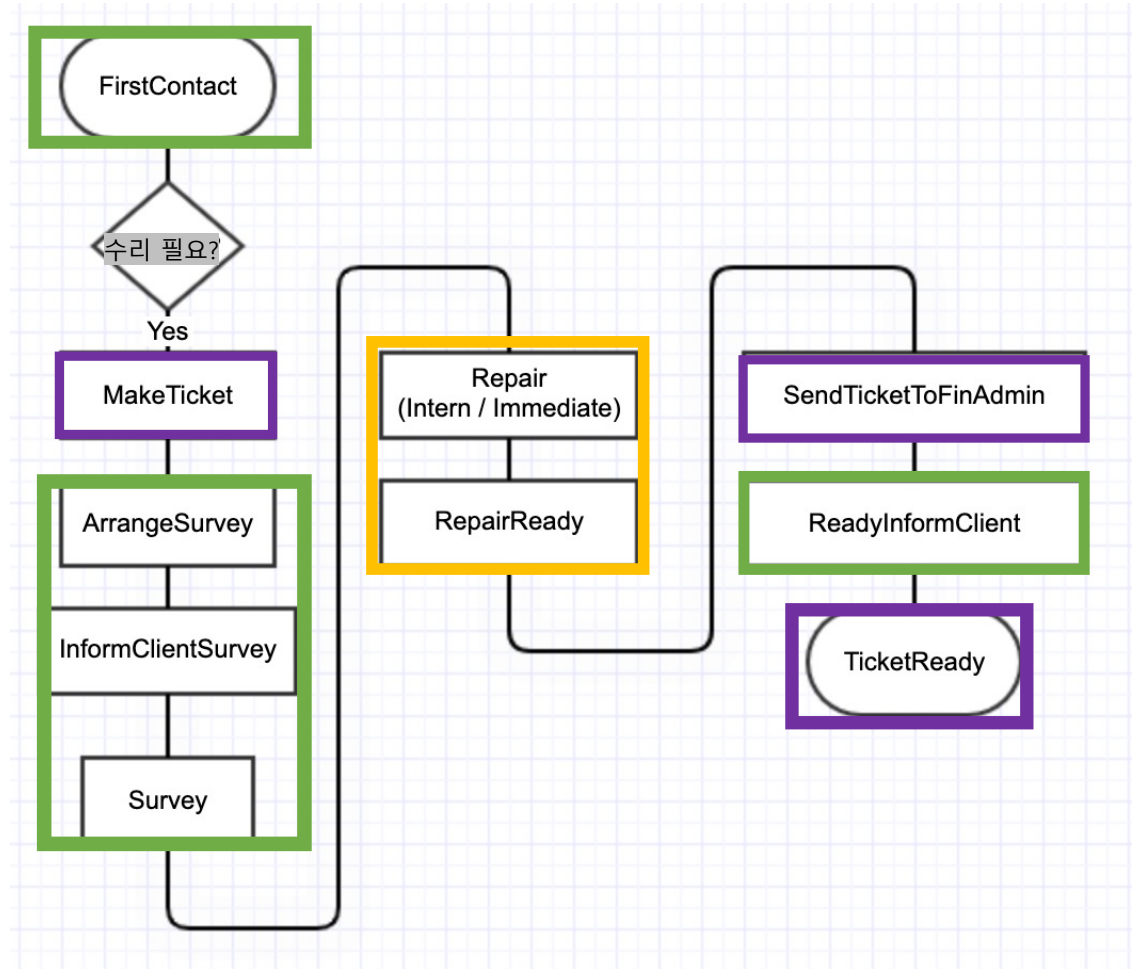
### 3. 해결 방법

업무를 세 종류로 명확히 구분

1. 고객 정보 관리 & 상담 업무

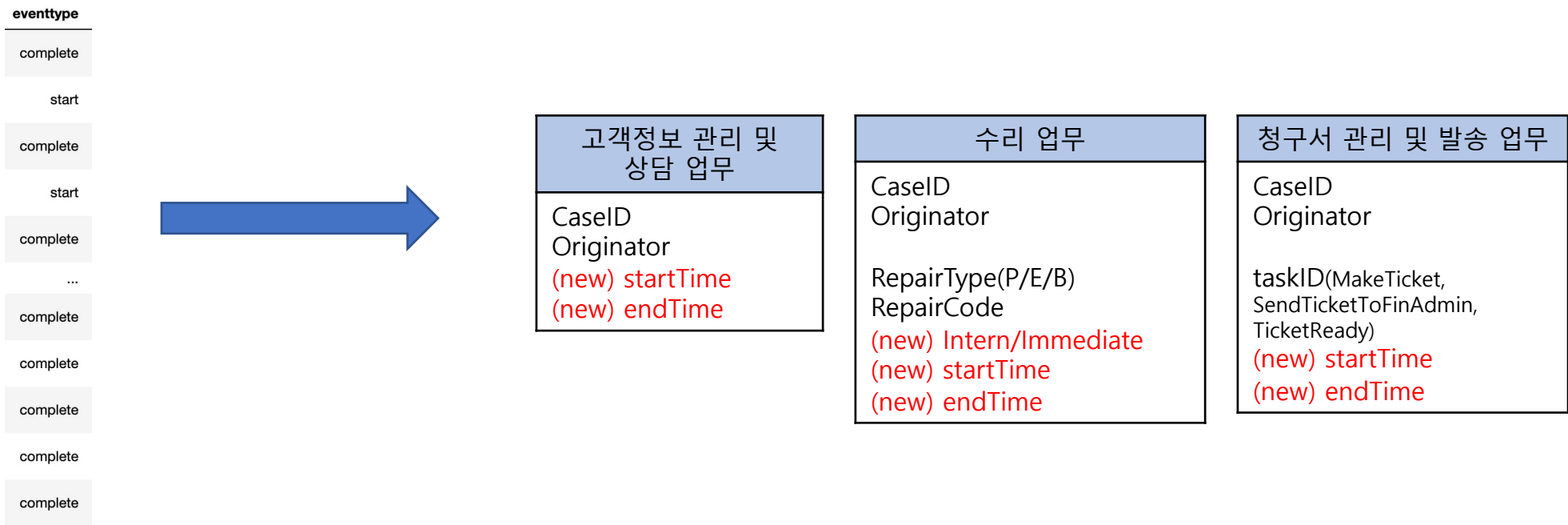
2. 직접 수리를 하는 업무

3. 청구서 관리 & 발송 업무



### 3. 해결 방법

- Eventtype column 대신, 업무 시작 시간과 종료 시간을 기록할 수 있는 column을 새로 구성(한 row에서 시간만 update하는 방식)
- 해당 데이터를 "자사 업무 기록"이란 주제에 맞게 외부 업체 관련 column을 제거하고, 수리 업무 종류를 Intern/Immediate로 구분할 수 있는 column추가





# 3. 해결 방법

확인한 문제점을 개선하기 위한 새로운 ERD 구성

