Data Analytics.. A way to look forward

Presentation by Raunak Jha

12th December, 2013, Gurgaon

Agenda

Institute of Actuaries of India

- What are actuarial challenges?
- What Analytics got to do with it?
- How Health Insurers can succeed?

WHAT ARE ACTUARIAL CHALLENGES?

Actuarial Challenges facing Health Insurance Business

- Shift in disease patterns impacting experience of insured population
- Lack of certainty with respect to course of treatment followed by providers
- Increasing market competition
- Fraud and abuse
- Increased awareness of customer has push for product innovation
- Changing medical and healthcare needs of customers

WHAT ANALYTICS GOT TO DO WITH IT?

What is analytics?

- In lay man terms… A science of ***analysis***
  - "Analysis is the process of breaking a complex topic or substance into smaller parts to gain a better understanding of it."

- So what is ***Advanced Analytics***?
  - Advanced analytics focuses on **forecasting** future events and behaviors, allowing businesses to conduct what-if analyses to **predict** the effects of potential changes in business strategies and **optimizing** resources accordingly to meet corporate goals.

How Analytics can help?

Shift in disease patterns (seasonal effect)

- Predicting how many members could contract the disease (e.g. in a group plan)
- How costly it would be for treating the disease for these member
- What would be split between standard course of treatment vs. advanced treatment

Lack of certainty with respect to course of treatment

- Develop simulations models to allow for variations in severity of illness
- Identify provider network which are likely to provide care to optimize outcomes and minimize costs
- Adjust claims projections on daily basis to get insured members needed treatment in cost-effective manner

How Analytics can help?

Market competition

- Cherry picking becomes important
- Build products, programs and service strategies based on **value delivered to customers** (not employer or agent)
- Portfolio optimization (for group plans)

Fraud and Abuse

- Developing dynamic rules engines to identify known types of frauds
- Setting up anomaly detection framework
- Developing predictive models (logit or probit regression models) to measure fraud propensity

HOW HEALTH INSURERS CAN SUCCEED?

Success Mantras…

- Commit to customer centric business model
- Use data from customer interactions to drive business decisions
- Invest in consumer engagement capabilities
- Deploy and monitor complex, automated, multichannel outbound interactions with consumers
- ***Fully leverage data and analytics***

GETTING BEHIND THE NUMBERS

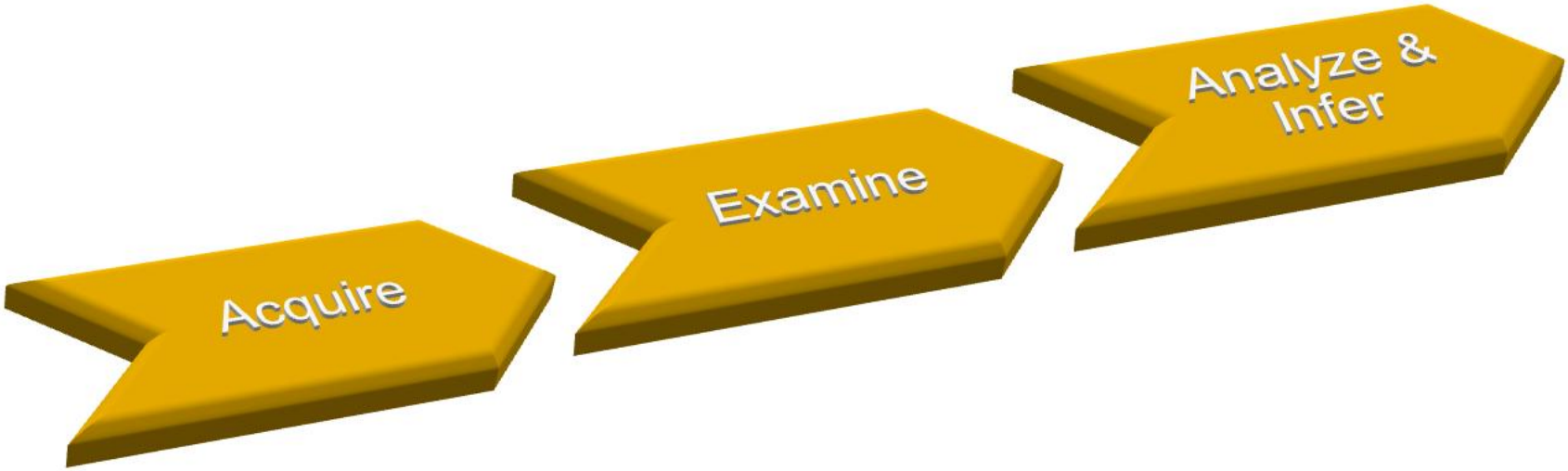
If you think data analytics is a mathematical exercise, you are missing the strategic value of tilting the business odds in your favor!!!

Data Analysis

- **What** is the objective
- **Who** are the end users
  - Level of prudence
  - Level of details
- **How** frequent will the report be generated

User	Frequency	Prudence Level	Report Structure	Detail
Management	High	Moderate	Tailor Made	High
Regulator	Low	High	Fixed	Moderate

Three Step Process

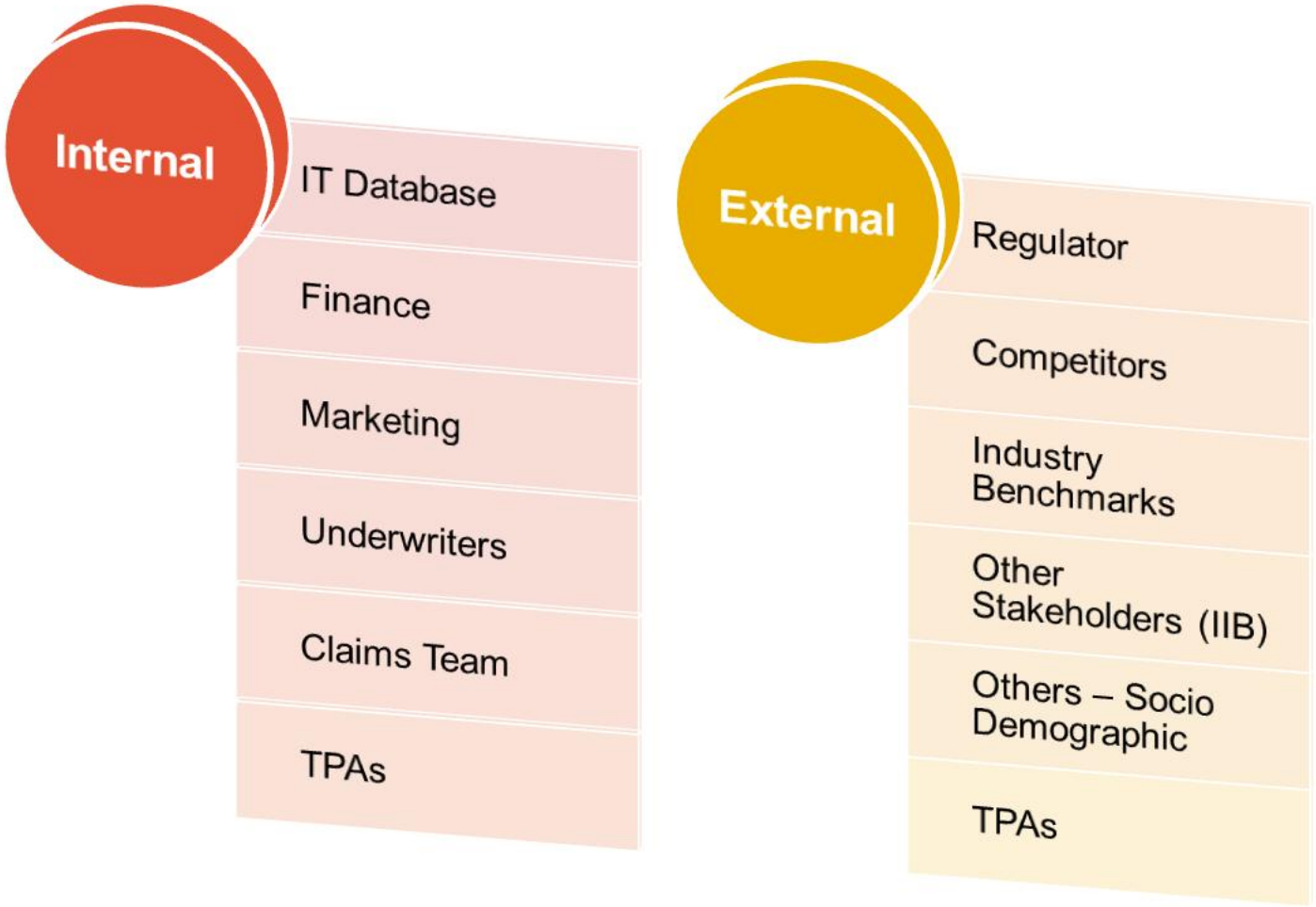


STEP 1 – ACQUIRING DATA

In God we trust, all others must bring Data

– W. Edwards Deming, Statistician

Sources of Data



Data Fields

Policy Characteristics

Insured Details
- Location
- Gender
- DoB & Age
- Previous/ Existing policy details

Medical Details
- Medical certificate
- Medical History
- Past Medical Procedures

Policy Details
- Proposer details
- Sum Insured
- Distribution Channel
- Adjustments/ Endorsements

Claims Characteristics

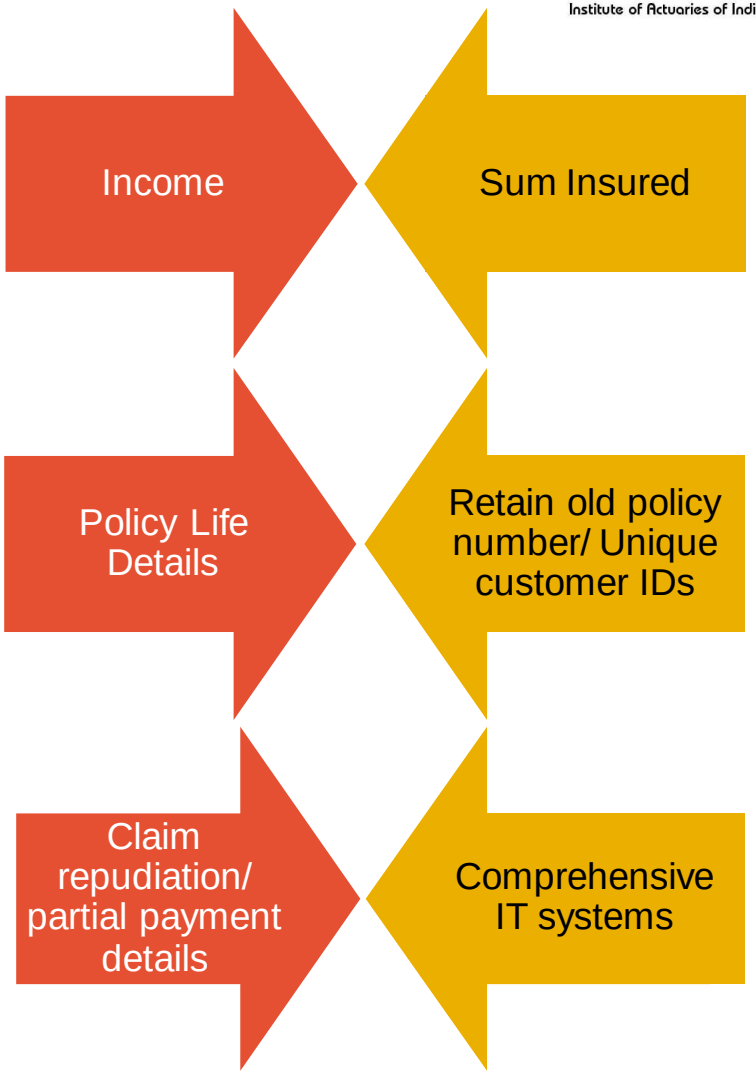
Documents
- Bills & Discharge summary
- All doctors' bills & certificates
- Domiciliary treatment certificates & bills
- Diagnostic test reports

Medical Details
- Nature of disease
- Hospital details
- Medical practitioner details
- Claim type (domiciliary/ hospitalization)

Claim Details
- Claim Amount
- Control/ Transaction dates
- Claims handling expenses
- Reinsurance recoveries

Additional Data

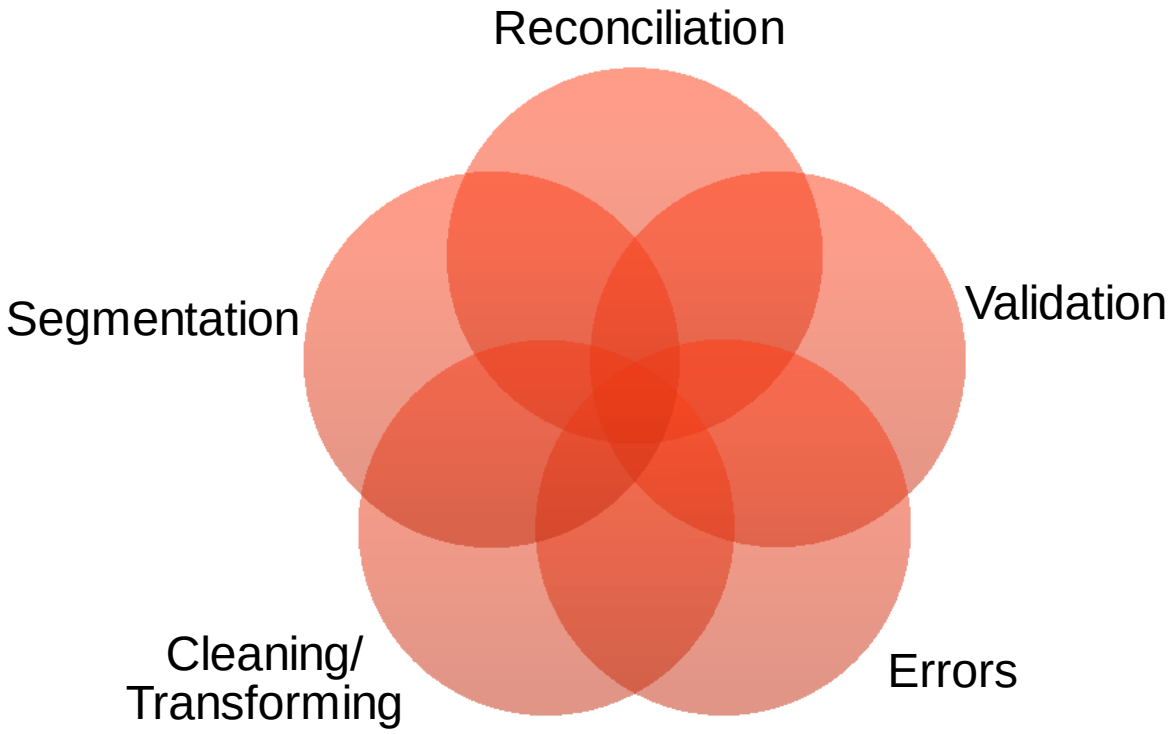
- Income
- Profession
- Family Physician details
- Policy Life details
- Detailed Medical History
  - Health status
  - Pre-existing disease
    - Diabetes
    - Hyper tension
    - Heart related, etc.
- Claim repudiation/ partial payment details



STEP 2 – EXAMINING

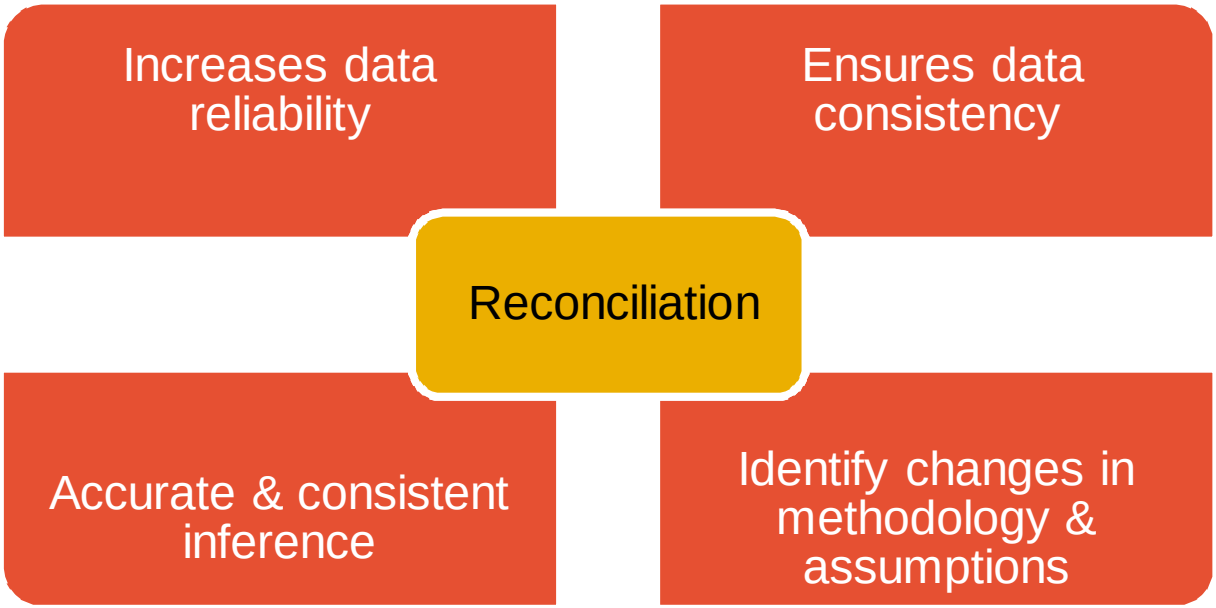
Analysis is only as good as the data on which it is based

Data Examination



Reconciliation

- Match information from multiple resources to address integrity or accuracy of data



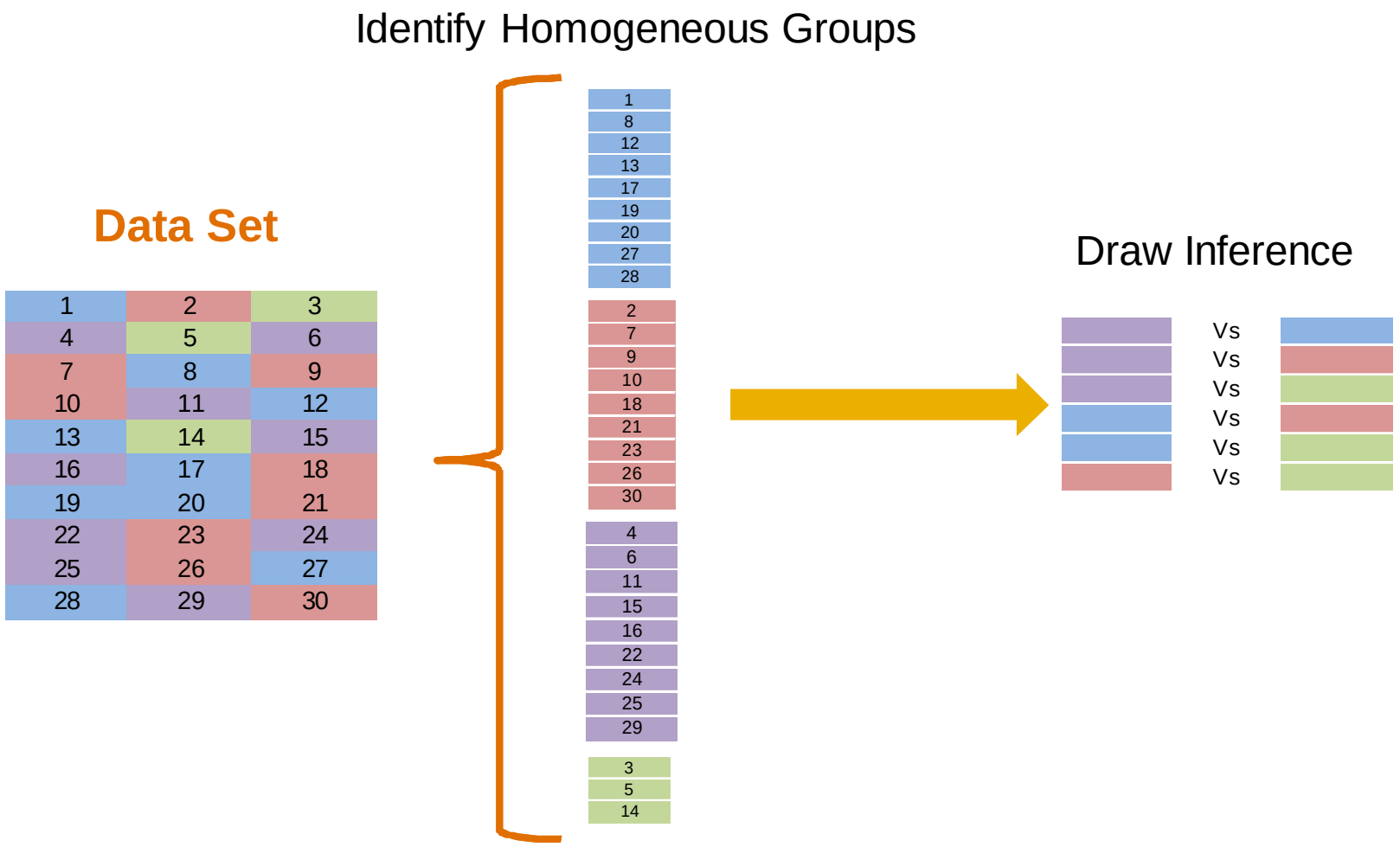
Validation

- Data validations plays a vital role in assessing the sensibility and credibility of the data. Some of the validations are:
  - Sensibility checks
    - Average Trends
    - Portfolio Mix
    - Claim Amounts Distribution
  - Other checks
    - Spot checks
    - Min and Max values
    - Number of records
    - Data formats
    - Other statistics

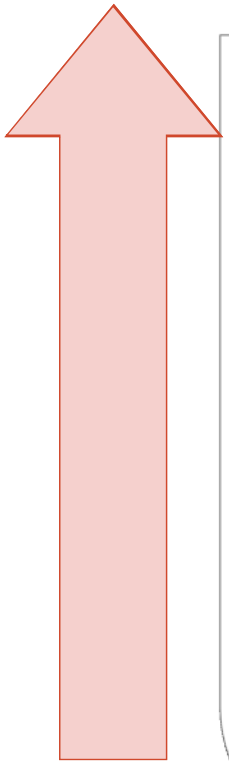
Errors & Distortions

- Most common data errors are
  - Wrong claim/ policy number
  - Wrong risk details
  - Wrong control dates
  - Inflated Claims
  - Missing/ incorrect claim type
- Various sources of data distortion include:
  - Changes in claim handling procedures
  - Case estimates
  - Processing delays
  - Large claims

Data Segmentation

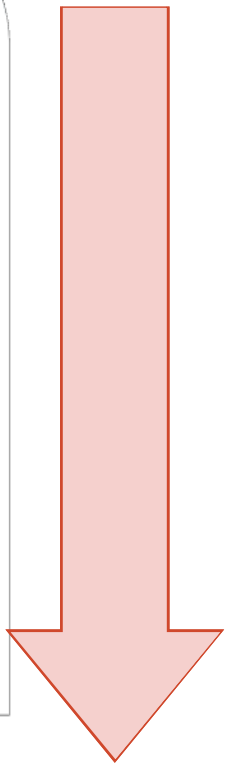


Segmenting the Data

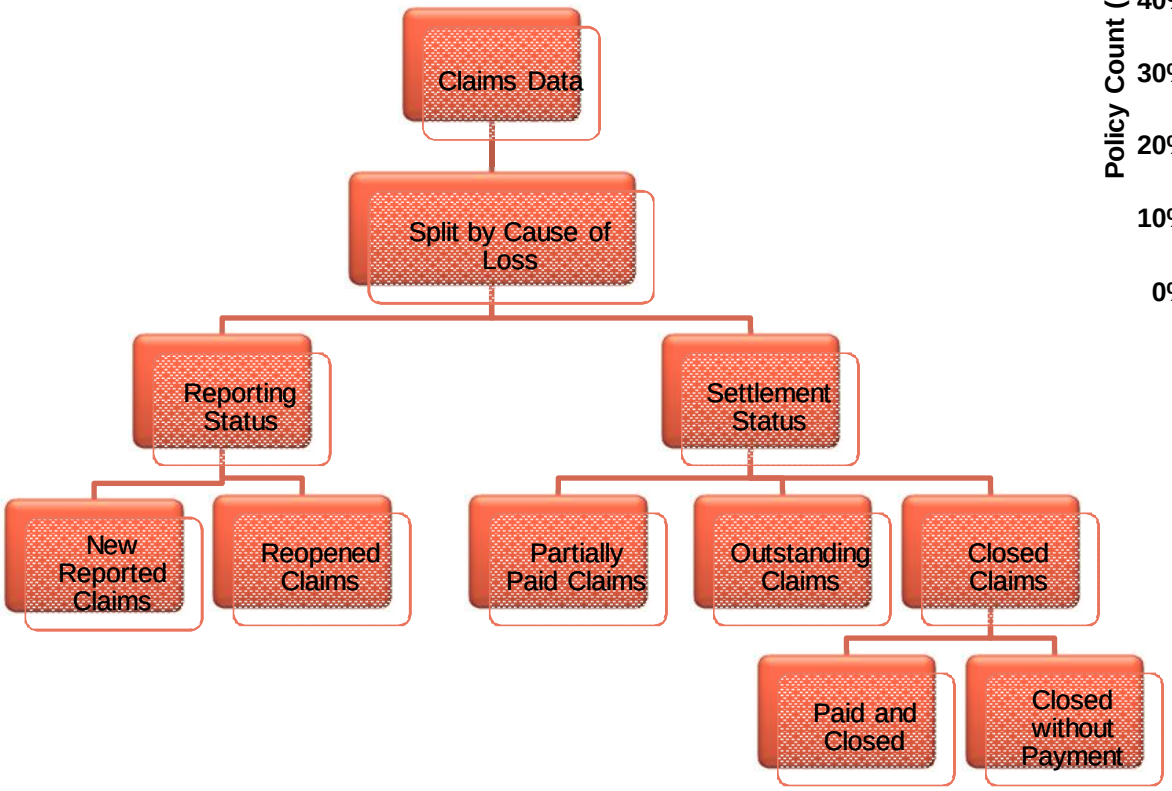


- Ease of interpretation
- Different event claim types have different risk drivers
- Can have very different trends
- Some segments can dominate the patterns
- Allows for changing mix of portfolio
- Ease of application of differing future inflation by claim type
- Allows explicit expense loading to be applied by claim type

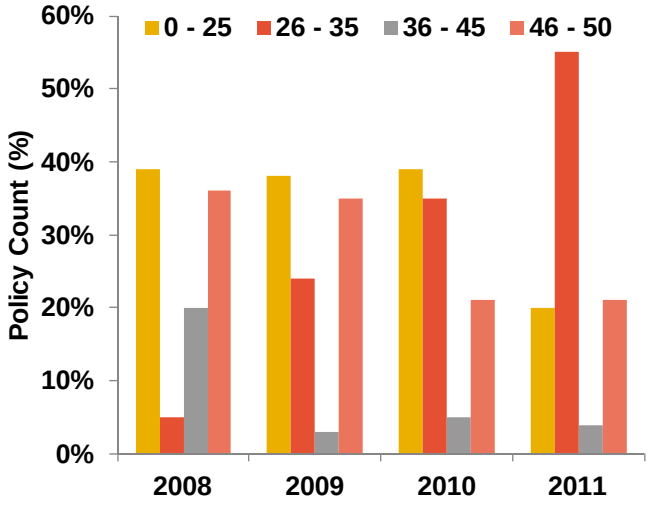
- More reports to produce
- Time/ Effort analyzing insignificant segments
- Reduces the amount of data available for each segment
- Possible system implementation issues
- Need to ensure some consistency in conclusion drawn between different reports



Data Segmentation – Examples



Claim Trends Analysis



Changing Business Mix

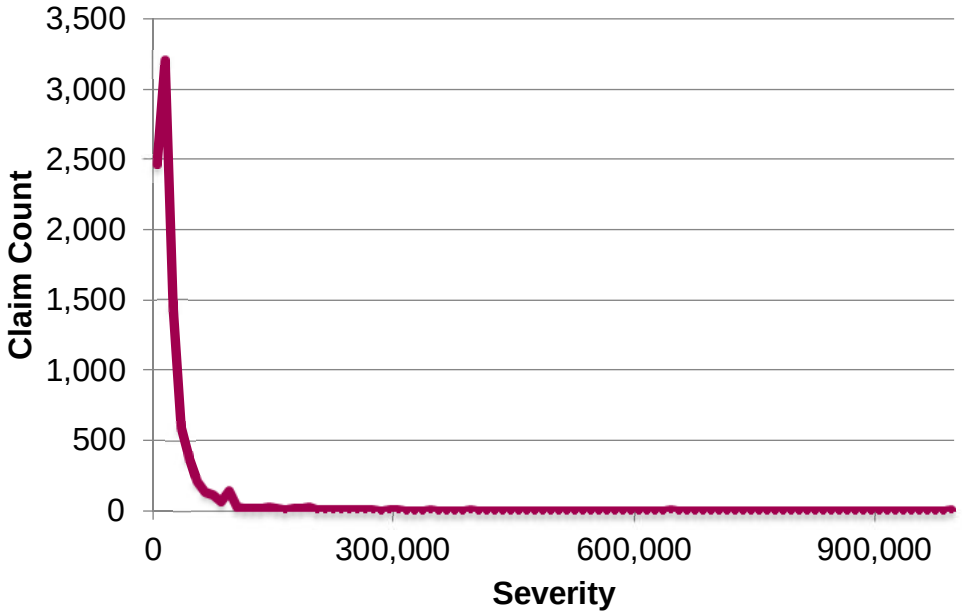
STEP 3 – ANALYZE & INFER

If you torture the data long enough, it will confess

– Ronald Coase, Economist

Data Snapshot – Tables & Graphs

Statistic	Value
Number of Observations	8,164
Mean	25,117
Standard Deviation	31,833
Coefficient of Variance	127%
Skewness	7
Kurtosis	132
Minimum	679
Median	16,869
Maximum	1,000,000

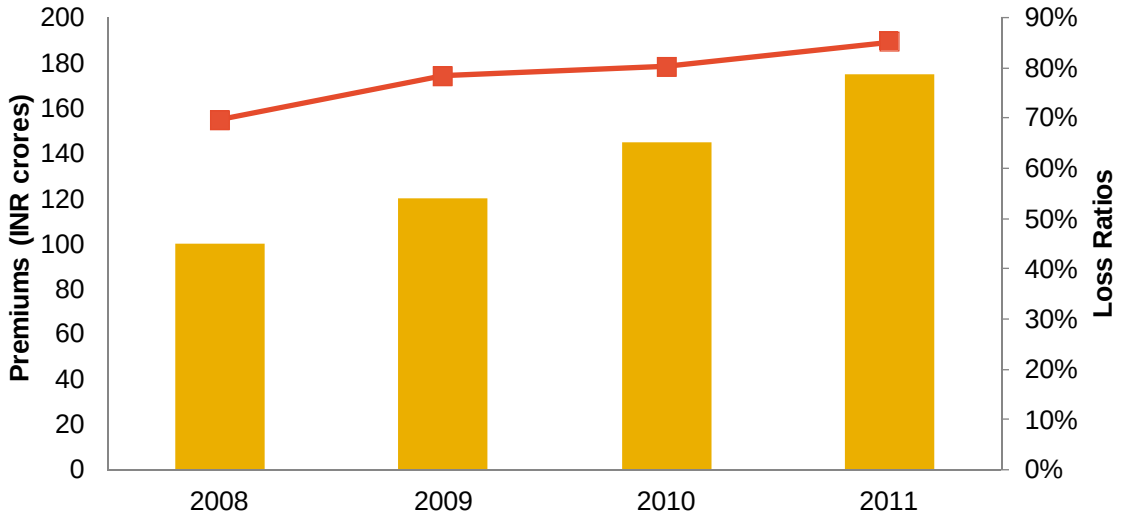


Analysis

- One-way analysis
- Two-way analysis
- Multi-way analysis

One Way Performance Report

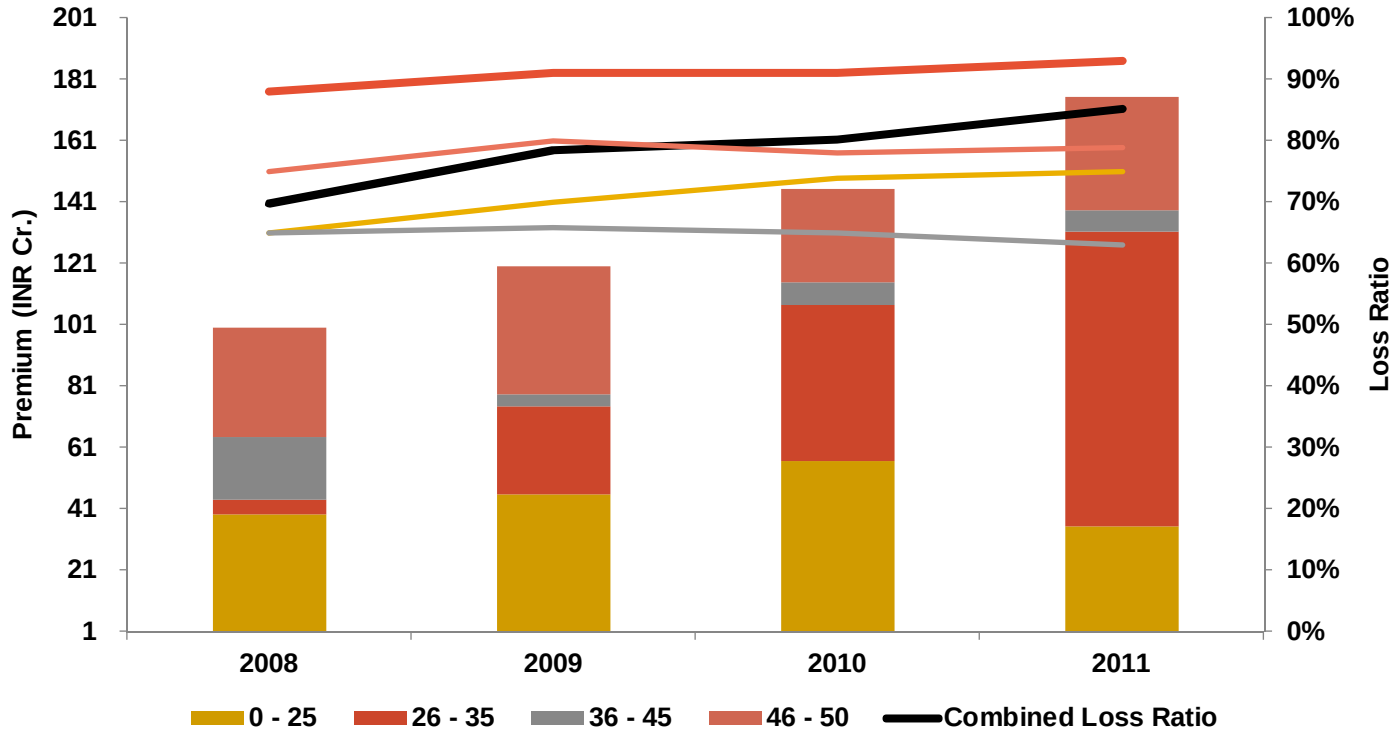
- Company X had targeted a loss ratio of 75% by 2013 in their 2010 annual business plan.



- Possible measures could include price revision or expense outgo controls.

Multi-Way Performance Report

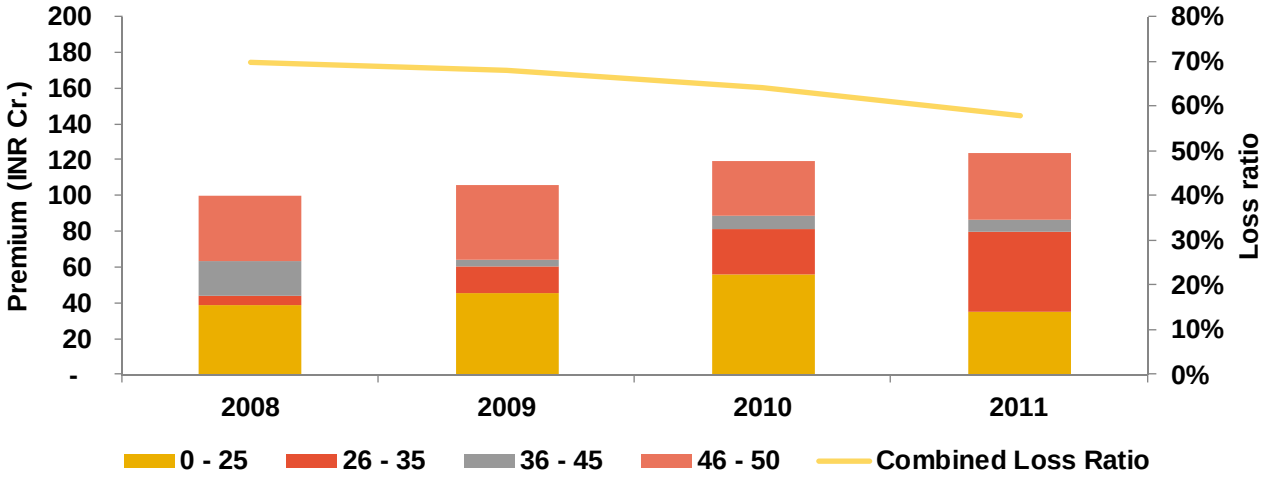
- A further detailed analysis suggests that we are increasing exposure towards 26-35 yrs Age Band which has a loss ratio of around 90%



Approach – 1

- Change in portfolio mix
  - Possible changes in marketing strategies
  - Re-pricing to attract better business

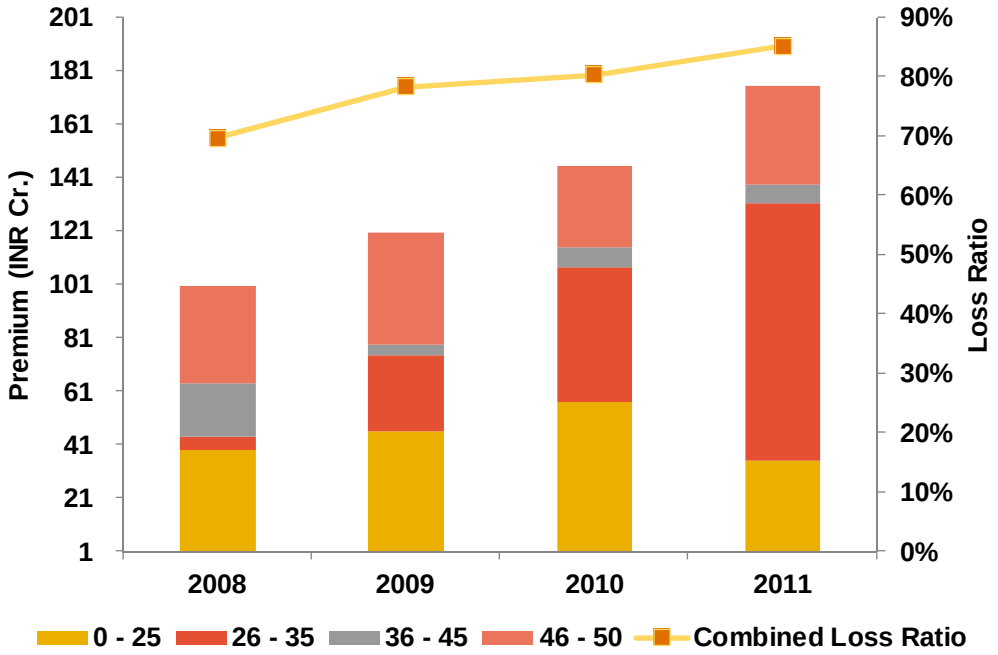
- Analyze historic data by reducing exposure for age band 26-35 by 50%, to assess the impact



Approach – 2

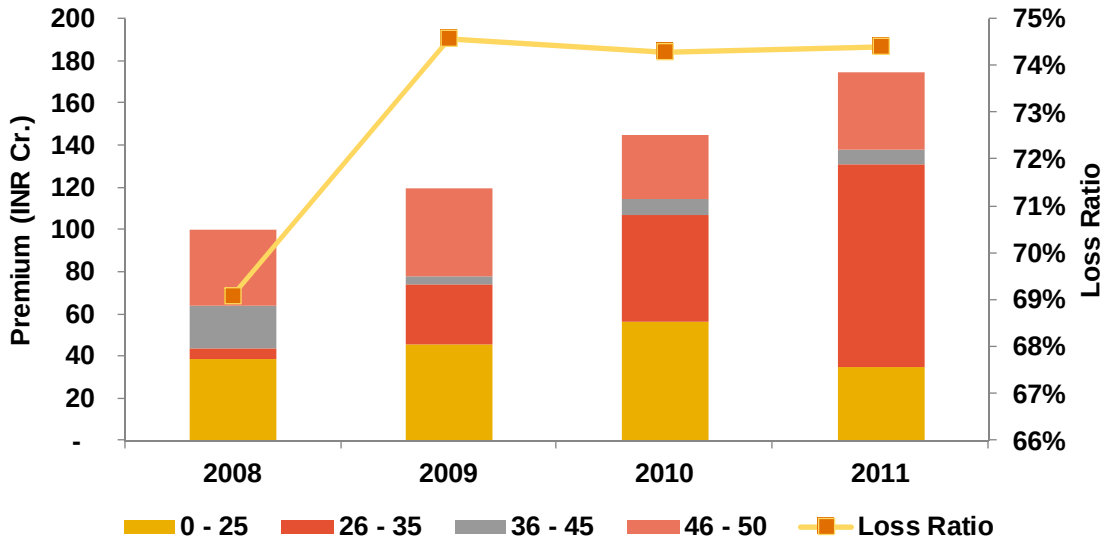
- Restructuring the benefits to control the loss ratio for 26-35 yrs age band.

Top 5 Diseases for 26 - 35 years	% Claim Count
Stress Related Diseases	35%
Accident	15%
Infections	12%
Clinical Findings	10%
Digestive	9%
Others	19%

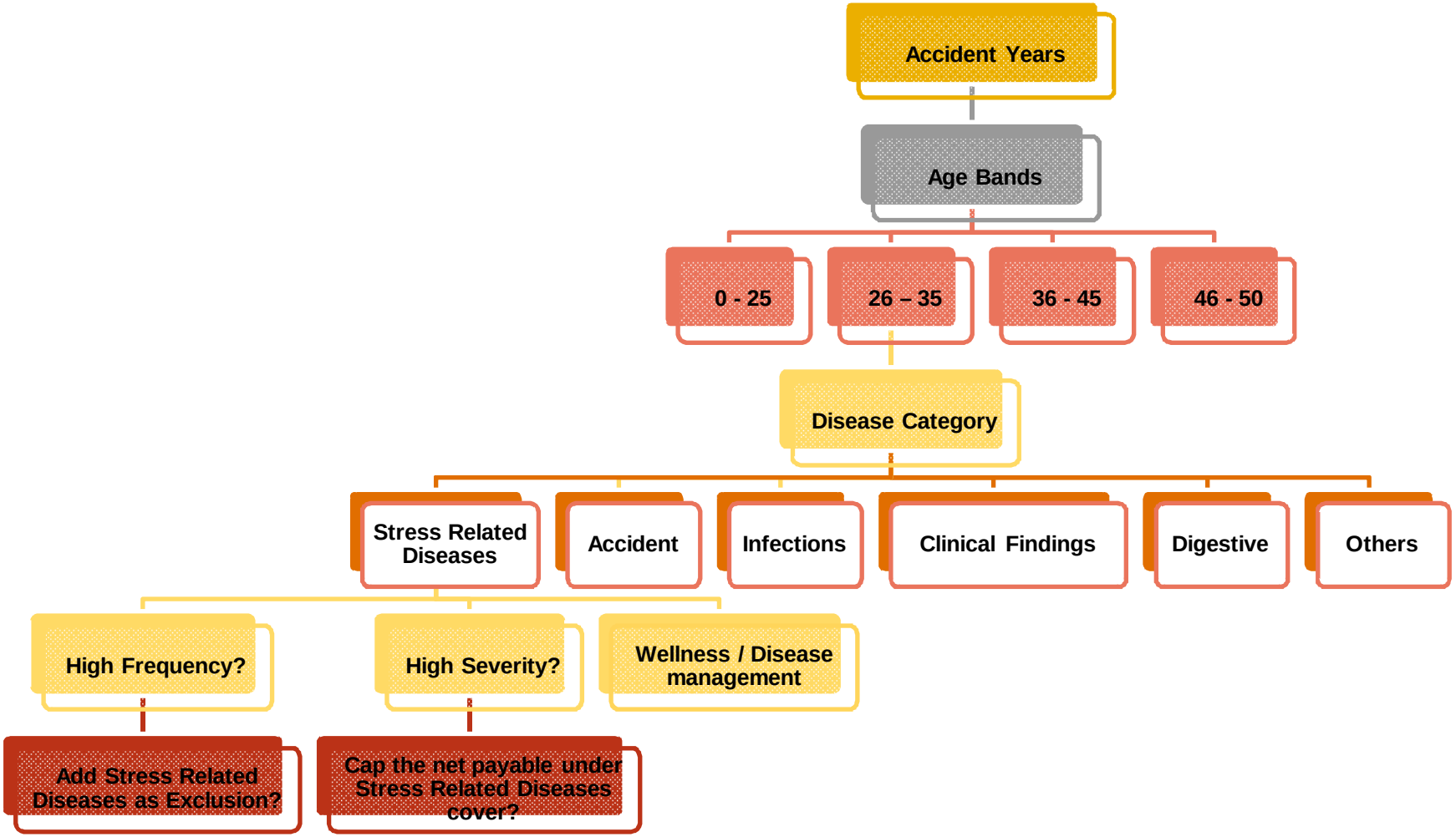


Approach – 2

- Inference – Loss ratios high because of “Stress Related Diseases” in the age group 26 – 35
- Solution – Study the impact on historic loss ratios by excluding claim instances for “Stress Related Diseases” for 26 – 35 year olds

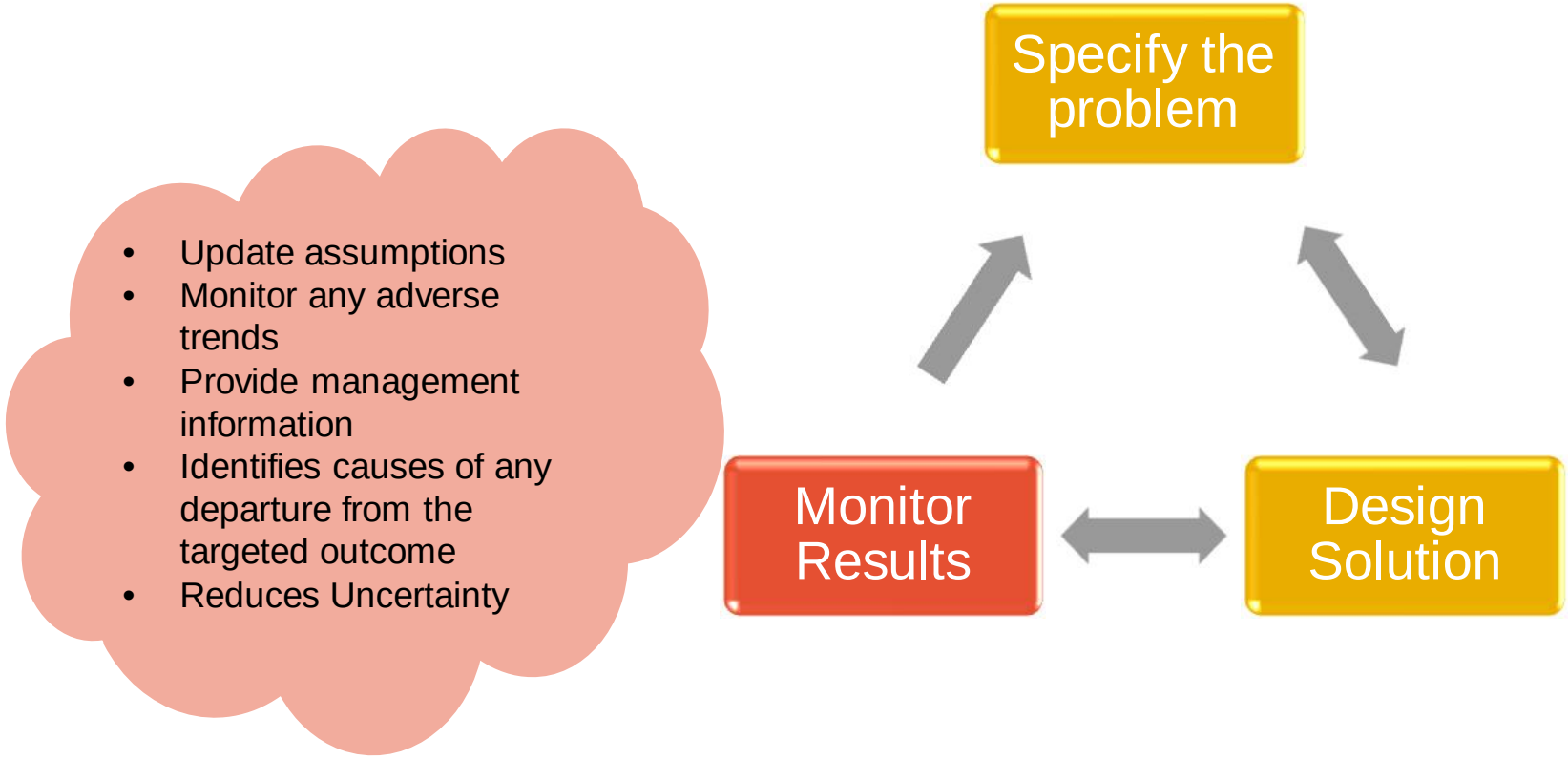


Drilling Down to the Root Cause

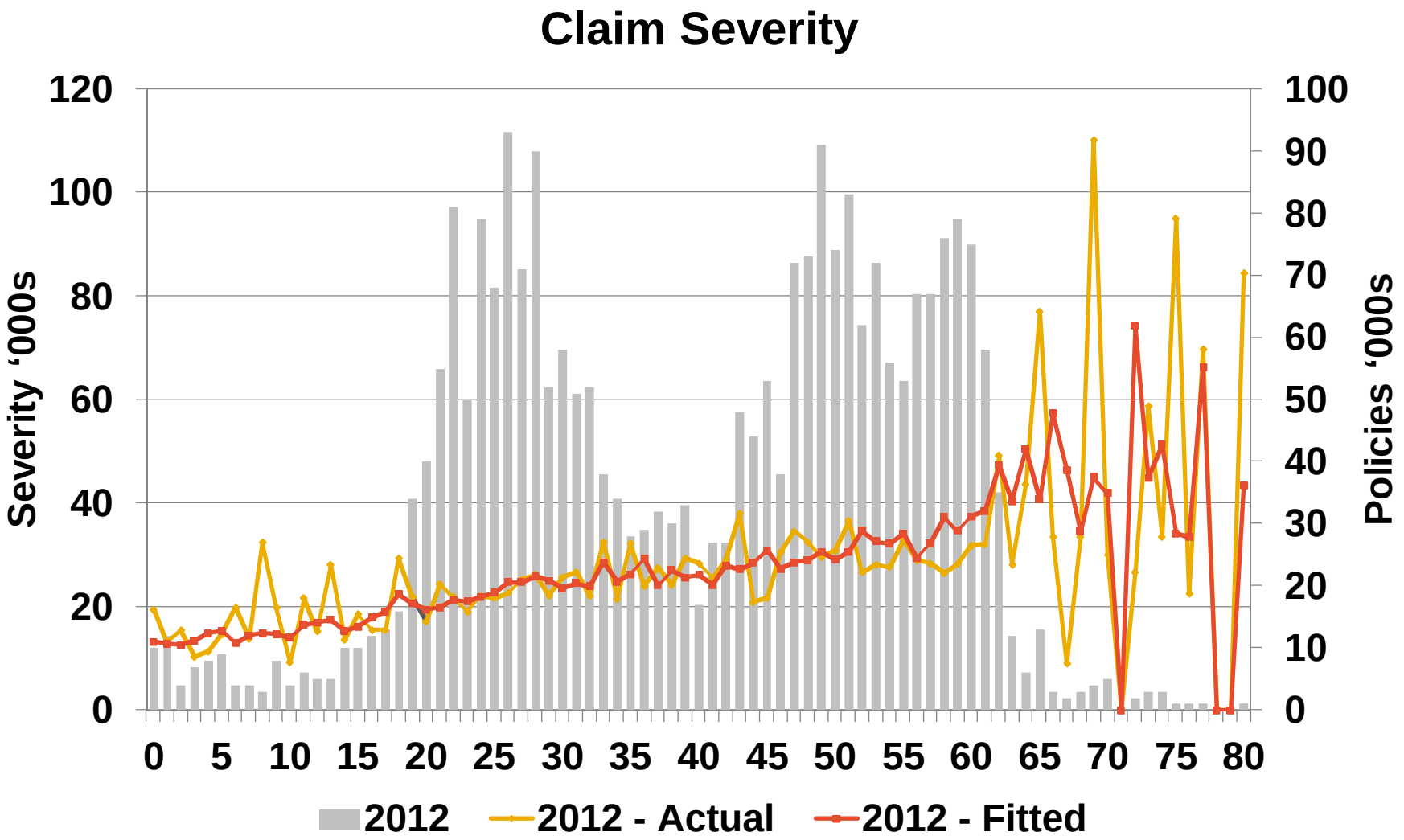


Monitoring

- Monitoring the experience is a fundamental part of the Actuarial Control Cycle.

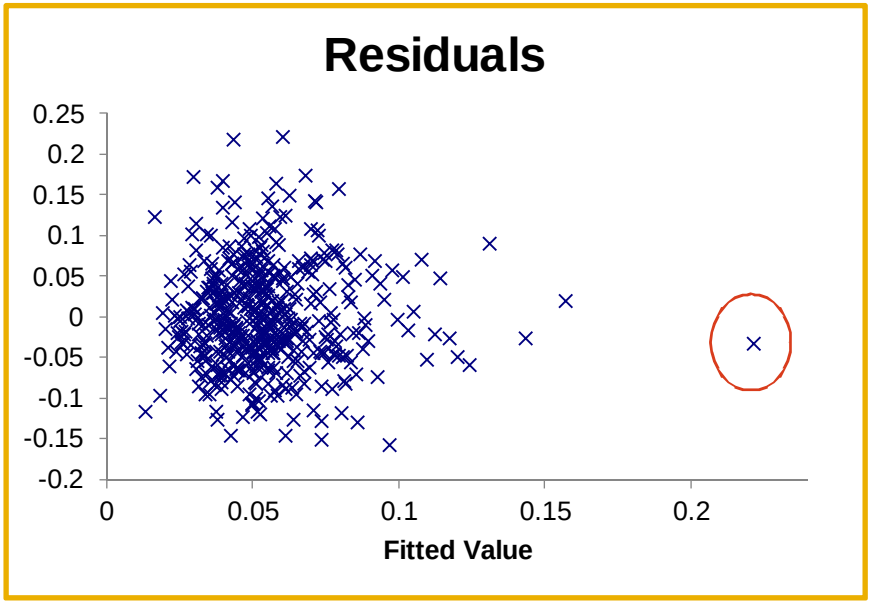


Monitoring – Example



Residual Analysis

- Measure of the difference between the observed value and the expected value
- A residual plot is a scatter plot of residuals against the fitted values
- Residuals that are unusually large (positive or negative) correspond to “outliers” i.e. data points that do not conform to the data set
- Outliers arise from several sources
  - Data errors
  - Biased Estimates
  - Incorrect Distribution
  - Inhomogeneous Groups



Various Analysis Parameters

Policy

- Portfolio Mix
- Average Premium
- Average Discount
- Average Sum Insured
- Expense analysis
- Level of persistency
- Cancellation/ Endorsement rates
- Commission rates
- Conversion rate

Claims

- Sum Insured utilization
- Frequency
- Severity
- Large claim propensity
- Loss Ratio
- Burning Cost
- Claim cost inflation
- Future projection
- Reinsurance optimization

Summary

- The goal is to turn data into information, and information into insight
- Can prove to be a strategic differentiator
- Quality data is a very important asset
- Anything that is measured and watched improves
- But keep in mind, Statistics are no substitute for judgment!

If the statistics are boring, then you've got the wrong numbers

— Edward R. Tufte

RETAIL HEALTH – A CASE STUDY

Hopeful Health Insurance Company

- Registration Date – 1st April, 2009
- Private establishment
- Market position – 6th largest health insurer (including GI Companies)
- Products offered (current)

Product Name	Sum Insured	Riders	Product Type
Hope I	50,000 – 10,00,000	Day Care Hospital Cash Ambulance Cover	Individual Health Insurance
Hope F	5,00,000 10,00,000 15,00,000	Day Care Ambulance Cover	Floater Health Insurance
Hope C	10,00,000 15,00,000		Critical Illness
Hope D	2,00,000 XS 3,00,000 3,00,000 XS 7,00,000		High Deductible

Exercise 1

- Can you identify a few reconciliations from different sources for a Health Insurance Company

Solution

Institute of Actuaries of India

Raw Data vs. Regulatory Filings

- Premiums
- Paid Amounts
- Outstanding Movements

Data from different MIS

- Average Cost
- Frequency
- Loss Ratio

IT Data vs. Finance Data

- Policy Count
- Premiums
- Commissions

Current vs. Previous reports

- Movement in portfolio mix
- Consistency
- Closing vs. Opening positions

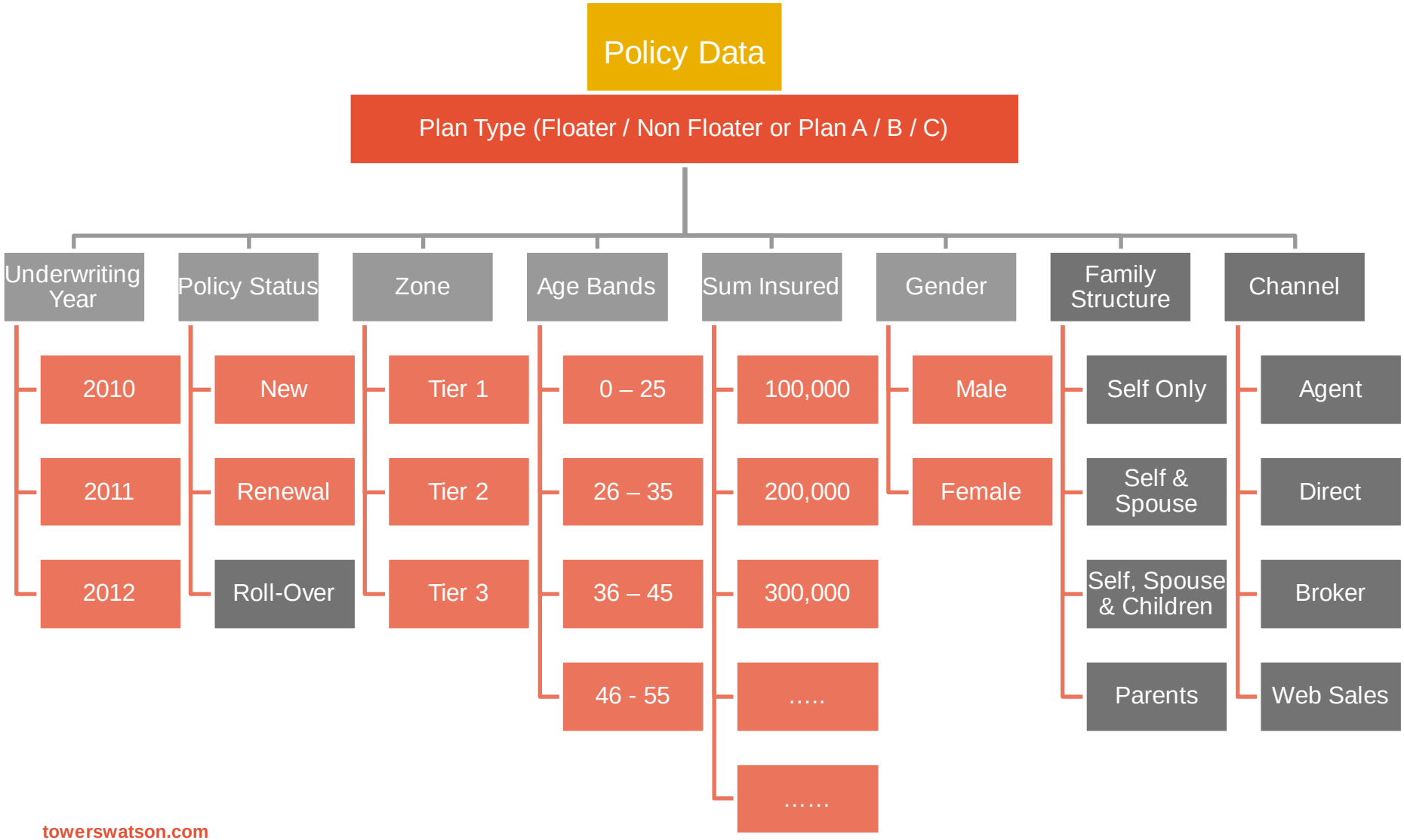
Data In Vs. Data Out

- Record Count
- Totals (Premium / claims)
- Variables

Exercise 2

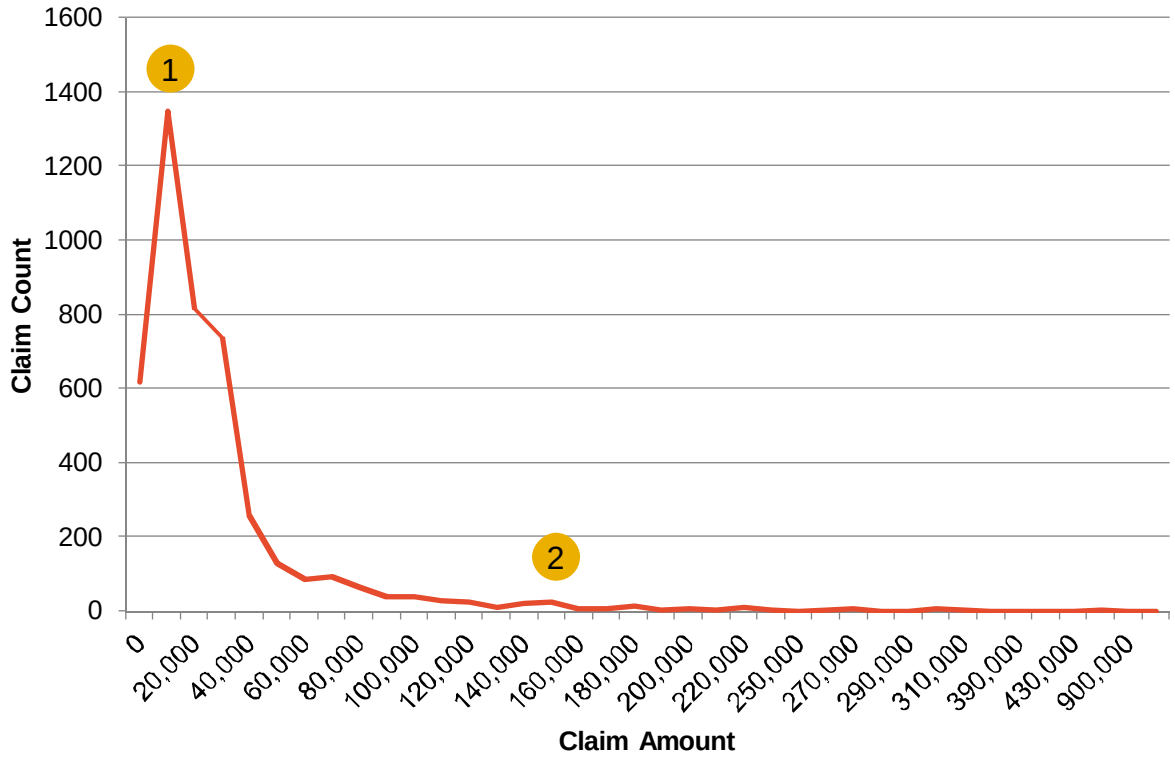
- Segment the available Policy Data

Data Segmentation – Policy Data



Exercise 3

- Construct a line chart for the claim amounts distribution & comment

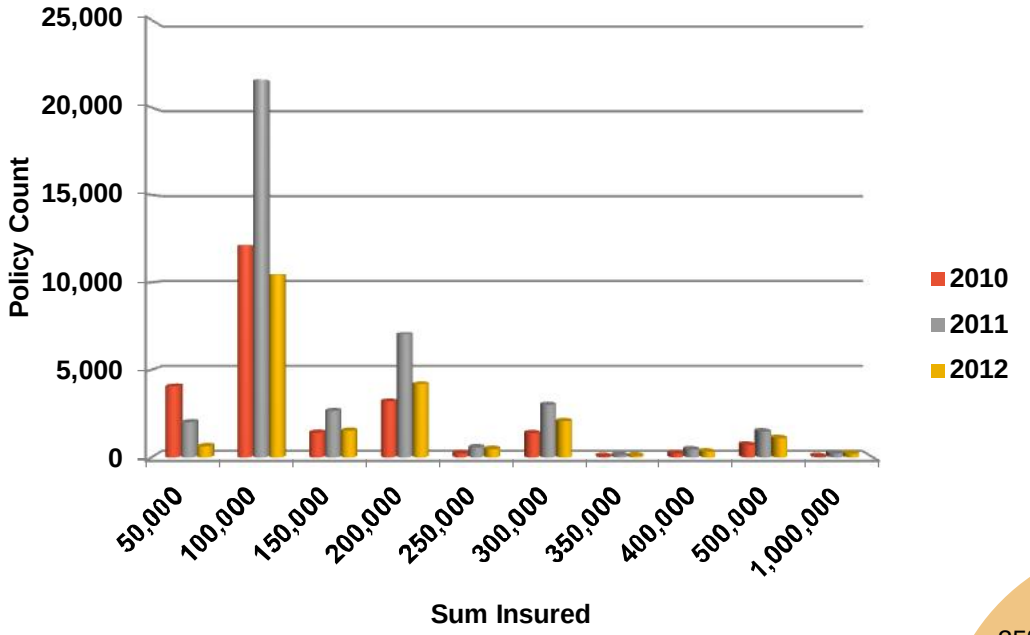


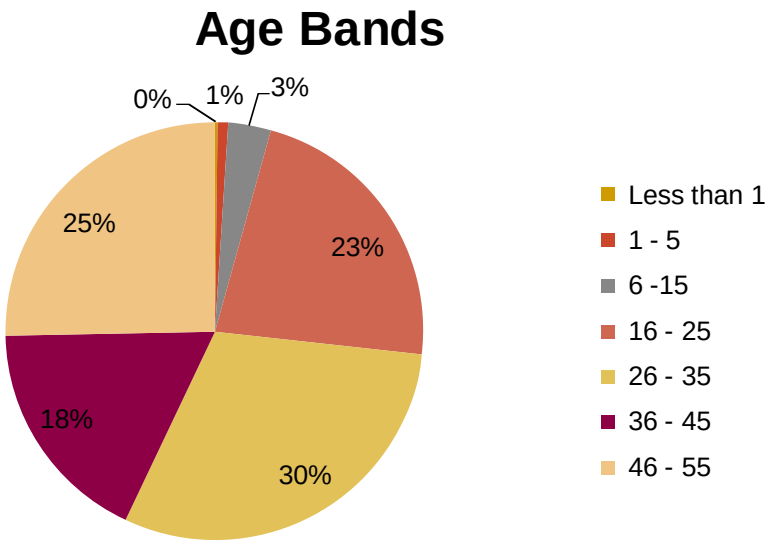
1. Maximum Claims for INR 30,000
2. Sudden Concentration of Claims at INR 1,50,000
3. Positively skewed
4. 99% claims below INR 2,25,000
5. Percentile Plot can help identify outliers as well

Exercise 4

- Comment on Hopeful company's portfolio mix and compare with the given strategy using one-way analysis

Solution

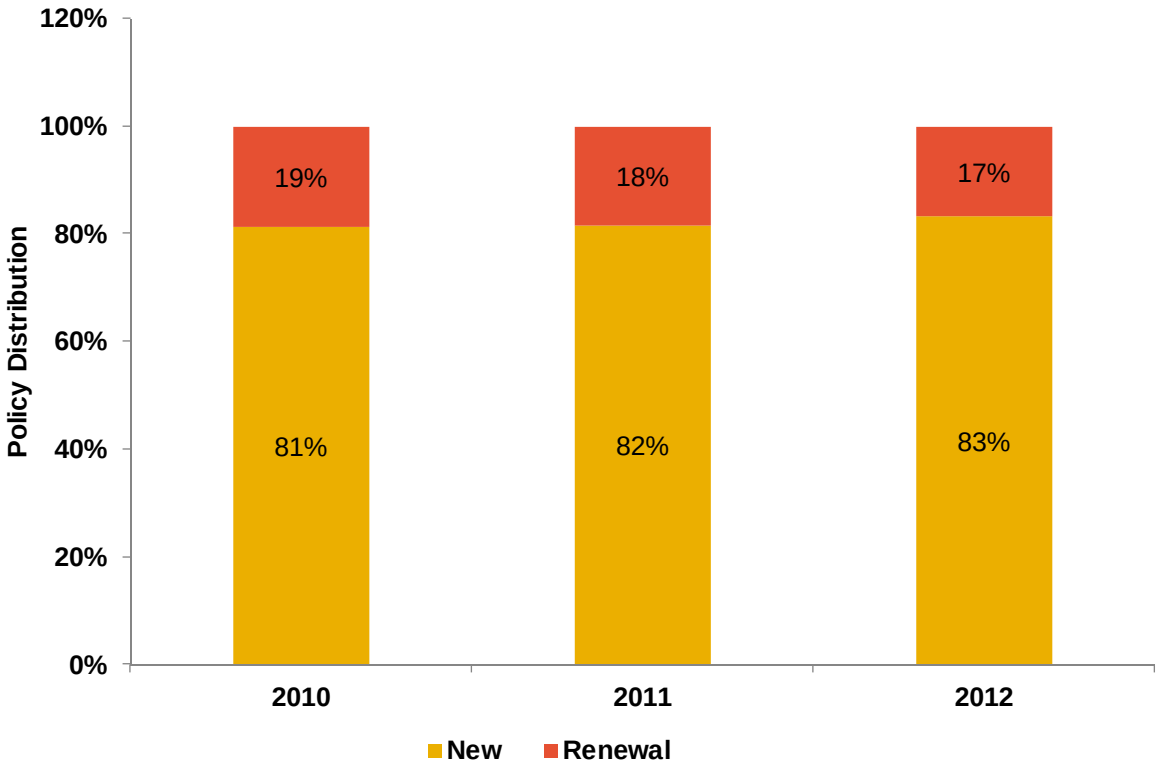




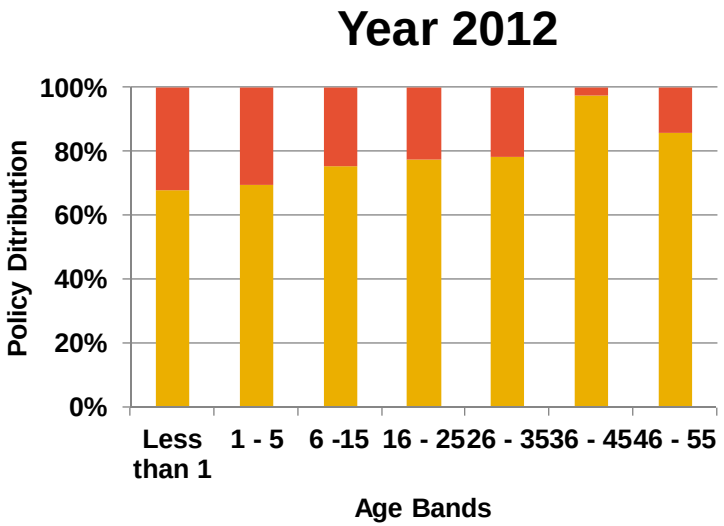
Exercise 5

- It seems that Hopeful Insurance Company's services are good as the new-renewal business proportion is as per industry benchmark (80% – 20%). Please comment if this is the case for all segments.

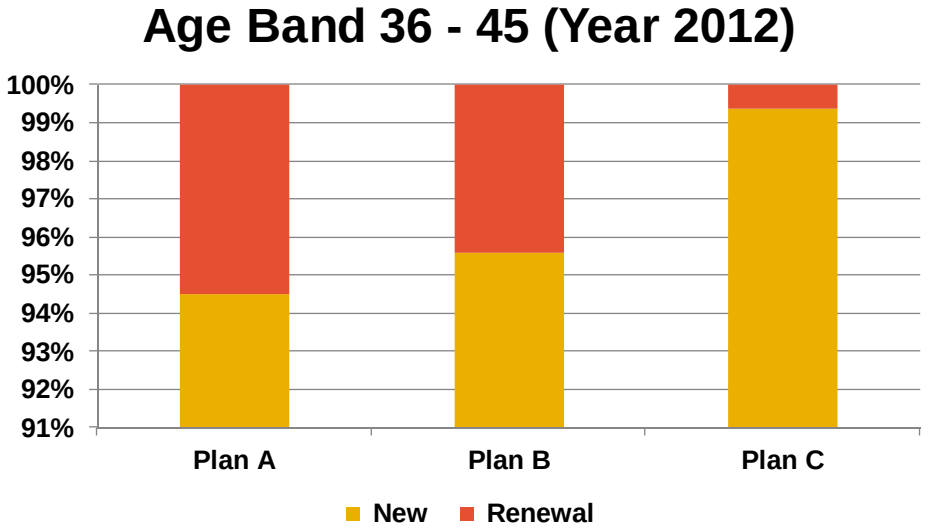
Solution – One Way Analysis



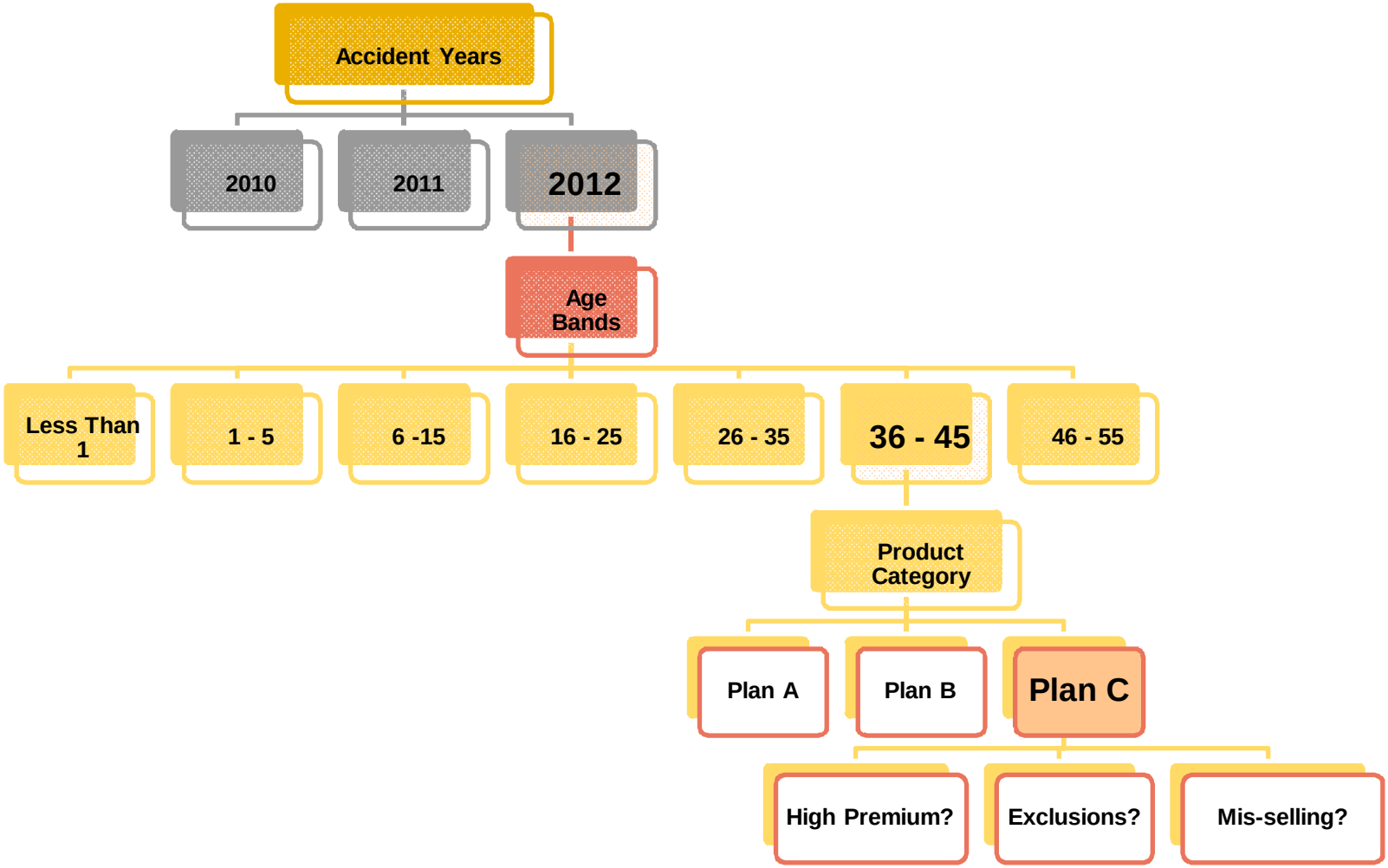
Solution – Multi Way Analysis



Age Band 36 – 45 years has minimum Renewals



Solution – Drilling Down to the Root Cause



Exercise 6

- Create a multi-way table to calculate the average claim cost across various Age groups, Gender and Product Plans
- Use this to project the average claim cost for 56 - 60 years old choosing Plan C

Solution – Multi Way Analysis – Plan / Age / Gender

Claim Severity		Gender		
Age	Plan	F	M	Grand Total
Less than 1	Plan A	13,469	19,082	16,977
	Plan B		13,029	13,029
	Plan C	33,750	13,521	23,636
Less than 1 Total		**15,497**	**17,593**	**16,895**
1 - 5	Plan A	14,942	21,610	19,081
	Plan B	13,356	16,472	14,914
	Plan C	26,393	16,686	20,263
1 - 5 Total		**18,297**	**19,268**	**18,887**
6 -15	Plan A	25,819	19,111	21,708
	Plan B	15,768	24,586	20,516
	Plan C	30,885	48,688	42,330
6 -15 Total		**25,633**	**26,071**	**25,901**
16 - 25	Plan A	25,519	27,815	27,074
	Plan B	28,658	32,174	31,150
	Plan C	35,709	25,018	28,128
16 - 25 Total		**28,764**	**30,711**	**30,134**
26 - 35	Plan A	38,070	37,200	37,440
	Plan B	33,846	30,784	31,638
	Plan C	50,700	34,740	38,996
26 - 35 Total		**36,121**	**32,068**	**33,192**
36 - 45	Plan A	31,794	29,294	30,528
	Plan B	34,858	51,093	42,892
	Plan C	41,252	34,261	37,892
36 - 45 Total		**38,688**	**38,057**	**38,381**
46 - 55	Plan A	29,551	39,732	34,517
	Plan B	33,653	45,860	38,642
	Plan C	37,680	40,952	39,073
46 - 55 Total		**35,855**	**42,049**	**38,502**
Grand Total		**34,637**	**34,058**	**34,295**

- Too many cells & hence a clumsy report
- Segments with missing or insufficient data
- Data can't be projected for new Age Bands
- Analysis dependent on Portfolio Mix
- Probable use of multi-regression methods

Exercise 7

- Discuss possible measures to achieve the following targets
  - To include higher age groups as well in their portfolio
  - Diversify its risk across all age groups
  - Maximum loss ratio of 80% across all products

Hopeful Health Insurance Company – Future Plans

- To include higher age groups as well in their portfolio
  - Better benefit structure
  - Underwriting guidelines (identify good and bad risk)
  - Data for higher age groups is unavailable and hence it may need to refer to some external source of data For eg. industry benchmarks or regulator's published reports
  - Project its experience using the information from lower age groups

- Diversify its risk across all age groups
  - Identify needs of each age group & develop health insurance solutions accordingly
  - Identify appropriate channels to target right customer with right products
  - Re-pricing/ restructuring existing products
  - By launching marketing campaigns/ differential incentives/ awareness programs/ target sales

Hopeful Health Insurance Company – Future Plans

- Maximum loss ratio of 80% across all products
  - Optimize the pricing and benefit structures
  - Reducing cross subsidies
  - Stringent underwriting guidelines
  - Better claim management systems

GROUP HEALTH STUDY – A CASE STUDY

Background

- Broker of Employer X has reached out to Insurer Y for renewing its annual group health insurance plan
- The client is not an existing client and is a big name in IT Industry
- The data is made in the form of summary tables which would be supplemented by additional data information later on request after Insurer confirms interest to accept the client proposal
- Policy pertaining to year 2012 has developed for only 6 months

Main Parties Involved

- Insurer
- Employer
- Broker
- Employee (The Ultimate Consumer)

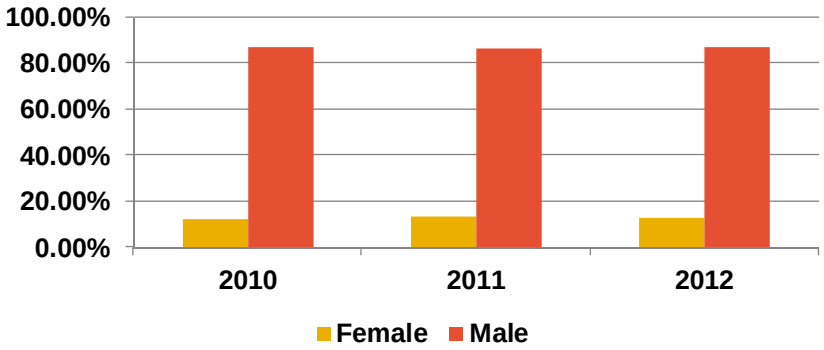
INSURER'S PERSPECTIVE

Portfolio Mix – By Premium

Table 1

Age	2010	2011	2012
18 – 25	37%	43%	41%
26 – 35	52%	48%	49%
36 – 45	10%	8%	9%
46 – 55	1%	1%	1%
56 – 65	0%	0%	0%
66 & Above	0%	0%	0%

Table 2



- Consistent portfolio mix over last 3 years, across many parameters
- Comprises of a high percentage of young population
- High concentration of risks at lower Sum Insured levels

Table 3

Sum Insured	2010	2011	2012
20,000	68.73%	66.81%	63.85%
50,000	17.23%	18.17%	19.40%
60,000	0.70%	0.72%	0.78%
1,00,000	4.91%	4.10%	4.35%
1,50,000	4.05%	6.43%	7.88%
2,00,000	2.39%	2.71%	2.66%
2,50,000	0.10%	0.07%	0.10%
3,00,000	0.79%	0.90%	0.96%
5,00,000	0.00%	0.00%	0.02%
15,00,000	0.00%	0.09%	0.00%

Business Overview

Table 4

Year / Coverage Type	Written Insured Lives	Premium Paid	Claim Count	Amount Claimed	Amount Paid	Frequency	Severity	Loss Ratio
2010	**23,749**	**18,000,000**	**760**	**20,000,000**	**14,564,677**	**3.20%**	**26,316**	**111%**
Floater	1,456	11,000,000	400	12,000,000	9,314,057	27.47%	30,000	109%
Non Floater	22,293	7,000,000	360	8,000,000	5,250,620	1.61%	22,222	114%
2011	**19,228**	**22,000,000**	**1,090**	**25,000,000**	**18,788,829**	**5.67%**	**22,936**	**114%**
Floater	2,736	11,500,000	490	12,000,000	9,983,801	17.91%	24,490	104%
Non Floater	16,493	10,500,000	600	13,000,000	8,805,029	3.64%	21,667	124%
2012	**26,439**	**27,000,000**	**690**	**16,000,000**	**12,612,298**	**2.61%**	**23,188**	**59%**
Floater	4,439	17,000,000	340	9,000,000	7,881,612	7.66%	26,471	53%
Non Floater	22,000	11,000,000	350	7,000,000	4,730,686	1.59%	20,000	64%

- Non-Floater cover has worse claims experience than floater
- Insured given floater cover forms 12% of the total insured base
- Better experience in 2011 along with significant increase in non-floater category could have resulted in lower premium per head for 2012
- The claims experience for 2012 is higher than expected which could result in steep premium hike next year

Experience Analysis – Frequency / Severity

Institute of Actuaries of India

Table 6	Claims Severity Analysis - 2012							
	Sum Insured							
Age Group	20,000	50,000	60,000	1,00,000	1,50,000	2,00,000	2,50,000	3,00,000
0-18	6,245	13,599	42,868	17,186	21,626	16,061	11,812	22,182
19-25	12,201	18,672	17,007	18,694	24,239	20,250	34,374	11,392
26-35	13,866	19,986	21,476	25,388	25,385	23,999	117,353	40,417
36-45	13,595	22,168		24,068	22,213	20,926		30,210
46-55	13,484	41,854		21,228	28,481	26,213		42,449
56-65	14,690	17,969		28,374				36,813
Above 65				28,188				94,048

Table 7	Frequency Analysis - 2012							
	Sum Insured							
Age Group	20,000	50,000	60,000	1,00,000	1,50,000	2,00,000	2,50,000	3,00,000
0-18	30%	67%	0%	0%	0%	0%	0%	0%
19-25	2%	6%	12%	16%	8%	20%	14%	6%
26-35	2%	4%	4%	12%	6%	12%	12%	18%
36-45	2%	4%	2%	6%	4%	4%	0%	16%
46-55	2%	4%	0%	2%	10%	4%	0%	26%
56-65	4%	4%	0%	94%	0%	0%	0%	266%
Above 65	0%	0%	0%	100%	0%	0%	0%	0%

Observations:

- Severity increasing with age
- Offsetting between age groups and sum insured levels
- Frequency is higher for lower age groups
- High frequency & severity is noticed for 3L sum insured

Consolidated Analysis

- The overall loss ratio for floater polices have been higher than non-floater
  - This could be because of significant difference between the premium charged for floater policies as compared to non-floater
- The estimated burning cost on the basis of policy year 2012 is 605
- Average severity has been observed to be higher than the lowest Sum Insured band, which comprises of more than 65% of population
- Probable need of restructuring the benefits provided
- A further investigation into data could provide useful insights including
  - Claim trend experience
  - Utilization trends
  - Need to rectify the current benefit structure

Other Information provided

Table 8

Payment Patterns

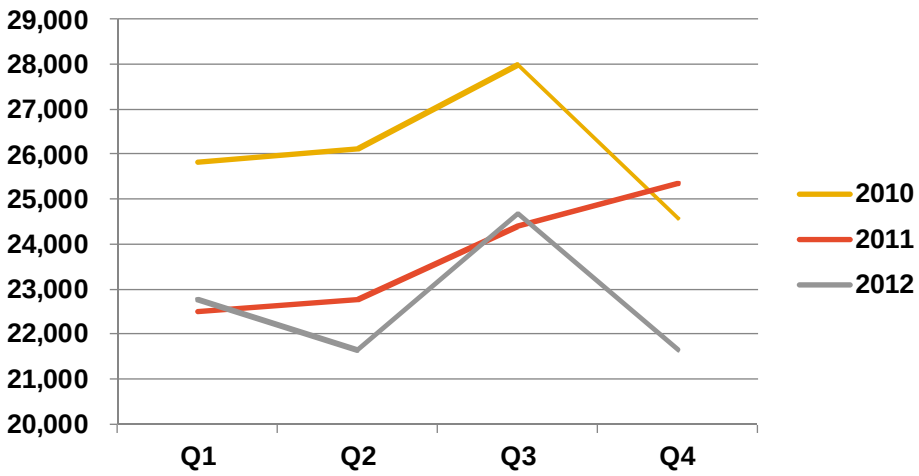
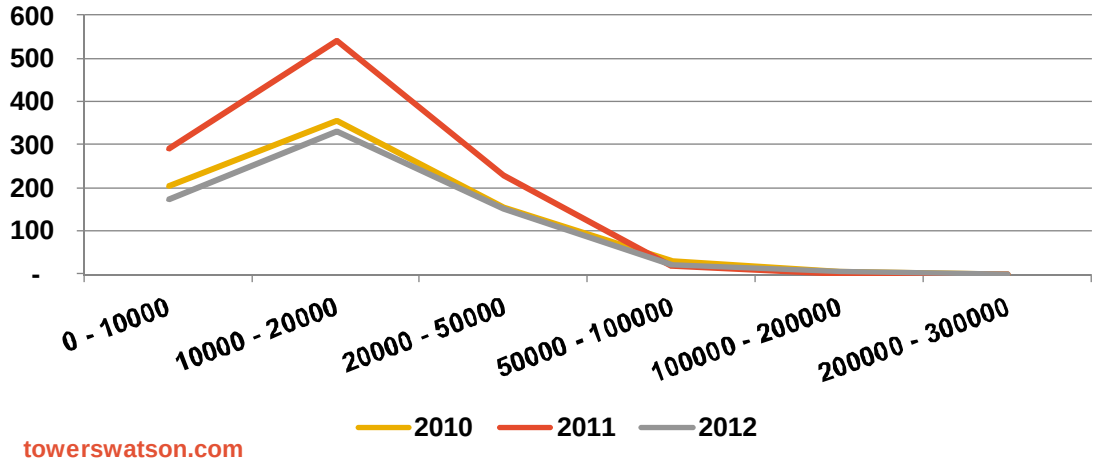


Table 9

Claim Amount Distribution

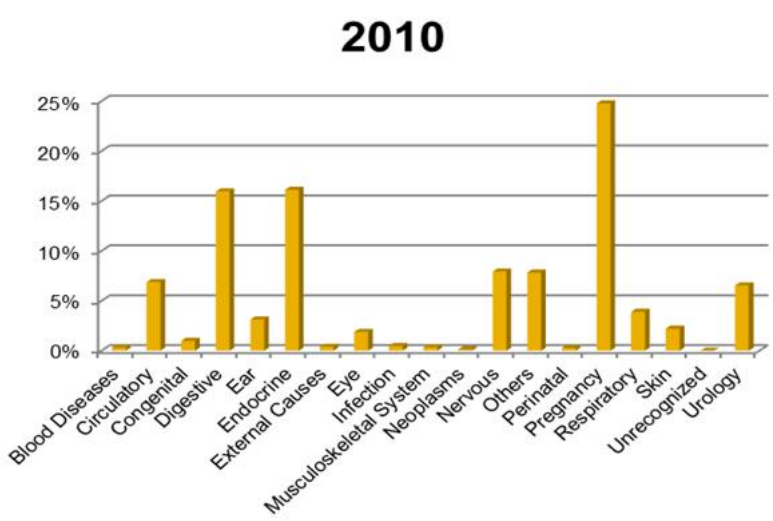


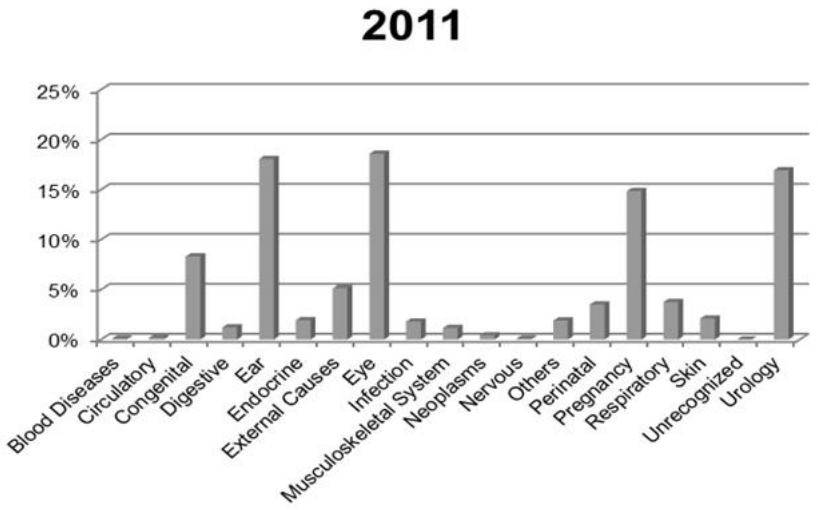
- Help identify trends/ seasonal effects in severity
- Similar analysis can help estimate IBNR or ultimate cost of an expiring policy
- Higher severity observed for Q3 payments as compared to Q2

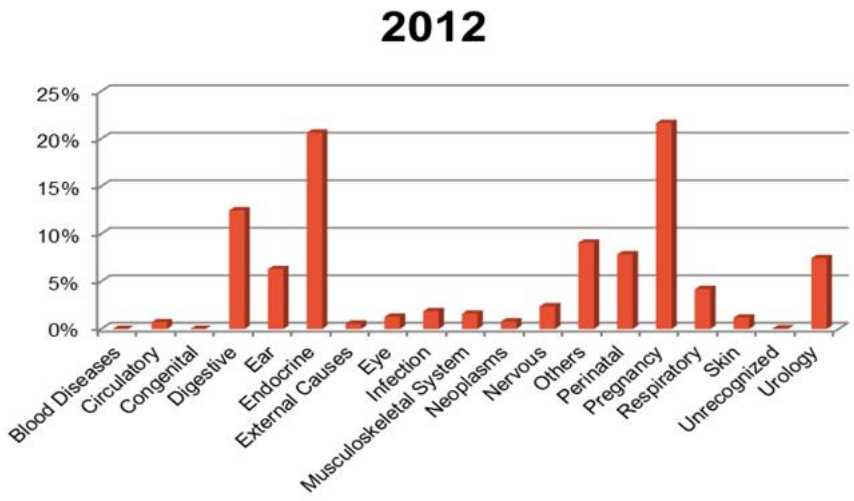
Data Preview- Frequency Trends by Disease

Institute of Actuaries of India

Table 10

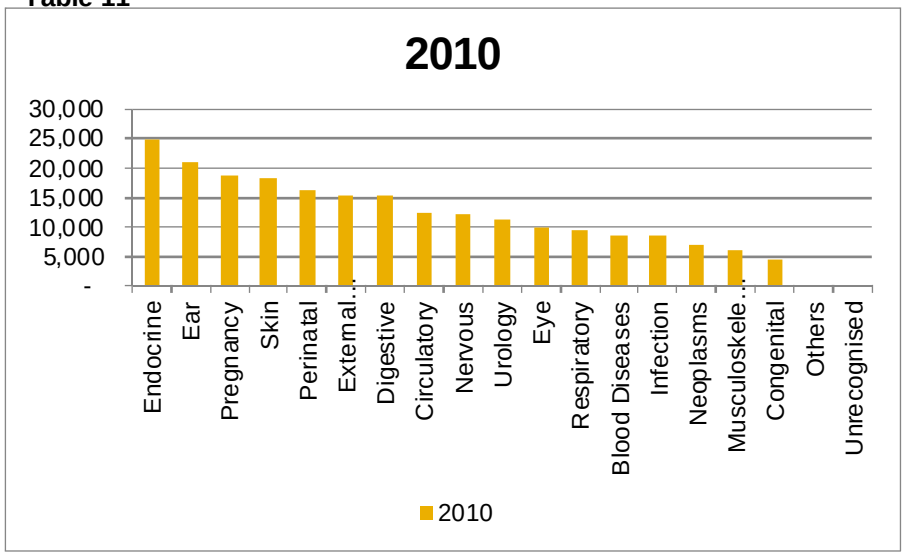


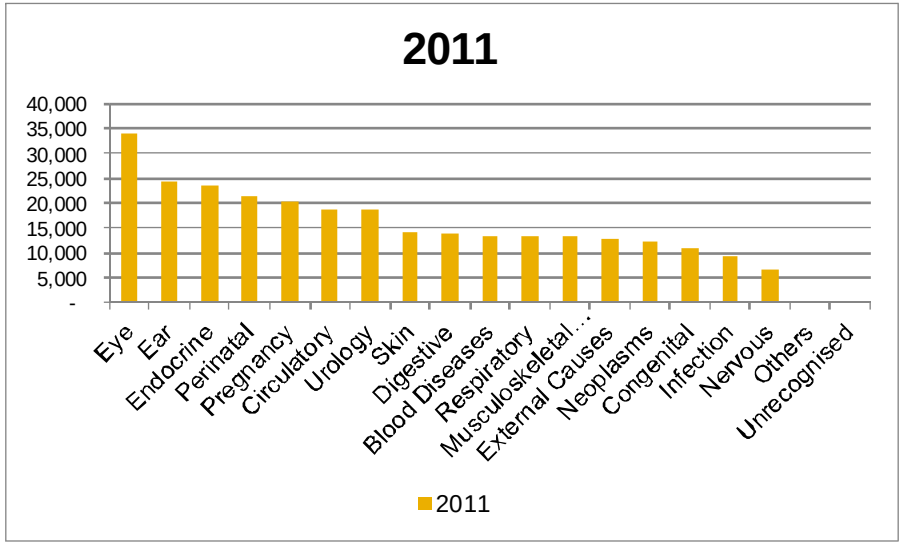


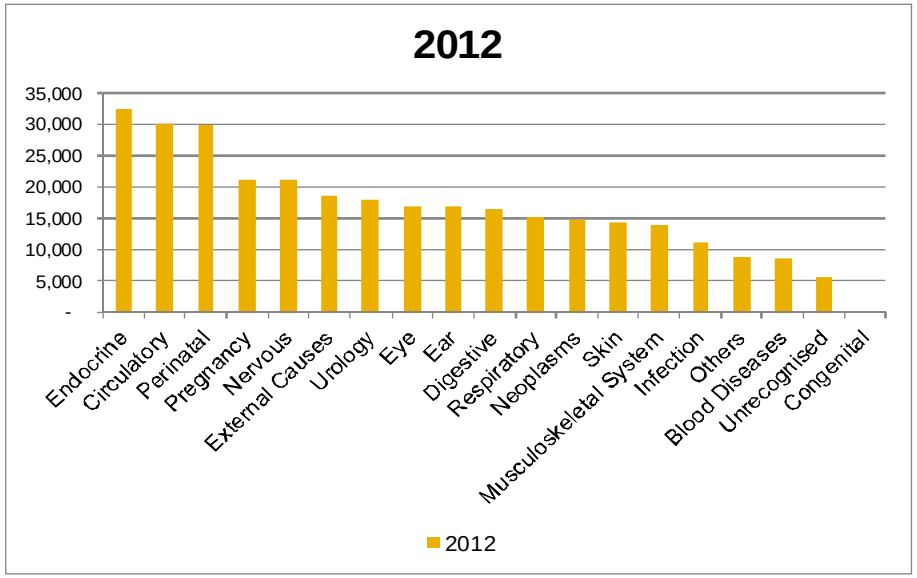


Data Preview- Severity Trends by Disease

Table 11





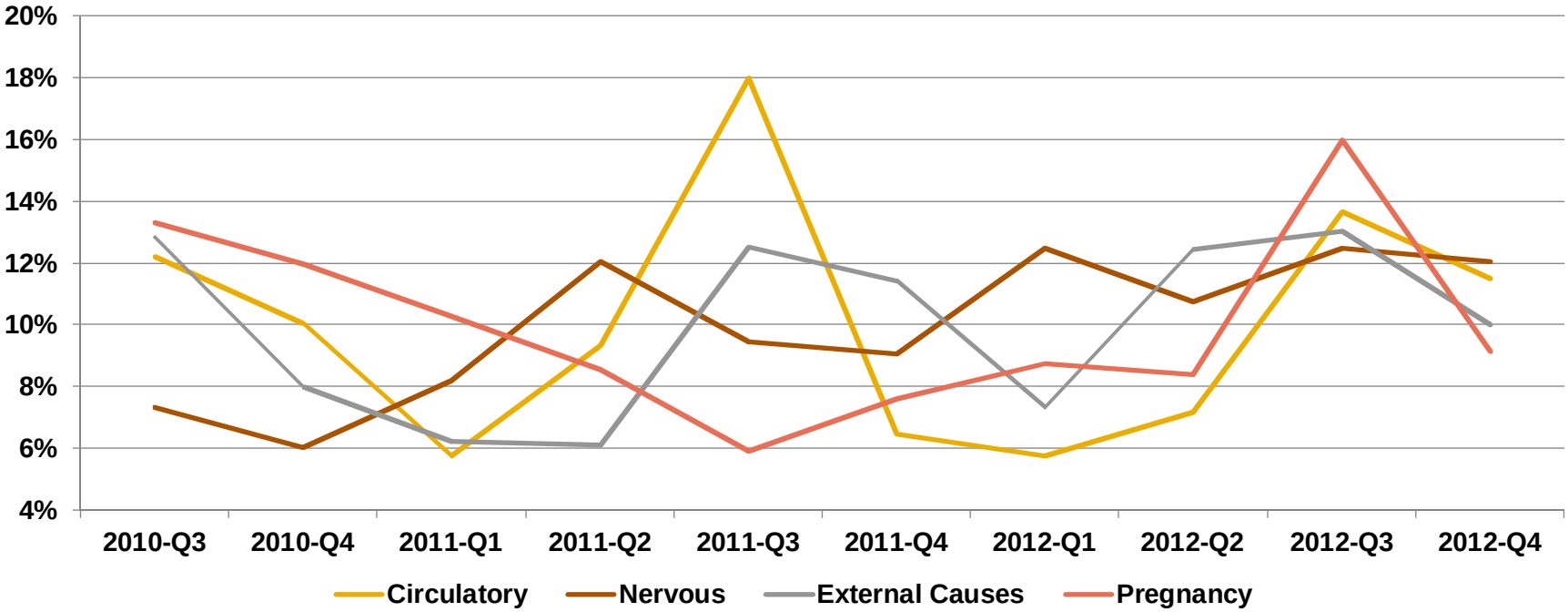


- Ear, Endocrine, Perinatal and Pregnancy are among top 5 diseases for 2010 and 2011
- While for 2012, Circulatory and Nervous system related ailments constitute lead the list
- Number of claims made against Injury (mainly fractures) has been very high for employees below 45 years

Claim Distribution – By Disease Category

- % of claims reported during each of the quarters by top 4 disease category

Table 12



- Claims due to circulatory problems are observed to be high for Q3 as compared to Q2

Additional Analysis

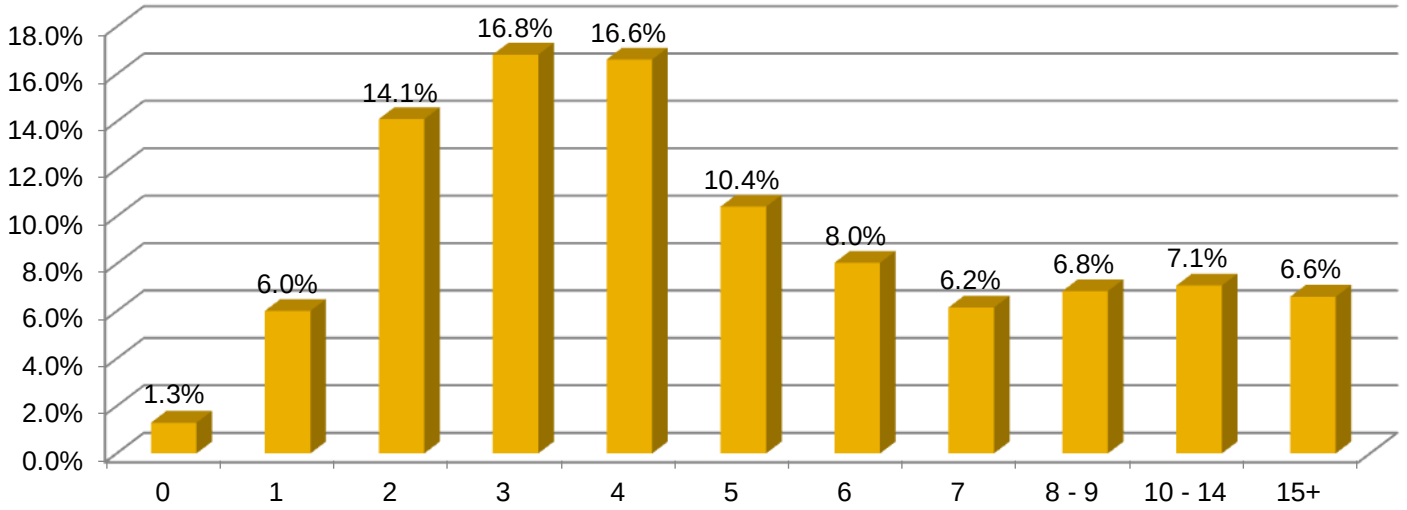
- Probable reasons for increased costs could be because of
  - Coverage of more small claims increased administrative costs
  - Lack of network hospitals increased expected claim costs
  - Utilization of room in higher categories than allowed lead to higher costs
  - Employees covered under lower Tier cities availed medical facilities in Tier I cities
- Improved employee education could have lead to reduced rejection of claims
  - Improved service quality
  - Lower administrative cost
- Premium sharing by employees may have lead to anti-selection

EMPLOYER'S PERSPECTIVE

Employer's Interests

- Loss of employee work days (sick leaves from HR System)

Table 13



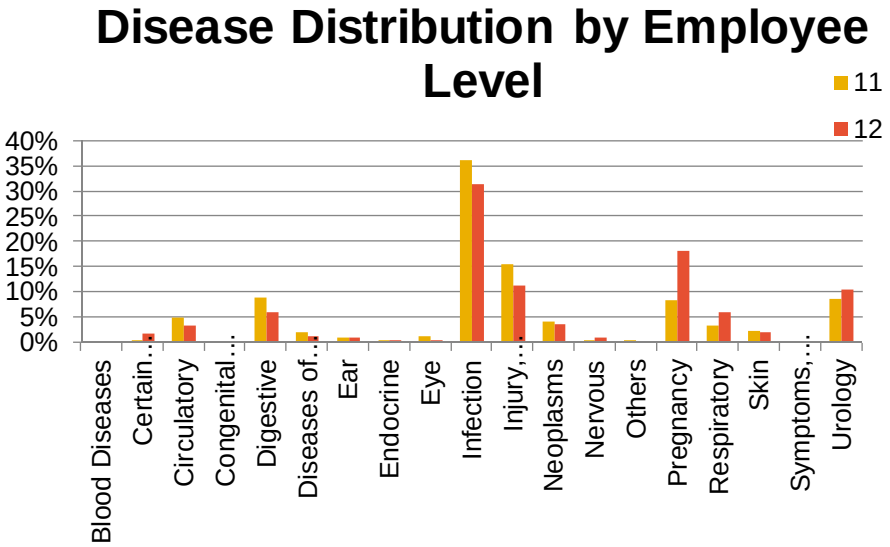
- Type of employees on Leave

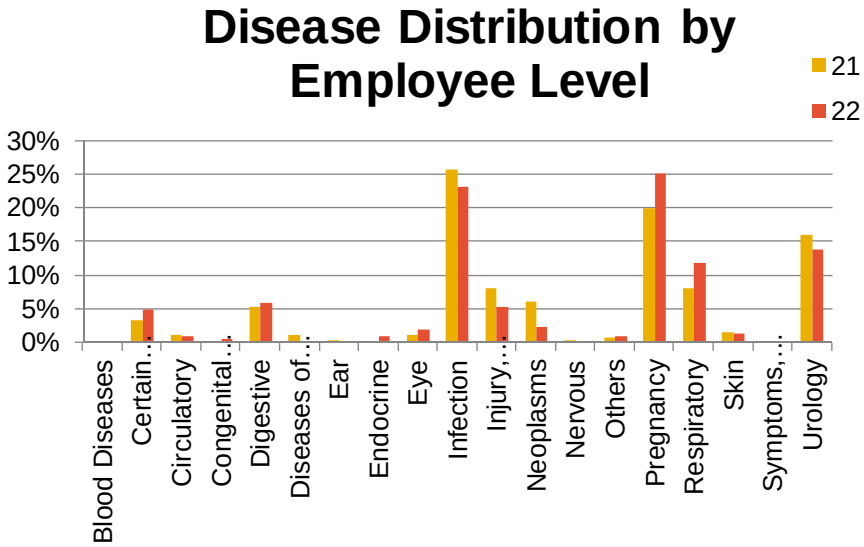
Table 14

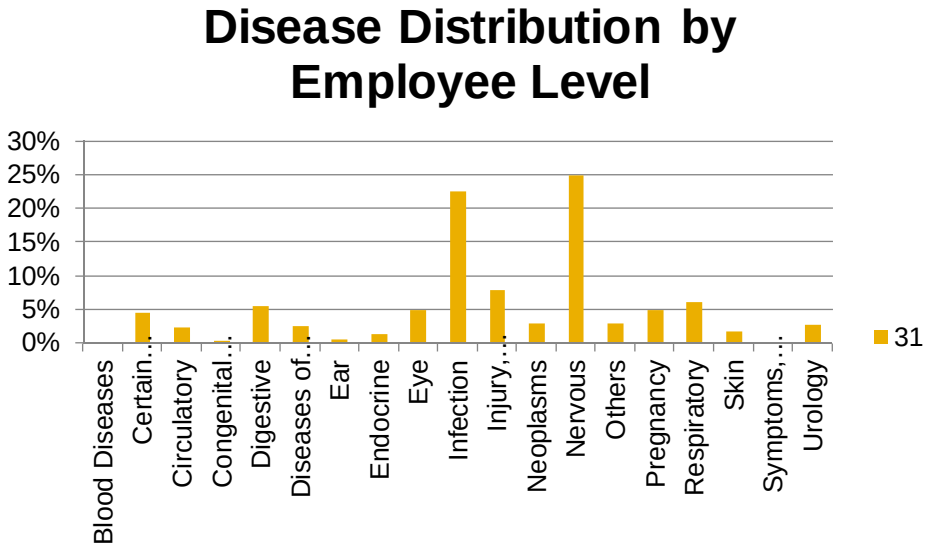
2012 (Ratio of Sick Employees to Total No. of Employees)				
Level of Employees	Q1	Q2	Q3	Q4
11	20%	10%	10%	20%
12	20%	10%	13%	33%
21	3%	6%	6%	9%
22	6%	10%	6%	11%
31	17%	23%	10%	20%
Total	**16%**	**10%**	**10%**	**22%**

Disease distribution by Employee Level

Table 15







Take Away from Tables

- Chart 1
  - Over 32% of the employee base who fell ill were hospitalized for duration of 3-4 days
  - Around 26% of the claimants were hospitalized for over one week duration
- Chart 2
  - Q1 and Q4 are impacted by highest level of sick leaves
  - Level 11, 12 (two lowest levels) and Level 31(highest level) have highest level of sick absenteeism
- Chart 3, 4 and 5
  - Infection, injury, pregnancy and urology related ailments form top four ailments for 1X level
  - Infection, pregnancy, urology and respiratory ailments form top four ailments for 2X level
  - For level 3X nervous system , infections, injury and respiratory related ailments form top four ailments

Employer's Interests

- Employer's Commercial interests lie in
  - Identifying total expense/revenue loss associated with loss of sick leaves and lack of productivity
  - Identifying controllable ailments and diseases impacting employee base
  - Limiting worsening effect on future health insurance premium
  - Providing for healthcare needs of its employees
  - Containing employee healthcare related cost by alternative ways like insurance plan restructuring/ disease management/ wellness programs/ resource management/ self-retention of medical costs

EMPLOYEE'S PERSPECTIVE

Employee

- Overall increased demand for quality care (preference for private healthcare hospitals)
- Increasingly wider coverage
- Able to meet cost of varying healthcare needs with time and able to provide for them
- Learn about personal health risks and how to control them
- Increased interest in remaining healthy rather than seeking healthcare services
- Reduce loss of income due to illness

FINAL TAKE-AWAYS…

The Changing Role of an Actuary

Move from a ***"What happened?"*** mentality to more of a ***"What will happen?"*** approach

New roles and responsibilities for us:

- Trusted Advisor
- Advanced analytics
- Looking Forward
- Predict
- Optimization
- Key Performance Predictors

In the end, these new advanced analytic capabilities will make actuaries even better business partners than they are today!