# Final Results & Insights Summary

In this project, I built a supervised machine learning model using XGBoost regression to predict medical insurance charges based on individual attributes such as age, gender, BMI, number of children, smoking status, and region.

Although I initially evaluated the model using MAE and RMSE, I found these metrics alone were not enough to explain the real-world impact. RMSE became more relevant when thinking about extreme high-cost cases, such as cancer. This led me to shift my focus away from abstract error metrics and instead explore the actual drivers behind high insurance charges.

## Key Insights from Analysis

- **Smoking status is the most significant driver of cost.**
 Smokers consistently had the highest average charges across all groups.
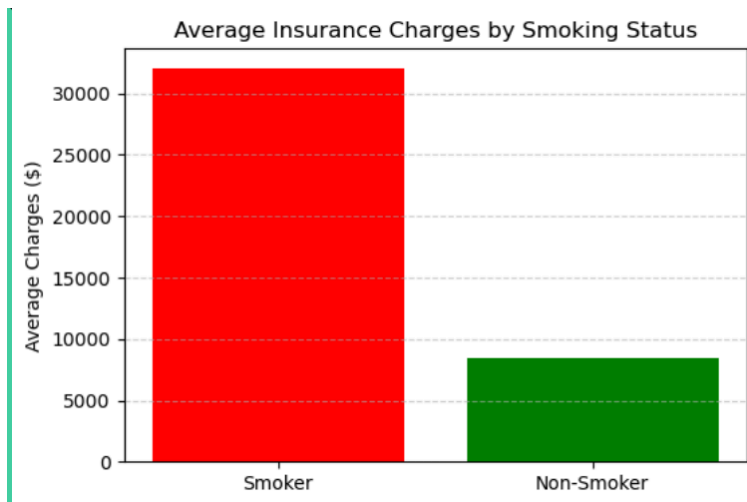
- **Obese smokers had the highest predicted costs.**
  All age, genders, and weight groups among non-smokers remained relatively stable in terms of charges, except for older individuals, who showed a slight increase. In contrast, smokers consistently had significantly higher charges across all age, gender, and weight categories, creating a wide gap even when compared to older non-smokers.

Average Insurance Charges by BMI Group and Smoking Status
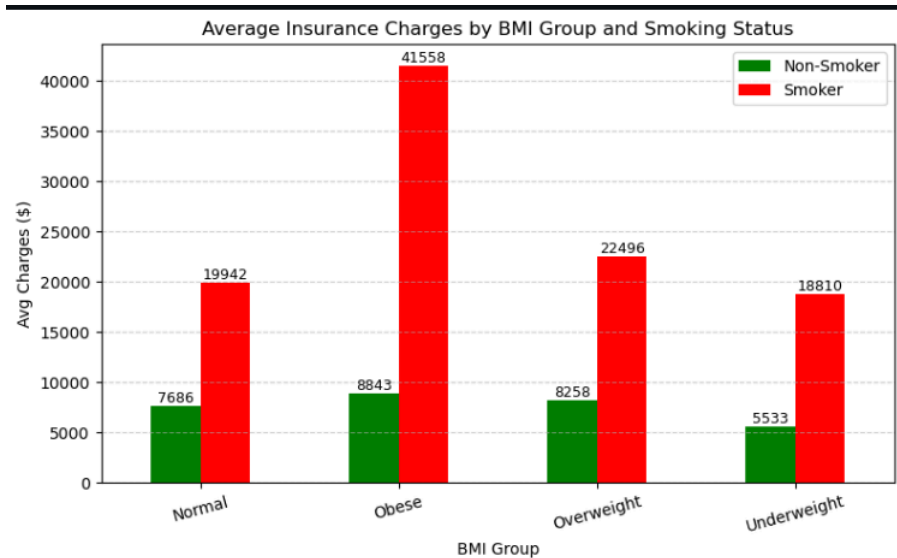
- **Non-smoker costs were generally stable.**
  Even among older or obese individuals, non-smokers had lower and more predictable average charges.

Average Insurance Charges by Age and Smoking Status
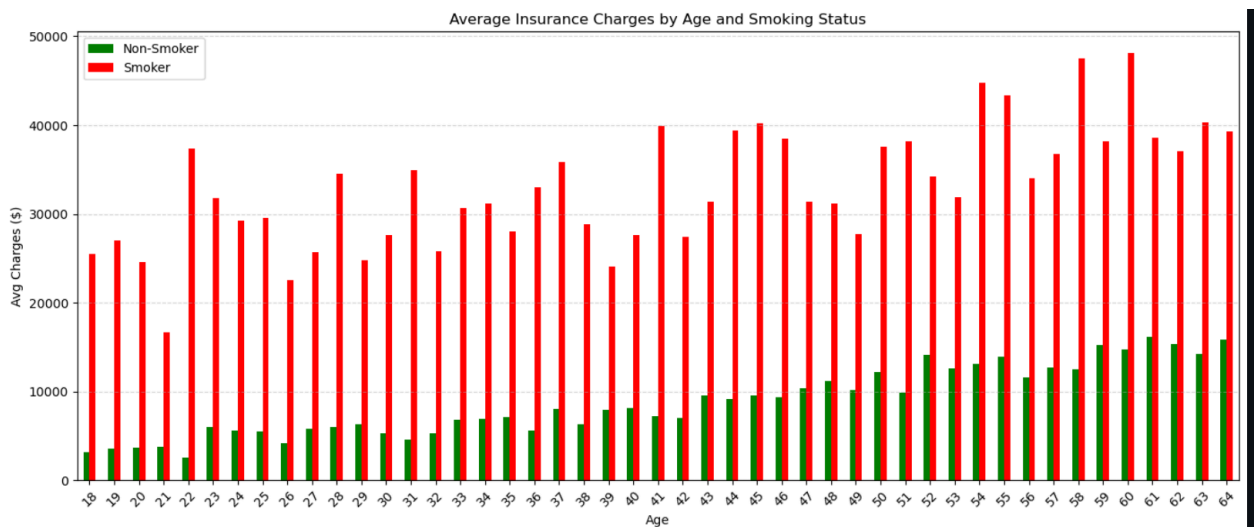
- **Gender, BMI, and age only became major cost factors when combined with smoking.**
  These factors showed limited impact alone, but had significant amplification when paired with smoker status.

Predicted Insurance Charges by Gender and Smoking Status
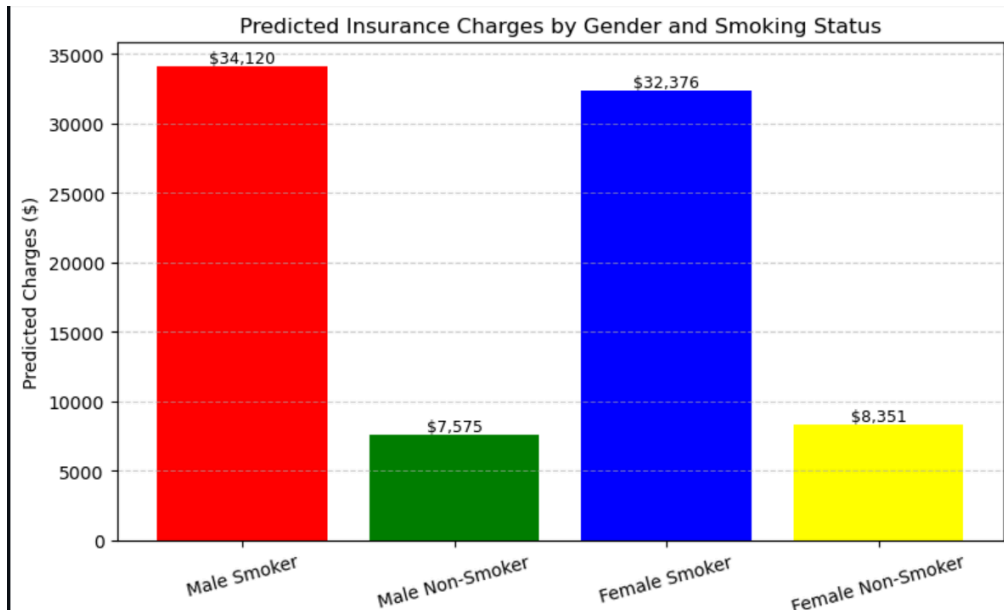
Recommendation

Rather than simply raising premiums, I recommend offering structured incentives:

- Discounts for verified non-smoking periods (30, 60, 90 days)
- Gift cards or wellness bonuses for continuous health improvements
- Tiered premium benefits based on participation in no-smoking programs
- Do not accept new applicants who are active smokers, as they present high-risk cost exposure
- Require a basic health screening or smoking verification test for new enrollees to confirm risk status

Model Interpretation

Initially, the AI model predicted slightly decreasing charges over time. I found this unrealistic. To address this, I introduced a 5% yearly adjustment to reflect the long-term rise in healthcare costs, which have consistently increased over the past 50 years. This sustained growth has been driven by rising medical technology costs, increased demand for care, and a growing population with chronic conditions. The adjustment also accounts for the expected increase in future customers, including the likelihood that some will be smokers and pose higher medical risk. This hybrid approach, combining AI prediction with human business insight, provided a more practical projection.

This project demonstrates how machine learning and human-centered analysis can work together to improve insurance pricing strategies, reduce risk exposure, and promote healthier behavior through incentive-driven programs.

Inflation-Adjusted Charges for 2024: $40489.11
Inflation-Adjusted Charges for 2025: $42513.56
Inflation-Adjusted Charges for 2026: $44639.24

Insurance Charges for Smoker with 5% Medical Inflation (2024–2026)