Hospital Healthcare Analysis

```
1 import pandas as pd
```

```
1 df = pd.read_excel('/content/drive/MyDrive/healthcare_dataset.xlsx')
```

```
1 df.head(10)
```

ID	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	
1	Bobby JacksOn	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18856.28	328	Urgent	2024-02-02	Paracetamol	
2	LesLie TErRy	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.33	265	Emergency	2019-08-26	Ibuprofen	In
3	DaNnY sMitH	76	Female	A-	Obesity	2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.10	205	Emergency	2022-10-07	Aspirin	
4	andrEw waTtS	28	Female	O+	Diabetes	2020-11-18	Kevin Wells	Hernandez Rogers and Vang,	Medicare	37909.78	450	Elective	2020-12-18	Ibuprofen	
5	adrIENNE bEll	43	Female	AB+	Cancer	2022-09-19	Kathleen Hanna	White-White	Aetna	14238.32	458	Urgent	2022-10-09	Penicillin	
6	EMILY JOHNSOn	36	Male	A+	Asthma	2023-12-20	Taylor Newton	Nunez-Humphrey	UnitedHealthcare	48145.11	389	Urgent	2023-12-24	Ibuprofen	
7	edwArD EDWaRDs	21	Female	AB-	Diabetes	2020-11-03	Kelly Olson	Group Middleton	Medicare	19580.87	389	Emergency	2020-11-15	Paracetamol	In
8	CHrisTInA MARtinez	20	Female	A+	Cancer	2021-12-28	Suzanne Thomas	Powell Robinson and Valdez,	Cigna	45820.46	277	Emergency	2022-01-07	Paracetamol	In
9	JASmINe aGuIlaR	82	Male	AB+	Asthma	2020-07-01	Daniel Ferguson	Sons Rich and	Cigna	50119.22	316	Elective	2020-07-14	Aspirin	
10	ChRISTopher BerG	58	Female	AB-	Cancer	2021-05-23	Heather Day	Padilla-Walker	UnitedHealthcare	19784.63	249	Elective	2021-06-22	Paracetamol	In

Next steps: Generate code with df | View recommended plots | New interactive sheet

Cleaning

```
1 df.dtypes
```

	0
ID	int64
Name	object
Age	int64
Gender	object
Blood Type	object
Medical Condition	object
Date of Admission	datetime64[ns]
Doctor	object
Hospital	object
Insurance Provider	object
Billing Amount	float64
Room Number	int64
Admission Type	object
Discharge Date	datetime64[ns]
Medication	object
Test Results	object

dtype: object

```
df.isnull().sum()
```

	0
ID	0
Name	0
Age	0
Gender	0
Blood Type	0
Medical Condition	0
Date of Admission	0
Doctor	0
Hospital	0
Insurance Provider	0
Billing Amount	0
Room Number	0
Admission Type	0
Discharge Date	0
Medication	0
Test Results	0

dtype: int64

```
int(df.duplicated().sum())
```

```
0
```

Summary

```
num_of_patients = df['ID'].count()
avg_patient_age = int(df['Age'].mean())
num_of_hospitals = df['Hospital'].nunique()
num_of_rooms = df['Room Number'].nunique()
```

```
num_of_doctors = df['Doctor'].nunique()
num_of_insurance_providers = df['Insurance Provider'].nunique()
total_billing_amount = df['Billing Amount'].sum()
print(f"Number of patients: {num_of_patients}")
print(f"Average patient age: {avg_patient_age:.2f}")
print(f"Number of hospitals: {num_of_hospitals}")
print(f"Number of rooms: {num_of_rooms}")
print(f"Number of doctors: {num_of_doctors}")
print(f"Number of insurance providers: {num_of_insurance_providers}")
print(f"Total billing amount: {total_billing_amount:.2f}")
```

```
Number of patients: 54966
Average patient age: 51.00
Number of hospitals: 39876
Number of rooms: 400
Number of doctors: 40341
Number of insurance providers: 5
Total billing amount: 1404068337.89
```

Patients

```
patient_by_gender = df['Gender'].value_counts()
patient_by_gender
```

Gender	count
Male	27496
Female	27470

dtype: int64

```
patient_by_hospital = df.groupby('Hospital')['ID'].count().sort_values(ascending=False)
patient_by_hospital
```

Hospital	ID
LLC Smith	44
Ltd Smith	39
Smith Ltd	37
Johnson PLC	37
Smith PLC	36
...	...
Jennings Caldwell Christensen, and	1
Jennings Jensen and Waters,	1
Jennings LLC	1
Jennings Murray and Davies,	1
Jenkins-Vasquez	1

39876 rows × 1 columns

dtype: int64

```
doctors_per_patient = df.groupby('Doctor')['ID'].count()
doctors_per_patient.sort_values(ascending=False)
```

Doctor	ID
Michael Smith	27
John Smith	22
Robert Smith	21
Michael Johnson	20
James Smith	20
...	...
Henry Ruiz	1
Henry Santiago	1
Henry Scott	1
Henry Shelton	1
Henry Perez	1

40341 rows × 1 columns

dtype: int64

Blood Type

```
common_blood_type = df['Blood Type'].mode()[0]
common_blood_type
```

```
'A-'
```

```
blood_type_count = df['Blood Type'].value_counts()
blood_type_count
```

Blood Type	count
A-	6898
A+	6896
B+	6885
AB+	6882
AB-	6874
B-	6872
O+	6855
O-	6804

dtype: int64

Admission Type

```
admission_type_count = df['Admission Type'].value_counts()
admission_type_count
```

Admission Type	count
Elective	18473
Urgent	18391
Emergency	18102

dtype: int64

```
admissionType_by_hospital = df.groupby('Hospital')['Admission Type'].value_counts()
admissionType_by_hospital.sort_values(ascending=False)
```

Hospital	Admission Type	count
LLC Smith	Emergency	19
Inc Smith	Urgent	16
Johnson Inc	Elective	16
Ltd Smith	Urgent	16
PLC Smith	Elective	16
...	...	...
Jackson-Evans	Urgent	1
Jackson-Fowler	Elective	1
Jackson-Fuller	Emergency	1
Jackson-Garcia	Emergency	1
Jackson-Carter	Urgent	1

44138 rows × 1 columns

dtype: int64

Length Of Stay

```
length_of_stay = df['Discharge Date'] - df['Date of Admission']
length_of_stay.value_counts()
```

	count
21 days	1943
20 days	1893
6 days	1889
14 days	1888
11 days	1872
19 days	1866
7 days	1864
27 days	1862
9 days	1856
30 days	1854
25 days	1850
4 days	1848
23 days	1845
13 days	1845
29 days	1844
28 days	1833
2 days	1830
3 days	1828
12 days	1823
5 days	1819
1 days	1808
8 days	1807
26 days	1800
18 days	1798
17 days	1793
10 days	1788
15 days	1771
22 days	1767
16 days	1745
24 days	1737

dtype: int64

```
length_of_stay.mean().days
```

15

Billing Amount

```
billing_amount_per_provider = df.groupby('Insurance Provider')['Billing Amount'].sum()
billing_amount_per_provider
```

Insurance Provider	Billing Amount
Aetna	2.764987e+08
Blue Cross	2.804091e+08
Cigna	2.843341e+08
Medicare	2.829110e+08
UnitedHealthcare	2.799154e+08

dtype: float64

```
billing_amount_per_hospital = df.groupby('Hospital')['Billing Amount'].sum()
billing_amount_per_hospital.sort_values(ascending=False)
```

Hospital	Billing Amount
Johnson PLC	1081477.32
LLC Smith	1030189.88
Smith PLC	1029424.47
Ltd Smith	1003365.53
Smith Ltd	970035.87
...	...
Clements-Bowman	-1277.65
Diaz-Bishop	-1353.70
Fitzpatrick, Nielsen and Mcdonald	-1774.04
Juarez-Clark	-2008.49
Medina and Elliott Stewart,	-2633.24

39876 rows × 1 columns

dtype: float64

```
max_billing = df['Billing Amount'].max()
patient_with_max_billing = df[df['Billing Amount']==max_billing]
patient_with_max_billing
```

	ID	Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Res
36349	36350	tOdd CARrlLIO	51	Female	A+	Hypertension	2023-09-08	Kathleen Griffin	Griffin Group	Blue Cross	52764.28	209	Elective	2023-10-04	Ibuprofen	No

```
df['Year'] = df['Date of Admission'].dt.year
billing_per_year = df.groupby('Year')['Billing Amount'].sum()
billing_per_year.astype(int)
```

Year	Billing Amount
2019	187511030
2020	283952664
2021	277178298
2022	278612039
2023	279670832
2024	97143472

dtype: int64