# Workers' Compensation

## Incorporating Data Analytics in Claims Management

Akshata Kanumuri, Dikshya Mohanty, Lei Huang, Muralidharan Singaravel

**DSBA 6100 Fall 2018 – Group Project**

# Table of Contents

## SUMMARY

In this paper, we have looked into how the workers' compensation claims management process works, key issues the insurers face in managing the claims, role of external data in better understanding the problem and how big data can have an impact on fraudulent claims.

Workers' compensation provides cash and medical benefits to workers who are injured or become ill in the course of their employment and provides benefits to the survivors of workers killed on the job. Some of the key issues in managing the workers' compensation claims are – handling bogus claims, the sheer expanse of unstructured data, current technology environment and enhancing consumer experience and engagement.

Leveraging real-time claim analytics solution and predictive analytics tools help an insurer to derive insights from the data to identify fraud sooner and more effectively at each stage of the claims cycle.

The data file used in data wrangling and data visualization is from one of the clients of a North Carolina company that manages workers' compensation claims. Key issues which we noticed in the files are missing values for various fields like Average Weekly Wage, Claimant Age at DOI, Date service began, Date service ended. Some of the missing values were replaced with mean and the values which could not be replaced, the fields were dropped in the final aggregated data.
In data visualization, by running various tableau reports on different claims dimensions and measures, we could see some interesting insights like the number of claims in the last quarter across all the years is less compared to the other quarters, lower back is the most frequently injured body part which also resulted in the most number of fatalities and Contagious Disease resulted in the highest average amount of indemnity settlement.

## SECTION I

### 1.1 WHAT IS WORKERS' COMPENSATION?

Workers' Compensation (WC) is insurance designed to provide medical care and wage replacement to employees for job-related injuries. Workers' compensation provides cash and medical benefits to workers who are injured or become ill in the course of their employment and provides benefits to the survivors of workers killed on the job. Benefits are provided without regard to fault and are the exclusive remedy for workplace injuries, illnesses, and deaths. Nearly all workers in the United States are covered by workers' compensation.

WC is designed to protect both employees - by providing medical coverage and wage replacement as well as the employer - by protecting them from liability (WC is the sole remedy for job-related injuries).

Overall Goal of Workers' Compensation is to rehabilitate the injured employee so they can return to work and be productive.

### 1.2 WORKERS' COMPENSATION CLAIMS MANAGEMENT PROCESS

After the job related injury occurs, there is a limited amount of time to submit paperwork and receive benefits, so the employee has to act promptly when an injury occurs. It's important to comply with regulations – otherwise the claim could be denied.

**Steps for filing a workers' compensation claim:**

**Step 1: The employee reports an injury to the employer**

To make a workers' compensation claim, the employee's injury or illness must be work-related. The employee must report the injury or illness to the employer within a certain number of days

after it occurs, it varies from one state to another for example in Colorado, the employee must provide written notice within four days of the injury whereas in Iowa, the deadline is 90 days.

**Step 2: The employer provides paperwork**

Once the employer or the department is notified about the injury, the employee is given the claims form, access to an approved medical practitioner, information on the employee's rights, workers' comp benefits and information about returning to work. Failure to give the injured employees this information could lead to lawsuits. As with the accident report, each state has its own laws for how much time can pass after an injury for a workers' comp claim to be filed.

**Step 3: The employer files the claim**

Usually, the employer is responsible for submitting the paperwork to the insurance carrier, and the employee's doctor needs to mail the medical report. There may be a time constraint for submitting these forms, so it should not be delayed

**Step 4: The insurer approves or denies the claim**

Once the claim is filed, the insurer will either approve the claim and contact the employee for payment details or deny the claim if it does not qualify for workers' compensation. The insurance company may choose a doctor to conduct an independent medical examination, which is used to create an appropriate compensation offer.

If the insurer approves the claim, the employee may:

- Accept the payment offer, which may cover costs for medical bills, medicine, disability payments, and a portion of lost wages.
- Negotiate for a lump-sum settlement or larger structured settlement.

If the insurer denies the claim:

- The employer can ask the insurer to review the decision if they believe the decision is wrong.
- The employee can appeal the decision.

**Step 5: The employee returns to work**

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

If employees at the workplace continue to experience occupational injuries, the workers' compensation premium may increase, so the employer should make sure to provide safety training, and hire workers with the skills to do their jobs safely.

## 1.3 KEY CHALLENGES IN MANAGING CLAIMS

**Some of the key issues in managing the workers' compensation claims are:**

a. **Bogus claims:** Dishonest employees will knowingly make bogus claims for workplace injuries against their employer's workers' compensation policies. One out of 10 insurance claims is fraudulent. Most fraud solutions on the market today are rules-based. Unfortunately, it's too easy for fraudsters to manipulate and get around the rules.
b. **The sheer expanse of unstructured data:** The healthcare industry is full of clinical, claim, patient, and hospital system data. Data is scattered and stored in multiple disparate formats, which are often overlooked or underutilized. As a result, the challenge is how to find the right data or how to position or the map the data to actionable KPIs.

c. **Current technology environment:** Insurance claims systems often consist of multiple, decades-old platforms, applications, and databases - each supporting a different part of the operation. New modern systems have been added to legacy platforms over the years to add functionality, further complicating the claims system. Different systems target different KPIs (Key Performance Indicators), causing huge process loopholes, and red flags. The vast divide in the technology landscape is not allowing systematic handshakes, hence other issues, such as security, are adding to the delays and confusion.

d. **Consumer experience and engagement:** The pace with which we want our services delivered and the high standard of excellence expected has led all sectors in business, including health and workers' compensation, to consider their definition of consumer. Failure to engage a consumer leads to complaints and negative PR, possibly lack of treatment adherence, whereas high levels of consumer experience leads to positive outcomes.

## 1.4 HOW BIG DATA CAN HAVE A BIG IMPACT ON FRAUDULENT CLAIMS

Big data isn't just useful for insurers looking to reduce fraud, it also helps society as a whole by keeping premiums affordable and ensuring that the insured receive accurate claims settlements. During the claims request, companies can leverage internal data sources with unstructured data along with combining big data with graph theory algorithms, to identify whether the claim is legitimate. Real time monitoring, through social media and digital channels provide greater insight throughout the claims cycle. Predictive analysis, on the other hand, uses a combination of rules, modeling, text mining, database searches and exception reporting to identify fraud sooner and more effectively at each stage of the claims cycle.

Fraud detection not only benefits the insurance company, but as a result legitimate claims can be processed more efficiently.

## 1.5 ROLE OF EXTERNAL DATA IN BETTER UNDERSTANDING THE PROBLEM

Information used to support the claims and provide simple and actionable outcomes can be derived from data both generated and held within organizations and from data that is generated and hosted by external data collectors and organizations. Some of the external data that can provide valuable insights to this particular datasets are like electronic health records (EHRs), past claims, and demographics. Analytics on these external data can help detect hidden patterns in information to deliver actionable insights.

- EHRs can be obtained from healthit.gov which contains a patient's medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results which the insurance providers can use to make decisions about a patient's claims.
- Workers Comp Claims History Reports can be obtained from Workers' Compensation Board or sites like national-employment-screening.com to check if the worker previously injured the same body part in another workplace accident or a condition that pre-dates the current industrial injury.
- Demographics data such as age, gender, state distribution, jurisdiction and cause of serious injury workers can be a reconsideration of estimates for future workers compensation insurance costs for workers suffering fatalities or disabling injuries in the workplace. This data can be obtained from Department of Labor.

## SECTION II

### 2.1 DESCRIPTION OF THE DATA WRANGLING AND DATA MERGING STEPS

1. Download the three data files – ClaimsDataFall2018.xlsx, Transactions_Part1.csv, and Transactions_Part2.csv and copy the files to a new folder in Google Drive.
2. Launch SAS Enterprise Guide and open a new project. Then, use the "Assign Project Library" under Tools menu to set up a new library and point it to the folder where the data files are copied.
3. Import Transactions_Part1.csv and Transactions_Part2.csv and make sure that all the column names and data type matches.
4. Append the two tables using Tasks→Data→Append Table option. Name the new concatenated dataset as Transactions_Append
5. Identify any data issues in Transactions_Append.sas7bdat dataset, by using Tasks→Describe→Characterize Data

   Following Data issues are identified in the Transactions_Append along with the plan of action:

   i. PayType has 4392 missing values, which is just 0.2% of the total records, so we can remove all the records with missing PayType.

| Variable | Label | Value | Frequency Count | Percent of Total Frequency |
|---|---|---|---|---|
| PayType | | MED | 1454734 | 56.7127 |
| | | ALE | 553992 | 21.5973 |
| | | CMP | 551975 | 21.5187 |
| | | NULL | 4392 | 0.1712 |

   ii. PayCategory has 4392 missing values, which is just 0.2% of the total records, so we can remove all the records with missing PayCategory.

| Variable | Label | Value | Frequency Count | Percent of Total Frequency |
|---|---|---|---|---|
| PayCategory | | Physician - Outpatient | 901037 | 35.1269 |
| | | Bill Review ALE | 508243 | 19.8138 |
| | | Temporary Total Disability | 447707 | 17.4538 |
| | | Rx | 307634 | 11.9931 |
| | | Hospital | 105011 | 4.0938 |
| | | Case Management | 76274 | 2.9735 |
| | | Other Medical | 64851 | 2.5282 |
| | | Permanent Partial Disability | 34614 | 1.3494 |
| | | Litigation | 18306 | 0.7137 |
| | | Other Permanency | 17971 | 0.7006 |
| | | Other Expenses | 16807 | 0.6552 |
| | | Other Indemnity | 15311 | 0.5969 |
| | | Settlement | 14163 | 0.5521 |
| | | Permanent Total Disability | 11974 | 0.4668 |
| | | Temporary Partial Disability | 10231 | 0.3989 |
| | | PRIOR EXP | 9610 | 0.3746 |
| | | NULL | 4392 | 0.1712 |

iii. ServiceFromDate has 11215 missing values, which is just 0.4% of the total records, so we can remove all the records with missing ServiceFromDate. We will not be using this column on the final merge list, so this is okay.

| Variable | Label | N | NMiss | Min | Mean | Median | Max |
|---|---|---|---|---|---|---|---|
| ServiceFromDate | | 2553878 | 11215 | 12JAN84 | 05NOV07 | 25MAR09 | 31OCT97 |

iv. ServiceToDate has 11214 missing values, which is just 0.4% of the total records, so we can remove all the records with missing ServiceToDate. We will not be using this column on the final merge list, so this is okay.

| Variable | Label | N | NMiss | Min | Mean | Median | Max |
|---|---|---|---|---|---|---|---|
| ServiceToDate | | 2553879 | 11214 | 15JAN87 | 29NOV07 | 31MAR09 | 31OCT97 |

6. Import ClaimsDataFall2018 and make sure that ClaimIdentifier column datatype matches the datatype in transactions_append table .
7. Identify any data issues in ClaimsDataFall2018.sas7bdat dataset, by using Tasks→Describe→Characterize Data

Following Data issues are identified in the ClaimsDataFall2018.sas7bdat along with the plan of action:

i. AverageWeeklyWage has 84924 missing values, which is 63% of the total records, so we will replace the null values with mean.

| Variable | Label | N | NMiss | Total | Min | Mean | Median | Max | StdMean |
|---|---|---|---|---|---|---|---|---|---|
| AverageWeeklyWage | | 49080 | 84924 | 28826777.18 | 0.01 | 587.34 | 491.97 | 2024000 | 50.34 |

ii. ClaimantAge_at_DOI has 45078 missing values, which is 34% of the total records, so we will replace the null values with mean. Further analyzing the column we see that there are close to 300 records with invalid workers' age ( age <14)

| Variable | Label | N | NMiss | Total | Min | Mean | Median | Max | StdMean |
|---|---|---|---|---|---|---|---|---|---|
| ClaimantAge_at_DOI | | 88926 | 45078 | 3543915.00 | -8000.00 | 39.85 | 42.00 | 94 | 0.40 |

iii. IncidentDescription has 18 missing values, so we can remove all the records with missing IncidentDescription

| Variable | Label | Value | Frequency Count | Percent of Total Frequency |
|---|---|---|---|---|
| IncidentDescription | | I | 55 | 0.0410 |
| | | Unknown | 27 | 0.0201 |
| | | EE was exposed to scabies | 24 | 0.0179 |
| | | Work exposure to inmate with bac | 22 | 0.0164 |
| | | NULL | 18 | 0.0134 |

iv. ReturnToWorkDate has 58637 missing values value which is 77 % of total value for that column, since we cannot determine what value to put, we will drop this column in the final merged dataset.

| Variable | Label | N | NMiss | Min | Mean | Median | Max |
|---|---|---|---|---|---|---|---|
| ReturnToWorkDate | | 75367 | 58637 | 29OCT76 | 01JUN06 | 09JUL07 | 07MAY15 |

8. Aggregated columns from transactions dataset to be added to claims data

i. TransactionID: This column will help identify all the transactions for a Claim.

ii. PayCategory: This column will help identify all the pay categories for each of the transactions for a Claim.

iii. PaymentAmount: This column will help identify the payment amount for each of the pay categories for a transaction in a claim.

iv. PaymentDate: We are tracking at the transaction level, this date will give the date on which the payment was made, which may be different from the claimant close date.

## 2.2 FINAL AGGREGATED DATA SET

1. The final merged dataset has the following columns:

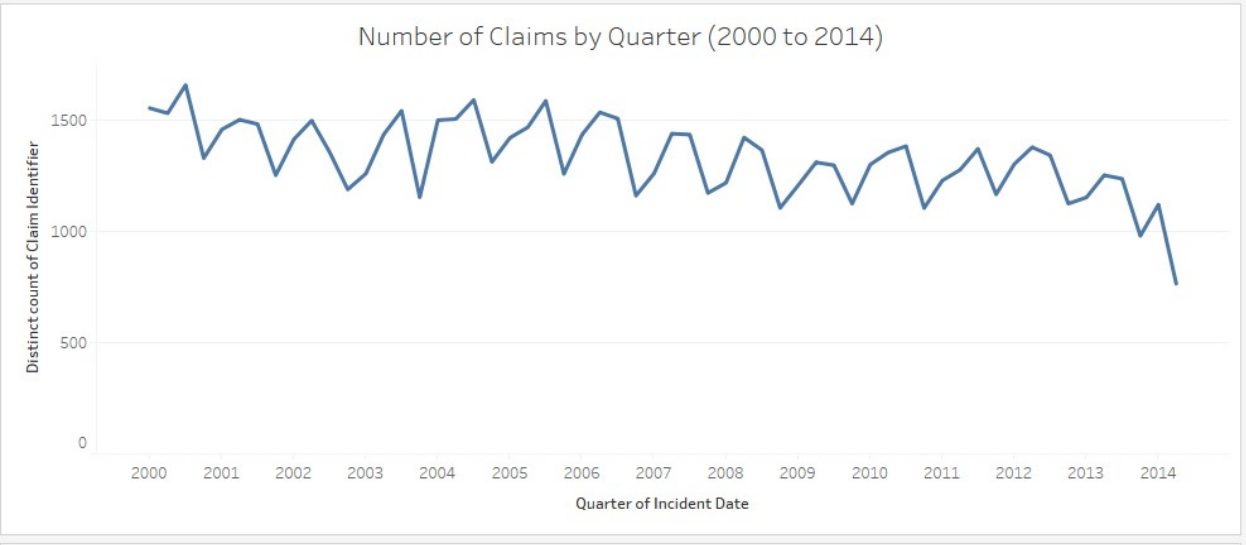
    1. ClaimIdentifier
    2. TotalPaid
    3. TotalReserves
    4. TotalRecovery
    5. IndemnityPaid
    6. OtherPaid
    7. ClaimStatus
    8. IncidentDate
    9. ClaimantOpenedDate
    10. ClaimantClosedDate
    11. ReceivedDate
    12. IsDenied
    13. IsFatality
    14. Claimant_Age_DOI
    15. Gender
    16. ClaimantType
    17. InjuryNature
    18. BodyPartRegion
    19. BodyPart
    20. TransactionID
    21. PayCategory
    22. PaymentDate
    23. PaymentAmount
    24. Time_To_Process

## SECTION III

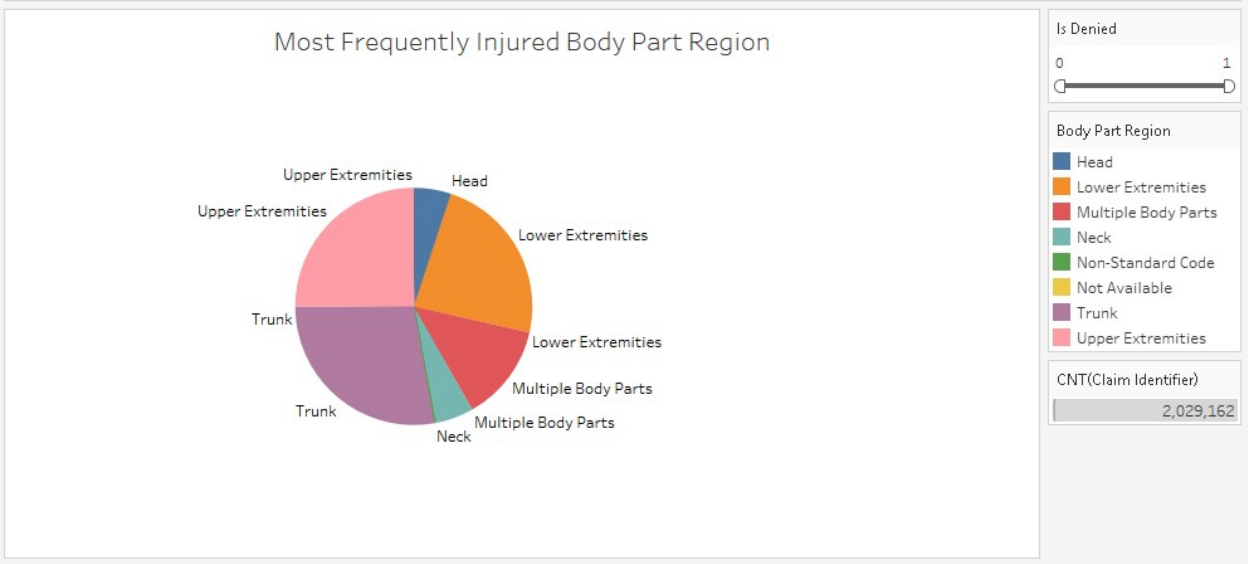
### 3.1 DATA VISUALIZATION IN TABLEAU

1. **Number of Claims by Quarter (Years 2000 to 2014):** From the plot we can see that there was an increase in the number of claims in the year 2004 and 2005. Also, we can see that the number of claims in Q4 across all the years is less compared to the other quarters, the reason may be less number of working days or holiday season.

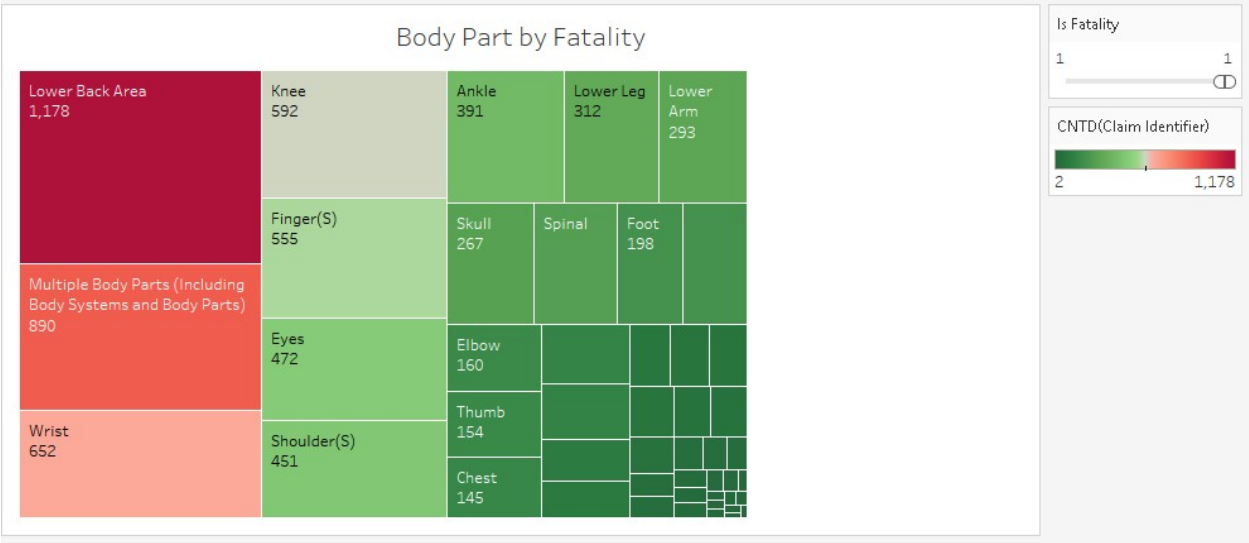
   The Insurer can work on reduced staffing capacity in the last quarter as the number of claims are comparatively less than Q1, Q2 and Q3.




2. **Most Frequently Injured Body Part Region:** From the plot we can see that Upper Extremities is the most frequently injured body part in all the claims.

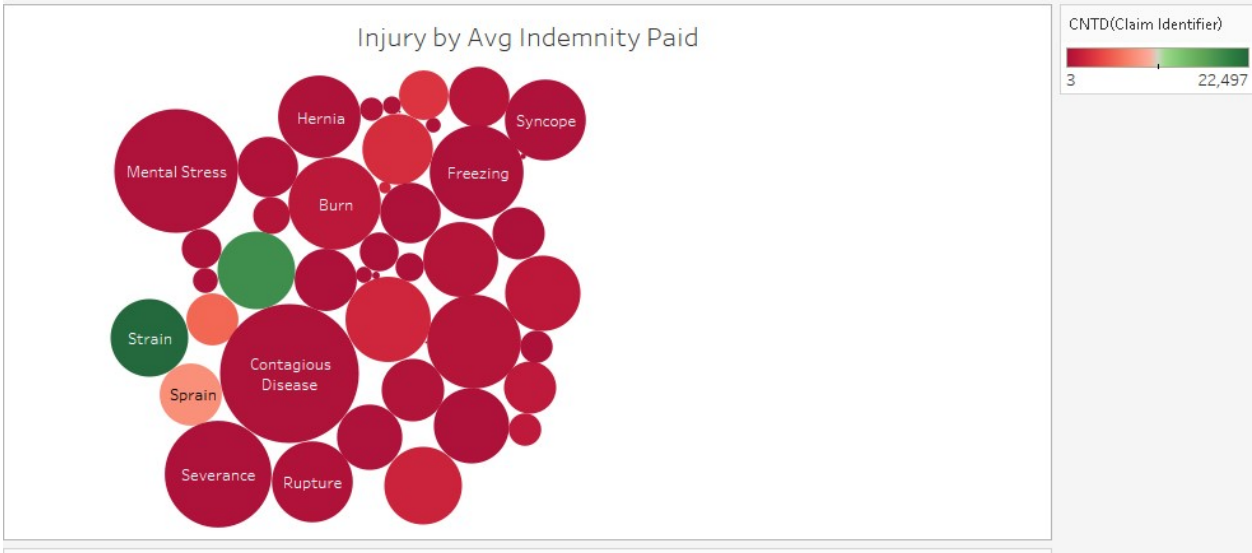



3. **Most Frequently Injured Body Part which resulted in Fatality:** From the plot we can see the most number of workers' death were due to the Lower Back Area Injury. Employee can take extra precaution to avoid strain Injury which is one of the major contributors to Lower Back Area Injury. An employer has legal obligations to provide safe and healthy workplaces for employees, so to reduce the risk of back pain in the workplace the employer should provide information, training and supervision on safe ways to work or provide special programs like provide occupational physiotherapists which can prevent long-term back pain and be highly cost-effective.




4. **Injury by Avg Indemnity Settlement:** From this Bubble map, we can see that the injury - Contagious Disease resulted in the highest avg amount of indemnity settlement.

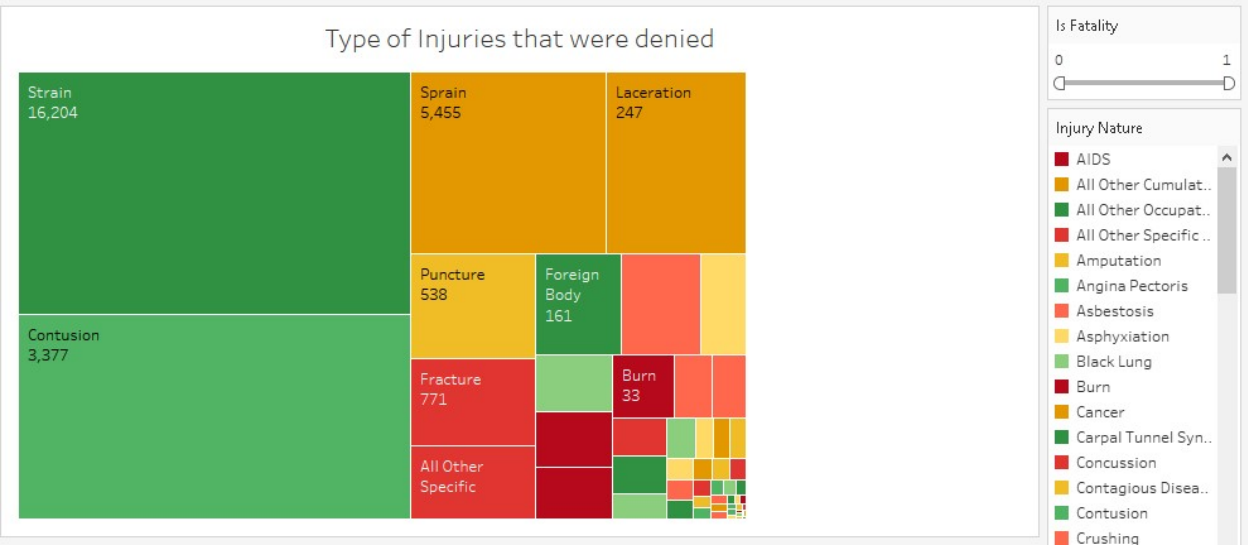
   Contagious illnesses in the workplace can harm productivity and quickly spread among staff. Addressing the issue of contagious diseases and ensuring employees know the best course of action if they are taken ill can help reduce the spread and the impact on overall productivity and efficiency.



5. **Type of Injury Claims that were denied:** From this Tree map, we can see that the type of Injury Nature - Strain resulted in the maximum of denials followed by Sprain for both fatal and non-fatal injuries.

   A claim can be denied for simply not meeting the eligibility requirements or other common reasons like injury was not reported in time or insufficient evidence that injury is work related.

   A claim denial is an additional work for all the parties involved in the claim cycle, the employee may want to appeal the decision which will result in added work for both the employer and the insurer.

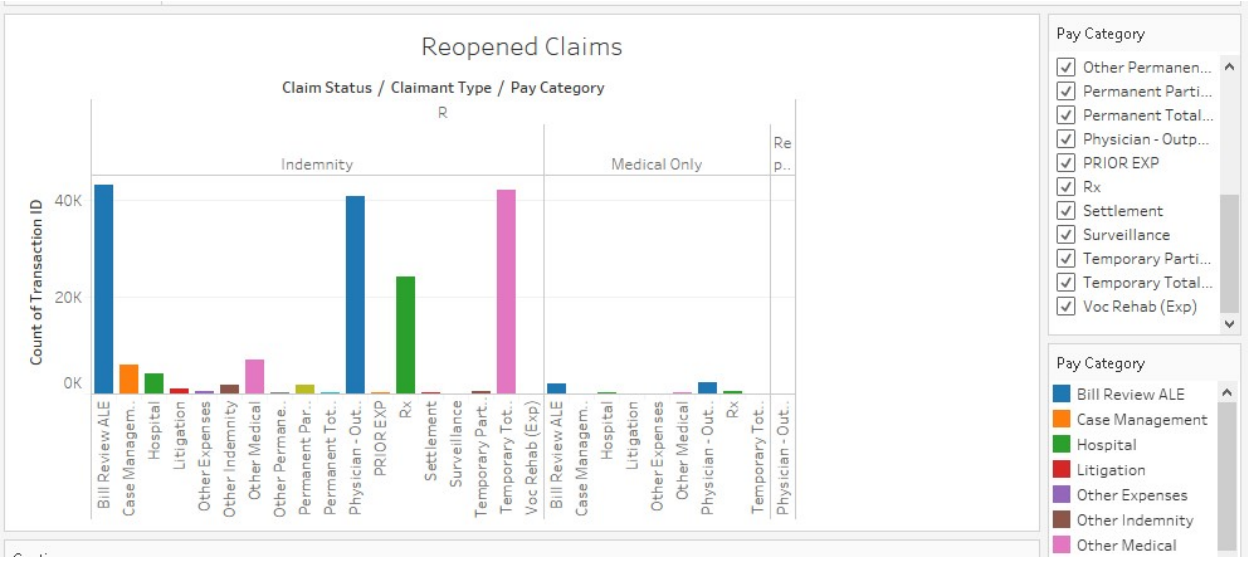



6. **Reopened Claims by Claimant Type and Pay Category:** From the plot, we can see that the 'Claimant Type' Indemnity has the highest number of reopened claims compared to Medical Only.

In Indemnity Claimant type, we can see that Bill Review ALE has the most number of reopened claims which suggests that the employee has new and material evidence to justify the reopening.

A reopened claim may be a claim filed for a benefit that was previously denied and the decision to deny has become final or it is evident that the current symptoms are associated with original injury.

A reopened claim is again an additional work for all the parties involved in the claim cycle which can be reduced by focusing on the employee physical recovery and the employer should not rush the employee back to work.



**REFERENCES**


1. Current Issues in Workers' Compensation James Robert Chelius, Rutgers University https://research.upjohn.org/up_press/192/
2. Workers' Compensation: Overview and Issues Scott D. Szymendera, https://digitalcommons.ilr.cornell.edu/key_workplace/1905/
3. http://www.genpact.com/insight/blog/five-ways-big-data-can-have-a-big-impact-on-insurance-claims
4. https://www.sas.com/en_us/insights/articles/risk-fraud/big-data-analytics-improves-claims-processing.html
5. https://www.employers.com/resources/faqs/claims-faqs
6. https://risk.lexisnexis.com/insights-resources/white-paper/the-value-of-incorporating-data-and-analytics-in-claims-handling
7. Advanced Claims Analytics With Big Data A Scalable Health White Paper https://www.scalablehealth.com/Resources/WP/SH_WP_ADVANCED%20CLAIM%20ANALYTICS.pdf