Massachusetts General Hospital

About Massachusetts General Hospital

- Academic medical center and teach hospital founded in **1811**
- **Harvard** Medical School **affiliated**, majority of staff faculty
- Highly ranked in the US
- Advanced research facilities
- **$1 billion** invested in research operations
- Conducted over **1,200 clinical trials**
- Leading hospital in treating complex and rare disorders
- Highly involved in COVID-19 research, testing for treatment discovery

Requirements

Questions to lead with

- How many patients have been admitted or readmitted over time?
- How long are patients staying in the hospital on average?
- How much is the average cost per visit?
- How many procedures are covered by insurance?
- What's the % of procedures / encounters covered and not covered by insurance
- What's the #1 cause of patients coming in for procedures?
- Any fluctuations or changes in data during real life events (COVID-19)

About the Dataset

Four main datasets

Encounters

Procedures

Patients

Payers

Four main datasets

Encounters

Procedures

Patients

Payers

Four main datasets

Encounters

patient encounter data: information that documents a patients diagnoses, treatments provided by a clinician the admission and discharge times for administrative purposes

Column	Description
Id	Primary Key. Unique Identifier of the encounter.
Start	The date and time (iso8601 UTC Date (yyyy-MM-dd'T'HH:mm'Z')) the encounter started
Stop	The date and time (iso8601 UTC Date (yyyy-MM-dd'T'HH:mm'Z')) the encounter concluded
Patient	Foreign key to the Patient.
Organization	Foreign key to the Organization.
Payer	Foreign key to the Payer.
EncounterClass	The class of the encounter, such as ambulatory, emergency, inpatient, outpatient, wellness, or urgent care
Code	Encounter code from SNOMED-CT
Description	Description of the type of encounter.
Base_Encounter_Cost	The base cost of the encounter, not including any line-item costs related to medications, immunizations, procedures, or other services.
Total_Claim_Cost	The total cost of the encounter, including all line items.
Payer_Coverage	The amount of cost covered by the Payer.
ReasonCode	Diagnosis code from SNOMED-CT, only if this encounter targeted a specific condition.
ReasonDescription	Description of the reason code.

Four main datasets

Encounters

Procedures

Patients

Payers

Four main datasets

Procedures

Patient procedure data: documenting the procedures patients had, including surgeries. Some procedures are from encounters.

Column	Description
Start	The date and time (iso8601 UTC Date (yyyy-MM-dd'T'HH:mm'Z')) the procedure was performed.
Stop	The date and time (iso8601 UTC Date (yyyy-MM-dd'T'HH:mm'Z')) the procedure was completed, if applicable.
Patient	Foreign key to the Patient.
Encounter	Foreign key to the Encounter where the procedure was performed.
Code	Procedure code from SNOMED-CT
Description	Description of the procedure.
Base_Cost	The line item cost of the procedure.
ReasonCode	Diagnosis code from SNOMED-CT specifying why this procedure was performed.
ReasonDescription	Description of the reason code.

Four main datasets

Encounters

Procedures

Patients

Payers

Four main datasets

Patients

patient encounter data: information that documents a patients diagnoses, treatments provided by a clinician the admission and discharge times for administrative purposes

Column	Description
Id	Primary Key. Unique Identifier of the patient.
BirthDate	The date (YYYY-MM-DD) the patient was born.
DeathDate	The date (YYYY-MM-DD) the patient died.
Prefix	Name prefix, such as Mr., Mrs., Dr., etc.
First	First name of the patient.
Middle	Middle name of the patient.
Last	Last or surname of the patient.
Suffix	Name suffix, such as PhD, MD, JD, etc.
Maiden	Maiden name of the patient.
Marital	Marital Status. M is married, S is single. Currently no support for divorce (D) or widowing (W)
Race	Description of the patient's primary race.
Ethnicity	Description of the patient's primary ethnicity.
Gender	Gender. M is male, F is female.
BirthPlace	Name of the town where the patient was born.

Four main datasets

Encounters

Procedures

Patients

Payers

Four main datasets

Payers

Insurance payers' data: contains information about medical insurance companies, their credentials, location and phone numbers

Column	Description
Id	Primary key of the Payer (e.g. Insurance company).
Name	Name of the Payer.
Address	Payer's street address without commas or newlines.
City	Street address city.
State_Headquartered	Street address state abbreviation.
Zip	Street address zip or postal code.
Phone	Payer's phone number.

Preprocessing

- ✓ Remove duplicates in PKs
- ✓ Fill nulls instead of remove
  - ✓ 0 in numeric columns
  - ✓ ‘Not mentioned’ in text columns
- ✓ Fix patient records
- ✓ Create dimension and fact tables
- ✓Verify relationships between PKs & FKs in model view
- ✓ Create bridge table where needed

✓ *Fill nulls instead of remove*

REASONCODE	REASONDESCRIPTION
Valid 42%	Valid 42%
Error 0%	Error 0%
Empty 57%	Empty 57%
27 distinct, 11 unique	27 distinct, 11 unique
null	
null	
null	
null	
55822004	Hyperlipidemia
null	
null	
363406005	Malignant tumor of colon
null	
254637007	Non-small cell lung cancer (disorder)

1.2 REASONCODE	REASONDESCRIPTION
Valid 100%	Valid 100%
Error 0%	Error 0%
Empty 0%	Empty 0%
27 distinct, 11 unique	27 distinct, 11 unique
0	Not mentioned
0	Not mentioned
0	Not mentioned
0	Not mentioned
55822004	Hyperlipidemia
0	Not mentioned
0	Not mentioned
363406005	Malignant tumor of colon
0	Not mentioned
254637007	Non-small cell lung cancer (disorder)

✓*Fix patient records*

FIRST	LAST
Valid 100%	Valid 100%
Error 0%	Error 0%
Empty 0%	Empty 0%
842 distinct, 721 unique	498 distinct, 219 unique
Nikita578	Erdman779
Zane918	Hodkiewicz467
Quinn173	Marquardt819
Abel832	Smitham825
Edwin773	Labadie908
Frankie174	Oberbrunner298
Eilene124	Walsh511
Lowell343	Price929
Adrian111	Gleason633
Eugene421	Abernathy524
Jay242	Wisozk929

→

FIRST.1	LAST.1
Valid 100%	Valid 100%
Error 0%	Error 0%
Empty 0%	Empty 0%
842 distinct, 721 unique	498 distinct, 219 unique
Nikita	Erdman
Zane	Hodkiewicz
Quinn	Marquardt
Abel	Smitham
Edwin	Labadie
Frankie	Oberbrunner
Eilene	Walsh
Lowell	Price
Adrian	Gleason
Eugene	Abernathy
Jay	Wisozk

✓ *Reference dimension and fact tables*

	Id	NAME
	Valid 100% Error 0% Empty 0% 10 distinct, 10 unique	Valid 100% Error 0% Empty 0% 10 distinct, 10 unique
1	b3221cfc-24fb-339e-823d-bc4136cbc4ed	Dual Eligible
2	7caa7254-5050-3b5e-9eae-bd5ea30e809c	Medicare
3	7c4411ce-02f1-39b5-b9ec-dfbea9ad3c1a	Medicaid
4	d47b3510-2895-3b70-9897-342d681c76...	Humana
5	6e2f1a2d-27bd-3701-8d08-dae202c58632	Blue Cross Blue Shield
6	5059a55e-5d6e-34d1-b6cb-d83d16e57bcf	UnitedHealthcare
7	4d71f845-a6a9-3c39-b242-14d25ef86a8d	Aetna
8	047f6ec3-6215-35eb-9608-f9dda363a44c	Cigna Health
9	42c4fca7-f8a9-3cd1-982a-dd9751bf3e2a	Anthem
10	b1c428d6-4f07-31e0-90f0-68ffa6ff8c76	NO_INSURANCE

Payers Dimension Table

	ENCOUNTERCLASS
	Valid 100% Error 0% Empty 0% 6 distinct, 6 unique
1	ambulatory
2	outpatient
3	wellness
4	urgentcare
5	inpatient
6	emergency

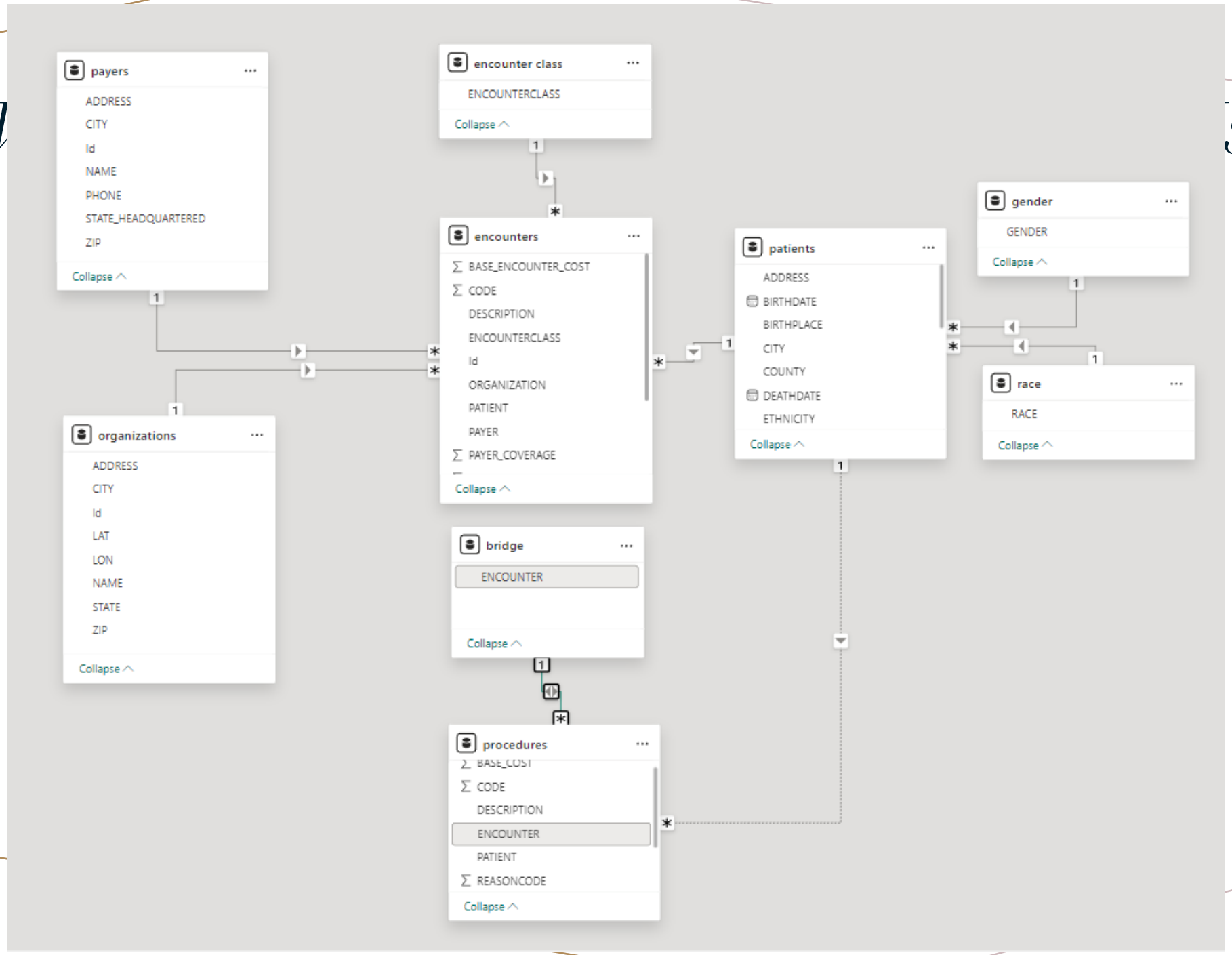
From Encounter

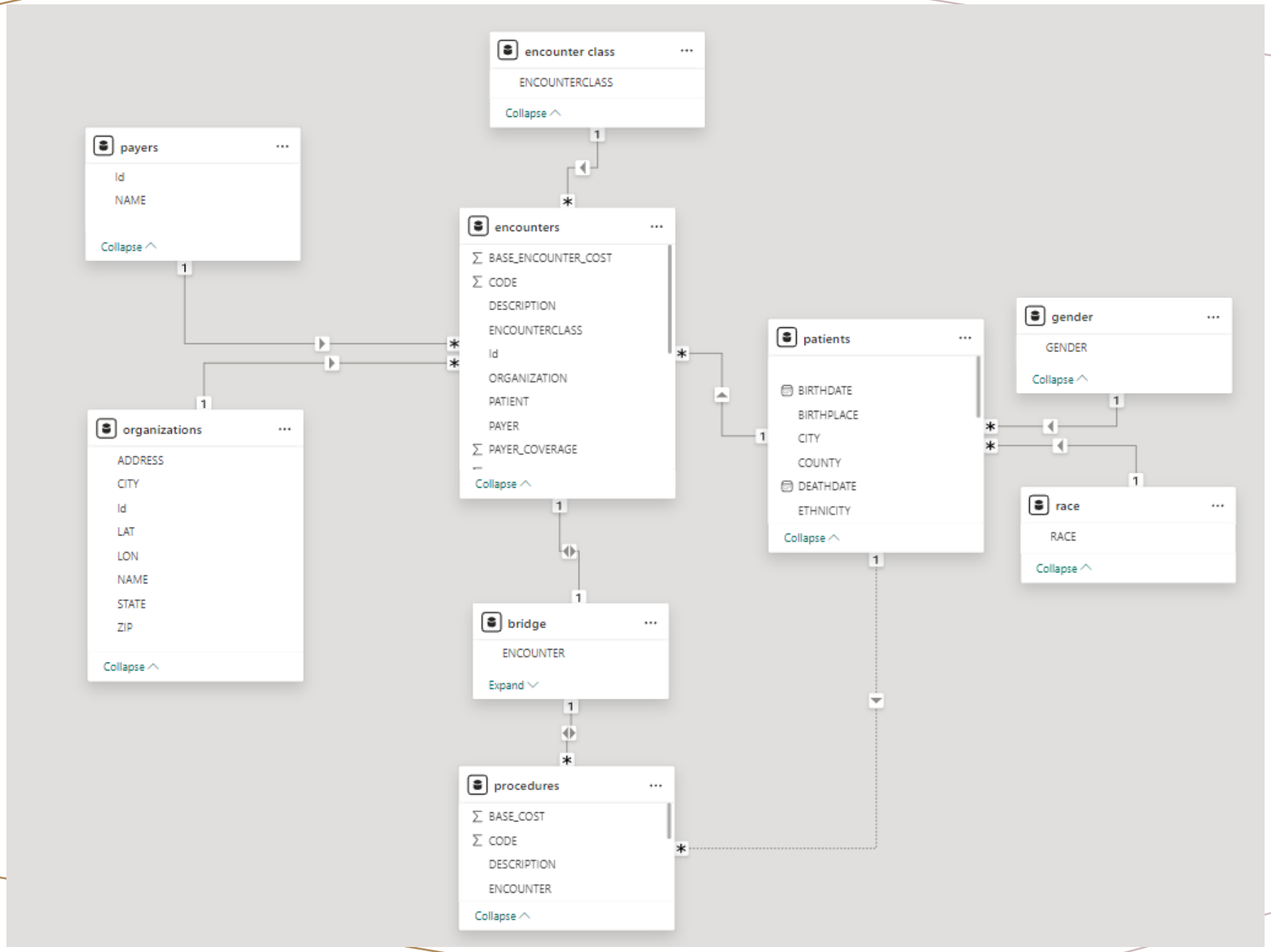
	GENDER
	Valid 100% Error 0% Empty 0% 2 distinct, 2 unique
1	F
2	M

	RACE
	Valid 100% Error 0% Empty 0% 6 distinct, 6 unique
1	white
2	asian
3	black
4	native
5	hawaiian
6	other

From Patient Data

- ✓ *Verify relationships between PKs & FKs*


payers
ADDRESS
CITY
Id
NAME
PHONE
STATE_HEADQUARTERED
ZIP
Collapse
encounter class
ENCOUNTERCLASS
Collapse
encounters
BASE_ENCOUNTER_COST
CODE
DESCRIPTION
ENCOUNTERCLASS
Id
ORGANIZATION
PATIENT
PAYER
PAYER_COVERAGE
Collapse
patients
ADDRESS
BIRTHDATE
BIRTHPLACE
CITY
COUNTY
DEATHDATE
ETHNICITY
Collapse
gender
GENDER
Collapse
race
RACE
Collapse
organizations
ADDRESS
CITY
Id
LAT
LON
NAME
STATE
ZIP
Collapse
bridge
ENCOUNTER
Collapse
procedures
CODE
DESCRIPTION
ENCOUNTER
PATIENT
REASONCODE
Collapse


encounter class
ENCOUNTERCLASS
Collapse
payers
Id
NAME
Collapse
encounters
BASE_ENCOUNTER_COST
CODE
DESCRIPTION
ENCOUNTERCLASS
Id
ORGANIZATION
PATIENT
PAYER
PAYER_COVERAGE
Collapse
patients
BIRTHDATE
BIRTHPLACE
CITY
COUNTY
DEATHDATE
ETHNICITY
Collapse
gender
GENDER
Collapse
race
RACE
Collapse
organizations
ADDRESS
CITY
Id
LAT
LON
NAME
STATE
ZIP
Collapse
bridge
ENCOUNTER
Expand
procedures
BASE_COST
CODE
DESCRIPTION
ENCOUNTER
Collapse

Proposed Dashboards

Massachusetts General Hospital - Visits and Procedures Overview

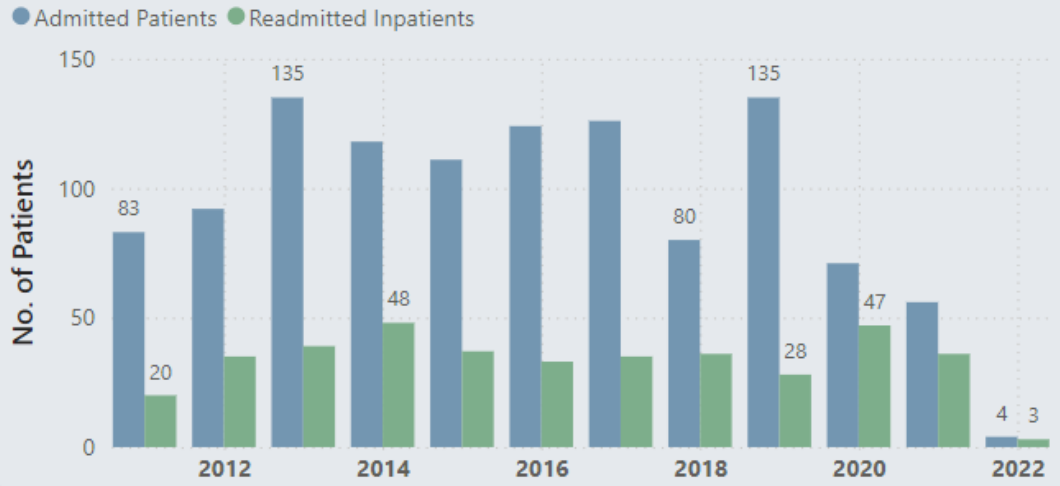
Hospital Visits

Insurance

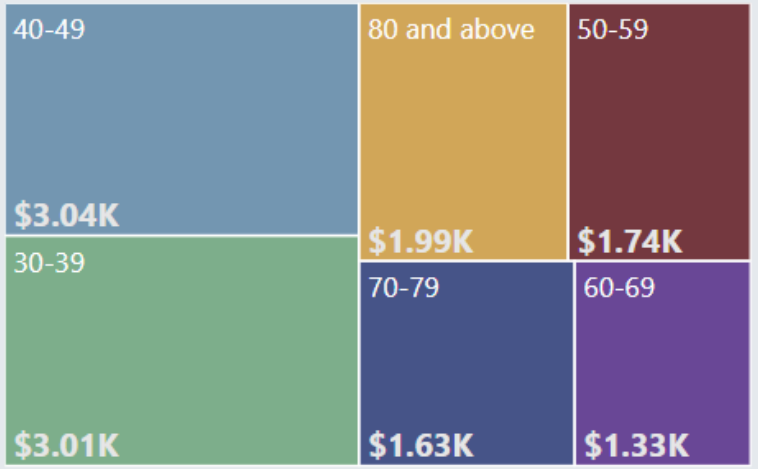
COVID-19

Total Revenue Encounters	Average Cost Per Visit	Avg. Patient Visit Time	Avg. Inpatient Stay
$101.5M	$3,465	6 hr(s)	37 hr(s)

Patient Admission – Readmission (Jan. 2011 – Feb. 2022)



Average Procedure Cost by Age Group



Massachusetts General Hospital - Visits and Procedures Overview

Hide Panel

Hospital Visits

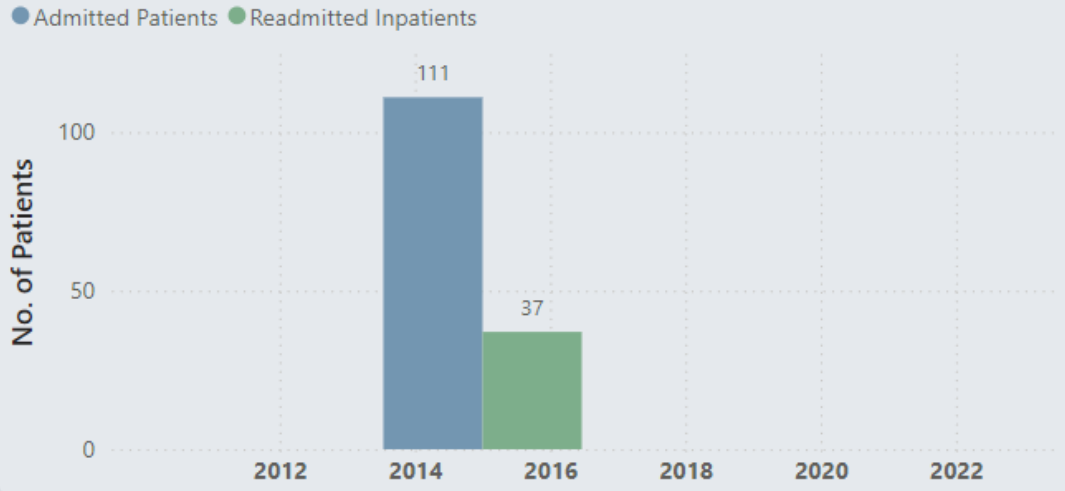
Insurance

COVID-19

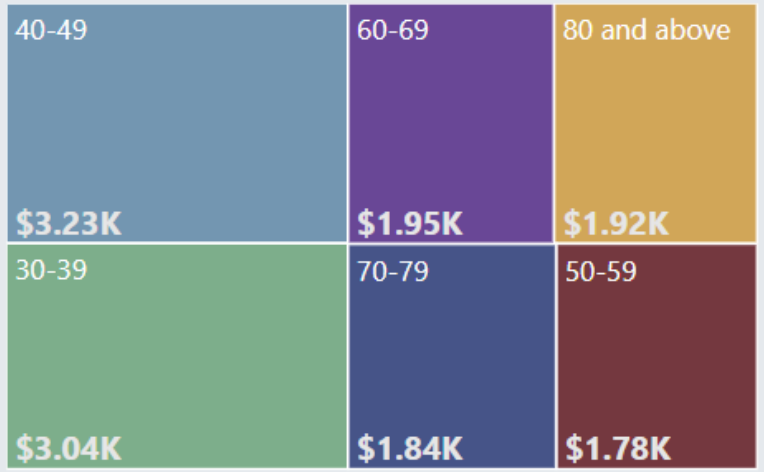
Total Revenue Encounters	Average Cost Per Visit	Avg. Patient Visit Time	Avg. Inpa
$10.1M	$3,839	20 hr(s)	25

Year
- 2014
- 2015
- 2016
- 2017

Patient Admission – Readmission (Jan. 2011 – Feb. 2022)



Average Procedure Cost by Age Group



Massachusetts General Hospital - Insurance Payers Overview

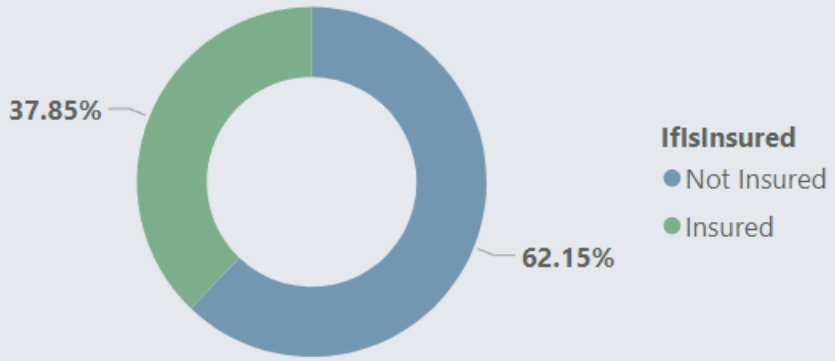
Hospital Visits

Insurance

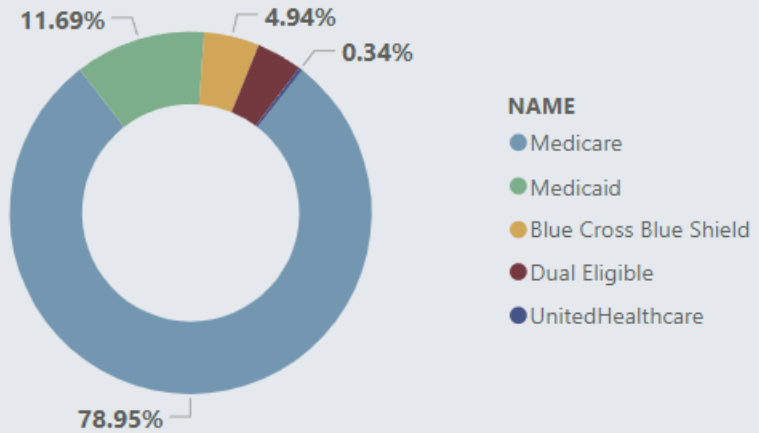
COVID-19

Total Procedures Done	Total Insured Procedures	Total Procedures Revenue	Avg. Cost per Procedure
48K	25K	$2.04K	$97.53M

Ratio of Insured – Not Insured Encounters



Top 5 Insurance Payers by Procedure



Massachusetts General Hospital - COVID-19 Impact

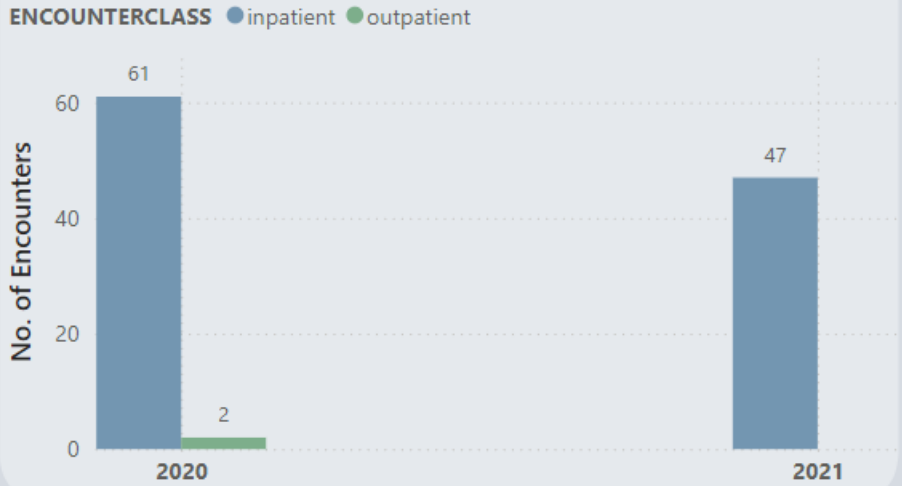
Hospital Visits

Insurance

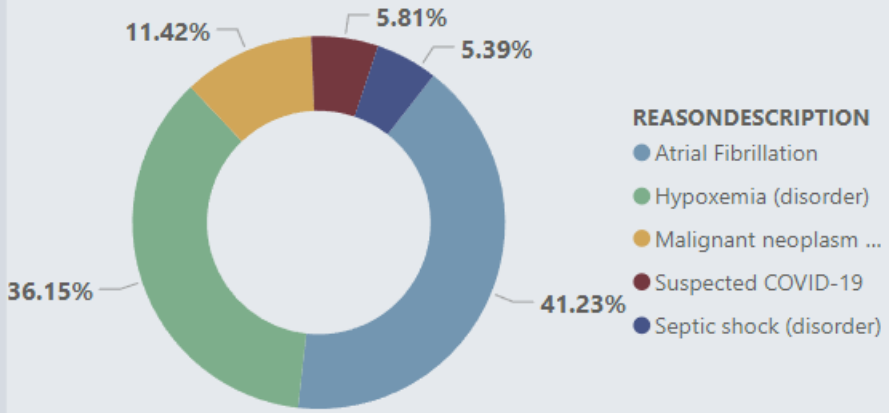
COVID-19

No. of Admitted Patients	No. of Readmitted Patients	Avg. Inpatient Stay Period	Morality Rate
127	62	57 hr(s)	16%

Inpatient – Outpatient Ratio (Jan. 2020 – Dec. 2021)

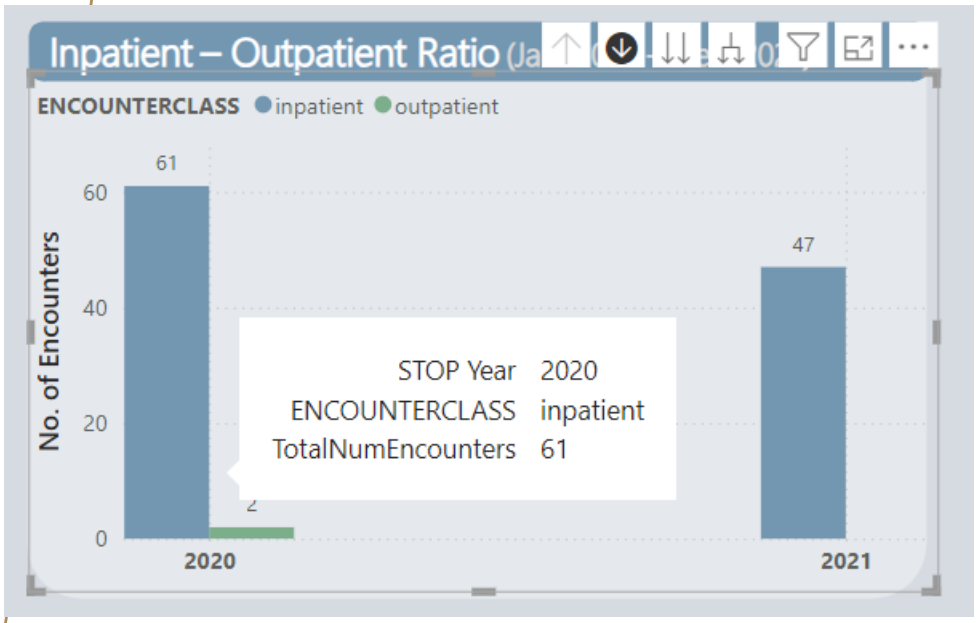


Main Reasons for Procedures

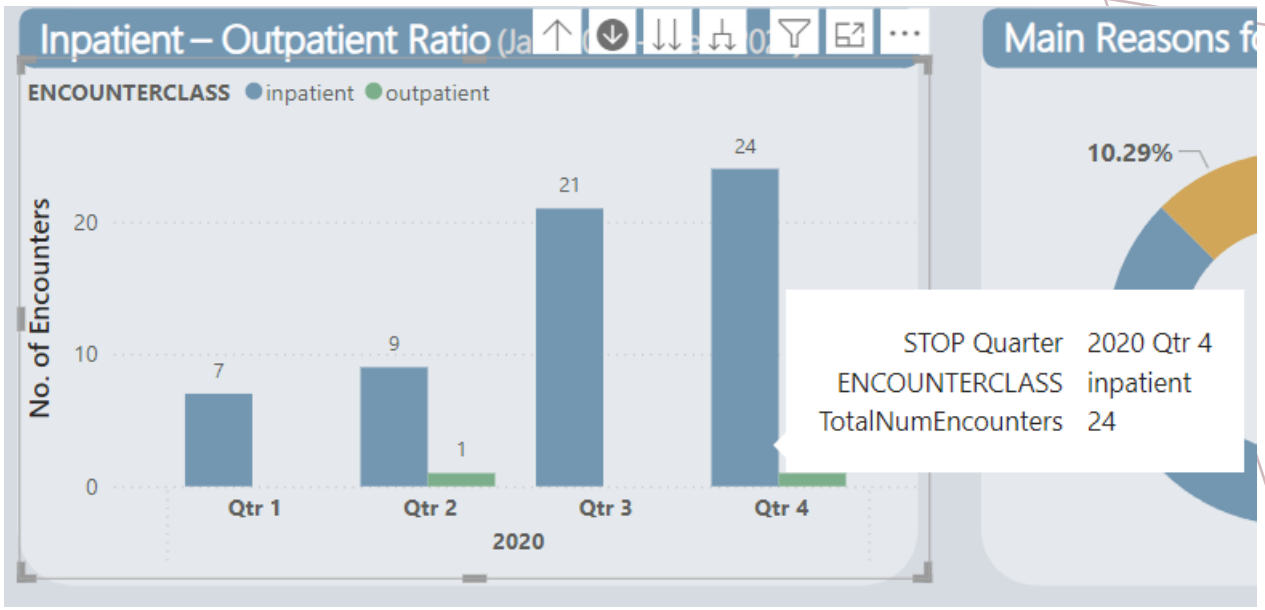
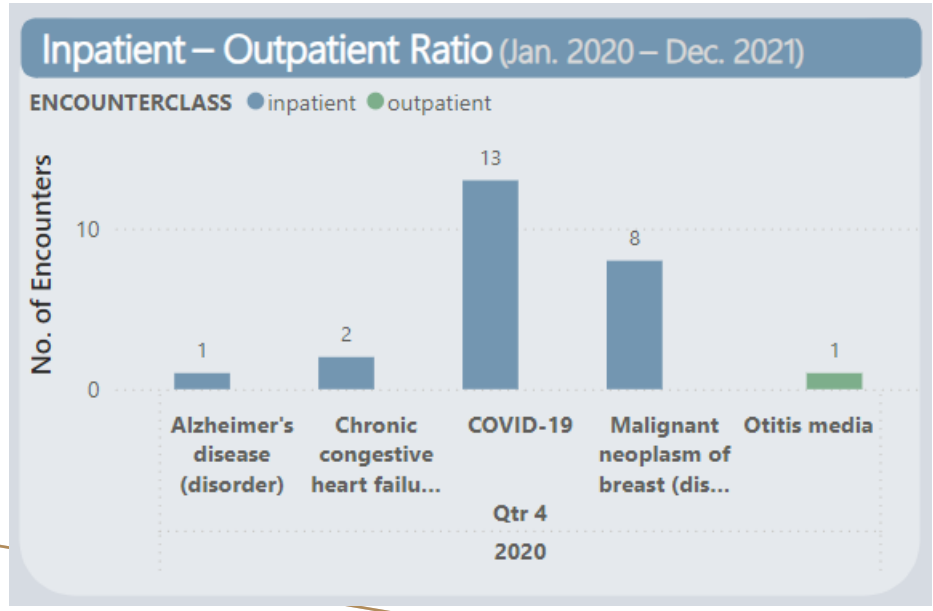


Drill Down / Up option

1



2



3