

Connecticut Claim Handling Complaints Data Analysis

Prepared by: Chris Smith

csmith3723@gmail.com

Table of Contents

Introduction.....	3
Methodology.....	3
Data Source.....	3
Data Analysis	4
Conclusion	8
References	9

Introduction

This data analysis report was developed as a portfolio project to demonstrate data analysis skills. The data pertains to Connecticut's insurance industry with a particular focus on complaints labeled as "Claim Handling" complaints. Among all insurance complaints filed in Connecticut from 2015 to 2025, more than 64% were related to Claim Handling, resulting in a recovery total of more than \$31million. The goal of this data analysis was to examine the Claim Handling complaints to identify any relationships between the specific type of Claim Handling complaint, claims processing times, and recovery amount.

This study was driven by the following questions:

1. What are the most common sub-reasons for "Claim Handling" complaints?
2. Do these sub-reasons result in the highest recovery, and if not, which ones do?
3. Is there any correlation between process time and recovery?
4. Is there any correlation between process time and sub-reason?

Methodology

Data Source

This data set, titled "Insurance Company Complaints, Resolutions, Status, and Recoveries," is taken from Connecticut's Data Catalog, available on data.ct.gov and updated as of March 25, 2025. It lists the complaints filed by 849 insurance companies and categorizes the complaints into four distinct categories: Underwriting, PolicyHolder Service, Marketing & Sales, and Claim Handling, with the latter making up most of the complaints. The data was loaded and cleaned with Excel and examined with Power BI.

Data Cleaning

The focus of this study is Claim Handling complaints. As such, I filtered out all complaints where the Reason category was not "Claim Handling," including uncategorized complaints. The study's focus was determining if there were any obvious relationships between complaint Sub-reason, Recovery, and Process Time; the data was cleaned following this framework. Furthermore, multiple sub-reasons are associated with a single

file number, which caused duplication of recovery amounts and caused erroneous calculations. To avoid these issues, the data was split into three separate tables: Processing Time & Recovery, Sub-reason & Recovery, and Company & Recovery. Without this split, data was inconsistent and inaccurate.

Process Time is not listed as part of the original data, but it was calculated using the Opened and Closed date of the complaint. To properly calculate the Process Time in days for all valid entries, complaints that had no Closed date were excluded, as well as a claim where a data entry error resulted in a Closed date earlier than the Opened date. A total of eight complaints out of over 19000 had a Process Time of over 1000 days, so they were excluded in any calculations where Process Time was a factor to avoid skewed outcomes. All columns related to coverage, disposition, and current claim status were also filtered out. Any complaints that had not been finalized and closed were excluded to avoid presenting data that was actively changing.

Data Analysis

I started by mapping each sub-reason to display how their recovery and complaint counts relate to other sub-reasons and the totals across all “Claim Handling” complaints. Claim Delay, Claim Denial, and Unsatisfactory Settlement/Offer combined accounted for over 50% of the total of both recovery and complaints. With this data mapped, I knew where to focus my attention to answer my remaining questions.

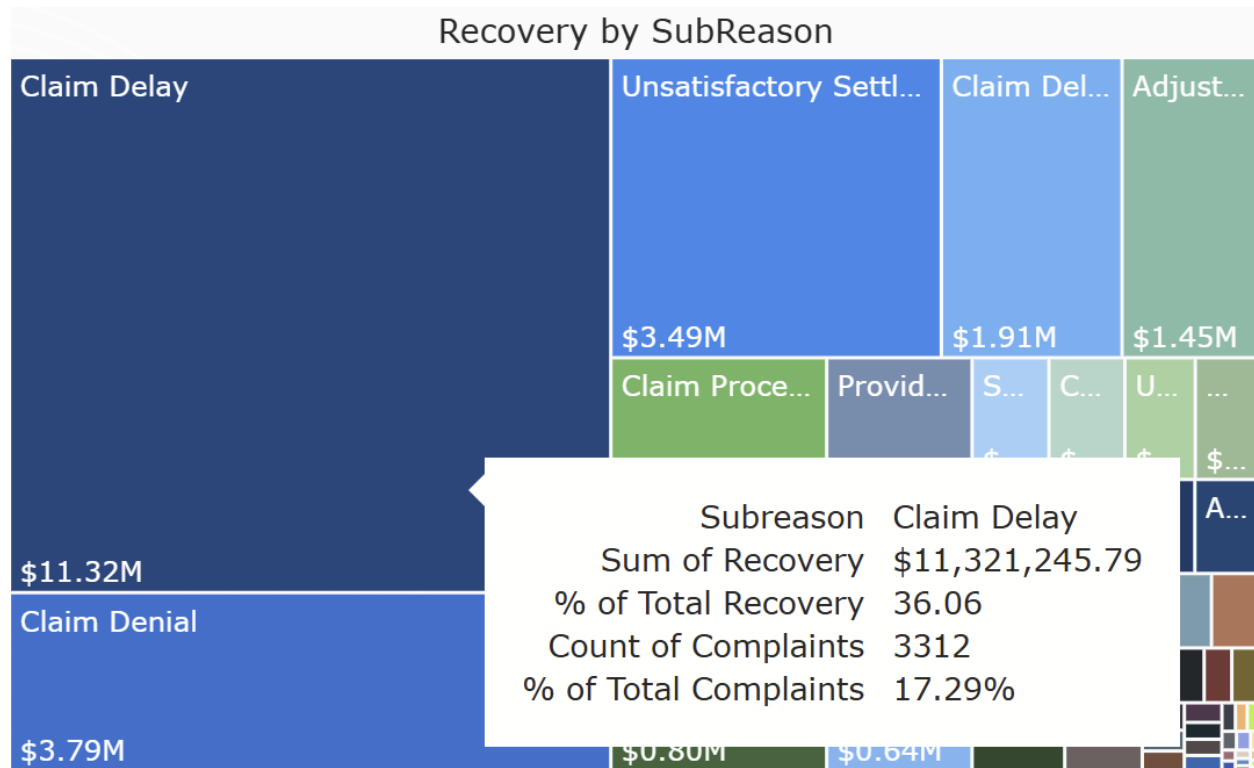


Figure 1: Tree map showing proportional breakdown of complaint sub-reasons, with count and percentage of total displayed.

Next, I compared the total number of complaints to both the total Recovery amount and the average Recovery amount for each sub-reason. Figure 2 shows that while ‘Claim Delay,’ the sub-reason with the highest Recovery, had an above average Recovery per Complaint, the most common sub-reason of ‘Claim Denial’ had a relatively low Recovery per Complaint. In fact, Figure 3 shows that certain sub-reasons with only a few complaints had far above average Recovery per Complaint amounts. This suggests that the high total Recovery for certain sub-reasons is a result of the volume of complaints, rather than other factors.

Subreason	Count of Subreason	Recovery per Complaint	Total Recovery
Claim Delay	3312	\$3,407.38	\$11,285,228
Claim Denial	4358	\$872.06	\$3,800,426
Unsatisfactory Settlement/Offer	3140	\$1,105.77	\$3,472,133
Claim Delays	435	\$4,384.46	\$1,907,242
Adjuster Handling	730	\$1,981.43	\$1,446,445
Claim Procedure	1305	\$1,045.04	\$1,363,772
Claim Handling Other	165	\$5,820.60	\$960,400
Coordination of Benefits	107	\$7,503.61	\$802,887
Total	19155	\$1,635.55	\$31,328,932

Figure 2: Table comparing number of complaints, recovery per complaint, and total recovery for each sub-reason.

Subreason	Count of Subreason	Recovery per Complaint	Sum of Recovery
Retro Denial	1	\$124,921.31	\$124,921
UR Case Management	10	\$30,179.30	\$301,793
Surprise Billing	43	\$7,654.99	\$329,165
Coordination of Benefits	107	\$7,503.61	\$802,887
Participating Provider Availability/Timely Access to Care	40	\$6,266.79	\$250,671

Figure 3: The table from Figure 2 shows the highest recovery per complaint occurs among subreasons with few complaints.

I also wanted to establish if there was any relationship between the most common complaint sub-reasons and how long it took for those complaints to be processed. Below, we can see that in terms of both median processing time and average processing time, the most common sub-reasons are either below or only slightly above the values for all sub-reasons. This data, used in conjunction with the table in Figure 2, demonstrates that the volume of complaints, processing time, and recovery amounts for each sub-reason do not appear to have any direct influence on each other.

Subreason	# of Complaints ▼	Median Processing Time	Avg Processing Time
Claim Denial	4358	26	36.84
Claim Delay	3312	26	39.07
Unsatisfactory Settlement/Offer	3140	26	37.81
Claim Procedure	1305	22	37.98
Adjuster Handling	730	27	36.99
UR MEDICALLY NECESSARY DENIAL	578	20	32.56
Actual Cash Value Dispute	531	23	36.00
Claim Delays	435	32	57.32
Medical Necessity Denial	421	20	27.98
Denial	361	30	54.75
Contributory Comparative Negligence	347	26	31.78
Eligibility of Provider	335	20	50.52
Total	19155	26	38.85

Figure 4: Table showing the number of complaints, median processing time, and average processing time for all sub-reasons.

Finally, I wanted to more directly determine if there was a relationship between Recovery and Process Time without consideration for sub-reason – particularly, did complaints that took longer to process result in high Recovery? To do this, I measured both the median Process Time (days) and the average Process Time for each sub-reason. As seen in Figure 5, there is no evidence that higher Recovery is directly related to how long it took to process the complaint. Additionally, there is a very low correlation between the data according to Power BI's quick measure feature.

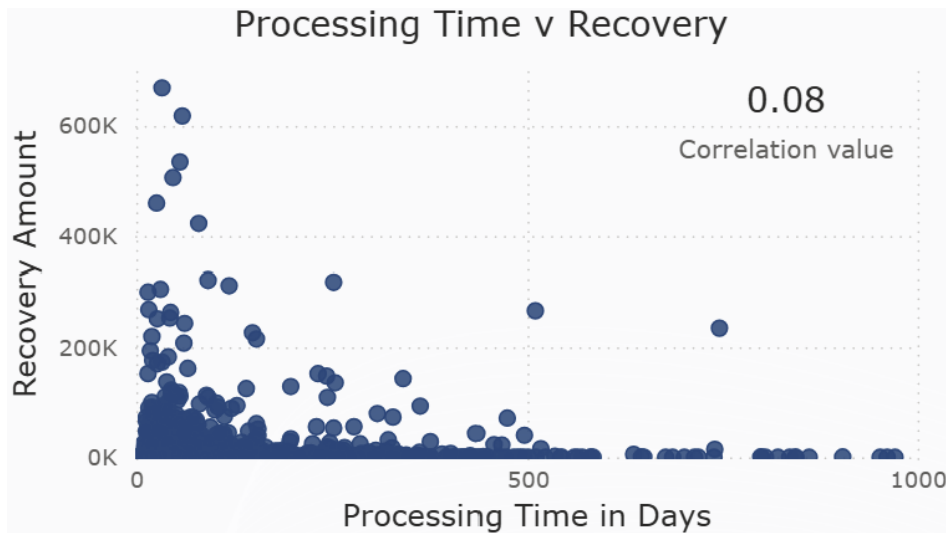


Figure 5: Scatter chart plotting the complaints by Process Time and Recovery with the correlation value.

Conclusion

After examining the data, there does not appear to be a clear or specific cause for high total Recovery on Claim Handling complaints other than the high volume. As noted in Figure 3, the average Recovery is highest on a sub-reason with very few complaints.

Claim Delay is the second most common sub-reason among complaints, but due to its higher-than-average recovery per complaint, it ends up with a total recovery that is over three times higher than the most common complaint sub-reason, Claim Denial.

Additionally, figure 4 shows average and median Process Time across all sub-reasons, and fewer than 10 sub-reasons have either an average or median Process time of more than 100 days. From this we can conclude that the specific complaint sub-reason is not an indicator of how long Process Time is likely to be. Furthermore, the scatter plot in Figure 5 displays Process Time in relation to Recovery. Using Power BI's calculation, the correlation factor between the two was only .08, a very weak positive correlation. In fact, most of the large sum Recovery points occur with less than 200 days of Process Time.

Overall, there appears to be no specific factor in particular, aside from the volume of complaints, that contributes to the total Recovery from all Insurance Claim Handling complaints in the state of Connecticut.

References

State of Connecticut. (2025, March 25). *Insurance Company Complaints, Resolutions, Status, and Recoveries*. Data.gov. <https://catalog.data.gov/dataset/insurance-company-complaints-resolutions-status-and-recoveries>