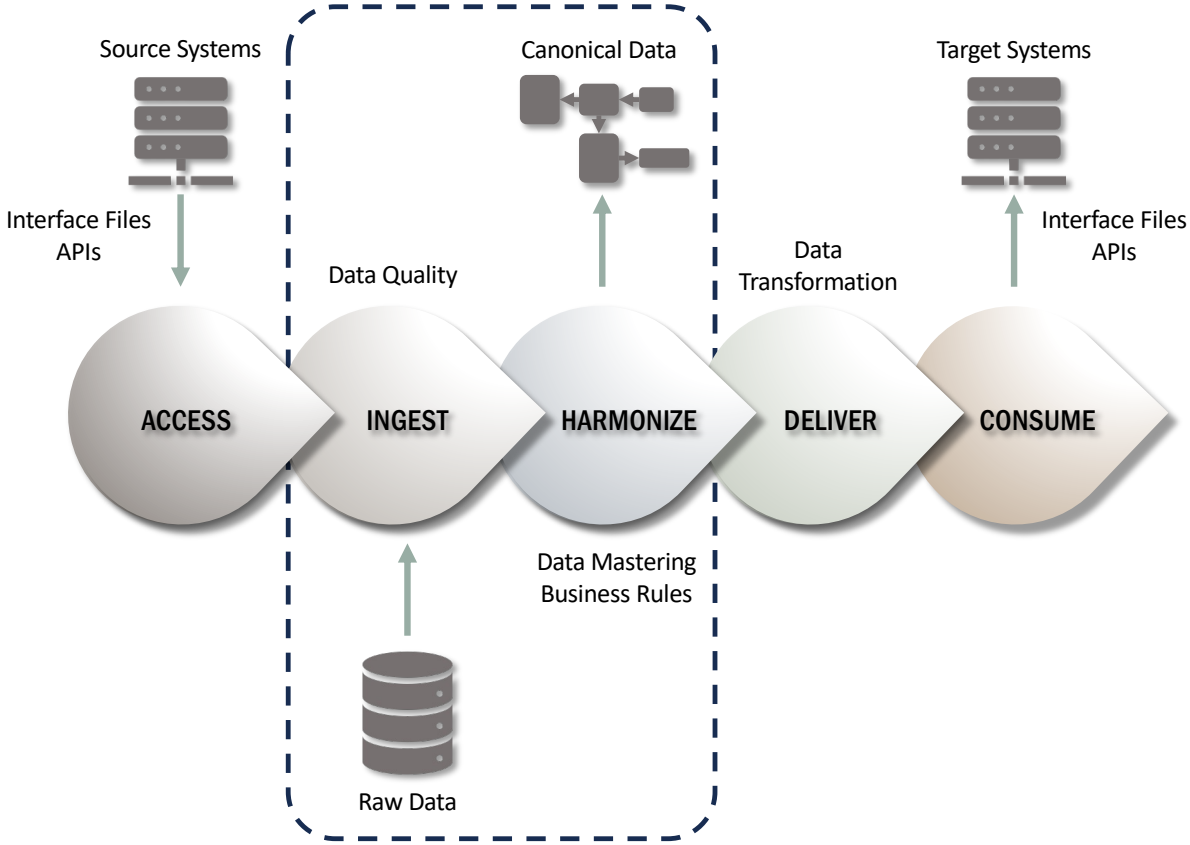
STOR 320: Introduction to Data Science

Gerhard Ungerer - MarketWatch

Overview – Gerhard Ungerer

- Chief Technology Officer at Random Bit LLC
- Bachelor of Science in Computer Science & Applied Mathematics from Stellenbosch University, 1993
- 32 years in software development and data management across Defense, Finance, Supply Chain, and Health & Human Services domains
- Specialize in Enterprise Architecture, Big Data Integration and Cloud Transformation
- UNC MADS program since Fall 2024
- Husband and father of 4 daughters
- Pilot

Data Management Pipeline



- Need for systematic processing and analysis of data is increasing exponentially.
- Not only in the real-time analytics environments, but also in the less exciting business operations areas.
- Tremendous value in machine learning, predictive analytics, and data insights.
- Also, in enterprise data integration and data quality remediation.

Use-cases for Statistics - VBP

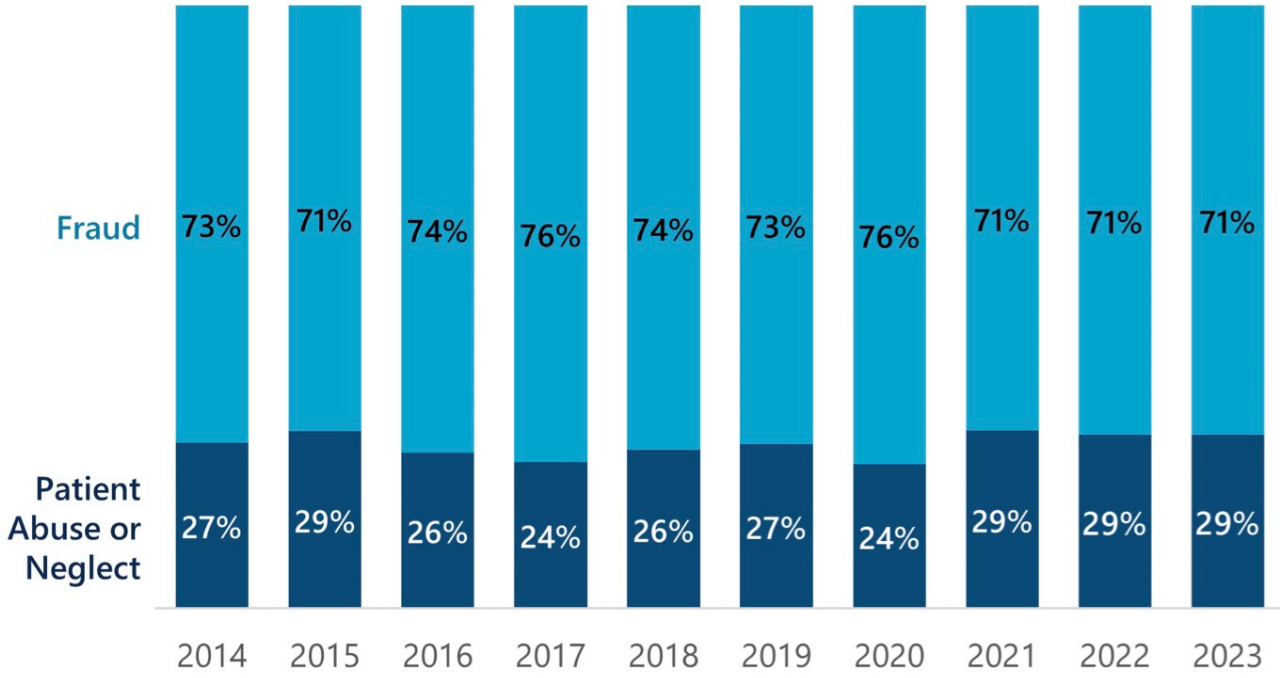
Table 1. Claimed-Based Outcome and Payment Measures in Chartbook

Measure Outcome	30-day risk-standardized mortality measures	30-day risk-standardized readmission measures	90-day risk-standardized complications measure	30-day excess days in acute care measures	Payment measures
AMI	♦	♦		♦	♦
CABG	♦	♦			
COPD	♦	♦			
HF	♦	♦		♦	♦
THA/TKA		♦	♦		♦
PN	♦	♦		♦	♦
Stroke	♦				
HW		♦			

https://www.cms.gov/medicare/quality/initiatives/hospital-quality-initiative/outcome-and-payment-measures

- Value-based payments
- Get paid based on outcomes in healthcare
- Need to decide which treatments have the best lasting effects
- Example: paying for physiotherapy and dietician services on heart patients have good outcomes so pay higher rates to keep patients engaged

Use-cases for Statistics - PI



Source: OIG analysis of Quarterly and Annual Statistical Reports for FYs 2014 through 2023.

https://oig.hhs.gov/documents/evaluation/9821/OEI-09-24-00200.pdf

- Program Integrity (Fraud, Waste & Abuse)
- The Improper Payments Information Act (IPIA) of 2002
- Look at patterns for medical services and see if that is within norm.
  - Flu cases in winter normal
  - High spike in spring might not be
  - Statistically abnormal referrals to a cardiologist

Use-cases for Statistics - Capitation

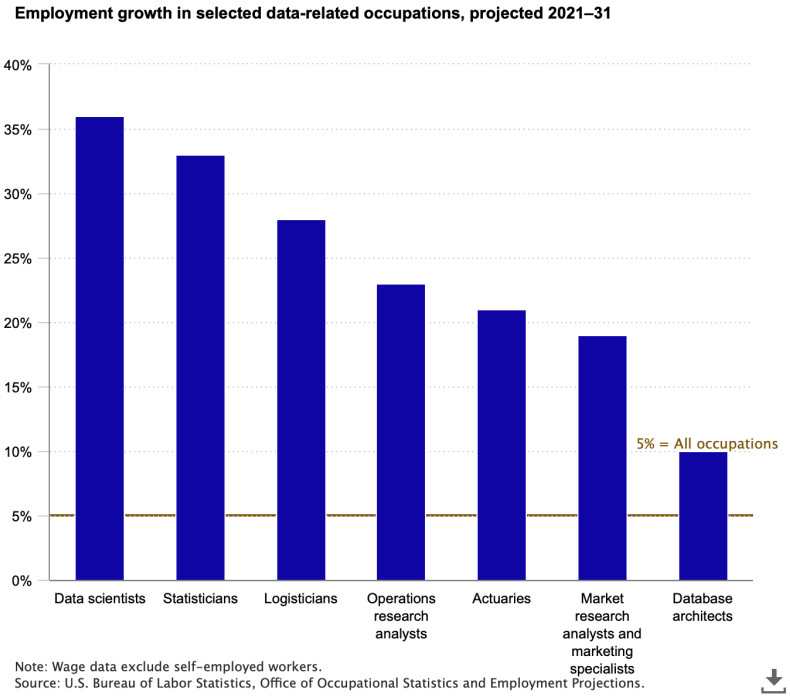
Table 1: Prospective LOB Adjustment Development

Financial Metric	Formula	Ind	SG	LG	Total
(A) Base Year Member Months	=	165,000	115,000	150,000	430,000
(B) Base Year Capitation Dollars	=	$45,000,000	$25,000,000	$40,000,000	$110,000,000
(C) Base Year Capitation PMPM	= (B) / (A)	$272.73	$217.39	$266.67	$255.81
(D) Base Year Average Age Adjusted Concurrent Risk Scores	=	1.100	1.300	1.250	1.206
(E) Relative Risk Adjusted Capitation Rate	= (C) x (D) / (D) Total Col	$248.79	$234.37	$276.44	$254.58
(F) Retrospective Revenue Neutral Risk Adjusted Capitation Rate	= (E) x (C) Total Col / (E) Total Col	$250.00	$235.51	$277.78	$255.81
(G) Prospective LOB Adjustment	= (F) / (C)	97.7%	108.3%	104.2%	100.0%

https://axenehp.com/intelligent-capitation-rates-strategically-aligning-risk-adjustment-costs/

- Capitation rate calculation
- Some groups are paid by visit
- Other groups get paid a fixed amount to provide all services for a group of people
  - 10,000 people
  - $300 per person per month
  - $3,000,000 per month regardless of how many visits
- Depends on how old people are and whether in good health or not.

Data Science Market Outlook



US Bureau of Labor and Statistics shows Data Scientists as the highest projected data-related occupation.

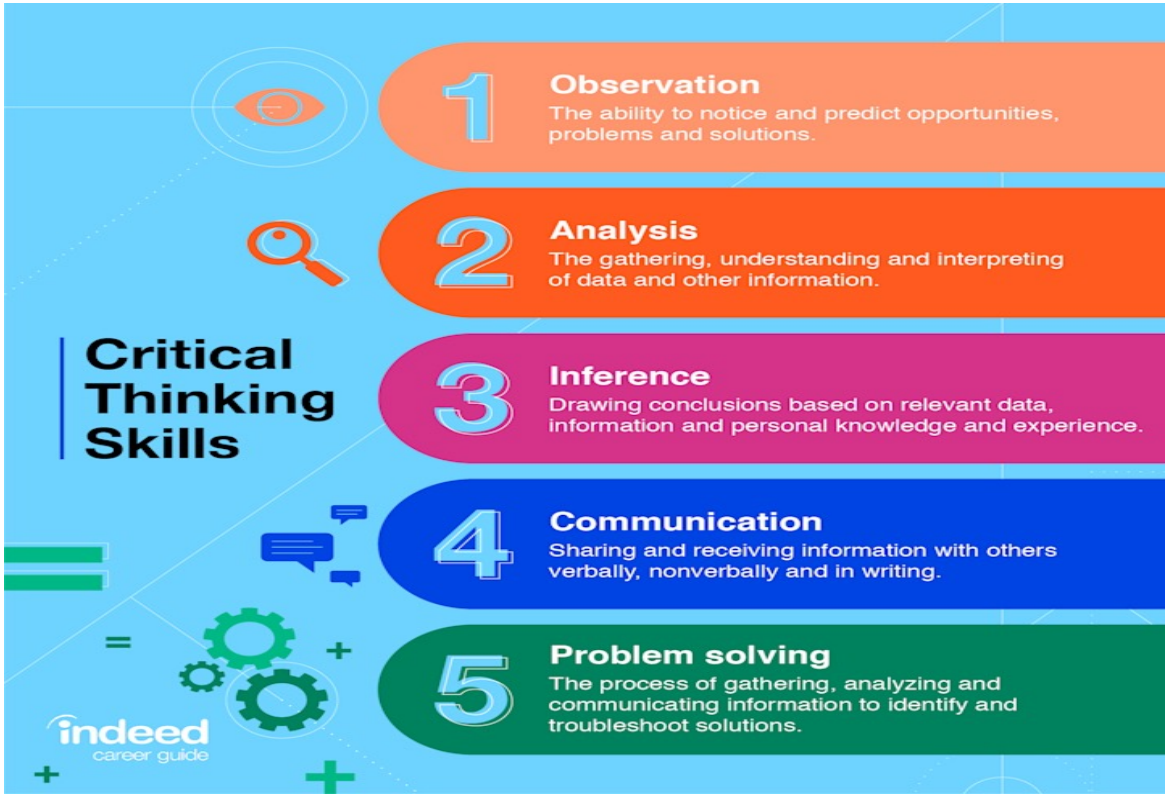
https://www.bls.gov/careeroutlook/2023/data-on-display/data-occupations.htm

Role	Entry-Level	Mid-Level	Senior-Level
Data Scientist	$95,000–$115,000	$130,000–$155,000	$160,000–$195,000
ML Engineer	$100,000–$125,000	$140,000–$170,000	$180,000–$210,000
Data Engineer	$90,000–$110,000	$120,000–$145,000	$150,000–$185,000
AI Researcher	$120,000–$140,000	$150,000–$180,000	$190,000–$230,000
BI Analyst	$80,000–$100,000	$110,000–$135,000	$140,000–$170,000
Data Analyst	$75,000–$95,000	$100,000–$125,000	$130,000–$155,000
CV Engineer	$100,000–$125,000	$135,000–$165,000	$175,000–$215,000
NLP Specialist	$105,000–$130,000	$140,000–$170,000	$180,000–$220,000
Data Architect	$125,000–$150,000	$160,000–$190,000	$200,000–$240,000
Deep Learning Engineer	$115,000–$140,000	$145,000–$175,000	$185,000–$230,000
Big Data Analyst	$90,000–$110,000	$120,000–$145,000	$150,000–$180,000
AI Product Manager	$105,000–$130,000	$140,000–$170,000	$180,000–$220,000
Robotics Engineer	$100,000–$125,000	$135,000–$165,000	$175,000–$215,000
Computational Biologist	$95,000–$115,000	$125,000–$150,000	$160,000–$200,000
Governance Specialist	$90,000–$110,000	$115,000–$140,000	$150,000–$180,000
Cloud Data Engineer	$100,000–$120,000	$130,000–$160,000	$170,000–$210,000
AI Consultant	$95,000–$120,000	$135,000–$165,000	$180,000–$220,000

Very lucrative field to be in.

https://www.linkedin.com/pulse/data-science-ai-salary-report-2025-what-you-need-know-efcif/

Critical thinking



- Builds transferrable experience
- Refines problem-solving capability
- Preparing you for a fast-paced career

https://www.indeed.com/career-advice/career-development/how-to-improve-critical-thinking

What employers are looking for

Aptitude

- Data management skills
- Machine Learning
- Data visualization
- Data mining
- AWS / Azure
- Python
- Communication skills
  - Presentation
  - Written
  - Verbal

Attitude

- INTEGRITY
- Willing to learn
- Great work ethic
- Invest in your skills
- Team member
- No ego, No drama
- Be interesting

Questions and Answers

gerhard@randombitcorp.com
gungerer@unc.edu
www.randombitcorp.com

Terminology

Term	Description
Data Pipeline	Set of steps for processing data
Canonical Model	A data model understood by everyone in the organization
Use-case	A scenario or example
Real-time analytics	Analysis and statistics that are built dynamically. Instagram followers, likes, etc.