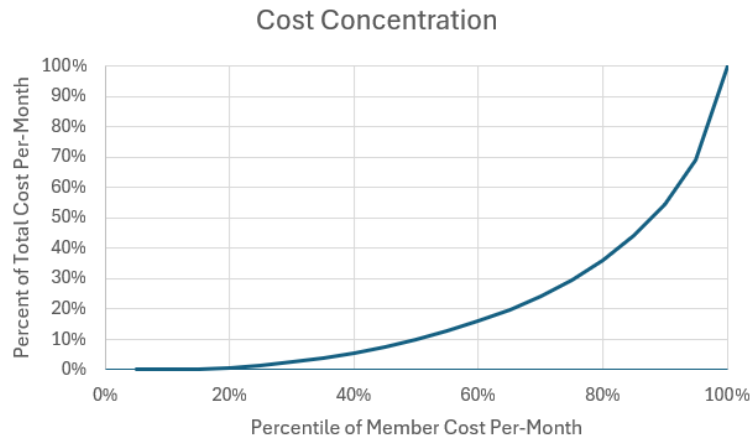
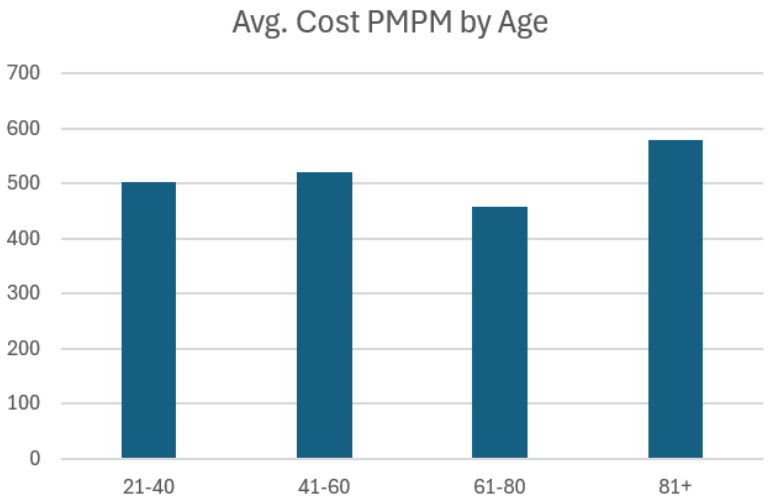
Executive Overview

This analysis applies risk segmentation methods to 114,612 Medicare beneficiaries from the 2009 CMS DE-SynPUF dataset to identify cost drivers and segment members by estimated risk. The model can successfully classify members, with members estimated in the top 10% of members capturing nearly 40% of total costs.

The experience analysis reveals a critical Medicare pattern: members falling within the 61-80-year age band exhibit lower average costs ($457 PMPM) compared to younger members ($504 - $521 PMPM), contradicting typical age-cost assumptions. This reflects the inflow of healthier members at age 65 eligibility, while younger members in the dataset consist exclusively of disabled populations. Additionally, analysis of the cost concentration shows the top 1% of members account for nearly 11% of the total cost, averaging $5,363 PMPM. This suggests that changes to the highest-risk members' plans could significantly impact total cost.





Objective and Methodology

Business Question

“Which member characteristics and utilization patterns drive cost variation in a Medicare population, and how concentrated is cost across risk segments?”

Data Engineering

The 2009 CMS DE-SynPUF dataset is a synthetic dataset created directly from real-world Medicare claims data, with values only changed for the purpose of anonymizing the individuals described in the dataset.
The data was stored in five tables: a beneficiary table and four types of claims (inpatient, outpatient, carrier, and Rx) tables. Relevant fields were identified from each claims table and then joined to the beneficiary table to create a single member-level experience file.

Modeling Approach

A Tweedie model was created to perform a concurrent analysis on the data. Decile analysis shows that the model captures 40% of cost in top 10% of members.

Model: Tweedie GLM (log-link, power=1.5)

Validation: 70/30 train-test split, decile lift analysis

Features:

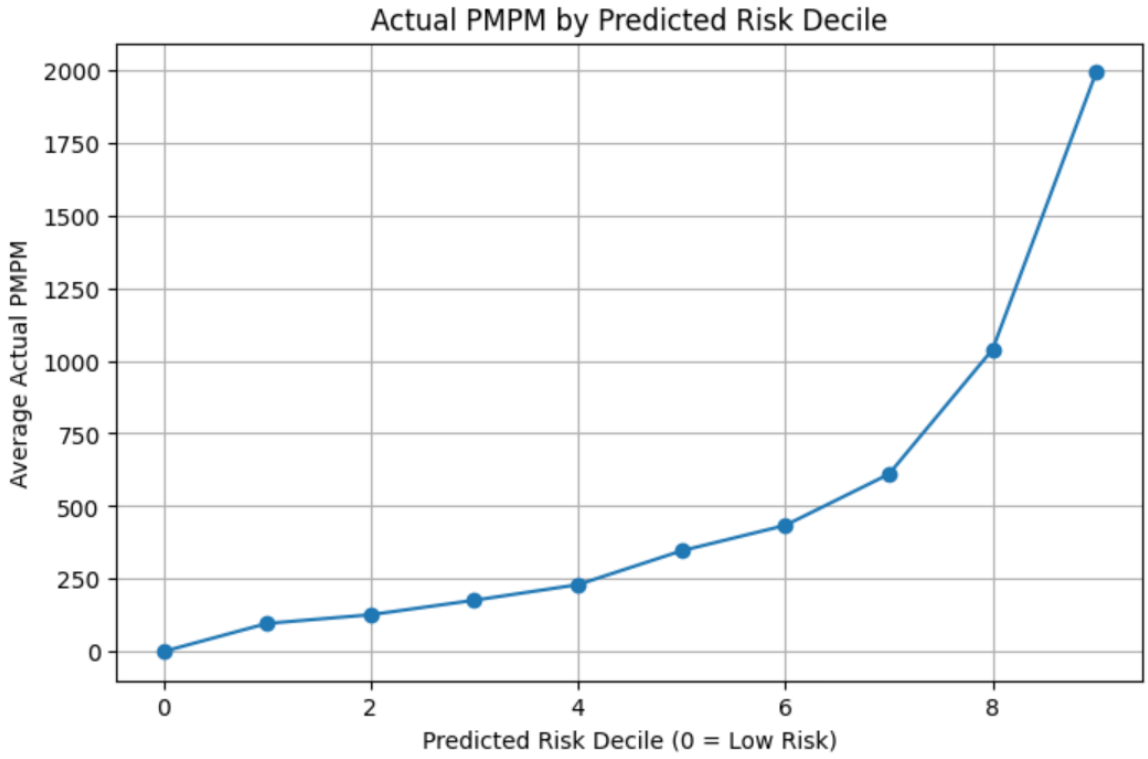
Feature	Relativity (Impact)
Any Rx Use	2.89
Any Inpatient Stay	2.43
Outpatient (6+ Visits)	1.91
Carrier Band	1.58
Outpatient (3-5 Visits)	1.55
Inpatient Count (Capped)	1.43
Outpatient (1-2 Visits)	1.39
Age 81+	1.31
Age 61-80	1.15

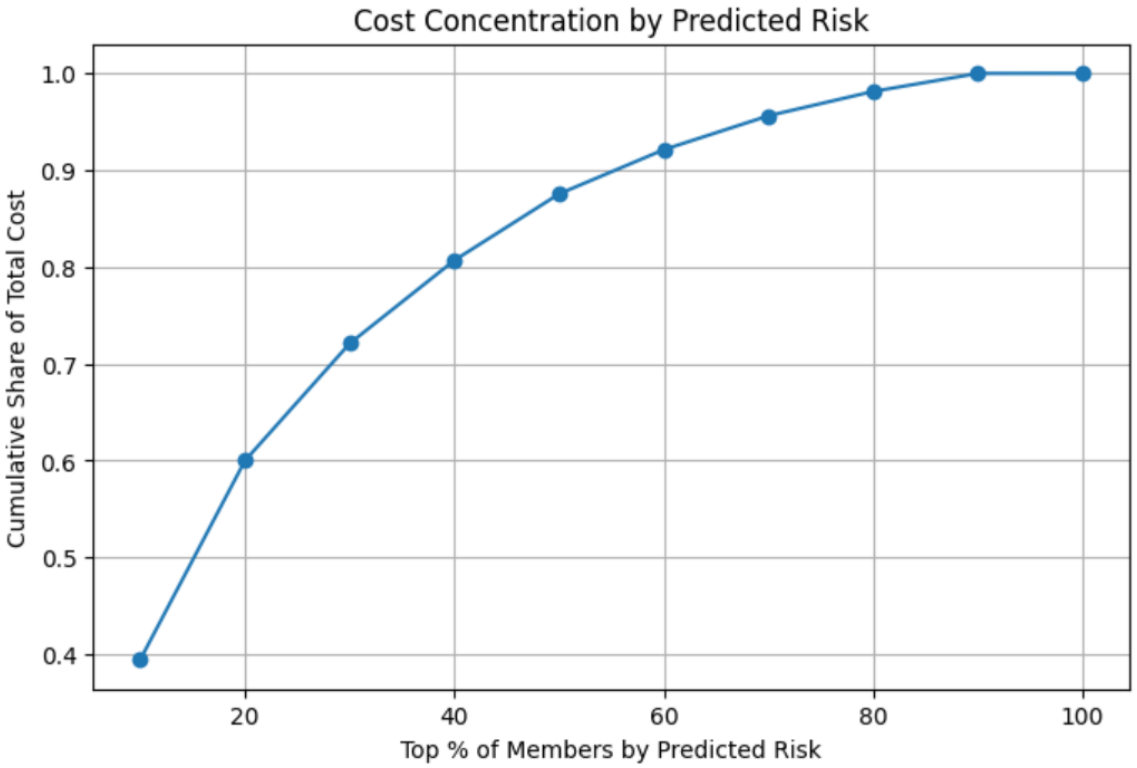
Age 41-60	1.10
Female	1.05
Rx and Age Interaction	0.99
Inpatient and Age 81+ Interaction	0.87

Key Findings

- Members falling within the 61-80-year age band exhibit lower average costs ($457 PMPM) compared to younger members ($504 - $521 PMPM)
- The top 1% of members account for nearly 11% of the total cost, averaging $5,363 PMPM
- Race/Ethnicity did not serve as a significant cost predictor
- Any Rx use increases estimated cost by a factor of 2.89x, while any inpatient admissions increase estimated cost by a factor of 2.43x

Shown below, 'Actual PMPM by Predicted Risk Decile' plots the actual average PMPM for the test data, with costs separated based on which cost decile the model estimated it to fall into. As expected, the higher deciles have higher actual average costs. 'Cost Concentration by Predicted Risk' plots the relationship between each decile and the total share of the cost. For example, the plot shows that the top 20% of members account for 60% of the cost.


Actual PMPM by Predicted Risk Decile
Average Actual PMPM
2000
1750
1500
1250
1000
750
500
250
0
0
2
4
6
8
Predicted Risk Decile (0 = Low Risk)


Cost Concentration by Predicted Risk
Cumulative Share of Total Cost
1.0
0.9
0.8
0.7
0.6
0.5
0.4
20
40
60
80
100
Top % of Members by Predicted Risk