

Insurance

CUSTOMER RISK
PROFILING

Dataset : [Insurance Data](#)

TABLE OF CONTENT

- Objective
- Introduction
- Data Processing and Segmentation
- Visualization
- Conclusion & Recommendations

OBJECTIVES

- **To find out which customer factors (like age, income, health) affect claim risk the most.**
- **To check which data (like smoking, job, or loyalty) doesn't really affect claim risk much.**
- **To help make better pricing and discount plans based on risk and customer type.**

INTRODUCTION

Dataset Shape:

Total Columns : 21

Total Rows : 800000

Data Analytics



DATA PROCESSING AND SEGMENTATION

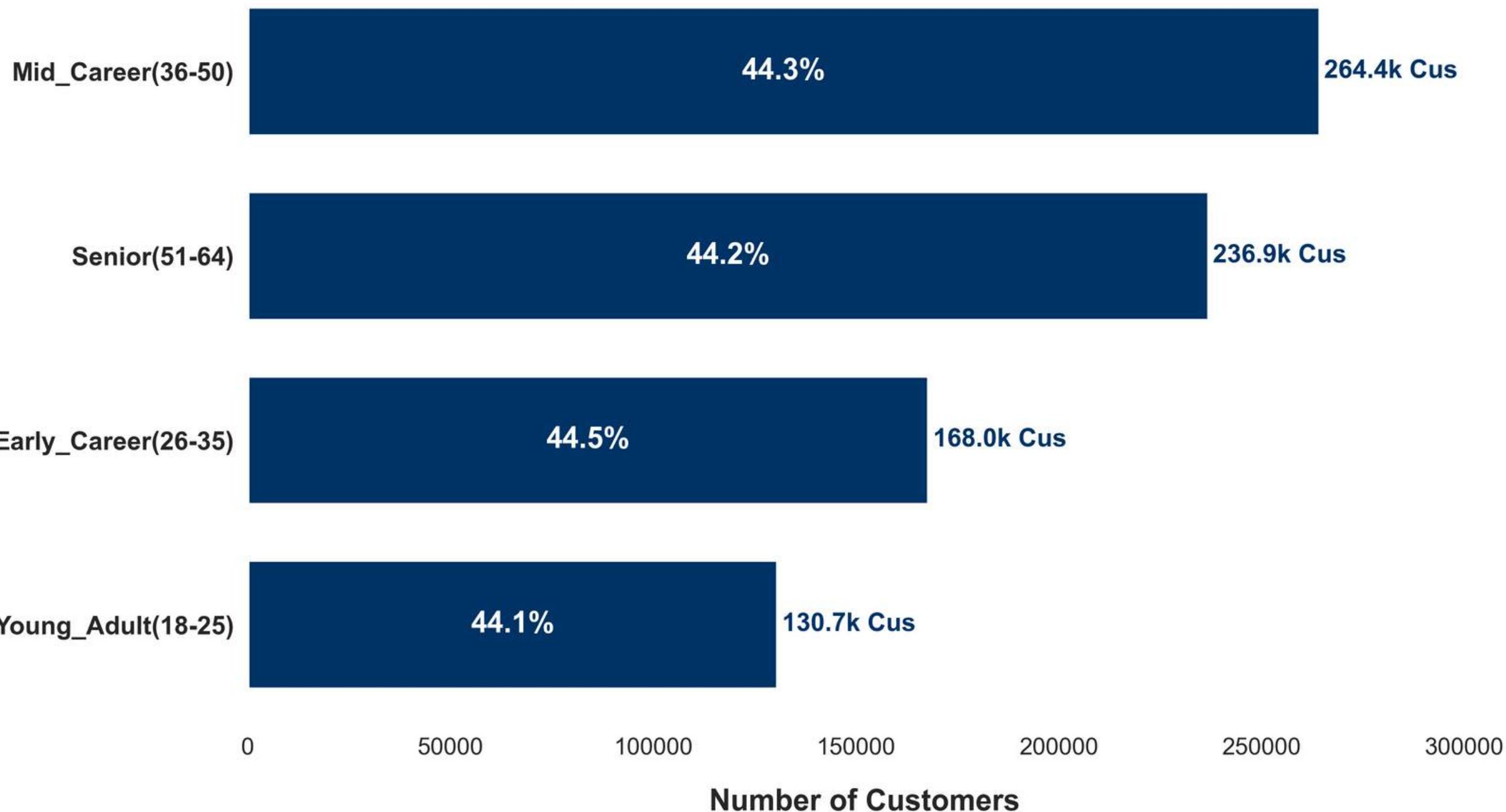
- Cleaned the data by handling missing values (like credit score and occupation) and checking for outliers (e.g., in previous claims).
- Standardized formats for fields like income, health scores, and dates.
- Segmented customers into groups for better comparison:
- Age Groups: Young Adult, Early Career, Mid Career, Senior
- Income Levels: Low, Lower-Middle, Middle, High
- Credit Scores: Poor, Fair, Good, Excellent
- Health Scores: Poor, Average, Good
- Other Segments: Smoking status, Occupation, Vehicle age, Customer tenure

Visualization

INSIGHTS:

- Mid-career (36–50) is the largest segment (264k customers) with 44.3% claim risk → highest volume impact.
- Young adults (130k) show similar risk (44.1%) but lower volume → secondary pricing focus.

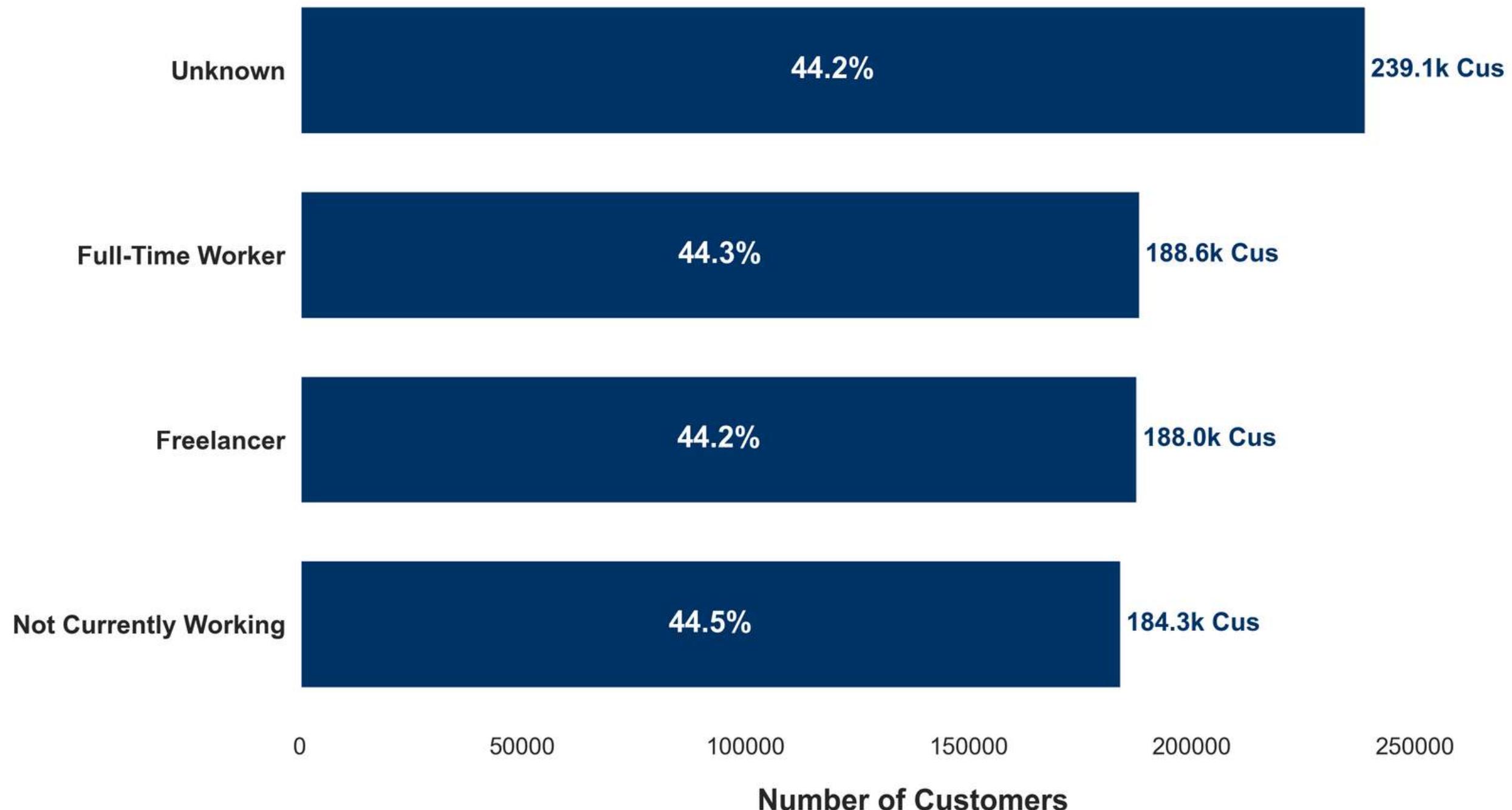
Customers & Claim Risk % by Age Group



INSIGHTS:

- Unknown occupation is the largest segment (239k) with 44.2% risk → data quality improvement required.
- Top 4 occupations (Unknown, Full-Time, Freelancer, Not Working) all within 0.3% risk → occupation not a strong driver.

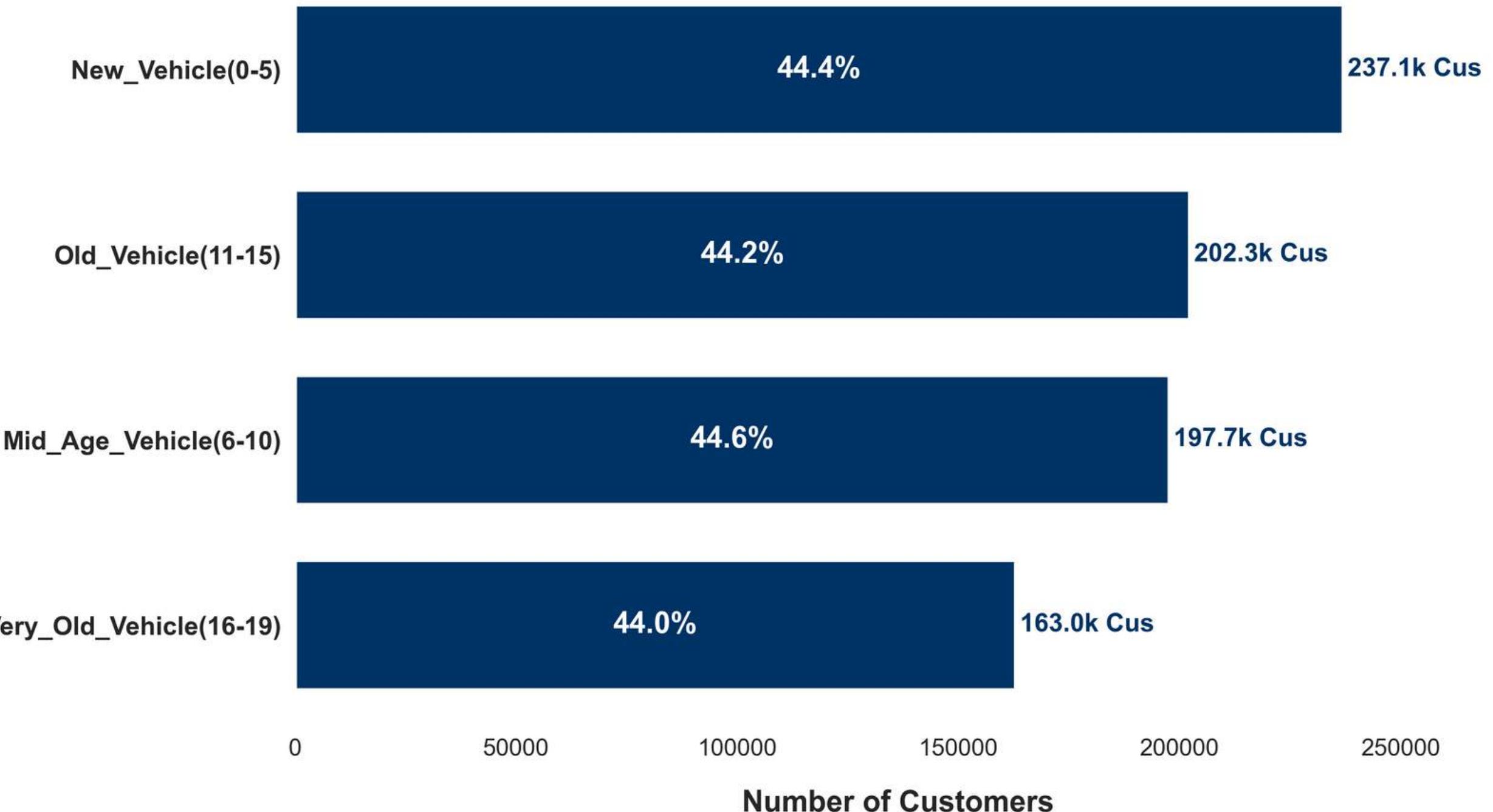
Customers & Claim Risk % by Occupation



INSIGHTS:

- New vehicles (0–5 yrs) largest group (237k) with 44.4% risk → biggest exposure segment.
- Risk difference across vehicle age groups <0.6% → vehicle age is not a meaningful rating factor.

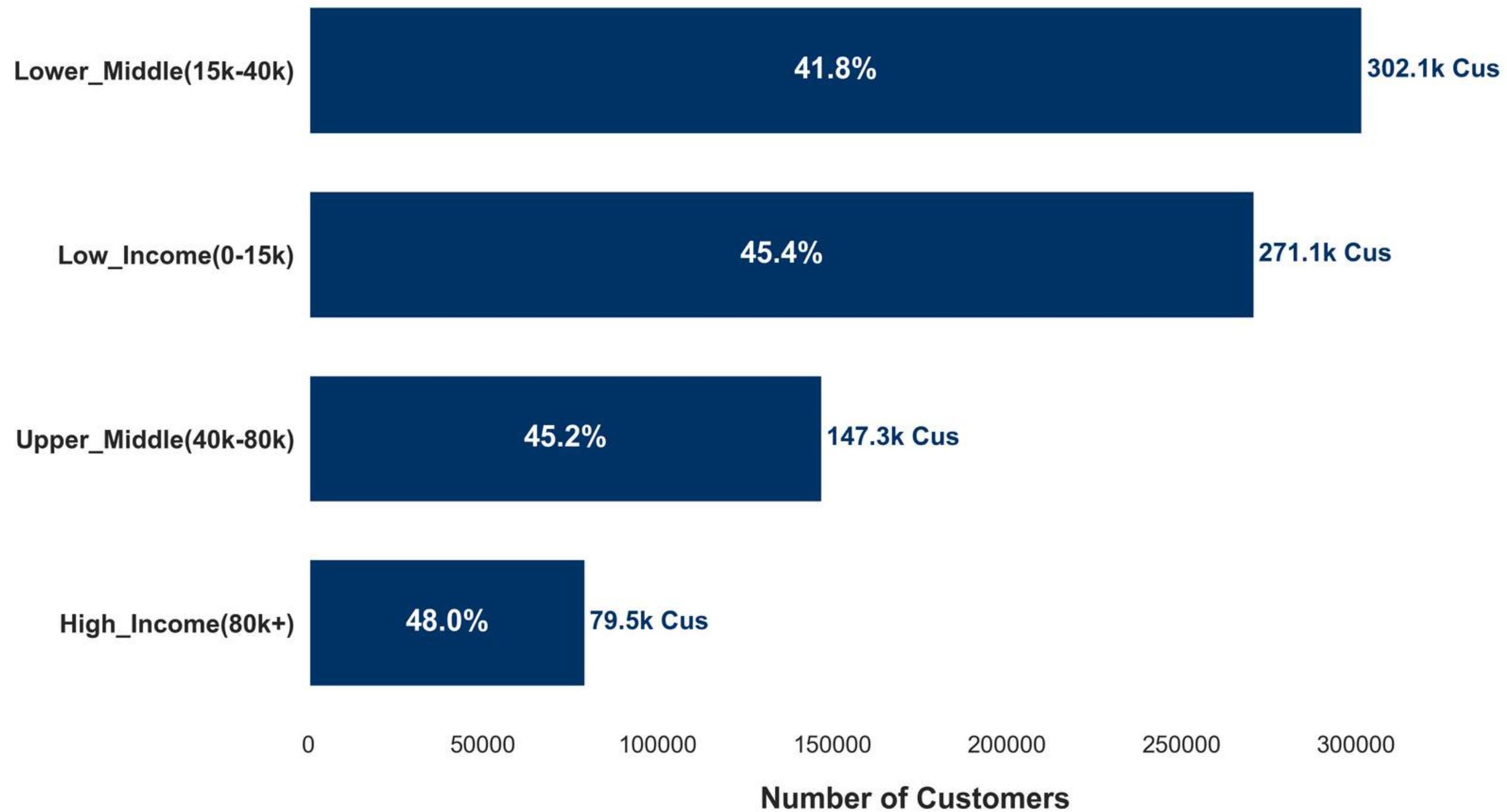
Customers & Claim Risk % by Vehicle Age Group



INSIGHTS:

- Lower-middle income (15k–40k) dominates with 302k customers but lowest risk (41.8%) → ideal for retention discounts.
- High-income (80k+) only 79k customers but highest risk (48.0%) → prime candidate for premium loading.

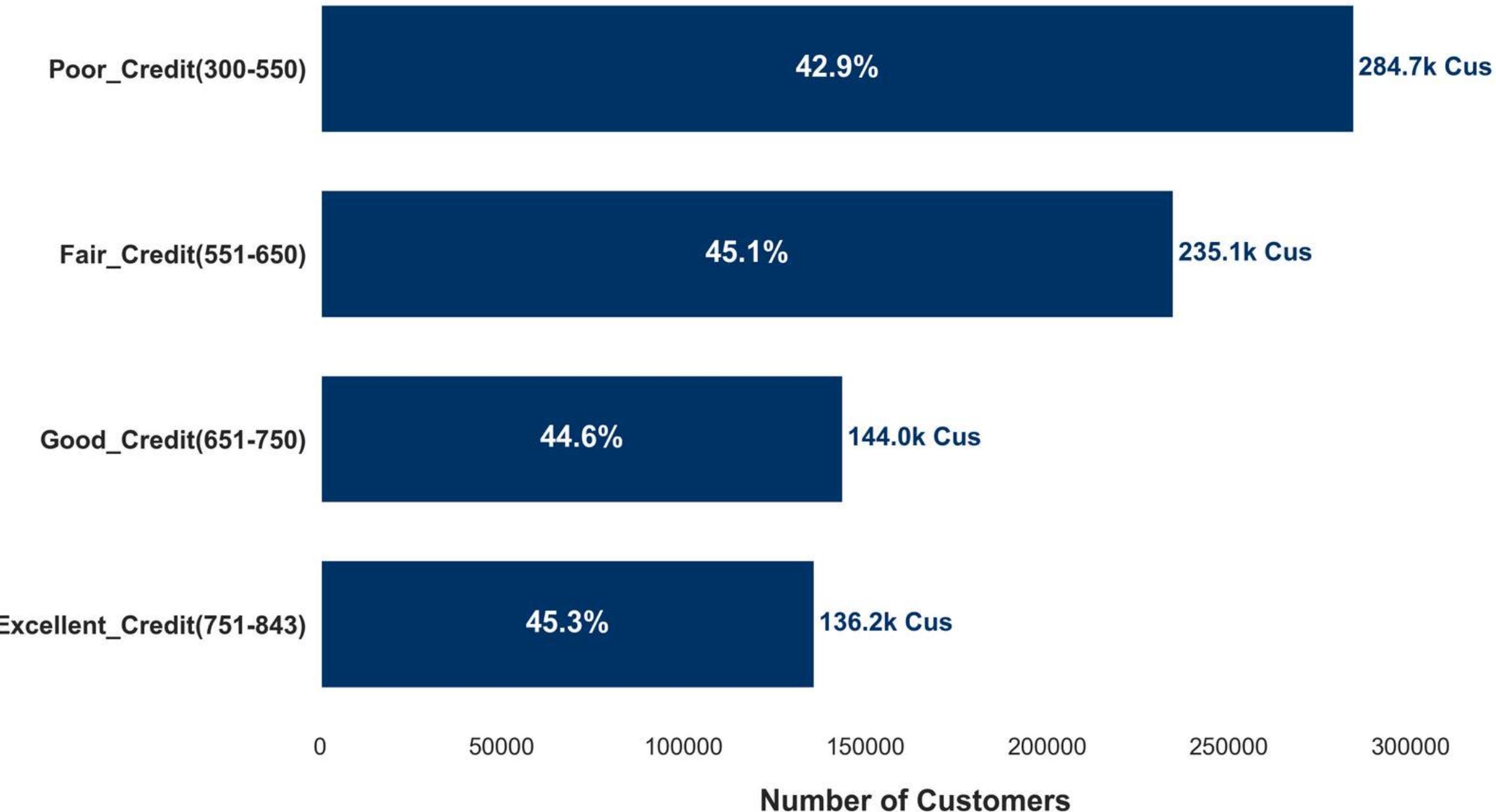
Customers & Claim Risk % by Income Level



INSIGHTS:

- Poor credit (300–550) is the biggest group (285k) with 42.9% risk → strong fraud/underwriting signal.
- Excellent credit customers (136k) still show 45.3% risk → credit score alone is not protective.

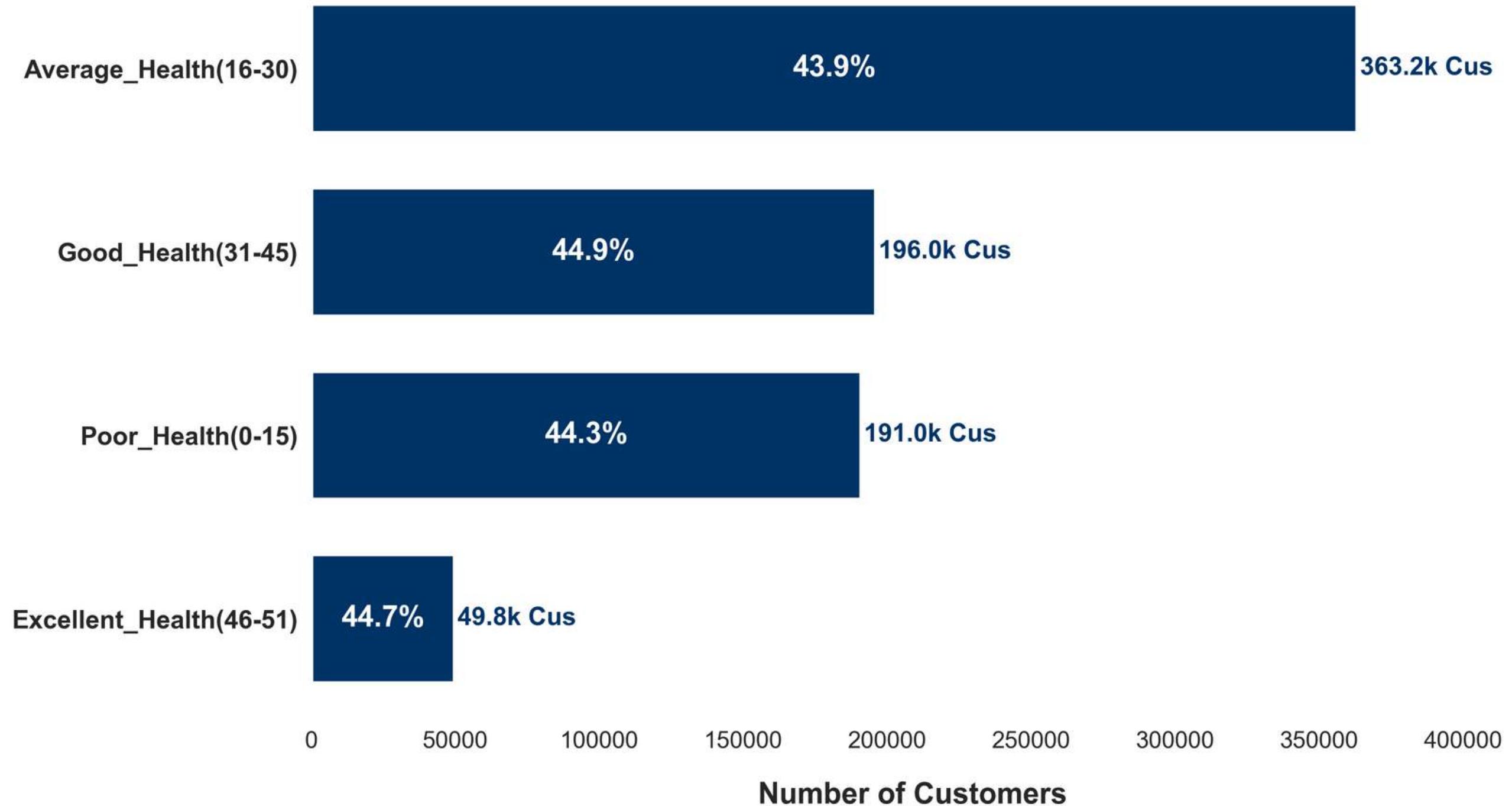
Customers & Claim Risk % by Credit Rating



INSIGHTS:

- Average health (16–30) dominates with 363k customers and lowest risk (43.9%) → largest safe segment.
- Good health (31–45) shows highest risk (44.9%) despite better score → counter-intuitive; needs investigation.

Customers & Claim Risk % by Health Category



CONCLUSION & RECOMMENDATIONS:

- Monitor Mid-Career Segment
- Ignore Smoking Status in Pricing
- Give Discounts to Low-Income Group
- No Need for Loyalty Discounts
- Increase Premiums for High-Income Customers
- Don't Rely on Credit Score Alone

Thank You

Welcome Any Question?

INSIGHTS:

*Lorem ipsum dolor sit amet,
consectetur the adipiscing elit.
Sed iaculis luctus varius.*

Model Building and Interpretation

MODEL

INTERPRETATION

RESULTS

Model	Accuracy	Precision	Recall	F1-Score
LightGBM (Best)	57.91%	56.27%	22.38%	32.02%
XGBoost	57.53%	54.83%	23.45%	nan
Random Forest	56.34%	51.52%	24.23%	32.96%
Logistic	55.82%	53.59%	1.94%	3.74%

INSIGHTS:

- The model correctly identified 76,794 safe customers (86% of safe customers correctly classified).
- Only 15,863 high-risk customers were detected (just 22% recall), meaning many risky customers were missed.
- 55,015 high-risk customers were labeled as “safe,” which in real life would cause the company a major claim loss.

Confusion Matrix – LightGBM (Best Model)

