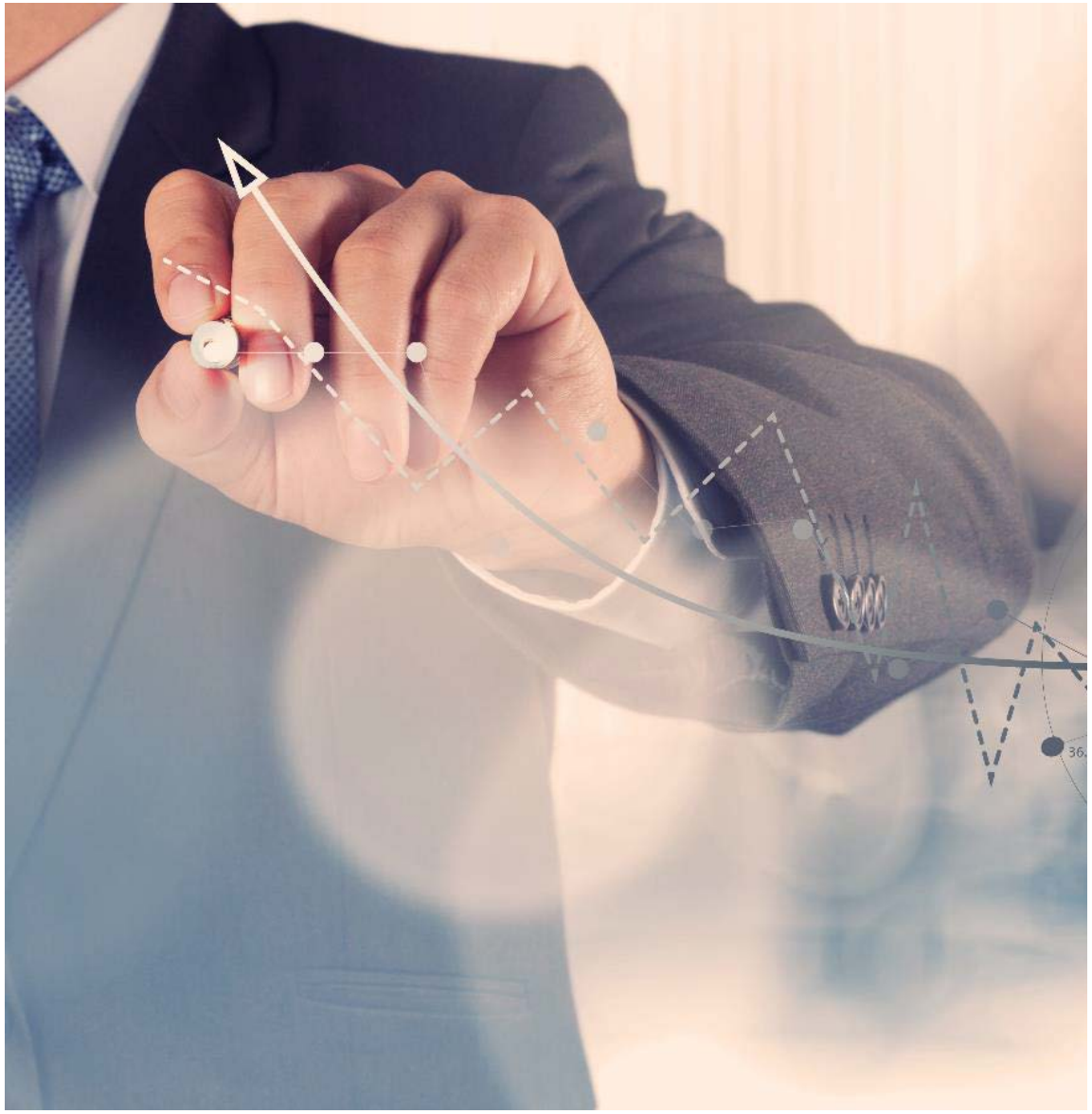
Analytics in Health Insurance
4th Seminar on Data Science and analytics, Indian Actuarial Institute(IAI)
11-2-2019

Prashant Rai, SAS consultant-Insurance



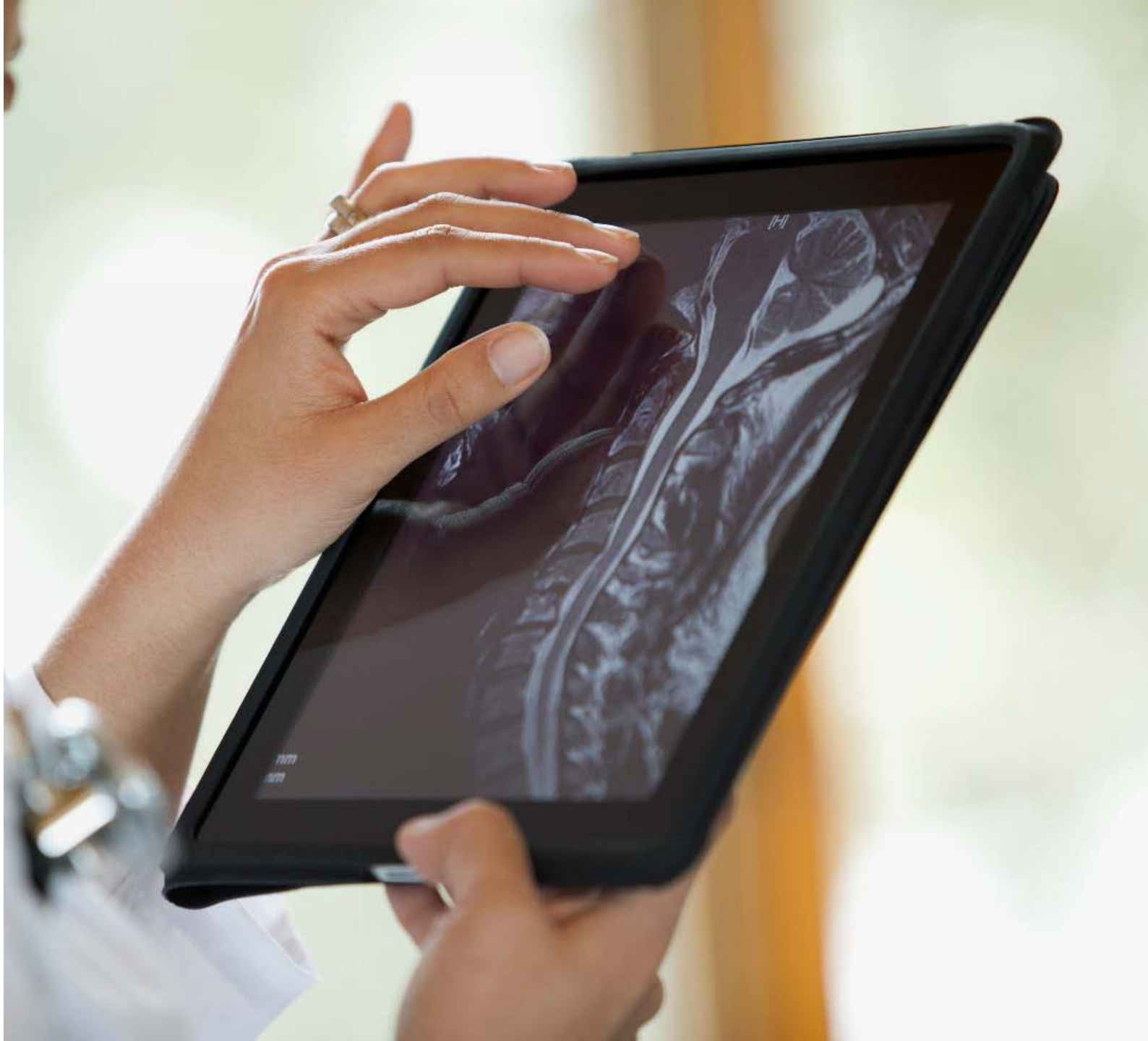
Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

AGENDA

1. Analytics - health insurance
2. Challenges and Opportunity
3. Strategy-People-Process
4. Insurance Analytics-Why Now!!!
5. Analytics-Area of application
6. How a complete analytics framework looks like- Elaboration through fraud management

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Healthcare Data is the new oil and Exponentials are the Refineries



Why Health Insurance needs data analytics?

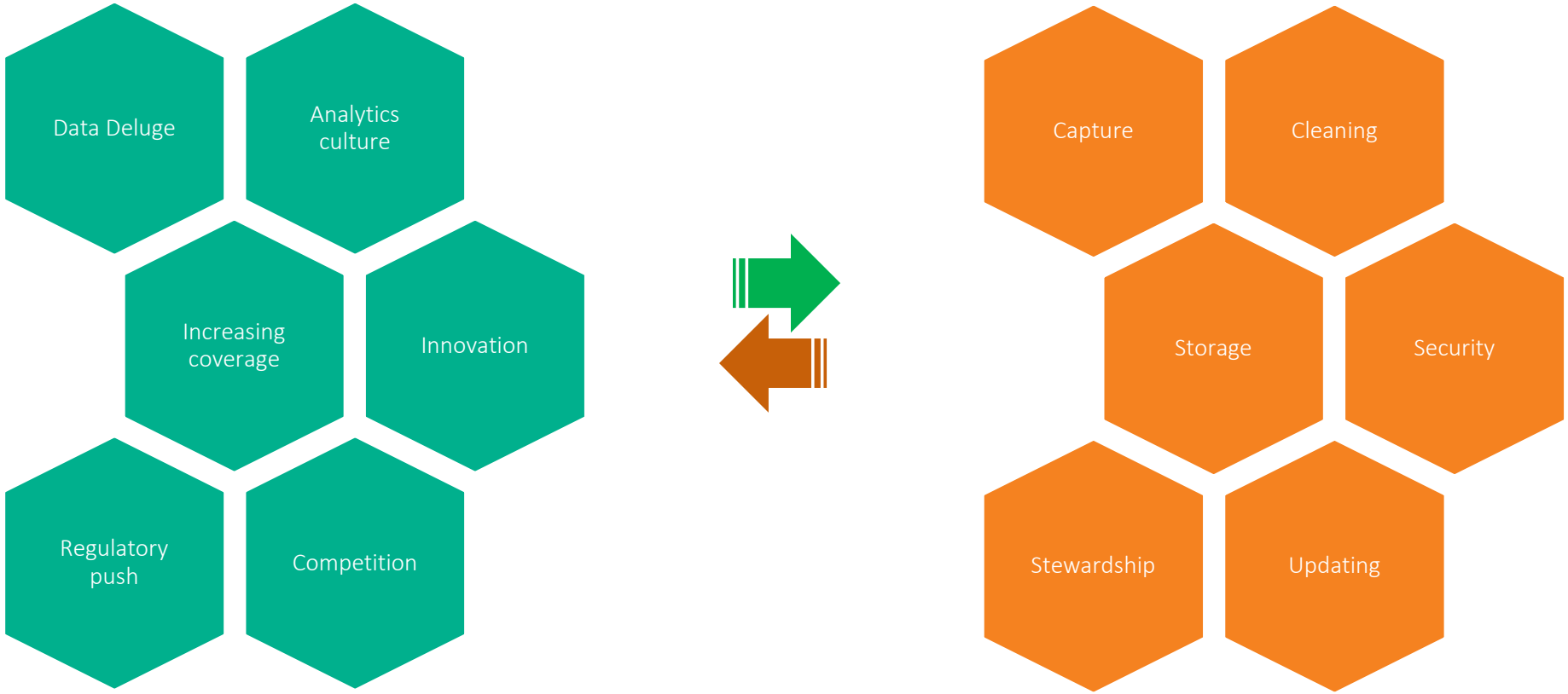
1. The cost of healthcare in India is increasing at 20% every year
2. There is a shortage of 1.5 million doctors and 2 million Hospital beds
3. With health insurance, only around 5% of the middle class have health coverage, and catastrophic coverage is even lower.
4. Ayushman Bharat has covered around 50 Cr. Indians but it is still inadequate and half of the population is still uncovered
5. Fraud Abuse and wastage is rampant and it is growing menacingly with respect to growing coverage
6. Innovation in digitalization and analytics is still underinvested and long term strategy for its consumption is pending since a long time

national health authority

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Opportunity and challenge in health insurance analytics

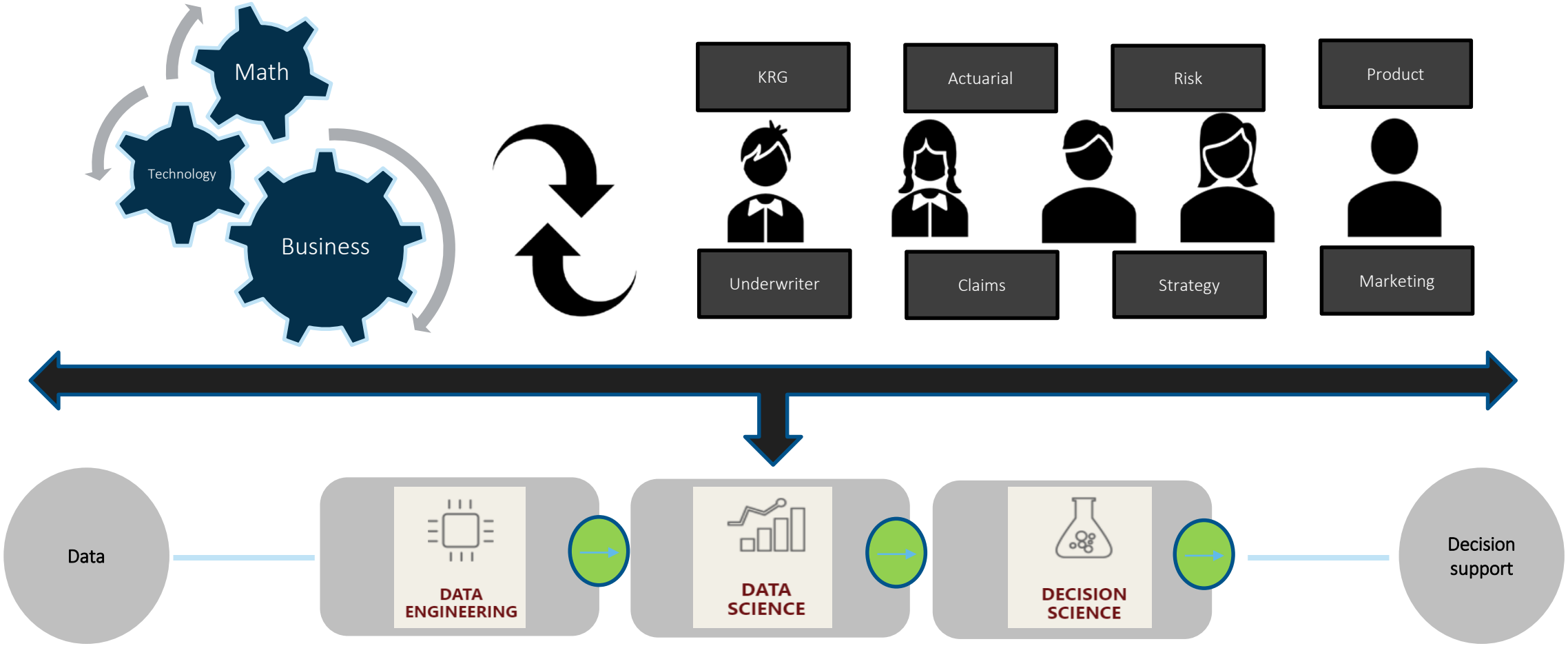
Analytics in healthcare comes with many challenges, including security, visualization, and a number of data integrity concerns.



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Framework- Successful analytics initiative

- Analytics is all about right strategy, analytics culture in decision making sitting at right process.
- Secondly, analytics adds to your great business sense which is already therewith great people working with the company.



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Exciting time for data analytics in Health Insurance

- Traditionally, for most organizations data analytics meant purchasing a Business Intelligence tool and creating reports. The impact of such tools was minimal as most healthcare data is both non-standard and unstructured (text, images) and data from multiple sources had to be combined to create value
- Natural Language Processing, image recognition and speech analysis combined with the availability of large-scale computing power, the ability to convert this type of information into analyzable signals is now within reach

Sources of the data deluge

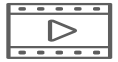
Mobile

EMRs

Paper / Text Documents

Social Media

Images


Videos

Sensors / Devices

40-50%
Annual growth in digital data volume

62%
Annual growth in unstructured data

~9X
of unstructured data vs. structured data by 2020

Advances in computing power and techniques

Smarter Algorithms

Faster Processing Speed

Improved Visualization

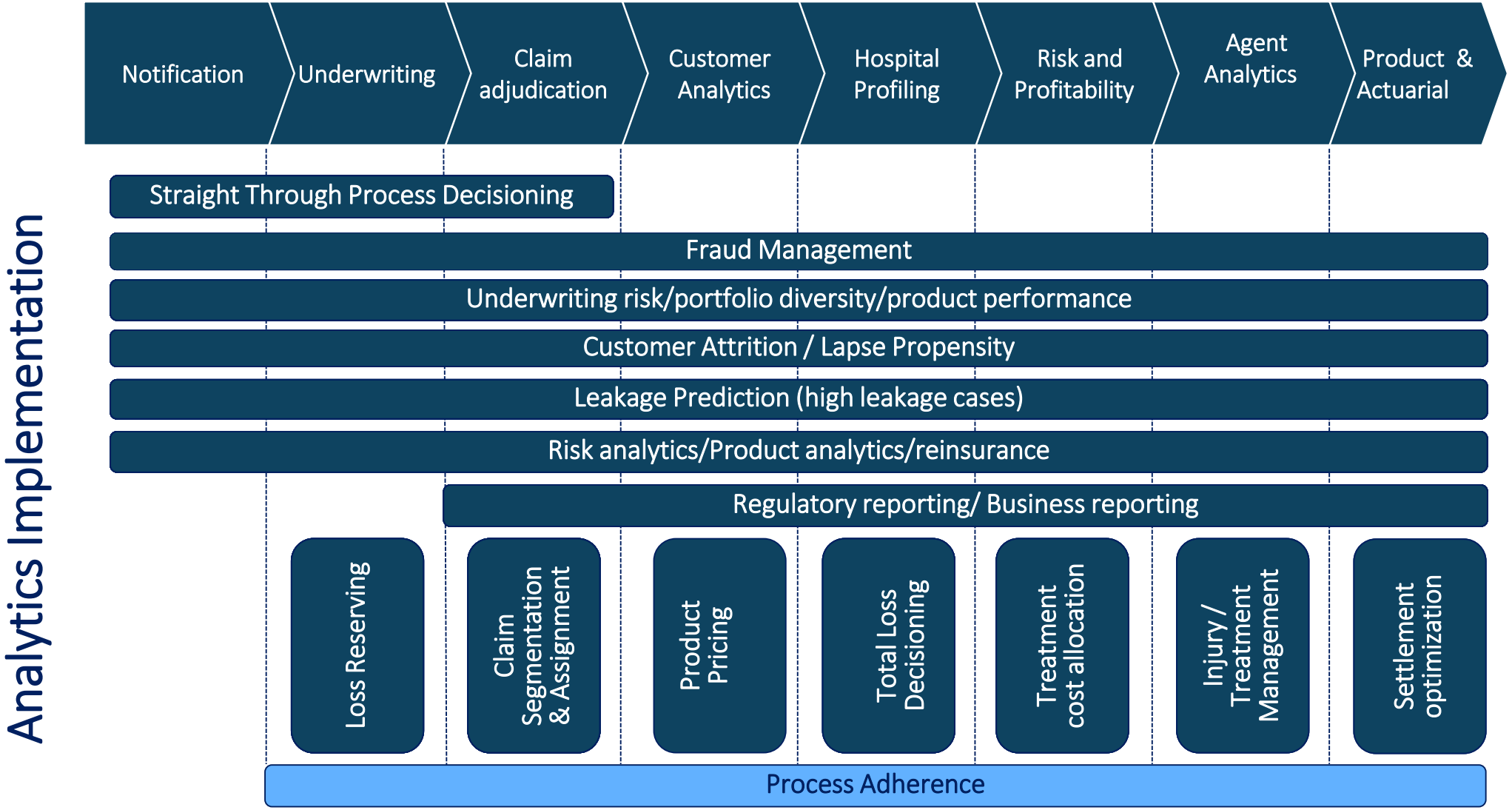
Retail Consumerism | Value Based Care | Personalized Digital Health | Preventative Medicine | Drug Trials and Discovery

Advances analytical and computing techniques coupled with the explosion of data in healthcare organizations can help uncover leading clinical practices, shrink research discovery time, streamline administration, and offer new personalized engagement paradigms at an industrial scale that align people's decisions and actions in ways that improve outcomes and add value

Customer Centric | Optimal Cost Structure | Adaptive Organization

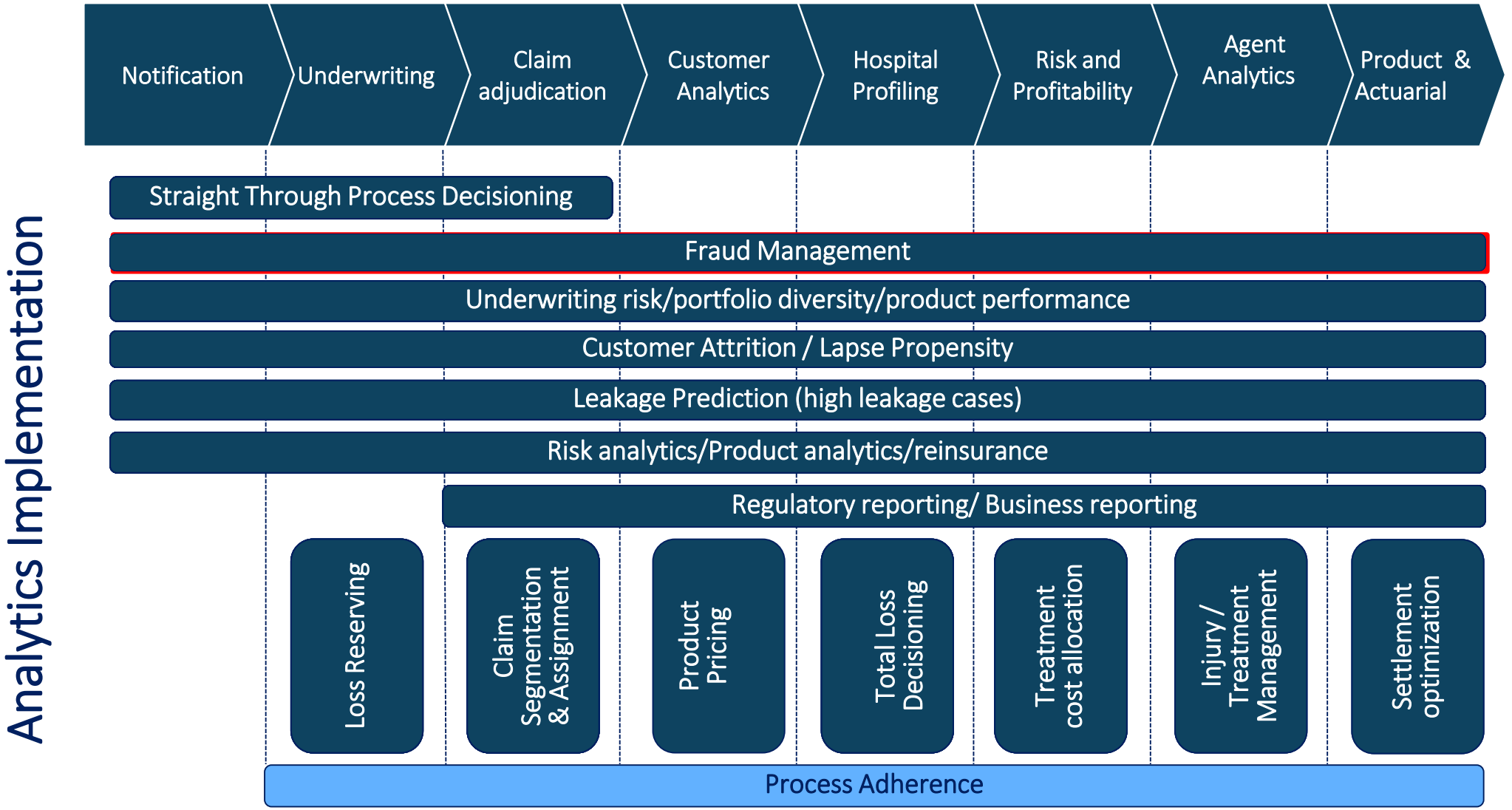
Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Analytics in the whole life cycle of health insurance



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Robust analytics framework-Fraud Management



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Fraud Management-Steps for effective decision making

- Fraud Management is not only about data analytics, it is much beyond that. It is a group of activity where all the stakeholders of claim life cycle will have to contribute into so that fraud management system tends to get wiser by every day.

1	2	3	4	5	6
Data Management	**Data Visualization and trigger**	**Entity Scoring and anomaly detection**	**Analytical Modeling**	**Entity Networking**	**Image and text analytics**
➓Data Relationship Relationships. ➓Structured data. ➓Unstructured data(text and Image).	➓Slicing & dicing ➓Dashboard-dynamic and presentation ➓Deduplication and text clustering ➓Dashboard for Monitoring	➓Selection of risk variables from the data set ➓Risk scoring of claim entities such as claim area, garage, surveyor, Driver, Vehicle ➓Risk Gradation	➓Loss Estimation(severity modeling) Model for possible profile ➓Live Comparative study of Loss vs estimation ➓Breach frequency calculation	• Highlight the risky garages through Trigger, Anomaly, Breach frequency. • Create opt graph for robust community relationship • Create suspicious network between entities	• Image analytics-Damage magnitude • Image matching • Unstructured data analysis and setting up dashboard for document received.

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Chips of Fraud Management-A successful initiative

- Rare Diseases
- Uncommon Diseases
- One Time Procedures
- Age Gender Related Procedures
- Treatment Specific configuration such as repeated admission, dose regulations, chronic / acute diseases
- Seasonal Diseases
- Diseases by Geography

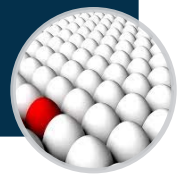
Policy Configurations

- Triggers to be Configured
- Risk Scoring of Private Hospitals by simulation of weight coefficient for risk variables
- Risk Score Ranges from 416 to 944 on a scale of 0 to 1000
- Reduction of false positive

Business Rules

- hospitals are Clustered into Clusters based Region, City Tier, Hospital Type (Private, Public), Specialty (Heart, Cancer, Both etc), Hospital Capability based on Bed Strength
- Anomaly Detection Rules comparing hospital behavior with Clusters. Anomaly comprises of Admission Rates, Package utilization, Treatment Costs los etc.

Anomaly Detection



- Fuzzy algorithm was used to calculate match score of demographic variables
- Analysis was extended further to compare TMS data on add member

Beneficiary validation

- Four Entities i.e Family, Beneficiary, Hospital and Treating Doctors are taken for analysis
- Identify Patterns of suspicious behavior of entities using Louvain Algorithm
- Networks comprising of # Cases amounting to INR # Crores were shared for investigation. Average value of the suspicious network is # Lakh

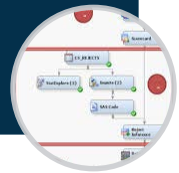
Network Analytics

- # Case were rejected and Reasons for rejections were analyzed
- # unique reason has been clustered into # group based on the similarity of the key words.
- Further these rejection reason has been categorized into fraud and abuse

Unstructured Data

- Predictive Models based on rejected Reasons
- Classification algorithms have been used to score the probability of the fraud based on the event rate on unstructured data.
- Fit statistics of the competing algorithm compare to decide the best fit

Predictive Models

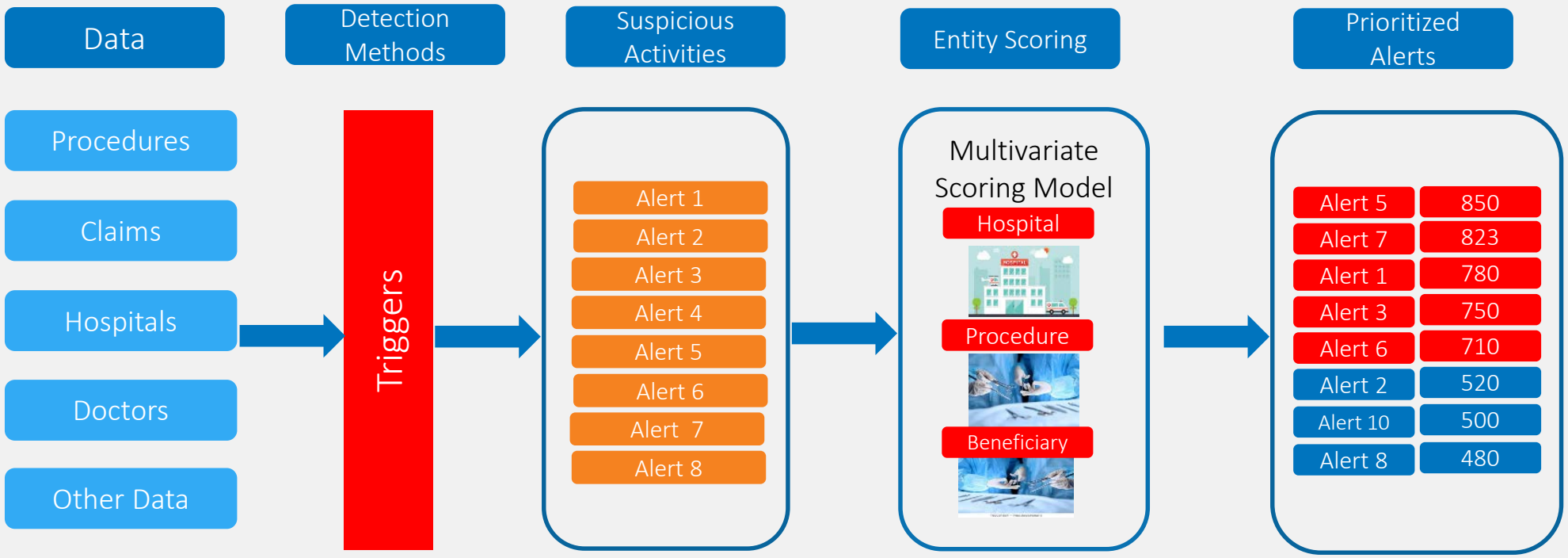


- De-duplication of the image and image comparison of patient.
- Reading the summary through OCR from Discharge summaries

Image Processing

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Skeleton-Fraud Management system



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Data Assessment

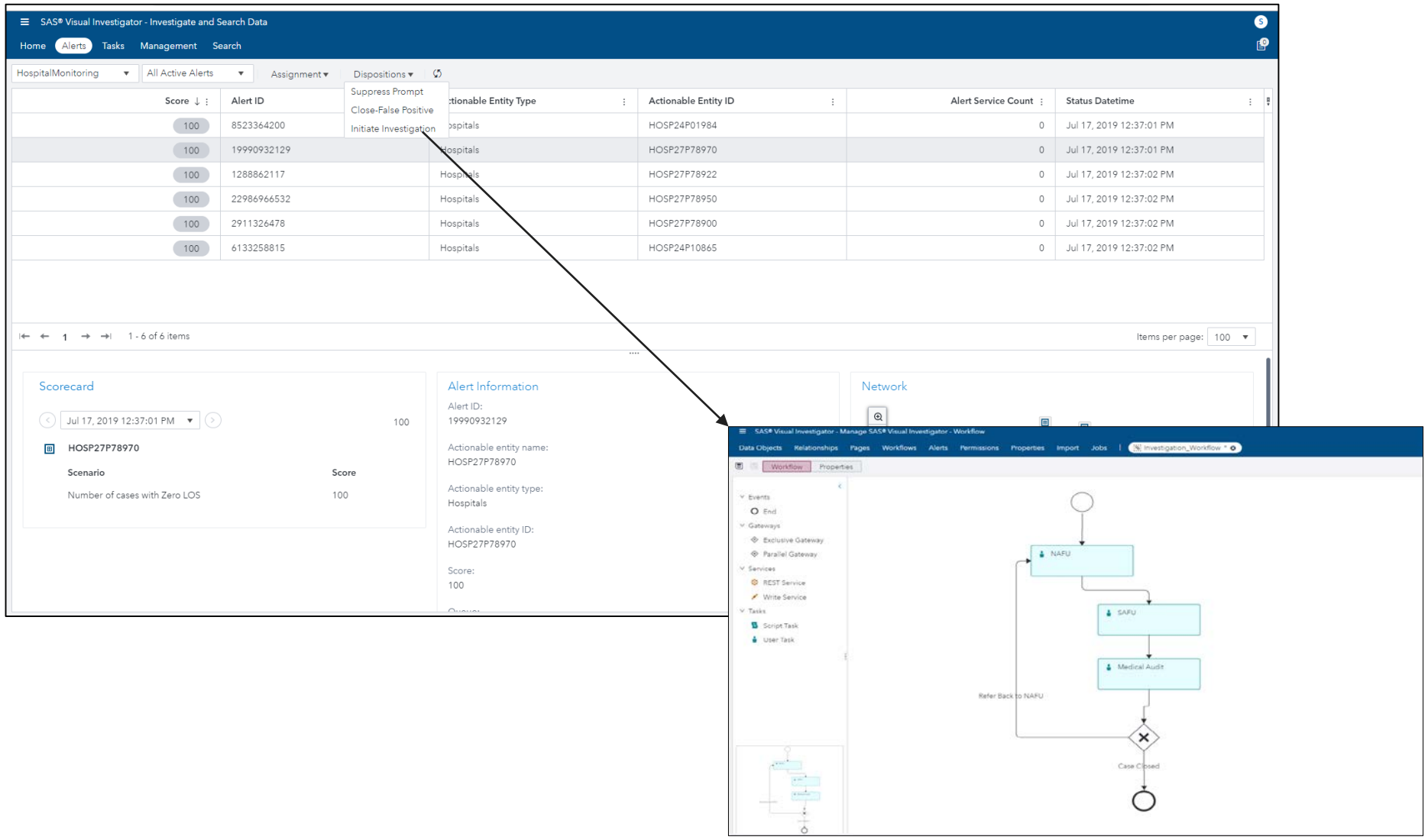
Configuring Triggers

Risk Scores

Suspicious Cases

Case Management

Trigger Summary



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Anomaly Detection

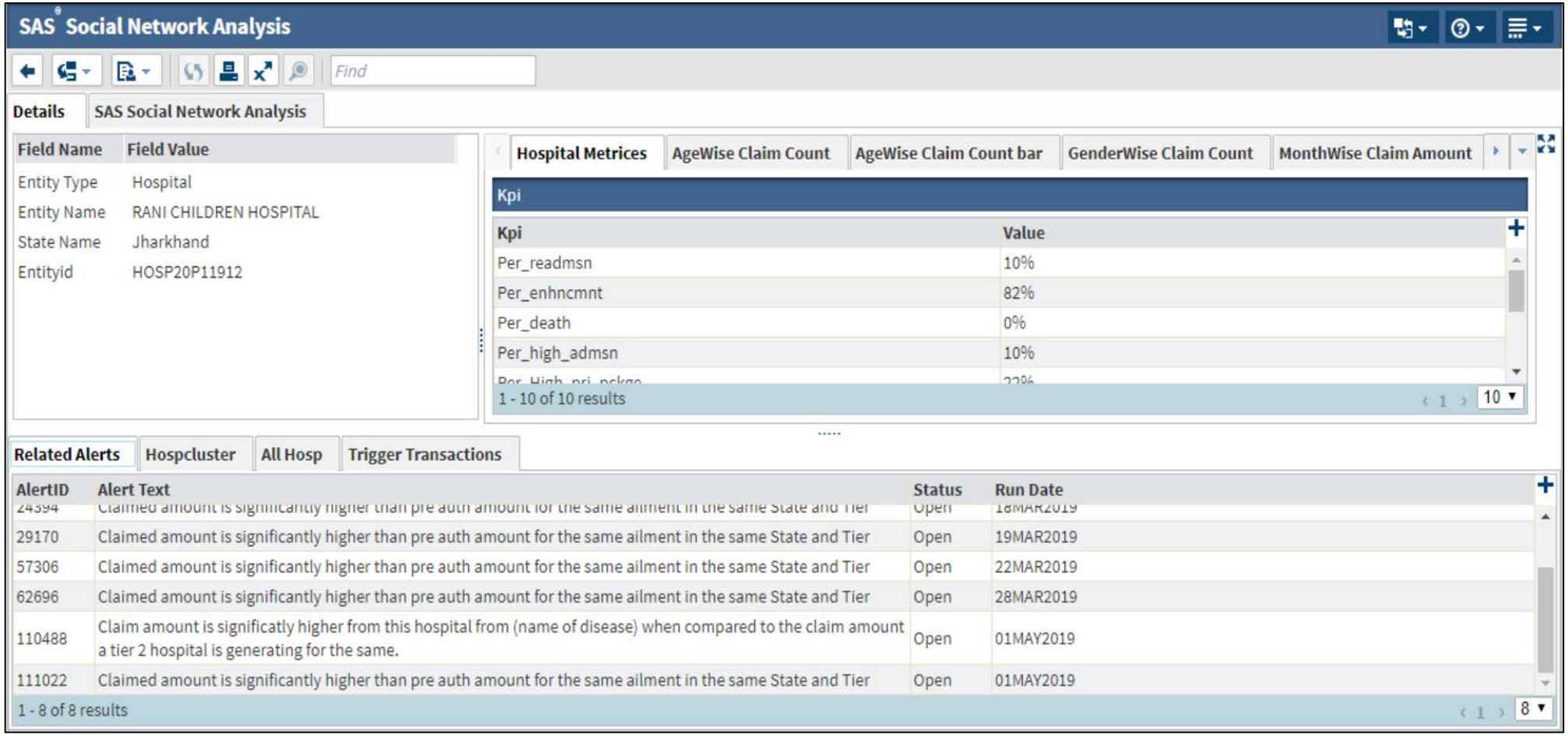
Clustering of Hospitals

Aggregation of Metrics

Anomaly Detection

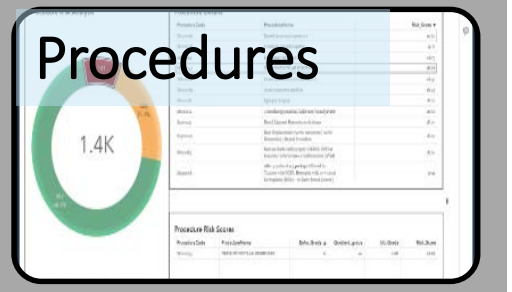
Risk Scoring

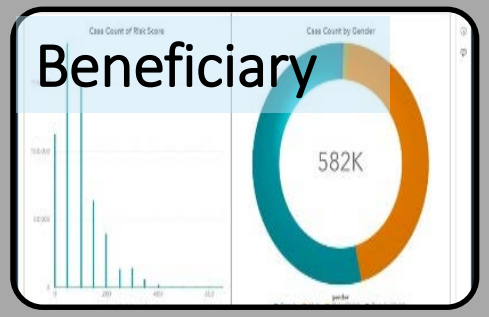
Suspicious Hospitals

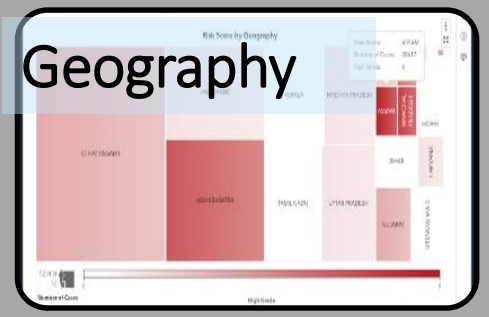


Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Risk Scoring







Hospital Risk Score

- Over 8k Private hospitals, scored on a scale of 0 to 1000 based on 10 different variables.
- About 2% private hospitals has been designated as highly risky.
- Gujrat has 43 hospitals coming as highly risky
- 75% of the total risky hospitals are concentrated in four states that includes Gujrat, Tamilnadu, UP and Jharkhand

Beneficiary Risk Score

- 98% of the beneficiary are in low or moderate risk category whereas 2% of beneficiary portfolio is high risk

Procedure Risk Score

- 1391 procedures have been scored on gradient based variables
- Around 3% of the procedures are changing rapidly over various metrics

Geography Risk Score

- Himachal Pradesh, Bihar, Uttarakhand are the geography which are observing high risk score among all the states.
- Smaller states in north such as Jharkhand, Himachal Pradesh, Uttarakhand and states from north east has relatively higher risk score compared to bigger states.

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

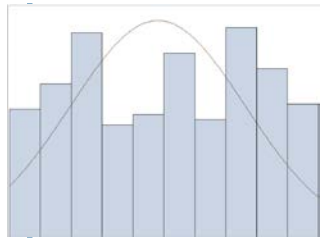
What Contributed to High Risk Score

- A Typical risk scoring process starts with identification of risk variables, testing its significance, sympathizing an equation and validating the equation for its ability to differentiate between high, moderate and low risk

Risk Scorecard

Variable	Gradation
LOS Risk Score	100
Claim Denied Risk Score	12
Death Risk Score	40
Enhancement Risk Score	100
Hi Admission Risk Score	20
Hi price package Risk Score	120
Hi utility Risk Score	12
Readmission Risk Score	80
Ticket Risk Score	240
Total Risk	704

Distribution comparison and moment statistics



The UNIVARIATE Procedure
Fitted Normal Distribution for Claim_Amt

Parameters for Normal Distribution

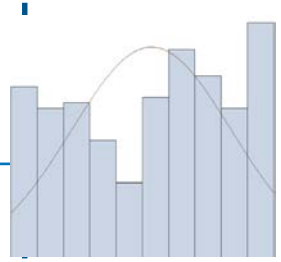
Parameter	Symbol	Estimate
Mean	Mu	64110.05
Std Dev	Sigma	36421.48

Goodness-of-Fit Tests for Normal Distribution

Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.09370398	Pr > D	<0.010
Cramer-von Mises	W-Sq	0.79522897	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	5.49007583	Pr > A-Sq	<0.005

Quantiles for Normal Distribution

Percent	Quantile	
	Observed	Estimated
1.0	2737.00	-20618.98
5.0	7129.50	4202.05
10.0	11677.50	17434.05
25.0	28639.50	39544.14
50.0	70191.00	64110.05
75.0	94995.50	88675.97
90.0	112296.00	110786.06
95.0	116448.00	124018.06
99.0	119261.50	148839.09



The UNIVARIATE Procedure
Fitted Normal Distribution for LOS

Parameters for Normal Distribution

Parameter	Symbol	Estimate
Mean	Mu	10.60667
Std Dev	Sigma	5.643117

Goodness-of-Fit Tests for Normal Distribution

Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.11520488	Pr > D	<0.010
Cramer-von Mises	W-Sq	0.72930674	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	4.87831523	Pr > A-Sq	<0.005

Quantiles for Normal Distribution

Percent	Quantile	
	Observed	Estimated
1.0	1.00000	-2.52119
5.0	2.00000	1.32456
10.0	3.00000	3.37472
25.0	6.00000	6.80044
50.0	11.00000	10.60667
75.0	15.00000	14.41289
90.0	18.00000	17.83861
95.0	19.00000	19.88877
99.0	20.00000	23.73452

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Rejected Reasons

NLP

Keywords

Topics

Categories

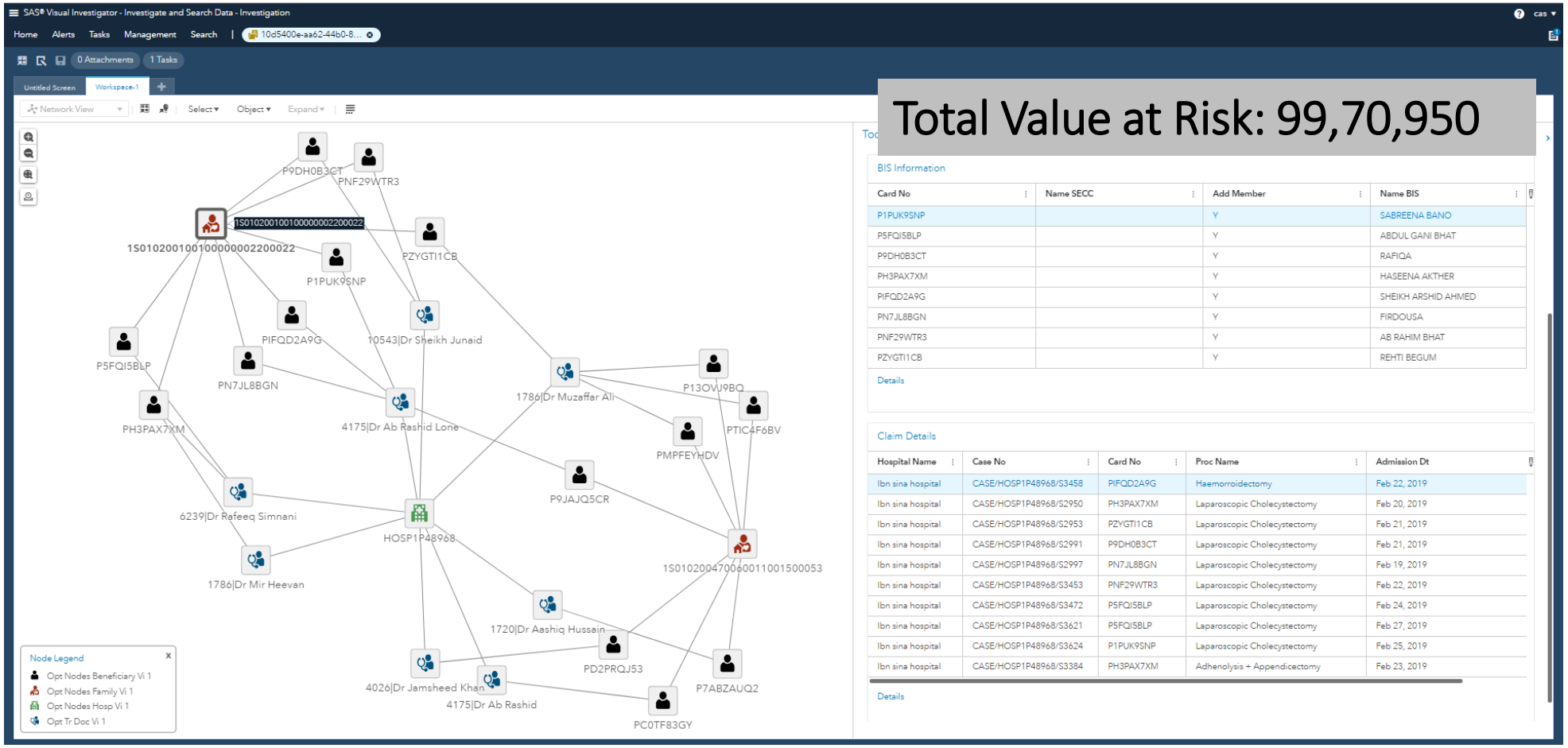
Clustering-Rejection Reason

Cluster ID	Topics	Cluster Name	Category
1	reject +report +require +justify +query +chart OPD +block less +reply +investigation +admit reject payable +submit	Hospital could not justify line of treatment given	Fraud
2	hospitalization +justify +case OPD basis 'OPD basis' +admission +patient +treat HR +complete +approve 'bed capacity' bed capacity	Hospitalization was not required. OPD was enough for treatment	Abuse
3	+intimation +late +approval company insurance +require 'late intimation' 'insurance company approval' 'intimation insurance company approval' 'insurance company approval' insurance +eat +await 'IC approval' IC	Late intimation of the claim to the insurance nodal agency and claim adjudicator	Fraud
4	mismatch +package +diagnosis +select +treatment +disease 'disease diagnosis treatment' kindly +attach package correct +block hydrocele excision 'correct package'	Incorrect booking/selection of packages	Abuse
5	mandatory +document +reminder +process +'mandatory document' unable +reject repeated +query +'repeated reminder' +reply +hospital multiple +raise +'multiple query'	Madatory document were not submitted despite repeated query/Reminder	Fraud
6	+case +reject +process unable +claim +note +discrepancy discharge clinical +treatment +summary +'clinical note' +surgery duplicate +overwrite	Discrepency in admission and discharge summary	Fraud
7	date +process +case unable registration +admission delay +delay +hour +intimation +day IP discharge +register	Inconsistency in Claim date,admission date discharge date and claim intimation dates	Fraud
8	+claim +day +hospital verification rejection right subject appeal +deny 'case verification' +stand +'claim stand' ehcp +document guidelines	Document was either not submitted according to PMJAY guidelines or it was blurred and unreadable	Fraud
9	+package +reject +block wrong kindly +treatment 'wrong package' package medical +tooth single +raise incorrect +select +procedure	Similar/incorrect/wrong packages are being blocked for	Fraud

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Network Analytics

- Entities
- Relationships
- Pattern Identification
- Suspicious Networks



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Image Deduplication

Images

Image Curing

Image Deduplication

Suspicious Cases

Mean Square Error & Structural Similarity Index

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

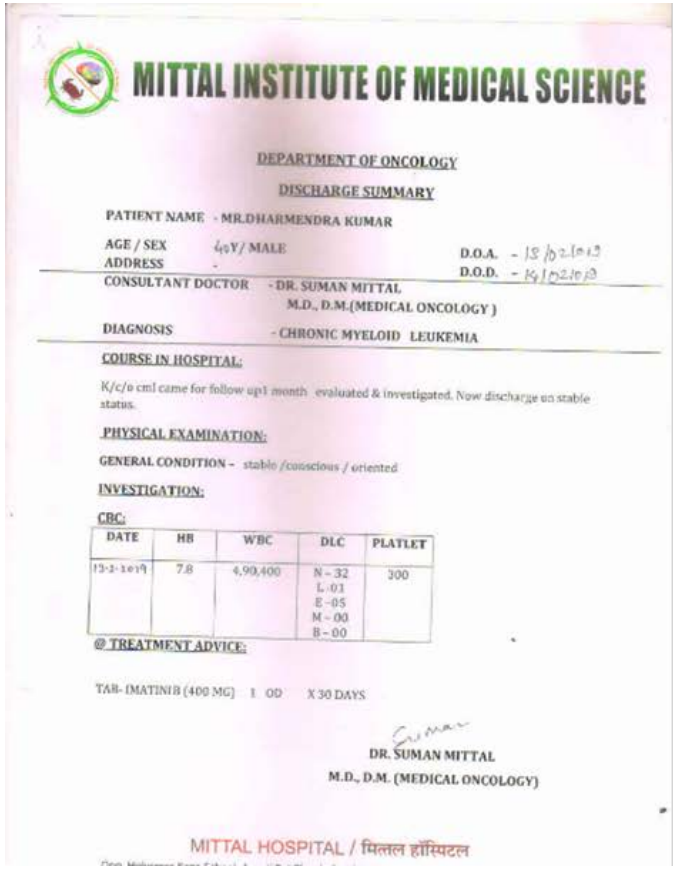
Image Deduplication

Discharge Summaries

OCR Conversion

Structure Data

Matching with Cases



MITTAL INSTITUTE OF MEDICAL SCIENCE

DEPARTMENT OF ONCOLOGY

DISCHARGE SUMMARY

PATIENT NAME - MR.DHARMENDRA KUMAR

AGE / SEX 40Y/ MALE D.O.A. - 18/02/2019

ADDRESS - D.O.D. - 28/02/2019

CONSULTANT DOCTOR - DR. SUMAN MITTAL
M.D., D.M.(MEDICAL ONCOLOGY)

DIAGNOSIS - CHRONIC MYELOID LEUKEMIA

COURSE IN HOSPITAL:

K/c/o cml came for follow up1 month evaluated & investigated. Now discharge on stable status.

PHYSICAL EXAMINATION:

GENERAL CONDITION - stable /conscious / oriented

INVESTIGATION:

CBC:

DATE	HB	WBC	DLC	PLATLET
13-2-2019	7.8	4,90,400	N - 32 L -01 E -05 M - 00 B - 00	300

@ TREATMENT ADVICE:

TAB- IMATINIB (400 MG) 1 OD X 30 DAYS

DR. SUMAN MITTAL
M.D., D.M. (MEDICAL ONCOLOGY)

MITTAL HOSPITAL / मित्तल हॉस्पिटल

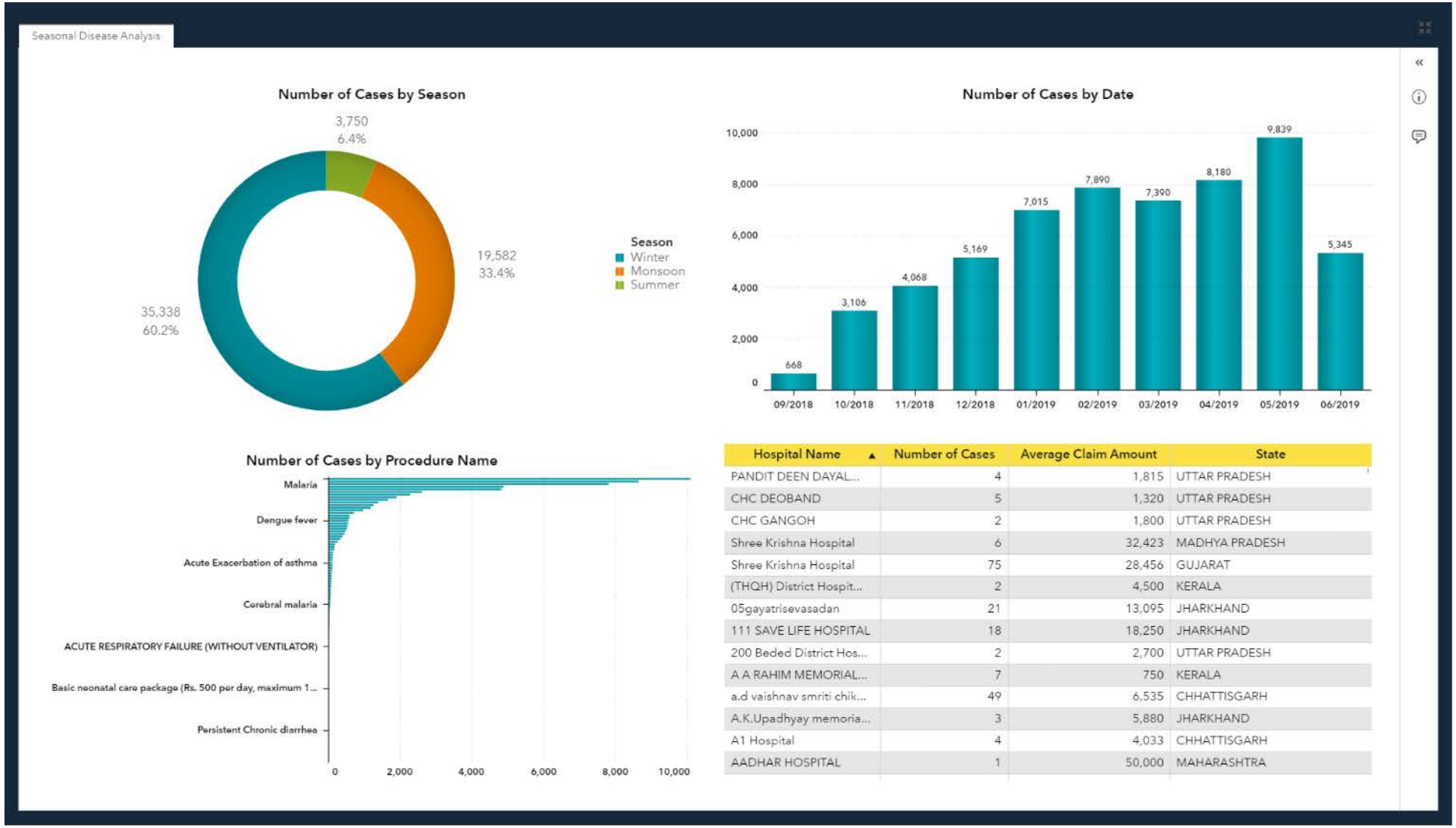
OCR

```
.- MITTM INSTITUTE llf MEIIIBAI SBIEHGE

 E
I2T§flARG].".§llMMARX
PATIENTNAME - MRDHARMENDRA KUMAR



AGE/SEX -36 Y/MALE D.D.A. -28/02/19
ADDRESS - I-IASDA D.O.D. - Z8/D2/19
CONSULTANT DOCTOR - DR. SUMAN MITTAL

M.D., D.M.(MEDlCAL ONCOLOGV)









DIAGNOSIS ~ CHRONIC MYELOID LEUKEMIA
§0.1.LRSli.IN.l'IflSI'_lIAL;

K/c/0 tml came for follow up] month evaluated (5: investigated. Now discharge on stable
status.


GENERAL CONDTTIDN~ stable /cnnscxous / urxemed

masnnarm;
gm _ _
DATE \ HB WBC FLATLET
28/02/19 12.1 41,200 486 '
muaxunmamzxcm

'rAE~IM/\TIN1H(4nn MO) I on XKUDAYS

DR. SUMAN MTTTAT.
M.D-. D.M. [MEDICAL ONCOLOGY)
```

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Prevention Insights

- Rare disease
- Not common disease
- Seasonal disease
- One time disease



Hospital Name	Number of Cases	Average Claim Amount	State
PANDIT DEEN DAYAL...	4	1,815	UTTAR PRADESH
CHC DEOBAND	5	1,320	UTTAR PRADESH
CHC GANGOH	2	1,800	UTTAR PRADESH
Shree Krishna Hospital	6	32,423	MADHYA PRADESH
Shree Krishna Hospital	75	28,456	GUJARAT
(THQH) District Hospit...	2	4,500	KERALA
05gayatrisevasadan	21	13,095	JHARKHAND
111 SAVE LIFE HOSPITAL	18	18,250	JHARKHAND
200 Beded District Hos...	2	2,700	UTTAR PRADESH
A A RAHIM MEMORIAL...	7	750	KERALA
a.d vaishnav smriti chik...	49	6,535	CHHATTISGARH
A.K.Upadhyay memoria...	3	5,880	JHARKHAND
A1 Hospital	4	4,033	CHHATTISGARH
AADHAR HOSPITAL	1	50,000	MAHARASHTRA

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Thank you

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.