Healthcare & Patients Management Analysis

Total Patients
55K

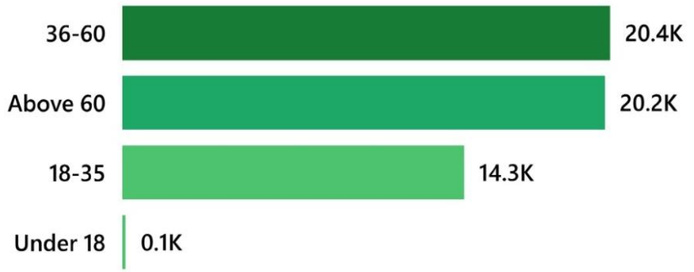
Total Revenue
1bn

Total Rooms
400

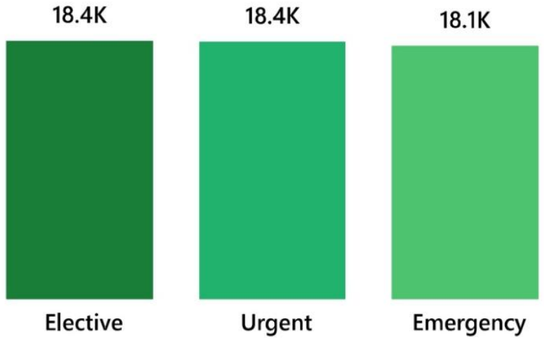
Avg. Billing Per Patient
25.62K

Demographic

Total Patients By Age Group



Total Patients By Admission Type



Total Patients By Gender

Total Patients By Insurance Provider

2019 | 2020 | 2021

2022 | 2023 | 2024

- Arthritis
- Asthma
- Cancer
- Diabetes
- Hypertension
- Obesity

Healthcare & Patients Management Analysis

Total Patients 55K

Total Revenue 1bn

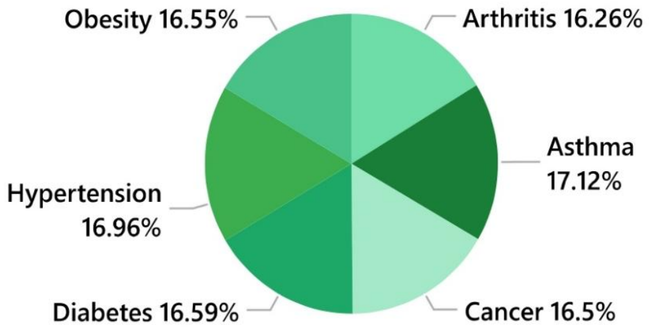
Total Rooms 400

Avg. Billing Per Patient 25.62K

Demographic

Total Revenue By Medical Condition

Recovery Rate By Medical Condition



Avg of Billing Amount By Medical Condition

Total Revenue By Insurance Provider

2019 | 2020 | 2021

2022 | 2023 | 2024

- Arthritis
- Asthma
- Cancer
- Diabetes
- Hypertension
- Obesity

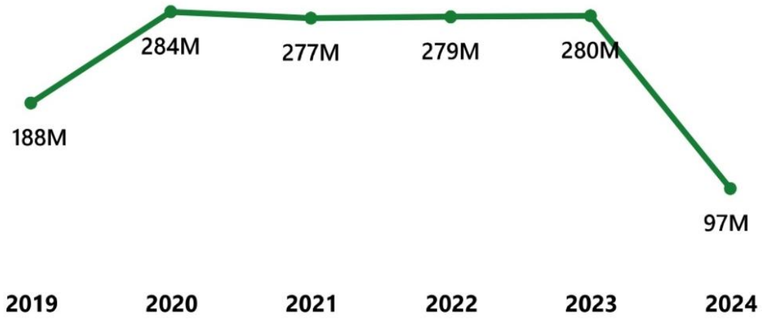
Healthcare & Patients Management Analysis

Demographic

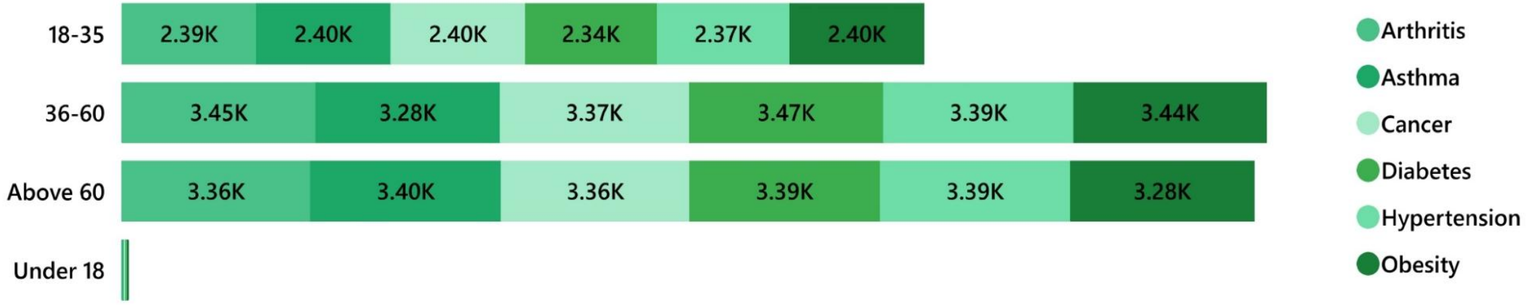
Total Patients	Total Revenue	Total Rooms	Avg. Billing Per Patient
55K	1bn	400	25.62K

Total Patients By Gender & Medical Condition

Total Revenue By Year



Total Patients By Age Group & Medical Condition



2019 | 2020 | 2021

2022 | 2023 | 2024

- Arthritis
- Asthma
- Cancer
- Diabetes
- Hypertension
- Obesity

Global Healthcare And Patient Management Analysis

Shaikh Saifuddin Ahmed

OVERVIEW

Objective

In this project dataset which is used is Global Healthcare And Patient Management Analysis, This project is designed to delve into global healthcare trends, providing a comprehensive overview of patient management, medical conditions, and healthcare costs. By analyzing patterns across various hospitals, regions, and patient demographics, the study aims to uncover insights into medical practices, resource allocation, and financial implications in the healthcare industry.

Dataset Includes:

- Patient Details: Name, Age, Gender, and Blood Type.
- Medical Information: Medical Conditions, Medications, Test Results.
- Admission and Hospital Details: Date of Admission, Discharge Date, Admission Type, Room Number, Hospital Name, and Doctor Name.
- Financial Information: Insurance Provider and Billing Amount

Data Source

This project utilizes an uncleaned dataset from Kaggle, which was preprocessed to extract valuable insights related to healthcare operations, patient demographics, and financial data.

Purpose

To uncover patterns in patient admissions, prevalent medical conditions, treatment protocols, and financial aspects of healthcare. The project focuses on identifying correlations between demographics, medical services, and patient outcomes to support informed decision-making, enhance operational efficiency, and guide healthcare policy improvements worldwide.

Skills Applied

Data Cleaning, Data Transformation with Power Query, Data Visualization techniques in Power Bi, Trend Analysis, Dashboard Design.

Tools Used

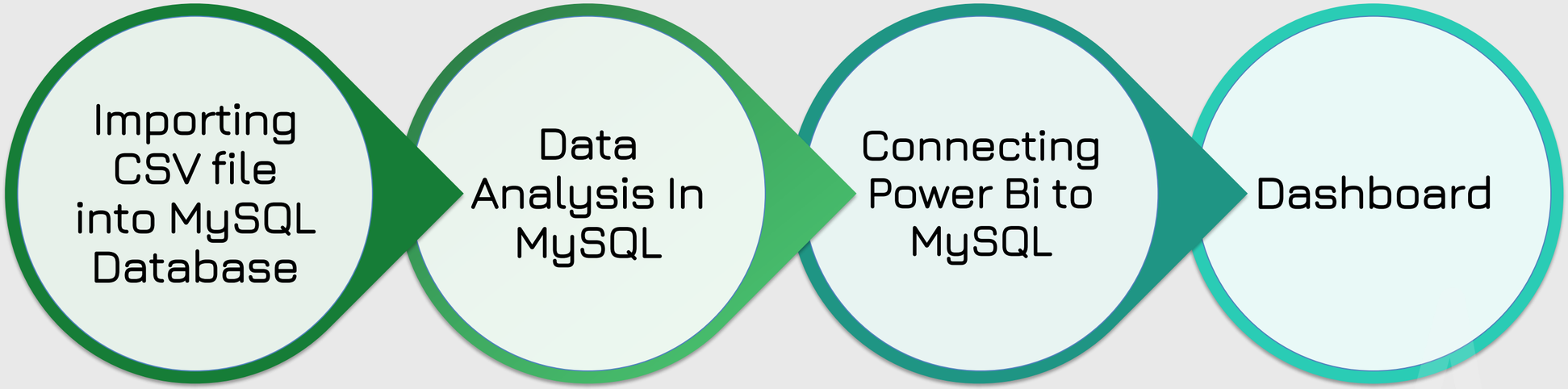
MYSQL

Power Query

Power Bi

Power Point

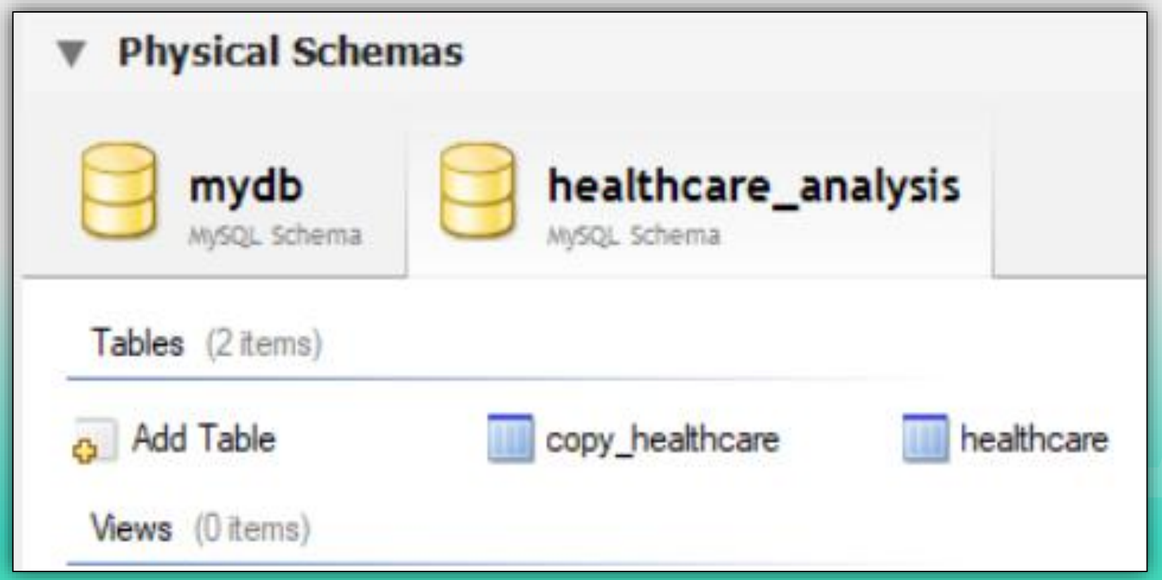
Project Roadmap



Importing CSV file into MYSQL Database

CSV

Creating Database



MYSQL Analysis

- Step 1 : It's a wise practice to make a copy of the original data and work on that copy data, if any mistake happen at least original data will be there.

```
CREATE TABLE copy_healthcare
LIKE healthcare;
```

Use this query to create the same table as original one.

```
INSERT INTO copy_healthcare
SELECT * FROM healthcare;
```

Use this query to insert same data as the original table is having.

Patient_Name	Age	Gender	Blood_Type	Medical_Condition	Date_of_Admission	Doctor_Name	Hospital_Name	Insurance_Provider	Billing_Amount	Room_Number	Admission_Type	Discharge_Date	Medication	Test_Results
Bobby JacksOn	30	Male	B-	Cancer	31-01-2024	Matthew Smith	Sons and Miller	Blue Cross	18856.28131	328	Urgent	02-02-2024	Paracetamol	Normal
LesLie TErRy	62	Male	A+	Obesity	20-08-2019	Samantha Davies	Kim Inc	Medicare	33643.32729	265	Emergency	26-08-2019	Ibuprofen	Inconclusive
DaNnY sMitH	76	Female	A-	Obesity	22-09-2022	Tiffany Mitchell	Cook PLC	Aetna	27955.09608	205	Emergency	07-10-2022	Aspirin	Normal
andrEw waTtS	28	Female	O+	Diabetes	18-11-2020	Kevin Wells	Hernandez Rogers and Vang,	Medicare	37909.78241	450	Elective	18-12-2020	Ibuprofen	Abnormal
adrIENNE bEll	43	Female	AB+	Cancer	19-09-2022	Kathleen Hanna	White-White	Aetna	14238.31781	458	Urgent	09-10-2022	Penicillin	Abnormal
EMILY JOHNSOn	36	Male	A+	Asthma	20-12-2023	Taylor Newton	Nunez-Humphrey	UnitedHealthcare	48145.11095	389	Urgent	24-12-2023	Ibuprofen	Normal
edwArD EDWaRDs	21	Female	AB-	Diabetes	03-11-2020	Kelly Olson	Group Middleton	Medicare	19580.87234	389	Emergency	15-11-2020	Paracetamol	Inconclusive
CHrisTInA MARtinez	20	Female	A+	Cancer	28-12-2021	Suzanne Thomas	Powell Robinson and Valdez,	Cigna	45820.46272	277	Emergency	07-01-2022	Paracetamol	Inconclusive
JASmINe aGuIlaR	82	Male	AB+	Asthma	01-07-2020	Daniel Ferguson	Sons Rich and	Cigna	50119.22279	316	Elective	14-07-2020	Aspirin	Abnormal
ChRISTopher BerG	58	Female	AB-	Cancer	23-05-2021	Heather Day	Padilla-Walker	UnitedHealthcare	19784.63106	249	Elective	22-06-2021	Paracetamol	Inconclusive
mIchElLe daniELs	72	Male	O+	Cancer	19-04-2020	John Duncan	Schaefer-Porter	Medicare	12576.79561	394	Urgent	22-04-2020	Paracetamol	Normal

➢ Step 2 : Removing duplicates records

In this dataset didn't have a unique identifier like a serial number, So I used the ROW_NUMBER window function along with a CTE to detect and eliminate duplicates.

```
WITH Row_Num_Cte AS (SELECT *,ROW_NUMBER()
OVER(PARTITION BY Patient_Name,Age,Gender,Blood_Type,Medical_Condition,
      Date_of_Admission,Doctor_Name,Hospital_Name,Insurance_Provider,
      Billing_Amount,Room_Number,Admission_Type,Discharge_Date,Medication,
      Test_Results) AS Row_Num
      FROM copy_healthcare)
      SELECT * FROM Row_Num_cte
      WHERE Row_Num > 1;
```

Patient_Name	Age	Gender	Blood_Type	Medical_Condition	Date_of_Admission	Doctor_Name	Hospital_Name	Insurance_Provider	Billing_Amount	Room_Number	Admission_Type	Discharge_Date	Medication	Test_Results	Row_Num
ABIgaIL YOung	41	Female	O+	Hypertension	15-12-2022	Edward Kramer	Moore-Mcdaniel	UnitedHealthcare	1983.568297	192	Elective	13-01-2023	Ibuprofen	Normal	2
AdAM tHOmaS	75	Male	O+	Hypertension	02-01-2022	Beverly Miller	Sanchez-Wilson	Cigna	16294.90772	180	Urgent	30-01-2022	Ibuprofen	Normal	2
aLEX BLAck	51	Male	O+	Diabetes	27-03-2022	Frank Williams	Short and Sons	Aetna	39525.6749	246	Emergency	21-04-2022	Aspirin	Normal	2
AlEXAndra sANTIAgO	53	Male	B-	Arthritis	17-01-2024	Donald Solis	and White, Wall Hurst	UnitedHealthcare	4177.715918	450	Urgent	15-02-2024	Penicillin	Normal	2
alFRED CERvAntEs	65	Male	AB-	Diabetes	24-12-2023	Linda Hall	Dorsey Ltd	Aetna	1903.879223	385	Elective	04-01-2024	Paracetamol	Inconclusive	2
aLiCE MoRRIs	80	Male	AB-	Arthritis	03-03-2021	Anthony Hawkins	Smith-Mcdowell	Cigna	7150.470789	293	Urgent	21-03-2021	Aspirin	Inconclusive	2
ALIcia taYLoR	78	Male	O+	Asthma	18-09-2022	Dawn Burton	Wright LLC	Aetna	31465.27498	149	Elective	15-10-2022	Aspirin	Inconclusive	2
AllisOn cocHRan	58	Female	B+	Hypertension	03-05-2022	Kevin Owens	Jackson-Hernandez	Cigna	28575.32191	234	Urgent	28-05-2022	Aspirin	Inconclusive	2
aLlisON MccIAiN	33	Male	O-	Diabetes	15-07-2022	Robert Thompson	Garcia, and Reyes Carey	Medicare	43591.38746	388	Emergency	14-08-2022	Lipitor	Normal	2
aLySsA jonES	31	Male	A+	Cancer	08-11-2023	Kevin Brown	and Weaver Morse, Daniel	Blue Cross	32853.04075	124	Urgent	27-11-2023	Ibuprofen	Normal	2

This query indicates that there are 534 duplicate records, which need to be removed.

This SQL Query creates a temporary table temp that stores only unique rows from the copy_healthcare table. It uses SELECT DISTINCT to eliminate duplicates and saves the cleaned data in the temp table, which exists only for the current session.

```
CREATE TEMPORARY TABLE temp AS SELECT DISTINCT * FROM copy_healthcare;
```

This SQL Query delete all rows from copy_healthcare.

```
DELETE FROM copy_healthcare;
```

This SQL Query inserts data from the temp table (which contains distinct rows) back into the copy_healthcare table, restoring the cleaned, duplicate-free data into the original table. This is done after deleting all the data from copy_healthcare.

```
INSERT INTO copy_healthcare SELECT * FROM temp;
```

➢ Step 3 : Standardizing Formats

Check for any spelling errors or extra spaces or datatype of columns.

```
SELECT Patient_Name,LENGTH(Patient_Name) AS Patient_Name_Len,
TRIM(Patient_Name) AS Correct_Patient_Name,LENGTH(TRIM(Patient_Name)) Correct_Trim_Len
FROM copy_healthcare
ORDER BY Patient_Name;
```

Patient_Name	Patient_Name_Len	Correct_Patient_Name	Correct_Trim_Len
aAroN ADaMS	11	aAroN ADaMS	11
aAROn aguIRRe	13	aAROn aguIRRe	13
AArOn AnderSoN	14	AArOn AnderSoN	14
aaRON AndeRSoN md	17	aaRON AndeRSoN md	17
AAron ArCHER	12	AAron ArCHER	12
AAron ArCHER	12	AAron ArCHER	12
AaROn bAkEr	11	AaROn bAkEr	11
AAroN bAker	11	AAroN bAker	11
AARON bAldWIN Jr.	17	AARON bAldWIN Jr.	17
aaROn bARNes	12	aaROn bARNes	12
aaROn bauTISta	14	aaROn bauTISta	14
aarOn bEnNetT	13	aarOn bEnNetT	13

No extra spaces found

```sql
SELECT Patient_Name,CONCAT(
UPPER(LEFT(Patient_Name, 1)),
LOWER(SUBSTRING(Patient_Name, 2))
) AS Correct_Patient_Name
FROM copy_healthcare
ORDER BY Patient_Name;
```

Here these Patient_Name are not written correctly, so we need to be corrected.

Patient_Name	Correct_Patient_Name
aAroN ADaMS	Aaron adams
aAROn aguIRRe	Aaron aguirre
AArOn AnderSoN	Aaron anderson
aaRON AndeRSoN md	Aaron anderson md
AAron ArCHER	Aaron archer
AAron ArCHER	Aaron archer
AaROn bAkEr	Aaron baker
AAroN bAker	Aaron baker
AARON bAldWIN Jr.	Aaron baldwin jr.
aaROn bARNes	Aaron barnes
aaROn bauTISta	Aaron bautista
aarOn bEnNetT	Aaron bennett

Use this Query to Fix it.

```sql
UPDATE copy_healthcare
SET Patient_Name = CONCAT(
UPPER(LEFT(Patient_Name, 1)),
LOWER(SUBSTRING(Patient_Name, 2)));
```

Fixed

Patient_Name	Age	Gender	Blood_Type
Aaron adams	38	Female	O-
Aaron aguirre	36	Male	A-
Aaron anderson	50	Female	A+
Aaron anderson md	20	Female	A-
Aaron archer	47	Female	B-
Aaron archer	49	Female	B-
Aaron baker	84	Female	A+
Aaron baker	73	Male	B+
Aaron baldwin jr.	20	Male	O-
Aaron barnes	85	Male	O-
Aaron bautista	21	Male	B-

	Patient_Name	Age	Gender	Blood_Type	Medical_Condition	Date_of_Admission
▶	Bobby JacksOn	30	Male	B-	Cancer	31-01-2024
	LesLie TErRy	62	Male	A+	Obesity	20-08-2019
	DaNnY sMitH	76	Female	A-	Obesity	22-09-2022
	andrEw waTtS	28	Female	O+	Diabetes	18-11-2020
	adrIENNE bEll	43	Female	AB+	Cancer	19-09-2022
	EMILY JOHNSOn	36	Male	A+	Asthma	20-12-2023
	edwArD EDWaRDs	21	Female	AB-	Diabetes	03-11-2020
	CHrisTInA MARtinez	20	Female	A+	Cancer	28-12-2021
	JASmINe aGuIlaR	82	Male	AB+	Asthma	01-07-2020
	ChRISTopher BerG	58	Female	AB-	Cancer	23-05-2021
	mIchElLe daniELs	72	Male	O+	Cancer	19-04-2020

MySQL stores and processes dates in the YYYY-MM-DD format by default. In this dataset the date store in a different format, like DD-MM-YYYY (e.g., 31-01-2024), MySQL won't recognize it, so we need to convert the date format into YYYY-MM-DD.

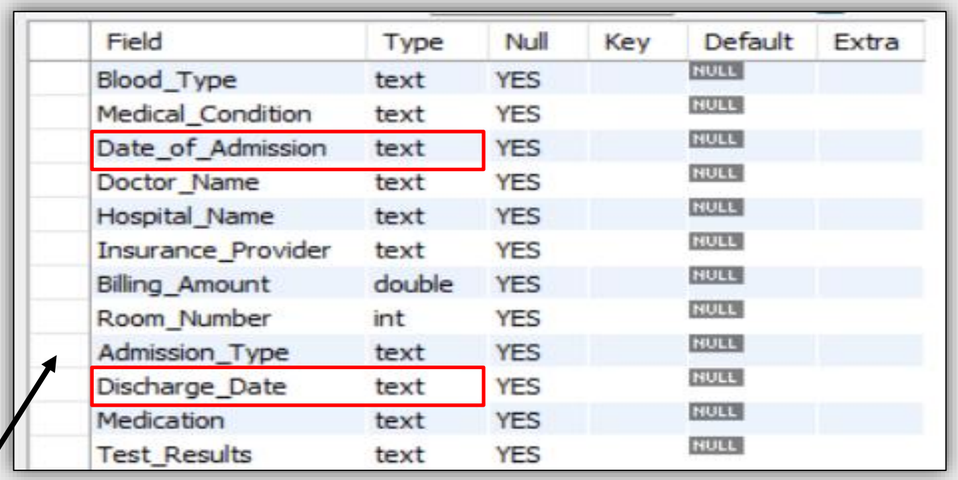
Use this query to Fix it.

```
UPDATE copy_healthcare
SET Date_of_Admission = STR_TO_DATE(Date_of_Admission, '%d-%m-%Y'),
Discharge_Date = STR_TO_DATE(Discharge_Date, '%d-%m-%Y');
```

	Patient_Name	Age	Gender	Blood_Type	Medical_Condition	Date_of_Admission
▶	Bobby jackson	30	Male	B-	Cancer	2024-01-31
	Leslie terry	62	Male	A+	Obesity	2019-08-20
	Danny smith	76	Female	A-	Obesity	2022-09-22
	Andrew watts	28	Female	O+	Diabetes	2020-11-18
	Adrienne bell	43	Female	AB+	Cancer	2022-09-19
	Emily johnson	36	Male	A+	Asthma	2023-12-20
	Edward edwards	21	Female	AB-	Diabetes	2020-11-03
	Christina martinez	20	Female	A+	Cancer	2021-12-28
	Jasmine aguilar	82	Male	AB+	Asthma	2020-07-01
	Christopher berg	58	Female	AB-	Cancer	2021-05-23
	Michelle daniels	72	Male	O+	Cancer	2020-04-19

Fixed

```
Desc copy_healthcare;
```



Field	Type	Null	Key	Default	Extra
Blood_Type	text	YES		NULL	
Medical_Condition	text	YES		NULL	
Date_of_Admission	text	YES		NULL	
Doctor_Name	text	YES		NULL	
Hospital_Name	text	YES		NULL	
Insurance_Provider	text	YES		NULL	
Billing_Amount	double	YES		NULL	
Room_Number	int	YES		NULL	
Admission_Type	text	YES		NULL	
Discharge_Date	text	YES		NULL	
Medication	text	YES		NULL	
Test_Results	text	YES		NULL	

Here these two-column datatype is wrong Date_of_Admission and Discharge_Date should be date.

Use this query to Fix it.

```
ALTER TABLE copy_healthcare
MODIFY COLUMN Date_of_Admission DATE,
MODIFY COLUMN Discharge_Date DATE;
```

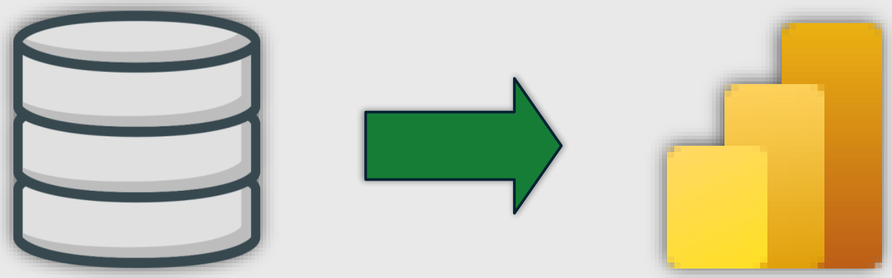
Fixed

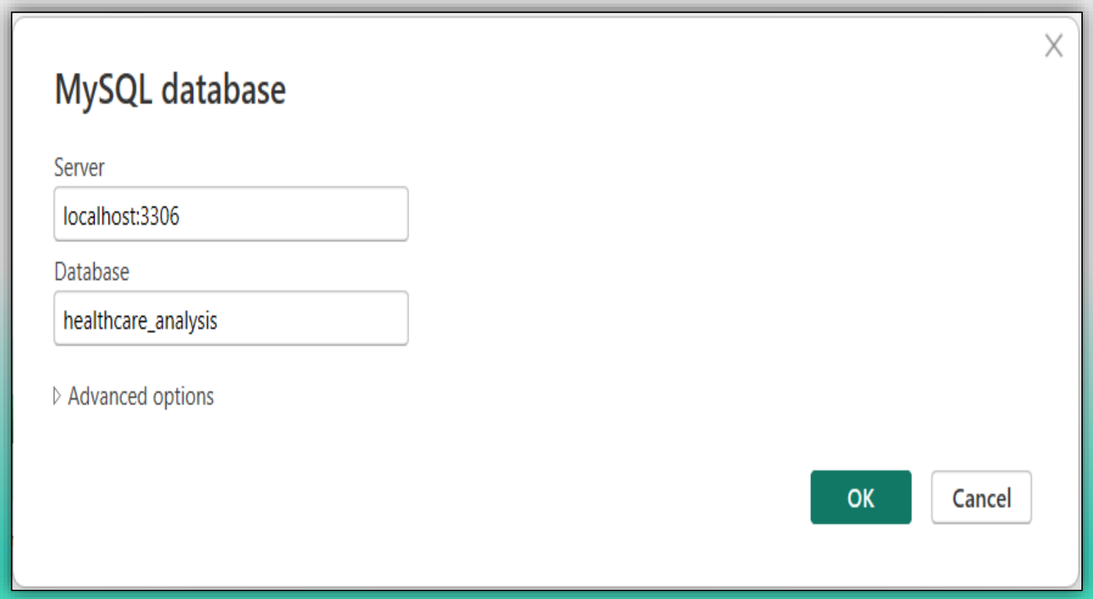
Field	Type	Null	Key	Default	Extra
Blood_Type	text	YES		NULL	
Medical_Condition	text	YES		NULL	
Date_of_Admission	date	YES		NULL	
Doctor_Name	text	YES		NULL	
Hospital_Name	text	YES		NULL	
Insurance_Provider	text	YES		NULL	
Billing_Amount	double	YES		NULL	
Room_Number	int	YES		NULL	
Admission_Type	text	YES		NULL	
Discharge_Date	date	YES		NULL	
Medication	text	YES		NULL	
Test_Results	text	YES		NULL	

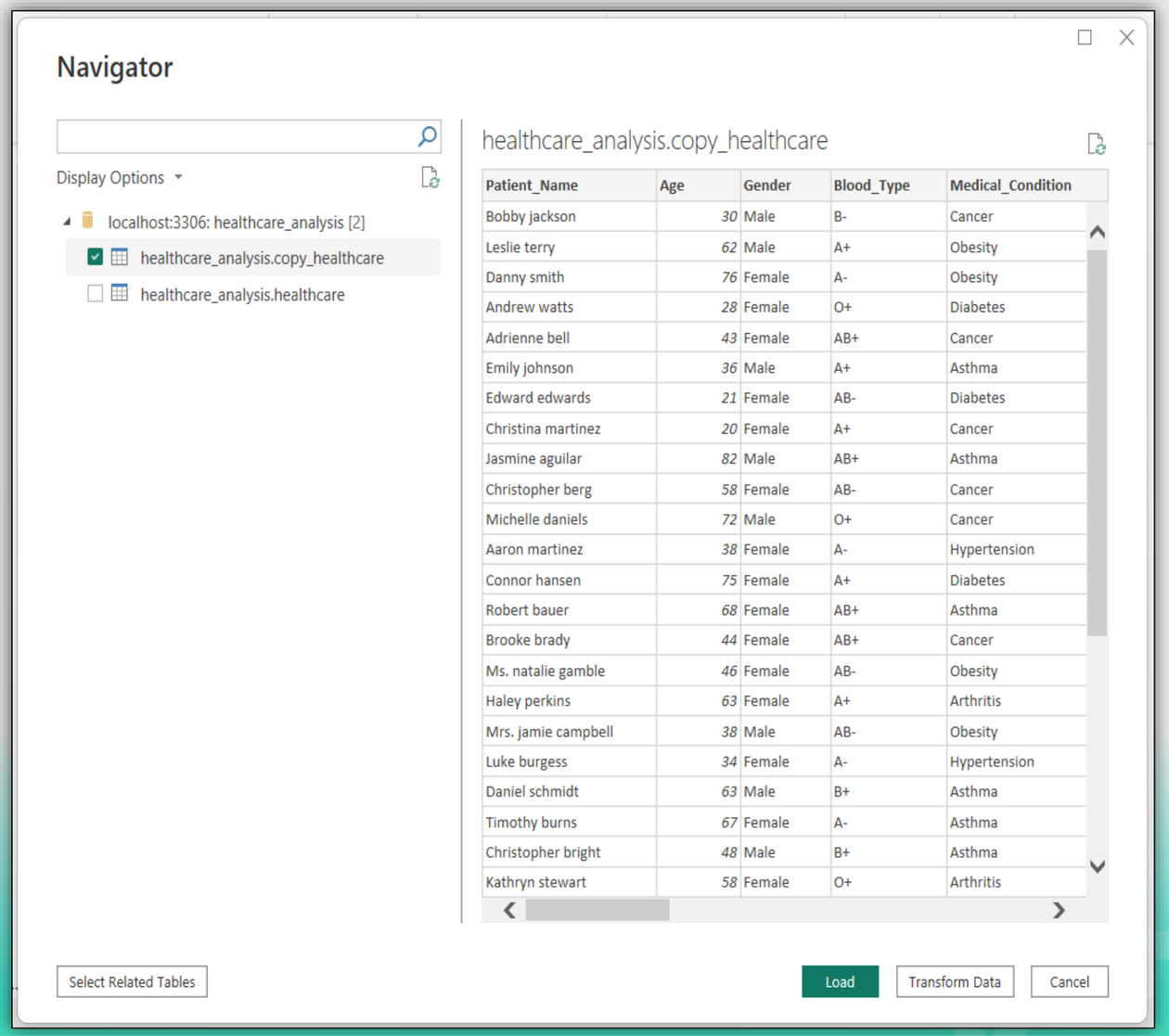
Patient_Name	Age	Gender	Blood_Type	Medical_Condition	Date_of_Admission	Doctor_Name	Hospital_Name	Insurance_Provider	Billing_Amount	Room_Number	Admission_Type	Discharge_Date	Medication	Test_Results
Bobby jackson	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18856.28131	328	Urgent	2024-02-02	Paracetamol	Normal
Leslie terry	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.32729	265	Emergency	2019-08-26	Ibuprofen	Inconclusive
Danny smith	76	Female	A-	Obesity	2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.09608	205	Emergency	2022-10-07	Aspirin	Normal
Andrew watts	28	Female	O+	Diabetes	2020-11-18	Kevin Wells	Hernandez Rogers and Vang,	Medicare	37909.78241	450	Elective	2020-12-18	Ibuprofen	Abnormal
Adrienne bell	43	Female	AB+	Cancer	2022-09-19	Kathleen Hanna	White-White	Aetna	14238.31781	458	Urgent	2022-10-09	Penicillin	Abnormal
Emily johnson	36	Male	A+	Asthma	2023-12-20	Taylor Newton	Nunez-Humphrey	UnitedHealthcare	48145.11095	389	Urgent	2023-12-24	Ibuprofen	Normal
Edward edwards	21	Female	AB-	Diabetes	2020-11-03	Kelly Olson	Group Middleton	Medicare	19580.87234	389	Emergency	2020-11-15	Paracetamol	Inconclusive
Christina martinez	20	Female	A+	Cancer	2021-12-28	Suzanne Thomas	Powell Robinson and Valdez,	Cigna	45820.46272	277	Emergency	2022-01-07	Paracetamol	Inconclusive
Jasmine aguilar	82	Male	AB+	Asthma	2020-07-01	Daniel Ferguson	Sons Rich and	Cigna	50119.22279	316	Elective	2020-07-14	Aspirin	Abnormal
Christopher berg	58	Female	AB-	Cancer	2021-05-23	Heather Day	Padilla-Walker	UnitedHealthcare	19784.63106	249	Elective	2021-06-22	Paracetamol	Inconclusive
Michelle daniels	72	Male	O+	Cancer	2020-04-19	John Duncan	Schaefer-Porter	Medicare	12576.79561	394	Urgent	2020-04-22	Paracetamol	Normal
Aaron martinez	38	Female	A-	Hypertension	2023-08-13	Douglas Mayo	Lyons-Blair	Medicare	7999.58688	288	Urgent	2023-09-05	Lipitor	Inconclusive

During the data cleaning process, 534 duplicate records were removed. Spelling errors in the Patient_Name column were identified and corrected. The dates in the Date_of_Admission and Discharge_Date columns, initially stored in the DD-MM-YYYY format (e.g., 31-01-2024), were corrected. Additionally, the datatypes for the Date_of_Admission and Discharge_Date columns were updated. The dataset is now ready for further in-depth analysis.

Connecting Power Bi to MYSQL







Patient_Name	Age	Gender	Blood_Type	Medical_Condition
Bobby jackson	30	Male	B-	Cancer
Leslie terry	62	Male	A+	Obesity
Danny smith	76	Female	A-	Obesity
Andrew watts	28	Female	O+	Diabetes
Adrienne bell	43	Female	AB+	Cancer
Emily johnson	36	Male	A+	Asthma
Edward edwards	21	Female	AB-	Diabetes
Christina martinez	20	Female	A+	Cancer
Jasmine aguilar	82	Male	AB+	Asthma
Christopher berg	58	Female	AB-	Cancer
Michelle daniels	72	Male	O+	Cancer
Aaron martinez	38	Female	A-	Hypertension
Connor hansen	75	Female	A+	Diabetes
Robert bauer	68	Female	AB+	Asthma
Brooke brady	44	Female	AB+	Cancer
Ms. natalie gamble	46	Female	AB-	Obesity
Haley perkins	63	Female	A+	Arthritis
Mrs. jamie campbell	38	Male	AB-	Obesity
Luke burgess	34	Female	A-	Hypertension
Daniel schmidt	63	Male	B+	Asthma
Timothy burns	67	Female	A-	Asthma
Christopher bright	48	Male	B+	Asthma
Kathryn stewart	58	Female	O+	Arthritis

Data Transformation

Queries [1]

healthcare_analysis copy...

```
= Table.TransformColumnTypes(#"Rounded Off",{{"Billing_Amount", Int64.Type}})
```

	Patient_Name	Age	Gender	Blood_Type	Medical_Condition	Date_of_Admission	Doctor_Nam
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid Error Empty
1	Bobby Jackson	30	Male	B-	Cancer	31-01-2024	Matthew Smi
2	Leslie Terry	62	Male	A+	Obesity	20-08-2019	Samantha Da
3	Danny Smith	76	Female	A-	Obesity	22-09-2022	Tiffany Mitch
4	Andrew Watts	28	Female	O+	Diabetes	18-11-2020	Kevin Wells
5	Adrienne Bell	43	Female	AB+	Cancer	19-09-2022	Kathleen Han
6	Emily Johnson	36	Male	A+	Asthma	20-12-2023	Taylor Newto
7	Edward Edwards	21	Female	AB-	Diabetes	03-11-2020	Kelly Olson
8	Christina Martinez	20	Female	A+	Cancer	28-12-2021	Suzanne Thor
9	Jasmine Aguilar	82	Male	AB+	Asthma	01-07-2020	Daniel Fergus
10	Christopher Berg	58	Female	AB-	Cancer	23-05-2021	Heather Day
11	Michelle Daniels	72	Male	O+	Cancer	19-04-2020	John Duncan
12	Aaron Martinez	38	Female	A-	Hypertension	13-08-2023	Douglas May
13	Connor Hansen	75	Female	A+	Diabetes	12-12-2019	Kenneth Fletc
14	Robert Bauer	68	Female	AB+	Asthma	22-05-2020	Theresa Free
15	Brooke Brady	44	Female	AB+	Cancer	08-10-2021	Roberta Stew
16	Ms. Natalie Gamble	46	Female	AB-	Obesity	01-01-2023	Maria Dough
17	Haley Perkins	63	Female	A+	Arthritis	23-06-2020	Erica Spencer
18	Mrs. Jamie Campbell	38	Male	AB-	Obesity	08-03-2020	Justin Kim
19	Luke Burgess	34	Female	A-	Hypertension	04-03-2021	Justin Moore
20	Daniel Schmidt	63	Male	B+	Asthma	15-11-2022	Denise Gallov
21	Timothy Burns	67	Female	A-	Asthma	28-06-2023	Krista Smith
22	Christopher Bright	48	Male	B+	Asthma	21-01-2020	Gregory Smit
23	Kathryn Stewart	58	Female	O+	Arthritis	12-05-2022	Vanessa New
24	Dr. Eileen Thompson	59	Male	A+	Asthma	02-08-2021	Donna Martin
25							

Query Settings

PROPERTIES

Name

healthcare_analysis copy_healthcare

All Properties

APPLIED STEPS

- Source
- Navigation
- Capitalized Each Word
- Rounded Off
- Changed Type

Data Transformation in Power Query Editor

- Previously, the patient names in the Patient_Name column were not written properly, so I updated them to capitalize each word.
- Previously, the billing amounts in the Billing_Amount column were written in decimals, so I updated them to Rounded off the value.
- Previously, the billing amounts in the Billing_Amount column the datatype were stored as decimals, so I updated their datatype to Whole Number.
- Extracted Day_Name_of_Admission, Month_Name_of_Admission & Year_of_Admission from Date_of_Admission column.

Data Transformation in Power Bi

- I created a New column to uniquely identify patients by concatenating their name, age, and gender, using a DAX function in Power BI, as the dataset did not have a unique identifier column.

- I created a New column, Age Group, using DAX function to categorize individuals into four groups: Under 18, 18-35, 36-60, and Above 60, based on the age column provided in the dataset.

- I created a New column, Readmission Within 30 Days, using a DAX function to identify patients who were readmitted within 30 days.

- Measures were created such as Total Revenue, Total Patients, Total Room, Avg. Billing Per Patient using a Dax Function which used was (Sum of total billing amount) for Total Revenue, (Count Distinct of Patients Id) for Total Patients, (Count Distinct of Room No.) for Total Room, ([Total Revenue] / [Total Patients]) for Avg. Billing Per Patient.

Data Ready For Analysis

Patient_Name	Age	Gender	Blood_Type	Medical_Condition	Date_of_Admission	Doctor_Name	Hospital_Name	Insurance_Provider	Billing_Amount	Room_Number
Eric Stewart	79	Female	O+	Diabetes	14-04-2023	Virginia Lynch	Duffy-Kelly	Aetna	3872	
Amanda Ponce	39	Male	AB-	Arthritis	06-07-2020	Christopher Merritt	Reid Inc	Aetna	23160	
Ann Duncan	24	Male	A-	Asthma	19-03-2024	Daniel Sanders	Inc Cherry	Aetna	18932	
Roy Jones	47	Female	AB-	Asthma	16-02-2023	David Stevens	and Harper Sons	Aetna	6252	
Christopher Weiss	49	Female	AB-	Asthma	16-02-2023	Kelly Thompson	Hunter-Hughes	Aetna	-1018	
Eduardo Houston	78	Male	AB+	Asthma	25-11-2022	Natalie Howell	Sons and Stephenson	Aetna	21875	
Dwayne Cole	37	Male	O-	Obesity	10-03-2022	John Day	Roberts-Sexton	Aetna	24500	
Holly Dunn	45	Male	O+	Obesity	31-10-2022	Carmen Johnson	PLC Glass	Aetna	19667	
Tommy Gaines	29	Male	B-	Diabetes	24-09-2019	Kathleen Diaz	Thomas-Nielsen	Aetna	19537	
Nancy Lopez	50	Male	O-	Arthritis	22-10-2023	Dr. Denise Diaz	Ltd Shelton	Aetna	7162	
Destiny Taylor	52	Male	A-	Diabetes	05-06-2023	Scott Baker	Bailey Inc	Aetna	3814	
Jill Munoz	78	Female	B+	Obesity	01-07-2020	Dennis Carter	Harris, White Benson and	Aetna	7953	
Donald Chase	34	Male	B+	Asthma	28-06-2020	Jeremy Rodriguez	Faulkner-Jones	Aetna	29307	
Alicia Morris	24	Male	A+	Arthritis	24-06-2019	Emily Wood	Sanchez-Campos	Aetna	11083	
Daniel Casey	29	Male	A+	Arthritis	01-02-2020	Charles Johnson	Jennings Jensen and Waters,	Aetna	38273	
Karen Duncan	67	Male	O-	Cancer	25-06-2019	Jamie Jensen	Schmitt Ltd	Aetna	35834	
Jeremy West	18	Female	A-	Asthma	20-09-2021	Phillip Brown	Brown Baker, Phillips and	Aetna	18814	
Alexis Lopez	57	Female	O+	Hypertension	04-12-2022	John Mathis	PLC Harrison	Aetna	21696	
Jessica Malone	66	Male	B+	Arthritis	31-12-2019	Carrie Hernandez	Inc Wyatt	Aetna	27813	
David Morgan	78	Male	AB+	Cancer	21-05-2020	Kelly Wilson	Farmer-Rogers	Aetna	15646	
Logan Horton	33	Female	B-	Obesity	24-07-2020	Jacqueline Thompson	and Scott Ward, Davis	Aetna	35865	
Sharon Franklin	64	Female	O+	Arthritis	22-08-2022	Fernando Gamble	Norton-White	Aetna	29688	
Maria Fleming	45	Female	AB+	Obesity	13-02-2023	Alice Green	Mendoza-Lester	Aetna	16343	
Michelle Skinner	35	Male	B+	Hypertension	18-03-2021	Christina Dyer	Reed Ltd	Aetna	6584	
Laura Cook	43	Male	AB-	Arthritis	14-03-2022	Lisa Hernandez	Pierce-Rosales	Aetna	25939	
Sheri Zamora	85	Male	A+	Arthritis	07-03-2024	Rebecca West	Davis Wood Kemp, and	Aetna	16075	
Tiffany Gonzales	35	Male	AB+	Obesity	06-01-2024	Brittney Gardner	Wilson and Morales Mccormick,	Aetna	4438	

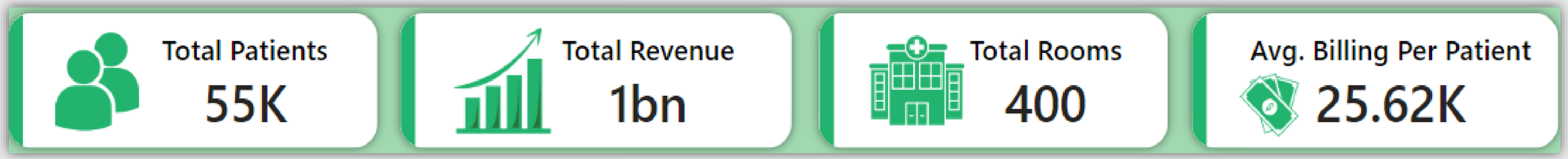
Data

Search

healthcare_analysis copy_...

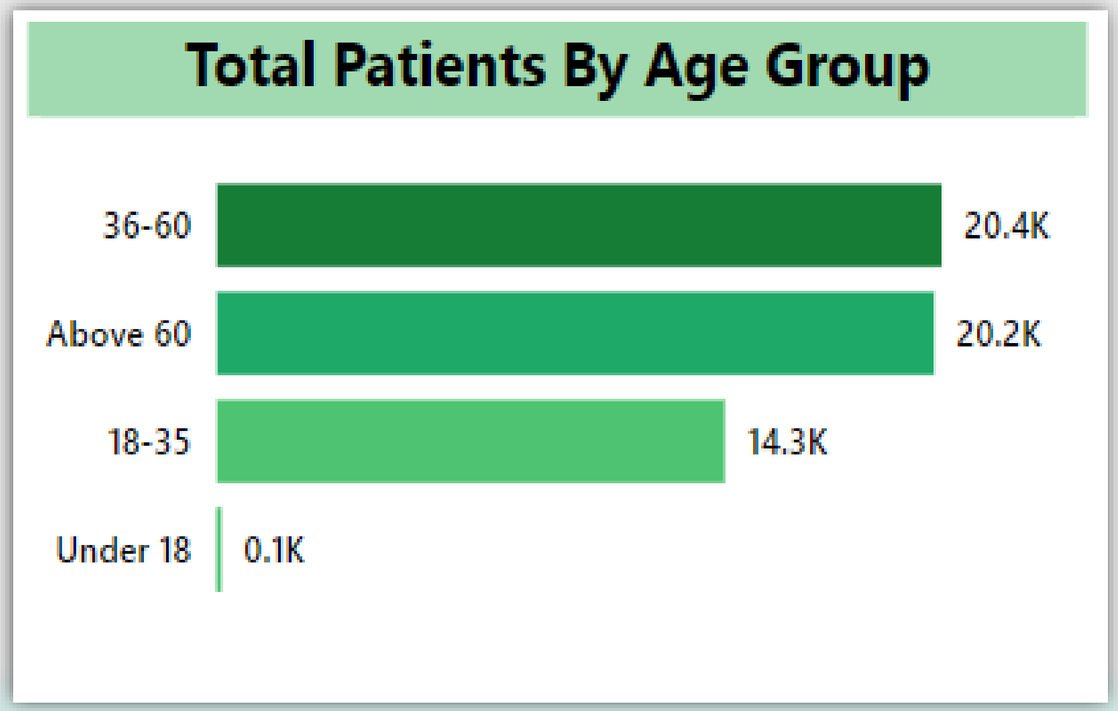
- Admission_Type
- Age
- Age Group
- Avg. Billing Per Patient
- Billing_Amount
- Blood_Type
- Date_of_Admission
- Day_Name_of_Admission
- Discharge_Date
- Doctor_Name
- Gender
- Hospital_Name
- Insurance_Provider
- Medical Condition Count
- Medical_Condition
- Medication
- Month_Name_of_Admission
- Patient Turnover Rate
- Patient_Name
- Patients_Id
- Readmission Within 30 Days
- Readmissions Count

Insights From KPI's



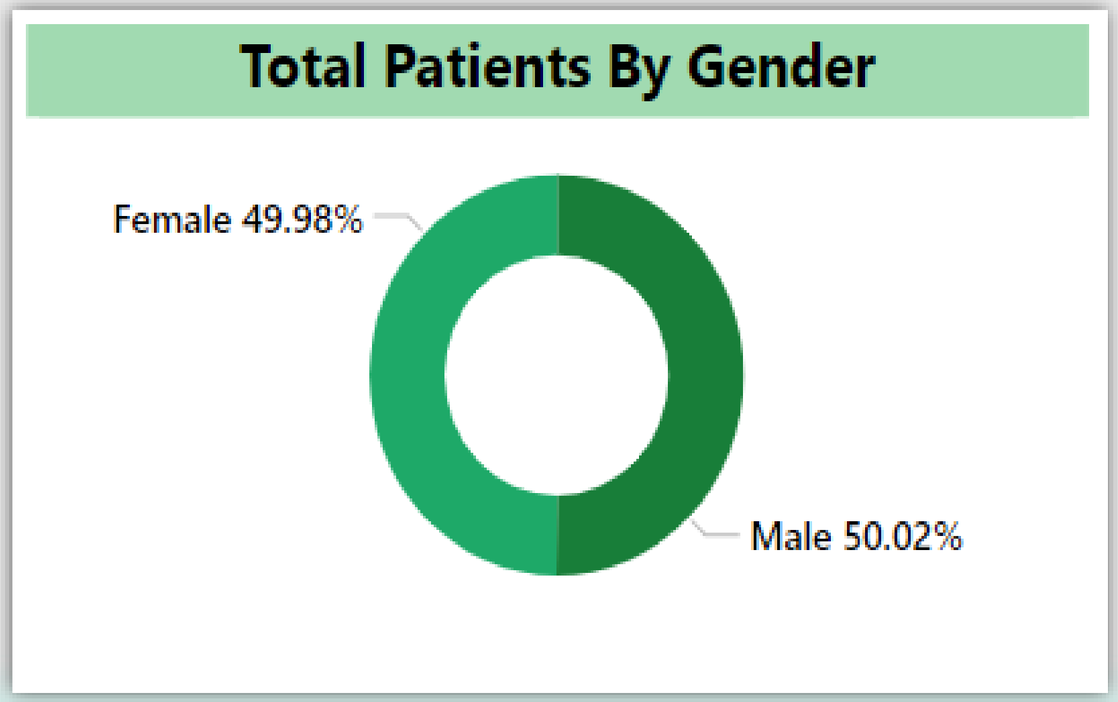
- The facility has served 55,000 Total Patients, indicating a large patient base and demand for healthcare services.
- The Total Revenue generated is 1 billion, reflecting strong financial performance and high service utilization.
- The facility has 400 Total Rooms, which suggests its capacity to handle a considerable number of inpatients and medical procedures.
- The Average Billing Per Patient is 25.62K, highlighting the average revenue generated per patient, which could reflect the pricing structure or the types of services provided.

Total Patients By Age Group



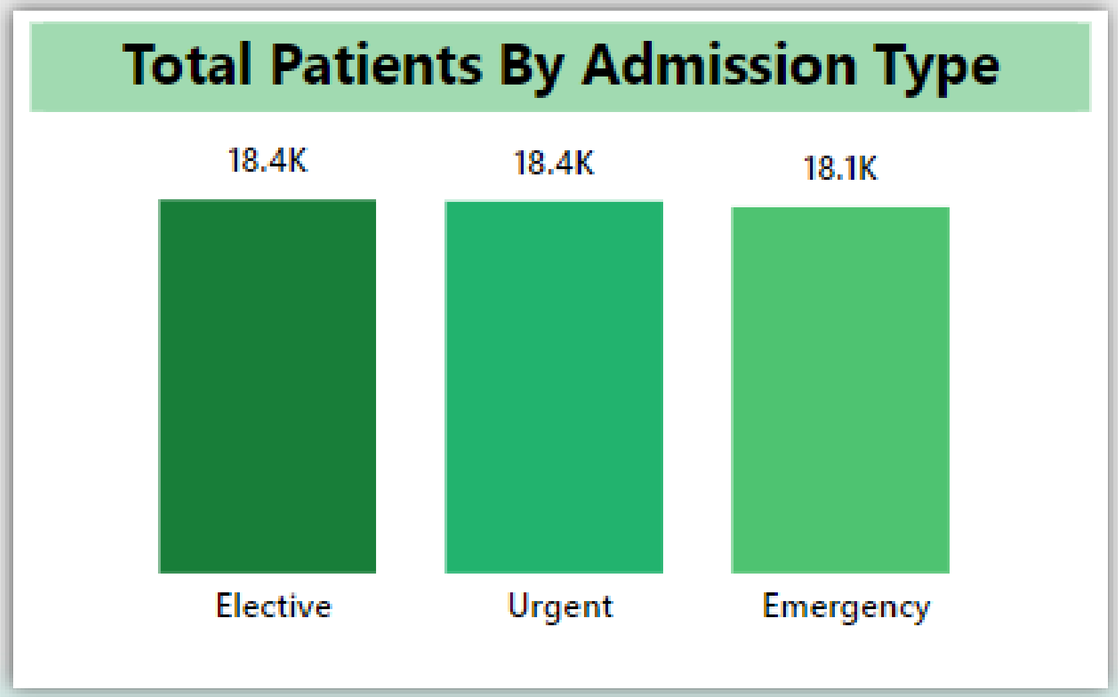
The age group with the highest number of patients is 36-60 years, accounting for 20.4K individuals, followed closely by patients above 60 years, who make up 20.2K. This indicates that middle-aged and elderly individuals are the primary users of healthcare services. In contrast, the 18-35 age group has significantly fewer patients at 14.3K, suggesting that younger adults may have fewer healthcare needs. Patients under 18 represent an extremely small proportion, with just 0.1K recorded, possibly reflecting either fewer health issues or specialized care elsewhere for pediatric cases.

Total Patients By Gender



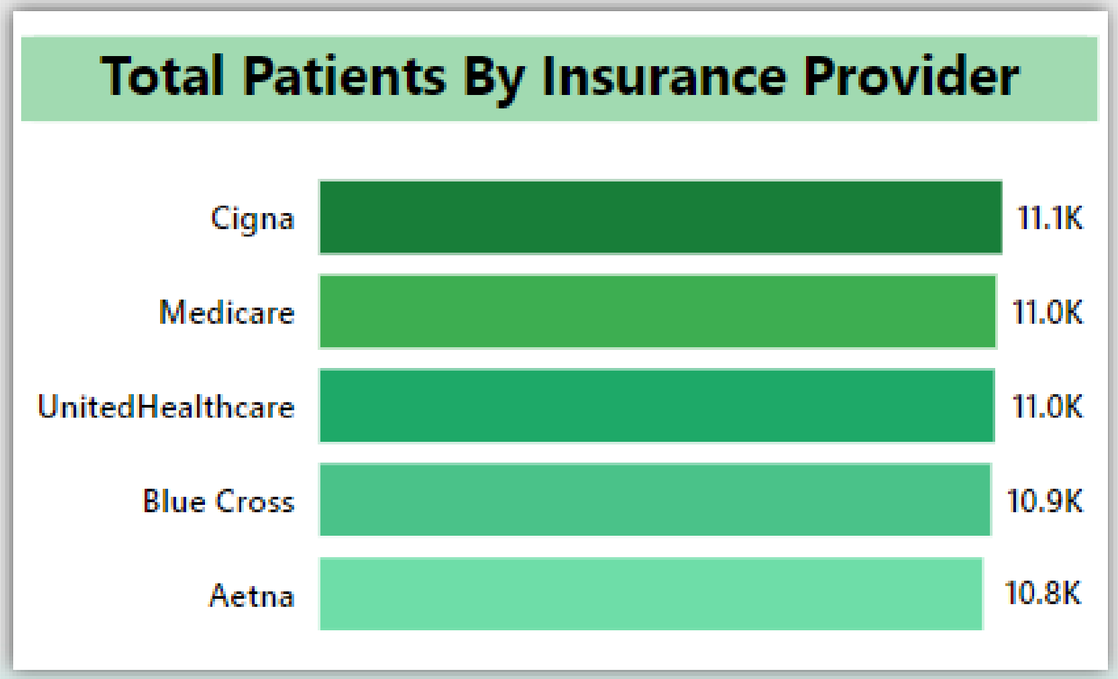
The gender distribution among patients is almost equal, with males slightly edging out at 50.02%, compared to females at 49.98%. This near parity indicates a balanced access or need for healthcare services across genders, showing no significant disparity between male and female patients.

Total Patients By Admission Type



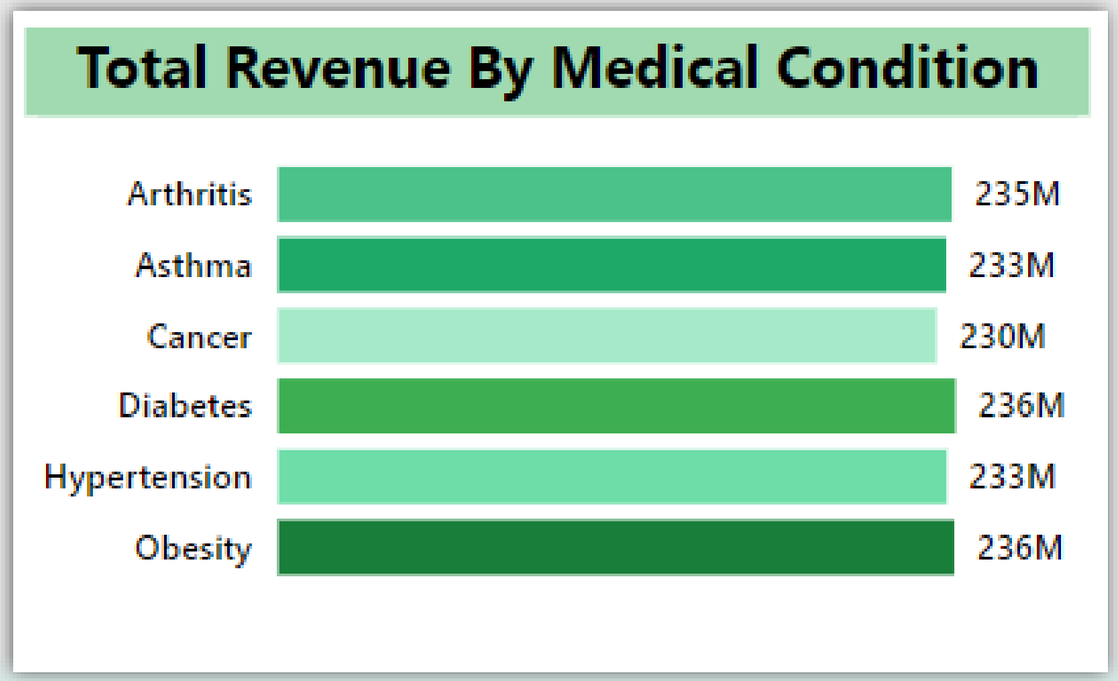
Patients are almost equally distributed across the three admission types. Elective procedures, which include planned treatments like surgeries or diagnostic tests, and urgent cases, which require prompt attention for non-life-threatening conditions, both have 18.4K patients. Emergency cases, involving critical and immediate care needs, follow closely with 18.1K. This balanced distribution suggests consistent demand across all admission types, requiring healthcare facilities to maintain readiness for all scenarios.

Total Patients By Insurance Provider



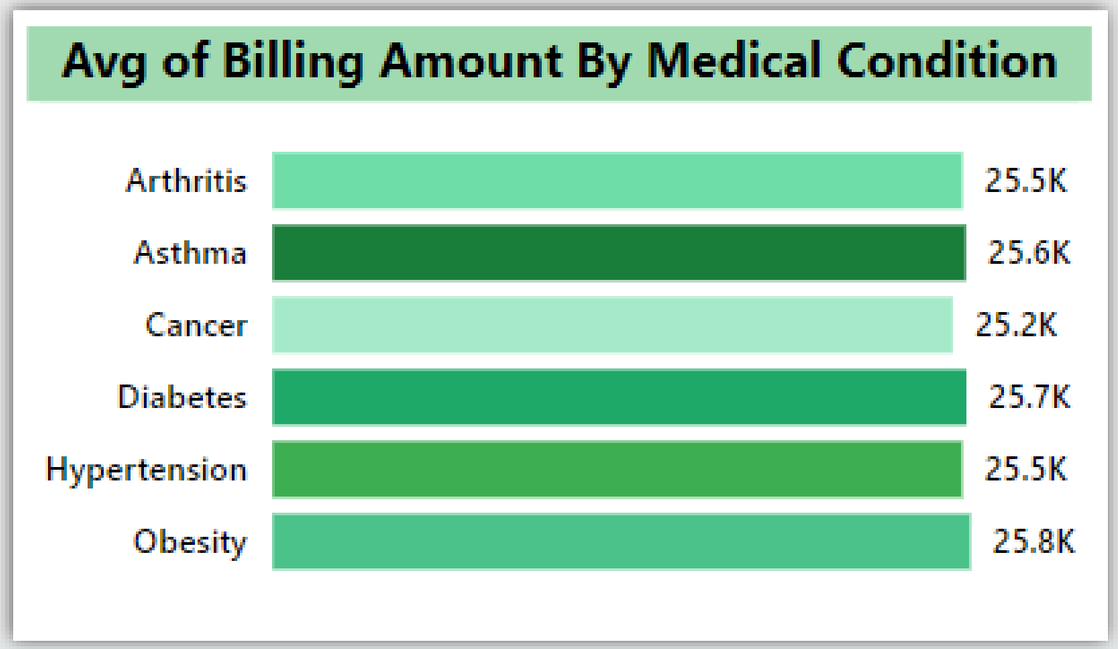
Cigna leads slightly with 11.1K patients, followed closely by Medicare and UnitedHealthcare, both with 11.0K each. Blue Cross is just behind at 10.9K, and Aetna accounts for 10.8K patients. These figures highlight a diverse distribution of patients across major insurance providers, indicating no single provider dominates the market. This diversity suggests the need for hospitals and clinics to accommodate various insurance networks effectively.

Total Revenue By Medical Condition



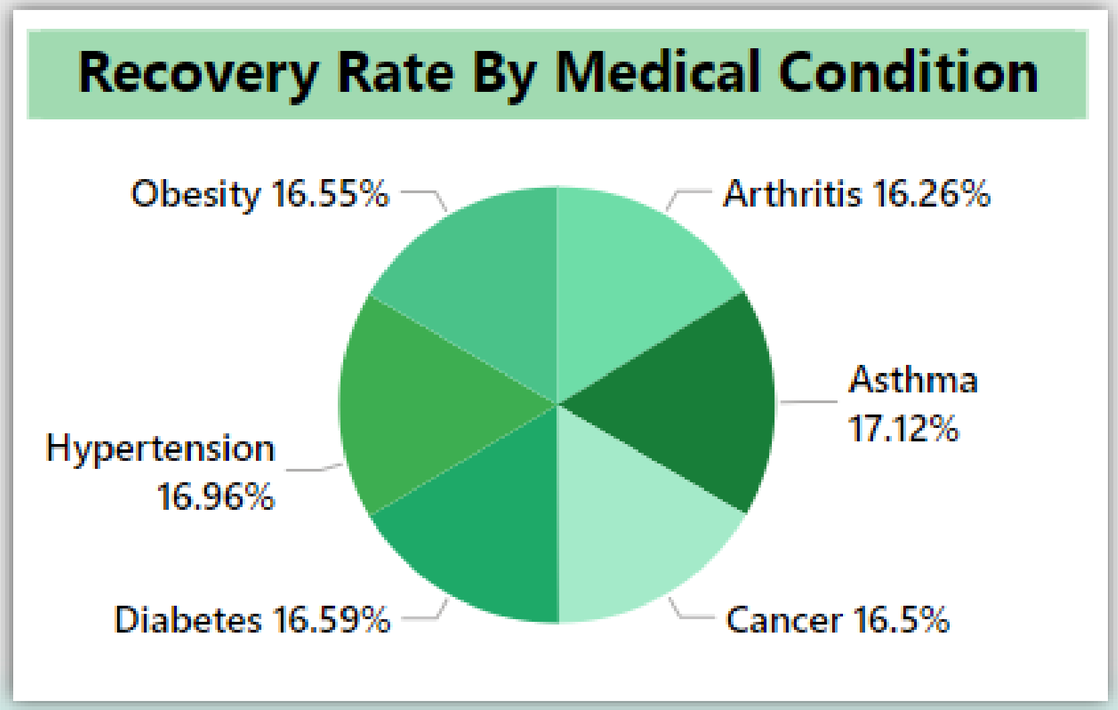
Obesity and diabetes generate the highest revenue, each contributing $236M. This indicates that these conditions might have a high prevalence or involve costly treatments. Following closely, arthritis brings in $235M, highlighting its significant role in revenue generation. Hypertension and asthma each contribute $233M, showing their importance in healthcare revenues. Lastly, cancer generates $230M, which, despite being slightly lower, remains a major source of revenue, likely due to the complexity and duration of its treatments.

Average Billing Amount By Medical Condition



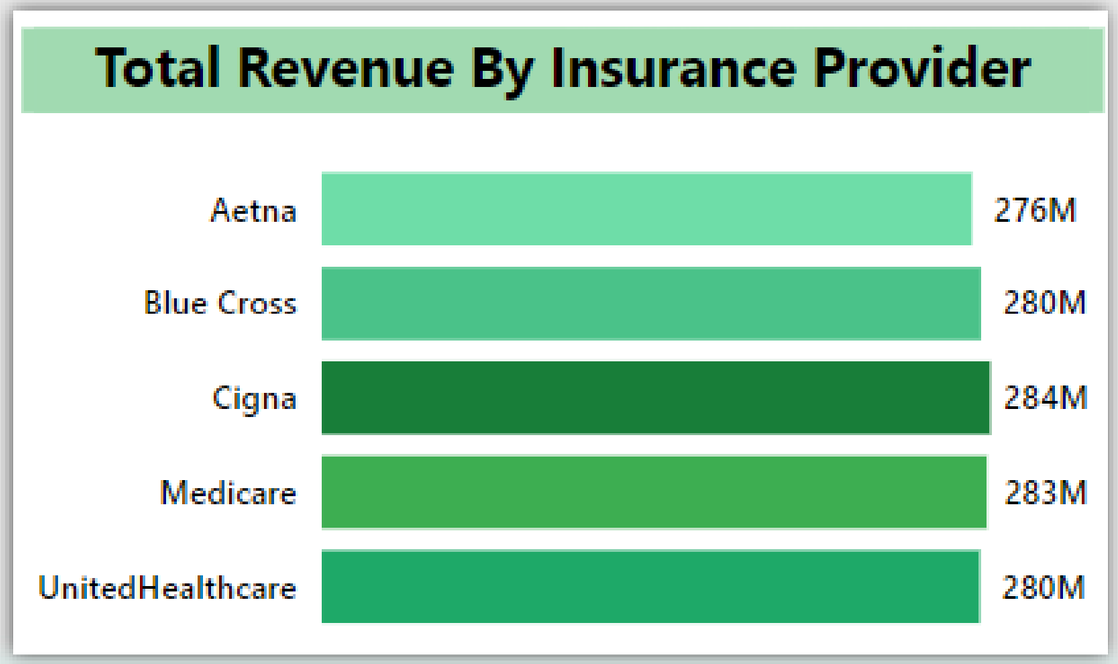
The highest average billing amount is associated with obesity at $25.8K, reflecting the complexity or extended care needed for managing this condition. Diabetes comes next with an average billing of $25.7K, suggesting intensive care protocols. Asthma follows closely with $25.6K, likely due to the frequent need for treatments and medications. Hypertension and arthritis each have an average billing of $25.5K, indicating consistent costs across these conditions. Cancer, at $25.2K, has the lowest average billing amount among the conditions, possibly due to varied treatment plans or insurance coverage.

Recovery Rate By Medical Condition



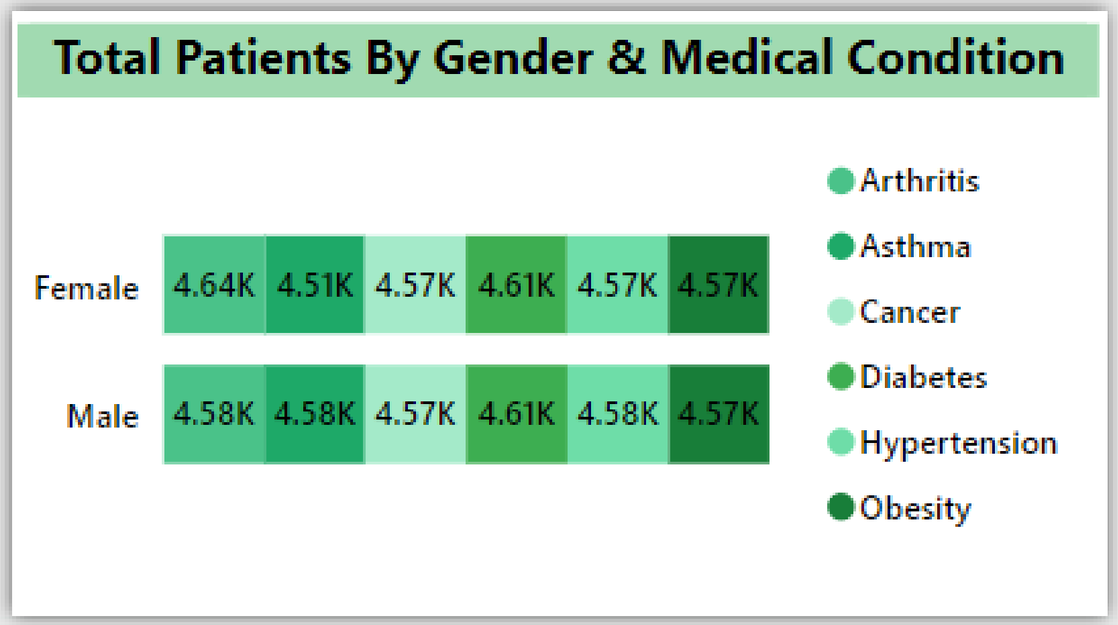
Asthma leads with the highest recovery rate at 17.12%, suggesting effective treatment methods and possibly earlier diagnosis. Hypertension shows a slightly lower recovery rate at 16.96%, reflecting steady progress in managing this condition. Diabetes and obesity have recovery rates of 16.59% and 16.55%, respectively, indicating moderate success in managing these chronic conditions. Cancer and arthritis have the lowest recovery rates at 16.5% and 16.26%, respectively, highlighting the challenges associated with these long-term or complex medical conditions.

Total Revenue By Insurance Provider



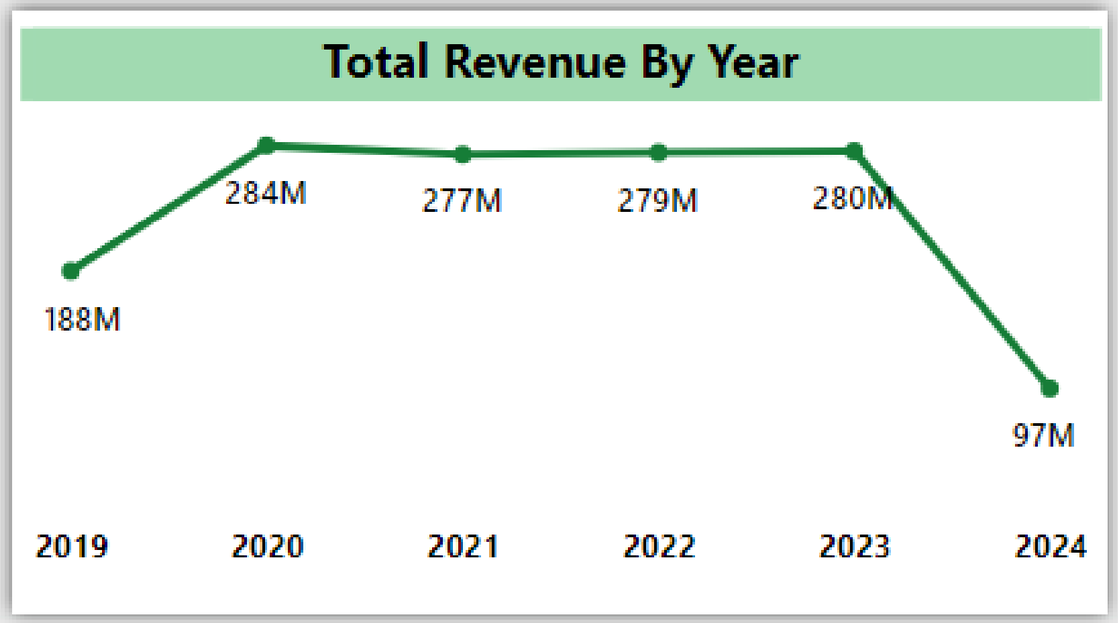
Cigna and Medicare generate the highest total revenues, with $284M and $283M, respectively, indicating their dominance in the healthcare market. Blue Cross and UnitedHealthcare follow closely, each contributing $280M, reflecting their significant market presence. Aetna generates slightly lower revenue at $276M, suggesting a slightly smaller customer base or market share compared to its competitors. These figures demonstrate a relatively balanced distribution of revenue across the major insurance providers, with no single provider overwhelmingly dominating.

Total Patients By Gender & Medical Condition



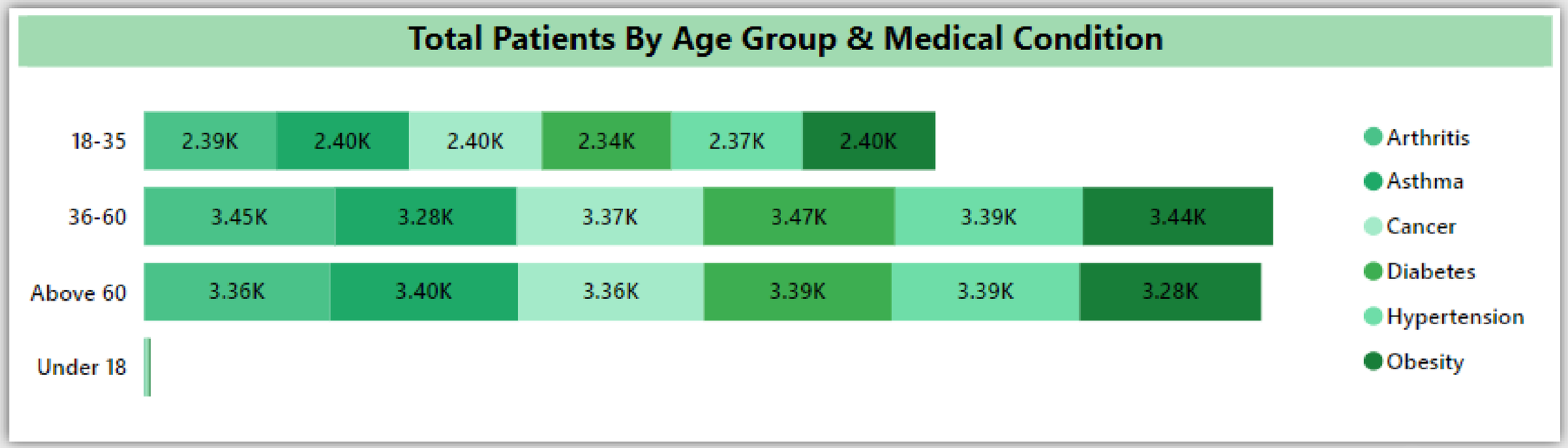
The data shows a relatively equal distribution of patients across genders for all medical conditions. For arthritis, the number of patients is slightly higher for females (4.64K) compared to males (4.58K). A similar trend is observed for asthma, cancer, diabetes, hypertension, and obesity, where female patients slightly outnumber male patients, but the difference is marginal (around 0.1K to 0.2K). This indicates that both genders are equally affected by these medical conditions, with a slight tilt towards females for most cases.

Total Revenue By Year



The trend for total revenue shows significant growth from 2019 to 2023 before a sharp decline in 2024. Revenue increased from $188M in 2019 to its peak at $284M in 2020, followed by a slight dip to $277M in 2021 and recovery to $279M in 2022. The year 2023 also saw a strong performance with $280M, but there was a dramatic drop to $97M in 2024, potentially due to reduced patient inflow, changes in healthcare policies, or other external factors.

Total Patients By Age Group & Medical Condition



The age group 36–60 has the highest number of patients across all medical conditions, with the highest concentration for diabetes (3.47K) and obesity (3.44K). The above 60 age group also has a significant patient population, particularly for arthritis (3.40K) and diabetes (3.39K), indicating the prevalence of these conditions in older individuals. The 18–35 age group has comparatively fewer patients, with asthma (2.40K) and cancer (2.40K) being the most common conditions. For the under 18 group, patient numbers are negligible, suggesting that these conditions are more prevalent in adult and older populations.

Recommendations

- **Target Gender-Inclusive Healthcare:-** Develop balanced healthcare campaigns for both genders, with a slight emphasis on females due to their higher patient numbers for most conditions. Focus on early diagnosis and management of chronic conditions.

- **Address Revenue Decline:-** Investigate the sharp revenue drop in 2024 by analyzing patient inflow, operational inefficiencies, or market shifts. Implement flexible payment plans, expand insurance partnerships, and boost marketing efforts to recover revenue.

- **Focus on Middle-Aged & Senior Patients:-** Prioritize chronic disease management and preventive care for patients aged 36-60 and above 60, as they represent the highest numbers for conditions like diabetes and hypertension. Encourage lifestyle changes for younger patients.

- **Allocate Resources to High-Revenue Conditions:-** Invest more in managing high-revenue conditions such as obesity, diabetes, and hypertension. Focus on advanced treatments and patient retention strategies to capitalize on these conditions.

- **Enhance Partnerships with Insurance Providers:-** Strengthen collaborations with key insurance providers like Cigna, Medicare, and UnitedHealthcare, ensuring better coverage and affordability. Expand insurance network and develop joint marketing campaigns.

Conclusion

The analysis shows key areas for improvement in healthcare and financial performance. Programs should focus on managing chronic conditions for middle-aged and older patients while ensuring gender-inclusive care. Addressing the 2024 revenue drop is critical, with strategies like better financial options, expanded services, and stronger marketing. Conditions like diabetes and hypertension should receive more resources, while recovery rates for cancer and arthritis need improvement. Partnering with top insurance providers and keeping treatment costs affordable will help sustain growth and support patients effectively. These steps will strengthen healthcare impact and financial stability.

Challenges faced throughout this Project

- **Data Quality and Consistency:-** The dataset required significant cleaning to address missing, incomplete, or inconsistent data, ensuring accuracy and reliability in the analysis.

- **Complexity of Healthcare Data:-** The healthcare dataset was highly complex, with multiple variables that needed to be carefully analyzed and interpreted to uncover meaningful insights.

- **Data Integration and Merging:-** Combining different data sources, such as patient demographics, medical conditions, and financial performance, was challenging due to variations in formats and structures.

- **Visualization and Communication:-** Presenting complex insights clearly and effectively through visualizations that could be easily understood by stakeholders was a key challenge.

- **Ethical Considerations:-** Maintaining patient privacy and adhering to ethical standards, including compliance with data protection regulations, was crucial throughout the project.

Thank You