WORKERS' COMPENSATION MANAGEMENT

Big Data Analytics | Fall 2017 | Group 8

Adhish Thite | Bhanu Sriram | Pranav Tambvekar | Ved Paranjape
UNC Charlotte | Big Data Analytics for Competitive Advantage – Fall 2017

Executive Summary

In this exercise, we studied working of worker's compensation claim management system, issues faced by the industry and how big data can be employed for solving problems and improving the system.

The Workers compensation is a program that provides benefits to certain workers or their dependents who suffer a work-related injury or disease. Claims can be of three types, report only, medical, or indemnity. Concepts like subrogation and litigation were new to us. In case a claim is disputed, it can go in litigation state where the law system- lawyers and judges are involved. If an employee gets injured due to some other party, then the employee's claims are approved, but at the same time, funds from the other party are collected to compensate.

The claims companies face few issues like fraudulent claims, cost increasing events like litigation, missed opportunities for subrogation, difficulty in deciding how much reserves to assign for each claim etc. These issues can be tackled with applying big data technologies and analytics.

A sample dataset consisting of claims data and transactions data was used for this activity. The given dataset had few issues. Many fields like claimant's age on incident date, claimant closed date etc. have missing values. Columns like indemnity paid, claimant age, total reserves etc. have values in a wrong range. The data needs to be studied well to decide a plan of action for these issues. For example, we can't simply remove the records that have a missing claimant Closed Date. The missing value for claimant closed date can mean that the claim is yet to be closed. A strategy like replacing the missing values with mean of remaining values of the sane column can be used to address this issue.

After wrangling the data, we used Tableau to visualize the data and find out some interesting information from it. It can be seen from the visualizations that the total costs for the company is much higher than the costs recovered by the company. The costs are seen to be higher in case of litigations. Persons of age between 40 to 55 are seen to have more claims than other ages. This fact can be used to take decisions like deciding premium amounts dependent on the age of the employees. The company can also sponsor health programs or other activities to ensure a safe work-environment. The best claim is the one which never happens. Hence avoid claims through safety and health programs.[1]

Apart from the data given, the data about premiums paid for the employers can provide more insights. The external data like police records, medical reports, can be used by the company to predict the fraudulent claims and to find the subrogation opportunities. The social media data and data from customer support centers can be useful to gain insights on the customer satisfaction as well as the time required for the claims.

1. http://www.theabdteam.com/blog/claims-management-101-primer-sound-decisions-and-cost-management/

Appendix A- Workers' Claim Management

Claim Management Process[2]

Workers' Claim Management includes the following processes –

- Getting Immediate Medical Action

  It is necessary to get primary medical attention before filing a workers' compensation claim. Some insurances require seeing a specific doctor. The doctor makes a medical report for the injury/sickness. This medical report is necessary in the further process. The medical report serves as an official record for the injuries and the basis for worker's compensation reimbursement.

- Notifying the Employer

  Although the workers' compensation laws differ by state, the injured employee must notify the employer about the injury or sickness due to work-related causes. This notification must be done within 30 to 45 days of an accident or injury becoming known. Any workplace accidents must be reported within the stipulated deadline. Just in case some injury surfaces not immediately after the accident but after the deadline expires. An official written form needs to be filled out after the injury which contains the information such as the type of injury, the parties involved in the accident, date, time and place, medical treatment received.

- Filing the Claim

  The employer informs the insurance company and the state's worker compensation board. After the claim is evaluated, the employees are notified whether their claim was accepted and the maximum amount of medical benefits they are entitled to.

- Post-claim Filing

  If the insurer approves the claim, it may cover the costs including medical bills, medicines, disability payments, and a portion of their lost wages. Behind the scenes, the insurer informs the board that the reimbursement payments have begun or the reasons if the payments are not being made.

- Return to Work

[2] http://injury.findlaw.com/workers-compensation/the-workers-compensation-claim-process.html

When the employee recovers and wants to return to work, they must alert the employer and the insurance company via a written notice. The insurance company may continue paying them the disability benefits depending on the severity of the injury.

Challenges Faced by Workers' Compensation Management

- Fraud

  The most important aspect in Workers' Compensation Management is fraudulent claims. It is extremely tough to distinguish between fraudulent and genuine claims.

- Low Premium Growth to Cover the Costs[3]

  Low to moderate pay wages hurt the worker's compensation market. If wages continue to stay low, soon there will not be enough premium growth to cover the costs associated with the medical component of the claim dollar.

- The Affordable Healthcare Act (AHCA)[4]

  Many people feel that the implantation of this act will lead to increase the leakage from group health to workers compensation, others feel it will have the modest effect. Though the initial research shows the modest changes, it's too early to predict the AHCA's effect on worker's compensation.

- Managing the Massive Amounts of Information

  Managing and making sense of massive growing amounts of information is the biggest challenge faced by worker's compensation claims management. This challenge increases daily with the introduction of new publicly available data sources, cheaper ways to store the information and new tools that enable new different ways to manipulate information.

- Technology and Innovation

  The worker's industry is way behind in terms of use of the technology compared to other industries. Since the data is being produced in huge volumes, you need technology and tools to store and analyze the data in an efficient manner.

- Workplace Safety

[3] http://www.insurancejournal.com/magazines/features/2014/05/05/327694.htm

[4] https://www.freewayinsurance.com/blog/auto-insurance/the-top-10-issues-facing-the-workers-compensation-industry/

The Occupational Safety and Health administration (OSHA) is focusing on working with the Judicial department and people are going to jail for doing the things that are willfully unsafe in the workplace. It is not something that happens overnight. It takes a unified effort change a culture of unsafe workplace.

- Problems in Claim Reporting

  These are the major challenges faced by the worker's compensation management. These include invalid claims, improper claim reporting, late reporting, failure to monitor claims, failure to have claims reviewed.

Role of Big Data Analytics[5]

Big Data offers a modern method of fulfilling the more than 100-year-old promise of the workers' compensation system: to make individuals whole again after a workplace illness or injury.

There are 3 major metrics which are important to track in workers' compensation –

- Employer's Injury Rate
- Claims Frequency & Claim Severity
- Average Claims Duration

Big Data Analytics can help the WCM (Workers' Compensation Management) to solve a couple of the major issues predominantly.

1. Fraud Detection

   Predictive Analysis combined with external data can help detect frauds quickly, and can also decipher a pattern in the fraudulent claims. Any abnormal pattern in the process can be picked out by analytics beforehand and the concerned teams can be alerted so as to take corrective measures.

2. Managing Massive Amounts of Information

   Along with managing the information, making use of the huge data is a humungous task. However, with the infrastructure that Big Data provides, managing and understanding the data is extremely affordable and easy. Additionally, Big Data Analytics will help the business get a better insight in the data.

3. Predicting Costs based on Heuristic Data

[5] https://www.sas.com/en_us/insights/articles/risk-fraud/big-data-analytics-improves-claims-processing.html

The expected total costs of the claims can be predicted using the existing data, and a new quote can be prescribed.

External Data Sources

To perform Big Data Analytics, the most necessary component is ‘data’. Data can be internal data from within the organization and subsidiaries, but it can also be aggregated from external sources.

The external data sources which can be useful for Big Data Analytics are –

- Police Records
- Medical Records
- Credit History
- Customer Service Center Data
- Social Media Analytics

Appendix B – Data Wrangling Analysis

Data Wrangling

While importing the claims data, there are fields with incorrect datatypes. We have changes the fields with incorrect datatypes to the datatypes according to claims data variables.

(AverageWeeklyAge from String data type to Number data type,
ClaimantAgeDOI from string to Number datatype,
ClaimantClosedDate from String to Number,
EmployerNotificationDate from String to Number,
ReturnToWorkDate from String to Number)

Apart from the above-mentioned datatype issues, there are some data issues as listed below:

Missing Values

1. Average Weekly Wage

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
AverageWeeklyWage		49080	84924	28826777.18	0.01	587.34	491.97	2024000.00	50.34

There are missing values in AverageWeeklyAge field. The frequency of the missing values is very high. Hence, we can't simply discard these values. We can replace these missing values by the mean of AverageWeeklyAge field. This can be done by using computational methods (recoded column) in the query builder. After applying these computational methods, the new column AverageWeeklyWage_Recoded is formed as shown below:

Summary of Numeric Variables for WORK.QUERY_FOR_CLAIMSDATAPALL_0002

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
AverageWeeklyWage		41932	83742	24255928.67	0.01	578.46	477.23	2024000.00	58.92

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
AverageWeeklyWage_Recoded		125674	0	72697214.05	0.01	578.46	578.46	2024000.00	19.66

2. Claimant Age DOI

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
ClaimantAge_at_DOI		88926	45078	3543915.00	-8000.00	39.85	42.00	94.00	0.40

There are missing values in ClaimantAge_at_DOI field. The frequency of the missing values is very high (around 60%). Hence, we can't simply discard these values. We can replace these missing values by the

mean of ClaimantAge_at_DOI field. This can be done by using computational methods (recoded column) in the query builder. After applying these computational methods, the new column ClaimantAge_at_DOI _Recoded is formed as shown below:

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
ClaimantAge_at_DOI		82340	43334	3242320.00	-8000.00	39.38	42.00	94.00	0.43

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
ClaimantAgeatDOI_Recoded		125674	0	4932346.00	-8000.00	39.25	39.00	94.00	0.28

3. Claimant_Close_Date

Variable	Label	N	NMiss	Min	Mean	Median	Max
ClaimantClosedDate		129326	4678	06JAN99	24MAY07	05APR06	06DEC14

There are few missing values in ClaimantClosedDate. If you observe the data set given, you can observe most of these records have missing values because the claim status is still in open and reopen state. Hence, they don't have any dates for claims which are still not closed.

We have replaced the missing values with 0. Below is the dataset after cleaning:

Variable	Label	N	NMiss	Min	Mean	Median	Max
ClaimantClosedDate_Recoded	ClaimantClosedDate_Recoded with 0	82134	0	01JAN60	28MAY07	12OCT08	30JUN14

4. Employer_Notification_Date

Variable	Label	N	NMiss	Min	Mean	Median	Max
EmployerNotificationDate		111716	22288	09OCT72	30AUG05	28DEC04	21JUL99

There are 22288 missing values in Employer_Notification_Date field. Most of these values are due to fatality. Since they expired, they can't notify the employer.
We have replaced the missing values with 0. Below is the dataset after cleaning:

Variable	Label	N	NMiss	Min	Mean	Median	Max
EmployerNotificationDate_Recoded	EmployerNotificationDate_Recoded	82134	0	01JAN60	08JUL01	30JUL06	21FEB88

5. ReturnToWorkDate

Variable	Label	N	NMiss	Min	Mean	Median	Max
ReturnToWorkDate		75367	58637	29OCT76	01JUN06	05JUL07	05JUL15

There are around 58000 missing values. Among these there are many people who are dead (i.e. isFatal is 1). Hence, they cannot return to work.
We've replaced these values with 0. Below is the dataset after cleaning.

Variable	Label	N	NMiss	Min	Mean	Median	Max
ReturntoWorkDate_Recoded		82134	0	01JAN60	15NOV91	09OCT05	07MAY15

Unacceptable values

1. Indemnity Paid

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
IndemnityPaid		134004	0	410127284.24	-475.00	3060.56	0.00	640732.32	51.83

The minimum value is negative. But the amount paid can never be negative value. Also, if you observe the maximum value of Indemnity Paid, it is also out of range. Hence, we need to clean this inconsistent data. To avoid these inconsistencies, advanced filter can be used where can filter records with IndemnityPaid value greater than or equal to 0 and less than 20000.

Indemnity Paid >=0 and Indemnity Paid <20000

After applying the filter, the data is cleaned as shown below.

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
IndemnityPaid		125674	0	31064852.78	0.00	247.19	0.00	19145.76	3.52

2. TotalPaid

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
TotalPaid		134004	0	903995008.11	-270.32	6746.03	235.32	4527290.68	107.96

The minimum value is negative. But the amount paid can never be negative value. Also, if you observe the maximum value of Total Paid, it is also out of range. Hence, we need to clean this inconsistent data. To avoid these inconsistencies, advanced filter can be used where can filter records with TotalPaid value greater than or equal to 0 and less than 20000.

TotalPaid >=0 and TotalPaid <20000
After applying the filter, the data is cleaned as shown below.

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
TotalPaid		125674	0	132235626.66	0.00	1052.21	208.00	19999.14	7.41

3. OtherPaid

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
OtherPaid		134004	0	493867723.87	-7820.46	3685.47	230.44	4129915.41	70.90

The minimum value is negative. But the amount paid can never be negative value. Also, if you observe the maximum value of Other Paid, it is also out of range. Hence, we need to clean this inconsistent data. To avoid these inconsistencies, advanced filter can be used where can filter records with OtherPaid value greater than or equal to 0 and less than 20000.

OtherPaid >=0 and OtherPaid <20000

After applying the filter, the data is cleaned as shown below

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
OtherPaid		125674	0	101170773.88	0.00	805.03	204.09	19906.79	5.17

4. TotalReserves

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
TotalReserves		134004	0	299215140.50	0.00	2232.88	0.00	2069575.10	80.16

If you observe the maximum value of TotalReserves, it is out of range. Hence, we need to clean this inconsistent data. To avoid these inconsistencies, basic filter can be used where can filter records with TotalReserves value less than 20000.

TotalReserves <20000

After applying the filter, the data is cleaned as shown below

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
TotalReserves		125674	0	7740256.09	0.00	61.59	0.00	19965.82	2.00

5. TotalRecovery

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
TotalRecovery		134004	0	9230090.27	0.00	68.88	0.00	130541.03	3.74

If you observe the maximum value of TotalRecovery, it is out of range. Hence, we need to clean this inconsistent data. To avoid these inconsistencies, basic filter can be used where can filter records with TotalReserves value less than 20000.

TotalRecovery <20000

After applying the filter, the data is cleaned as shown below

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
TotalReserves		125674	0	7740256.09	0.00	61.59	0.00	19965.82	2.00

6. ClaimantAge_at_DOI

Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
ClaimantAge_at_DOI		88926	45078	3543915.00	-8000.00	39.85	42.00	94.00	0.40

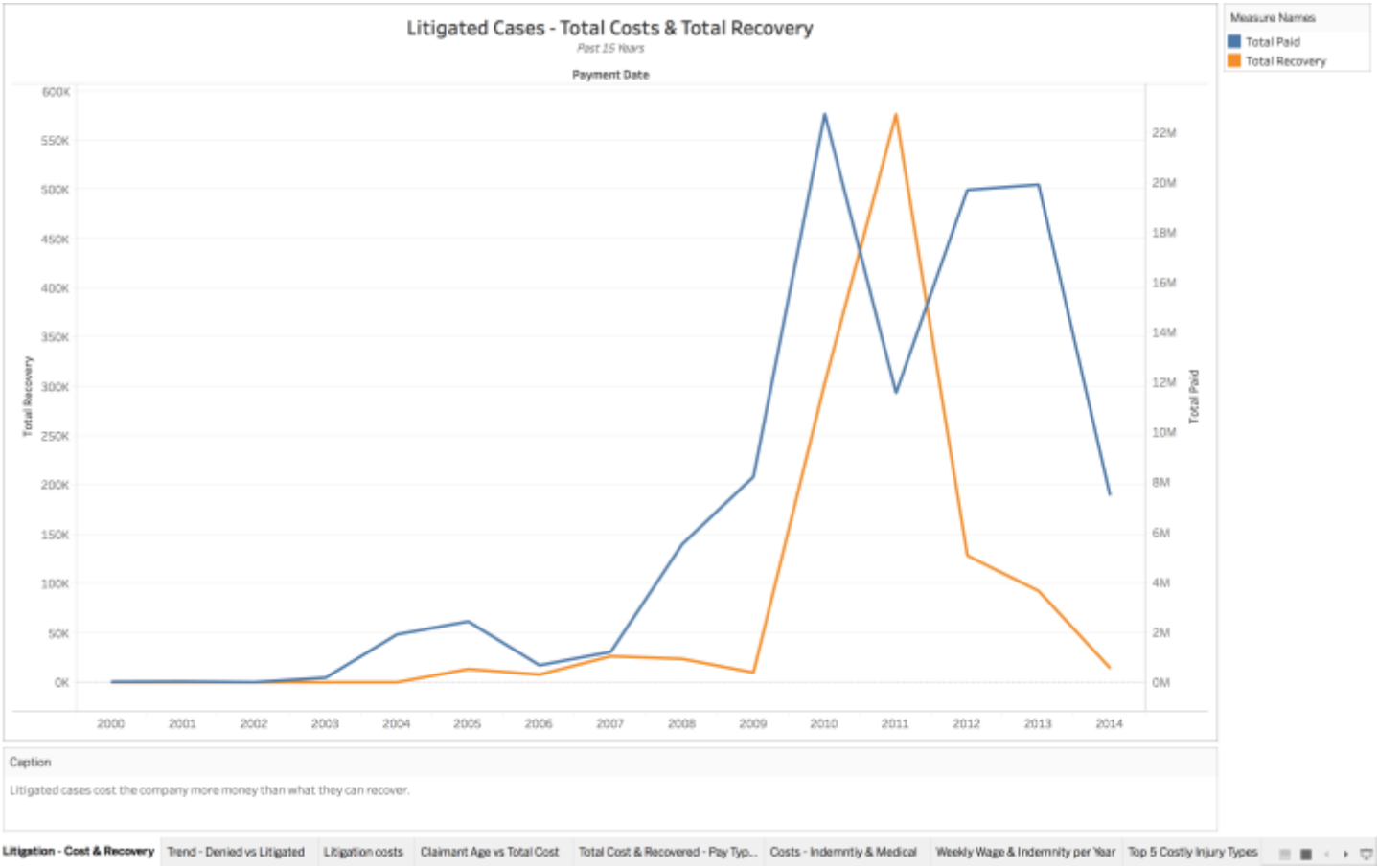
If you observe the minimum value, it is -8000. Age can never be a negative value. To avoid this inconsistent data, basic filter can be used where can filter records with ClaimantAge_at_DOI value greater than 0.

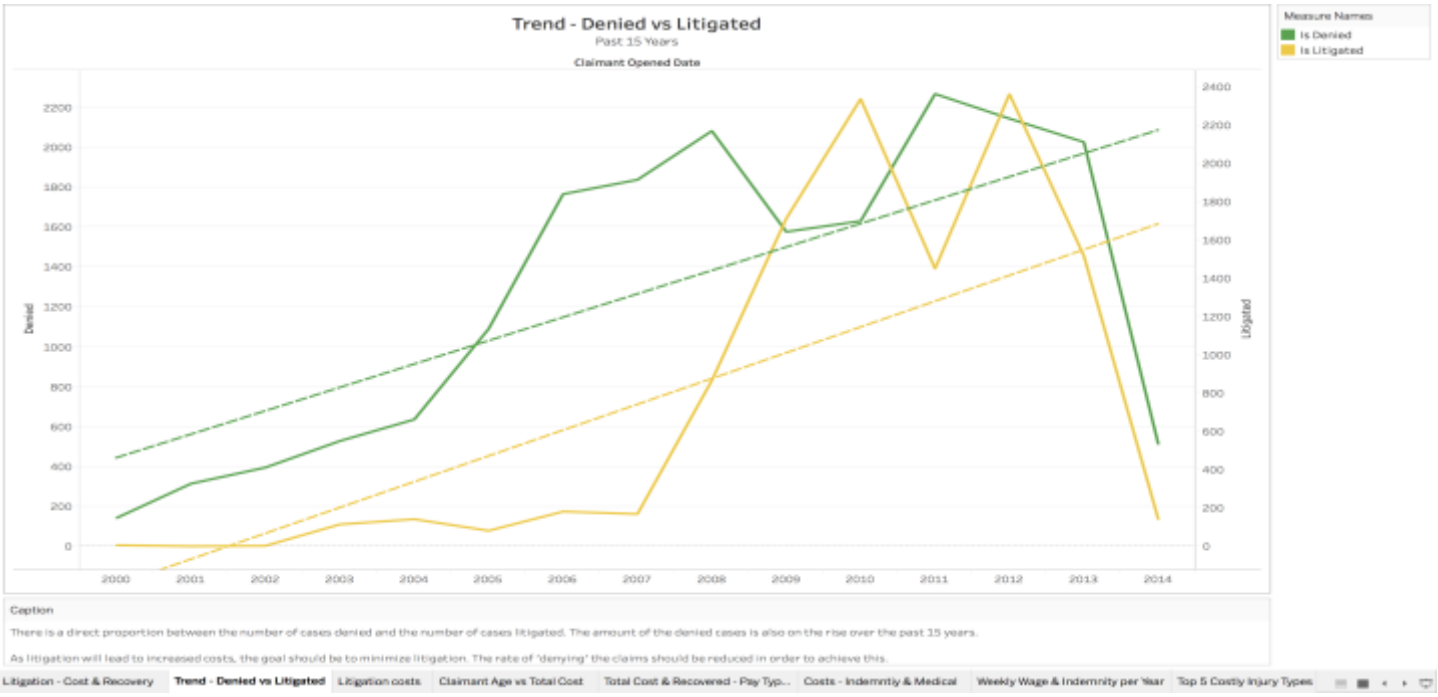
ClaimantAge_at_DOI > 0

After applying the filter, the data is cleaned as shown below

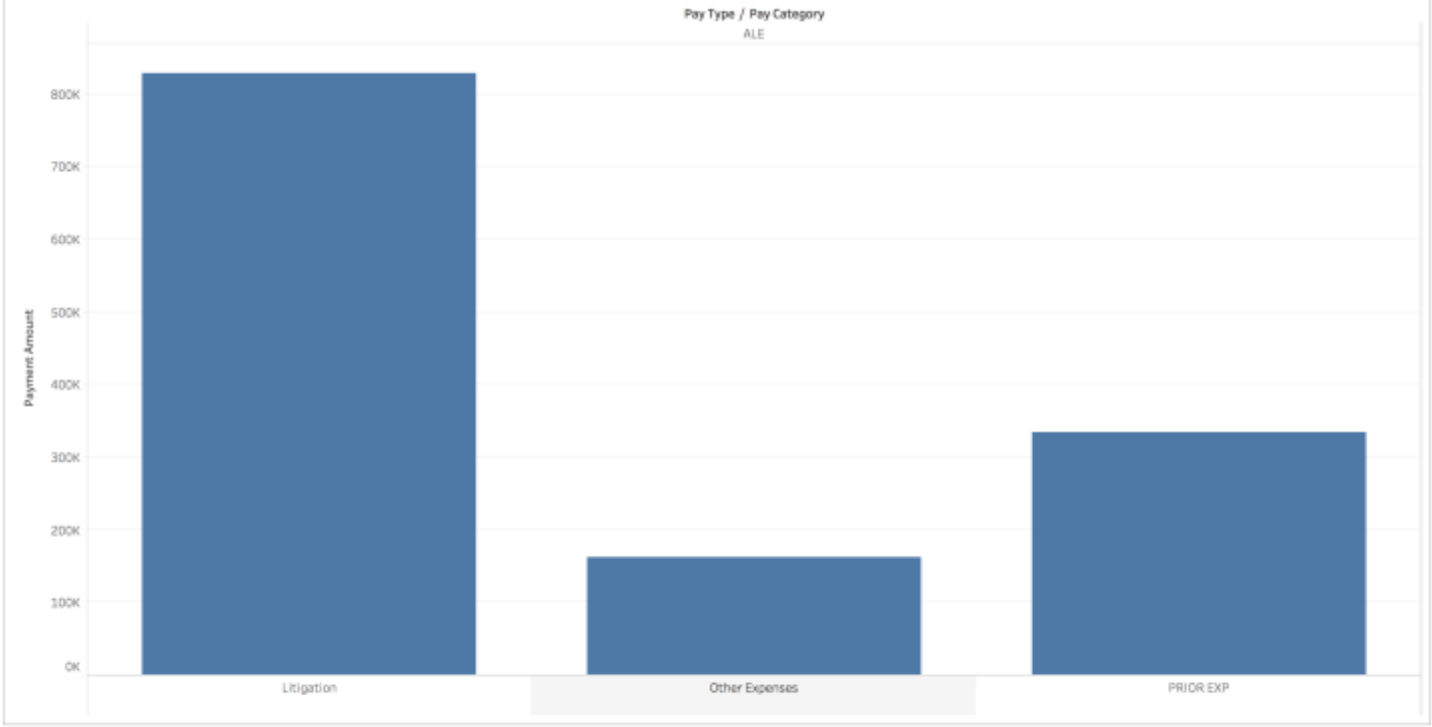
Variable	Label	N	NMiss	Total	Min	Mean	Median	Max	StdMean
ClaimantAge_at_DOI		82134	0	3433387.00	1.00	41.80	42.00	94.00	0.04

Appendix C – Visualizations





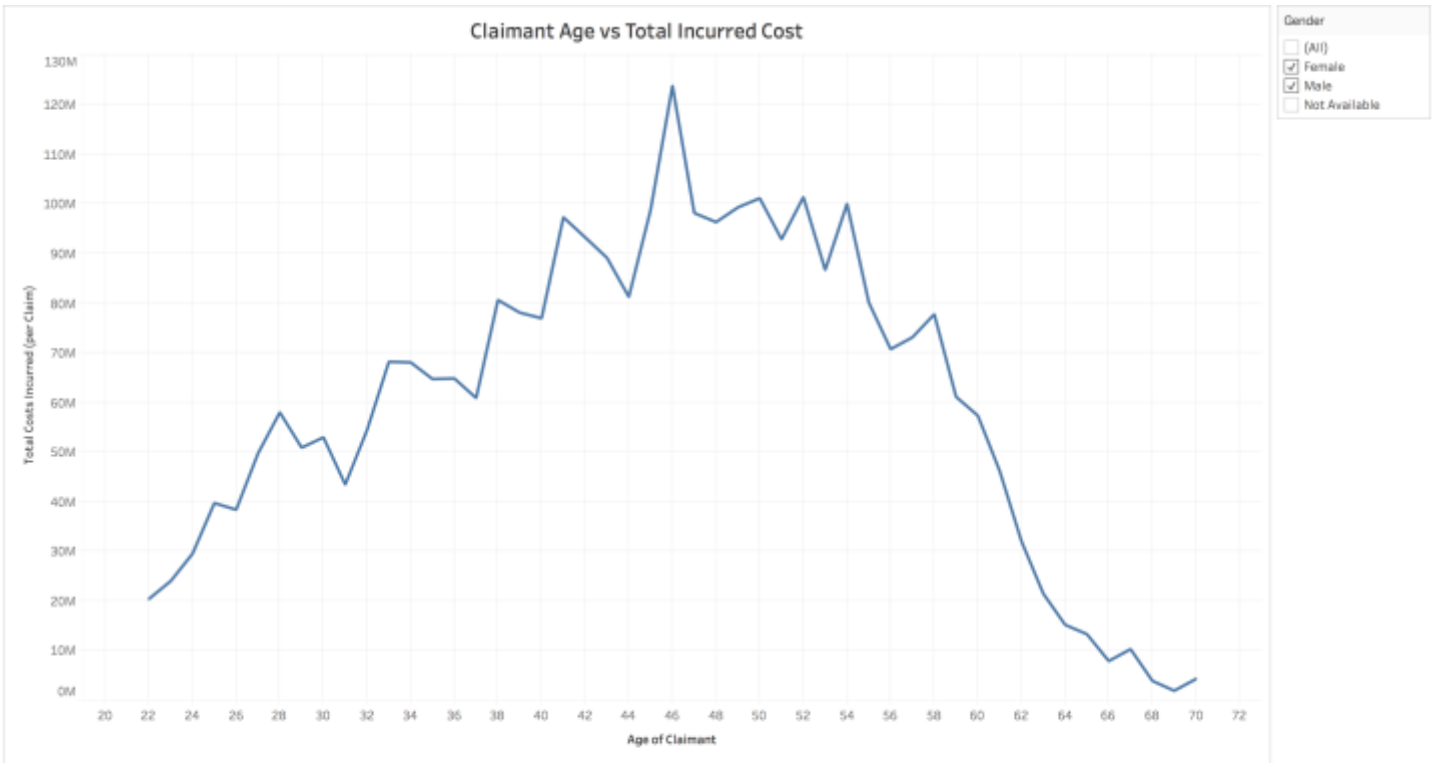
Comparisons of Expense Sums due to Litigation and Attorney



Caption
Insurance company should strike the right balance between finding the right attorneys to represent them in court and should not compromise with that. Else, as observed, the company would end up losing much more in litigation.

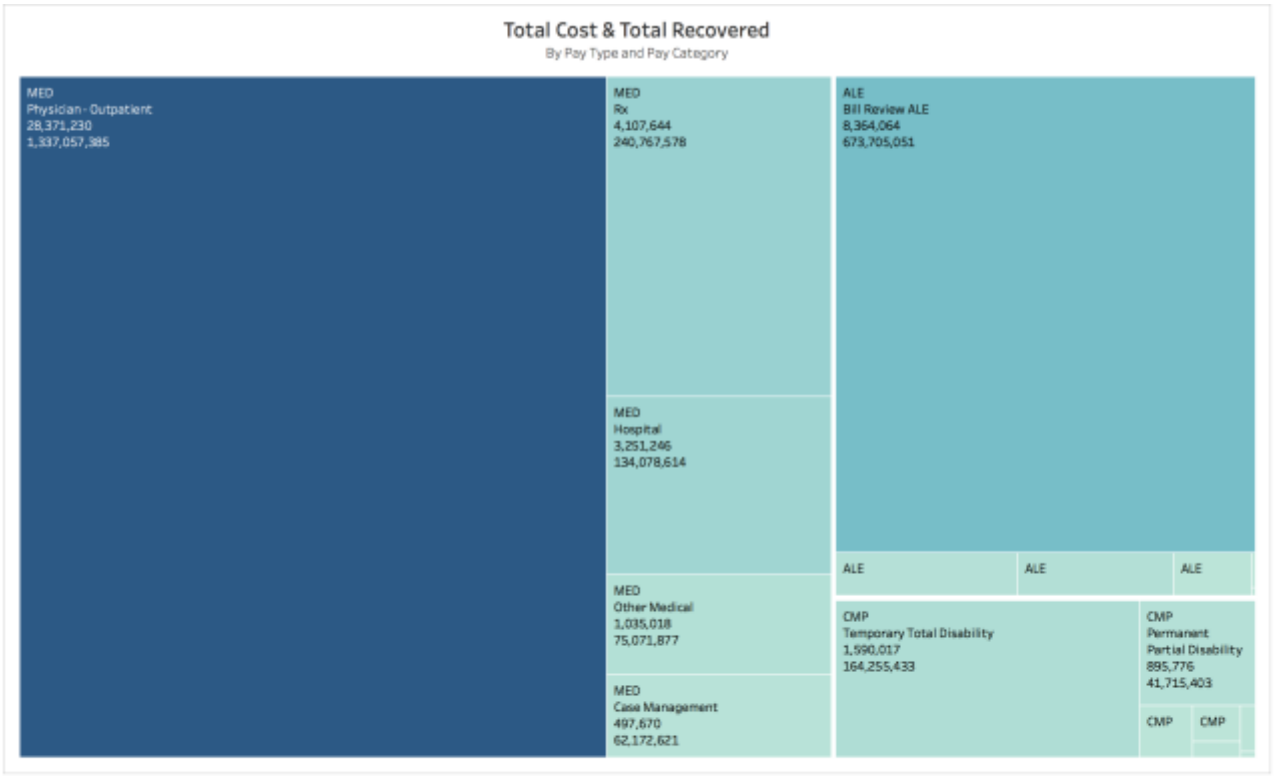
Litigation - Cost & Recovery | Trend - Denied vs Litigated | **Litigation costs** | Claimant Age vs Total Cost | Total Cost & Recovered - Pay Typ... | Costs - Indemnity & Medical | Weekly Wage & Indemnity per Year | Top 5 Costly Injury Types

Claimant Age vs Total Incurred Cost



Caption	Company needs to bear extra costs for workers 40 to 50 years of age. The Company should target these age groups and revise the premium accordingly. Additionally, the Company should also undertake Health Care Programs to take precautionary measures.

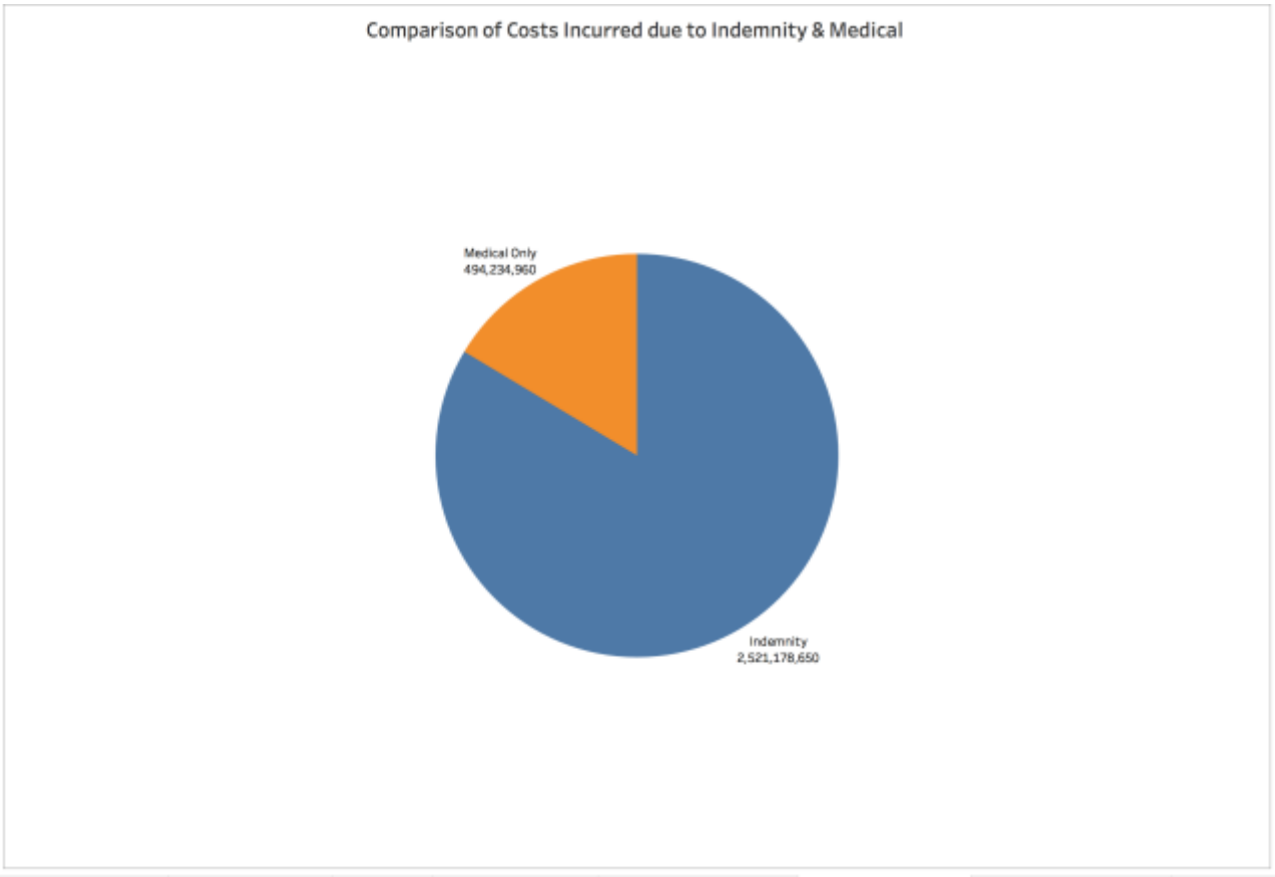
Litigation - Cost & Recovery | Trend - Denied vs Litigated | Litigation costs | **Claimant Age vs Total Cost** | Total Cost & Recovered - Pay Typ... | Costs - Indemnity & Medical | Weekly Wage & Indemnity per Year | Top 5 Costly Injury Types

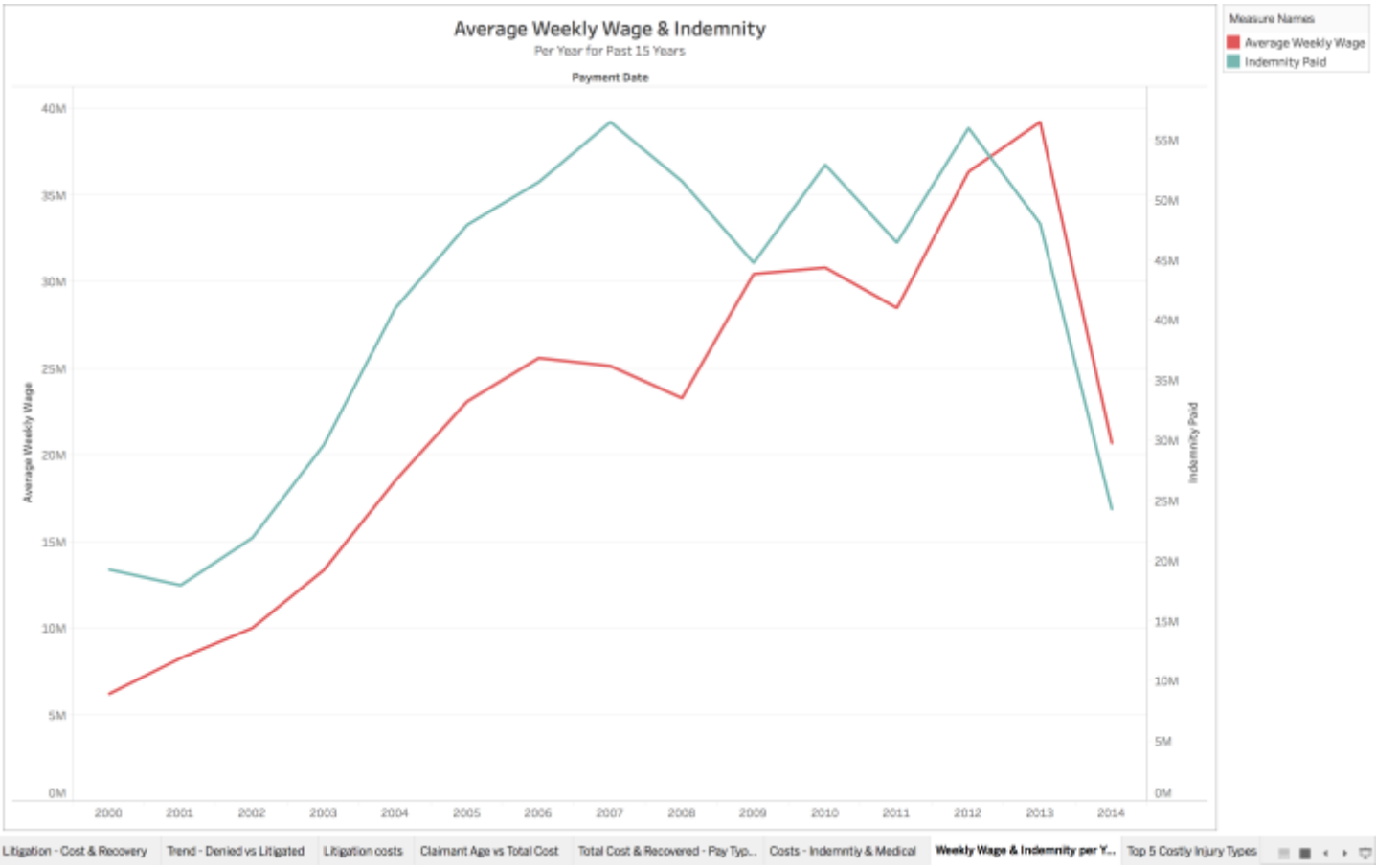


Caption

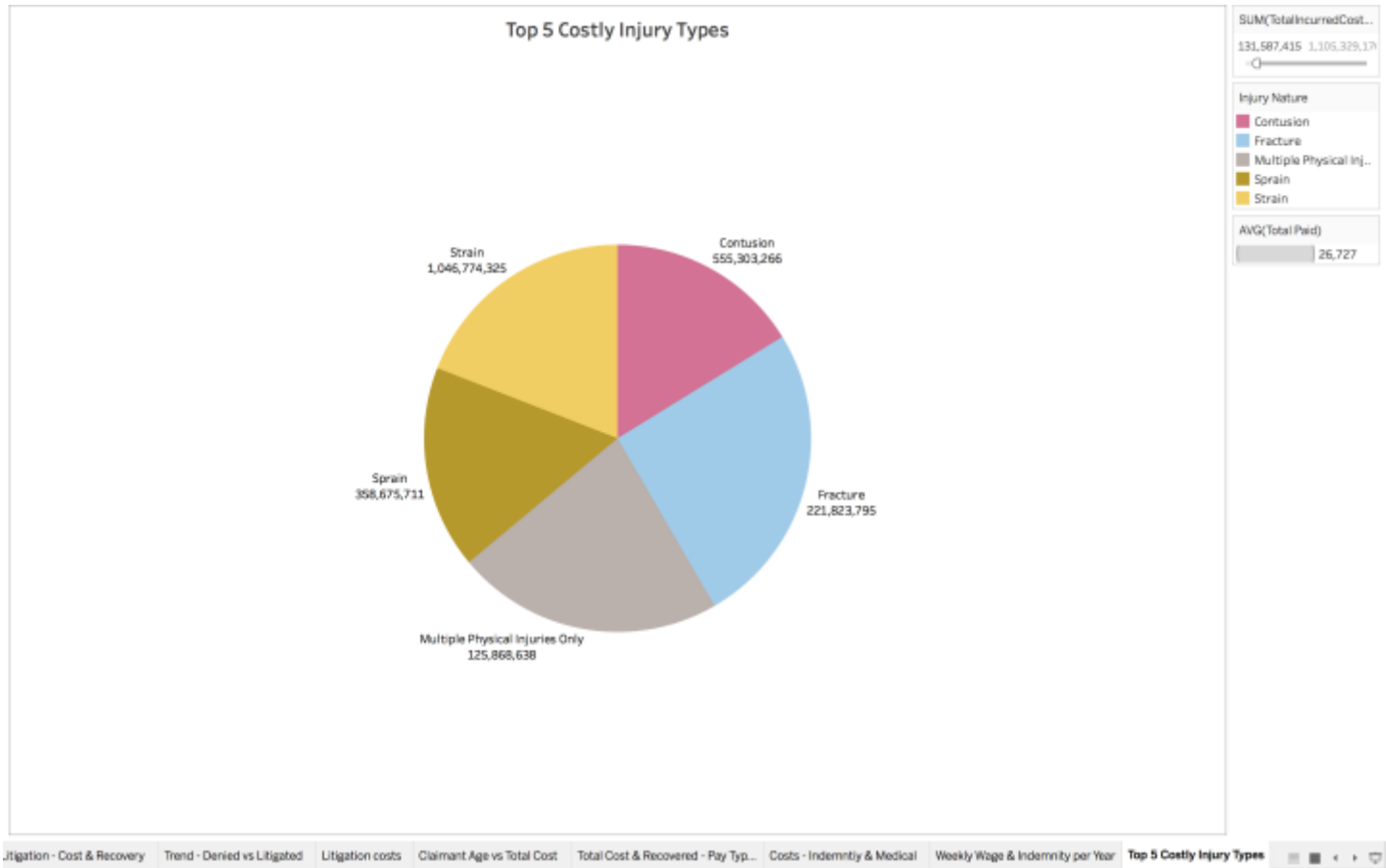
Although there is a direct proportion in the Total Paid and Total Recovered Cost, the amount recovered is very less as compared to the cost paid.

The company should take measures (increasing the premiums, etc) to increase the amount of cost recovery.



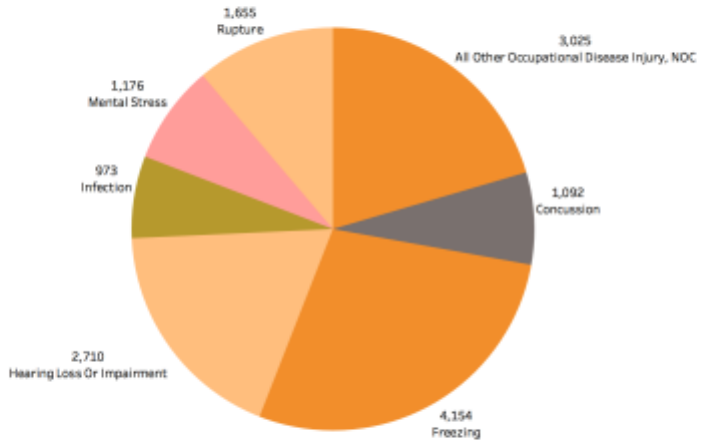


Litigation - Cost & Recovery | Trend - Denied vs Litigated | Litigation costs | Claimant Age vs Total Cost | Total Cost & Recovered - Pay Typ... | Costs - Indemnity & Medical | **Weekly Wage & Indemnity per Y...** | Top 5 Costly Injury Types



Litigation - Cost & Recovery | Trend - Denied vs Litigated | Litigation costs | Claimant Age vs Total Cost | Total Cost & Recovered - Pay Typ... | Costs - Indemnity & Medical | Weekly Wage & Indemnity per Year | **Top 5 Costly Injury Types**

Largest Fund Reserves according to injuries

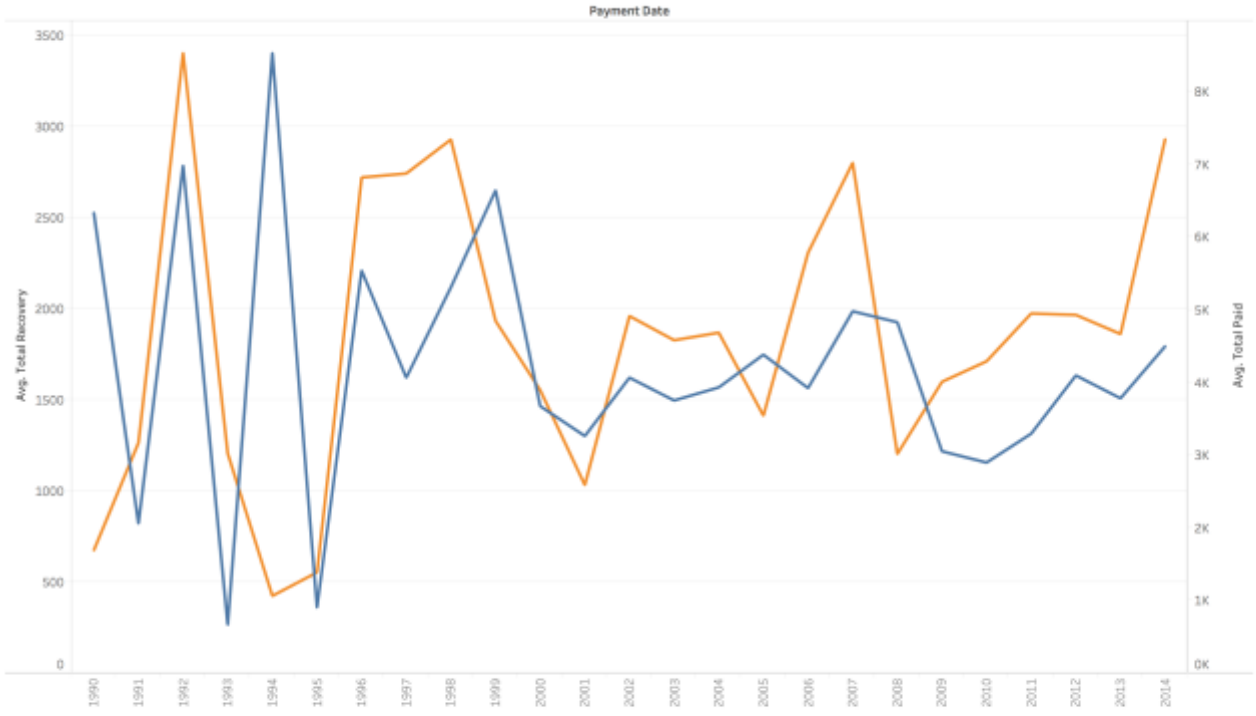


AVG(Total Reserves)
956 4,154

Injury Nature
- All Other Occupation..
- Concussion
- Freezing
- Hearing Loss Or Imp..
- Infection
- Mental Stress
- Rupture

Recovery | Trend - Denied vs Litigated | Litigation costs | Claimant Age vs Total Cost | Total Cost & Recovered - Pay Typ... | Costs - Indemnity & Medical | Weekly Wage & Indemnity per Year | Top 5 Costly Injury Types | **Fund Reserves**

Comparison of amount recovered and amount paid in Subrogations



Measure Names
- Avg. Total Paid
- Avg. Total Recovery

Caption

Company gains money in subrogation, doesn't it? Take a closer look.

enied vs Litigated | Litigation costs | Claimant Age vs Total Cost | Total Cost & Recovered - Pay Typ... | Costs - Indemnity & Medical | Weekly Wage & Indemnity per Year | Top 5 Costly Injury Types | Fund Reserves | **Costs-Subrogation**