Assignment 3
Gathering, Scraping, Munging and Cleaning Data

Members:
Yash Revadekar - 002738776
Krutik Kanakia - 002787847
Ikra Bagwan - 002794307

Sample data from every table:

Doctor Table:

Doc_ID	lst_nm	frst_nm	pri_spec	cty	st	phn_numbr	age
1407814791	PUGH	ROBERT	INTERNAL MEDICINE	MURRELLS INLET	SC		35
1407491400	MCCLAIN	REBECCA	NURSE PRACTITIONER	CROCKER	MO	8751452635	44
1407164437	KORENCHEN	PATRICIA	CLINICAL SOCIAL WORKER	ALBUQUERQUE	NM	5053858496	34
1407256159	MANN	EDITH	CLINICAL SOCIAL WORKER	PENN YAN	NY	3155367329	56
1407505571	KOWITZ	JACQUELINE	INTERNAL MEDICINE	ROCKVILLE	MD		45
1407254881	SAUCIER	KESHA	NURSE PRACTITIONER	WILMINGTON	CA	8751452655	45

Payment Table:

Pay_ID	Pay_Method	Pay_Date
1	Cash	11/28/2021
2	Card	11/19/2021
3	Cash	1/13/2022
4		2/20/2022
5	Card	2/28/2022
6	Cash	11/15/2021
7	Card	10/15/2021

Bill_payment Table:

Bill_ID	Bill_Date	P_ID	Doc_Fees	Room_Charge
1	11/25/2021	1	50000	10000
2	11/15/2021	2	80000	15000
3	1/10/2022	3	75000	12000
4	2/12/2022	4	45000	8000
5	3/20/2022	5	60000	7500
6	11/15/2021	6	57500	4500
7	10/5/2021	7	56000	8500
8	1/10/2022	8	54500	4500
9	3/12/2022	9	53000	6500

Patient Table:

P_ID	F_name	L_name	Age	Gender	Phone	Email	B_Grp	Address	State	City	Admit_Date	Discharge_Date
1	Chris	Brown	56	M	8578452154	cb@gmail.com	A+	56, Washington Street	MA	Boston	11/10/2021	11/25/2021
2	Nick	Jonas	45	M	8756984125	nj@gmail.com	B-	45, Lambert Street	TX	Dallas	10/20/2021	11/15/2021
3	Nikki	Brook	67	F	8796521453	nb@gmail.com	O+	15, Huntington Ave	FL	Tampa	1/7/2022	1/10/2022
4	Robert	Costa	79	M	8763214589	rc@gmail.com	A-	25, Breaking Rocks	MA	Lowell	2/6/2022	2/12/2022
5	Kelly	Mcardle	35	F	8796542314	km@gmail.com	A+	88, Center Street	MA	Boston	3/15/2022	3/20/2022

Record Table:

R_ID	Doc_ID	P_ID	Bill_ID
1	1407160575	1	1
2	1407164437	2	2
3	1407172521	3	3
4	1407182421	4	4
5	1407230097	5	5
6	1407234792	6	6

Any code and scripts you used :

All the code and scripts that were used in this assignment is provided in the GitHub link below.

You must find sources of data :

https://data.cms.gov/provider-data/topics/doctors-clinicians

Brief ReadMe File :

Project Description
An integrated system comprising numerous entities makes up the hospital management system. The management of all these elements would be difficult. We are therefore creating a single database for hospitals to address this issue and improve system efficiency by maintaining all the information needed to administer hospitals. Additionally, with the aid of this project, we will use visualization tools to exhibit and convey the data for prevalent trends for the breakouts. This would make it easier to comprehend analytics based on data or information collected from several hospitals.

Doctor Table:
We used python programs to web scrape the data for the doctor table from the Browse AI website. Our table has eight columns, and using Jupyter, we were able to read the doctor's csv file and remove all of the null entries. By grouping by one column (primary spec) and count (doc id), we were able to see the data. Next, we used matplotlib to plot the following bar graph. Finally, we group by primary spec once more, got the mean for the age, and plot the graph for age. After finishing everything, we entered this information into our database.

Patient Table:
We manually input the data for the patients table. We have 13 columns in our table, and using Jupyter, we were able to read the patient's csv file and remove all of the null values. We have verified the accuracy of the data for this table by displaying all the rows with null values. By converting the admit date into a month and year using the date-time format, we then grouped the data by monthyear, counted the patient IDs, and generated a line graph. After finishing everything, we entered this information into our database.

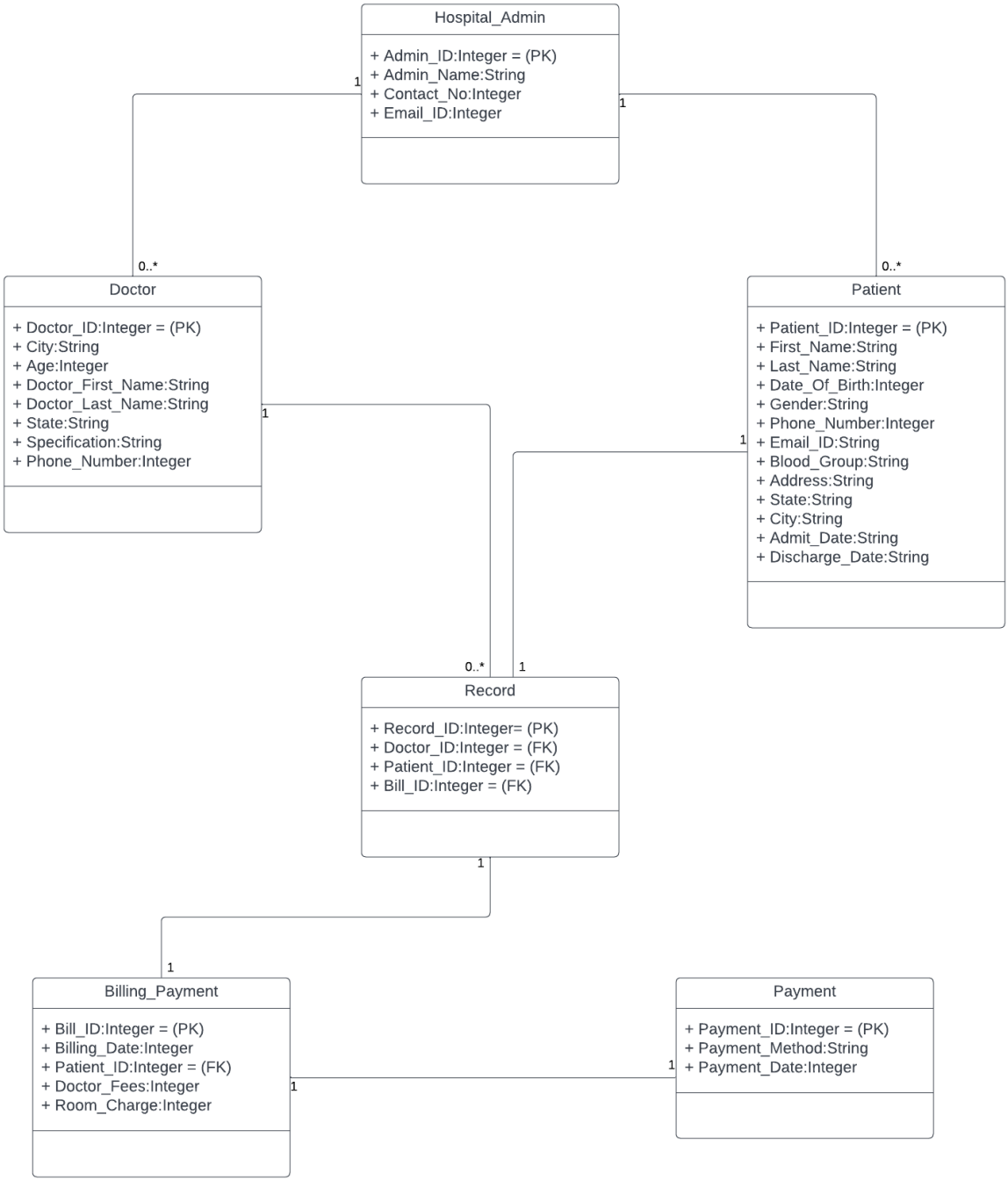
Payment Table:
We manually entered the data for the Payments table. Our database has three columns, and using Jupyter, we were able to read the payment's csv file and remove all of the null entries. After finishing everything, we entered this information into our database.

Bill Payment Table:
We manually entered the data for the bill payment table. Our database has five columns, and using Jupyter, we were able to read the bill payment csv file and remove all of the null entries. When grouping by monthyear and taking the count of the room charge, we turned the bill date into a month year by date-time before plotting the line graph. After finishing everything, we entered this information into our database.

Record Table:
We manually entered the data into the Record table. Our database has three columns, and using Jupyter, we were able to read the payment's csv file and remove all of the null entries. After finishing everything, we entered this information into our database.

Respond to any feedback you received in assignment 2:

According to the feedback we received, we restructured our ERD diagram.



Use-Cases :

1.To show number of patients a doctor serves in each city :

```
Select
        p.F_name
        ,p.L_name
        ,p.City
        ,count(*) as count_of_doctors
From patient p
Left join record r on
        r.P_ID = p.P_ID
Group by 1,2,3
;
```

2. To show amount of room charged in each city for Patient :

```
Select
        p.F_name
        ,p.L_name
        ,p.City
        ,sum(room_charge) as room_charge
From patient p
Left join record r on
        r.P_ID = p.P_ID
Left join billpayment bp on
        bp.bill_id = r.bill_id
Group by 1,2,3
;
```

3. Show billing date of patients:

```
Select
        p.F_name
        ,p.L_name
       ,(Bill_Date) as bill_date
From patient p
Left join billpayment bp on
        bp.P_ID = p.P_ID
Group by 1,2,3
;
```

4. View the number of patients served by doctors in a particular state:

```
select p.state, count(*) as count_of_doctors From Patient p LEFT JOIN record r
on r.P_ID=p.P_ID Group by 1;
```

5. To show highest-paid medical professional in each city:

```
Select
        d.frst_nm
        ,d.lst_nm
        ,d.cty
        ,sum(DOC_Fees) as doc_fees
From doctor d
Left join record r on
        r.Doc_ID = d.Doc_ID
Left join billpayment bp on
        bp.bill_id = r.bill_id
Group by 1,2,3;
```

6. Number of patients treated in each speciality :

```
select d.pri_spec, count(*) as count_of_patients From Doctor d LEFT JOIN
record r on r.Doc_ID=d.Doc_ID Group by 1;
```

7. Specialty bringing in highest room charges per state:

```
Select
        d.pri_spec
        ,P.state
        ,sum(room_charge) as doc_fees
From doctor d
Left join record r on
        r.Doc_ID = d.Doc_ID
Left join billpayment bp on
        bp.bill_id = r.bill_id
Group by 1,2
;
```

8. Doctor with highest fees in each state:

```
Select
        d.frst_nm
        ,d.lst_nm
        ,d.st
        ,sum(DOC_Fees) as doc_fees
From doctor d
Left join record r on
        r.Doc_ID = d.Doc_ID
Left join billpayment bp on
        bp.bill_id = r.bill_id
Group by 1,2,3;
```

9. Doctor served the highest patients in every state :

```
select d.frst_nm,d.lst_nm,d.st,count(*) as count_of_patients From Doctor d Left
JOIN record r on r.Doc_ID=d.Doc_ID GROUP BY 1,2,3;
```

10. Average bill per patient per state:

```
Select
        d.pri_spec
        ,p.state
        ,avg(room_charge) as doc_fees
From patient p
Left join record r on
        r.P_ID = p.P_ID
Left join doctor d on
        d.Doc_ID = r.Doc_ID
Left join billpayment bp on
        bp.bill_id = r.bill_id
Group by 1,2
;
```

11. Monitor Patient Admission Records:

```
Select p.admit_date, r.record_id
From Patient
Inner join record r on p.admit_date = r.record_id;
```

12. View for Doctors Revenue

```
Select  d.doctor_id , bp.doctor_fees
     count(billingdays) as doctor_billing_days,
        sum(r.billingdays*r.billingpayments) as doctor_hospital_revenue,
        count(p.patient_id) as patient_count
```

```
From doctor d
Inner join record r on r.doctor_id = d.doctor_id
Inner join billing_payment bp on bp.doctor_fees = d.doctor_id
Group by bp.doctor_fees, d.doctor_id;
```

13. View Doctors Record

```
Select d.doctor_id, r.record_id
 From Doctor
Inner join record r on d.doctor_id= r.record_id;
```

14. Highest number of patients in every state :

```
Select p.first_name, p.last_name ,p.state,
        count(*) as count_of_patients
From patient p
Left join record r on r.patient_id = p.patient_id
Group by 1,2,3;
```

15. Show billing date of patients

```
Select  p.f_name ,p.l_name ,count(billing_date) as bill_date
From patient p
Left join billing_payment bp on bp.P_ID = p.P_ID
Group by 1,2;
```

GitHub Link :
https://github.com/Krutikkanakia-neu/Hospital-Management