**ETL Project:**

**Extraction(E):**

I used 2 data set from two Insurance Company (Integra Managed Care and Visiting Nurse Services of New York).

All of my data is for the patient of the Insurance from 08/25/2019 to 08/25/2020.

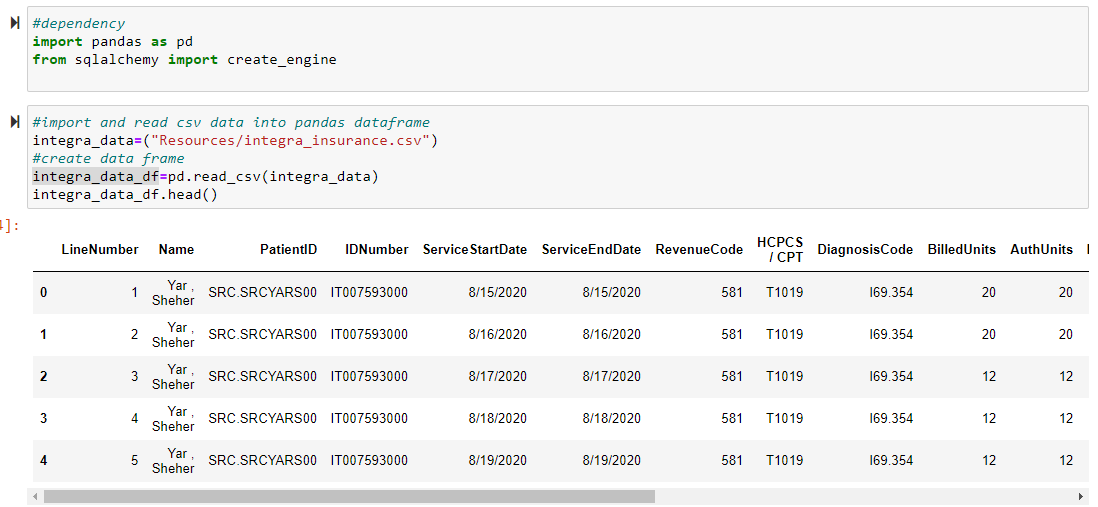
The source of dataset for the Company Sarah Care USA. NY, USA

* Integra\_insurance date from Integra Managed Care
* Vnsny\_insurance data from Visiting Aide Of New York

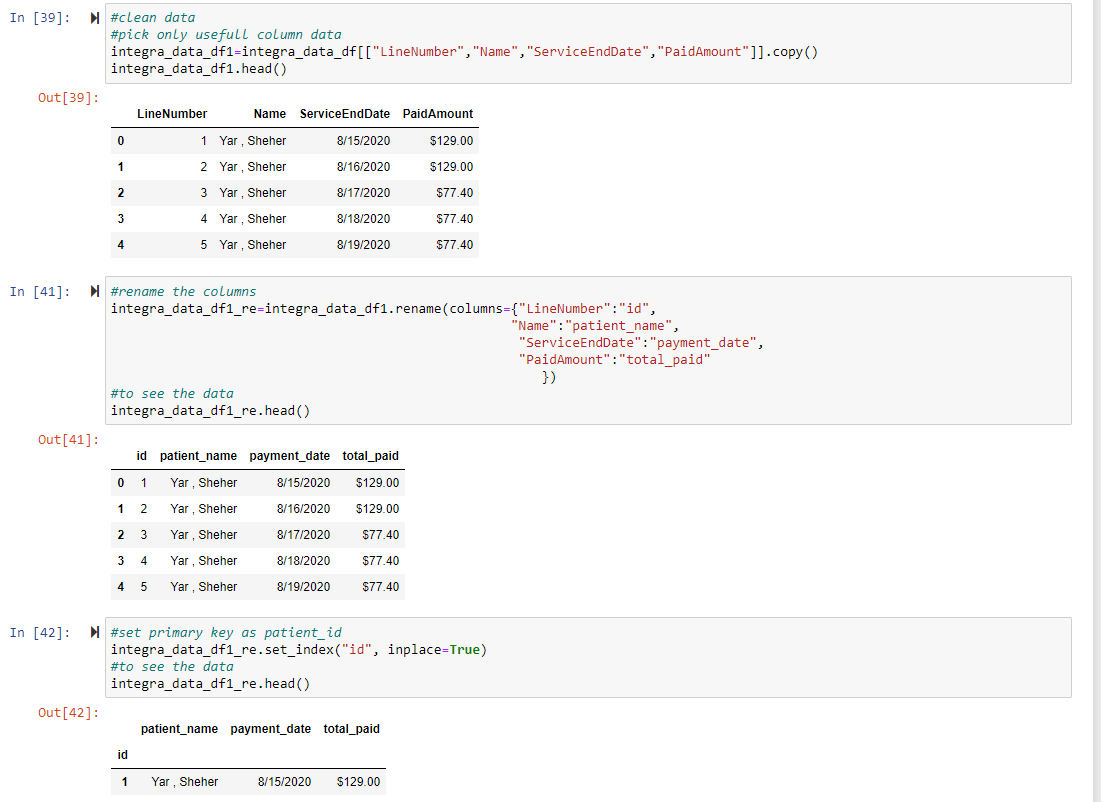
**Transformation(T):**

First, I import the dependency.

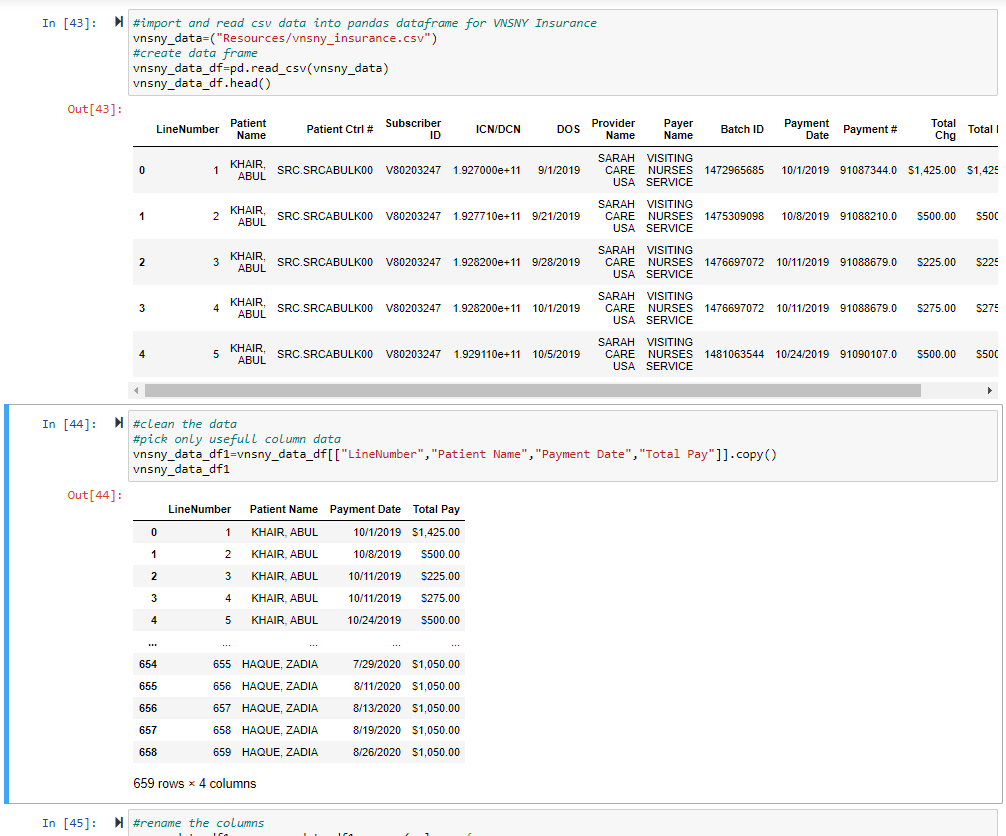
Second, I clean up the data set and drop which one are not relevant to the output object for Integra Insurance Data. (image)

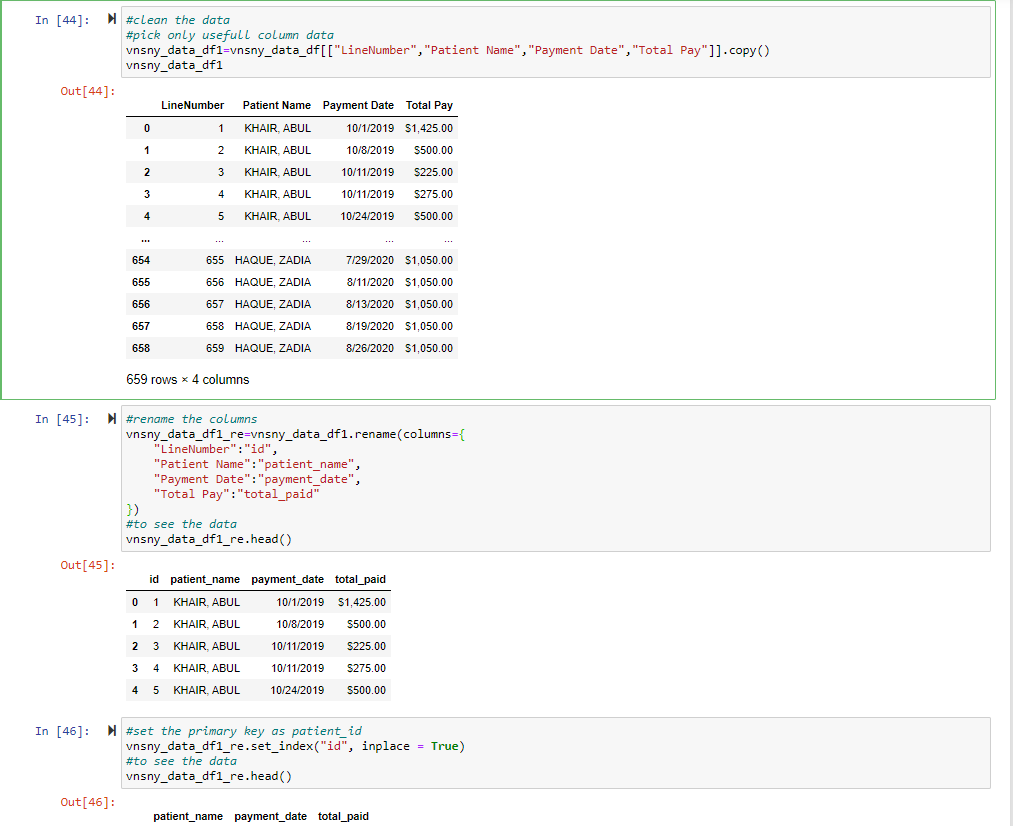


After cleaning the dataset, I drop the unnecessary data and rename the column. And make the column name as I can easily do query to fine the billed amount received from the Insurance company.



Same procedure (Reading and Cleaning) I did for the VNSNY insurance data:

­

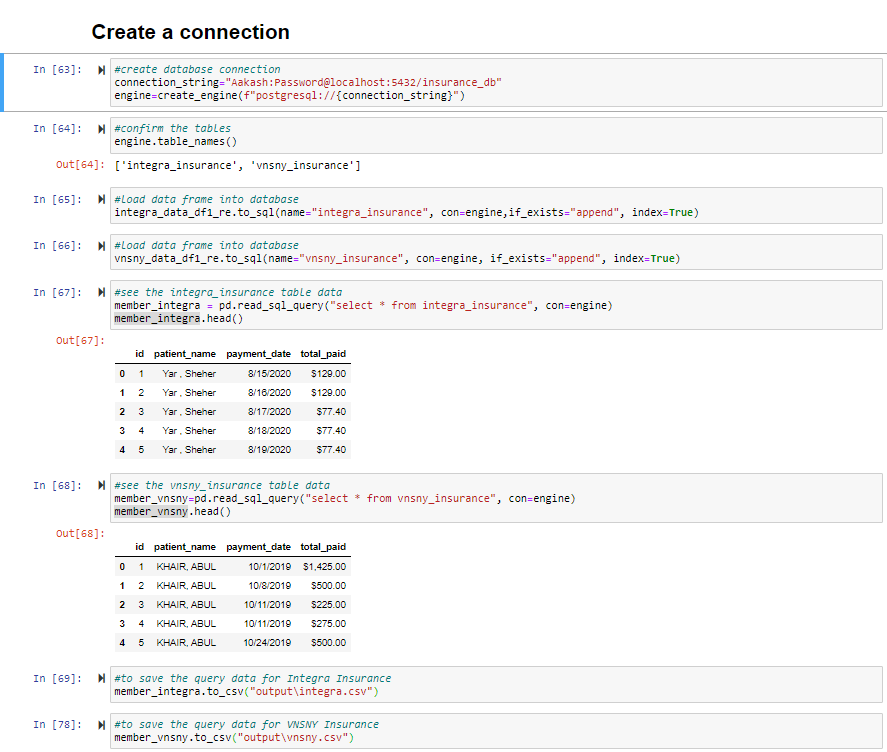
****

**Load(L)**

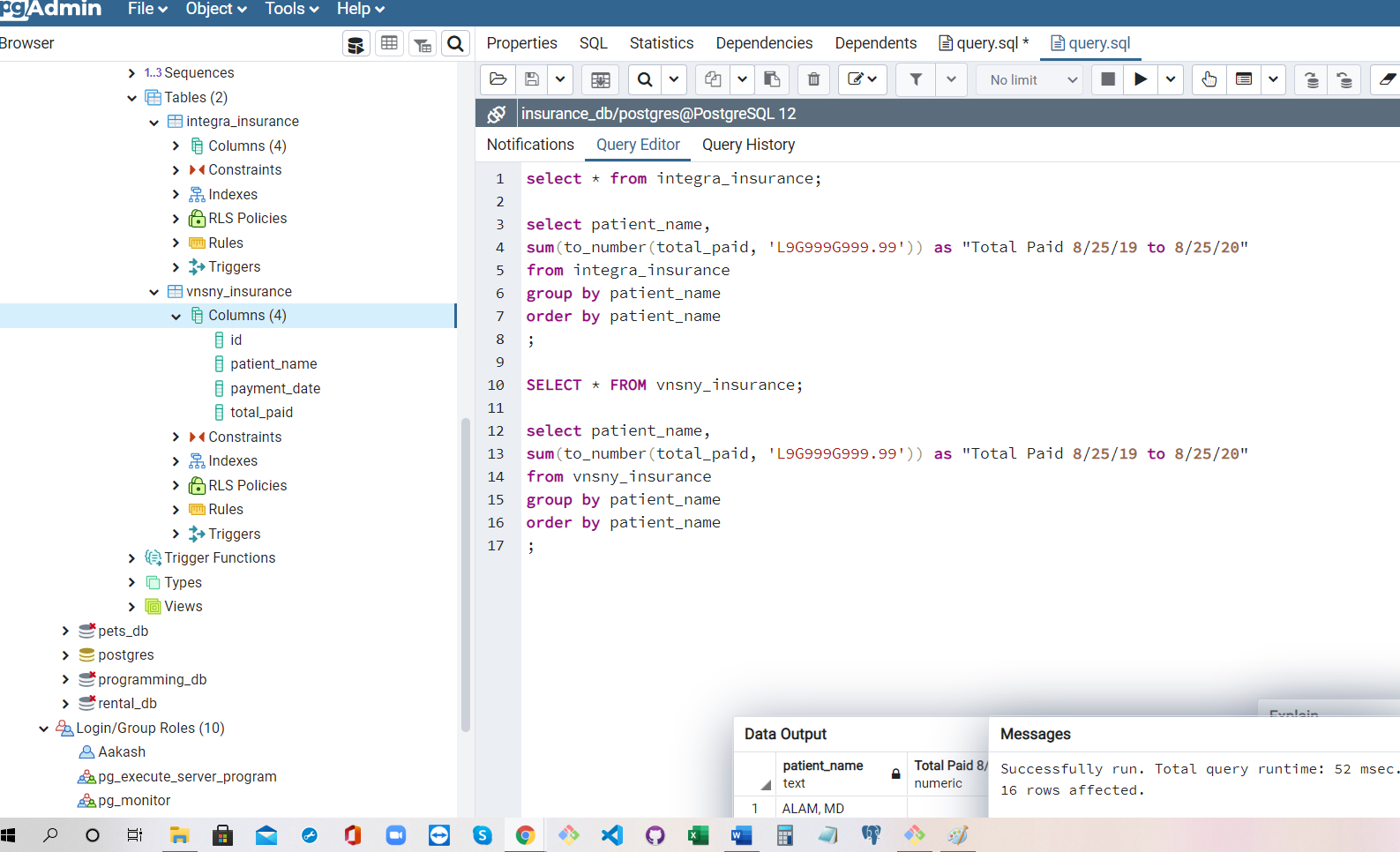
At the last step to load the data set, I created a database using MYSQL to match the column with Panda’s Data Frame.

And connected to the database using SQLAlchemy and loaded the result. Also did the query to find the total receivable amount from the insurance patient wise.

And save the output as csv file



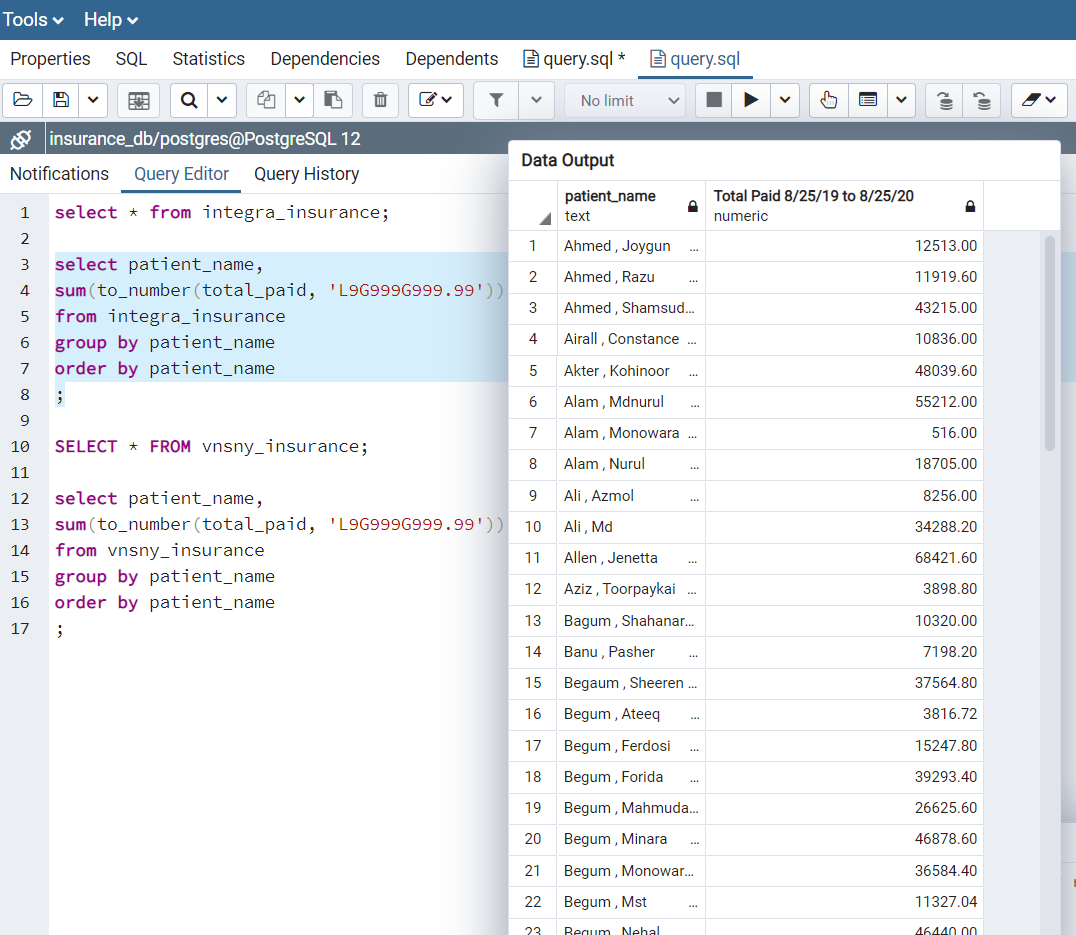
Sample queries:



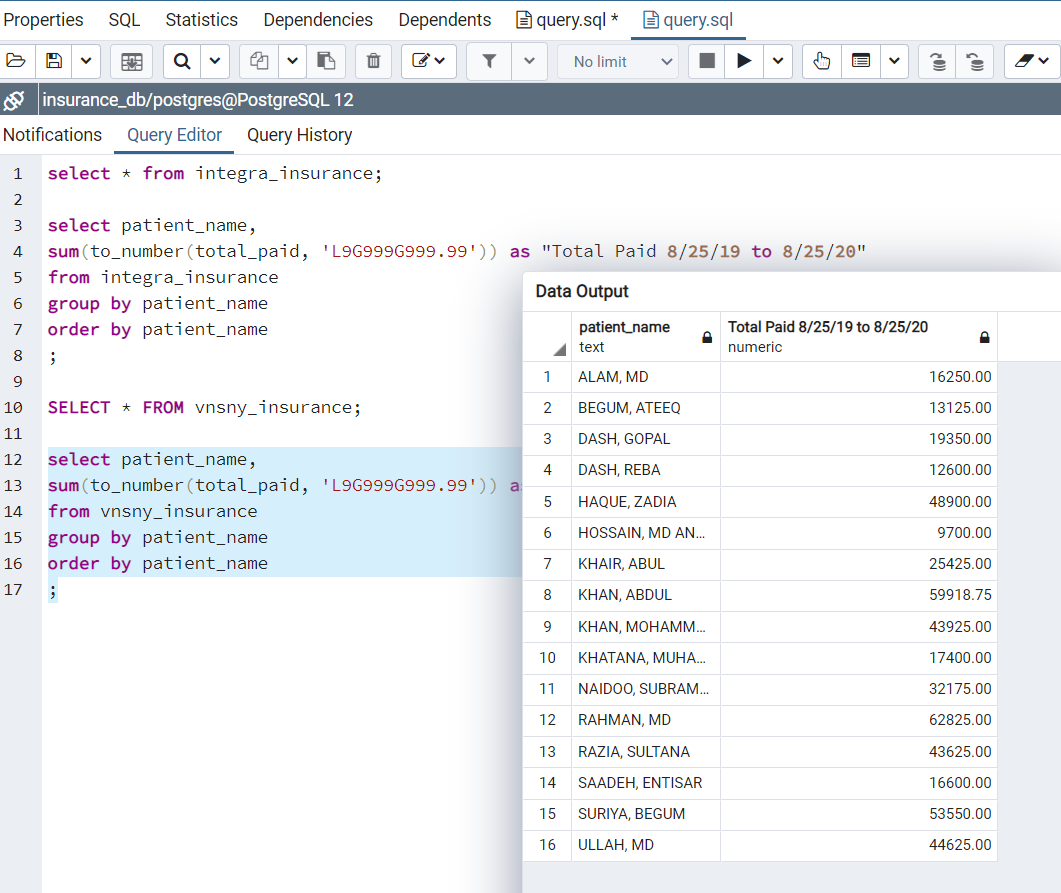
Summary:

Finally, I found the total amount received from the insurance individual Member wise from the both Insurance company.

**Integra Insurance:**



**VNSNY Insurance:**



And our management is very happy to see the receivable amount for individual Members.