Explanation of Coding Sample

This coding sample is from an investigative project I completed during my master's program. The project focused on identifying home health agencies potentially over-billing Medicare, while distinguishing between fraud and legitimately higher costs due to sicker patients. The data on provider utilization is organized by codes reflecting patient Clinical, Functional, and Service Levels.

For this analysis, I used three datasets:

1. Provider Data (provider) – Contains general information on home health agencies.
2. Home Health Resource Group Data (provider_hhrg) – Includes financial and service utilization data for providers.
3. Case-Mix Weight Data (case_mix_weight) – Includes case-mix weights indicating patient complexity and expected care costs.

I loaded each dataset into Python, carefully exploring the columns and handling missing or non-numeric values. I cleaned the data where necessary and merged the three datasets by creating unique identifiers, sometimes splitting columns or using loops to ensure accurate alignment. From there, I conducted billing analysis by calculating new metrics from existing columns, generating visualizations, and highlighting patterns that could indicate over-billing. This project demonstrates both data wrangling and analytical skills applied to a real-world policy problem.