

Navigating the Numbers: Insights into Insurance Claim Management

Guotao Ma

1. Executive Summary

This report delineates the classification of insurance claim entries, and analyzes the data for the Managed Service Organization (MSO) at North East Medical Services (NEMS). Central to this analysis is a payment summary table derived from NEMS's medical management database, which encompasses individual insurance claim summaries from various providers to HPSM from July 2024 to March 2025. Each entry incorporates key data fields, and the raw dataset has been omitted to ensure HIPAA compliance. The analysis highlights critical areas for strategic improvement, including a high prevalence of "Balance Due" claims, significant seasonal fluctuations, and data inconsistencies.

1.1 Key Findings

- **Financial Strain from Outstanding Balances:** A substantial financial constraint was identified, with over 68% of claims (279 out of 409 records) classified as "Balance Due". This indicates potential issues with patient financial responsibility, initial claim submission accuracy, or the efficiency of billing and collection processes.
- **Time-Series Dynamics and Surges:**
 - A significant surge in claim volume occurred in December 2024, with 216 claims, possibly due to a major billing cycle culmination or a year-end push to finalize collections.
 - Conversely, March 2025 showed an unprecedented cumulative outstanding balance of \$11,268, suggesting a potential bottleneck in payment resolution for that period.
 - Several months (July-August 2024 and January-March 2025) had a **TOTAL_PAID** value of zero while maintaining high balances, indicating a severe bottleneck where claims are processed but payments are not being recorded.
- **Data Peculiarities:** The **TRUE_BILL_AMT**, **TOTAL_PAID**, and **BALANCE** columns showed a limited number of discrete values, suggesting a tiered service model or predefined payment scenarios.

1.2 Recommendations

To address these findings and improve operational efficiency, the report recommends:

- **Streamlining Claim Approval:** Enhance workflow automation and communication with insurance providers to accelerate reimbursements.
- **Refining Data Management:** Implement robust data validation rules and standardize input procedures to ensure data integrity and accurate reporting.

2. Dataset Components

Here is a rephrased and reorganized list of the data fields:

- **MEMBID:** Member Identification Number (string, nominal)
- **YRMO:** Year and Month (string, ordinal)
- **DOS:** Date of Service (string, ordinal)
- **MODIF:** Confidential Internal Indicator (string)
- **TRUE_BILL_AMT:** Actual Billed Amount (float)
- **TOTAL_PAID:** Total Amount Paid (float)
- **BALANCE:** Remaining Balance (float)
- **FILE_REVIEW:** Status Label (string)
- **ADJCODES:** Adjustment Codes (string)
- **IS_RESUBMIT:** Resubmission Status (string, binary)

For the **FILE_REVIEW** column:

- **Paid:** The claim has been settled.
- **PMPM Paid:** The claim has been settled, and within the same year-month, modification, and member ID, at least one additional paid claim exists.
- **Balance Due:** The claim currently possesses an outstanding balance.
- **DUP:** The claim is a duplicate.
- **Needs manual review:** The claim exhibits irregular circumstances that necessitate a manual review.

3. Guidelines for Data Classification

For HIPAA compliance, classification must be conducted exclusively using Microsoft Excel or Power BI on an on-site desktop device. Alternative methods, such as Google Colab and Jupyter

Notebook, are strictly prohibited. The subsequent instructions were executed to update the status label within the **FILE_REVIEW** column.

Step 3.1: Filter raw data where **MODIF** is equal to U1 or U2.

Step 3.2: Locate all records matching the criteria from Step 3.1 within the same month/year (**YRMO**).

Step 3.3: From Step 3.2, identify records where **TRUE_BILL_AMT** equals **TOTAL_PAID** and **BALANCE** is 0; mark "Paid" in the **FILE_REVIEW** column.

Step 3.4: For "Paid" records of the same member within the same **YRMO**, insert "PMPM Paid" in the **FILE_REVIEW** column.

Step 3.5: From Step 3.2, if **TRUE_BILL_AMT** is greater than **TOTAL_PAID** and **BALANCE** is greater than 0, enter "Balance Due" in the **FILE_REVIEW** column.

Step 3.6: If **ADJCODES** contains a note and **BALANCE** is greater than 0, enter "Needs manual review" in the **FILE_REVIEW** column.

Step 3.7: If a record has no note and **BALANCE** is greater than 0, enter "Balance Due" in the **FILE_REVIEW** column.

Step 3.8: For "Balance Due" records of the same member within the same **YRMO**, if **IS_RESUBMIT** equals 1, mark "DUP" in the **FILE_REVIEW** column.

Step 3.9: Filter raw data where **MODIF** is equal to U8.

Step 3.10: Locate all records matching the criteria from Step 3.9 within the same month/year (**YRMO**).

Step 3.11: Locate records where **TRUE_BILL_AMT** equals **TOTAL_PAID** and **BALANCE** is 0; mark "Paid" in the **FILE_REVIEW** column.

Step 3.12: For "Paid" records of the same member within the same **YRMO**, insert "PMPM Paid" in the **FILE_REVIEW** column.

Step 3.13: From Step 3.10, if **TRUE_BILL_AMT** is greater than **TOTAL_PAID** and **BALANCE** is greater than 0, enter "Balance Due" in the **FILE_REVIEW** column.

Step 3.14: If **ADJCODES** contains a note and **BALANCE** is greater than 0, enter "Needs manual

review" in the **FILE_REVIEW** column.

Step 3.15: For "Balance Due" records of the same member within the same **YRMO**, if **IS_RESUBMIT** equals 1, mark "DUP" in the **FILE_REVIEW** column.

4. Exploratory Data Analysis (EDA)

4.1 Statistical summary of the numerical columns

Statistic	TRUE_BILL_AMT	TOTAL_PAID	BALANCE
Mean	92.30	18.06	74.24
Median	50	0	50
Standard Deviation	126.89	50.31	127.51
Minimum	50	0	0
Maximum	472	472	472
Sum	37752	7388	30364
Count	409	409	409

Table 1: Statistics of **TRUE_BILL_AMT**, **TOTAL_PAID**, and **BALANCE**

4.2 Distribution Analysis

The **TRUE_BILL_AMT** column in Microsoft Excel shows only two values: \$50.00 and \$472.00, forming a bimodal distribution. The accompanying histogram (Figure 1) clearly indicates that most claims are for \$50.00, suggesting it's a standard or default amount. The \$472.00 value represents a much smaller subset, possibly for more complex or less frequent billing scenarios.

A histogram (Figure 2) of the **TOTAL_PAID** column revealed three distinct claim amounts: \$0, \$50, and \$472. The majority of claims were for \$0, followed by 100 claims at \$50, and a small remainder at \$472. This indicates most claims receive no payment or a small standard amount, with few larger payouts.

The **BALANCE** column, like **TOTAL_PAID**, consistently shows values of \$0, \$50, and \$472. As Figure 3 illustrates, \$50 is the most common balance, followed by \$0, with \$472 being the least frequent, indicating a strong concentration around \$50.

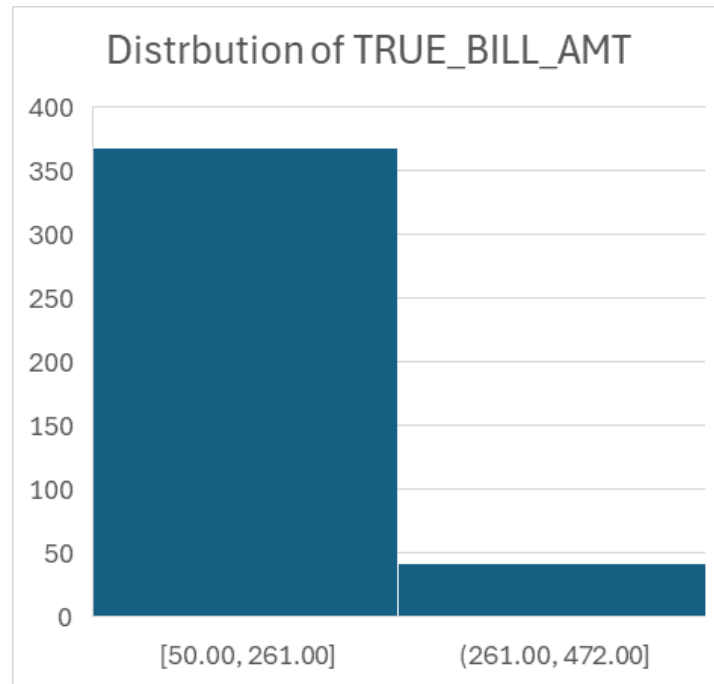


Figure 1: Distribution of TRUE_BILL_AMT

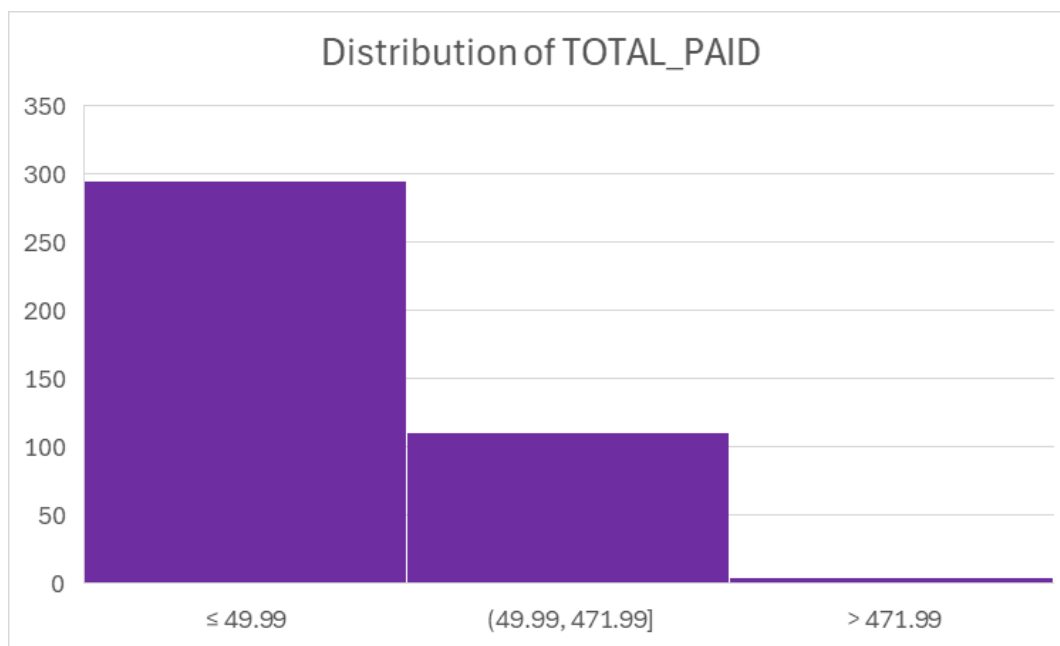


Figure 2: Distribution of TOTAL_PAID

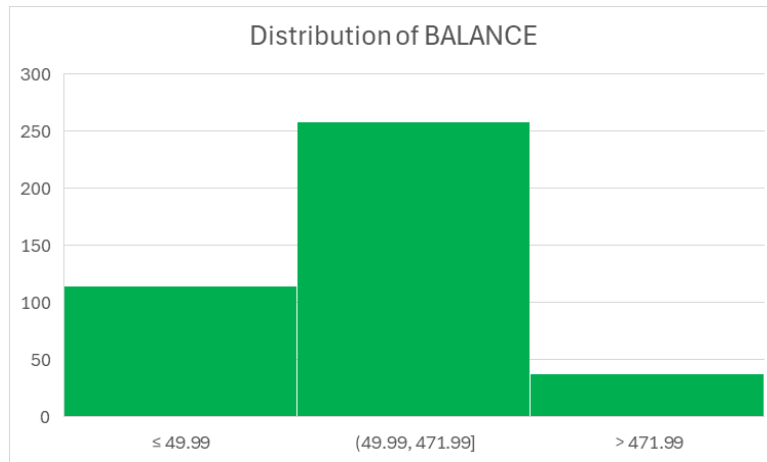


Figure 3: Distribution of **BALANCE**

4.3 Categorical Analysis

FILE_REVIEW	Count	Sum of TOTAL_PAID	Sum of TRUE_BILL_AMT
Balance Due	279	\$0.00	\$27454.00
DUP	7	\$0.00	\$2038.00
Needs manual review	9	\$0.00	\$872.00
Paid	61	\$4316.00	\$4316.00
PMPM Paid	53	\$3072.00	\$3072.00
Grand Total	409	\$7388.00	\$37752.00

Table 2: Pivot table of **FILE_REVIEW**

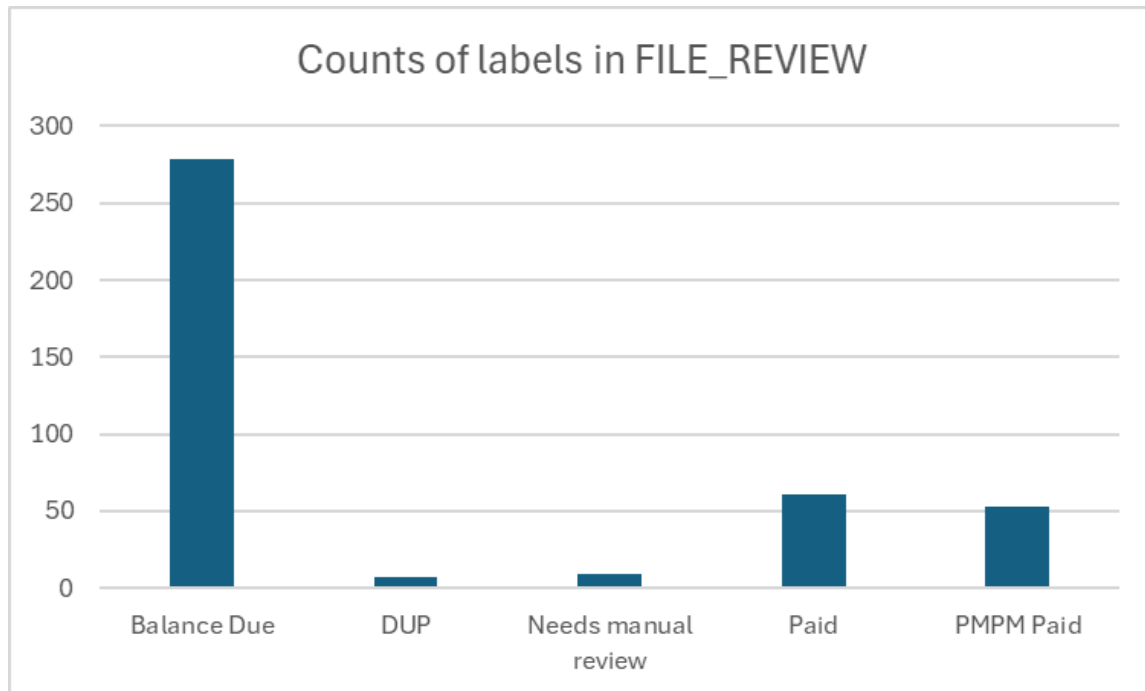


Figure 5: Counts of labels in **FILE_REVIEW**

Table 2 clearly illustrates the distribution of various labels within the dataset, highlighting the prevalence of certain financial statuses. The label "Balance Due" stands out as the most frequently occurring, suggesting that a significant portion of the data reflects outstanding payments. This high frequency could indicate a large number of claims where payment has not yet been received. The dominance of "Balance Due" warrants further investigation to understand the underlying reasons for its high frequency. This could include analyzing payment terms, billing cycles, or potential issues with payment processing.

Following "Balance Due," labels such as "Paid" and "PMPM Paid" exhibit the next highest frequencies. The "Paid" label signifies that the full amount owed has been settled, while "PMPM Paid" (Per Member Per Month Paid) likely refers to a specific payment model, perhaps in a subscription or insurance context, where recurring monthly payments have been successfully processed. The combined prevalence of these "paid" statuses, although lower than "Balance Due," still represents a substantial segment of the dataset, indicating a healthy rate of collection for some accounts.

4.4 Time-series Analysis

YRMO	Count	Sum of BALANCE	Sum of TOTAL_PAID
202407	17	1272.00	0.00
202408	21	3160.00	0.00
202409	5	50.00	200.00
202410	6	250.00	50.00
202411	5	100.00	994.00
202412	216	7610.00	6144.00
202501	2	522.00	0.00
202502	72	6132.00	0.00
202503	65	11268.00	0.00
Grand Total	409	30364.00	7388.00

Table 3: Pivot table of YRMO

Table 3 and Figure 6 present a comprehensive and dynamic perspective on NEMS's claim processing workflow, revealing crucial insights into operational efficiencies and financial health. A prominent observation is the notable fluctuation in claim volume throughout the period, culminating in a significant surge experienced in December 2024 and a dramatic decline in January 2025. This spike warrants further investigation to understand its underlying causes, whether attributable to seasonal trends, specific events, or changes in operational parameters.

A critical aspect of the analysis involves the identification of distinct periods marked by substantial outstanding balances. For instance, March 2025 exhibits a notably high outstanding balance, indicative of a backlog in claim resolution or payment collection during that month. Conversely, the data clearly illustrates months in which payments were actively recorded and processed; December 2024 again emerges as a period of significant financial activity, not only in terms of claim volume but also in successful payment processing. The correlation between the December 2024 claim surge and the active payment recording suggests an efficient processing capacity during that period, or potentially a concerted effort to resolve outstanding claims.

A significant diagnostic indicator within the figure is the observation of months presenting zero "TOTAL_PAID" values concurrently with elevated "BALANCE" figures. This specific constellation strongly implies that claims processed during these intervals are predominantly in an unresolved state, awaiting either authorization, subsequent processing, or remittance. The identification and analysis of these particular months can elucidate chokepoints in the claims lifecycle, such as

protracted adjudication periods, impediments in payer communications, or difficulties in patient invoicing and collection. Comprehending the underlying causes of these outstanding balances is imperative for enhancing NEMS's cash flow and overall fiscal stability. Further granular examination into the characteristics of these outstanding claims (e.g., aging, service classification, responsible payer) would yield invaluable insights for targeted interventions.

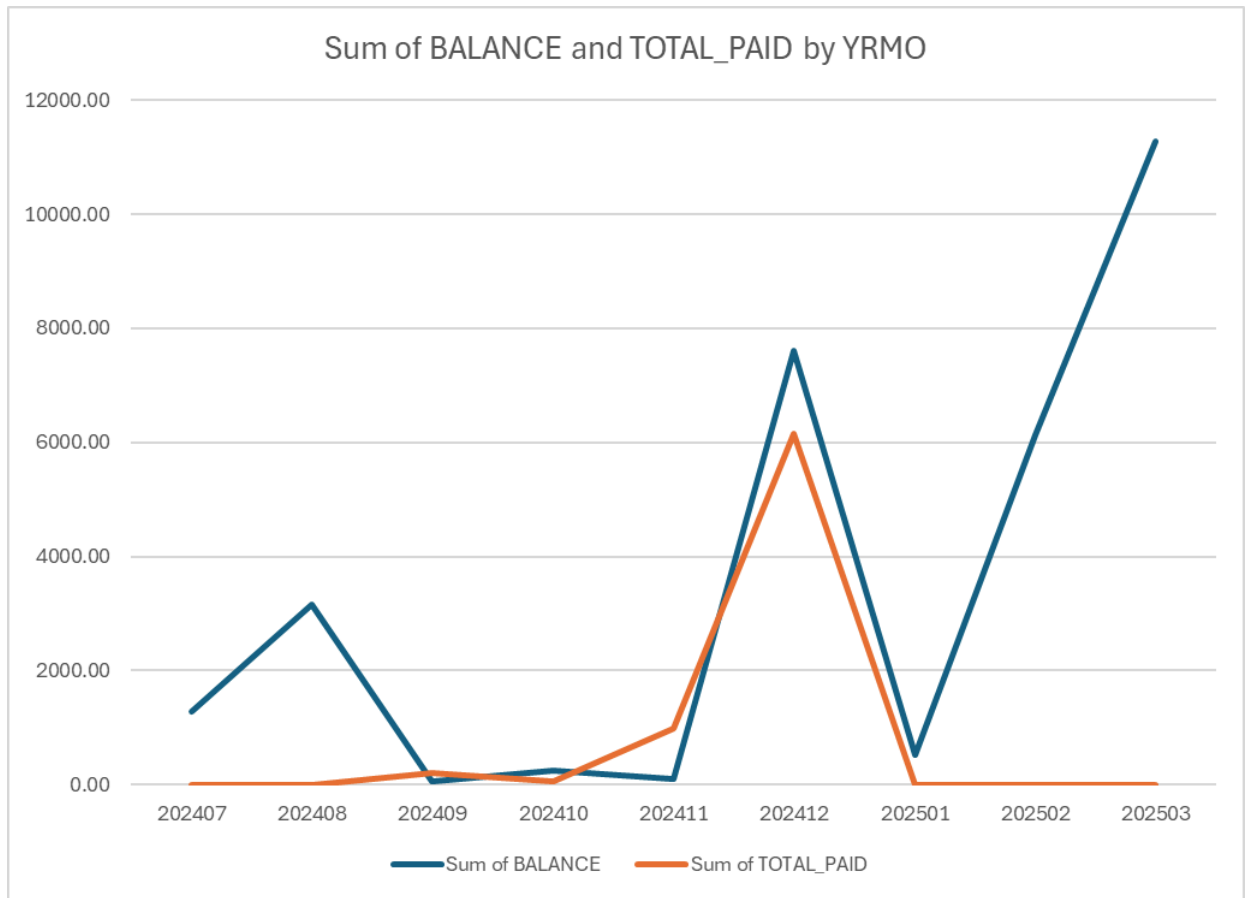


Figure 6: Monthly sum of **BALANCE** and **TOTAL_PAID**

5. Interpretation of the Findings

This section consolidates the principal findings from the Exploratory Data Analysis (EDA), converting statistical observations into practical interpretations concerning NEMS's insurance claim processing.

5.1 Dominance of Outstanding Balances and Its Implications

A substantial financial constraint has been identified, with more than 68% of claims (279 out of 409 records) classified as "Balance Due." This significant proportion highlights a critical issue concerning unresolved claims and uncollected revenue. The ramifications of this are extensive,

directly affecting the organization's financial performance and overall operational effectiveness. Addressing this pervasive challenge is not merely advantageous but imperative for sustained growth and stability. By resolving these outstanding balances, the organization can unlock substantial revenue, enhance cash flow, and reallocate resources currently dedicated to managing these overdue accounts. Furthermore, optimizing the claims process will improve operational efficiency, mitigate administrative burdens, and prevent future accumulations of "Balance Due" claims.

5.2 Understanding Paid Claims and Efficiency

While "Balance Due" claims are numerous, the successful processing of other claims, particularly where **TOTAL_PAID** matches **TRUE_BILL_AMT**, proves the billing system's effectiveness. This parity signifies accurate processing, full payment, and settled accounts, highlighting efficient administrative processes. Consistently achieving a zero balance on these claims demonstrates a robust system for managing financial transactions, minimizing disputes, and ensuring revenue flow. These successful closures validate current billing and collection procedures.

5.3 Time-Series Dynamics: Highlighting Key Periods

5.3.1 December 2024 Surge

In December 2024, the count of claims totaled 216, notably exceeding that of any other month. This particular period within the billing cycle appears to correspond with a significant peak in activity, potentially indicating a major billing cycle culmination or a substantial batch processing event. The simultaneous observation of a large volume being paid suggests a highly efficient and well-managed handling of this increased workload. This could also be indicative of a year-end push to finalize collections, a common practice in many organizations.

Considering that the majority of NEMS patients are from AAPI (Asian American and Pacific Islander) communities, this surge in billing or patient activity could be directly influenced by a confluence of Western and Asian holidays. For instance, increased patient visits or a concentrated effort in billing might occur around Christmas and the New Year, common holidays for many in Western cultures. Additionally, the Lunar New Year, a pivotal celebration for many Asian communities, often sees increased social activity and declining non-urgent healthcare appointments, which could contribute to a rise in patient interactions and subsequent billing in December. The interplay of these cultural calendars likely plays a significant role in the observed billing patterns, necessitating a nuanced understanding of patient behavior and operational demands during these specific times.

5.3.2 March 2025 High Balance

In March 2025, a significant trend emerged in our claims data. While the absolute number of claims processed stood at 65, a decrease compared to December 2024, the cumulative outstanding balance for these claims reached an unprecedented \$11,268. This striking discrepancy suggests a critical shift in the nature of claims being handled.

This observation points to two primary possibilities, both warranting immediate and thorough investigation. Firstly, it could indicate that the individual claims processed in March, though fewer in quantity, are significantly higher in their outstanding monetary value, which might suggest a higher concentration of complex, high-cost claims during this period, or perhaps a backlog of such claims finally being processed. Secondly, and perhaps more critically, the elevated total balance could signify a substantial bottleneck in our payment resolution process specifically for claims initiated or due for payment in March. This bottleneck could be due to various factors, including administrative delays, issues with payer adjudication, or problems within our internal accounts receivable workflows.

A focused investigation into claims from this particular period is essential. This inquiry should include a comprehensive review of individual claim values, the underlying reasons for their substantial outstanding balances, their processing durations, and the precise stages where impediments may be arising. Grasping the fundamental cause of this anomaly is paramount for enhancing our operational effectiveness, optimizing cash flow administration, and ultimately, bolstering our overall financial stability.

5.3.3 "Zero Paid" Months

A review of financial records from July-August 2024 and January-March 2025 revealed a consistent issue: **TOTAL_PAID** is zero while **BALANCE** is high, indicating a severe bottleneck in claims and payment processing. Claims are entering the system, but payments aren't being recorded. This points to potential delays in claim approval by payers (due to complexity, missing paperwork, or payer backlog), and issues with patient billing and collections (confusing bills, inadequate follow-up, third-party problems). Addressing these requires faster claim approvals (automation, complete paperwork, escalation), and streamlined patient billing/collections (proactive communication, smarter collection strategies). By concentrating on these aspects, financial risk will be mitigated, and operational efficiency will be enhanced.

5.3.4 Data Peculiarities and Their Meaning

The **TRUE_BILL_AMT** column in the dataset consistently shows only two values: \$50 and \$472, suggesting that NEMS uses a simplified, possibly tiered, service model, with these amounts representing either two standard service tiers or a base charge (\$50) consistently combined with additional fees to reach \$472 for specific services. Conversely, a more detailed

examination of the financial data reveals that both the **TOTAL_PAID** and **BALANCE** columns exhibit a limited range of variability, each containing only three distinct numerical values: \$0, \$50, and \$472, which suggests a highly specific and potentially restricted set of transaction outcomes or payment statuses within the dataset being analyzed. The presence of these discrete values, rather than a continuous spectrum, could indicate a system with predefined payment tiers, fixed charge amounts, or a small number of typical financial scenarios. Further investigation into the nature of these values and the context from which they arise would be necessary to fully understand their implications for the overall financial landscape represented by the data.

Although data entries labeled with "DUP" (duplicate) and "Needs manual review" constitute a relatively small subset of the overall dataset, a thorough examination of their underlying behaviors and patterns is crucial. Understanding the characteristics of these specific entries can provide valuable insights into potential data collection issues, system errors, or user input inconsistencies. This analysis will not only help in the immediate task of data cleaning and validation but also inform future strategies for improving data accuracy and efficiency, and identifying the root causes of duplicates and entries requiring manual review is essential for developing robust data management protocols and enhancing the reliability of the entire dataset.

6. Conclusion

The analysis reveals several critical findings, including the high prevalence of "Balance Due" claims, a distinct seasonal surge in claim volume, and significant inconsistencies observed within various column values of the dataset. The predominance of "Balance Due" claims suggests potential issues in patient financial responsibility, initial claim submission accuracy, or the efficiency of the billing and collection processes, which could be indicative of patients not fully understanding their financial obligations, insufficient communication regarding co-pays and deductibles, or delays in insurance reimbursements that lead to the patient being billed for the outstanding amount.

Furthermore, the observed seasonal surge in claim volume points to cyclical patterns in healthcare utilization, which could be influenced by factors such as flu seasons, holiday periods affecting elective procedures, or specific enrollment periods for insurance plans. Understanding these seasonal fluctuations is crucial for resource allocation, staffing, and proactive management of claim processing workflows.

The inconsistencies identified within the column values highlight fundamental data integrity issues. Such inconsistencies can lead to inaccurate reporting, hinder effective data analysis, and potentially result in incorrect billing or claim denials. These could stem from manual data entry errors, lack of standardized data input protocols, or integration challenges between different systems.

In light of these findings, it is imperative for NEMS and similar healthcare organizations to implement strategic improvements across several key areas. Firstly, streamlining claim approval processes is essential to reduce administrative burdens, accelerate reimbursements, and minimize the incidence of "Balance Due" claims. This could involve automating certain aspects of the approval workflow, enhancing pre-authorization procedures, and improving communication with insurance providers.

Secondly, enhancing patient billing practices is crucial for ensuring financial stability. This includes clear and transparent communication of patient responsibilities upfront, offering flexible payment options, and providing easily understandable billing statements. Educating patients on their insurance benefits and out-of-pocket costs can also contribute to reducing "Balance Due" issues.

Finally, refining data management practices is paramount for improving operational effectiveness and fiscal stability. This should encompass implementing robust data validation rules, standardizing data input procedures, and regularly auditing data for accuracy and completeness. Investing in advanced data analytics tools can also provide deeper insights into claim patterns and operational efficiencies, ultimately leading to more informed decision-making and sustainable growth for the organization.

7. References

Adhikari, A., DeNero, J., & Wagner, D. (2022). *Computational and Inferential Thinking: The Foundations of Data Science*. University of California, Berkeley. <https://inferentialthinking.com/>

Data Science Undergraduate Studies. (2025). *Principles and Techniques of Data Science*, 2025. University of California, Berkeley. <https://ds100.org/course-notes/>