

## Dataset information

### Metadata

#### Potential Analysis Questions for the Healthcare Dataset

## Data Transformation

## Analysis

### Patient Demographics and Characteristics in relation to medical conditions

1. What is the age distribution of patients?
1. What is the gender ratio of patients and its relation with medical conditions?
3. What is the distribution of blood types among patients and different medical conditions?
2. What is the distribution of blood types among patients?

### Disease Prevalence and Patterns

1. What are the most common medical conditions?
2. How has the prevalence of specific diseases changed over time?

### Hospital Utilization

1. What is the average length of stay for different medical conditions?
2. What is the distribution of admission types (emergency, elective, urgent)?
3. How do medical conditions correlate with admission type?

### Financial Analysis

1. What is the total billing amount for different medical conditions?
2. How does the average billing amount vary by insurance provider?
3. What is the distribution of billing amounts?

### Medication Patterns

1. What are the most commonly prescribed medications?
2. Are there any correlations between medications and medical conditions?
3. How does medication usage vary by age or gender?

## Dashboard

# Dataset information

## Healthcare Dataset

Source: <https://www.kaggle.com/datasets/prasad22/healthcare-dataset>

This synthetic healthcare dataset has been created to serve as a valuable resource for data science,

machine learning, and data analysis enthusiasts. It is designed to mimic real-world healthcare data, enabling users to practice, develop, and showcase their data manipulation and analysis skills in the context of the healthcare industry.

The inspiration behind this dataset is rooted in the need for practical and diverse healthcare data for educational and research purposes. Healthcare data is often sensitive and subject to privacy regulations, making it challenging to access for learning and experimentation. To address this gap, Python's Faker library was leveraged to generate a dataset that mirrors the structure and attributes commonly found in healthcare records.

## Metadata

Each column provides specific information about the patient, their admission, and the healthcare services provided, making this dataset suitable for various data analysis and modeling tasks in the healthcare domain. Here's a brief explanation of each column in the dataset -

- **Name:** This column represents the name of the patient associated with the healthcare record.
- **Age:** The age of the patient at the time of admission, expressed in years.
- **Gender:** Indicates the gender of the patient, either "Male" or "Female."
- **Blood Type:** The patient's blood type, which can be one of the common blood types (e.g., "A+", "O-", etc.).
- **Medical Condition:** This column specifies the primary medical condition or diagnosis associated with the patient, such as "Diabetes," "Hypertension," "Asthma," and more.
- **Date of Admission:** The date on which the patient was admitted to the healthcare facility.
- **Doctor:** The name of the doctor responsible for the patient's care during their admission.
- **Hospital:** Identifies the healthcare facility or hospital where the patient was admitted.
- **Insurance Provider:** This column indicates the patient's insurance provider, which can be one of several options, including "Aetna," "Blue Cross," "Cigna," "UnitedHealthcare," and "Medicare."
- **Billing Amount:** The amount of money billed for the patient's healthcare services during their admission. This is expressed as a floating-point number.
- **Room Number:** The room number where the patient was accommodated during their admission.
- **Admission Type:** Specifies the type of admission, which can be "Emergency," "Elective," or "Urgent," reflecting the circumstances of the admission.
- **Discharge Date:** The date on which the patient was discharged from the healthcare facility, based on the admission date and a random number of days within a realistic range.
- **Medication:** Identifies a medication prescribed or administered to the patient during their admission. Examples include "Aspirin," "Ibuprofen," "Penicillin," "Paracetamol," and "Lipitor."
- **Test Results:** Describes the results of a medical test conducted during the patient's admission. Possible values include "Normal," "Abnormal," or "Inconclusive," indicating the outcome of the test.

## Potential Analysis Questions for the Healthcare Dataset

### Patient Demographics and Characteristics

- What is the age distribution of patients?
- What is the gender ratio of patients and its relation with medical conditions?
- What is the distribution of blood types among patients and different medical conditions?
- What is the distribution of blood types among patients?

### Disease Prevalence and Patterns

- What are the most common medical conditions?
- How has the prevalence of specific diseases changed over time?

### Hospital Utilization

- What is the average length of stay for different medical conditions?
- What is the distribution of admission types (emergency, elective, urgent)?
- How do medical conditions correlate with admission type?

### Financial Analysis

- What is the total billing amount for different medical conditions?
- How does the average billing amount vary by insurance provider?
- What is the distribution of billing amounts?

### Medication Patterns

- What are the most commonly prescribed medications?
- Are there any correlations between medications and medical conditions?
- How does medication usage vary by age or gender?

## Data Transformation

1. Remove extra white spaces:

healthcare\_dataset - Power Query Editor

File Home Transform Add Column View

Group By Use First Row as Headers Count Rows

Table

Any Column

lowercase  
UPPERCASE  
Capitalize Each Word

Trim  
Clean  
Add Prefix  
Add Suffix

Remove leading and trailing whitespaces from each cell in the selected columns.

Number Column

Date & Time Column

Structured Column

Queries [1]  
healthcare\_dataset

Table.TransformColumns(#"Promoted Headers", {"Room Number", {"Age", Text.Trim, type text}, {"Admission Type", {"Discharge Date", type text}}

15 COLUMNS, 999+ ROWS

Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 1:47 AM

	Hospital	Insurance Provider	Room Number	Admission Type	Discharge Date
1	Sons and Miller	Blue Cross	1885	Clean	2024-02
2	Kim Inc	Medicare	3364	Add Prefix	2019-08
3	Cook PLC	Aetna	2795	Add Suffix	2022-10
4	Hernandez Rogers and Vang,	Medicare	37909.78240987528	450	2020-12
5	White-White	Aetna	14238.317813937623	458	2022-10
6	Nunez-Humphrey	UnitedHealthcare	48145.11095104189	389	2023-12
7	Group Middleton	Medicare	19580.87234486093	389	2020-11
8	Powell Robinson and Valdez,	Cigna	45820.46272159459	277	2022-01
9	Sons Rich and	Cigna	50119.222791548505	316	2020-07
10	Padilla-Walker	UnitedHealthcare	19784.63106221073	249	2021-06
11	Schaefer-Porter	Medicare	12576.79560950234	394	2020-04
12	Lyons-Blair	Medicare	7999.586879604188	288	2023-09
13	Powers Miller, and Flores	Cigna	43282.2835770435	134	2019-12
14	Rivera-Gutierrez	UnitedHealthcare	33207.706633729606	309	2020-06
15	Morris-Arellano	UnitedHealthcare	40701.599227308754	182	2021-10
16	Cline-Williams	Blue Cross	12263.357425021362	465	2023-01
17	Cervantes-Wells	UnitedHealthcare	24499.847903736576	114	2020-07
18	Torres, and Harrison Jones	Cigna	17440.465444124675	449	2020-04
19	Houston PLC	Blue Cross	18843.02301783416	260	2021-03
20	Hammond Ltd	Cigna	23762.203579059587	465	2022-11
21	Jones LLC	Blue Cross	42.514588533243	115	2023-07
22	Williams-Davis	Aetna	17695.911622343818	295	2020-02

## 2. Change letter casing

Changed casing to capitalize each word

change case - Power Query Editor

File Home Transform Add Column View

Group By Use First Row as Headers Count Rows

Table

Any Column

lowercase  
UPPERCASE  
Capitalize Each Word

Trim  
Clean  
Add Prefix  
Add Suffix

Convert the first letter of each word in the selected columns into an uppercase letter.

Number Column

Date & Time Column

Structured Column

Queries [2]  
healthcare\_dataset  
change case

Table.TransformColumns(Table.TransformColumns(#"Promoted Headers", {"Name", {"Age", type text}, {"Date of Admission", type text}}, {"Room Number", {"Age", type text}, {"Admission Type", {"Discharge Date", type text}})

15 COLUMNS, 999+ ROWS

Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 1:47 AM

	Name	Age	Gender	Medical Condition	Date of Admission
1	Bobby Jackson	30	Male	Cancer	1/31/2024
2	Leslie Terry	62	Male	Obesity	8/20/2019
3	Danny Smith	76	Female	Obesity	9/22/2022
4	Andrew Watts	28	Female	Diabetes	11/18/2020
5	Adrienne Bell	43	Female	Cancer	9/19/2022
6	Emily Johnson	36	Male	Asthma	12/20/2023
7	Edward Edwards	21	Female	Diabetes	11/3/2020
8	Christina Martinez	20	Female	Cancer	12/28/2021
9	Jasmine Aguilar	82	Male	Asthma	7/1/2020
10	Christopher Berg	58	Female	Cancer	5/23/2021
11	Michelle Daniels	72	Male	Cancer	4/19/2020
12	Aaron Martinez	38	Female	Hypertension	8/13/2023
13	Connor Hansen	75	Female	Diabetes	12/12/2019
14	Robert Bauer	68	Female	Asthma	5/22/2020
15	Brooke Brady	44	Female	Cancer	10/8/2021
16	Ms. Natalie Gamble	46	Female	Obesity	1/1/2023
17	Haley Perkins	63	Female	Arthritis	6/23/2020
18	Mrs. Jamie Campbell	38	Male	Obesity	3/8/2020
19	Luke Burgess	34	Female	Hypertension	3/4/2021
20	Daniel Schmidt	63	Male	Asthma	11/15/2022
21	Timothy Burns	67	Female	Asthma	6/28/2023
22	Christopher Bright	48	Male	Asthma	1/21/2020

## 3. Change Datatype

- Age:

The screenshot shows the Power Query Editor interface. The 'Transform' tab is active, and the 'Data Type: Whole Number' dropdown is open, showing options like Percentage, Date/Time, Date, Duration, Text, True/False, Binary, and Number. The 'Age' column is selected, and the transformation formula is visible: `TransformColumnTypes("#Capitalized Each Word",{"Age", Int64.Type})`. The data table shows columns: Age, Gender, Blood Type, Medical Condition, and Date of Admission.

	Age	Gender	Blood Type	Medical Condition	Date of Admission
1	30	Male	B-	Cancer	2024-01-31
2	62	Male	A+	Obesity	2019-08-20
3	76	Female	A-	Obesity	2022-09-22
4	28	Female	O+	Diabetes	2020-11-18
5	43	Female	Ab+	Cancer	2022-09-19
6	36	Male	A+	Asthma	2023-12-20
7	21	Female	Ab-	Diabetes	2020-11-03
8	20	Female	A+	Cancer	2021-12-28
9	82	Male	Ab+	Asthma	2020-07-01
10	58	Female	Ab-	Cancer	2021-05-23
11	72	Male	O+	Cancer	2020-04-19
12	38	Female	A-	Hypertension	2023-08-13
13	75	Female	A+	Diabetes	2019-12-12
14	68	Female	Ab+	Asthma	2020-05-22
15	44	Female	Ab+	Cancer	2021-10-08
16	46	Female	Ab-	Obesity	2023-01-01
17	63	Female	A+	Arthritis	2020-06-23
18	38	Male	Ab-	Obesity	2020-09-08
19	34	Female	A-	Hypertension	2021-03-04
20	63	Male	B+	Asthma	2022-11-15
21	67	Female	A-	Asthma	2023-06-28
22	48	Male	B+	Asthma	2020-01-21

- Date of admission, Date of Discharge:

The screenshot shows the Power Query Editor interface. The 'Transform' tab is active, and the 'Data Type: Date' dropdown is open, showing options like Decimal Number, Currency, Whole Number, Percentage, Date/Time, Date, Duration, Text, True/False, Binary, and Number. The 'Date of Admission' column is selected, and the transformation formula is visible: `TransformColumnTypes("#Capitalized Each Word",{"Age", Int64.Type}, {"Date of Admission", type date})`. The data table shows columns: Date of Admission, Doctor, Hospital, Insurance Provider, and Billing Amount.

	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount
1	1/31/2024	Matthew Smith	Sons And Miller	Blue Cross	18856.28
2	8/20/2019	Samantha Davies	Kim Inc	Medicare	33643.32
3	9/22/2022	Tiffany Mitchell	Cook Plc	Aetna	27955.09
4	11/18/2020	Kevin Wells	Hernandez Rogers And Vang,	Medicare	37909.78
5	9/19/2022	Kathleen Hanna	White-White	Aetna	14238.31
6	12/20/2023	Taylor Newton	Nunez-Humphrey	Unitedhealthcare	48145.11
7	11/3/2020	Kelly Olson	Group Middleton	Medicare	19580.87
8	12/28/2021	Suzanne Thomas	Powell Robinson And Valdez,	Cigna	45820.46
9	7/1/2020	Daniel Ferguson	Sons Rich And	Cigna	50119.22
10	5/23/2021	Heather Day	Padilla-Walker	Unitedhealthcare	19784.63
11	4/19/2020	John Duncan	Schaefer-Porter	Medicare	12576.79
12	8/13/2023	Douglas Mayo	Lyons-Blair	Medicare	7999.586
13	12/12/2019	Kenneth Fletcher	Powers Miller, And Flores	Cigna	43282.28
14	5/22/2020	Theresa Freeman	Rivera-Gutierrez	Unitedhealthcare	33207.70
15	10/8/2021	Roberta Stewart	Morris-Arellano	Unitedhealthcare	40701.59
16	1/1/2023	Maria Dougherty	Cline-Williams	Blue Cross	12263.35
17	8/23/2020	Erica Spencer	Cervantes-Wells	Unitedhealthcare	24499.84
18	3/8/2020	Justin Kim	Torres, And Harrison Jones	Cigna	17440.46
19	3/4/2021	Justin Moore Jr.	Houston Plc	Blue Cross	18843.02
20	11/15/2022	Denise Galloway	Hammond Ltd	Cigna	23762.20
21	6/28/2023	Krista Smith	Jones Llc	Blue Cross	42.51458
22	1/21/2020	Gregory Smith	Williams-Davis	Aetna	17695.91

- Billing amount:

healthcare\_dataset - Power Query Editor

File Home Transform Add Column View

Data Type: Currency

Replace Values

Unpivot Columns

Move

Convert to List

Split Column

Format

Parse

Merge Columns

Extract

Statistics

Standard Scientific

Trigonometry

10<sup>2</sup>

Rounding

Information

Date

Time

Duration

Expand

Aggregate

Extract Values

Create Data Type

Queries [1] healthcare\_dataset

Table

TransformColumnTypes(\*"Capitalized Each Word",{"Age", Int64.Type}, {"Date of Admission", type date},

	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date
1	Blue Cross	18,856.28	328	Urgent	
2	Medicare	33,643.33	265	Emergency	
3	Aetna	27,955.10	205	Emergency	
4	Medicare	37,909.78	450	Elective	
5	Aetna	14,238.32	458	Urgent	
6	Unitedhealthcare	48,145.11	389	Urgent	
7	Medicare	19,580.87	389	Emergency	
8	Well Robinson And Valdez,	45,820.46	277	Emergency	
9	Cigna	50,119.22	316	Elective	
10	Unitedhealthcare	19,784.63	249	Elective	
11	Medicare	12,576.80	394	Urgent	
12	Medicare	7,999.59	288	Urgent	
13	Cigna	43,282.28	134	Emergency	
14	Unitedhealthcare	33,207.71	309	Urgent	
15	Unitedhealthcare	40,701.60	182	Urgent	
16	Blue Cross	12,263.36	465	Elective	
17	Unitedhealthcare	24,499.85	114	Elective	
18	Cigna	17,440.47	449	Urgent	
19	Blue Cross	18,843.02	260	Elective	
20	Cigna	23,762.20	465	Elective	
21	Blue Cross	42.51	115	Elective	
22	Aetna	17,695.91	295	Urgent	

15 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 1:48 AM

#### 4. Add custom Columns

- Length of stay:

healthcare\_dataset - Power Query Editor

File Home Transform Add Column View

Conditional Column

Index Column

Function

Duplicate Column

General

From Text

From Number

From Date & Time

Queries [1] healthcare\_dataset

Table

Custom Column

Add a column that is computed from the other columns.

New column name

Length of Stay

Custom column formula

= [Discharge Date] - [Date of Admission]

Available columns

Age

Gender

Blood Type

Medical Condition

Date of Admission

Doctor

Hospital

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.

OK Cancel

Query Settings

PROPERTIES

Name

healthcare\_dataset

APPLIED STEPS

Source

Promoted Headers

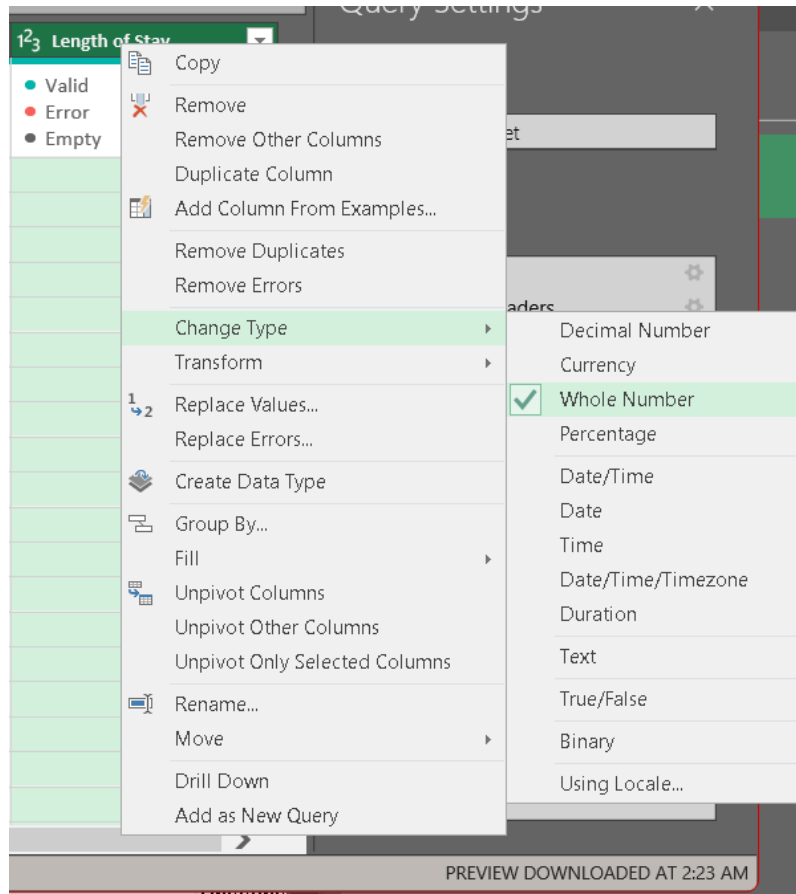
Trimmed Text

Capitalized Each Word

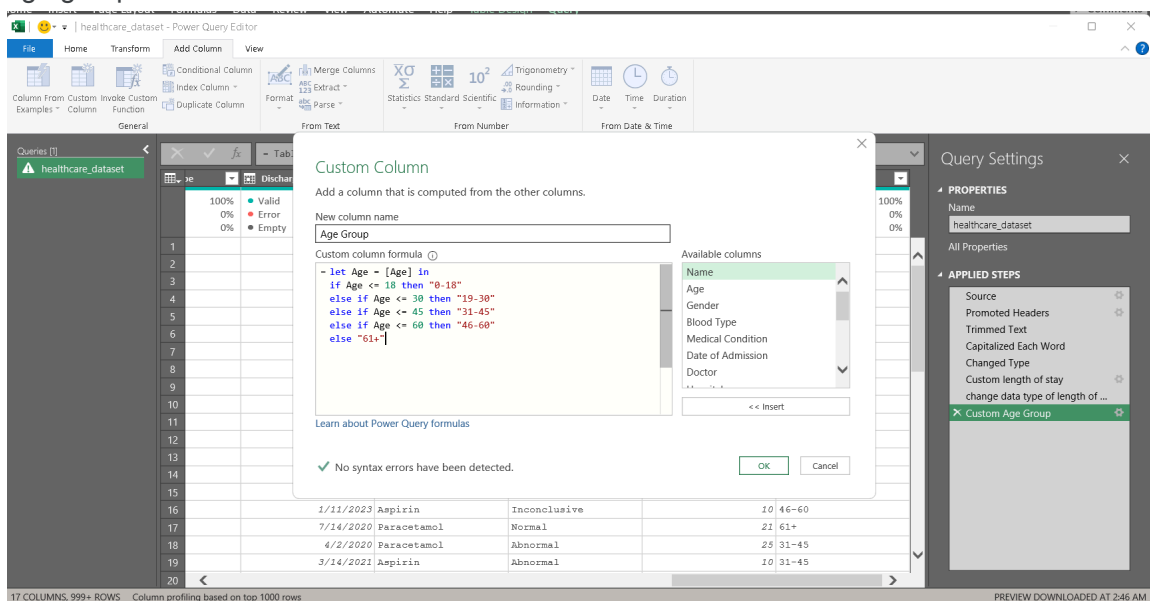
Changed Type

15 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 2:11 AM



- Age group:

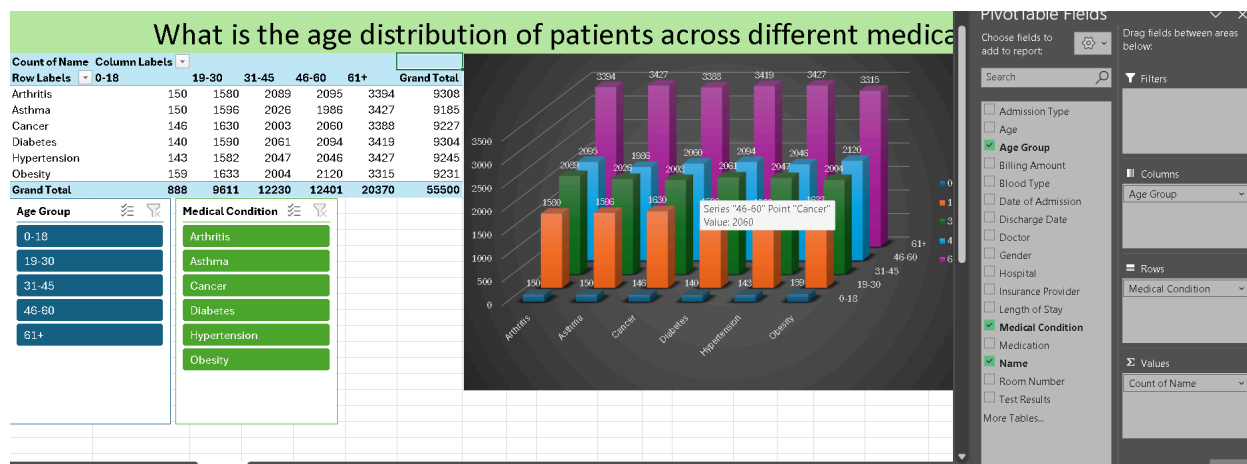


# Analysis

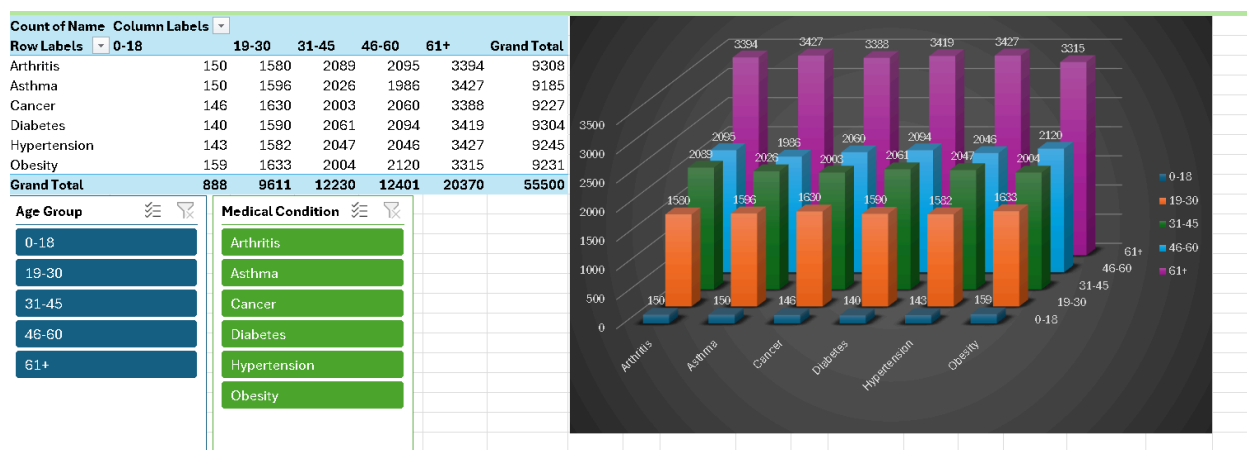
## Patient Demographics and Characteristics in relation to medical conditions

### 1. What is the age distribution of patients?

Pivot Table1:



Set “Age Group” to columns, “Medical Condition” to Rows and “Count of Name” for the Value field. Added Pivot chart for Visualization.





Change count of names to show value as percentage of Column Total. Added another pivot chart for visualization.

Value Field Settings

Source Name: Name

Custom Name:

Count of Name

Summarize Values By

Show Values As

Summarize value field by

Choose the type of calculation that you want to use to summarize data from the selected field

Sum

Count

Average

Max

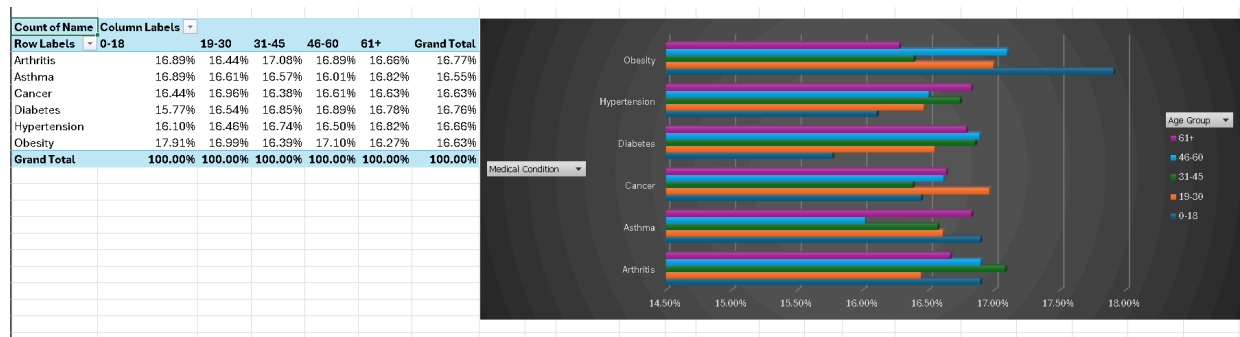
Min

Product

Number Format

OK

Cancel



Slicers linked with both charts

Age Group		Medical Condition	
0-18		Arthritis	
19-30		Asthma	
31-45		Cancer	
46-60		Diabetes	
61+		Hypertension	
		Obesity	

## Analysis:

### Key Findings

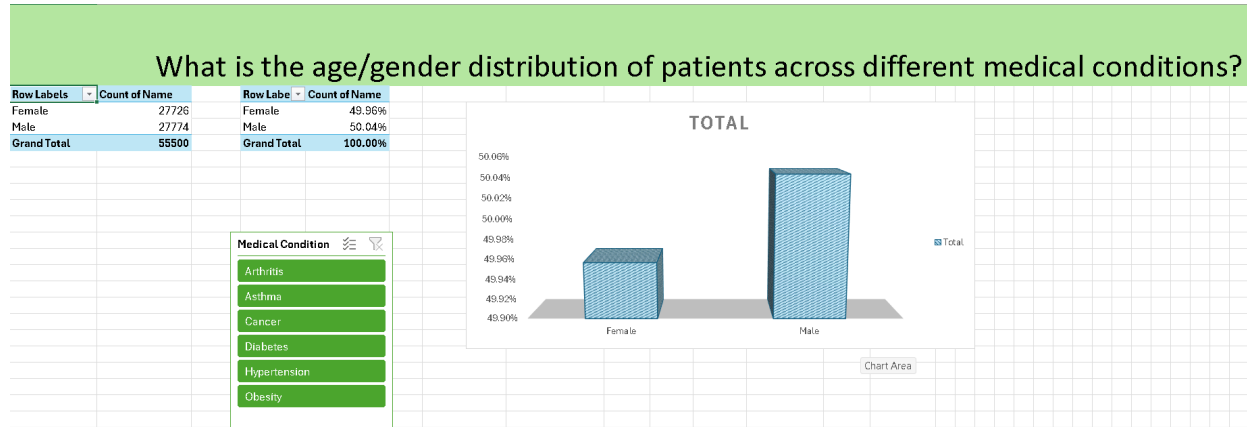
- The age groups 31-45 and 46-60 exhibit the highest patient counts across all medical conditions. There's a noticeable decline in patient numbers for the older age group (61+).
- Obesity stands out as the most prevalent condition across all age groups. In contrast, arthritis and hypertension primarily affect the older population (46-60 and 61+). Asthma demonstrates a relatively even distribution across age groups. Cancer, diabetes, and hypertension show a higher concentration in the 31-45 and 46-60 age brackets.

### Implications

- The high prevalence of obesity across all age groups underscores the need for comprehensive public health initiatives targeting prevention and management.
- The concentration of chronic conditions like arthritis and hypertension in older populations highlights the importance of specialized geriatric care and chronic disease management programs.
- The significant presence of cancer and diabetes in the 31-45 and 46-60 age groups emphasizes the need for early detection and screening programs.
- The relatively even distribution of asthma across age groups suggests potential environmental or lifestyle factors contributing to the condition.

# 1. What is the gender ratio of patients and its relation with medical conditions?

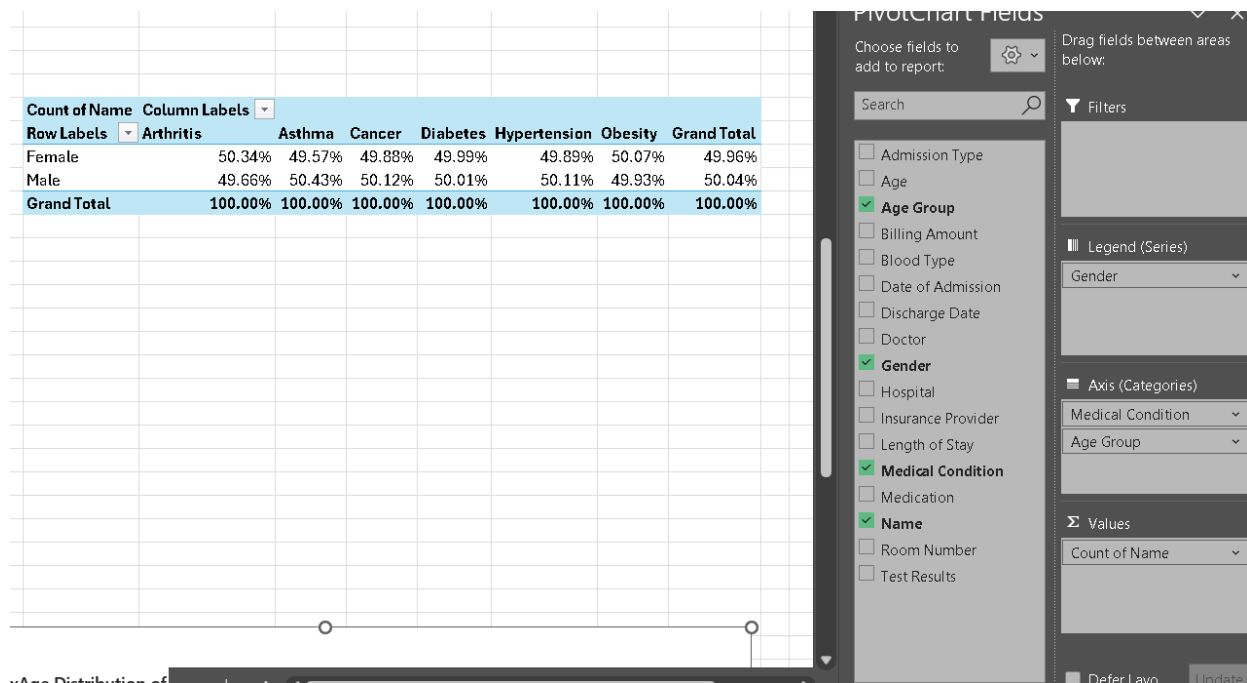
## Pivot Table for male and female patient distribution in numbers and percentage



## Gender Breakdown:

- Females account for 27,726 individuals, representing approximately 49.96% of the total.
- Males account for 27,774 individuals, representing approximately 50.04% of the total.
- There is a near-equal distribution of genders, with a slight majority of males.

## Pivot table for percentage gender distribution of medical conditions



- The data indicates a near-equal distribution of male and female patients across all medical conditions. The percentage values for both genders fluctuate slightly around 50% for each condition.
- There is no apparent significant gender bias evident in any specific medical condition based on the provided data.

**Pivot Table for Pivot table for gender distribution of medical conditions**

Count of Name	Column Labels		
Row Labels	Female	Male	Grand Total
<b>Arthritis</b>	<b>16.90%</b>	<b>16.64%</b>	<b>16.77%</b>
0-18	0.30%	0.24%	0.27%
19-30	2.87%	2.82%	2.85%
31-45	3.75%	3.77%	3.76%
46-60	3.82%	3.73%	3.77%
61+	6.15%	6.08%	6.12%
<b>Asthma</b>	<b>16.42%</b>	<b>16.68%</b>	<b>16.55%</b>
0-18	0.28%	0.26%	0.27%
19-30	2.81%	2.94%	2.88%
31-45	3.65%	3.65%	3.65%
46-60	3.52%	3.63%	3.58%
61+	6.15%	6.20%	6.17%
<b>Cancer</b>	<b>16.60%</b>	<b>16.65%</b>	<b>16.63%</b>
0-18	0.25%	0.28%	0.26%
19-30	2.96%	2.92%	2.94%
31-45	3.65%	3.57%	3.61%
46-60	3.75%	3.68%	3.71%
61+	5.99%	6.21%	6.10%
<b>Diabetes</b>	<b>16.77%</b>	<b>16.75%</b>	<b>16.76%</b>
0-18	0.26%	0.24%	0.25%
19-30	2.95%	2.78%	2.86%
31-45	3.60%	3.83%	3.71%
46-60	3.78%	3.77%	3.77%
61+	6.19%	6.14%	6.16%
<b>Hypertension</b>	<b>16.63%</b>	<b>16.68%</b>	<b>16.66%</b>
0-18	0.28%	0.24%	0.26%
19-30	2.83%	2.87%	2.85%
31-45	3.60%	3.77%	3.69%
46-60	3.58%	3.79%	3.69%
61+	6.34%	6.01%	6.17%
<b>Obesity</b>	<b>16.67%</b>	<b>16.59%</b>	<b>16.63%</b>
0-18	0.27%	0.30%	0.29%
19-30	2.90%	2.98%	2.94%
31-45	3.62%	3.60%	3.61%
46-60	3.94%	3.70%	3.82%
61+	5.94%	6.00%	5.97%
<b>Grand Total</b>	<b>100.00%</b>	<b>100.00%</b>	<b>100.00%</b>

### PivotTable Fields

Choose fields to add to report:

Drag fields between areas below:

**Filters**

**Columns**

Gender

**Rows**

Medical Condition

Age Group

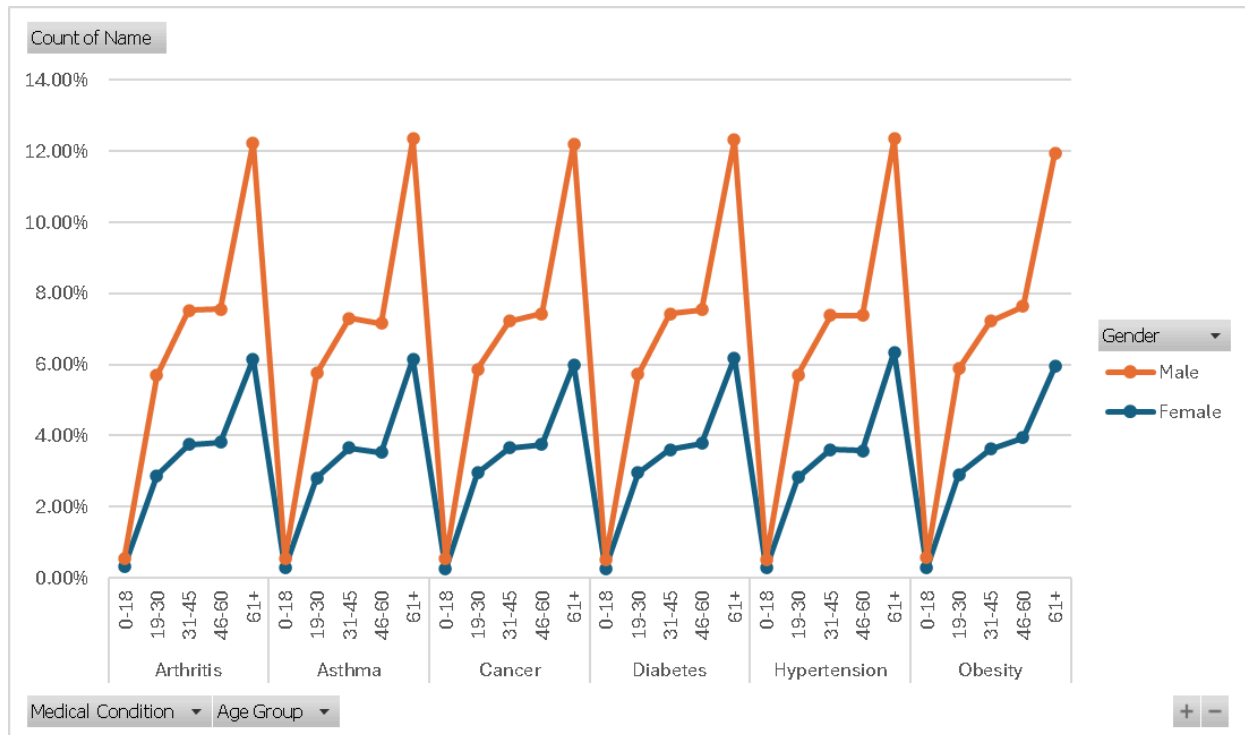
**Σ Values**

Count of Name

☐ Defer Layo...

- ☐ Admission Type
- ☐ Age
- ☒ Age Group
- ☐ Billing Amount
- ☐ Blood Type
- ☐ Date of Admission
- ☐ Discharge Date
- ☐ Doctor
- ☒ Gender
- ☐ Hospital
- ☐ Insurance Provider
- ☐ Length of Stay
- ☒ Medical Condition
- ☐ Medication
- ☒ Name
- ☐ Room Number
- ☐ Test Results

More Tables...



- There is a relatively even distribution of male and female patients across most medical conditions and age groups.
- Obesity shows the highest percentage across all age groups for both genders.
- Arthritis and Hypertension are predominantly prevalent in older age groups (46-60 and 61+) for both genders.
- Asthma and Cancer show a relatively even distribution across age groups for both genders.
- The 31-45 and 46-60 age groups have a higher percentage of patients for most medical conditions compared to younger and older age groups.

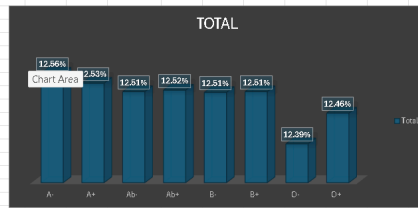
3. What is the distribution of blood types among patients and different medical conditions?

**Pivot Chart for blood type distribution among patients**

## What is the distribution of blood types among patients and different medical conditions

Row Labels	Count of Name	Row Labels	Count of Name
A-	6969	A-	12.56%
A+	6956	A+	12.53%
Ab-	6945	Ab-	12.51%
Ab+	6947	Ab+	12.52%
B-	6944	B-	12.51%
B+	6945	B+	12.51%
O-	6877	O-	12.39%
O+	6917	O+	12.46%
<b>Grand Total</b>	<b>55500</b>	<b>Grand Total</b>	<b>100.00%</b>

Medical Condition
Arthritis
Asthma
Cancer
Diabetes
Hypertension
Obesity

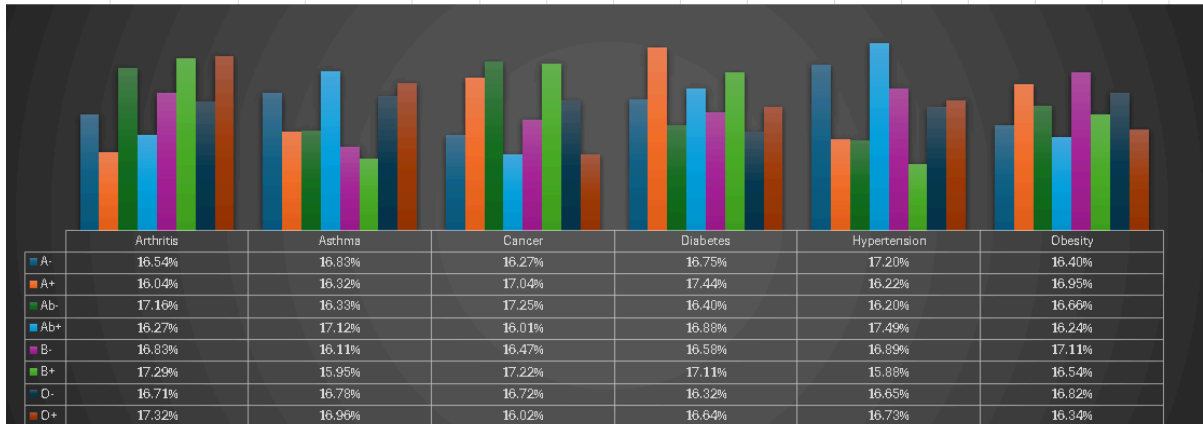


PivotTable Fields
Choose fields to add to report:
Search
Filters
<input type="checkbox"/> Admission Type <input type="checkbox"/> Age <input type="checkbox"/> Age Group <input type="checkbox"/> Billing Amount <input checked="" type="checkbox"/> Blood Type <input type="checkbox"/> Date of Admission <input type="checkbox"/> Discharge Date <input type="checkbox"/> Doctor <input type="checkbox"/> Gender <input type="checkbox"/> Hospital <input type="checkbox"/> Insurance Provider <input type="checkbox"/> Length of Stay <input type="checkbox"/> Medical Condition <input type="checkbox"/> Medication <input checked="" type="checkbox"/> Name <input type="checkbox"/> Room Number <input type="checkbox"/> Test Results
Drag fields between areas below:
Columns
Rows
Blood Type Count of Name

- The distribution of blood types appears relatively even across all medical conditions. There are minor fluctuations, but no significant overrepresentation or underrepresentation of any blood type for a specific condition.

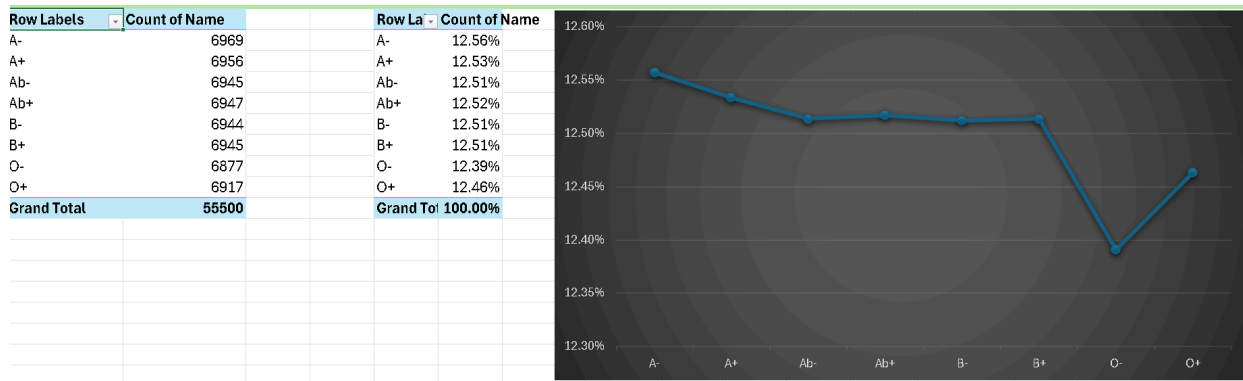
## Pivot table for blood distribution of patients in relation medical conditions

Count of Name	Column Labels								
Row Labels	A-	A+	Ab-	Ab+	B-	B+	O-	O+	Grand Total
Arthritis	16.54%	16.04%	17.16%	16.27%	16.83%	17.29%	16.71%	17.32%	16.77%
Asthma	16.83%	16.32%	16.33%	17.12%	16.11%	15.95%	16.78%	16.96%	16.55%
Cancer	16.27%	17.04%	17.25%	16.01%	16.47%	17.22%	16.72%	16.02%	16.63%
Diabetes	16.75%	17.44%	16.40%	16.88%	16.58%	17.11%	16.32%	16.64%	16.76%
Hypertension	17.20%	16.22%	16.20%	17.49%	16.89%	15.88%	16.65%	16.73%	16.66%
Obesity	16.40%	16.95%	16.66%	16.24%	17.11%	16.54%	16.82%	16.34%	16.63%
Grand Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%



- The most striking feature is the relatively even distribution of blood types across all medical conditions. There are minor fluctuations, but no significant overrepresentation or underrepresentation of any blood type for a specific condition.
- The data suggests a lack of correlation between blood type and medical condition based on the provided information.

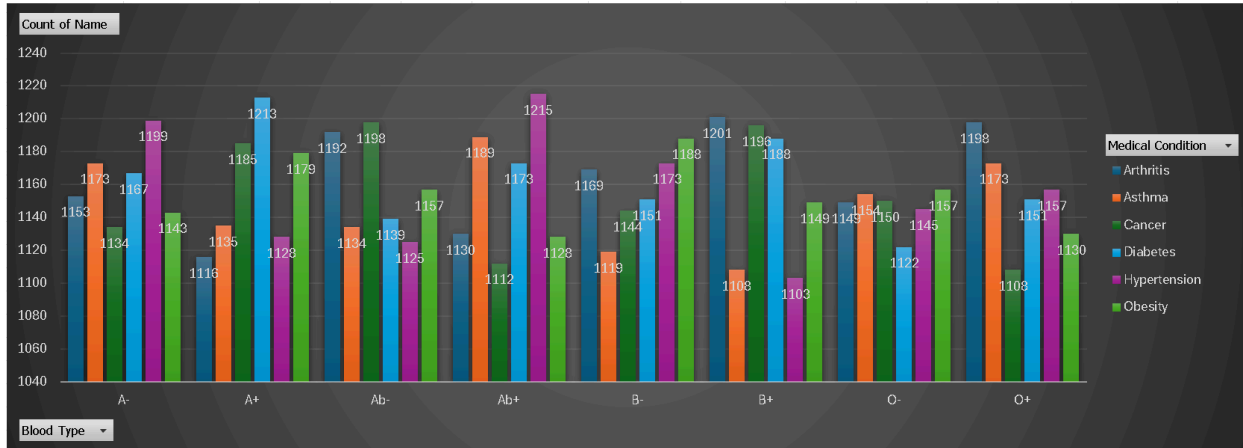
## 2. What is the distribution of blood types among patients?



- The data reveals a relatively even distribution of blood types among the patient population.
- There are minor fluctuations in the percentages of different blood types, but no significant predominance of any particular type.
- While the total number of patients represented in the dataset is 55500, the most frequent blood type appears to be O+, followed closely by A+.

Count of Name	Column Labels						
Row Labels	Arthritis	Asthma	Cancer	Diabetes	Hypertension	Obesity	Grand Total
A-	1153	1173	1134	1167	1199	1143	6969
A+	1116	1135	1185	1213	1128	1179	6956
Ab-	1192	1134	1198	1139	1125	1157	6945
Ab+	1130	1189	1112	1173	1215	1128	6947
B-	1169	1119	1144	1151	1173	1188	6944
B+	1201	1108	1196	1188	1103	1149	6945
O-	1149	1154	1150	1122	1145	1157	6877
O+	1198	1173	1108	1151	1157	1130	6917
<b>Grand Total</b>	<b>9308</b>	<b>9185</b>	<b>9227</b>	<b>9304</b>	<b>9245</b>	<b>9231</b>	<b>55500</b>





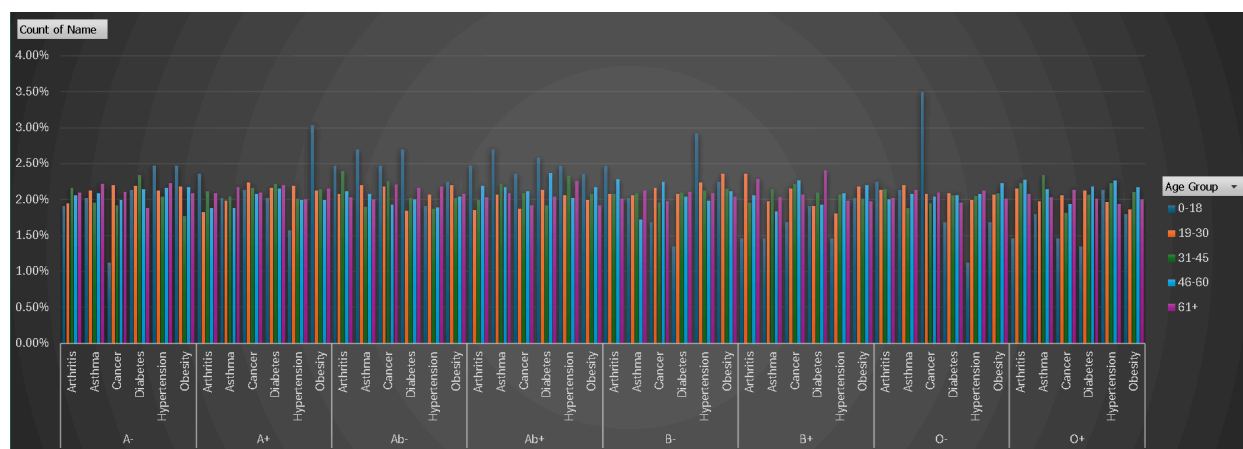
- Obesity is particularly high in the AB+ and A- blood groups.
- Diabetes is most prevalent in the A+ group.
- Hypertension has a high occurrence in B+ and O+ blood groups.
- Asthma is consistently present across most blood groups, with the highest counts in A- and O+.
- Arthritis shows a relatively uniform distribution across blood groups, with slightly higher counts in B+ and AB-.
- Cancer varies more widely, with A+ and B+ showing the highest counts.
- The distribution of medical conditions does not vary drastically between blood groups, but there are some slight variations.
- AB+ blood type shows a notably higher count for Obesity compared to other conditions.
- B+ blood type shows a consistently higher count across most conditions, particularly for Arthritis and Hypertension.
- The O- group has the lowest overall count, particularly for Hypertension and Cancer.

Count of Name	Column Labels					
Row Labels	0-18	19-30	31-45	46-60	61+	Grand Total
A-	12.16%	12.78%	12.21%	12.62%	12.64%	12.56%
Arthritis	1.91%	1.95%	2.17%	2.06%	2.10%	2.08%
Asthma	2.03%	2.12%	1.95%	2.09%	2.22%	2.11%
Cancer	1.13%	2.21%	1.92%	1.99%	2.11%	2.04%
Diabetes	2.14%	2.20%	2.35%	2.14%	1.89%	2.10%

Hypertension	2.48%	2.12%	2.04%	2.16%	2.23%	2.16%
Obesity	2.48%	2.18%	1.77%	2.17%	2.09%	2.06%
<b>A+</b>	<b>13.18%</b>	<b>12.54%</b>	<b>12.71%</b>	<b>12.01%</b>	<b>12.72%</b>	<b>12.53%</b>
Arthritis	2.36%	1.83%	2.12%	1.89%	2.09%	2.01%
Asthma	2.03%	1.99%	2.04%	1.89%	2.17%	2.05%
Cancer	2.14%	2.24%	2.17%	2.08%	2.10%	2.14%
Diabetes	2.03%	2.16%	2.22%	2.15%	2.20%	2.19%
Hypertension	1.58%	2.20%	2.01%	2.00%	2.01%	2.03%
Obesity	3.04%	2.12%	2.14%	2.00%	2.15%	2.12%
<b>Ab-</b>	<b>14.53%</b>	<b>12.59%</b>	<b>12.48%</b>	<b>12.07%</b>	<b>12.68%</b>	<b>12.51%</b>
Arthritis	2.48%	2.08%	2.40%	2.12%	2.03%	2.15%
Asthma	2.70%	2.21%	1.90%	2.08%	2.00%	2.04%
Cancer	2.48%	2.18%	2.26%	1.93%	2.21%	2.16%
Diabetes	2.70%	1.84%	2.03%	2.01%	2.16%	2.05%
Hypertension	1.91%	2.07%	1.87%	1.90%	2.18%	2.03%
Obesity	2.25%	2.21%	2.03%	2.04%	2.08%	2.08%
<b>Ab+</b>	<b>14.98%</b>	<b>11.99%</b>	<b>12.63%</b>	<b>13.05%</b>	<b>12.27%</b>	<b>12.52%</b>
Arthritis	2.48%	1.85%	2.00%	2.19%	2.03%	2.04%
Asthma	2.70%	2.07%	2.22%	2.17%	2.09%	2.14%
Cancer	2.36%	1.87%	2.09%	2.12%	1.92%	2.00%
Diabetes	2.59%	2.13%	1.92%	2.37%	2.04%	2.11%
Hypertension	2.48%	2.06%	2.33%	2.02%	2.25%	2.19%
Obesity	2.36%	2.00%	2.08%	2.17%	1.92%	2.03%
<b>B-</b>	<b>12.73%</b>	<b>12.99%</b>	<b>12.51%</b>	<b>12.39%</b>	<b>12.35%</b>	<b>12.51%</b>
Arthritis	2.48%	2.08%	2.08%	2.28%	2.01%	2.11%

Asthma	2.03%	2.06%	2.09%	1.73%	2.13%	2.02%
Cancer	1.69%	2.16%	1.96%	2.25%	1.97%	2.06%
Diabetes	1.35%	2.08%	2.10%	2.04%	2.11%	2.07%
Hypertension	2.93%	2.24%	2.13%	1.98%	2.09%	2.11%
Obesity	2.25%	2.36%	2.15%	2.11%	2.04%	2.14%
<b>B+</b>	<b>10.02%</b>	<b>12.40%</b>	<b>12.49%</b>	<b>12.39%</b>	<b>12.76%</b>	<b>12.51%</b>
Arthritis	1.46%	2.36%	1.95%	2.06%	2.29%	2.16%
Asthma	1.46%	1.98%	2.14%	1.84%	2.04%	2.00%
Cancer	1.69%	2.15%	2.22%	2.27%	2.07%	2.15%
Diabetes	1.91%	1.91%	2.10%	1.93%	2.41%	2.14%
Hypertension	1.46%	1.81%	2.07%	2.09%	1.98%	1.99%
Obesity	2.03%	2.18%	2.01%	2.20%	1.97%	2.07%
<b>O-</b>	<b>12.39%</b>	<b>12.58%</b>	<b>12.17%</b>	<b>12.50%</b>	<b>12.37%</b>	<b>12.39%</b>
Arthritis	2.25%	2.13%	2.14%	2.01%	2.03%	2.07%
Asthma	2.14%	2.21%	1.88%	2.08%	2.14%	2.08%
Cancer	3.49%	2.08%	1.95%	2.04%	2.10%	2.07%
Diabetes	1.69%	2.09%	2.06%	2.06%	1.96%	2.02%
Hypertension	1.13%	2.00%	2.05%	2.08%	2.13%	2.06%
Obesity	1.69%	2.07%	2.09%	2.23%	2.01%	2.08%
<b>O+</b>	<b>10.02%</b>	<b>12.14%</b>	<b>12.80%</b>	<b>12.97%</b>	<b>12.21%</b>	<b>12.46%</b>
Arthritis	1.46%	2.15%	2.23%	2.27%	2.08%	2.16%
Asthma	1.80%	1.98%	2.34%	2.14%	2.04%	2.11%
Cancer	1.46%	2.06%	1.82%	1.94%	2.14%	2.00%
Diabetes	1.35%	2.12%	2.07%	2.19%	2.02%	2.07%
Hypertension	2.14%	1.97%	2.23%	2.27%	1.94%	2.08%

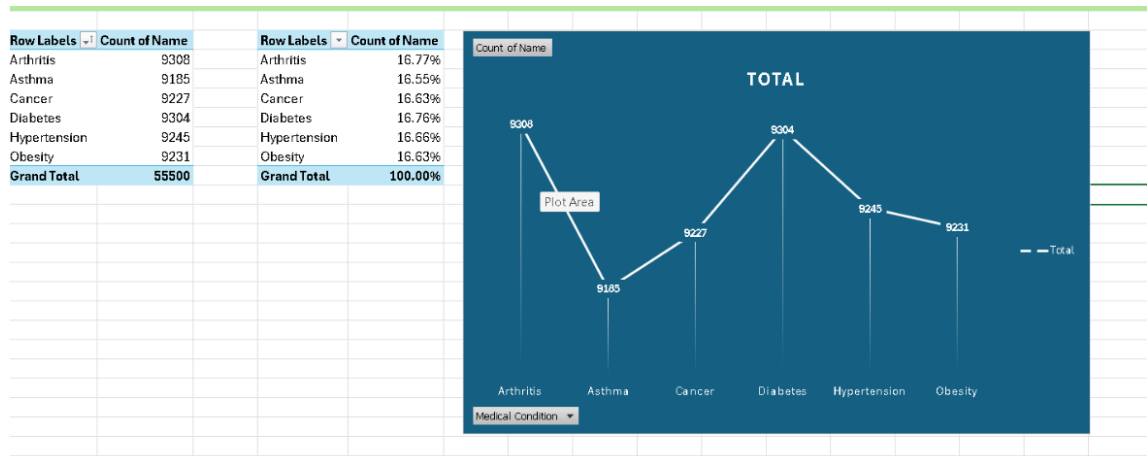
Obesity	1.80%	1.86%	2.11%	2.17%	2.00%	2.04%
<b>Grand Total</b>	<b>100.00%</b>	<b>100.00%</b>	<b>100.00%</b>	<b>100.00%</b>	<b>100.00%</b>	<b>100.00%</b>



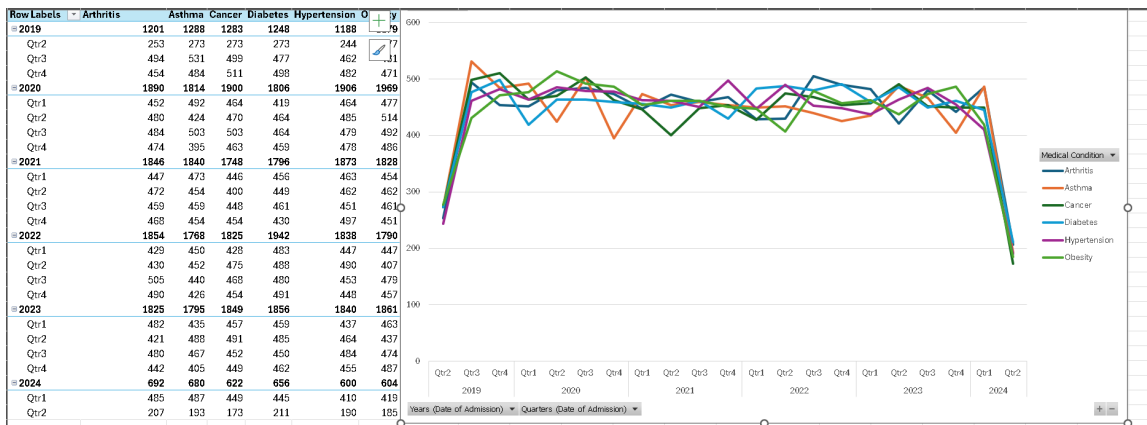
- Hypertension is the most prevalent condition across all blood types and is particularly common in the 46-60 age group.
- Asthma and Cancer have lower prevalence across most blood types, especially in younger age groups.
- Diabetes and Obesity are consistently present but show varying levels of prevalence depending on the age group and blood type.
- The 46-60 age group appears to be the most critical in terms of health concerns, particularly for Hypertension and Diabetes.

## Disease Prevalence and Patterns

## 1. What are the most common medical conditions?



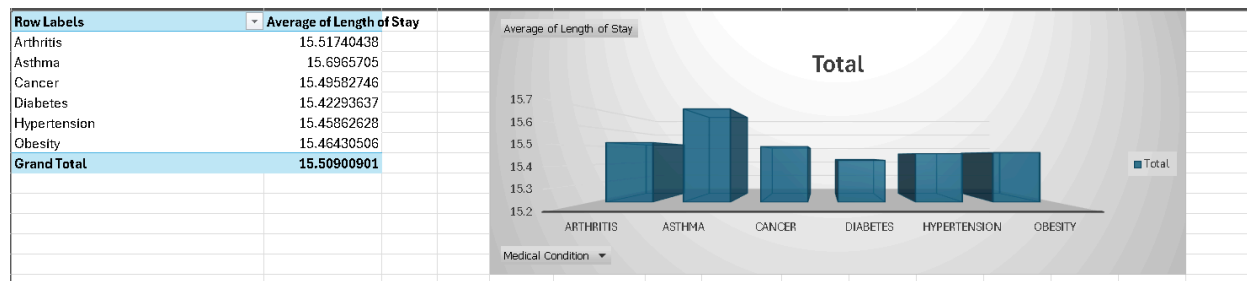
## 2. How has the prevalence of specific diseases changed over time?



- The data shows relatively stable trends across all conditions, with minor fluctuations. This suggests that these medical conditions are persistently prevalent across the population sampled.
- There may be slight seasonal variations, especially noticeable in conditions like Arthritis and Hypertension, where counts often peak in Qtr3 or Qtr4.
- The data for 2024 (partial year) suggests a possible decline in cases, especially in Qtr2, which might indicate either a real decrease or lack of further data.

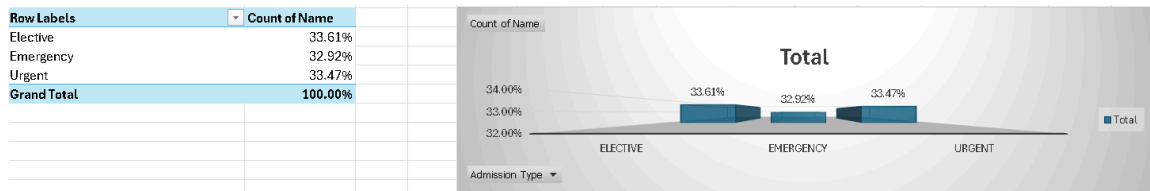
## Hospital Utilization

## 1. What is the average length of stay for different medical conditions?



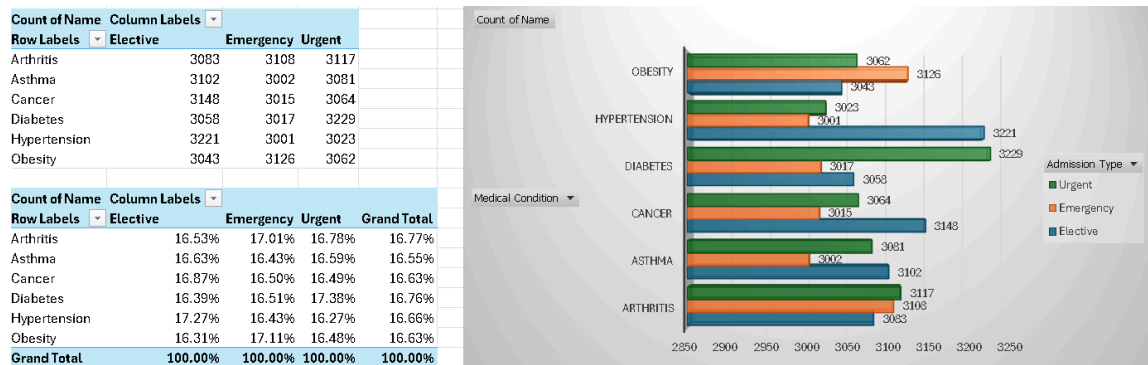
The average length of stay for most conditions is close to the grand total, indicating consistency across conditions. Asthma shows a slightly longer average stay, while Diabetes shows a shorter average stay. The minimal variation among the conditions suggests that the differences in average length of stay are not very significant.

## 2. What is the distribution of admission types (emergency, elective, urgent)?



- The percentages for Elective, Emergency, and Urgent categories are quite similar, ranging between 32.92% and 33.61%. This indicates that the distribution of cases among these categories is fairly even.
- Elective has the highest percentage at 33.61%. This suggests that among the three categories, elective cases make up the largest share, though the difference is minimal.
- Emergency has the lowest percentage at 32.92%. This is only slightly lower than the other two categories.
- The percentages are very close to each other, with each category making up roughly one-third of the total. This indicates a balanced distribution with no category significantly dominating or being underrepresented.

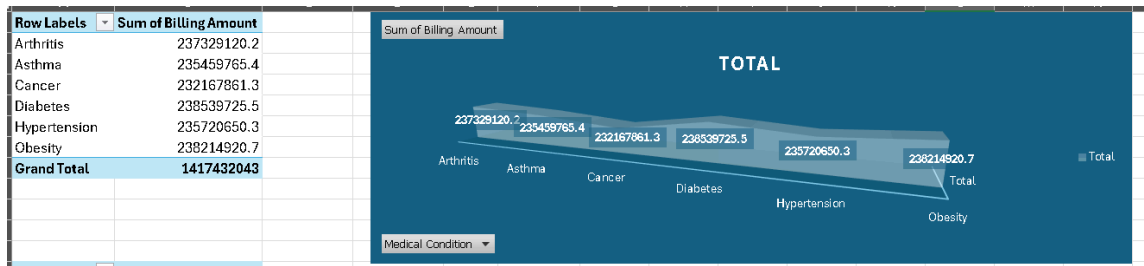
### 3. How do medical conditions correlate with admission type?



- The counts for each condition across categories (Elective, Emergency, Urgent) are relatively close. The highest count for each condition varies between categories but generally stays within a few hundred of each other.
- 
- The percentages reflect the distribution of each condition across the categories relative to the total for that category.
- Hypertension has the highest percentage for Elective (17.27%) and the lowest for Urgent (16.27%).
- Diabetes has the highest percentage for Urgent (17.38%) and is relatively high in Emergency as well (16.51%).
- Asthma shows a balanced distribution with similar percentages across all categories.
- 
- Elective Category: Conditions with higher counts are Hypertension (3,221) and Cancer (3,148), and their percentages are also relatively high (16.87% for Cancer and 17.27% for Hypertension).
- Emergency Category: Conditions with higher counts are Obesity (3,126) and Arthritis (3,108). Their percentages are also higher (17.11% for Obesity and 17.01% for Arthritis).
- Urgent Category: Conditions with higher counts are Diabetes (3,229) and Arthritis (3,117). Diabetes has the highest percentage in the Urgent category (17.38%).

## Financial Analysis

## 1. What is the total billing amount for different medical conditions?



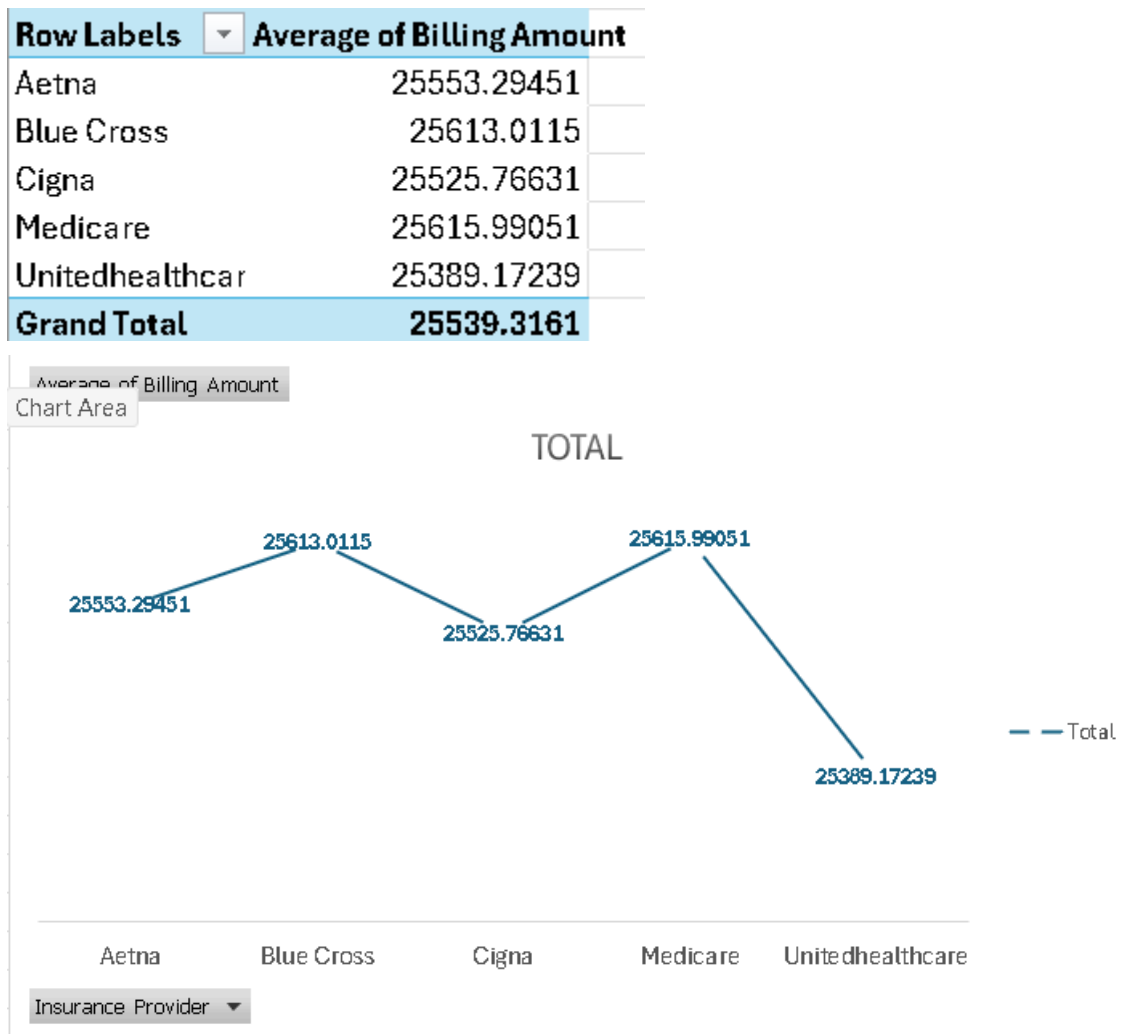
- 
- Diabetes has the highest total billing amount at 238,539,725.5. This suggests that, in terms of total revenue generated, diabetes cases are the highest among the listed conditions.
- Cancer has the lowest total billing amount at 232,167,861.3. Despite being a significant condition, it generates the least total revenue among the listed conditions.
- Obesity (238,214,920.7) and Diabetes (238,539,725.5) have the highest billing amounts, closely followed by Arthritis (237,329,120.2) and Hypertension (235,720,650.3).
- The difference in billing amounts between Diabetes and Cancer is approximately 6,371,864.2, which is the largest gap among the conditions.
- Diabetes and Obesity together account for a significant portion of the total billing amount ( $238,539,725.5 + 238,214,920.7 = 476,754,646.2$ ), which is about 33.6% of the grand total (1,417,432,043).

### Summary

- Diabetes leads in total billing amount, followed closely by Obesity.
- Cancer has the lowest total billing amount among the listed conditions, though the difference is not extreme.
- The distribution of billing amounts shows that some conditions, like Diabetes and Obesity, contribute more significantly to the total revenue compared to others like Cancer.



2. How does the average billing amount vary by insurance provider?



- Medicare has the highest average billing amount at 25,615.99. This indicates that, on average, billing amounts under Medicare are slightly higher compared to other insurance providers.
- Unitedhealthcare has the lowest average billing amount at 25,389.17. This suggests that billing amounts under Unitedhealthcare are slightly lower on average compared to other providers.
- The differences between the average billing amounts across insurance providers are relatively small, with the range being approximately 226.82 (25,615.99 - 25,389.17).
- The average billing amount for Blue Cross (25,613.01) and Medicare (25,615.99) are the highest, while Unitedhealthcare (25,389.17) has the lowest.
- The average billing amounts across all providers are quite close to the grand total average of 25,539.32, indicating a consistent billing pattern among the providers.

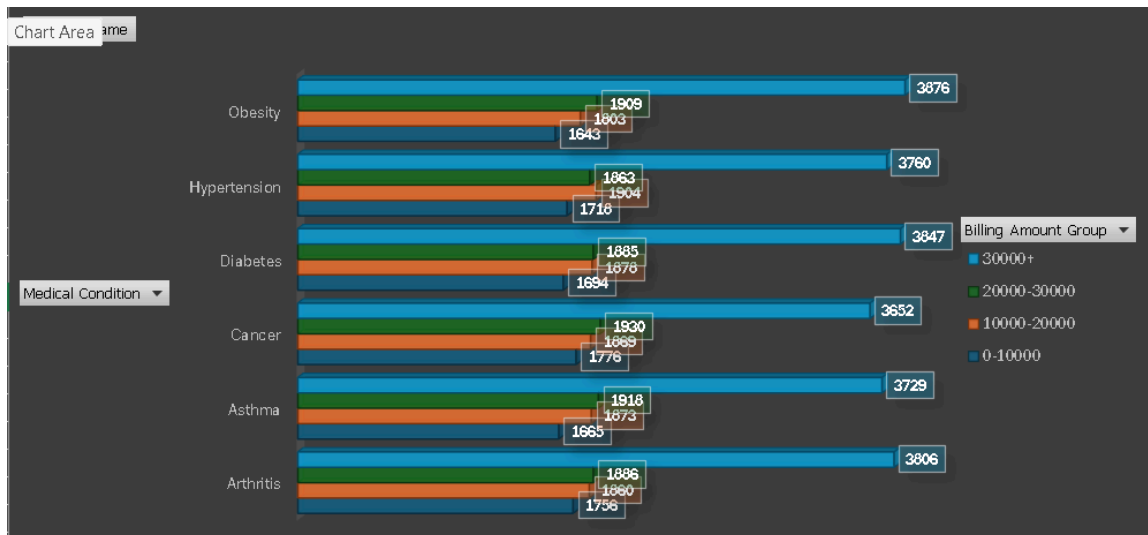
## Summary

- Medicare and Blue Cross have the highest average billing amounts, while Unitedhealthcare has the lowest.
- The variation in average billing amounts among the providers is minimal, suggesting that billing practices are relatively consistent across these insurance companies.

### 3. What is the distribution of billing amounts?

Added a new column called Billing Amount group in Power query.

Count of Name	Column Labels			
Row Labels	0-10000	10000-20000	20000-30000	30000+
Arthritis	1756	1860	1886	3806
Asthma	1665	1873	1918	3729
Cancer	1776	1869	1930	3652
Diabetes	1694	1878	1885	3847
Hypertension	1718	1904	1863	3760
Obesity	1643	1803	1909	3876



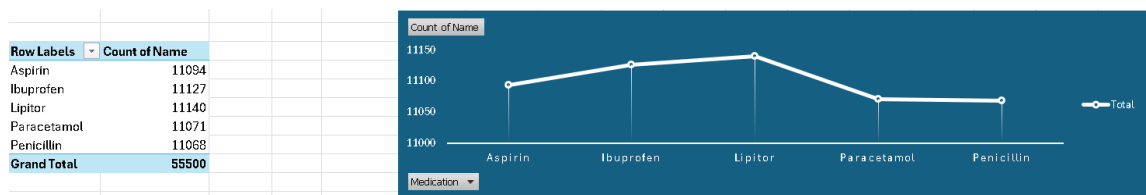
- Obesity has the highest count in the 30,000+ range with 3,876 cases.
- Diabetes follows closely with 3,847 cases.
- Arthritis has the third highest count in this range with 3,806 cases.
- Cancer has the lowest count in this range with 3,652 cases.
- Arthritis has the highest count in the 0-10,000 range with 1,756 cases.
- Obesity has the lowest count in this range with 1,643 cases.
- All conditions show a higher number of cases in the highest billing range (30,000+) compared to the lower ranges.
- The 30,000+ range generally has the most cases, indicating that a significant portion of the cases for each condition involve higher billing amounts.
- The 0-10,000 range has the least number of cases across all conditions, reflecting fewer cases with lower billing amounts.
- The counts in the 10,000-20,000 and 20,000-30,000 ranges are relatively consistent across different conditions, with slight variations.

## Summary

- Obesity and Diabetes have the highest counts in the 30,000+ billing range, indicating that these conditions tend to result in higher billing amounts.
- Arthritis has the highest count in the 0-10,000 range, while Obesity has the lowest.
- The overall trend shows that higher billing amounts are more common, with the 30,000+ range consistently having the highest counts across all conditions.

## Medication Patterns

### 1. What are the most commonly prescribed medications?



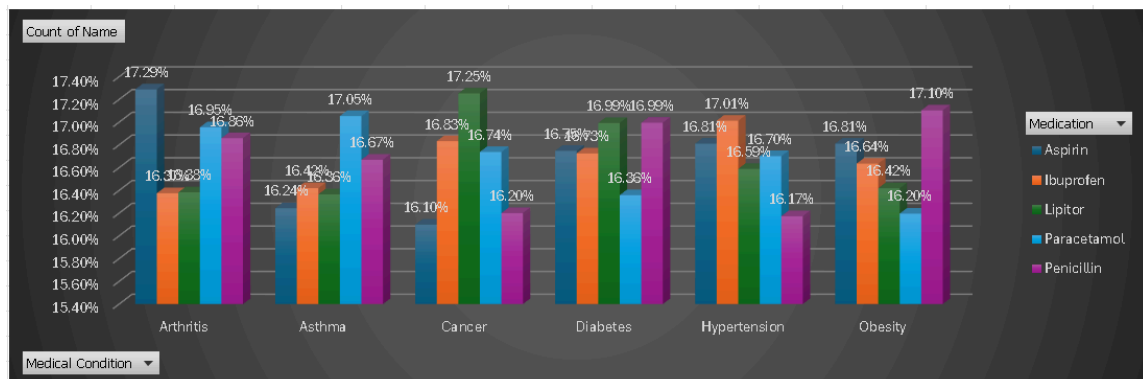
- Lipitor has the highest count with 11,140 occurrences. This suggests that Lipitor is the most frequently recorded medication among the listed options.
- Penicillin has the lowest count with 11,068 occurrences. This is still quite close to the counts of the other medications.
- The counts for Aspirin (11,094), Ibuprofen (11,127), and Paracetamol (11,071) are all quite close to each other and to the highest count.
- The difference in counts among the medications is minimal, ranging from 62 (between Lipitor and Penicillin) to 136 (between Lipitor and Paracetamol).
- Each medication contributes a similar proportion to the grand total of 55,500. The differences in counts are relatively small compared to the total, indicating a fairly balanced distribution.

## Summary

- Lipitor is the most frequently recorded medication, while Penicillin has the fewest occurrences among the listed medications.
- The differences in counts are minor, suggesting that all medications are represented fairly evenly in the dataset.
- The data shows a balanced distribution with no single medication dominating the counts significantly.

## 2. Are there any correlations between medications and medical conditions?

Count of Name	Column Labels				
Row Labels	Aspirin	Ibuprofen	Lipitor	Paracetamol	Penicillin
Arthritis	17.29%	16.37%	16.38%	16.95%	16.86%
Asthma	16.24%	16.42%	16.36%	17.05%	16.67%
Cancer	16.10%	16.83%	17.25%	16.74%	16.20%
Diabetes	16.75%	16.73%	16.99%	16.36%	16.99%
Hypertension	16.81%	17.01%	16.59%	16.70%	16.17%
Obesity	16.81%	16.64%	16.42%	16.20%	17.10%



- Arthritis: Highest percentage is for Aspirin (17.29%).
- Asthma: Highest percentage is for Paracetamol (17.05%).
- Cancer: Highest percentage is for Lipitor (17.25%).
- Diabetes: Highest percentage is for Lipitor and Penicillin (16.99%).
- Hypertension: Highest percentage is for Ibuprofen (17.01%).
- Obesity: Highest percentage is for Penicillin (17.10%).
- Arthritis: Lowest percentage is for Ibuprofen (16.37%).
- Asthma: Lowest percentage is for Aspirin (16.24%).

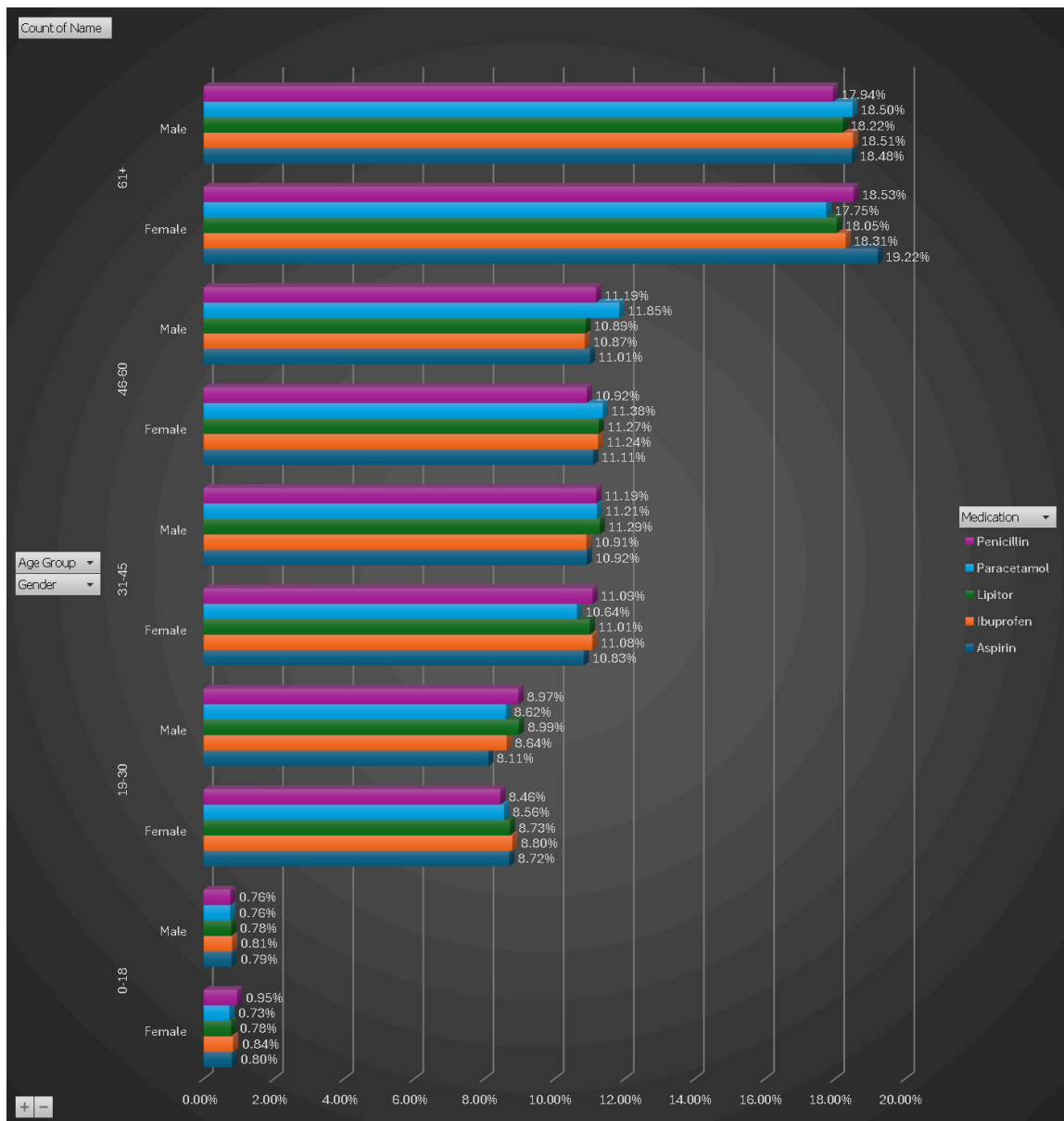
- Cancer: Lowest percentage is for Aspirin (16.10%).
- Diabetes: Lowest percentage is for Paracetamol (16.36%).
- Hypertension: Lowest percentage is for Penicillin (16.17%).
- Obesity: Lowest percentage is for Paracetamol (16.20%).
- Lipitor and Penicillin are frequently among the top medications for several conditions, with varying preferences across conditions.
- Paracetamol shows the highest percentage in Asthma and Diabetes, while Aspirin shows the highest in Arthritis.
- The percentages are fairly balanced, with none of the medications showing a significant dominance across all conditions.

### Summary

- Different conditions have varying preferences for medications, with some medications consistently showing higher percentages for certain conditions.
- Lipitor and Penicillin often rank among the top for multiple conditions.
- The distribution of medication usage is fairly balanced, with no single medication overwhelmingly favored across all conditions.

### 3. How does medication usage vary by age or gender?

Count of Name		Column Labels				
Row Labels		Aspirin	Ibuprofen	Lipitor	Paracetamol	Penicillin
0-18		1.60%	1.64%	1.56%	1.49%	1.71%
Female		0.80%	0.84%	0.78%	0.73%	0.95%
Male		0.79%	0.81%	0.78%	0.76%	0.76%
19-30		16.83%	17.44%	17.71%	17.18%	17.43%
Female		8.72%	8.80%	8.73%	8.56%	8.46%
Male		8.11%	8.64%	8.99%	8.62%	8.97%
31-45		21.76%	21.99%	22.30%	21.85%	22.28%
Female		10.83%	11.08%	11.01%	10.64%	11.09%
Male		10.92%	10.91%	11.29%	11.21%	11.19%
46-60		22.12%	22.11%	22.15%	23.23%	22.11%
Female		11.11%	11.24%	11.27%	11.38%	10.92%
Male		11.01%	10.87%	10.89%	11.85%	11.19%
61+		37.70%	36.82%	36.27%	36.25%	36.47%
Female		19.22%	18.31%	18.05%	17.75%	18.53%
Male		18.48%	18.51%	18.22%	18.50%	17.94%



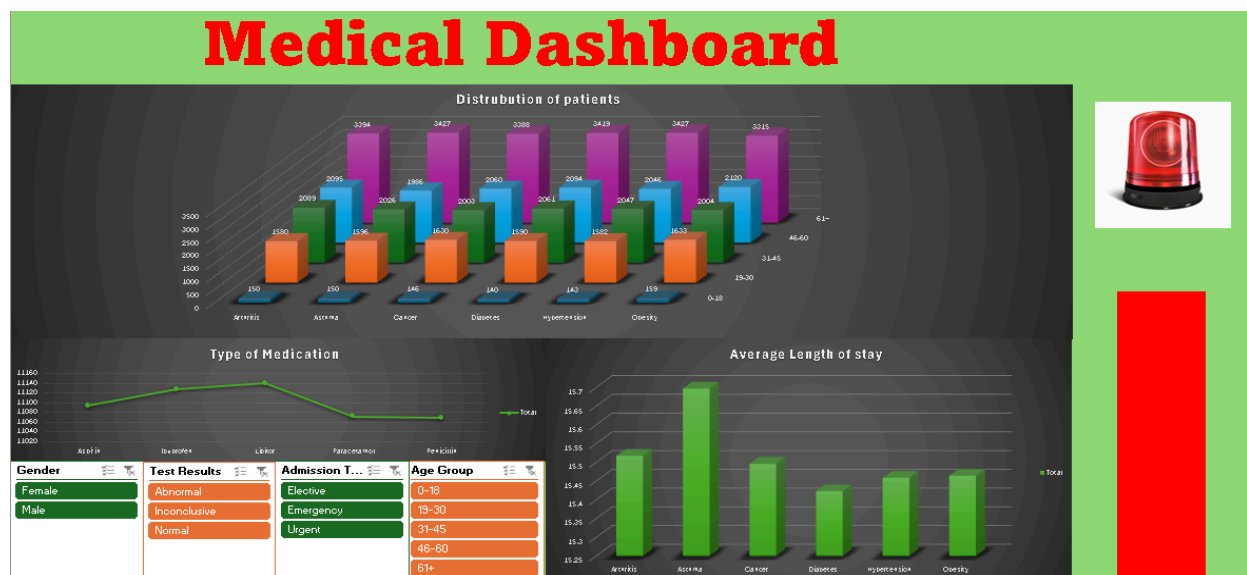
- **0-18:** The percentage usage of medications is relatively low for this age group across all medications, with **Penicillin** having the highest percentage (1.71%).
- **19-30:** Medication usage increases significantly. **Lipitor** has the highest usage (17.71%) in this age group.
- **31-45:** This age group shows the highest percentage usage for most medications, with **Lipitor** leading (22.30%).
- **46-60:** This group also shows high usage, with **Paracetamol** having the highest percentage (23.23%).
- **61+:** The highest percentage usage overall, with **Aspirin** leading (37.70%).
- **Female:** Generally has slightly higher medication usage percentages in most age groups, especially in the 46-60 age group for **Paracetamol**.

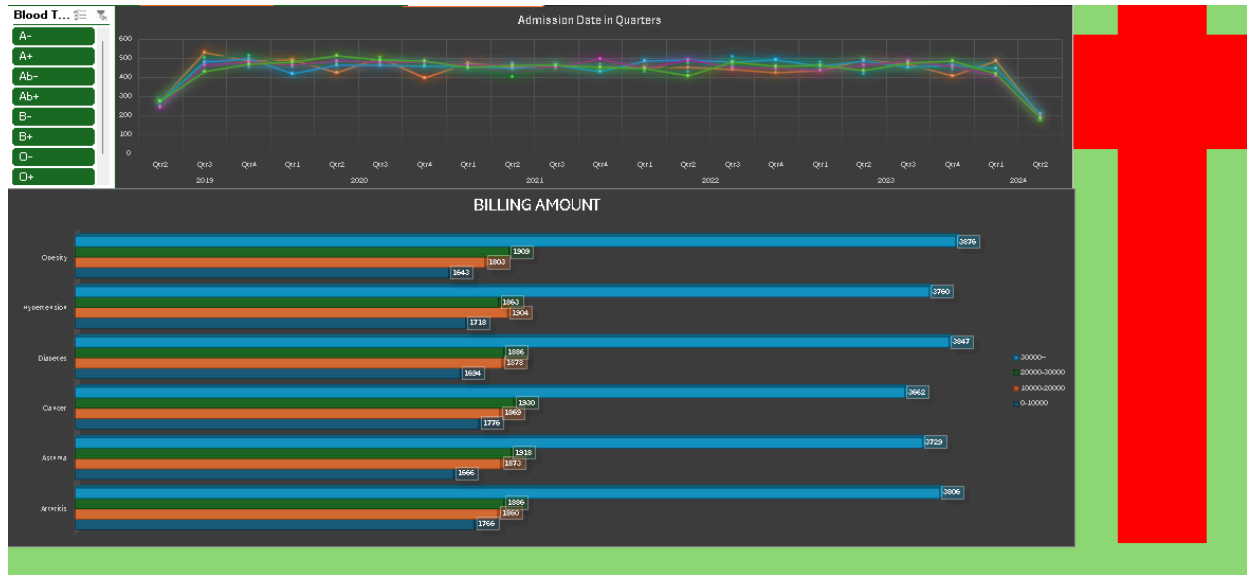
- **Male:** Shows similar usage patterns but with slightly lower percentages compared to females in most categories.
- **Aspirin** is most frequently used in the 61+ age group.
- **Ibuprofen** and **Lipitor** also show high usage in the 61+ age group but slightly lower than **Aspirin**.
- **Paracetamol** shows high usage in the 46-60 age group.
- **Penicillin** has the highest usage among the younger age groups, particularly 0-18.

## Summary

- **Medication usage** increases with age, with the 61+ age group showing the highest percentages for all medications.
- **Aspirin** is most prevalent among older adults, while **Penicillin** is more common among younger age groups.
- **Females** generally show slightly higher medication usage percentages compared to males across most age groups.
- The distribution highlights a clear trend of increased medication use with advancing age.

## Dashboard





A custom dashboard was created by combining pivot charts together and filtering them using custom slicers linking each of the slicers with the charts in the sheet.

Visual information like distribution of patients, type of medication used, average length of stay for patients, time of admission and billing amount can be interpreted using individual or simultaneous filters like patient Gender, Test result, Admission Type, Age group and Blood Type.