

Smart Doctor & Hospital Finder: Insurance Coverage & Specialty-Based Search

DSC 202 Group 11:

Arya Gaikwad (A69035152)

Devana Perupurayil (A69034326)

Harini Gurusankar (A16264685)

Shruti Sawant (A69033524)

Table of Contents

3	Introduction
6	Data Sources
12	Methodology
23	Demonstration

Introduction

- Problem Statement
- Objectives

Problem Statement

In California, Medi-Cal is the health insurance option for low-income individuals and their dependents, and this insurance network is financed by the state and the federal government. Healthcare providers, insurance companies, and government agencies often struggle to manage large-scale datasets that span multiple domains, including hospitals, providers, insurance plans, and specializations. This fragmented data leads to inefficiencies in provider management, and difficulty for patients to access the healthcare information needed.



Objectives

- Create an application to combine Medi-Cal information such as providers, hospitals and their ratings, insurance plans and taxonomy data to provide users a centralized source of information to easily access a variety of healthcare information.
- By leveraging relational (PostgreSQL) and graph (Neo4j) databases, the application will facilitate efficient data retrieval, improve network transparency, and enhance healthcare decision-making for all stakeholders.

Data Sources

- Data Sources
- Structure of Data

Data Source: Provider Information

Medi-Cal Managed Care and Provider Listing: 3,268,559 rows and 33 columns

Purpose: A dataset of providers that are under Medi-Cal and their personal information such as specialization, license type, if they see children or not. It also shows each hospital and the address of the hospital including latitude and longitude.

Link: <https://data.ca.gov/dataset/medi-cal-managed-care-provider-listing>

ManagedCarePlan ⬆️	SubNetwork ⬆️	PlanCode ⬆️	RecordType ⬆️	NPI ⬆️	FacilityName ⬆️	LastName ⬆️	FirstName ⬆️	Taxonomy ⬆️	MCNAProviderGroup ⬆️
ANTHEM BLUE CROSS PARTNERSHIP PLAN	ANTHEM BLUE CROSS PARTNERSHIP PLAN	340	Provider	1811955438	Ucsf Sfgh Medical Group	Gupta	Nalin	207T00000X	Adult Core Specialist
ANTHEM BLUE CROSS PARTNERSHIP PLAN	ANTHEM BLUE CROSS PARTNERSHIP PLAN	340	Provider	1811967920	Santa Clara Valley Medical Center	Nguyen	Michelle	207R00000X	Non-ANC Provider Type
ANTHEM BLUE CROSS PARTNERSHIP PLAN	ANTHEM BLUE CROSS PARTNERSHIP PLAN	340	Provider	1811970973	Scvmc Physician Services	Dai Biller	Jenny	207V00000X	OB/GYN Primary and Specialty Care
ANTHEM BLUE CROSS PARTNERSHIP PLAN	ANTHEM BLUE CROSS PARTNERSHIP PLAN	340	Provider	1811970973	Santa Clara Valley Medical Center	Dai Biller	Jenny	207V00000X	OB/GYN Primary and Specialty Care

Data Source: Hospital Information

Current California Healthcare Facility Listing: 10,908 rows and 18 columns

Purpose: A dataset of California healthcare facilities with information about hospitals, locations, and hospital description (if it has an emergency room, the number of beds).

Link:

<https://data.ca.gov/dataset/licensed-healthcare-facility-listing/resource/9e038458-a91e-40ef-9c78-07f1ff6fb839>

_id ↕	OSHPD_ID ↕	FACILITY_NAME ↕	LICENSE_NUM ↕	FACILITY_LEVEL_DESC ↕	DBA_ADDRESS1 ↕	DBA_CITY ↕	DBA_ZIP_CODE ↕	COUNTY_CODE ↕	COUNTY_NAME ↕	ER_SERVI
1	106010042	JOHN GEORGE PSYCHIATRIC HOSPITAL	140000046	Distinct Part Facility	2060 FAIRMONT DRIVE	SAN LEANDRO	94578	01	Alameda	None
2	106010735	ALAMEDA HOSPITAL	140000002	Parent Facility	2070 CLINTON AVE	ALAMEDA	94501	01	Alameda	Emergency
3	106010739	ALTA BATES SUMMIT MEDICAL CENTER-ALTA BATES CAMPUS	140000004	Parent Facility	2450 ASHBY AVENUE	BERKELEY	94705	01	Alameda	Emergency
4	106010776	UCSF BENIOFF CHILDREN'S HOSPITAL OAKLAND	140000015	Parent Facility	747 52ND STREET	OAKLAND	94609	01	Alameda	Emergency

Data Source: Hospital Ratings Information

California Hospital Performance Ratings: 25975 rows and 13 columns

Purpose: This dataset provides the risk-adjusted ratings of hospitals in California for certain events and the number of adverse events of each type.

Link: <https://catalog.data.gov/dataset/california-hospital-performance-ratings-c3975>

OSHPDID	system	Type of Report	Performance Measure	# of Adverse Events	# of Cases	Risk-adjusted Rate	Hospital Ratings	Longitude	Latitude
106010735	Alameda Health System	IMI	Pneumonia	2	76	3	As Expected	-122.253	37.76266
106010735	Alameda Health System	IMI	Heart Failure	2	111	2.1	As Expected	-122.253	37.76266
106010735	Alameda Health System	IMI	GI Hemorrhage	5	83	4.6	As Expected	-122.253	37.76266
106010735	Alameda Health System	IMI	PCI					-122.253	37.76266
106010735	Alameda Health System	IMI	Acute Stroke Subarachnoid					-122.253	37.76266
106010735	Alameda Health System	IMI	Acute Stroke Hemorrhagic	5	9	48.9	Worse	-122.253	37.76266
106010735	Alameda Health System	IMI	AMI	3	17	16.1	As Expected	-122.253	37.76266

Data Source: Taxonomy Codes

Medicare Provider and Supplier Taxonomy Crosswalk: 561 rows and 4 columns

Purpose: Has a list of unique taxonomy codes and their specialization (description). Used to connect to providers.

Link:

<https://data.cms.gov/provider-characteristics/medicare-provider-supplier-enrollment/medicare-provider-and-supplier-taxonomy-crosswalk/data>

MEDICARE SPECIALTY CODE	MEDICARE PROVIDER/SUPPLIER TYPE DESCRIPTION	PROVIDER TAXONOMY CODE	PROVIDER TAXONOMY DESCRIPTION: TYPE...
01	Physician/General Practice	208D00000X	Allopathic & Osteopathic Physicians/Gene...
02	Physician/General Surgery	208600000X	Allopathic & Osteopathic Physicians/Surg...
02	Physician/General Surgery	2086H0002X	Allopathic & Osteopathic Physicians/Surg...
02	Physician/General Surgery	2086S0120X	Allopathic & Osteopathic Physicians/Surg...
02	Physician/General Surgery	2086S0122X	Allopathic & Osteopathic Physicians/Surg...
02	Physician/General Surgery	2086S0105X	Allopathic & Osteopathic Physicians/Surg...
02	Physician/General Surgery	2086S0102X	Allopathic & Osteopathic Physicians/Surg...
02	Physician/General Surgery	2086X0206X	Allopathic & Osteopathic Physicians/Surg...

Data Source: Insurance Plan Information

CY 2023 Two-Plan Model Rates: 644 rows and 10 columns

Purpose: This dataset gives cost information for health plans in 2023 based on their category of aid and gives a range of cost (lower bound, midpoint, upper bound).

Link:

<https://data.ca.gov/dataset/medi-cal-managed-care-capitation-rates-two-plan-model/resource/c7036f04-6203-4de4-8fe1-025350978470>

_id ↕	Rating Period ↕	Calendar Year ↕	Model ↕	County ↕	Health Plan Name ↕	Category of Aid ↕	Lower Bound ↕	Midpoint ↕	Upper Bound ↕
1	1/23-12/23	2023	Two Plan	Alameda	Alameda Alliance for Health	Child - UIS Federal	\$31.98	\$33.88	\$35.91
2	1/23-12/23	2023	Two Plan	Alameda	Alameda Alliance for Health	Adult - UIS Federal	\$159.56	\$165.24	\$171.19
3	1/23-12/23	2023	Two Plan	Alameda	Alameda Alliance for Health	ACA Optional Expansion - UIS Federal	\$262.52	\$271.24	\$280.36
4	1/23-12/23	2023	Two Plan	Alameda	Alameda Alliance for Health	SPD - UIS Federal	\$777.28	\$801.25	\$826.24
5	1/23-12/23	2023	Two Plan	Alameda	Alameda Alliance for Health	SPD/Full-Dual - UIS Federal	\$121.65	\$126.08	\$130.73

Methodology

- Technical Steps
- Application Structure

Data Pre-Processing

Used Python programming language to clean the obtained data sources.

- Hospitals Data Processing
 - Extracts relevant columns such as OSHPD_ID, FacilityName, FacilityType, Address, ZIP, County, Total_number_of_beds and so on.
 - Removes duplicates based on the facility details and ensures unique OSHPD_ID.
- Provider Data Processing
 - Merges the providers.csv file with hospital_data.csv using the facility name and associate providers with hospitals using OSHPD_ID.
 - Cleans missing values and applies standardization for fields like SeesChildren and Telehealth using a replacement dictionary.
- Taxonomy Data Preprocessing
 - Extracts taxonomy details, like the Taxonomy_Code, Medicare_provider_description, Provider_Taxonomy_Description, and Medicare_Speciality_Code.
 - Removes duplicates to ensure unique Taxonomy_Codes.
- Hospitals_Ratings Data Processing
 - Updated the column names to be consistent across other datasets and retain required columns.
- Insurance Data Processing
 - Dropped columns which were not up to date and ensures there were no duplicate rows.

Updated Provider Network

- Removed FacilityName and address from provider data and added OSHPD_ID as a foreign key to reference hospital dataset
- Changed values for SeesChildren and Telehealth to Both or No, instead of B and N

licensure_type	primarycare	specialist	seeschildren	telehealth	bhindicator	oshp_id
CSW	N	Y	Both	No		306370021
OTH	N	Y	Both	No		306370856
OTH	N	Y	Both	No		106190782
NPA	Y	N	Both	Both		306374450
MD	Y	N	Both	No		306374553
MD	N	Y	Both	No		306374260
NPA	Y	N	Both	No		306370856
MD	Y	Y	Both	No		306370856
NPA	Y	N	Both	No		306370024
OTH	N	Y	Both	No		306370024

providernetwork.csv

OSHPD_ID	FacilityName
306370021	SAN YSIDRO HEALTH MOUNTAIN HEALTH FAMILY MEDICINE
306370856	VISTA COMMUNITY CLINIC
106190782	TARZANA TREATMENT CT
306374450	NESTOR COMMUNITY HEALTH CENTER
306374553	SAN DIEGO FAMILY CARE
306374260	GROSSMONT SPRING VALLEY FAMILY HLTH CTRS INC
306370024	CITY HEIGHTS FAMILY HEALTH CENTERS INC
306370507	SAN YSIDRO HEALTH MATERNAL AND CHILD HEALTH CTR
306374018	ST VINCENT DE PAUL VILLAGE FAMILY HEALTH CENTER
306374160	NORTH PARK FAMILY HEALTH CENTERS

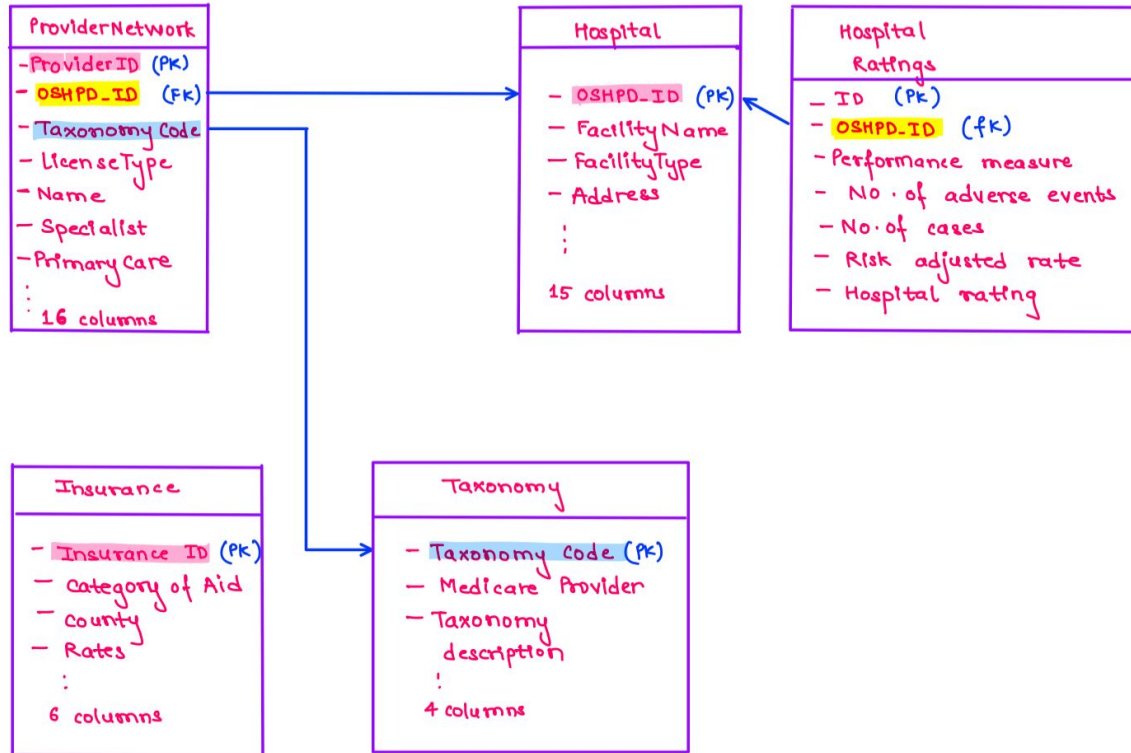
final_hospitals_data.csv

Updated Taxonomy Codes

- Only selected four columns for taxonomy and removed codes that were not present in provider network data

PROVIDER TAXONOMY CODE	MEDICARE PROVIDER/SUPPLIER TYPE DESCRIPTION	PROVIDER TAXONOMY DESCRIPTION: TYPE, CLASSIFICATION, SPECIALIZATION	MEDICARE SPECIALTY CODE
208D00000X	Physician/General Practice	Allopathic & Osteopathic Physicians/General Practice	1
208600000X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery	2
2086H0002X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Hospice and Palliative Medicine	2
2086S0120X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Pediatric Surgery	2
2086S0122X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Plastic and Reconstructive Surgery	2
2086S0105X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Surgery of the Hand	2
2086S0102X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Surgical Critical Care	2
2086X0206X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Surgical Oncology	2
2086S0127X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Trauma Surgery	2
2086S0129X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Vascular Surgery	2
208G00000X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Thoracic Surgery (Cardiothoracic Vascular Surgery)	2
204F00000X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Transplant Surgery	2
208C00000X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Colon & Rectal Surgery	2
207T00000X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Neurological Surgery	2
204E00000X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Oral & Maxillofacial Surgery	2
207X00000X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Orthopedic Surgery	2
207XS0114X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Orthopedic Surgery/Adult Reconstructive Orthopedic Surgery	2
207XX0004X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Orthopedic Surgery/Foot and Ankle Surgery	2
207XS0106X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Orthopedic Surgery/Hand Surgery	2
207XS0117X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Orthopedic Surgery/Orthopedic Surgery of the Spine	2
207XX0801X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Orthopedic Surgery/Orthopedic Trauma	2

Relations



Creating Tables in Postgres

Insurance_data

```
CREATE TABLE insurance_data (
```

```
  ID INT Primary Key,
```

```
  County VARCHAR(100),
```

```
  Health_Plan_Name VARCHAR(255),
```

```
  Category_of_Aid VARCHAR(255),
```

```
  Lower_Bound VARCHAR(20),
```

```
  Midpoint VARCHAR(20),
```

```
  Upper_Bound VARCHAR(20)
```

```
);
```

id	county	health_plan_name	category_of_aid	lower_bound
1	Alameda	Alameda Alliance for Health	Child - UIS Federal	\$31.98
2	Alameda	Alameda Alliance for Health	Adult - UIS Federal	\$159.56
3	Alameda	Alameda Alliance for Health	ACA Optional Expansion - ...	\$262.52
4	Alameda	Alameda Alliance for Health	SPD - UIS Federal	\$777.28
5	Alameda	Alameda Alliance for Health	SPD/Full-Dual - UIS Fede...	\$121.65

taxonomy_data







```
CREATE TABLE taxonomy_data (  
    PROVIDER_TAXONOMY_CODE TEXT Primary Key,  
    MEDICARE_PROVIDER_SUPPLIER_TYPE_DESCRIPTION VARCHAR(255),  
    PROVIDER_TAXONOMY_DESCRIPTION VARCHAR(255),  
    MEDICARE_SPECIALTY_CODE VARCHAR(255)  
);
```

provider_taxonomy_code	medicare_provider_supplier_type_description	provider_taxonomy_description
208D00000X	Physician/General Practice	Allopathic & Osteopathic Physicians/General Practice
208600000X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery
2086H0002X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Hospice...
2086S0120X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Pediatr...
2086S0122X	Physician/General Surgery	Allopathic & Osteopathic Physicians/Surgery/Plastic...

Hospitals

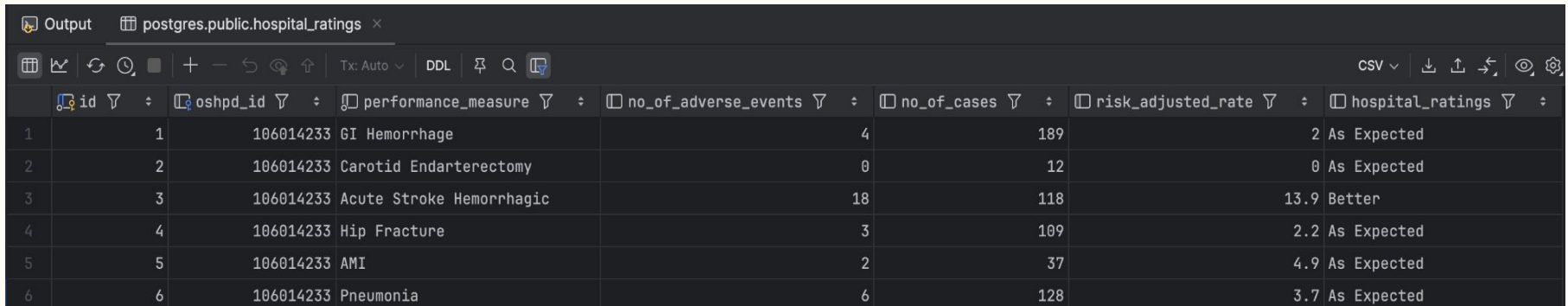
```
CREATE TABLE Hospitals (  
  OSHPD_ID INT PRIMARY KEY,  
  FacilityName VARCHAR(255) NOT NULL,  
  FacilityType VARCHAR(255),  
  Address VARCHAR(255),  
  Address2 VARCHAR(255),  
  City VARCHAR(255),  
  State VARCHAR(2),
```

```
  ZIP INT,  
  County VARCHAR(255),  
  Facility_level_desc VARCHAR(255),  
  Total_number_beds INT,  
  Er_service_level_desc VARCHAR(255),  
  Facility_status_desc VARCHAR(255) CHECK (Facility_status_desc  
  IN ('Open', 'Suspense')),  
  License_type_desc VARCHAR(255) CHECK (License_type_desc IN  
  ('Clinic', 'Home Health Agency/Hospice', 'Hospital', 'Long Term Care  
  Facility')),  
  License_category_desc VARCHAR(255));
```

	 oshpd_id ▾	 facilityname ▾	 facilitytype ▾	 address ▾	 address2 ▾	 city
1	306370021	SAN YSIDRO HEALTH MOUNTAIN HEA...	Clinic - Federally Qualified H...	1388 BUCKMAN SPRINGS RD	<null>	CAMPO
2	306370856	VISTA COMMUNITY CLINIC	Clinic - Federally Qualified H...	1000 VALE TERRACE DR	<null>	VISTA
3	106190782	TARZANA TREATMENT CT	Clinic - Hospital Based or Ind...	1130 1/2 S HOOVER ST	<null>	LOS ANGE
4	306374450	NESTOR COMMUNITY HEALTH CENTER	Clinic - Federally Qualified H...	1016 OUTER RD	<null>	SAN DIEG
5	306374553	SAN DIEGO FAMILY CARE	Clinic - Federally Qualified H...	7011 LINDA VISTA RD	<null>	SAN DIEG
6	306374260	GROSSMONT SPRING VALLEY FAMILY...	Clinic - Federally Qualified H...	8788 JAMACHA RD	<null>	SPRING V
7	306370024	CITY HEIGHTS FAMILY HEALTH CEN...	Clinic - Federally Qualified H...	5454 EL CAJON BLVD	<null>	SAN DIEG
8	306370507	SAN YSIDRO HEALTH MATERNAL AND...	Clinic - Federally Qualified H...	4050 BEYER BLVD	<null>	SAN YSID

Hospital_Ratings

```
CREATE TABLE hospital_ratings (  
    id SERIAL PRIMARY KEY,  
    OSHPD_ID INT REFERENCES  
    hospitals(OSHPD_ID),  
    Performance_measure VARCHAR(255) NOT  
    NULL,  
    No_of_adverse_events INT,  
    No_of_cases INT,  
    Risk_adjusted_rate FLOAT,  
    Hospital_ratings VARCHAR(255)  
);
```



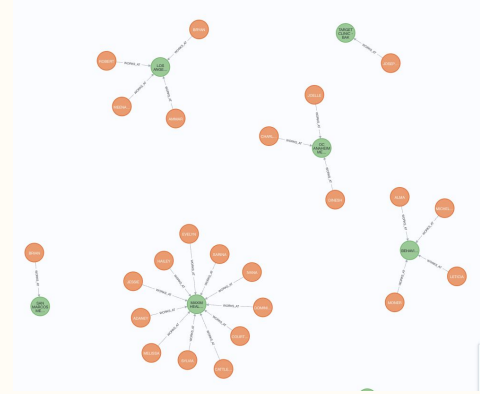
	id	oshpd_id	performance_measure	no_of_adverse_events	no_of_cases	risk_adjusted_rate	hospital_ratings
1	1	106014233	GI Hemorrhage	4	189	2	As Expected
2	2	106014233	Carotid Endarterectomy	0	12	0	As Expected
3	3	106014233	Acute Stroke Hemorrhagic	18	118	13.9	Better
4	4	106014233	Hip Fracture	3	109	2.2	As Expected
5	5	106014233	AMI	2	37	4.9	As Expected
6	6	106014233	Pneumonia	6	128	3.7	As Expected

Provider Network

```
CREATE TABLE ProviderNetwork (  
    ProviderID SERIAL PRIMARY KEY,  
    FirstName TEXT NOT NULL,  
    LastName TEXT NOT NULL,  
    ManagedCarePlan TEXT NOT NULL,  
    SubNetwork TEXT,  
    NPI INTEGER NOT NULL,  
    Taxonomy VARCHAR,  
    MCNAPProviderGroup TEXT,  
    MCNAPProviderType TEXT,  
    LicensureType VARCHAR(4),  
    PrimaryCare CHAR(1) CHECK (PrimaryCare IN ('Y', 'N')),  
    Specialist CHAR(1) CHECK (Specialist IN ('Y', 'N')),  
    SeesChildren TEXT CHECK (SeesChildren IN ('Both', 'Only', 'No')),  
    Telehealth TEXT CHECK (Telehealth IN ('Both', 'Only', 'No')),  
    BHIndicator TEXT,  
    OSHPD_ID INTEGER REFERENCES hospitals(OSHPD_ID)  
);
```

	🔍 firstname 🔽	🔍 lastname 🔽	🔍 managedcareplan 🔽	🔍 subnetwork 🔽	🔍 npi 🔽	🔍 taxonomy 🔽	🔍
1	DANI	CASTLEBERRY	Blue Shield of California Promise	Blue Shield of California Promise	1053706853	1041C0700X	Mer
2	Yvonne	Scarlett	Blue Shield of California Promise	Blue Shield of California Promise	1083026223	171100000X	Nor
3	NINOSKA	FONSECA	Blue Shield of California Promise	Blue Shield of California Promise	1558493221	106H00000X	Mer
4	Carolyn	Martinez	Blue Shield of California Promise	Blue Shield of California Promise	1609101997	363L00000X	Nor

Creating Nodes and Edges in Neo4j



Nodes	Relationships
Provider	:PART_OF (Subnetwork) - Directed :WORKS_AT (Hospital) - Directed :SPECIALIZES_IN (Taxonomy) - Directed
ManagedCarePlan	
SubNetwork	:UNDER (ManagedCarePlan) - Directed
Hospital	
Taxonomy	

Demonstration

- Relational Queries and Use-Cases
- Graph Queries and Use-Cases
- Application Queries and Use-Cases

Application Query Demos

1. Get the shortest path from a hospital to a specialty the user needs and returns shortest path of relationships. This result is then used in the Postgres query to return provider and hospital information regarding that specialty.
2. Get the best/worst rated hospital (using average of risk-adjusted ratings) and use facility names in Neo4j to find specialties provided.
3. Get the hospital that covers the most insurance (Neo4j) and get the insurance that has the most variety of category of aid and display its average cost of all categories of aid.
4. Identify healthcare providers offering services across networks within a specific cost range. It will return information about the provider, health care plan and associated subnetwork.

Postgres Query Example 1

Percentage of Providers Offering Telehealth in Each County

```
SELECT h.County,  
       ROUND(COUNT(CASE WHEN p.Telehealth IN ('Both', 'Only') THEN 1 END) *  
100.0 / COUNT(*), 2) AS Telehealth_Percentage  
FROM ProviderNetwork p  
JOIN Hospitals h ON p.OSHPD_ID = h.OSHPD_ID  
GROUP BY h.County  
ORDER BY Telehealth_Percentage DESC;
```

county	telehealth_percentage
Stanislaus	100
Santa Barbara	91.3
Kings	88.24
San Joaquin	70
Sacramento	63.48
San Diego	62.26
Kern	53.57
Placer	50
Tulare	42.74
Contra Costa	33.33

Postgres Query Example 2

Rank the hospitals by the number of primary care physicians they have

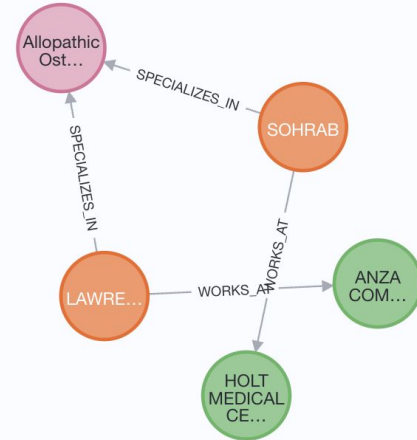
```
SELECT
    h.FacilityName,
    COUNT(p.Specialist) FILTER (WHERE p.PrimaryCare = 'Y') AS primary_count,
    RANK() OVER (ORDER BY COUNT(p.PrimaryCare) FILTER (WHERE p.PrimaryCare =
'Y') DESC) AS rank
FROM
    hospitals h
JOIN
    providernetwork p ON p.OSHPD_ID = h.OSHPD_ID
GROUP BY
    h.FacilityName;
```

	facilityname ▾	primary_count ▾	rank ▾
1	SAC HEALTH SYSTEM	154	1
2	MAIN CAMPUS COMMUNITY HEALTH C...	98	2
3	CLINICA DEL VALLE DEL PAJARO	57	3
4	SALUD PARA LA GENTE	57	3
5	TRUECARE	45	5
6	LOS ANGELES COMMUNITY HOSPITAL	39	6
7	CITY HEIGHTS FAMILY HEALTH CEN...	30	7
8	NORTH PARK FAMILY HEALTH CENTE...	28	8
9	EAST CLIFF FAMILY HEALTH CENTER	27	9
10	ALISAL HEALTH CENTER	27	9
11	VISTA COMMUNITY CLINIC	26	11
12	FAIR OAKS HEALTH CENTER	22	12
13	SANTA CRUZ WOMEN'S HEALTH CENT...		13

Neo4j Query Example 1

Pathfinding Query to find the connection between 2 hospitals through relationships.

```
MATCH p=shortestPath((h1:Hospital
{FacilityName: 'HOLT MEDICAL
CENTER'})-[:WORKS_AT|SPECIALIZES_IN*]-
(h2:Hospital {FacilityName: 'ANZA
COMMUNITY HEALTH'}))
RETURN p
```



Neo4j Query Example 2

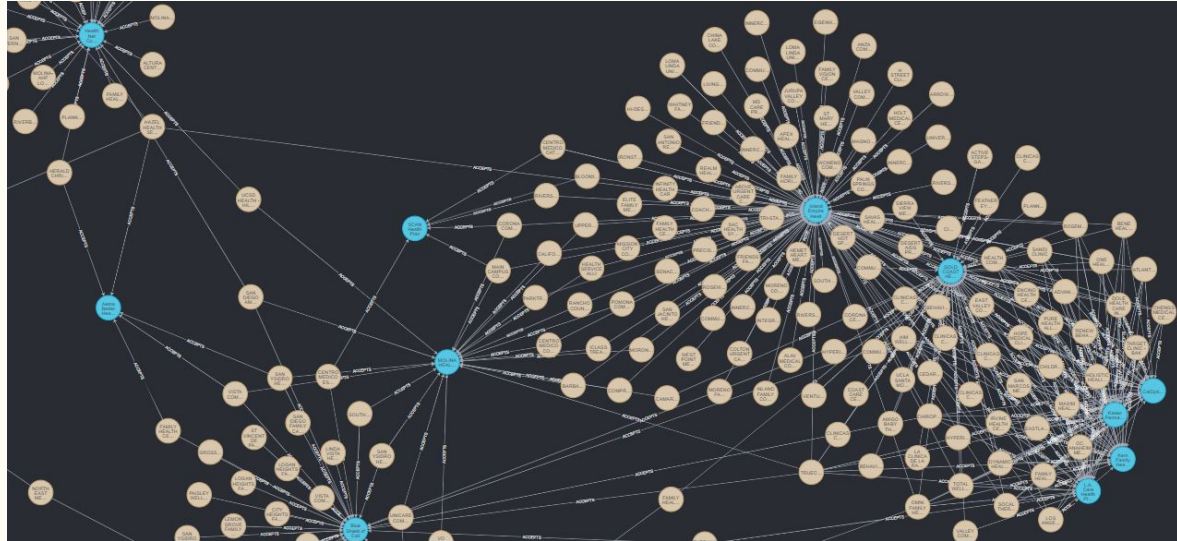
The most connected doctor based on the number of locations he worked at:

```
MATCH (p:Provider)-[:WORKS_AT]->(h:Hospital)
WITH p.FirstName + ' ' + p.LastName AS providerName,
COLLECT(DISTINCT h) AS hospitals,
COUNT(DISTINCT h) AS hospitalCount
RETURN providerName, hospitalCount,
[hospital IN hospitals | hospital.FacilityName] AS hospitalNames
ORDER BY hospitalCount DESC
```

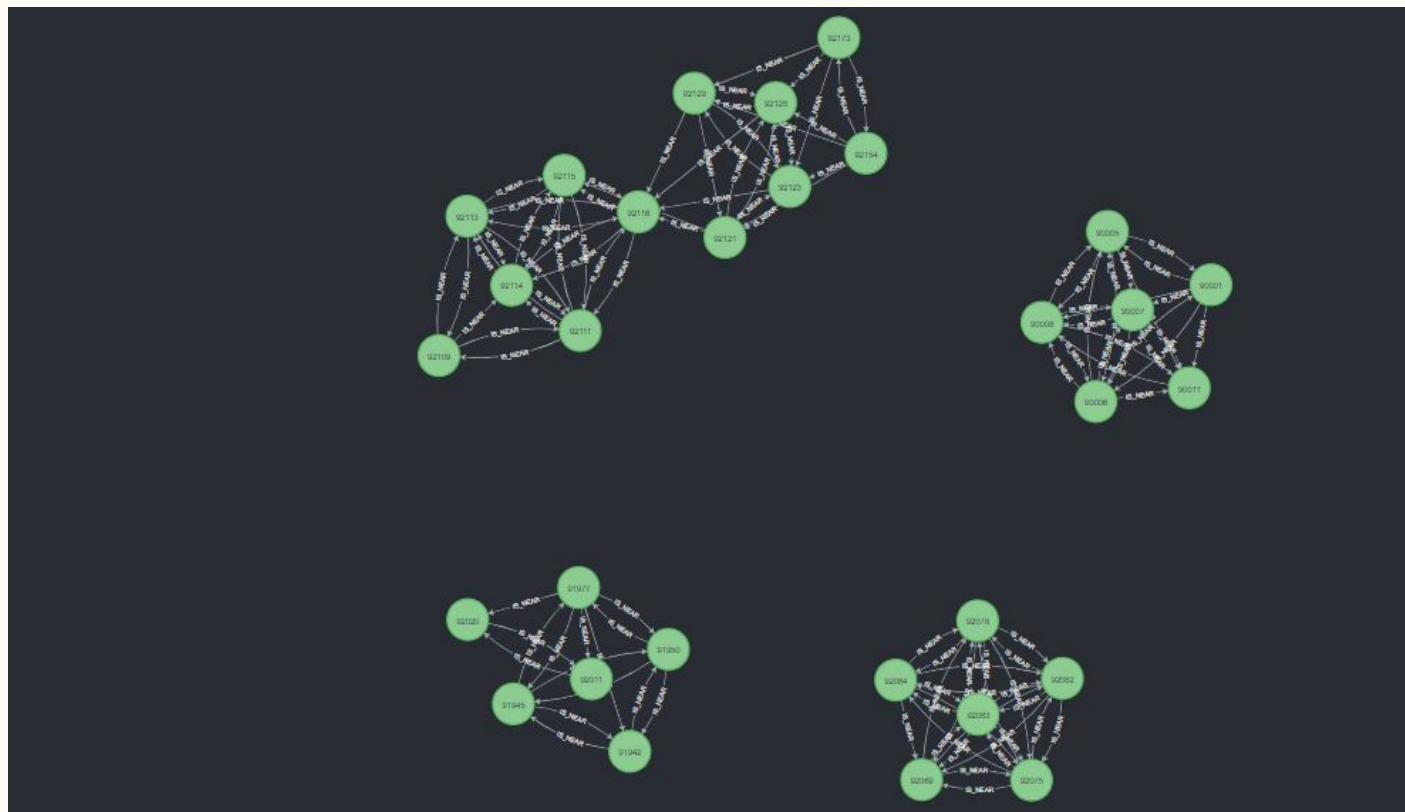
	providerName	hospitalCount	hospitalNames
1	"FIDEL SANTACRUZ"	6	["LOS ANGELES COMMUNITY HOSPITAL", "ST. FRANCIS MEDICAL CENTER", "MONTEREY PARK HOS"]
2	"ANTOINE MANSOUR"	6	["AHF HEALTHCARE CENTER - HOLLYWOOD", "GREATER EL MONTE COMMUNITY HOSPITAL", "ALH"]
3	"EDGARDO CAPITULO"	5	["LOS ANGELES COMMUNITY HOSPITAL", "MONTEREY PARK HOSPITAL", "EAST LOS ANGELES DOC"]
4	"CRAIG SMITH"	5	["LOS ANGELES COMMUNITY HOSPITAL", "CENTINELA HOSPITAL MEDICAL CENTER", "PROVIDENCI"]
5	"AUGUSTO ZABLAN"	5	["ST. FRANCIS MEDICAL CENTER", "GREATER EL MONTE COMMUNITY HOSPITAL", "HOLLYWOOD PF"]
6	"NANCY ELLERBROEK"	4	["PROVIDENCE SAINT JOSEPH MEDICAL CENTER", "LOS ANGELES COMMUNITY HOSPITAL", "PROV"]

Care Plan Within a Zip Code using Graph Topology

- Establishes a LOCATED_IN relationship between each hospital and its corresponding ZIP code.
- Creates an ACCEPTS relationship between hospitals and the managed care plans they accept.



IS_NEAR



- Creates a graph named zipPlanGraph for analysis.
- Includes ZIP, Hospital, and ManagedCarePlan nodes.
- Defines LOCATED_IN, ACCEPTS and IS_NEAR relationships for flexible traversal.

nodeProjection	relationshipProjection	graphName	nodeCount	relationshipCount	projectMillis
<pre>{ "ZIP": { "label": "ZIP", "properties": { "type": "ZIP" } }, "ManagedCarePlan": { "label": "ManagedCarePlan", "properties": { "type": "ManagedCarePlan" } }, "Hospital": { "label": "Hospital", "properties": { "type": "Hospital" } } }</pre>	<pre>{ "ACCEPTS": { "aggregation": "DEFAULT", "orientation": "UNDIRECTED", "indexInverse": false, "properties": { "type": "ACCEPTS" } }, "LOCATED_IN": { "aggregation": "DEFAULT", "orientation": "UNDIRECTED", "indexInverse": false, "properties": { "type": "LOCATED_IN" } }, "IS_NEAR": { "aggregation": "DEFAULT", "orientation": "UNDIRECTED", "indexInverse": false, "properties": { "type": "IS_NEAR" } } }</pre>	"zipPlanGraph"	915	2620	134

- Uses Dijkstra's shortest path algorithm on zipPlanGraph.
- Computes the shortest path from ZIP codes to Managed Care Plans.
- Converts distance into a similarity score (higher score = closer relationship).
- Orders results by ZIP code and plan name for better readability.

	zip_code	plan_name	similarity_score
1	"90001"	"AIDS Healthcare Foundation"	0.16666666666666666
2	"90001"	"Aetna Better Health"	0.2
3	"90001"	"Alameda Alliance for Health"	0.11111111111111111
4	"90001"	"Blue Shield of California Promise"	0.25



Thank you!