Specialty Drug Cost Analysis in Medicare Advantage Plans

Data Navigators

1. **Data Cleaning & Preprocessing**

* To ensure a robust analysis, we performed several key data cleaning steps. First, we standardized column names across both datasets, ensuring that Medicare and Medicaid drug data had consistent formatting (e.g., using Brnd\_Name for drug names). This step was essential for accurate merging and comparison.
* Next, we addressed missing values in spending columns by replacing NaN values with 0. This prevented data loss and ensured that all drugs were included in the analysis. Additionally, we filtered out drugs with zero spending across all years (2018-2022) to focus only on relevant specialty drugs.
* Handling duplicate entries was another crucial step. Initially, some drugs appeared multiple times due to different formulations or variations. To resolve this, we removed redundant records while ensuring that key data points were retained. In cases where the same drug had multiple formulations, we aggregated spending to get a more accurate representation of overall costs.

1. **Data Integration (Medicare vs. Medicaid)**

* Once the data was cleaned, we integrated Medicare and Medicaid specialty drug spending to perform a comparative analysis. The merging was done based on Brnd\_Name, ensuring that each drug's spending under both programs could be analyzed side by side.To facilitate year-over-year comparisons, we ensured that spending data was aligned across 2018-2022. This allowed us to identify cost trends over time.One of the most critical aspects of the integration was calculating cost differences between Medicare and Medicaid:

**Total spending difference per year:**

* (Medicare\_Spending - Medicaid\_Spending), which helped us determine where Medicare might be overpaying.

**Per-unit cost differences:**

* (Medicare\_Cost\_Per\_Unit - Medicaid\_Cost\_Per\_Unit), which provided insights into pricing inefficiencies.

**During this phase, we encountered two major challenges:**

* Initially empty merged dataset: This issue was resolved by handling missing values before merging.

**Duplicate entries in the final dataset:**

* We carefully removed redundant rows and, where applicable, aggregated spending across different formulations to avoid double counting.

1. **Ready for Analysis & Visualization**

With the cleaned and integrated dataset, we are now well-positioned for analysis.

The next steps include:

1. **Identifying the Top 10 Cost-Driving Drugs**

Analyzing drugs with the highest Medicare vs. Medicaid price gaps.

1. **Spending Trends (2018-2022)**

Visualizing how specialty drug costs have evolved over time.

1. **Policy Implications**

Investigating whether Medicare overpays compared to Medicaid, and exploring

the reasons behind these pricing differences.

1. **Optimization Strategies**

Recommending potential cost reduction strategies under Medicare Advantage.