DATA IMPORTING TO SQL

1. Imported the data as CSV file in SQL server to avoid data loss as the data set is too large to be imported as excel file
2. During the importing process I faced issue in few data types that is mismatched and I have rectified it imported successfully

DATA UNDERSTNADING

**Medicare\_charge\_inpatient**

**DRG Definition** - The **Diagnosis-Related Group (DRG)**, which categorizes hospital cases into groups for billing and reimbursement.

**Total discharge** – is the no. of inpatients discharges under this DRG

**Avg covered charges** – is the total bill amount of the patient

**Avg total payment** – is the payment done out of the bill amount this includes (payment by the patient, insurance claim, secondary insurance claim amount)

**Avg medicare payment** - is the part of the total payment that is provide by the insurance company for the service taken by the patient)

**Medicare\_charge\_outpatient**

**APC (Ambulatory Payment Classification)**

**Outpatient Services** - Represents how many times a particular outpatient procedure was performed.

**Average\_Estimated\_Submitted\_Charges -** The **average amount billed by hospitals** for that outpatient service. This is the **initial charge submitted** before Medicare adjustments.

**Average\_Total\_Payments -** The **actual amount received by the hospital** for that outpatient service. Includes **Medicare reimbursements + patient payments + secondary insurance payments**. Usually **much lower than the submitted charges** due to negotiated rates and Medicare payment rules.

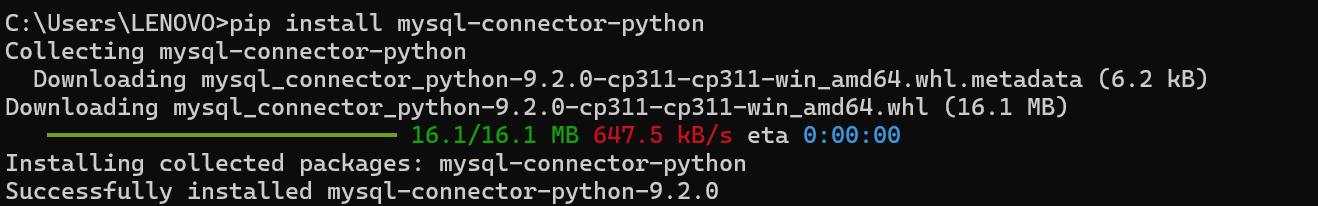
I have cleaned the data and Analysed the Data and made the data ready for further analysis,

I have performed the EDA steps to understand the data

Then I have imported the cleaned data from SQL Server to Power BI desktop, as the data is clean to perform the visualization task, I have directly loaded the data

I have found the KPI and wrote few DAX formulas to make the visualization interactive and understandable

**Connecting SQL server to Python IDE’s**

****

**PART-1**

**Part-1A:** Highest Cost Variation – procedure wise cost variation

Here I use std to find the cost variation for a particular procedure as std deviation tells us how much each provider ‘s cost deviates from the average cost for a procedure

I din use coefficient of variation because it is used to compare between different procedure wise deviation from the mean, as I need to find the variation for a particular procedure, I used STD

**Part-1B:** Highest cost claims by provider

explanation – For a particular procedure which provider is charging high

here I have used the covered charges to find which provider is claiming highest cost

**Part 1D:** Highest Number of Procedures and Largest Differences between Claims and Reimbursements

Explain: I have found the procedure that occurs more frequently and their max claim difference accordingly they have been arranged

MY POWER BI PUBLISHED LINK:

<https://app.powerbi.com/view?r=eyJrIjoiZTEzMTVjNDgtNjk1OC00NDg5LTk0NDgtMzliZWQyMzZkODBmIiwidCI6Ijg2NjY2OTljLTI2NTktNDE0MS05YzdhLTIxYzlkMzI0NTkxZCJ9>

## 🚀 Live Demo

Check out the live app: [Streamlit App on Render](<https://medicare-fraud-prediction-9.onrender.com>)

The **Z-score (standard score)** measures how far a data point is from the mean in terms of standard deviations. It helps in understanding how unusual or typical a value is within a dataset.