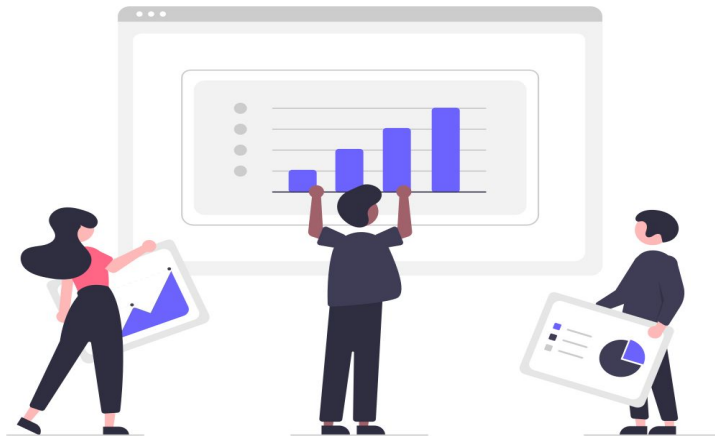


# Healthy Data



By Jason Scott, Tyler Koon, Ankush Garg, and Daniel Brown

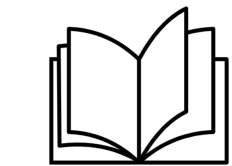
# Who are we?



## Data Engineering Team supporting Healthcare Data Scientists

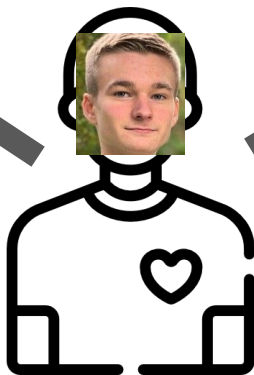
- Explore the use of three different NoSQL Databases
- Support new hospital price strategy, insurance, and data management
- Uncover hidden relationships and patterns in complex datasets using graph algorithms

# MongoDB



Patient History

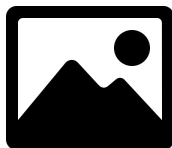
Patient



Doctors Notes



Prescriptions



Imaging Reports



Lab Results

## Patient Medical Records System

patient_histories	
id	string pk
historyDate	date
description	string
patientId	string fk

doctor_notes	
id	string pk
noteDate	date
noteText	string
patientId	string fk

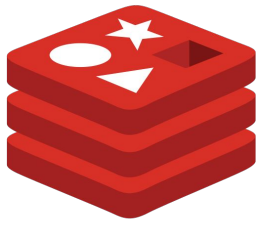
prescriptions	
id	string pk
patientId	string fk
prescriptionDate	date
medication	string
dosage	string

lab_results	
id	string pk
patientId	string fk
resultDate	date
testType	string
result	string

imaging_reports	
id	string pk
patientId	string fk
reportDate	date
imageType	string
findings	string

patients	
id	string pk
name	string
dateOfBirth	date
gender	string

Redis



# United States Healthcare, Price Transparency, Neo4j

## Key Questions

- Who are the most connected insurance providers in the state of illinois?
- Which insurance companies are similar to each other?
- Are there patterns to how hospitals and insurances work together?



# Dataset



# Chargemaster files



{.json}



# Hospital table and Rates table

rate	
hospital_id <sup>o</sup>	varchar(6)
row_id <sup>o</sup>	bigint unsigned
line_type	varchar(40)
description	varchar(2000)
rev_code	varchar(4)
local_code	varchar(40)
code	varchar(50)
ms_drg	varchar(3)
apr_drg	varchar(5)
eapg	varchar(5)
hcpcs_cpt	varchar(5)
modifiers	varchar(14)
alt_hcpcs_cpt	varchar(5)
thru	varchar(10)
apc	varchar(4)
icd	varchar(15)
ndc	varchar(13)
drug_hcpcs_multiplier	int
drug_quantity	varchar(25)
drug_unit_of_measuremen...	varchar(25)
drug_type_of_measuremen...	enum
billing_class	enum
setting	enum
rate_category	enum
payer_name	varchar(1000)
plan_name	varchar(1000)
standard_charge	decimal(14,2)
standard_charge_percent	decimal(6,2)
contracting_method	enum
additional_generic_note...	text
additional_payer_specif...	text

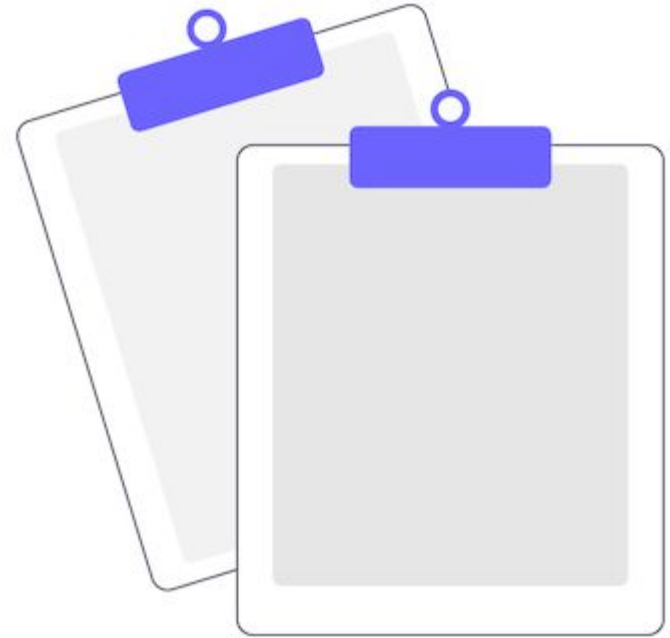
>

hospital	
id <sup>o</sup>	varchar(6)
ein	varchar(11)
name	varchar(100)
alt_name	varchar(100)
system_name	varchar(100)
addr	varchar(100)
city	varchar(50)
state	char(2)
zip	char(5)
phone	varchar(14)
urban_rural	char(1)
category	varchar(50)
control_type	varchar(50)
medicare_termination_st...	varchar(100)
last_updated	date
file_name	varchar(1000)
mrf_url	varchar(2048)
permalink	varchar(2048)
transparency_page	varchar(2048)
additional_notes	text

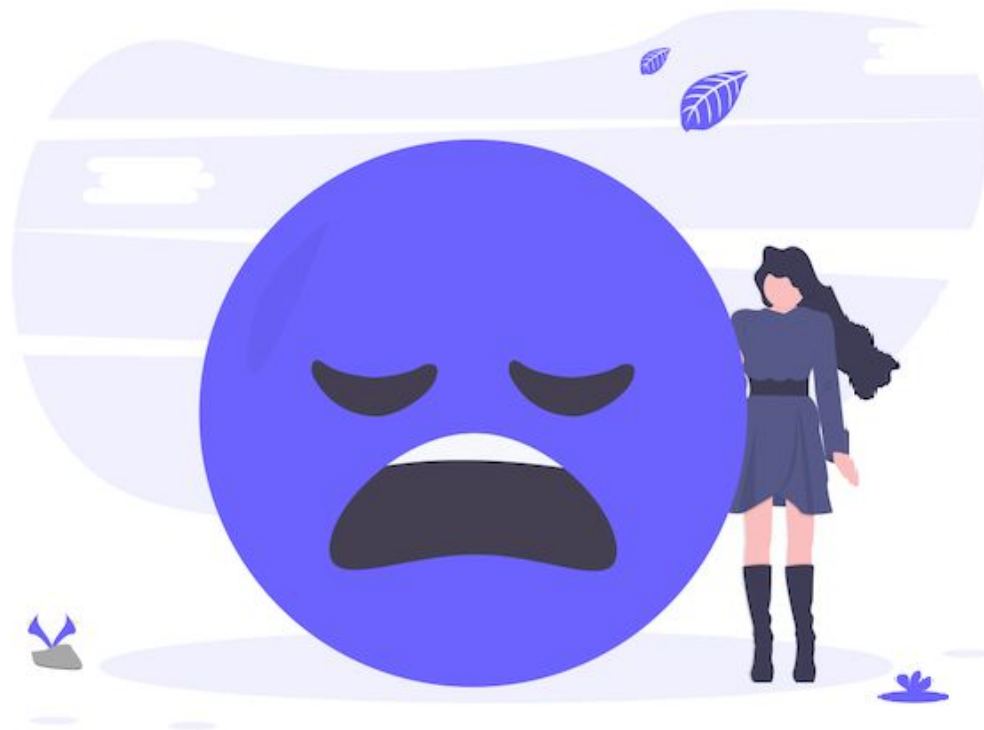


# Assumptions

- Pricing is correct
- Dataset is complete
- Insurance pays what is asked
- Data refresh



# Problems with the dataset



# Approach for Data Processing

## Rate Type

Focus on the rates charged for **entire operations**

---

12 Hospitals

---

## Procedure

Only consider **knee and hip replacement surgeries**

---

55 Insurance Plans

---

## Region

Restrict our focus to the state of **Illinois**

---

22 Insurance Providers

---

# Page Rank and Degree Centrality

## Objective

Identify most prominent insurance providers in our area

### Most Well Connected

Provider	PageRank	Degree Centrality
Blue Cross	0.558781	209.0
UnitedHealthcare	0.366083	113.0
Other	0.364395	54.0
Cigna	0.337618	92.0
Humana	0.290281	72.0

### Least Well Connected

Provider	PageRank	Degree Centrality
Wellcare	0.156711	3.0
Caterpillar	0.156324	3.0
Ascension	0.153355	2.0
Amita Health	0.153355	2.0
Clear Spring	0.153355	4.0

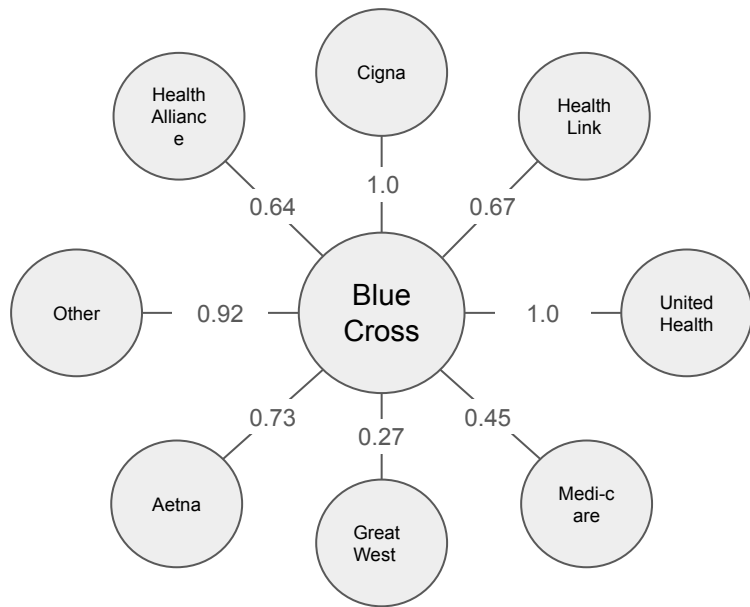
## Key Takeaway

Blue Cross is most prominent insurer in the community

# Node Similarity

## Objective

Better understanding of similarities in price and coverage between insurance providers



## Pricing Consistency

Maintaining consistent and predictable rates for standard procedures and operations, across locations and providers

## Key Takeaway

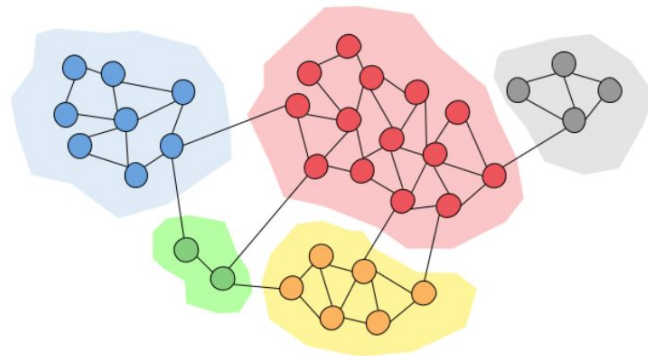
This model ensures prices remain consistent with local hospitals and between similar plans

# Louvain (Community Detection)

## Objective

Identify clusters of similar hospitals and insurances within our graph network

	Type	Name	Community
0	Hospital	PRESENCE MERCY MEDICAL CENTER	12
1	Hospital	PRESENCE SAINT JOSEPH HOSPITAL - ELGIN	12
2	Hospital	PRESENCE SAINT JOSEPH MEDICAL CENTER	12
3	Hospital	ST MARY & ELIZABETH MED CTR-CLAREMONT CAMPUS	12
4	Insurance	Ascension	12
5	Insurance	Blue Cross	12
6	Insurance	Caterpillar	12
7	Insurance	Cigna	12
8	Insurance	Coventry	12
9	Insurance	Golden Rule/UnitedHealthcare	12
10	Insurance	Great West	12
11	Insurance	Health Alliance	12
12	Insurance	Healthspring	12
13	Insurance	Self Pay	12
14	Insurance	Swedish Covenant	12
15	Insurance	Blue Cross	12

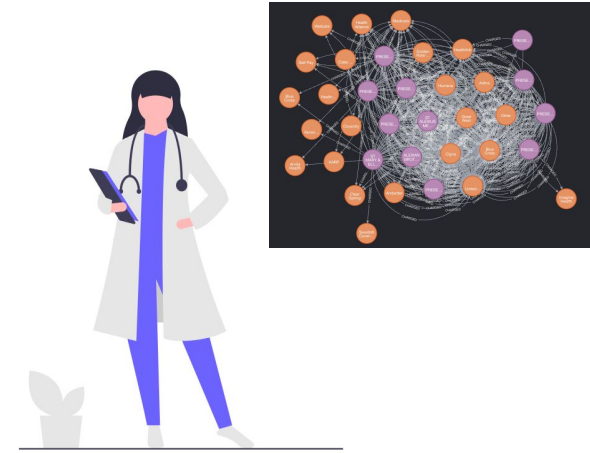


## Key Takeaway

Hospitals and insurance providers are grouped in 3 distinct clusters

# Key Takeaways

- **PageRank:** Blue Cross is the most prominent insurance provider in Illinois, followed by United Healthcare
- **Node Similarity:** Price variation exists for the same procedures across insurance providers in the state
- **Louvain:** 3 distinct clusters of hospitals and insurances exist across the state



# Opportunities for Future Research

- Geographic expansion
- Include additional DRG codes
- Incorporate revenue code data
- Augment with secondary datasets

