

10th Tech talk on Employee Benefits

06-June-2024

Bytes to Breakthroughs : The Power of Big Data

Jayesh Pandit

Designation Partner M/S.K.A.Pandit



Agenda

- Introduction to Big Data
- Big Data in Employee Benefits
- Data Analysis and Results
- Looking into future trends and Acknowledgements



Introduction to Big Data

- Structured, semi-structured and unstructured data characterized by variety, increasing volumes and more velocity.
- The large data sets are analyzed to reveal patterns, trends and develop analytical capabilities for well-informed decisions.

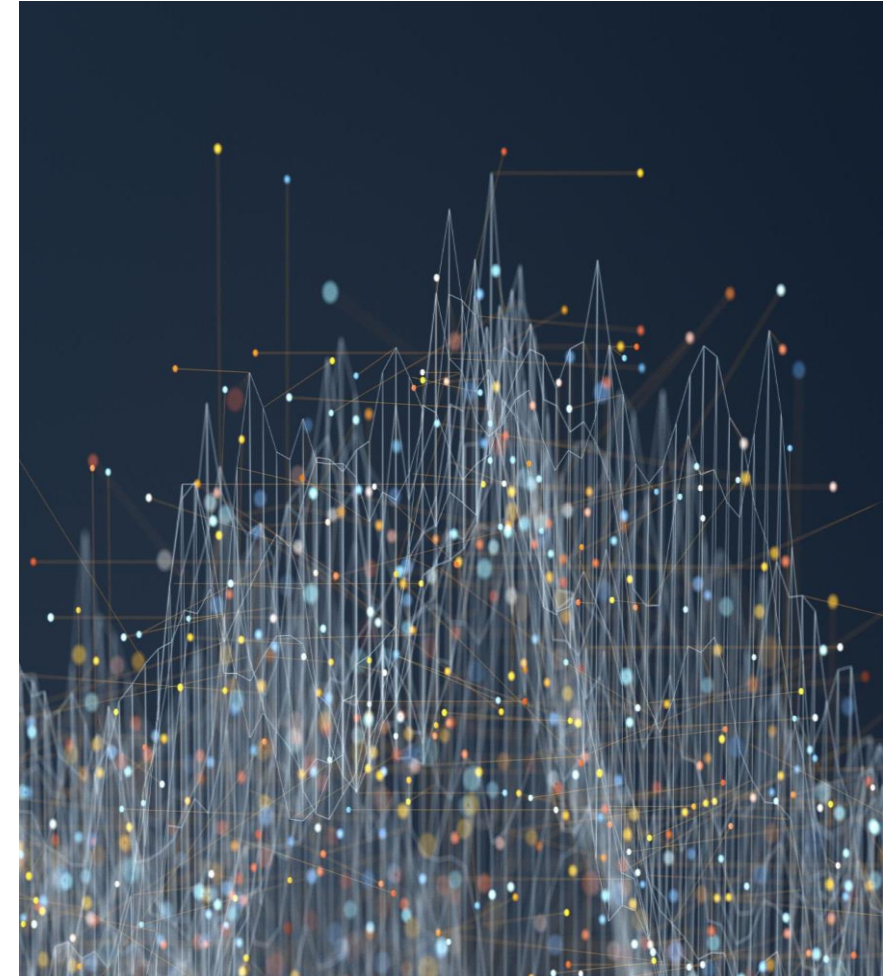
The 4 V's of Big Data:

Volume: Trillions of Gigabyte of data is generated everyday across the world

Velocity : The speed at which the data is generated, processed and analyzed

Variety : The characteristic of big data to have diverse information which is structured for analyzing

Veracity – Accuracy and Trustworthiness of data



Power of Big Data

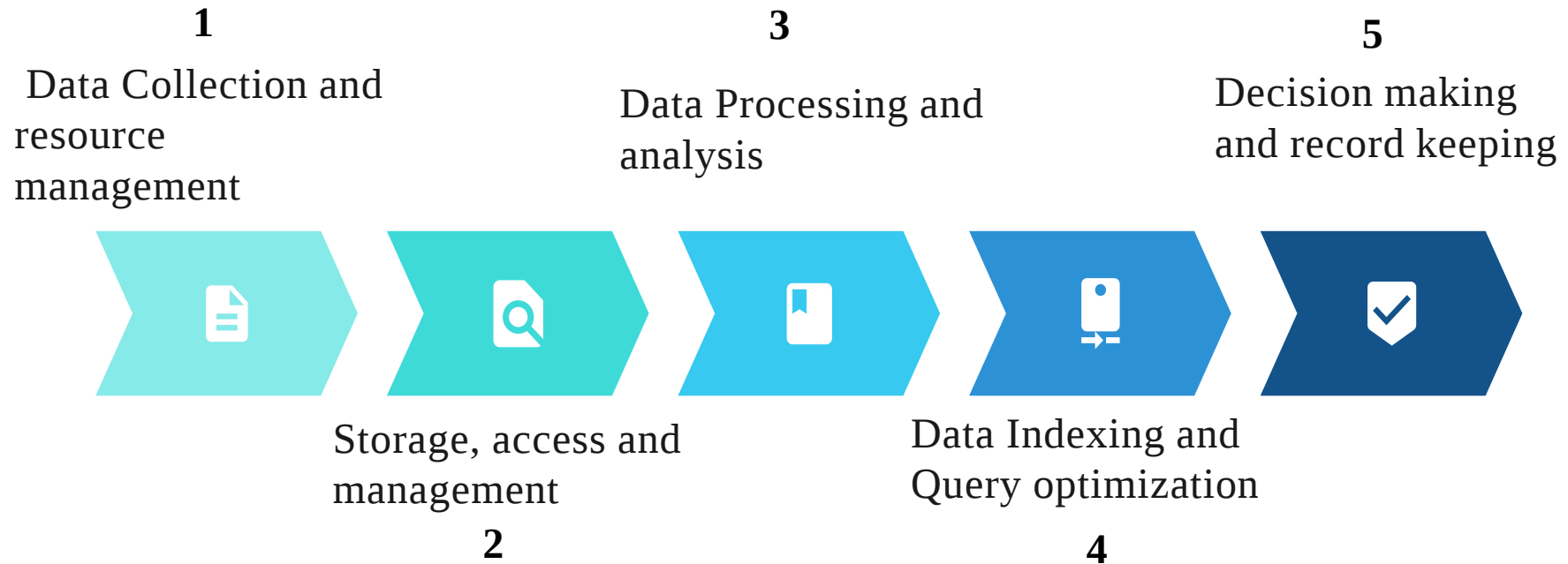
The access to big data and the fast-paced technology can contribute in revolutionizing the actuarial field by addressing complex problems and enhance the decision making.



- Increased Accuracy and stabilization of assumption analysis.
- Predicting future trends of attrition and new employment.
- Tailoring products adhering to unique customer needs and preferences. E.g. Savings and insurance
- Enhance the modelling for infrequent and landmark events in the industry.
- Increase the accuracy of provisioning and thus, consulting.



Process of optimizing Big Data



- Optimizing Big data is a continuous process that requires technological solutions along with strategic planning and organizational adjustments to ensure that the data is being used efficiently and effectively.

Organizational Evolution to Leverage Big data



Process	Traditional	Evolution	Future?
Data Structure	• Templates in Excel/Access/csv stored • Multiple sources and different clients • Data Manipulation for valuation program & software • Manual collection & aggregation of data	• Client data stored in a common large database format (e.g. SQL) • Programs trained for data manipulation • Single aggregated database to make the analysis easier	• Portals to execute data retrieved directly from client system • AI and ML to manipulates the formats and feeds directly into programs • Real time Analysis with the access • Aggregated database still to store data?
Data Analysis and Indexing	• Excel Modelling and similar tools	• Statistical programs (R, Python) • Excel contributes to validating the tests for training programs • Proprietary software	• Advanced Data Science techniques deployed for assumption setting • Using multiple sources for benchmarks and context setting (e.g. Reports, Macro Economic Indices and Factors)
Optimization and Visualisation	• Standard Graphs and Charts	• Software like Tableau / Power BI to integrate larger data sets with advanced visualization. • Story-telling and Infographics	• Generative AI story-telling and Retrieval-Augmented Generation

Looking from the Employee Benefits Lens

Data Velocity	Data Volume	Data Variety & Veracity
Periodic and static data (Qtrly, Semi-annual, Annual)	Low from Data capture perspective ?	Structured and single source; validated for accuracy

- Traditionally the employee data for New joining, Separates, Retirees, etc. is collected periodically
- With the growing technology, in the future real time API connections to HR information can contribute to the frequency of data collection.

- Traditional Valuations (Grat / Leave) is low. For DBO <10 columns?
- Pension (e.g. PSUs) - <20 columns
- Social Security e.g. EPS1995 (<20 columns)
- For assumption setting time series data can be more extensive

- Collection of data is structured to the set format and sources are determined pre-hand.
- Multi-source data from a industry and geographic perspective shall be a contributor for assumption setting in the near future.
- Considering the technology deployed, true unstructured unlikely for Employee benefits?

Data Analysis and Results

Overview of the Data Analysis Done

Some key findings and trends from example case study analyses performed on high volume data sets.

- Attrition analysis when there is big data and control figures available.
- Salary escalation trend across geographies based on 10 million data.
- Salary escalation duration wise based on big data.
- Medical claim trend post retirement based on 1 lakh+ data.



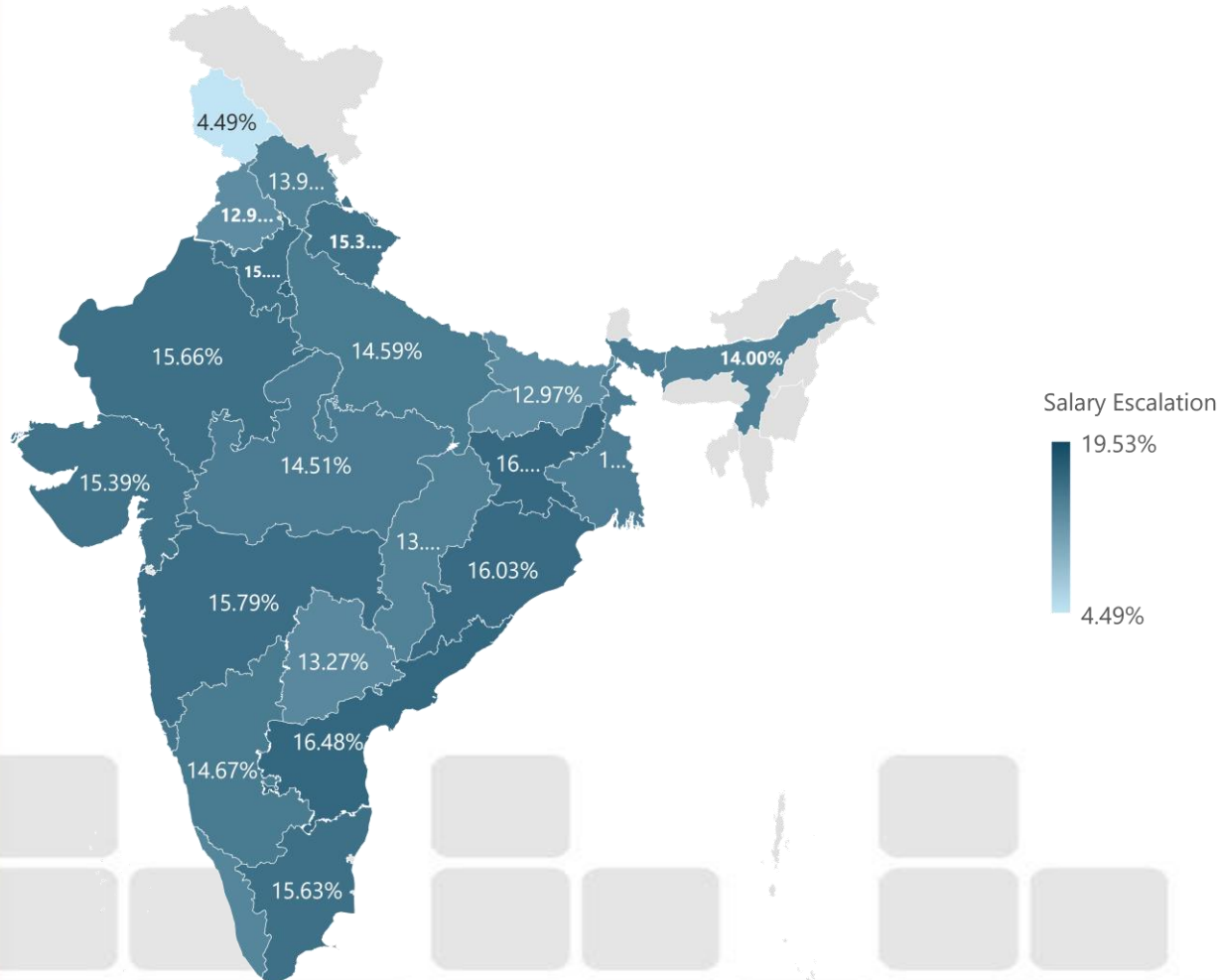
Comprehensive Attrition Analysis



Members as on 31-03-2022	4,12,37,384	
Pensioners including beneficiary as on 31-03-2022	56,49,797	
Members as on 31-03-2024	4,68,89,911	
Pensioners including beneficiary as on 31-03-2024	64,51,746	
Increase in Members	56,52,527	
Increase in pensioners	8,01,949	
Members joined after 31-03-2022	1,92,04,012	
Members as on 31-03-2022	4,12,37,384	
Less transferred to pensioner	8,01,949	
	4,04,35,435	
Members as on 31-03-2024	4,68,89,911	
Less Members joined after 31-03-2022	1,92,04,012	
Members as on 31-03-2024 which were there as on 31-03-2022	2,76,85,899	
Total Depletion	1,27,49,536	
Depletion as a % of members as on 31-03-2017		31.53%
		131.53%
		14.69%

Salary Escalation Observed Across Geographies

Geographic Salary Escalation Distribution

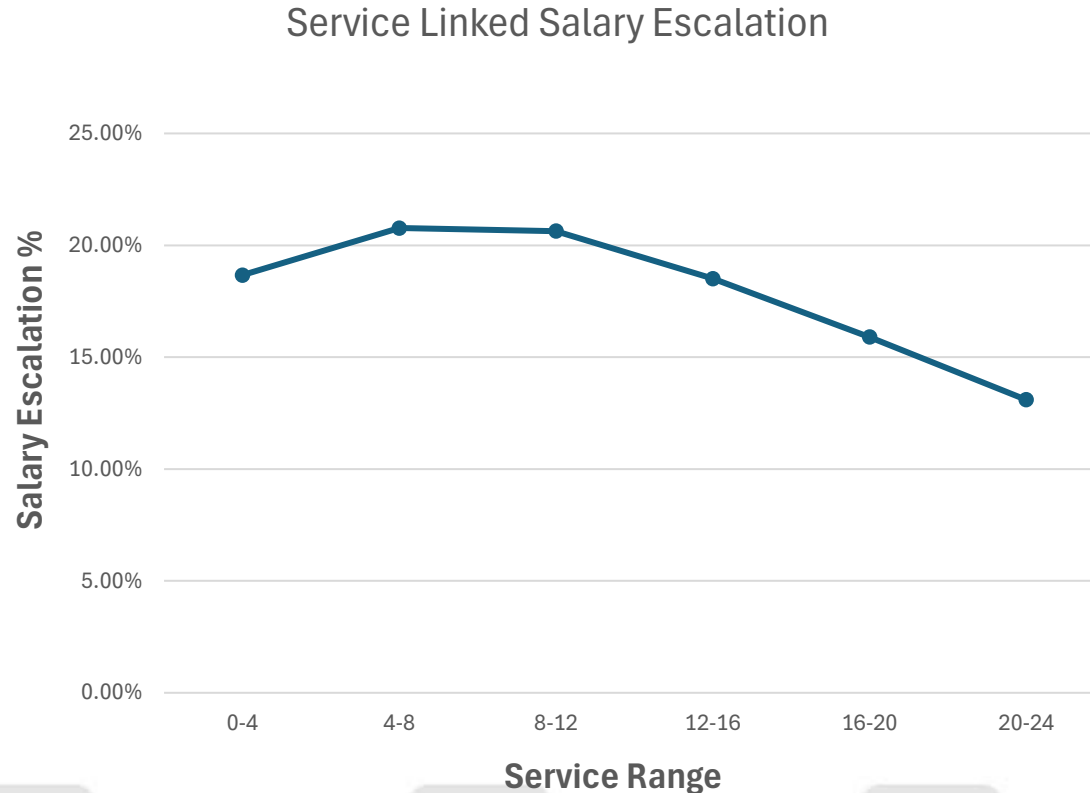


- The outcomes of big data can be highly insightful and influential.
- Here with the huge volume of data available, we are able to observe the increase in the salaries across geographies.
- The metropolitan cities have a higher SE comparatively.
- With the availability of larger, time series and niche data we can possibly capture trends and predict the industry movements.

Salary Escalation Observed Across Geographies

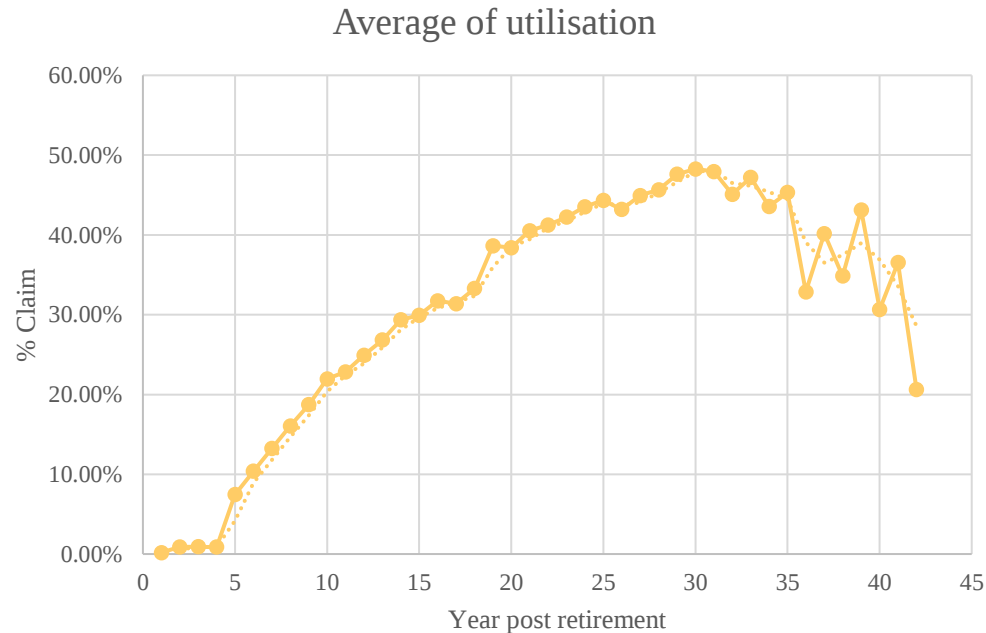
STATES	Salary escalation	STATES	Salary escalation
ANDHRA PRADESH	18.01%	MADURAI	12.90%
BANGALORE	19.43%	MAHARASHTRA (BANDRA)	16.38%
BIHAR	12.97%	MADHYA PRADESH	14.51%
COIMBATORE	13.88%	MEERUT	15.82%
CHATTISGARH	13.98%	NORTH EAST	14.00%
DELHI (NORTH)	16.21%	NAGPUR	13.78%
DELHI (SOUTH)	16.30%	NIZAMABAD	13.27%
GOA	15.19%	ODISHA	16.03%
GULBARGA	14.80%	PUNJAB	14.52%
GUJARAT	16.36%	PUNE	17.21%
GURUGRAM	17.55%	PEENYA (BANGALORE)	19.62%
GUNTUR	14.95%	RAJASTHAN	15.66%
HIMACHAL PRADESH	13.96%	SURAT	15.21%
HARYANA (FARIDABAD)	15.50%	TAMBARAM	17.21%
JHARKHAND	16.25%	THANE	15.64%
J&K and Ladakh	4.49%	TAMIL NADU (CHENNAI)	20.10%
JALPAIGURI	14.05%	UTTARAKHAND	15.37%
KANDIVALI	15.92%	UTTAR PRADESH	13.36%
KARNATAKA (MANGALORE)	14.54%	VADODARA	14.60%
KERALA	13.58%	WEST BENGAL (KOLKATA)	14.60%
LUDHIANA	11.28%	AVERAGE	15.10%

Capturing Movement of Salary Escalation



- Using the data of over 20 million **employees for a time span of 24 years with moving data**, we were able to observe the movement of increase in salary based on the service. Time to use escalation curve rather than fixed salary increase %
- The initial years of service line with the increase in percentage and as later captures, the percentage decreases as the quantum of the salary figures is increased.

Capturing Medical claim trend post retirement



- Using the data of over 1 Lakh **Post retirement members of movement of claims in past 20 years.**
- The rise in the claims is observed in the initial years post retirement and post 30 years reflecting the reducing life expectancy.

Looking into Future trends

Harnessing the Future : Final Thoughts on Big Data

- Leveraging the Big Data is crucial for achieving operational efficiency and enhancing predictive capabilities.
- Bridging the Skill Gap and investing in the technology are the next stepping stones of this evolution.
- The primary challenges to address are the Data Governance, Privacy, Security and integration.
- The AI and Machine Learning Integration along with quantum computing are the upcoming reins to harness the big data.



- Increased Accuracy and stabilization of assumption analysis.
- Predicting future trends of attrition and new employment.
- Tailoring products adhering to unique customer needs and preferences.
- Enhance the modeling for infrequent and landmark events in the industry.
- Increase the accuracy of provisioning and thus, consulting.
- Generative AI and Big Data.

Generative AI is a part of Big Data Analytics that has good potential in Employee Benefits Actuarial Valuation.

The AI algorithms can generate data-driven insights by processing large volumes of data.

The analytical capabilities of these algorithms can prove to be highly valuable in actuarial assumptions and valuation models.

Thank You!!