

# Sistema de recomendação de Hospitais Trabalho Final

Marcelo Vilela RA: 202619

Pedro Morelli RA: 204737

Renan Amorim RA: 186454

# Motivação Geral

---

Oferecer um banco de dados com a capacidade de recomendar os hospitais mais próximos, dado uma latitude e longitude.

Além disso, é possível pesquisar os melhores hospitais de uma especialidade médica, ou afim de uma informação de caráter público, informações sobre os estados com maior número de médicos de uma especialidade ou simplesmente com maior número de hospitais.

# Bases de Dados

---

Bases de Dados:

- Hospital General Information de [Data.Medicare.gov](https://data.medicare.gov)
- Physician Compare National de [Data.Medicare.gov](https://data.medicare.gov)
- World Cities Database de [simplemaps.com](https://simplemaps.com)

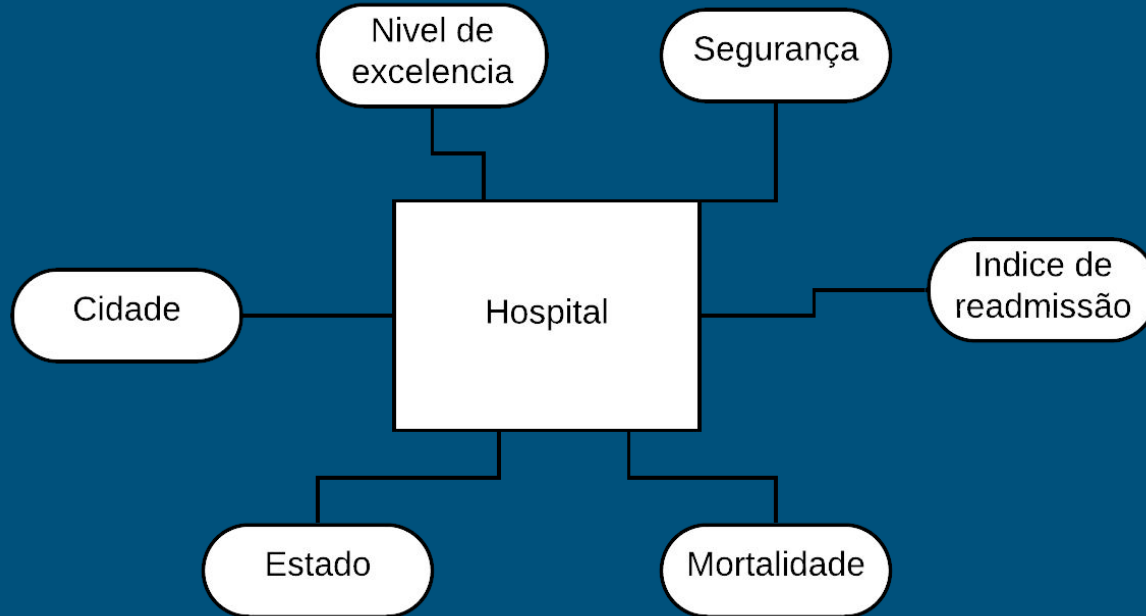
Informações baseadas em hospitais dos EUA.

# Etapa 1 - Análise Relacional com SQL

---

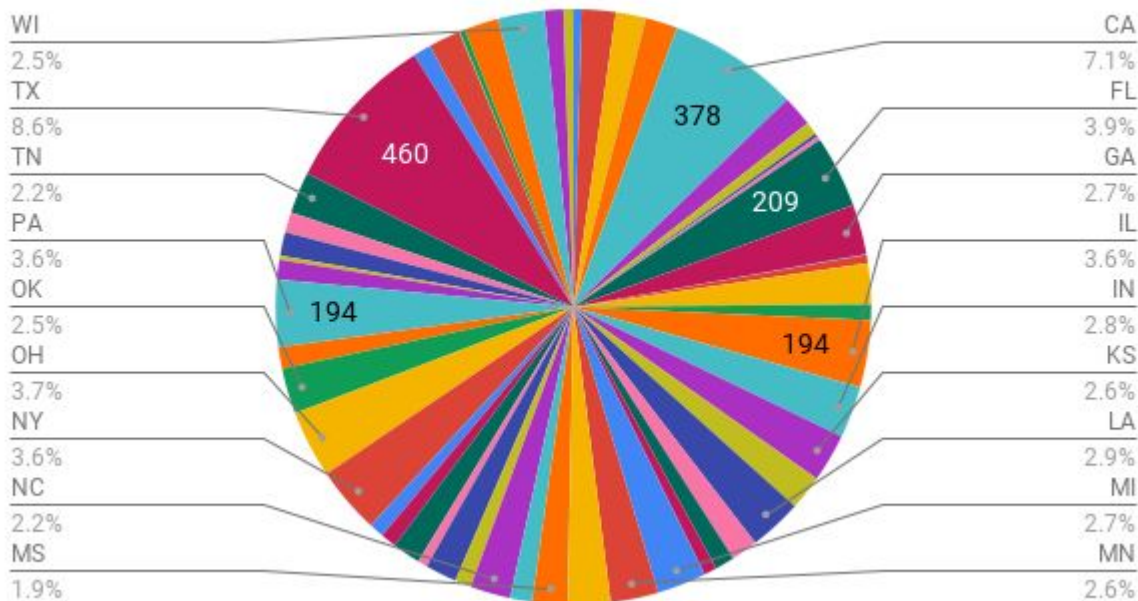
# Modelo conceitual inicial

---



# Relação Número de Hospitais x Excelência

Número de Hospitais por Estado



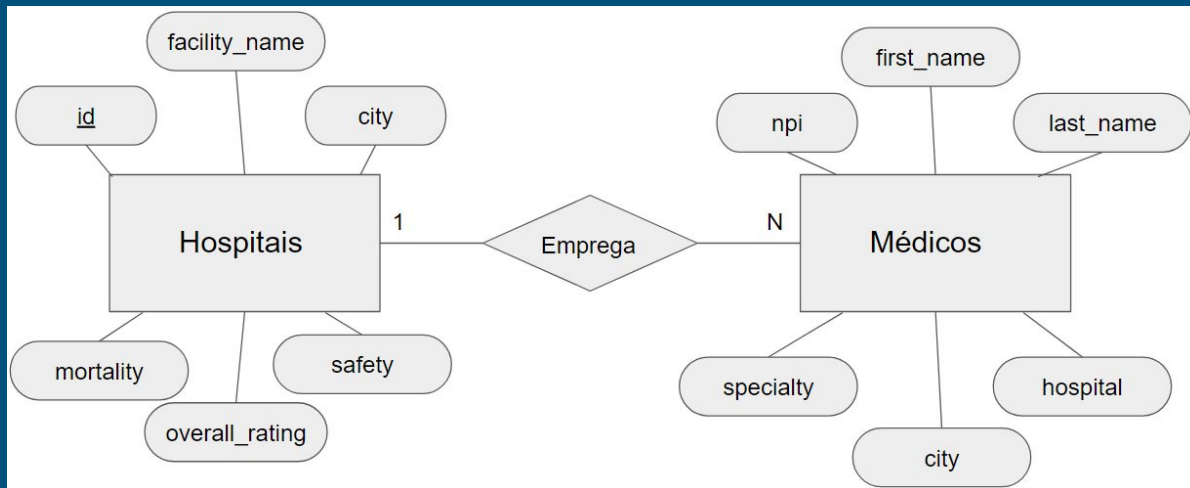
STATE	EXCELLENCE
FL	1
IN	1
MS	1
OH	1
SC	1
VA	1

# Etapa 1 - Análises em SQL

---

- Incremento na complexidade
  - Avaliação de hospitais agora com dados sobre os médicos
- Objetivo:
  - Estabelecer relações de qualidade de hospitais, especialidades médicas e ranqueamento para que se possa extrair dados visuais que auxiliem em recomendações.

# Etapa 1 - Análises em SQL



- **Hospitais** (id, facility\_name, city, county, emergency\_services, patient\_experience, mortality, safety\_of\_care, readmission, effectiveness\_of\_care, overall\_rating)
- **Médicos** (npa, first\_name, last\_name, gender, specialty, city, hospital)



# Etapa 1 - Análises em SQL

- Listar as cidades com mais hospitais

```
SELECT city, COUNT(*) qtd
FROM Hospitais
GROUP BY city
ORDER BY qtd DESC
```

index	CITY	QTD
0	HOUSTON	35
1	CHICAGO	33
2	PHILADELPHIA	25
3	COLUMBUS	23
4	LOS ANGELES	22
5	SPRINGFIELD	19
6	DALLAS	19
7	PHOENIX	19
8	WASHINGTON	18
9	BALTIMORE	18
10	LAS VEGAS	18
11	JACKSON	17
12	OKLAHOMA CITY	17

# Etapa 1 - Análises em SQL

- Listar os hospitais com mais médicos

```
SELECT Medicos.hospital, COUNT(*) aux
FROM Medicos, hospitais
WHERE Medicos.hospital = hospitais.facility_name
GROUP BY Medicos.hospital
ORDER BY aux DESC
```

# index	HOSPITAL	AUX
0	MEMORIAL HOSPITAL	210
1	MEMORIAL MEDICAL CENTER	70
2	ST JOSEPH'S HOSPITAL	30
3	COMMUNITY HOSPITAL	24
4	GOOD SAMARITAN HOSPITAL	24
5	COMMUNITY MEMORIAL HOSPITAL	21
6	ST JOSEPH HOSPITAL	21
7	ST JOSEPH MEDICAL CENTER	20
8	HIGHLAND HOSPITAL	18
9	ST LUKES HOSPITAL	16
10	ST MARY MEDICAL CENTER	16
11	EDWARD HOSPITAL	16
12	CLEVELAND CLINIC	16
13	MEDSTAR WASHINGTON HOSPITAL CENTER	15
14	CJW MEDICAL CENTER	14
15	UNIVERSITY OF VIRGINIA MEDICAL CENTER	14
16	MASSACHUSETTS GENERAL HOSPITAL	12
17	ST FRANCIS HOSPITAL	12
18	TRISTAR CENTENNIAL MEDICAL CENTER	12
19	SAINT JOSEPH HOSPITAL	12
20	UNIVERSITY OF ALABAMA HOSPITAL	12

# Etapa 1 - Análises em SQL

- Listar os 10 melhores hospitais com base em uma especialidade

```
SELECT h.facility_name, h.city
  FROM Hospitais h, Medicos m
 WHERE m.specialty = 'PATHOLOGY'
    and m.hospital = h.facility_name
    and h.overall_rating != 'Not Available'
 ORDER BY h.overall_rating DESC
 LIMIT 10
```

index	FACILITY_NAME	CITY
0	MASSACHUSETTS GENERAL HOSPITAL	BOSTON
1	MILTON S HERSHEY MEDICAL CENTER	HERSHEY
2	UNIVERSITY OF WASHINGTON MEDICAL CTR	SEATTLE
3	GOOD SAMARITAN REGIONAL MEDICAL CENTER	CORVALLIS
4	OCEAN MEDICAL CENTER	BRICK
5	OCEAN MEDICAL CENTER	BRICK
6	ADVOCATE BROMENN MEDICAL CENTER	NORMAL
7	MEMORIAL MEDICAL CENTER	ASHLAND
8	MEMORIAL MEDICAL CENTER	ASHLAND
9	ADVOCATE BROMENN MEDICAL CENTER	NORMAL

# Etapa 1 - Análises em SQL

- Listar os médicos que trabalham nos hospitais com melhor experiência para os pacientes

```
SELECT DISTINCT m.first_name, m.last_name, m.specialty
FROM Medicos m, Hospitais h
WHERE m.hospital = h.facility_name
and h.patient_experience = 'Above the national average'
```

index	FIRST_NAME	LAST_NAME	SPECIALTY
0	DANA	RUBIS	NURSE PRACTITIONER
1	SHARLENE	MADANES	NURSE PRACTITIONER
2	BRYCE	PASCHOLD	PODIATRY
3	LYNN	HOLLIDAY	FAMILY MEDICINE
4	MADAIAH	LOKESHWARI	INTERNAL MEDICINE
5	NUMAAN	MALIK	PULMONARY DISEASE
6	HOWARD	SCHULTZ	PHYSICIAN ASSISTANT
7	ANTHONY	CALABRETTA	GENERAL SURGERY
8	JOSELYN	JOSEPH	INTERNAL MEDICINE
9	DEEPTI	BAHL	ENDOCRINOLOGY
10	KRISTIN	JOHNSON	NURSE PRACTITIONER

# Etapa 1 - Análises em SQL

- Selecione uma especialidade, listar os hospitais com mais médicos daquela especialidade

```
SELECT h.facility_name, m.specialty, COUNT (*) aux
FROM Medicos m, Hospitais h
WHERE m.specialty = 'NEUROLOGY' AND h.facility_name = m.hospital
GROUP BY h.facility_name
ORDER BY aux DESC
```

index	FACILITY_NAME	SPECIALTY	AUX
0	COMMUNITY HOSPITAL	NEUROLOGY	16
1	SUMMIT HEALTHCARE REGIONAL MEDICAL CENTER	NEUROLOGY	6
2	ADVOCATE CHRIST HOSPITAL & MEDICAL CENTER	NEUROLOGY	3
3	INOVA ALEXANDRIA HOSPITAL	NEUROLOGY	3
4	MEMORIAL HEALTHCARE	NEUROLOGY	2
5	BLESSING HOSPITAL	NEUROLOGY	2
6	MASSACHUSETTS GENERAL HOSPITAL	NEUROLOGY	2
7	MARSHFIELD MEDICAL CENTER	NEUROLOGY	2
8	DOCTORS HOSPITAL AT RENAISSANCE	NEUROLOGY	2
9	RONALD REAGAN U C L A MEDICAL CENTER	NEUROLOGY	2
10	DUKE UNIVERSITY HOSPITAL	NEUROLOGY	2
11	SIERRA VISTA REGIONAL MEDICAL CENTER	NEUROLOGY	1
12	SARASOTA MEMORIAL HOSPITAL	NEUROLOGY	1

# Como expandir a query anterior?

---

Query: Para cada especialidade, listar o hospital com o maior número de médicos que atendem àquela especialidade.

- Lógica relacional em SQL é inadequada, precisamos de uma estrutura hierárquica.

# Etapa 2 - Análise Hierárquica com XQuery

---

# Desafios

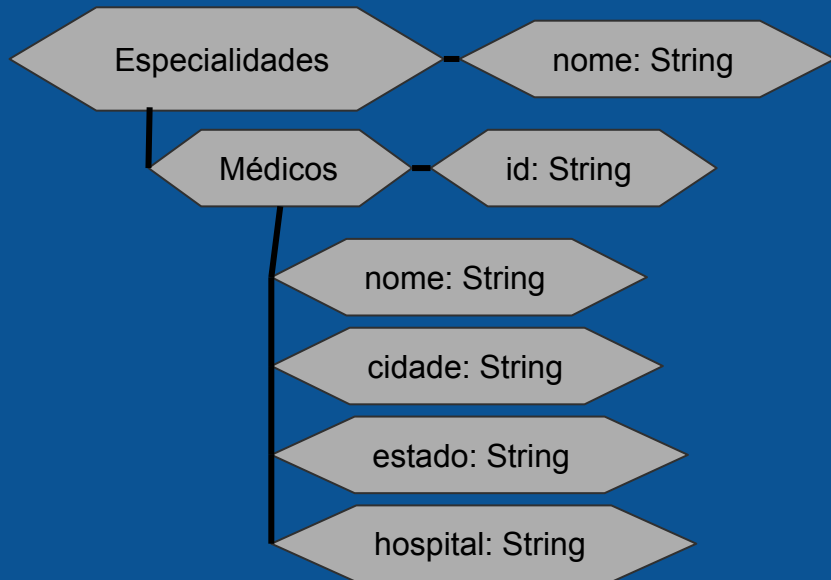
---

- Os datasets de temática hospitalar disponíveis na web não apresentam um nível de hierarquia adequado.
- Portanto, precisamos desenvolver um script em python para conversão de datasets em .csv para .xml cuja hierarquia permitiu as seguintes análises.

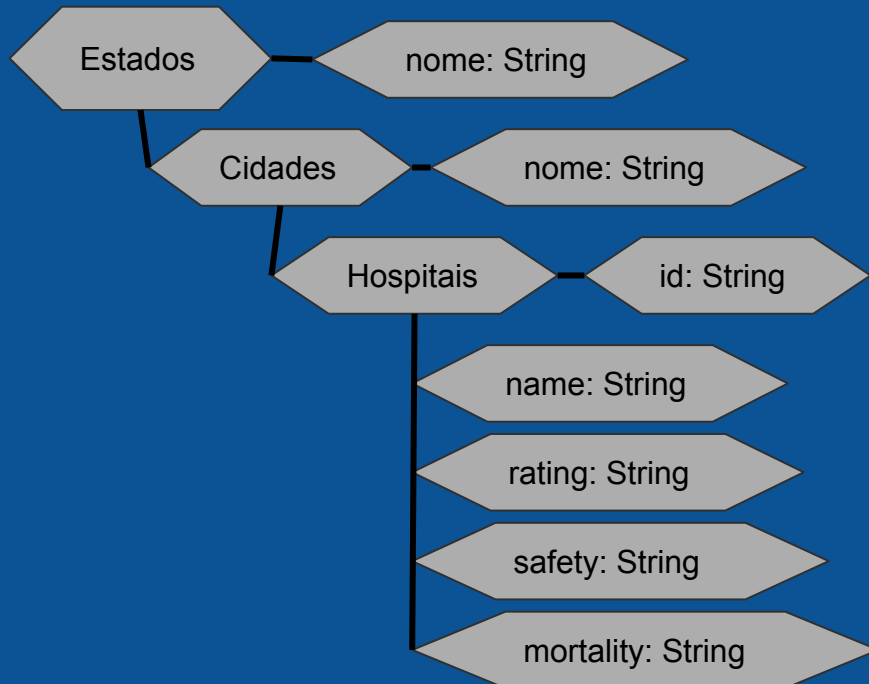


# Modelo Conceitual

## Médicos



## Hospitais



## Etapa 2 - Análises em XQuery

---

- Para cada especialidade, listar o hospital com o maior número de médicos que atendem àquela especialidade.

```
let $hospital := doc('http://www.students.ic.unicamp.br/~ra186454/hospital2.xml')
let $doctors := doc('http://www.students.ic.unicamp.br/~ra186454/doctor2_mini.xml')

for $x in $doctors//especialidades
for $y in distinct-values($x//hospital1)

let $count := count($x//medico[@hospital1 eq $y])
order by $count descending

return concat("(", " ", $y, " ", $count, ")")
```

# Etapa 2 - Análises em XQuery

---

- Contagem do número de médicos por especialidade

```
let $doctors := doc('http://www.students.ic.unicamp.br/~ra186454/doctor2_mini.xml')
for $i in ($doctors//especialidade)
for $j in $i//count(medico)
return if ($j != 0) then ($i/@nome/string(), $j) else ()
```

## Etapa 2 - Análises em XQuery

---

- Seleciona os hospitais de excelência (melhores parâmetros de avaliação) e os lista por ordem decrescente da melhor classificação

```
let $hospital := doc('http://www.students.ic.unicamp.br/~ra186454/hospital2.xml')
for $x in $hospital/estados/estado/cidades/cidade/hospitais/hospital
where $x/mortality="&#34;Below the national average&#34;" and $x/safety="&#34;Above the national average&#34;"
and $x/effectiveness="&#34;Above the national average&#34;"
order by $x/hospital_rating descending
return ( " - Nome do Hospital" , $x/facility_name/string(), " Rating:" , $x/hospital_rating/string())
```

## Etapa 2 - Análises em XQuery

---

- Retorna o nome dos médicos que operam em hospitais cuja efetividade é acima da média nacional

```
let $hospital := doc('http://www.students.ic.unicamp.br/~ra186454/hospital2.xml')
let $doctors := doc('http://www.students.ic.unicamp.br/~ra186454/doctor2_mini.xml')

for $x in $hospital/estados/estado/cidades/cidade/hospitais/hospital
for $y in $doctors/especialidades/especialidade/medicos/medico
where $x/effectiveness="&#34;Above the national average&#34;" and $x/facility_name=$y/hospital1
return ("Dr.", $y/terceiro_nome/string(), ", ", $y/primeiro_nome/string(), ". ")
```

## Etapa 2 - Análises em XQuery

---

- Para cada cidade, listar qual a especialidade com mais médicos

```
let $hospital := doc('http://www.students.ic.unicamp.br/~ra186454/hospital2.xml')
let $doctors := doc('http://www.students.ic.unicamp.br/~ra186454/doctor2_mini.xml')

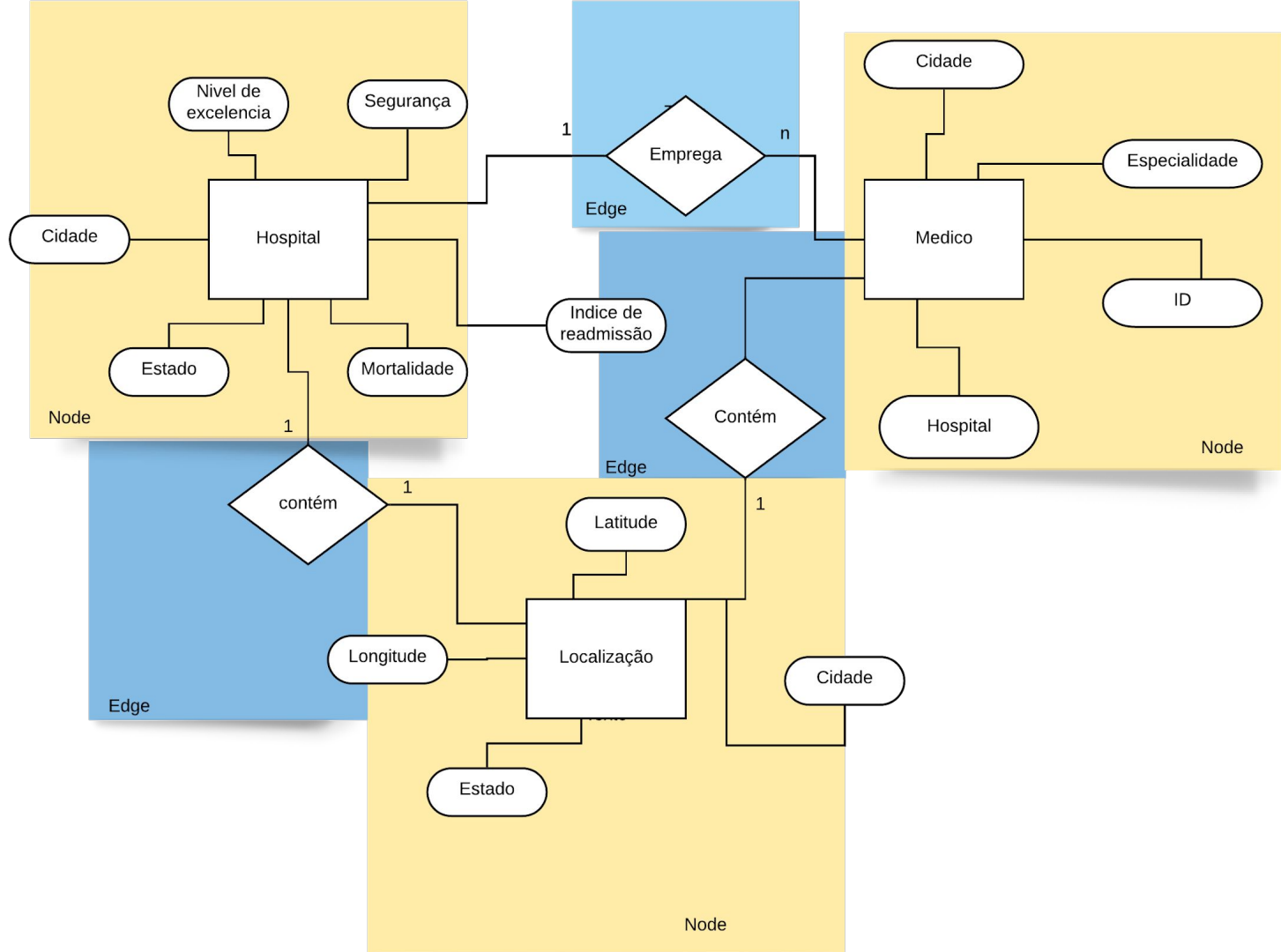
for $y in distinct-values($doctors//especialidade)
for $x in $doctors//cidade

let $count := count($x//medico[@cidade eq $y/@nome])
order by $count descending

return concat("(", " ", $y/@nome, " ", $count, ")")
```

# Etapa 3 - Análise de Rede

---





City(25)

HUNTS...

CULLM...

FORT  
PAYNE

SHEFFI...

ALABAS...

ANNIST...

GADSD...

BOAZ

LUVER...

BIRMIN...

HAMILT...

ENTER...

Hospital(25)

WEDO...

CULLM...

ANDAL...

COMMU...

DEKALB  
REGIO...

CALLAH...

SHELBY  
BAPTIST

STRING...

HUNTS...

MARION  
REGIO...

MARSH...

CRENS...

ST  
VINCE...

# Objetivo Final

---

Através do sistema de redes com o Neo4J, conseguimos estabelecer uma relação entre hospitais e localidade.

Dessa forma, podemos recomendar os hospitais mais próximos de acordo com uma localização determinada.

city	NumberOfHospitals
"CHICAGO"	29
"HOUSTON"	26
"LOS ANGELES"	20
"DALLAS"	19
"PHILADELPHIA"	19
"COLUMBUS"	18
"SPRINGFIELD"	17
"BALTIMORE"	17
"WASHINGTON"	16
"PHOENIX"	16
"JACKSON"	16
"OKLAHOMA CITY"	16
"GREENVILLE"	14
"SAN ANTONIO"	13
"LAS VEGAS"	13

# Etapa 3 - Análises em Redes/Neo4J

---

- Simulação dos 10 hospitais mais próximos de Liberty Island

```
with "1 Liberty Island, New York" as myLocation
call apoc.spatial.geocodeOnce(myLocation) yield location
WITH point({longitude: location.longitude, latitude: location.latitude}) as myPosition, 100 as distanceInKm
MATCH (h:Hospital)-->(rating:Rating)
WHERE exists(h.latitude) and
distance(myPosition, point({longitude:h.longitude,latitude:h.latitude})) < (distanceInKm * 100)
RETURN h,rating.name as rating,distance(myPosition, point({longitude:h.longitude,latitude:h.latitude})) as distance order by distance limit 10
```

# Etapa 3 - Análises em Redes/Neo4J

- Número de hospitais por cidade

```
MATCH (h:Hospital)-[:IS_IN*3..3]->(city)
return city.name as city,count(h) as NumberOfHospitals order by NumberOfHospitals desc limit 15
```

city	NumberOfHospitals
"CHICAGO"	29
"HOUSTON"	26
"LOS ANGELES"	20
"DALLAS"	19
"PHILADELPHIA"	19
"COLUMBUS"	18
"SPRINGFIELD"	17
"BALTIMORE"	17
"WASHINGTON"	16
"PHOENIX"	16
"JACKSON"	16
"OKLAHOMA CITY"	16
"GREENVILLE"	14
"SAN ANTONIO"	13
"LAS VEGAS"	13

# Etapa 3 - Análises em Redes/Neo4J

---

- Retornar melhores 10 estados por avaliação

```
MATCH (r)<-[:HAS_RATING]-(h:Hospital)-[:IS_IN*5..5]->(state)
where not r.name="Not Available"
return state.name as state,avg(toINT(r.name)) as averageRating,count(h) as numberOfHospitals order by averageRating desc limit 15
```

## Etapa 3 - Análises em Redes/Neo4J

---

- Estados com hospitais acima da média em efetividade de tratamento

```
MATCH (h:Hospital)-[:IS_IN*5..5]->(state) where h.effectiveness = "Above the National average"  
return state.name as state,h.effectiveness,count(h) as numberOfHospitals order by numberOfHospitals desc limit 15
```

# Etapa 3 - Análises em Redes/Neo4J

---

- Estados com hospitais abaixo da média em mortalidade

```
MATCH (h:Hospital)-[:IS_IN*5..5]->(state) where h.mortality = "Below the National average"  
return state.name as state,h.mortality,count(h) as numberOfHospitals order by numberOfHospitals desc limit 15
```