DSBA 6100 Big Data Fall 2017

Report on

Claims Analysis

Submitted by

Clay Collins

Piyusha Biswas

Shravya Maroli

Sonika Rajan

DSBA 6100 Big Data Fall 2017

Executive summary

Initial Research

Our initial research was more inclined towards understanding the Insurance domain in handling claims. We did some study of the variables in the dataset to understand what all of them stand for and how they affect an insured's claim specific to worker's compensation. Visualization furthered our insights and we were clearly able to see some trends and understand patterns which very well aligned with our understanding on the topic.

Data issues and actions

Following were the issues that were observed:Many observations with "NULL" variables. Observations with "NULL" age values were removed. Two columns in the two transaction datasets ServiceToDate and ServiceFromDate were classified differently in part 1 and 2. These variables were removed before appending the tables together and merging the transaction data with the claims data.Unusual observations removed. Aggregate variables created for number of transactions performed per claimant, sum of payment amount, and total incurred cost per claim.

Findings from Visualization

Visualization highlighted many more interesting patterns in data. Following are some of the key findings:2009 has seen highest claims and total payment, total recovery etc. Average weekly wage of claimants also rise.There is a decrease in claims after 2009.Men are more likely to make a claim then women. The majority of payouts are given to claimants with strains, but

severe injuries receive higher payouts. We thought of concentrating on the potential reasons of why this could be possible and what areas needs to be concentrated on.

Appendix

Research Report

Our first step towards the project was to understand claims handling in Insurance better. During our research we particularly focused on Workers' compensation type of Insurance, since our dataset was a Workers' compensation claims dataset. Despite a company's best efforts to create a safe work environment, accidents can happen. So it is very important to have a Workers' compensation insurance as soon as one starts working. Under the law in most states, every business must have some form of workers' compensation insurance to cover injured employees. Filing a workers' compensation claim is similar to filing an insurance claim; it isn't a lawsuit against an employer, but rather a request for benefits. We found that not all claims are paid. There are certain criteria based on which a claim is handled. Let us understand what a claim really means: An insurance claim is a formal request to an insurance company asking for a payment based on the terms of the insurance policy. This means there could be various reasons in ones policy because of which the claim was handled in a certain way. More specifically Workers' compensation can pay for three things when an employee is injured at work: medical bills, recovery costs, and partial missed wages. The claim has to be filed as soon as a worker become injured on the job or develop an illness that is related to your work. a significant amount of time between when the injury allegedly occurred and when it is reported can throw up red flags. With workers' comp, indemnity describes payments made to an injured or sick employee whose injury or illness occurred because of employment. Workers' compensation indemnity attempts to

compensate the employee for lost wages and make the employee financially whole. The indemnity payment further depends on the type of injury. The policy may also cover other expenses like medical bills, conveyance depending on individual policy structure. It is of utmost importance that the employer is notified of the injury with 30 days of occurrence to be eligible to file a claim. The workers' compensation benefits vary as per state rules/ company rules. For our project, the formulae for certain metrics were provided for handling the claims. Claimant age and health also plays a crucial role in claim payment. Any latent disease that had been growing and worsened during working hours can jeopardize a claim payment.

Typically, there are three basic eligibility requirements for workers' compensation benefits:

- The person or company you were working for must carry workers' compensation insurance or be legally required to do so.
- You must be an employee of that person or company.
- Your injury or illness must be work-related.

Apart from private insurers the federal government also has some Workers' benefits. The Department of Labor's OWCP administers four major disability compensation programs which provides to federal workers (or their dependents) who are injured at work or acquire an occupational disease:

- Wage replacement benefits
- Medical treatment
- Vocational rehabilitation
- Other benefits

Other specific groups are covered by:

- Energy Employees Occupational Illness Compensation Program
- Federal Employees' Compensation Program
- Longshore and Harbor Workers' Compensation Program
- Black Lung Benefits Program

There are special rules for certain workers:

- Domestic workers
- Agricultural and farm workers
- Leased or loaned workers
- Casual or seasonal workers
- Undocumented workers

Because of such variation, not all employers are required to have workers' compensation coverage. State laws vary, but an employer's responsibility to provide coverage usually depends on how many employees it has, what type of business it is, and what type of work the employees are doing.

Data wrangling steps

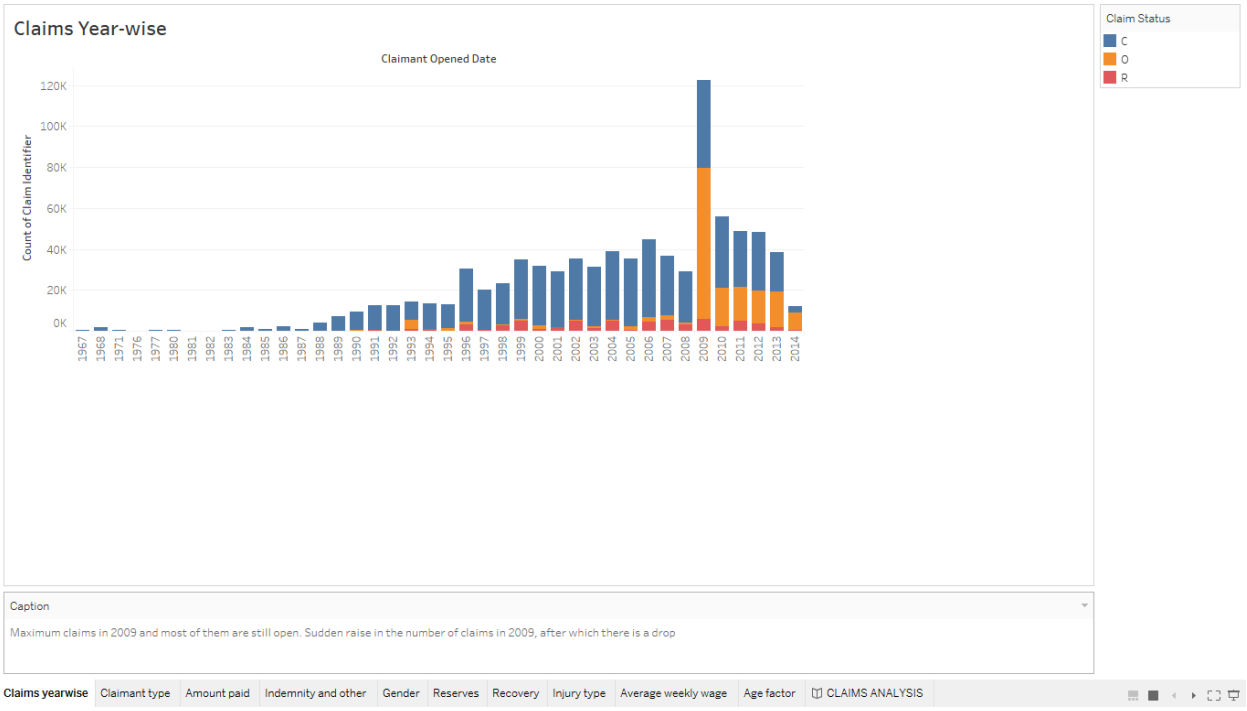
- First, all three datasets (claims, transactions parts 1 and 2) were uploaded into SAS Enterprise Guide.
- The data was characterized to see the issues with the dataset.
    - Some claimants received negative total payouts. This seemed unreal.
    - There were no explicit “missing” values, many entries were marked “NULL”.
    - Some individuals were making less than $1 a week. We are assuming this

dataset comes from the United States or at least a developed country and that these results were errors.

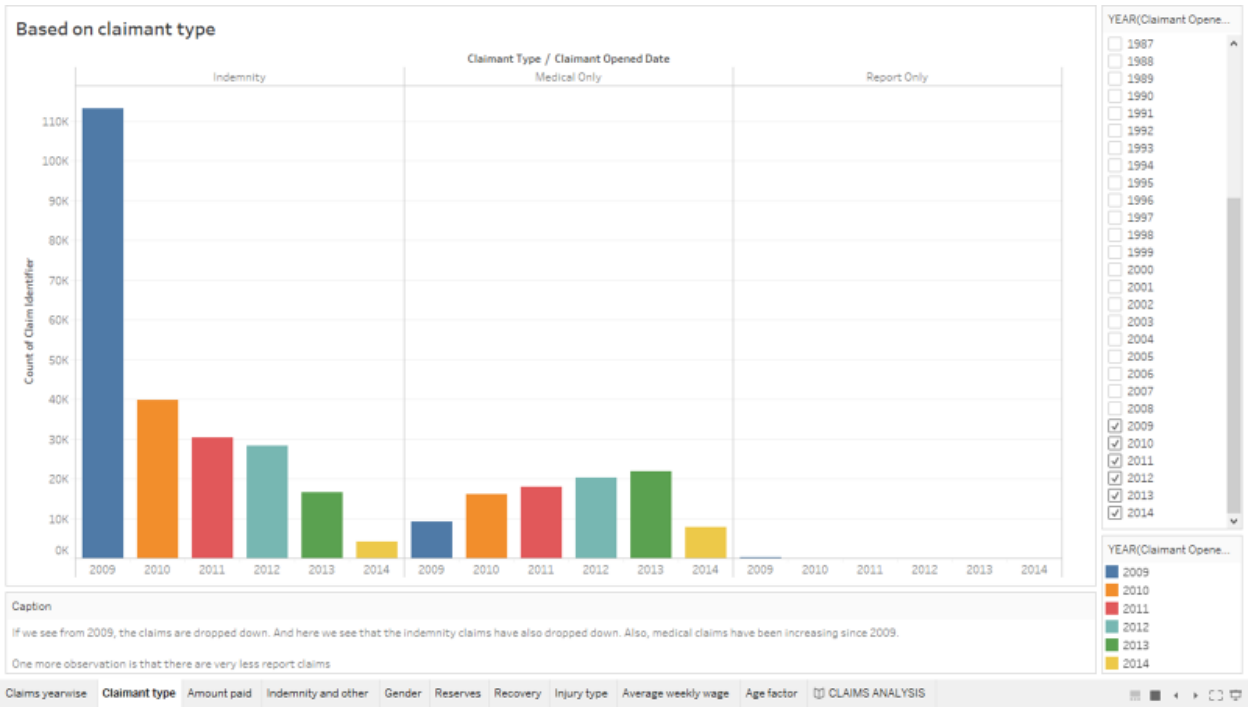
- After Characterization of the dataset, we realized that Part 1 of the transaction data contained several “NULL” variables in the columns ServiceToDate and ServiceFromDate while part 2 did not. This caused these two variables to be labeled as character variables in part 1 but as date variables in part 2. This prevented the datasets from being appended or joined to another dataset. These variables were chosen to be omitted in both datasets, since they were found to be irrelevant to our analysis.
- Both transaction datasets were merged together using the append table command in SAS Enterprise Guide. This created one consolidated transactions dataset.
- The merged transaction dataset was joined with the claims dataset using the query builder with claimant’s ID number being the unique identifier.

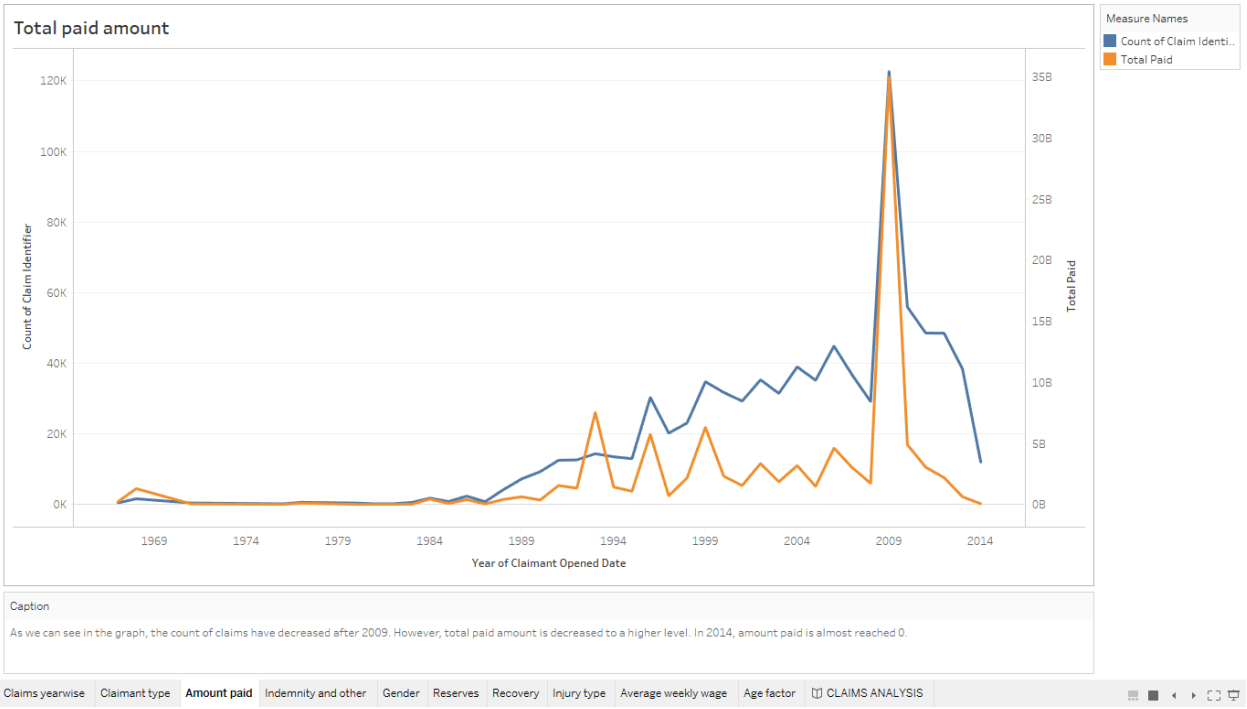
DSBA 6100 Big Data Fall 2017

Visualizations

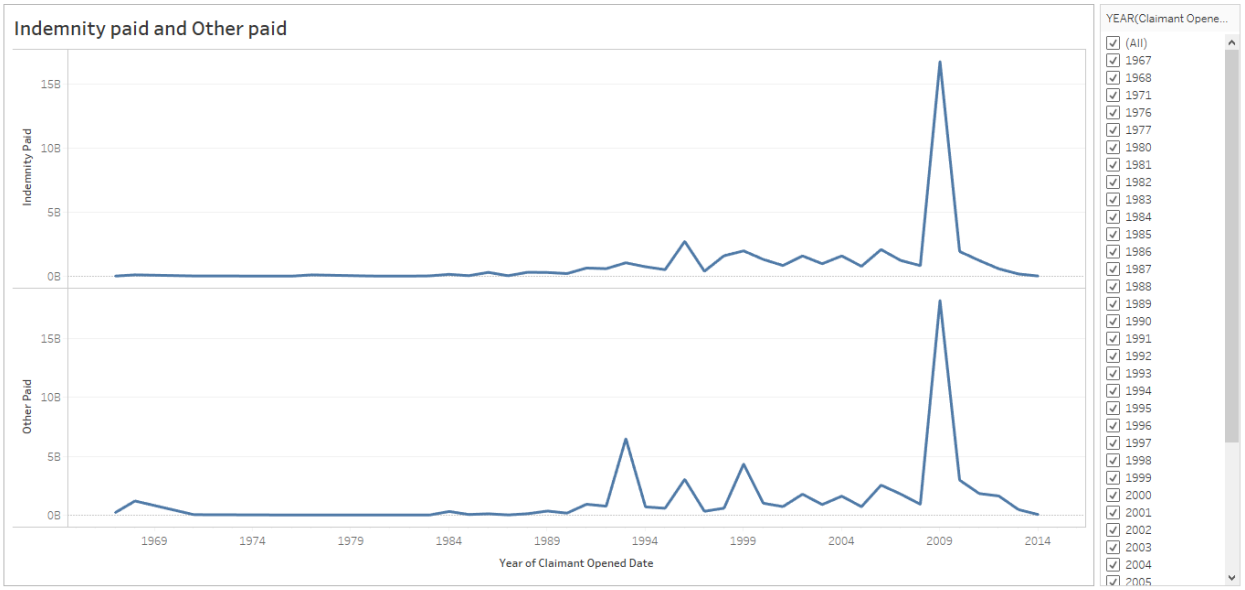


DSBA 6100 Big Data Fall 2017





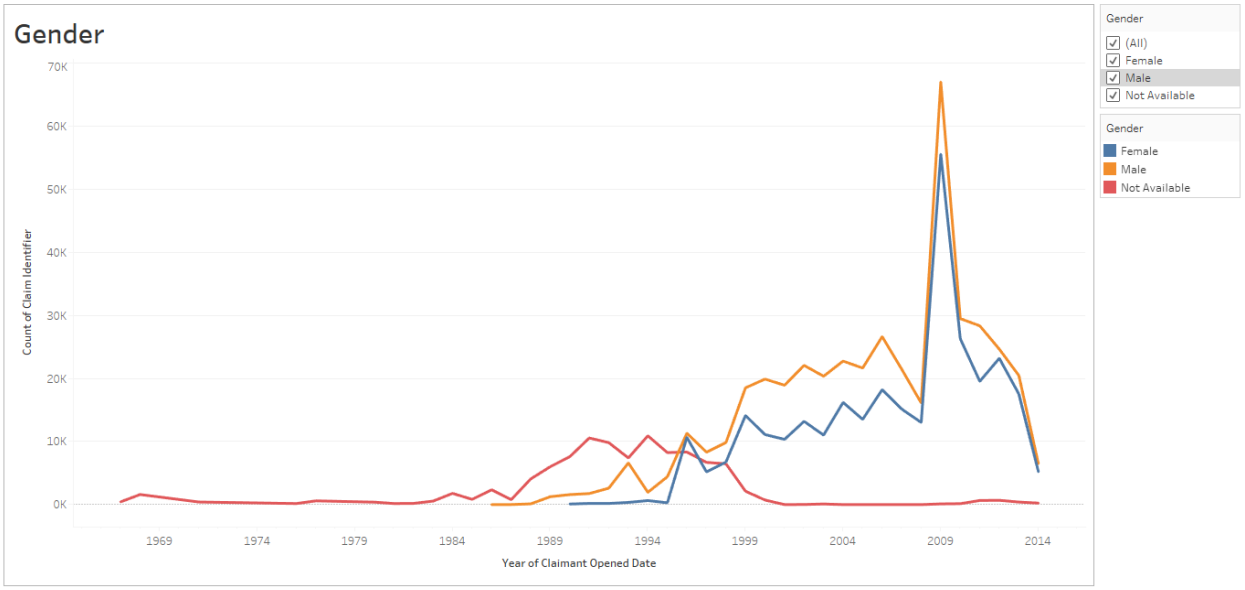
DSBA 6100 Big Data Fall 2017



Caption

There doesnot seem to be any variations in indemnity and other paid w.r.t the date

Claims yearwise | Claimant type | Amount paid | **Indemnity and other** | Gender | Reserves | Recovery | Injury type | Average weekly wage | Age factor | CLAIMS ANALYSIS



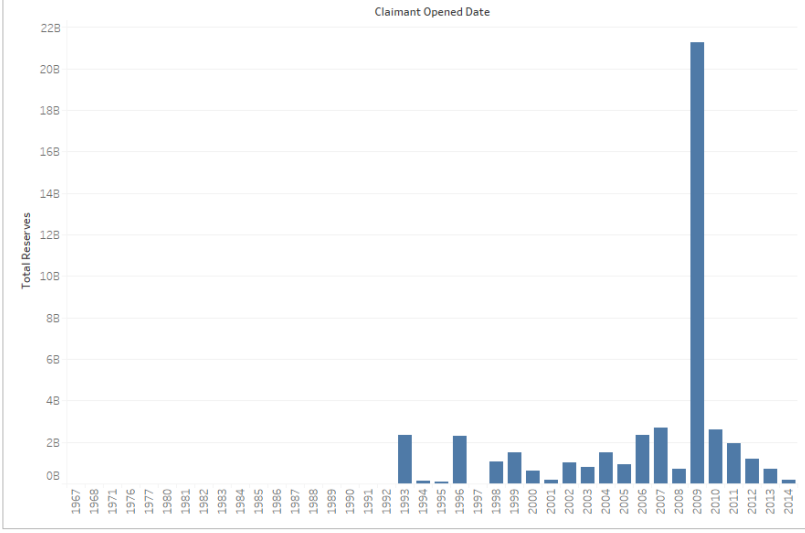
Caption

We could see that upto 1985 gender information was not available. After 1985, it was available

Claims yearwise | Claimant type | Amount paid | Indemnity and other | **Gender** | Reserves | Recovery | Injury type | Average weekly wage | Age factor | CLAIMS ANALYSIS

DSBA 6100 Big Data Fall 2017

Based on Reserves

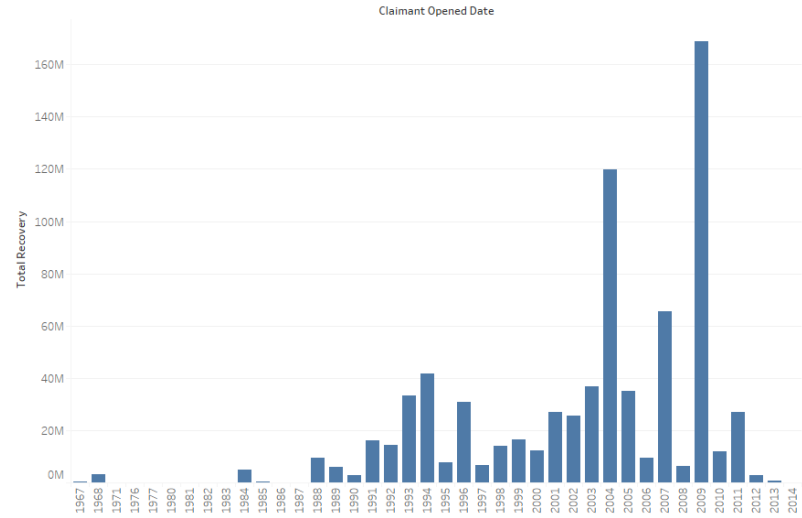


Caption

Reserves made by the company are more in 2009, as claims are more

Claims yearwise | Claimant type | Amount paid | Indemnity and other | Gender | Reserves | Recovery | Injury type | Average weekly wage | Age factor | CLAIMS ANALYSIS

Recovery

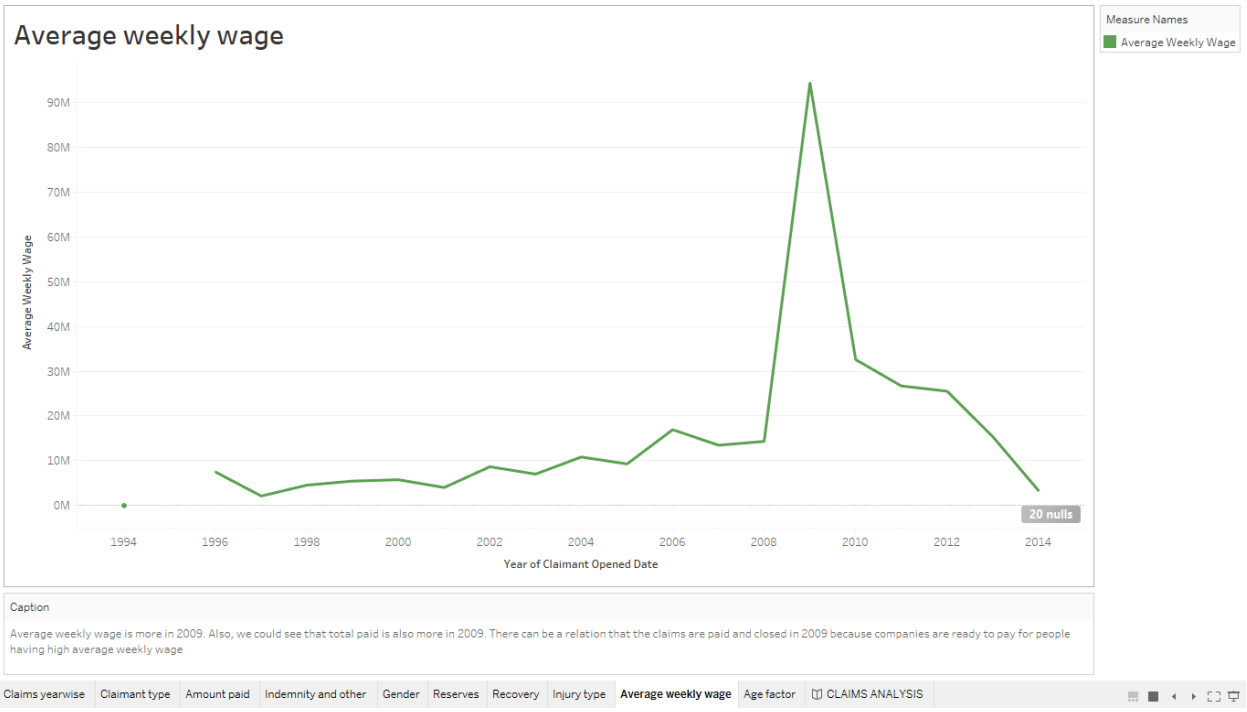
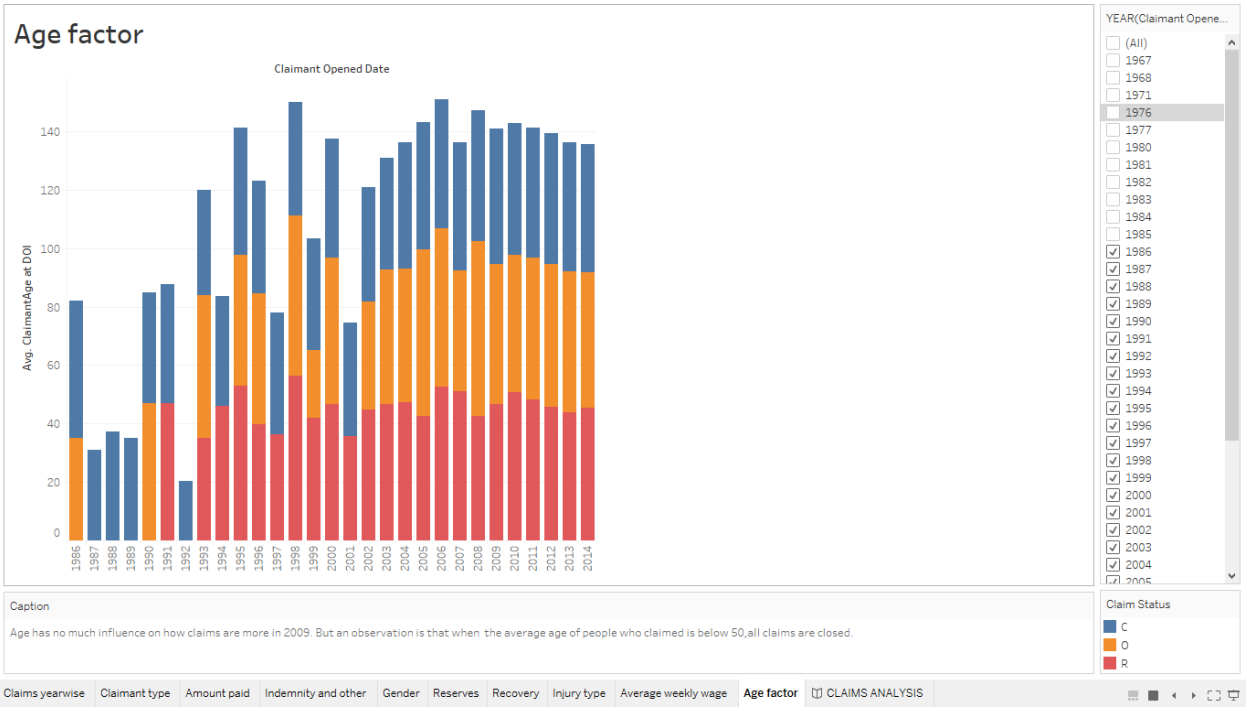


Caption

Sum of total recovery is more in 2009. Also, for 2004 we see an increase. Increase in 2004 was not seen in any other variables except recovery.

Claims yearwise | Claimant type | Amount paid | Indemnity and other | Gender | Reserves | Recovery | Injury type | Average weekly wage | Age factor | CLAIMS ANALYSIS

DSBA 6100 Big Data Fall 2017



DSBA 6100 Big Data Fall 2017