Introduction

This report summarizes the analysis conducted on healthcare charge data, focusing on inconsistencies, calculated fields, feature enhancements, and insights derived from visualizations.

Data Inconsistencies

Observations:

Pharmacy Charges:

Encountered erroneous values and abnormally large figures in scientific format.

AR-DRG Values:

Inconsistencies noted with values such as DRG001, DRG002, and DRG003 combined into "DRGX" for easier handling.

AR-DRG Values:

Inconsistencies noted with values such as DRG001, DRG002, and DRG003 combined into "DRGX" for easier handling.

Calculated Fields

Calculated Age:

Derived from the date of birth. Discrepancies were identified, revealing 2,411 episodes where the age difference exceeded one year.

Length of Stay:

Calculated based on admission and separation dates.

Calendar Year of Admission:

Computed to facilitate further analysis.

Feature Enhancements

To enhance analytical capabilities, several features were created:

- Total Charge: Total charge for each episode.
- Revenue Per Day: Daily revenue generated per episode.
- ICU and CCU Visit Flags: Indicators for episodes involving ICU and CCU visits.
- Surgical Episode Flag: Indicated surgical charges from the theater charge column.
- Overnight Flag: Identified whether the episode was a same-day or overnight stay.

Data Analysis and Visualizations

Year-on-Year Comparison

Analyzed the top 10 revenue-generating AR-DRG categories, noting that Neonate, Orthopedic, GI, and Renal interventions accounted for significant revenue shares. This finding aligns with the understanding that these procedures typically require theater use and ICU visits, which are associated with higher costs.

Cumulative Year Analysis

Conducted a comparison of the top 10 Average Length of Stay (LOS) versus the top 10 Mode LOS by DRG.

Mode vs. Median: The mode was selected as a measure since it is less influenced by outliers. This approach provides a clearer picture of typical LOS, especially in skewed datasets where extreme values might distort the mean.

Conclusion

The analysis highlighted critical data inconsistencies and introduced calculated fields and features that enhance the analytical framework. The findings from both year-on-year comparisons and cumulative analyses provide valuable insights into revenue generation and length of stay trends, supporting informed decision-making in healthcare management.