Data Analysis Report

of

The Processing of Home Repair Requests of
Netherlands Rental Housing Authority

Tran Thi Ngoc Diem

Outline

I. **Task Description**

II. **Data Analysis Report**

diemtit96@gmail.com

Outline

I. **Task Description**

II. **Data Analysis Report**

diemtit96@gmail.com

I. Task Description

1. Data Description:

- This data is the process data of the Netherlands Rental Housing Authority's processing of home repair requests, represented in Excel format.
- Each **caseID** refers to one home repair request. The repair process for each **caseID** composes of multiple sequential steps, each assigned by a unique **taskID**. The description for other columns is as follows:

caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time
419	InformClientSurvey	System	complete								1/6/2020	11:51
561	ArrangeSurvey	Monica	complete								1/7/2020	13:48
902	ReadyInformClient	System	complete							TRUE	1/8/2020	16:20
18	Survey	Edd	start								1/9/2020	6:55
852	Survey	Edd	complete				TRUE	240	1		1/10/2020	11:33
450	TicketReady	System	complete								1/11/2020	21:48
521	MakeTicket	Dian	complete		E	13					1/12/2020	20:30
89	Survey	Anne	start								1/13/2020	14:40
604	MakeTicket	Monica	start								1/14/2020	18:37
610	InternRepair	Cindy	start								1/15/2020	8:33
460	ArrangeSurvey	Dian	complete								1/16/2020	8:37
870	TicketReady	System	complete								1/17/2020	15:49
340	ArrangeSurvey	Dian	start								1/18/2020	23:18
516	FirstContact	Dian	complete	Phone							1/19/2020	12:40
325	RepairReady	System	complete							TRUE	1/20/2020	20:00
752	SendTicketToFinAdmin	System	complete								1/21/2020	13:38
65	Survey	Cindy	start								1/22/2020	4:02
447	FirstContact	Monica	complete	Phone							1/23/2020	22:19
863	SendTicketToFinAdmin	System	complete								1/24/2020	0:21
882	FirstContact	Dian	complete	Personal							1/25/2020	11:37
934	Survey	John	start								1/26/2020	11:25
100	ArrangeSurvey	Monica	complete								1/27/2020	23:21
897	MakeTicket	Dian	start								1/28/2020	4:09
533	TicketReady	System	complete								1/29/2020	22:14
724	ArrangeSurvey	Dian	start								1/30/2020	9:12
545	InformClientSurvey	System	complete								1/31/2020	5:59
712	ArrangeSurvey	Dian	start								2/1/2020	12:23
815	MakeTicket	Monica	start								2/2/2020	16:40

1. Categorical Variables:
- caseID: House repair request reception number.
- taskID: Task.
- originator: Task Representative.
- eventtype: Displays the start and complete of the task.
- contact: The request channel of the home repair request.
- RepairType: Repair method.
- objectKey: The house to be repaired.
- RepairInternally: Internal repair.
- RepairCode: Repair type.
- RepairOK: Repair normally terminated.
- date: Task performance date.
- time: Task performance time.

2. Continuous Variables:
- EstimatedRepairTime: Estimated repair time.

diemtit96@gmail.com

I. Task Description

2. Target Description:

- **Question 1**: Perform basic analysis and data preprocessing and explain the results
- **Question 2**: Detailed analysis and process analysis:
  - i. Describe how the repair process is currently progressing
  - ii. Analyze the problems of the current process (bottlenecks, repetitive tasks, etc.)
  - iii. If there is a data element that effects the time required for the case, present it and explain the rationale
- **Question 3**: Describe ways to improve the current process

diemtit96@gmail.com

Outline

I. Task Description

II. **Data Analysis Report**

diemtit96@gmail.com

II.1. Performance of basic analysis and data preprocessing and the explanation of the results

1. **Missing Value Analysis**
2. **Abnormal Data Observation**

diemtit96@gmail.com

II.1. Performance of basic analysis and data preprocessing and the explanation of the results

❑ Missing Value Explanation and How to fill missing value:

- ➢ This picture displays the information about the number of non-missing values, unique values, and missing values for each column

	non_nan	unique	missing
caseID	13262	1000	0
taskID	13260	13	2
originator	13260	16	2
eventtype	13257	2	5
contact	1000	4	12262
RepairType	927	3	12335
objectKey	927	890	12335
RepairInternally	927	2	12335
EstimatedRepairTime	927	7	12335
RepairCode	927	4	12335
RepairOK	1854	1	11408
date	13249	881	13
time	13249	1427	13
datetime	13249	6329	13

diemtit96@gmail.com

II.1. Performance of basic analysis and data preprocessing and the explanation of the results

	non_nan	unique	missing
caseID	13262	1000	0
taskID	13260	13	2
originator	13260	16	2
eventtype	13257	2	5
contact	1000	4	12262
RepairType	927	3	12335
objectKey	927	890	12335
RepairInternally	927	2	12335
EstimatedRepairTime	927	7	12335
RepairCode	927	4	12335
RepairOK	1854	1	11408
date	13249	881	13
time	13249	1427	13
datetime	13249	6329	13

1. Missing Value Explanation

- This dataset has 13262 observations (rows) for 1000 unique ***caseID.***
- For each *caseID*, these information are collected and inserted only once during the whole process: *contact*, *RepairType*, *objectKey*, *RepairInternally*, *EstimatedRepairTime*, *RepairCode*, (except *RepairOK* are inserted twice) => if all *caseID* are fully filled, there are only 1000 observations for each column (except *RepairOK* has 2000 observations)
- However, there are ***73 caseID*** that were notes as *InformClientWrongPlace*, thus those values mentioned above are not collected. Only *contact* task are collected for these cases (1000 values), and the rest only contains 927 values (1000 – 73), except RepairOK contains 2000-73*2=1854 values.
- **These calculated values matched what we observed from the dataset, therefore, despite having a lot of missing values, this is considered normal and will not be processed to filled or removed. The processing steps for other columns will be mentioned in the next slide.**

diemtit96@gmail.com

II.1. Performance of basic analysis and data preprocessing and the explanation of the results

	non_nan	unique	missing
caseID	13262	1000	0
taskID	13260	13	2
originator	13260	16	2
eventtype	13257	2	5
contact	1000	4	12262
RepairType	927	3	12335
objectKey	927	890	12335
RepairInternally	927	2	12335
EstimatedRepairTime	927	7	12335
RepairCode	927	4	12335
RepairOK	1854	1	11408
date	13249	881	13
time	13249	1427	13
datetime	13249	6329	13

2. How to fill missing value:

Because the whole process was conducted step-by-step based on taskID and taskID, originator, eventtype, date, and time are related to each other. Therefore, the rest of columns contain missing values, can be filled using some logical rules. For example:

- **taskID**: 2 rows. Since the taskID are processed step-by-step, and some taskID must exist once per workflow, these 2 values can be easily filled by summarizing unique values and fill the missing one.
- **originator**: 2 rows. Some specific tasks must be done by System, thus they can be easily filled if missing originator. Other than that, they can be filled if a taskID don't have start or completed originator. If both are missing, it can be randomly filled from 1 of 15 unique originator.
- **eventtype**: 5 rows. Since they contain only "start" or "completed", they can be filled easily from the processing order.
- Similar for **data** and **time**

diemtit96@gmail.com

II.1. Performance of basic analysis and data preprocessing and the explanation of the results

2. How to fill missing value:

Because the whole process was conducted step-by-step based on taskID and taskID, originator, eventtype, date, and time are related to each other. Therefore, the rest of columns contain missing values, can be filled using some logical rules. For example:

- **taskID**: 2 rows. Since the taskID are processed step-by-step, and some taskID must exist once per workflow, these 2 values can be easily filled by summarizing unique values and fill the missing one.

	caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
1950	15	FirstContact	Monica	complete	Phone	NaN	NaN	NaN	NaN	NaN	NaN	1/5/1970	9:21	1970-01-05 09:21:00
11469	15	MakeTicket	Monica	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/5/1970	9:21	1970-01-05 09:21:00
2479	15	MakeTicket	Monica	complete	NaN	P	7031.0	NaN	NaN	NaN	NaN	1/5/1970	9:23	1970-01-05 09:23:00
7058	15	ArrangeSurvey	Monica	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/5/1970	9:23	1970-01-05 09:23:00
3612	15	ArrangeSurvey	Monica	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/5/1970	9:26	1970-01-05 09:26:00
7785	15	NaN	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/5/1970	9:26	1970-01-05 09:26:00
3161	15	Survey	Barbara	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	5:27	1970-01-06 05:27:00
3393	15	ImmediateRepair	Barbara	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	6:24	1970-01-06 06:24:00
6263	15	Survey	Barbara	complete	NaN	NaN	NaN	True	120.0	1.0	NaN	1/6/1970	6:24	1970-01-06 06:24:00
4764	15	ImmediateRepair	Barbara	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	8:12	1970-01-06 08:12:00
13025	15	ExternRepair	Dolt	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	8:12	1970-01-06 08:12:00
7243	15	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/6/1970	10:00	1970-01-06 10:00:00
10668	15	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	16:13	1970-01-06 16:13:00
4709	15	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	16:23	1970-01-06 16:23:00
12085	15	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/6/1970	16:23	1970-01-06 16:23:00

II.1. Performance of basic analysis and data preprocessing and the explanation of the results

2. How to fill missing value:

Because the whole process was conducted step-by-step based on taskID and taskID, originator, eventtype, date, and time are related to each other. Therefore, the rest of columns contain missing values, can be filled using some logical rules. For example:

- **taskID**: 2 rows. Since the taskID are processed step-by-step, and some taskID must exist once per workflow, these 2 values can be easily filled by summarizing unique values and fill the missing one.

	caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
1950	15	FirstContact	Monica	complete	Phone	NaN	NaN	NaN	NaN	NaN	NaN	1/5/1970	9:21	1970-01-05 09:21:00
11469	15	MakeTicket	Monica	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/5/1970	9:21	1970-01-05 09:21:00
2479	15	MakeTicket	Monica	complete	NaN	P	7031.0	NaN	NaN	NaN	NaN	1/5/1970	9:23	1970-01-05 09:23:00
7058	15	ArrangeSurvey	Monica	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/5/1970	9:23	1970-01-05 09:23:00
3612	15	ArrangeSurvey	Monica	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/5/1970	9:26	1970-01-05 09:26:00
7785	15	InformClientSurvey	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/5/1970	9:26	1970-01-05 09:26:00
3161	15	Survey	Barbara	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	5:27	1970-01-06 05:27:00
3393	15	ImmediateRepair	Barbara	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	6:24	1970-01-06 06:24:00
6263	15	Survey	Barbara	complete	NaN	NaN	NaN	True	120.0	1.0	NaN	1/6/1970	6:24	1970-01-06 06:24:00
4764	15	ImmediateRepair	Barbara	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	8:12	1970-01-06 08:12:00
13025	15	ExternRepair	Dolt	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	8:12	1970-01-06 08:12:00
7243	15	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/6/1970	10:00	1970-01-06 10:00:00
10668	15	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	16:13	1970-01-06 16:13:00
4709	15	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/6/1970	16:23	1970-01-06 16:23:00
12085	15	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/6/1970	16:23	1970-01-06 16:23:00

II.1. Performance of basic analysis and data preprocessing and the explanation of the results

	non_nan	unique	missing
caseID	13262	1000	0
taskID	13260	13	2
originator	13260	16	2
eventtype	13257	2	5
contact	1000	4	12262
RepairType	927	3	12335
objectKey	927	890	12335
RepairInternally	927	2	12335
EstimatedRepairTime	927	7	12335
RepairCode	927	4	12335
RepairOK	1854	1	11408
date	13249	881	13
time	13249	1427	13
datetime	13249	6329	13

2. How to fill missing value:

Because the whole process was conducted step-by-step based on taskID and taskID, originator, eventtype, date, and time are related to each other. Therefore, the rest of columns contain missing values, can be filled using some logical rules. For example:

- **taskID**: 2 rows. Since the taskID are processed step-by-step, and some taskID must exist once per workflow, these 2 values can be easily filled by summarizing unique values and fill the missing one.
- **originator**: 2 rows. Some specific tasks must be done by System, thus they can be easily filled if missing originator. Other than that, they can be filled if a taskID don't have start or completed originator. If both are missing, it can be randomly filled from 1 of 15 unique originator.
- **eventtype**: 5 rows. Since they contain only "start" or "completed", they can be filled easily from the processing order.
- Similar for **data** and **time**

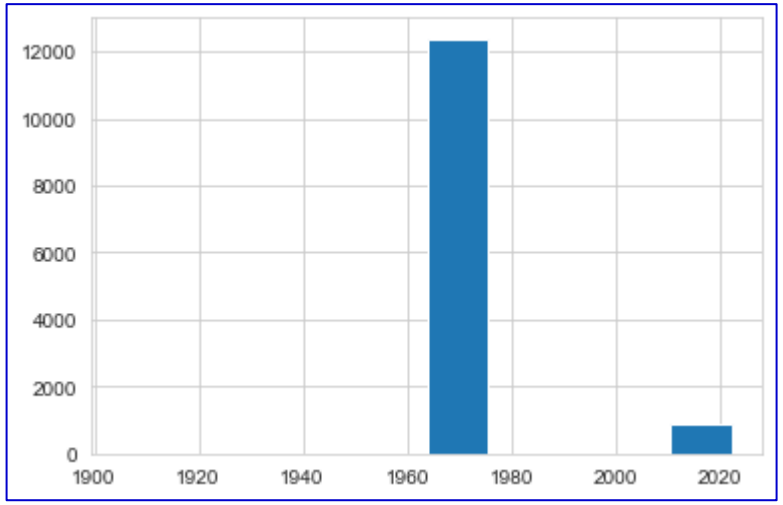
diemtit96@gmail.com

II.1. Performance of basic analysis and data preprocessing and the explanation of the results

- **Observations have abnormal data:** There are some requests (caseID), which have the abnormal datetime, most of observations is done in 1970, but some in 1905 or 2020, 2021, 2022

	caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
5903	1	FirstContact	Dian	complete	Phone	NaN	NaN	NaN	NaN	NaN	NaN	5/23/1905	0:00	1905-05-23 00:00:00
11087	1	MakeTicket	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:08	1970-01-02 08:08:00
4035	1	MakeTicket	Dian	complete	NaN	E	1340.0	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
5098	1	ArrangeSurvey	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
4561	1	ArrangeSurvey	Dian	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
9171	1	InformClientSurvey	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
5133	1	Survey	Cindy	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/11/1970	21:33	1970-01-11 21:33:00
3835	1	Survey	Cindy	complete	NaN	NaN	NaN	True	240.0	1.0	NaN	1/11/1970	21:56	1970-01-11 21:56:00
4078	1	InternRepair	John	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	4:36	1970-01-17 04:36:00
1994	1	InternRepair	John	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	8:12	1970-01-17 08:12:00
7465	1	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	8:12	1970-01-17 08:12:00
5325	1	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	14:03	1970-01-17 14:03:00
3676	1	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	15:44	1970-01-17 15:44:00
12047	1	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	15:44	1970-01-17 15:44:00

	caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
8008	2	MakeTicket	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/8/1970	5:17	1970-01-08 05:17:00
9321	2	FirstContact	Dian	complete	Personal	NaN	NaN	NaN	NaN	NaN	NaN	1/8/1970	5:17	1970-01-08 05:17:00
1073	2	MakeTicket	Dian	complete	NaN	P	8161.0	NaN	NaN	NaN	NaN	1/8/1970	5:21	1970-01-08 05:21:00
2563	2	ArrangeSurvey	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/8/1970	5:21	1970-01-08 05:21:00
2911	2	ArrangeSurvey	Dian	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/8/1970	5:25	1970-01-08 05:25:00
10505	2	InformClientSurvey	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/8/1970	5:25	1970-01-08 05:25:00
11800	2	Survey	Anne	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/12/1970	4:19	1970-01-12 04:19:00
3425	2	Survey	Anne	complete	NaN	NaN	NaN	True	120.0	1.0	NaN	1/12/1970	4:57	1970-01-12 04:57:00
5726	2	ImmediateRepair	Anne	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/12/1970	4:57	1970-01-12 04:57:00
6279	2	ExternRepair	Dolt	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/12/1970	6:57	1970-01-12 06:57:00
4582	2	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/12/1970	11:17	1970-01-12 11:17:00
3574	2	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/12/1970	15:14	1970-01-12 15:14:00
5056	2	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/12/1970	15:14	1970-01-12 15:14:00
80	2	ImmediateRepair	Anne	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3/26/2020	6:57	2020-03-26 06:57:00
495	2	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	5/15/2021	8:57	2021-05-15 08:57:00



- **How to deal**:
  1. Cases with abnormal dates can be removed
  2. We can rely on the time before and after each taskID to replace abnormal date and time

diemtit96@gmail.com

II.2.i. Describe how the repair process is currently progressing

The repair process will be conducted based on the order of **taskID** and operated by about 15 staffs at the Netherlands Rental Housing Authority and System. It has 2 processes, the **taskID** will be executed them in the following order:

I. **The normal repair process**

II. **In addition, in case the customer inputs the wrong address information**

diemtit96@gmail.com

II.2.i. Describe how the repair process is currently progressing

The repair process will be conducted based on the order of **taskID** and operated by about 15 staffs at the Netherlands Rental Housing Authority and System. It has 2 processes, the **taskID** will be executed them in the following order:

I. **The normal repair process and explanation:**

1. ***FirstContact:*** First, the clients make a request to the staff at the Netherlands Rental Housing Authority through 4 contact methods: ***Phone***, ***Web***, ***Letter***, ***Person***. The staff creates a reception number (**caseID**). Operated by staff and marked in **eventtype** as ***complete*** status, no ***start*** status.
2. ***MakeTicket***: Then, the staff makes a ticket for the request to confirm the repair method (**RepairType**: ***B, E, P***) and the house to be repaired (**objectKey**).
3. ***ArrangeSurvey***: The staff will arrange a time for people to come to survey the house.

caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
1	FirstContact	Dian	complete	Phone	NaN	NaN	NaN	NaN	NaN	NaN	5/23/1905	0:00	1905-05-23 00:00:00
1	MakeTicket	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:08	1970-01-02 08:08:00
1	MakeTicket	Dian	complete	NaN	E	1340.0	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	InformClientSurvey	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	Survey	Cindy	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/11/1970	21:33	1970-01-11 21:33:00
1	Survey	Cindy	complete	NaN	NaN	NaN	True	240.0	1.0	NaN	1/11/1970	21:56	1970-01-11 21:56:00
1	InternRepair	John	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	4:36	1970-01-17 04:36:00
1	InternRepair	John	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	8:12	1970-01-17 08:12:00
1	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	8:12	1970-01-17 08:12:00
1	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	14:03	1970-01-17 14:03:00
1	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	15:44	1970-01-17 15:44:00
1	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	15:44	1970-01-17 15:44:00

II.2.i. Describe how the repair process is currently progressing

The repair process will be conducted based on the order of **taskID** and operated by about 15 staffs at the Netherlands Rental Housing Authority and System. It has 2 processes, the **taskID** will be executed them in the following order:

I. **The normal repair process and explanation:**

4. ***InformClientSurvey***: Then the ***System*** will confirm the information to the client before going to client's house.

5. ***Survey***: Surveyor comes and checks the status of the house. After that, the surveyor will determine the repair type (**RepairCode: *1, 2, 3, 4***), whether the house needs to be repaired immediately, repaired internally (**RepairInternally: *True*** or ***False***), externally, or both (***ImmediateRepair***, ***ExternRepair***, or ***InternRepair***), and estimate the repair time (**EstimatedRepairTime**).

caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
1	FirstContact	Dian	complete	Phone	NaN	NaN	NaN	NaN	NaN	NaN	5/23/1905	0:00	1905-05-23 00:00:00
1	MakeTicket	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:08	1970-01-02 08:08:00
1	MakeTicket	Dian	complete	NaN	E	1340.0	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	InformClientSurvey	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	Survey	Cindy	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/11/1970	21:33	1970-01-11 21:33:00
1	Survey	Cindy	complete	NaN	NaN	NaN	True	240.0	1.0	NaN	1/11/1970	21:56	1970-01-11 21:56:00
1	InternRepair	John	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	4:36	1970-01-17 04:36:00
1	InternRepair	John	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	8:12	1970-01-17 08:12:00
1	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	8:12	1970-01-17 08:12:00
1	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	14:03	1970-01-17 14:03:00
1	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	15:44	1970-01-17 15:44:00
1	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	15:44	1970-01-17 15:44:00

II.2.i. Describe how the repair process is currently progressing

The repair process will be conducted based on the order of **taskID** and operated by about 15 staffs at the Netherlands Rental Housing Authority and System. It has 2 processes, the **taskID** will be executed them in the following order:

I. **The normal repair process and explanation:**

6. Repair Step **(*ImmediateRepair*, *ExternRepair*, or *InternRepair*):** Based on the survey provided in step 5, then **taskIDs*: ImmediateRepair*, *ExternRepair*, or *InternRepair*** will be conducted.

7. ***RepairReady***: After the repair staff is done, they will report to the system to check repair normally terminated (**RepairOK**), if it's okay, marked ***True.***

caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
1	FirstContact	Dian	complete	Phone	NaN	NaN	NaN	NaN	NaN	NaN	5/23/1905	0:00	1905-05-23 00:00:00
1	MakeTicket	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:08	1970-01-02 08:08:00
1	MakeTicket	Dian	complete	NaN	E	1340.0	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	InformClientSurvey	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	Survey	Cindy	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/11/1970	21:33	1970-01-11 21:33:00
1	Survey	Cindy	complete	NaN	NaN	NaN	True	240.0	1.0	NaN	1/11/1970	21:56	1970-01-11 21:56:00
1	InternRepair	John	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	4:36	1970-01-17 04:36:00
1	InternRepair	John	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	8:12	1970-01-17 08:12:00
1	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	8:12	1970-01-17 08:12:00
1	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	14:03	1970-01-17 14:03:00
1	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	15:44	1970-01-17 15:44:00
1	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	15:44	1970-01-17 15:44:00

II.2.i. Describe how the repair process is currently progressing

The repair process will be conducted based on the order of **taskID** and operated by about 15 staffs at the Netherlands Rental Housing Authority and System. It has 2 processes, the **taskID** will be executed them in the following order:

I. **The normal repair process and explanation:**

8. ***SendTicketToFinAdmin***: Send ticket to admin before informing to client.
9. ***ReadyInformClient***: After that, everything is okay, switching to ready to inform to client status.
10. ***TicketReady***: Make ***TicketReady*** is complete, the **caseID** is done.

caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
1	FirstContact	Dian	complete	Phone	NaN	NaN	NaN	NaN	NaN	NaN	5/23/1905	0:00	1905-05-23 00:00:00
1	MakeTicket	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:08	1970-01-02 08:08:00
1	MakeTicket	Dian	complete	NaN	E	1340.0	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	InformClientSurvey	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	Survey	Cindy	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/11/1970	21:33	1970-01-11 21:33:00
1	Survey	Cindy	complete	NaN	NaN	NaN	True	240.0	1.0	NaN	1/11/1970	21:56	1970-01-11 21:56:00
1	InternRepair	John	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	4:36	1970-01-17 04:36:00
1	InternRepair	John	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	8:12	1970-01-17 08:12:00
1	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	8:12	1970-01-17 08:12:00
1	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	14:03	1970-01-17 14:03:00
1	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	15:44	1970-01-17 15:44:00
1	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	15:44	1970-01-17 15:44:00

II.2.i. Describe how the repair process is currently progressing

The repair process will be conducted based on the order of **taskID** and operated by about 15 staffs at the Netherlands Rental Housing Authority and System. It has 2 processes, the **taskID** will be executed them in the following order:

I. **The normal repair process and explanation:**

**Note:*

i. **Step 2, 3, 5, 6:** Operated by ***staff***. Conducting process follows 2 steps, which are ***start*** and ***complete*** status should be marked in **eventtype**.

ii. **Step 4, 7, 8, 9, 10:** Operated by ***System***. Auto marked in **eventtype** as ***complete*** status, no ***start*** status

caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
1	FirstContact	Dian	complete	Phone	NaN	NaN	NaN	NaN	NaN	NaN	5/23/1905	0:00	1905-05-23 00:00:00
1	MakeTicket	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:08	1970-01-02 08:08:00
1	MakeTicket	Dian	complete	NaN	E	1340.0	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	InformClientSurvey	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	Survey	Cindy	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/11/1970	21:33	1970-01-11 21:33:00
1	Survey	Cindy	complete	NaN	NaN	NaN	True	240.0	1.0	NaN	1/11/1970	21:56	1970-01-11 21:56:00
1	InternRepair	John	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	4:36	1970-01-17 04:36:00
1	InternRepair	John	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	8:12	1970-01-17 08:12:00
1	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	8:12	1970-01-17 08:12:00
1	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	14:03	1970-01-17 14:03:00
1	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	15:44	1970-01-17 15:44:00
1	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	15:44	1970-01-17 15:44:00

II.2.i. Describe how the repair process is currently progressing

The repair process will be conducted based on the order of **taskID** and operated by about 15 staffs at the Netherlands Rental Housing Authority and System. It has 2 processes, the **taskID** will be executed them in the following order:

I. **The normal repair process and explanation:**

II. **In addition, in case the customer inputs the wrong address information, the repair process will be operated as follows:**

1. ***FirstContact***:
   - Operated by ***staff***
   - Marked in **eventtype** as ***complete*** status, no ***start*** status
2. ***InformClientWrongPlace***:
   - Operated by ***staff***
   - Marked in **eventtype** as ***complete*** status, no ***start*** status

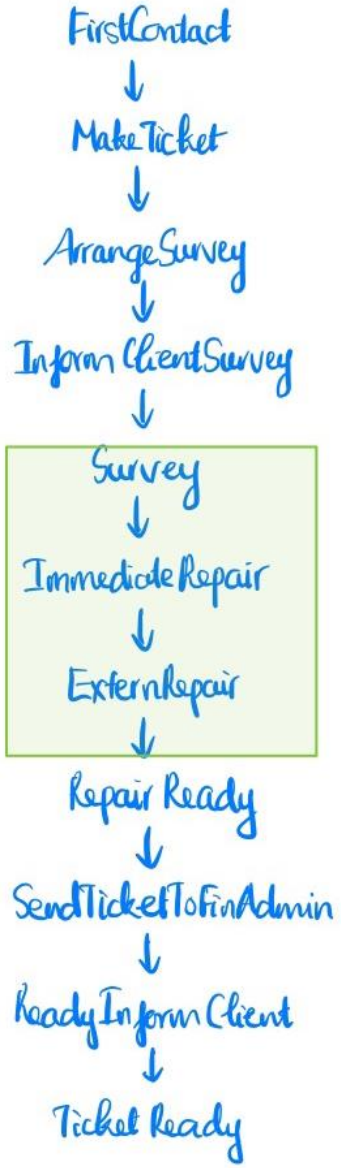
caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
809	FirstContact	Dian	complete	Letter	NaN	NaN	NaN	NaN	NaN	NaN	1/4/1970	7:43	1970-01-04 07:43:00
809	InformClientWrongPlace	Dian	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2/3/2020	7:43	2020-02-03 07:43:00

diemtit96@gmail.com

II.2.ii. Analyze the problems of the current process (bottlenecks, repetitive tasks, etc.)

- This is an example of the whole step of the current process

	caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
8008	2	MakeTicket	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/8/1970	5:17	1970-01-08 05:17:00
9321	2	FirstContact	Dian	complete	Personal	NaN	NaN	NaN	NaN	NaN	NaN	1/8/1970	5:17	1970-01-08 05:17:00
1073	2	MakeTicket	Dian	complete	NaN	P	8161.0	NaN	NaN	NaN	NaN	1/8/1970	5:21	1970-01-08 05:21:00
2563	2	ArrangeSurvey	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/8/1970	5:21	1970-01-08 05:21:00
2911	2	ArrangeSurvey	Dian	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/8/1970	5:25	1970-01-08 05:25:00
10505	2	InformClientSurvey	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/8/1970	5:25	1970-01-08 05:25:00
11800	2	Survey	Anne	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/12/1970	4:19	1970-01-12 04:19:00
3425	2	Survey	Anne	complete	NaN	NaN	NaN	True	120.0	1.0	NaN	1/12/1970	4:57	1970-01-12 04:57:00
5726	2	ImmediateRepair	Anne	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/12/1970	4:57	1970-01-12 04:57:00
6279	2	ExternRepair	Dolt	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/12/1970	6:57	1970-01-12 06:57:00
4582	2	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/12/1970	11:17	1970-01-12 11:17:00
3574	2	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/12/1970	15:14	1970-01-12 15:14:00
5056	2	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/12/1970	15:14	1970-01-12 15:14:00
80	2	ImmediateRepair	Anne	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3/26/2020	6:57	2020-03-26 06:57:00
495	2	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	5/15/2021	8:57	2021-05-15 08:57:00



❑ **There are some problems, I think**:

1. Check status of repair process (RepairOK): duplicate checking step, I think we only need to insert that information one time
2. Some unnecessary taskID: ArrangeSurvey, TicketReady, …
3. The way to store information is cumbersome, confusing, and takes a lot of memory

diemtit96@gmail.com

II.2.iii. If there is a data element that effects the time required for case, present it and explain the rationale

- To determine a data element that effects the time required or not, first of all

caseID	taskID	originator	eventtype	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode	RepairOK	date	time	datetime
1	FirstContact	Dian	complete	Phone	NaN	NaN	NaN	NaN	NaN	NaN	5/23/1905	0:00	1905-05-23 00:00:00
1	MakeTicket	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:08	1970-01-02 08:08:00
1	MakeTicket	Dian	complete	NaN	E	1340.0	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:11	1970-01-02 08:11:00
1	ArrangeSurvey	Dian	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	InformClientSurvey	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/2/1970	8:16	1970-01-02 08:16:00
1	Survey	Cindy	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/11/1970	21:33	1970-01-11 21:33:00
1	Survey	Cindy	complete	NaN	NaN	NaN	True	240.0	1.0	NaN	1/11/1970	21:56	1970-01-11 21:56:00
1	InternRepair	John	start	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	4:36	1970-01-17 04:36:00
1	InternRepair	John	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	8:12	1970-01-17 08:12:00
1	RepairReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	8:12	1970-01-17 08:12:00
1	SendTicketToFinAdmin	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	14:03	1970-01-17 14:03:00
1	ReadyInformClient	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	True	1/17/1970	15:44	1970-01-17 15:44:00
1	TicketReady	System	complete	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1/17/1970	15:44	1970-01-17 15:44:00

	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode
419	Phone	P	4689.0	True	120.0	1.0
561	Phone	P	3095.0	True	120.0	1.0
902	Web	P	6093.0	True	120.0	1.0
18	Phone	E	9679.0	True	240.0	1.0
852	Phone	E	4508.0	True	240.0	1.0
...	...	...	...	...	...	...
703	Web	E	2591.0	True	240.0	1.0
474	Personal	E	5563.0	True	240.0	1.0
833	Phone	E	8638.0	True	240.0	1.0
301	Letter	E	9121.0	True	220.0	2.0
794	Phone	E	8152.0	True	220.0	2.0

II.2.iii. If there is a data element that effects the time required for case, present it and explain the rationale

- To determine a data element that effects the time required or not, I used the hypothesis test – Chi-square goodness of fit test:
  - This test allows us to test whether the observed proportions for a categorical variable differ from hypothesized proportions. The chi-square statistical test is used to determine whether there's a significant difference between an expected distribution and an actual distribution.
  - This test will give a hypothesis, for example:
    - H0(Null Hypothesis): There is no relationship between EstimatedRepairTime and RepairType
    - H1(Alternative Hypothesis): There is a relationship between EstimatedRepairTime and RepairType

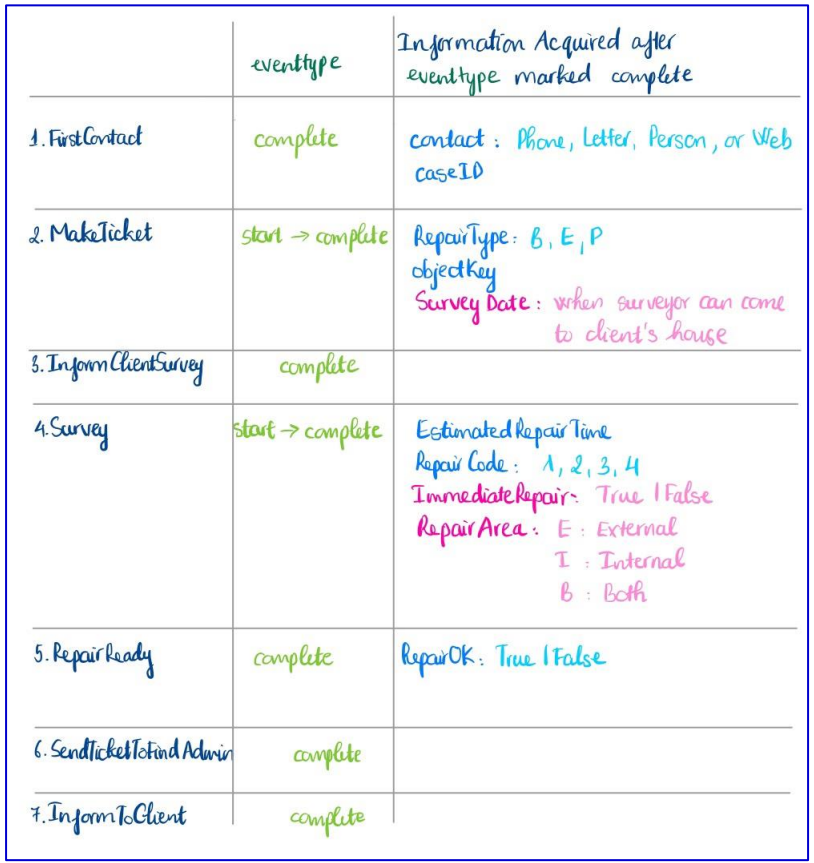
```
If Statistic >= Critical Value: significant result, reject null hypothesis (H0), dependent.
If Statistic < Critical Value: not significant result, fail to reject null hypothesis (H0), independent.
```

  - With the help of Chi-Squared test, I can conclude that is some dependency of 'RepairType, RepairInternally, RepairCode' attribute on the target variable 'EstimatedRepairTime'

	contact	RepairType	objectKey	RepairInternally	EstimatedRepairTime	RepairCode
419	Phone	P	4689.0	True	120.0	1.0
561	Phone	P	3095.0	True	120.0	1.0
902	Web	P	6093.0	True	120.0	1.0
18	Phone	E	9679.0	True	240.0	1.0
852	Phone	E	4508.0	True	240.0	1.0
...	...	...	...	...	...	...
703	Web	E	2591.0	True	240.0	1.0
474	Personal	E	5563.0	True	240.0	1.0
833	Phone	E	8638.0	True	240.0	1.0
301	Letter	E	9121.0	True	220.0	2.0
794	Phone	E	8152.0	True	220.0	2.0

II.3. Describe the ways to improve the current process

- The current process has been conducting based on the order of about 10 **taskID** for each **caseID**
- After each **taskID**, some information such as: **contact, RepairType,** …will be collected. The purpose is to get information for the next **taskID**.
- However, I found out that the current process can be improved to save data storage space, as well as save more time for the whole process, therefore I propose a modification for the current process as following:



	eventtype	Information Acquired after eventtype marked complete
1. FirstContact	complete	contact : Phone, Letter, Person, or Web caseID
2. MakeTicket	start → complete	RepairType: B, E, P objectKey Survey Date : when surveyor can come to client's house
3. InformClientSurvey	complete	
4. Survey	start → complete	Estimated Repair Time Repair Code : 1, 2, 3, 4 ImmediateRepair: True / False RepairArea : E : External I : Internal B : Both
5. RepairReady	complete	RepairOK : True / False
6. SendTicketToFindAdmin	complete	
7. InformToClient	complete	

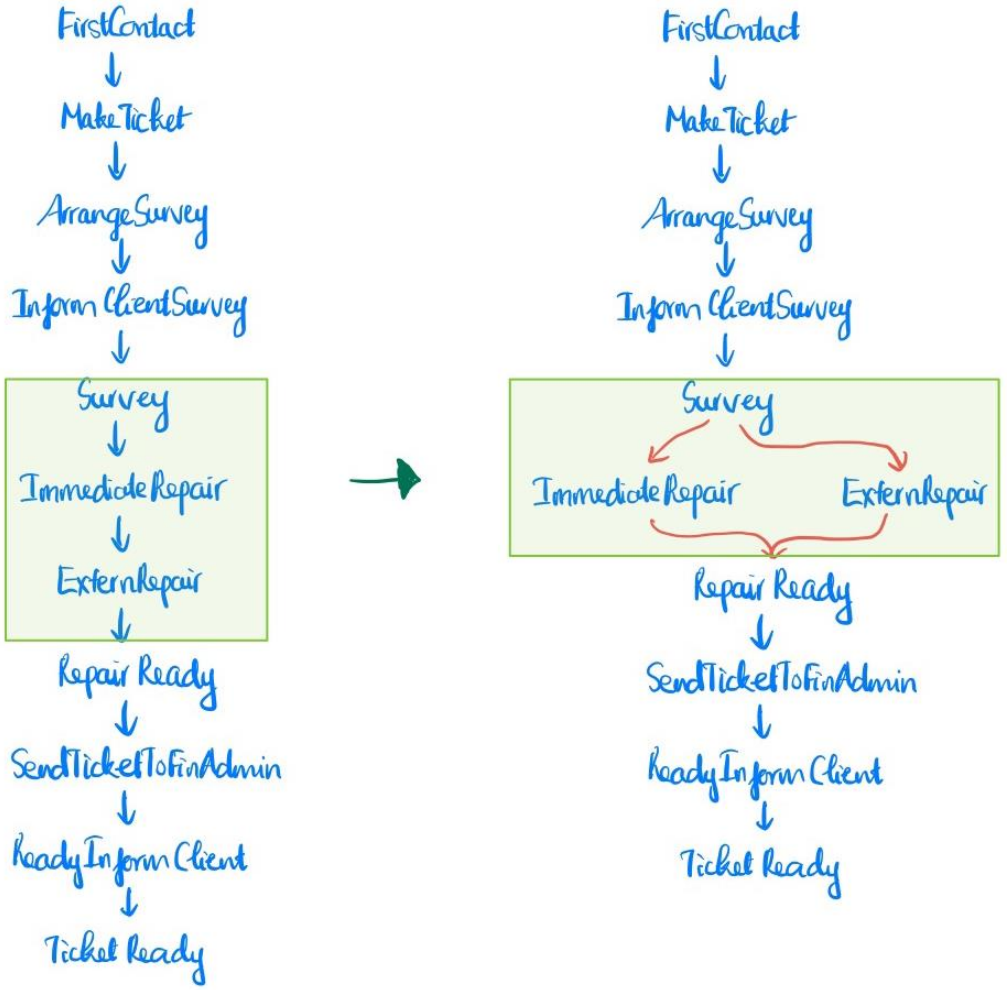
Explanation

- **First, in step 2 – *MakeTicket***: After finishing this step, I suggest adding the information **SurveyDate** – when the surveyor can come to the client's house. Then we inform all information to the client once again before going to the next step.
- **Second, in step 5 – *Survey***: I think we should collect **ImmediateRepair** (receive 2 values: ***True*** or ***False***) and **RepairArea** (receive 3 values: ***E*** - external repair, ***I*** - internal repair, ***B*** - external and internal repair) instead of **RepairInternally**. This will help staff receive information clearly and conduct quickly in case of urgent repair and will be more convenient in the future if need to check again.
- **Third, eliminate some taskID:** check **RepairReady** step is complete then send ticket to Admin to check again, and then send conformation to the client and finish. Do not need to input the information for **RepairOk** again (It's checked in step 5).

diemtit96@gmail.com

II.3. Describe the ways to improve the current process

- Change the way conducting the taskID:



diemtit96@gmail.com

II.3. Describe the ways to improve the current process

- Change the way information is stored ➔ Input all the information of a request (caseID) in one row, for example:

CaseID : number

DateReceive : datetime

ContactMethod : Phone, Mail, Web, Person, ...

RepairType : B, E, P

ObjectKey : number

SurveyDate : datetime

ImmediateRepair : True or False

RepairArea : E, I, B

RepairCode : 1, 2, 3, 4

RepairStatus : Todo, InProcess, Finish

AdminCheck : Yes or No

InformClient : Yes or No

DateFinish : datetime

etc.....

diemtit96@gmail.com

Thank you for your reading!