```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sea
import numpy as np

df=pd.read_excel("C:\\Users\\ishav\\Downloads\\Papollo-Healtcare-
Dataset.xlsx")

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7157 entries, 0 to 7156
Data columns (total 13 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Patient_ID               7157 non-null   int64
 1   Admit_Date               7157 non-null   datetime64[ns]
 2   Discharge_Date           7157 non-null   datetime64[ns]
 3   Diagnosis                7157 non-null   object
 4   Bed_Occupancy            7157 non-null   object
 5   Test                     7157 non-null   object
 6   Doctor                   7157 non-null   object
 7   Followup Date            7157 non-null   object
 8   Feedback                 7157 non-null   float64
 9   Billing Amount           7157 non-null   int64
 10  Health Insurance Amount  7157 non-null   float64
 11  month                    7157 non-null   int32
 12  year                     7157 non-null   int32
dtypes: datetime64[ns](2), float64(2), int32(2), int64(2), object(5)
memory usage: 671.1+ KB
```

```
print(df.describe())
```

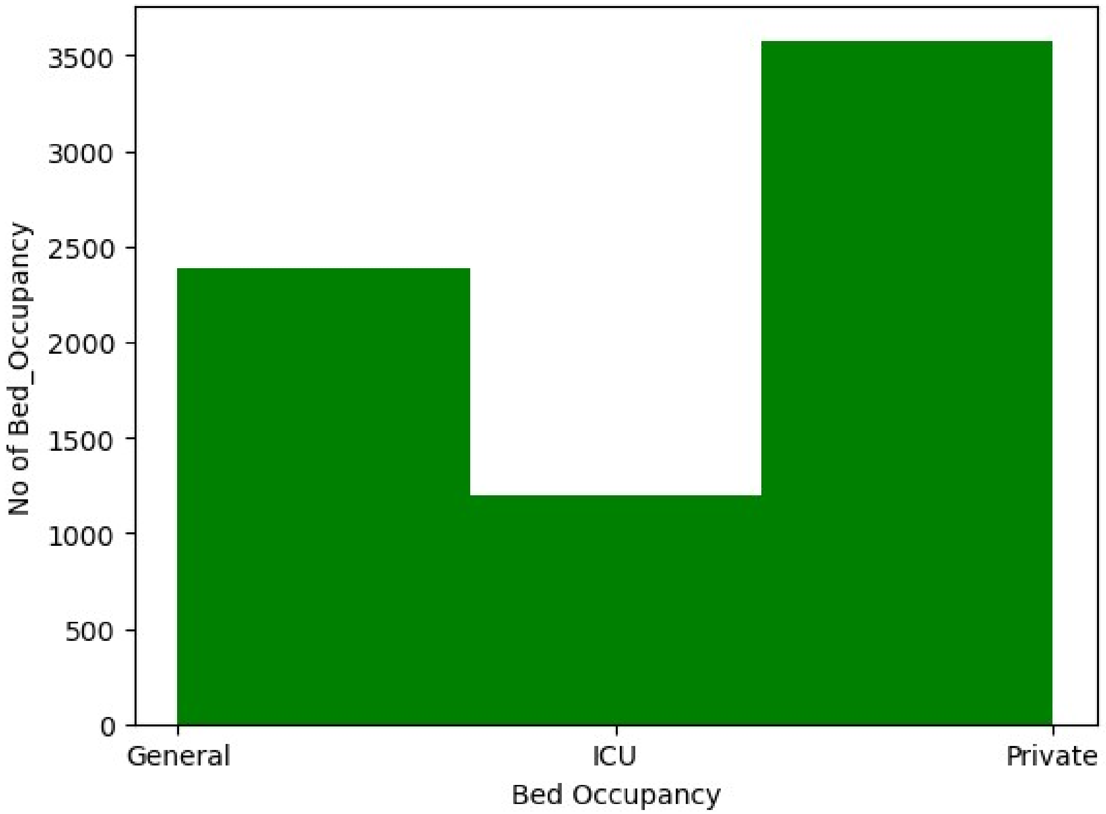
```
         Patient_ID                     Admit_Date  \
count   7157.000000                           7157
mean   27149.000000  2023-07-24 09:14:42.772111360
min    23571.000000            2022-12-05 00:00:00
25%    25360.000000            2023-04-03 00:00:00
50%    27149.000000            2023-07-22 00:00:00
75%    28938.000000            2023-11-11 00:00:00
max    30727.000000            2024-03-06 00:00:00
std     2066.192271                            NaN

                      Discharge_Date     Feedback  Billing Amount  \
count                           7157  7157.000000     7157.000000
mean   2023-08-01 15:14:27.681989376     4.723543    26607.891854
min              2023-01-12 00:00:00     3.500000     1223.000000
25%              2023-04-11 00:00:00     4.500000     5454.000000
50%              2023-07-31 00:00:00     5.000000    12348.000000
75%              2023-11-20 00:00:00     5.000000    43522.000000
```

```
max                2024-03-07 00:00:00     5.000000    95867.000000
std                                NaN     0.479901    25791.931009

       Health Insurance Amount        month         year
count                7157.000000  7157.000000  7157.000000
mean                23947.102669     5.754227  2023.125192
min                  1100.700000     1.000000  2022.000000
25%                  4908.600000     2.000000  2023.000000
50%                 11113.200000     5.000000  2023.000000
75%                 39169.800000     9.000000  2023.000000
max                 86280.300000    12.000000  2024.000000
std                 23212.737908     3.689429     0.355785

df

      Patient_ID Admit_Date Discharge_Date         Diagnosis
Bed_Occupancy  \
0          23571 2022-12-31     2023-01-12  Viral Infection
General
1          27781 2023-01-04     2023-01-12           Typhoid
ICU
2          24413 2023-01-05     2023-01-12          Malaria
General
3          27360 2023-01-05     2023-01-12               Flu
Private
4          26097 2023-01-06     2023-01-12  Viral Infection
General
...          ...        ...            ...               ...
...
7152       29885 2024-03-03     2024-03-07           Typhoid
Private
7153       26517 2024-03-04     2024-03-07        Pneumonia
Private
7154       28201 2024-03-05     2024-03-07               Flu
General
7155       24833 2024-03-06     2024-03-07           Typhoid
Private
7156       30727 2024-03-06     2024-03-07  Viral Infection
Private

                Test       Doctor        Followup Date  Feedback  Billing
Amount  \
0                MRI    Jay Sinha  2023-01-20 00:00:00       5.0
13234
1                MRI  Jaya Yaadav  2023-01-16 00:00:00       4.0
12241
2            CT Scan    Jay Sinha  2023-01-15 00:00:00       5.0
35255
3              X-Ray  Jaya Yaadav  2023-01-19 00:00:00       5.0
6582
```

```
4     Blood Test  Jaya Yaadav  2023-02-06 00:00:00       4.9
4567
...          ...          ...                    ...       ...
...
7152  Blood Test    Jay Sinha  2024-03-10 00:00:00       5.0
2342
7153     CT Scan  Jaya Yaadav  2024-03-14 00:00:00       5.0
24943
7154  Blood Test    Jay Sinha             00:00:00       3.5
2343
7155       X-Ray     Mark Joy  2024-03-15 00:00:00       4.9
3475
7156  Ultrasound  Jaya Yaadav  2024-03-11 00:00:00       5.0
66537

      Health Insurance Amount  month  year Patinets_stay_duration
0                     11910.6     12  2022                12 days
1                     11016.9      1  2023                 8 days
2                     31729.5      1  2023                 7 days
3                      5923.8      1  2023                 7 days
4                      4110.3      1  2023                 6 days
...                       ...    ...   ...                    ...
7152                   2107.8      3  2024                 4 days
7153                  22448.7      3  2024                 3 days
7154                   2108.7      3  2024                 2 days
7155                   3127.5      3  2024                 1 days
7156                  59883.3      3  2024                 1 days

[7157 rows x 14 columns]
```

```
df["Patinets_stay_duration"]=df["Discharge_Date"]-df["Admit_Date"]
```

```
plt.hist(df["Bed_Occupancy"],bins=3,color=["green"])
plt.xlabel("Bed Occupancy")
plt.ylabel("No of Bed_Occupancy")
plt.show()
```

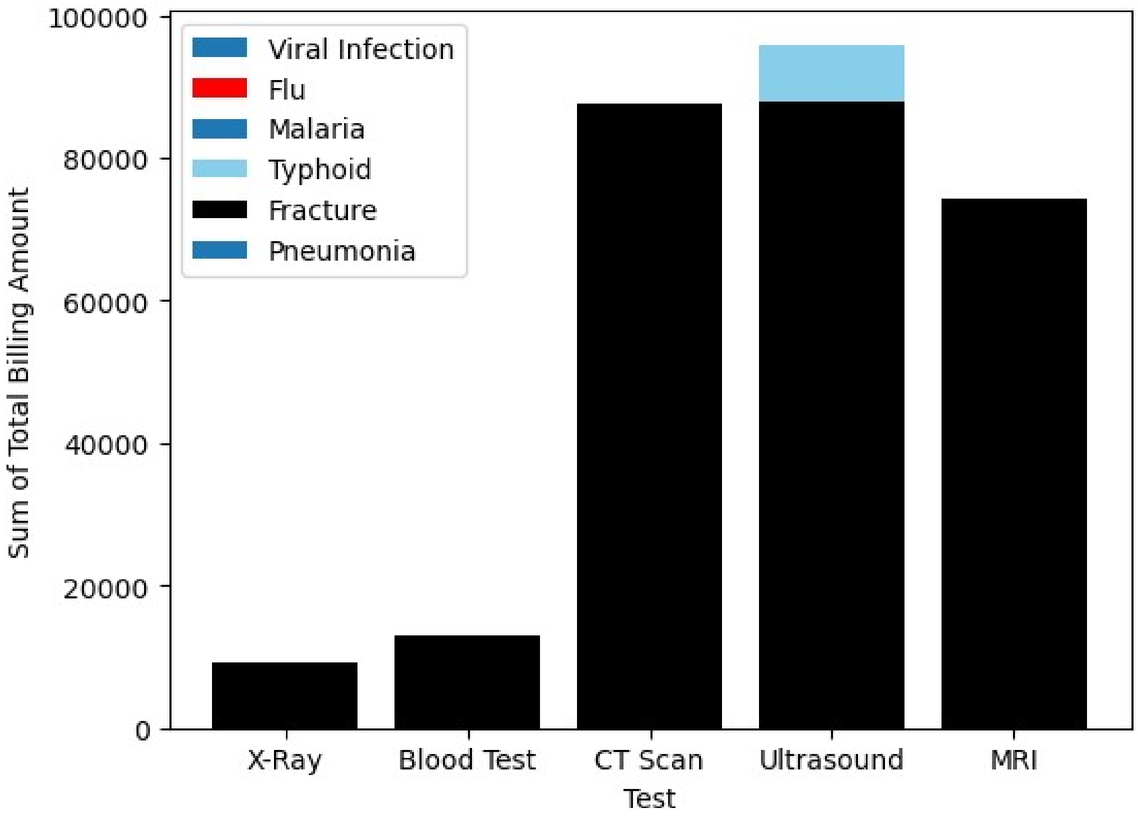


```
df=df.drop(columns=["Unnamed: 11","Unnamed: 12"])

df_group_by=df.groupby("Bed_Occupancy")["Billing
Amount"].sum().reset_index()
print(df_group_by)
```

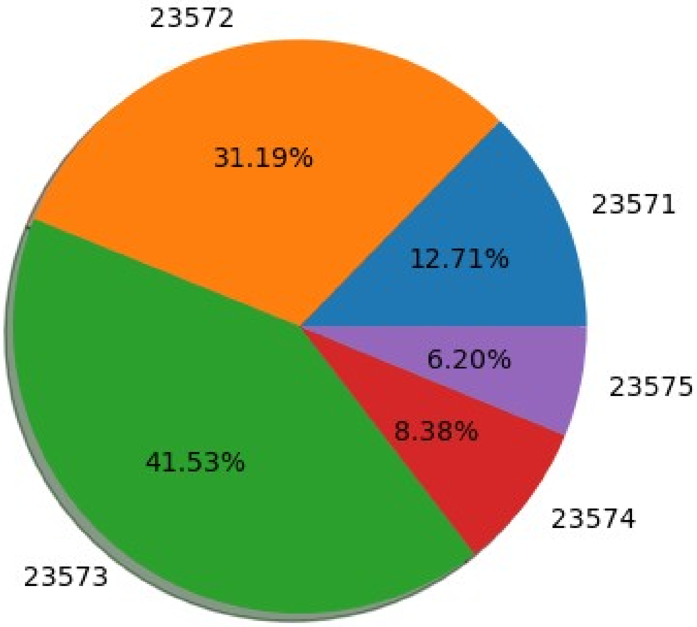
```
  Bed_Occupancy  Billing Amount
0       General        63164925
1           ICU        40447981
2       Private        86819776
```

```
df_group_by=df.groupby("Test")["Billing Amount"].sum().reset_index()

color_map={
    "Viral Infection": "green",
    "Flu": "red",
    "Malaria": "white",
    "Typhoid": "skyblue",
    "Fracture": "black",
    "Pneumonia": "yellow"
}
for Diagnosis,col in color_map.items():
    df_dia=df[df["Diagnosis"]==Diagnosis]
    plt.bar(df_dia["Test"],df_dia["Billing
Amount"],label=Diagnosis,color=col)
```

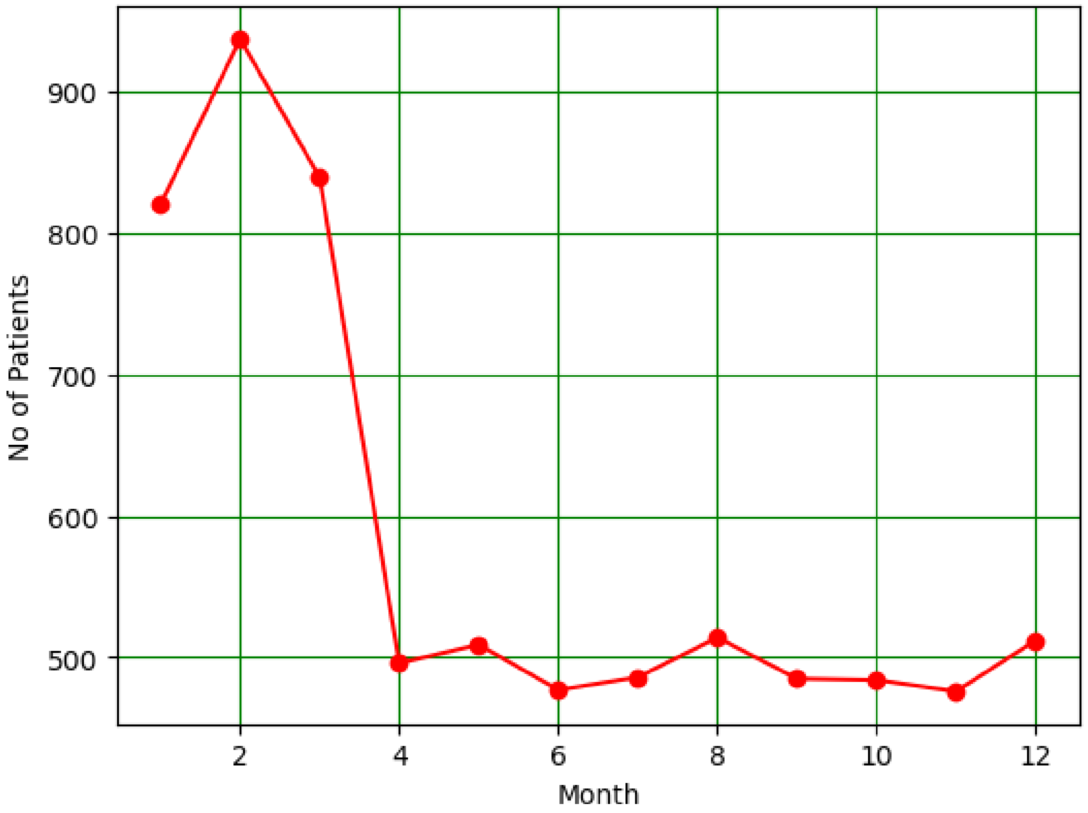
```
plt.legend()
plt.xlabel("Test")
df_group_by=df_group_by.sort_values(by="Billing
Amount",ascending=False)
plt.ylabel("Sum of Total Billing Amount")
plt.show()
```



```
df_group_by_Patient_ID=(
    df.groupby("Patient_ID")["Health Insurance
Amount"].sum().reset_index())
top3=df_group_by_Patient_ID.head(5)
plt.pie(top3["Health Insurance
Amount"],labels=top3["Patient_ID"],autopct="%1.2f%%",shadow=True)
plt.show()
```



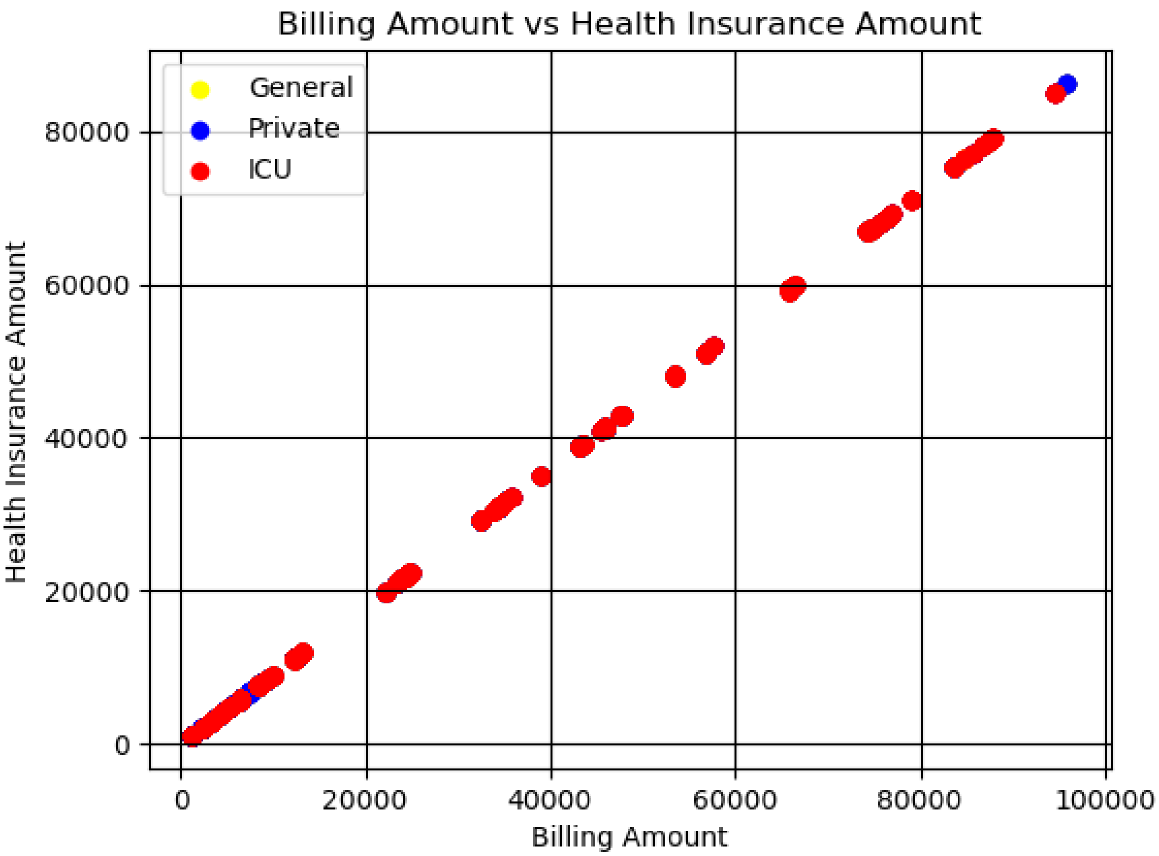
```
df["month"]=month=pd.to_datetime(df["Admit_Date"]).dt.month
df["year"]=month=pd.to_datetime(df["Admit_Date"]).dt.year

number_of_patients_month=df.groupby("month")
["Patient_ID"].count().reset_index()
plt.plot(

number_of_patients_month["month"],number_of_patients_month["Patient_ID
"],
    marker="o",color="red")
plt.xlabel("Month")
plt.ylabel("No of Patients")
plt.grid(color="green")
plt.show()
```



```
df

       Patient_ID Admit_Date Discharge_Date         Diagnosis
Bed_Occupancy  \
0           23571 2022-12-31     2023-01-12  Viral Infection
General
1           27781 2023-01-04     2023-01-12           Typhoid
ICU
2           24413 2023-01-05     2023-01-12          Malaria
General
3           27360 2023-01-05     2023-01-12               Flu
Private
4           26097 2023-01-06     2023-01-12  Viral Infection
General
...           ...        ...            ...               ...
...
7152        29885 2024-03-03     2024-03-07           Typhoid
Private
7153        26517 2024-03-04     2024-03-07        Pneumonia
Private
7154        28201 2024-03-05     2024-03-07               Flu
General
7155        24833 2024-03-06     2024-03-07           Typhoid
Private
```

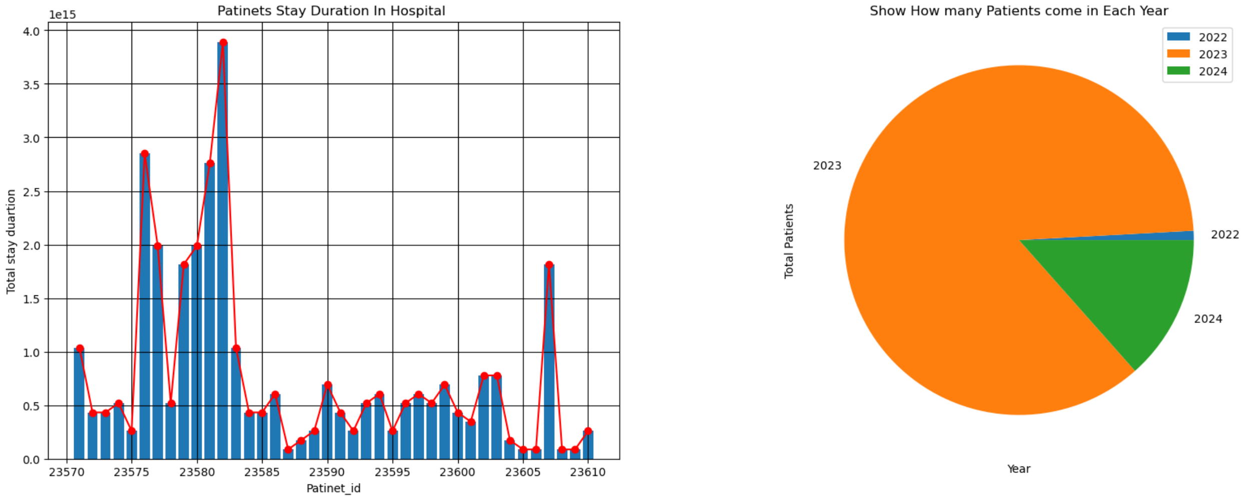
```
7156       30727 2024-03-06     2024-03-07  Viral Infection
Private

            Test       Doctor        Followup Date  Feedback  Billing
Amount  \
0            MRI    Jay Sinha  2023-01-20 00:00:00       5.0
13234
1            MRI  Jaya Yaadav  2023-01-16 00:00:00       4.0
12241
2        CT Scan    Jay Sinha  2023-01-15 00:00:00       5.0
35255
3          X-Ray  Jaya Yaadav  2023-01-19 00:00:00       5.0
6582
4     Blood Test  Jaya Yaadav  2023-02-06 00:00:00       4.9
4567
...          ...          ...                  ...       ...
...
7152  Blood Test    Jay Sinha  2024-03-10 00:00:00       5.0
2342
7153     CT Scan  Jaya Yaadav  2024-03-14 00:00:00       5.0
24943
7154  Blood Test    Jay Sinha             00:00:00       3.5
2343
7155       X-Ray     Mark Joy  2024-03-15 00:00:00       4.9
3475
7156  Ultrasound  Jaya Yaadav  2024-03-11 00:00:00       5.0
66537

      Health Insurance Amount  month  year Patinets_stay_duration
0                     11910.6     12  2022                12 days
1                     11016.9      1  2023                 8 days
2                     31729.5      1  2023                 7 days
3                      5923.8      1  2023                 7 days
4                      4110.3      1  2023                 6 days
...                       ...    ...   ...                    ...
7152                   2107.8      3  2024                 4 days
7153                  22448.7      3  2024                 3 days
7154                   2108.7      3  2024                 2 days
7155                   3127.5      3  2024                 1 days
7156                  59883.3      3  2024                 1 days

[7157 rows x 14 columns]
```

```
try:
    color={"General":"yellow","Private":"Blue","ICU":"Red"}
    for bad_occupancy,color in color.items():
        df_deft=df[df["Bed_Occupancy"]==bad_occupancy]
        plt.scatter(df_deft["Billing Amount"],df_deft["Health
Insurance Amount"],c=color,label=bad_occupancy)
    plt.legend()
    plt.xlabel("Billing Amount")
    plt.ylabel("Health Insurance Amount")
    plt.title("Billing Amount vs Health Insurance Amount")
    plt.grid(color="black")
    plt.show()
except:
    print(" ")
```

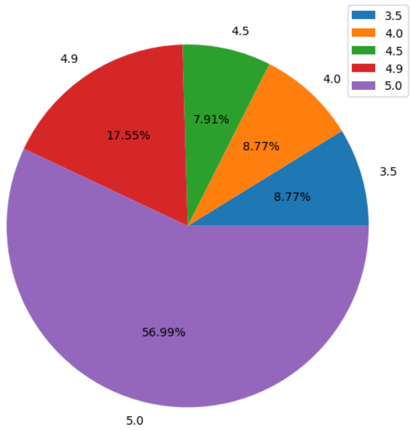


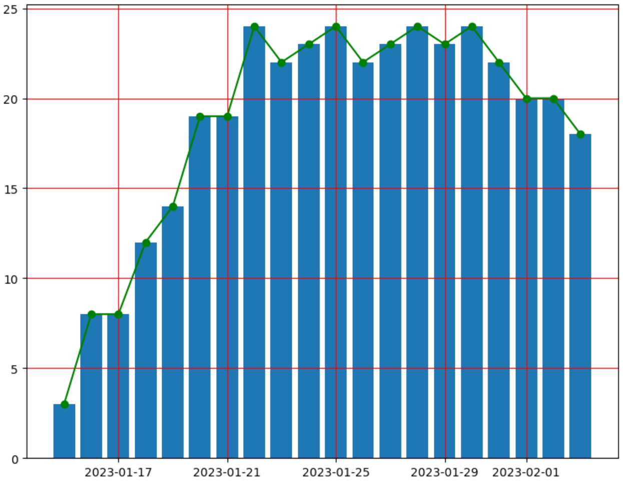
```
df_group_by=df.groupby("Patient_ID")
["Patinets_stay_duration"].sum().reset_index()
Top_5=df_group_by.head(40)
fig,axs=plt.subplots(1,2,figsize=(20,7))
df_group_by=df_group_by.sort_values(by="Patient_ID",ascending=False)
axs[0].bar(Top_5["Patient_ID"],Top_5["Patinets_stay_duration"])
axs[0].plot(Top_5["Patient_ID"],Top_5["Patinets_stay_duration"],marker
="o",color="Red")
axs[0].set_title("Patinets Stay Duration In Hospital ")
axs[0].set_xlabel("Patinet_id")
axs[0].set_ylabel("Total stay duartion")
axs[0].grid(color="black")

df_group_by_date=df.groupby("year")["Patient_ID"].sum().reset_index()
```

```
axs[1].pie(df_group_by_date["Patient_ID"],labels=df_group_by_date["yea
r"],autopct="%1.2f%%",)
axs[1].legend()
axs[1].set_title("Show How many Patients come in Each Year")
axs[1].set_xlabel("Year")
axs[1].set_ylabel("Total Patients")
axs[1].grid()
plt.show()
```



```
fig,axis=plt.subplots(1,2,figsize=(20,7))
df_group_by_date=df.groupby("Feedback")
["Doctor"].count().reset_index()

axis[0].pie(df_group_by_date["Doctor"],labels=df_group_by_date["Feedba
ck"],autopct="%1.2f%%")
axis[0].legend()
axs[0].set_title("Top Doctors")
axs[0].set_xlabel("Rating")



df_group_by_followup_date=df.groupby("Follow_up_date")
["Patient_ID"].count().reset_index()
top_5=df_group_by_followup_date.head(20)
axis[1].bar(top_5["Follow_up_date"],top_5["Patient_ID"])
axis[1].plot(top_5["Follow_up_date"],top_5["Patient_ID"],marker="o",co
lor="green")
axis[1].grid(color="red")
axs[1].set_title("Patinets Stay Duration In Hospital ")
axs[1].set_xlabel("Patinet_id")
plt.show()
```

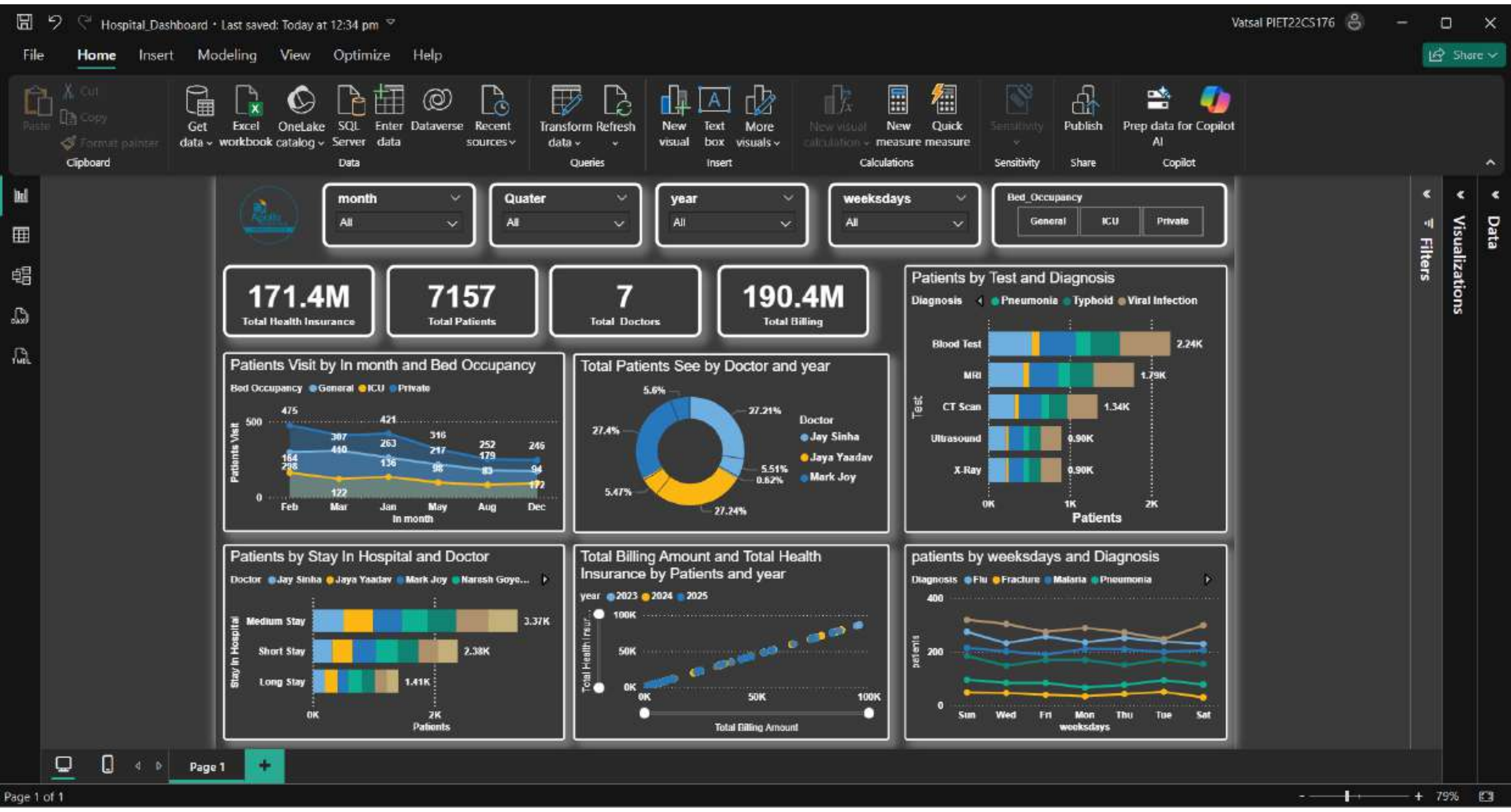

3.5
4.0
4.5
4.9
5.0
4.5
4.9
4.0
3.5
5.0
7.91%
8.77%
17.55%
8.77%
56.99%


25
20
15
10
5
0
2023-01-17
2023-01-21
2023-01-25
2023-01-29
2023-02-01

```
select count(*) from healthcare_dateset where `Followup Date` is null;

update healthcare_dateset set `Followup Date`="2025-05-11" where `Followup Date` is null;

select * from healthcare_dateset where year(`Followup Date`)=2025;

-- 1□ Write a query to count total number of patients in the hospital.
SELECT
        COUNT(*) AS total_number_patients
FROM
        healthcare_dateset;

-- Q-2 Display total billing amount collected per diagnosis type.
select * from healthcare_dateset;
select diagnosis,sum(`billing amount`) as total_billing_amount from healthcare_dateset
group by diagnosis;

-- Q-3 Find the average feedback rating for each doctor.
select doctor,round(avg(Feedback),1) from healthcare_dateset
group by doctor;

-- Q-4 Show all patients admitted in the ICU only.
select count(Patient_ID) as all_patients,Bed_Occupancy from healthcare_dateset where
Bed_Occupancy="ICU"
group by Bed_Occupancy;

-- Q-5 Get patients whose discharge date is later than '2023-01-10'.
select * from healthcare_dateset where date(`Followup Date`)>'2023-01-10';

-- Q-6 List the top 5 highest billing patients.
select Patient_ID,sum(`billing amount`) as higest_billing from healthcare_dateset
group by Patient_ID  order by higest_billing  desc limit 5;

-- Q-7 Display total health insurance claim amount per bed type.
select Bed_Occupancy,sum(`Health Insurance Amount`) as total_health_amount from healthcare_dateset
group by Bed_Occupancy;

--  Q-8 Get the patient ID, doctor, and billing amount for viral infection diagnosis.
select Patient_ID,Doctor,`Billing Amount`,Diagnosis from healthcare_dateset  where Diagnosis="viral
infection";

-- Q-9 Count how many follow-ups are scheduled for each month.
select count(`Followup Date`) as follow_ups ,month(`Followup Date`) as month from healthcare_dateset
group by month(`Followup Date`);


-- Q-10 Calculate the total days each patient stayed in the hospital.
select  Patient_ID,datediff(discharge_date,admit_date) as patient_stayed_total_days from
healthcare_dateset;

-- Q-11 Retrieve the top 3 doctors whose average billing amount per patient is the highest.
select Patient_ID,doctor,avg(`Billing Amount`) as avg_billing from healthcare_dateset
group by Patient_ID,doctor order by doctor desc limit 3;
```

```
-- Q-12 Find the percentage of total patients in each diagnosis category.
select round((COUNT(*) * 100.0 / (SELECT COUNT(*) FROM healthcare_dateset)),1) AS
Percentage_of_Patients,Diagnosis from healthcare_dateset
group by diagnosis;

-- Q-13 Identify patients who stayed longer than the average stay duration of the hospital.
with stayed_longer as
(select Patient_ID,datediff(discharge_date,admit_date)as stayed_longer_duration from healthcare_dateset
order by stayed_longer_duration desc )

select Patient_ID,stayed_longer_duration from stayed_longer where stayed_longer_duration>(select
avg(stayed_longer_duration) from stayed_longer) limit 1 ;


-- Q-14 Get patients with more than one follow-up scheduled and show count of follow-ups.
select Patient_ID,count(`Followup Date`) as follow_up from healthcare_dateset
group by Patient_ID having follow_up>1;

-- Q-15 Retrieve the patients admitted on a weekend and calculate their average feedback rating.
select dayofweek(admit_date) as week_days,Patient_ID,avg(Feedback) as avg_feedback_rating from
healthcare_dateset where dayofweek(admit_date)=7
group by Patient_ID,week_days;

-- Q-16 Find the longest continuous sequence of bed occupancy days per patient.

select * from
(
select Patient_ID,Bed_Occupancy,datediff(discharge_date,admit_date) as day_diff,
row_number() over (partition by Bed_Occupancy order by datediff(discharge_date,admit_date) desc)
as sequence_of_bed
from healthcare_dateset order by day_diff desc
)
as e where e.sequence_of_bed=1;

--  Q-17 Display only those doctors whose total insurance claim amount from patients is greater than 80%
of total billing.
with total_billing as
(
select doctor,round(sum(`health insurance amount`),1)
as total_claim_amount,
round(sum(`billing amount`)*0.80,1) as `80% of total billing` from healthcare_dateset group by Doctor)
select * from total_billing where  total_claim_amount>`80% of total billing` ;

-- Q-18 Rank doctors based on total number of patients handled per month using RANK() window
function.

select doctor,count(Patient_ID) as total_number_of_patients
,month(admit_date) as month,
dense_rank() over(partition by month(admit_date) order by count(doctor) desc )
as rank_dotor from healthcare_dateset
group by month(admit_date),doctor;

-- Q-19 Create a pivot-style output: Count of patients by Diagnosis (rows) and Bed Type (columns).

SELECT
```

```
    Diagnosis,
    SUM(CASE WHEN Bed_Occupancy = 'General' THEN 1 ELSE 0 END) AS General,
    SUM(CASE WHEN Bed_Occupancy = 'ICU' THEN 1 ELSE 0 END) AS ICU,
    SUM(CASE WHEN Bed_Occupancy = 'Private' THEN 1 ELSE 0 END) AS Private
FROM healthcare_dateset
GROUP BY Diagnosis
ORDER BY Diagnosis;
```

Hospital_Dashboard • Last saved: Today at 12:34 pm
Vatsal PIET22CS176
File
Home
Insert
Modeling
View
Optimize
Help
Share
Paste
Cut
Copy
Format painter
Clipboard
Get data
Excel workbook
OneLake catalog
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
New visual calculation
New measure
Quick measure
Calculations
Sensitivity
Sensitivity
Publish
Share
Prep data for AI
Copilot
Copilot
month
All
Quater
All
year
All
weeksdays
All
Bed_Occupancy
General
ICU
Private
171.4M
Total Health Insurance
7157
Total Patients
7
Total Doctors
190.4M
Total Billing
Patients by Test and Diagnosis
Diagnosis ● Pneumonia ● Typhoid ● Viral Infection
Blood Test
2.24K
MRI
1.79K
CT Scan
1.34K
Ultrasound
0.90K
X-Ray
0.90K
Test
0K
1K
2K
Patients
Patients Visit by In month and Bed Occupancy
Bed Occupancy ● General ● ICU ● Private
475
421
307
440
263
316
217
252
179
246
164
298
136
98
83
94
122
172
500
0
Patients Visit
Feb
Mar
Jan
May
Aug
Dec
In month
Total Patients See by Doctor and year
5.6%
27.21%
27.4%
5.51%
0.62%
5.47%
27.24%
Doctor
● Jay Sinha
● Jaya Yaadav
● Mark Joy
Patients by Stay In Hospital and Doctor
Doctor ● Jay Sinha ● Jaya Yaadav ● Mark Joy ● Naresh Goye...
Medium Stay
3.37K
Short Stay
2.38K
Long Stay
1.41K
Stay In Hospital
0K
2K
Patients
Total Billing Amount and Total Health Insurance by Patients and year
year ● 2023 ● 2024 ● 2025
100K
50K
0K
Total Health Insur...
0K
50K
100K
Total Billing Amount
patients by weeksdays and Diagnosis
Diagnosis ● Flu ● Fracture ● Malaria ● Pneumonia
400
200
0
patients
Sun
Wed
Fri
Mon
Thu
Tue
Sat
weeksdays
Filters
Visualizations
Data
Page 1
+
Page 1 of 1
79%