MedicalMatch

One-time Shifts. Anytime. Anywhere.

Capstone Project: Healthcare Attrition Analysis

By: Tanya Seegmiller

As an incoming data scientist on MedicalMatch's product solutions team, my objective is to analyze healthcare attrition data to identify key trends and at-risk groups. By quantifying the financial impact of attrition on hospitals, clinics and nursing homes, I will provide critical insights that with empower our marketing team to develop targeted campaigns enhancing retention and operational efficiency in the healthcare sector.

MedicalMatch

Problem: Healthcare Facilities are understaffed and clinicians are burnt out. Resignations have increased 66% in the last 3 years and there is a projected shortfall of 3.2 Million workers by 2026

Ratio of labor costs to total facility expenses	Cost to facilities to hire and onboard new clinicians	Clinicians that report a "high level" of burnout.
62%	CNA: $16,000 RN: $56,000 Doctor: $500k- $1M	46%

Source: Deloitte Insights, Addressing healthcare's talent emergency 2022, Association of American Medical Colleges, Becker Hospital CFO Report.

MedicalMatch

Table of Contents

Data Collection and Cleaning

Data Source:
Excel Bi Analytics

Data Info:
Rows: 60,000
Columns: 25

Data Cleaning:
Research & customize healthcare specific feature ratios.

Exploratory Data Analysis

EDA Goals:

- Identify key trends and at-risk groups.
- Quantify the financial impact of attrition on the market to develop a targeted marketing strategy.

Model Building & Scoring Metrics

Model Type:
This is a binary classification problem. All classification models will be tested.

Model Performance Criteria:

- False Negatives will contribute to high attrition costs.
- Recall will be our performance metric

Conclusions and Recommendation

No spoilers folks!
You'll have to wait until the end of the show to find out

🤭

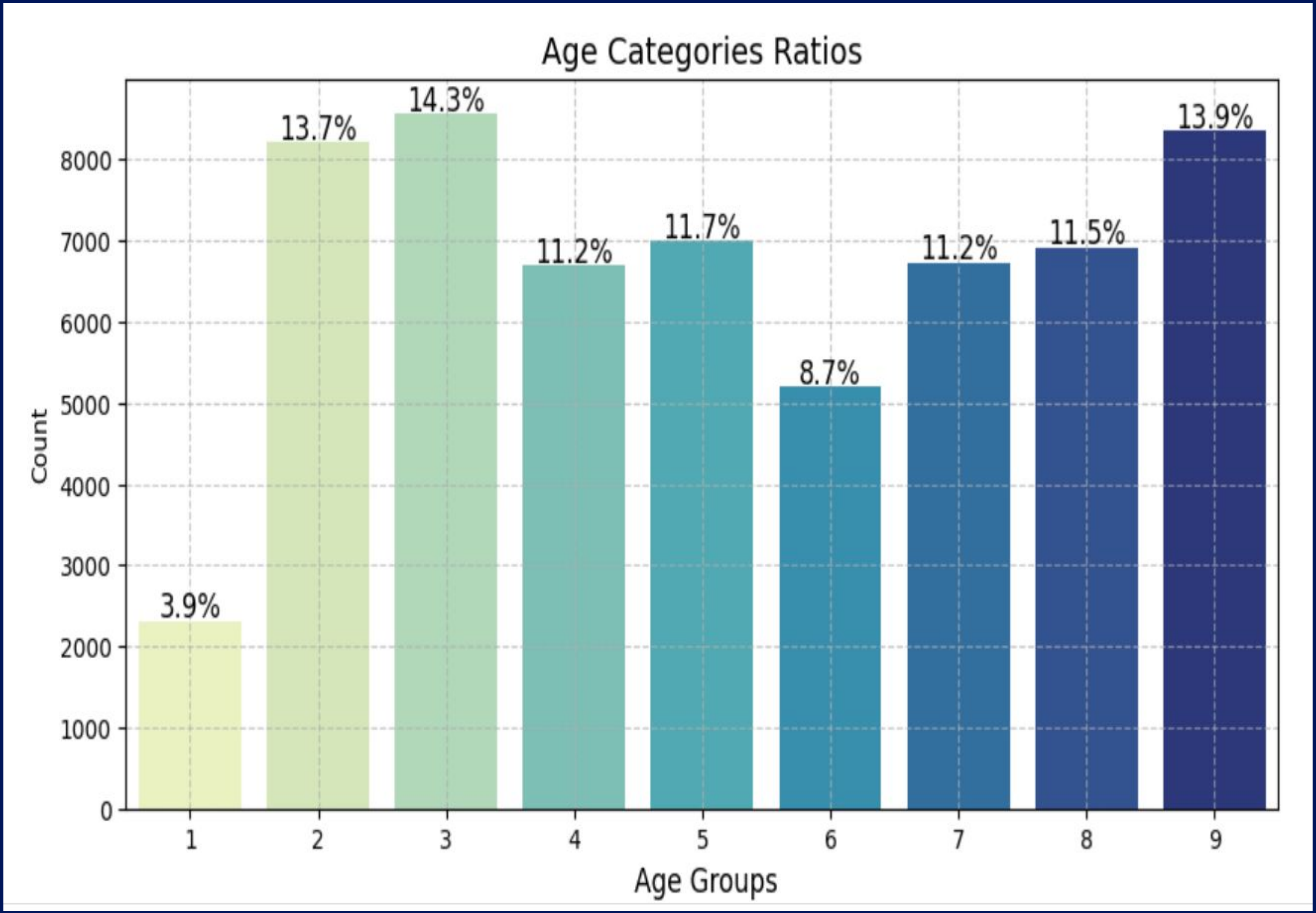
Age Categories:

1: >30 yrs
2: 30-34 yrs
3: 35-39 yrs
4: 40-44 yrs
5: 45-49 yrs
6: 50-54 yrs
7: 55-59 yrs
8: 60-64 yrs
9: 65+ yrs

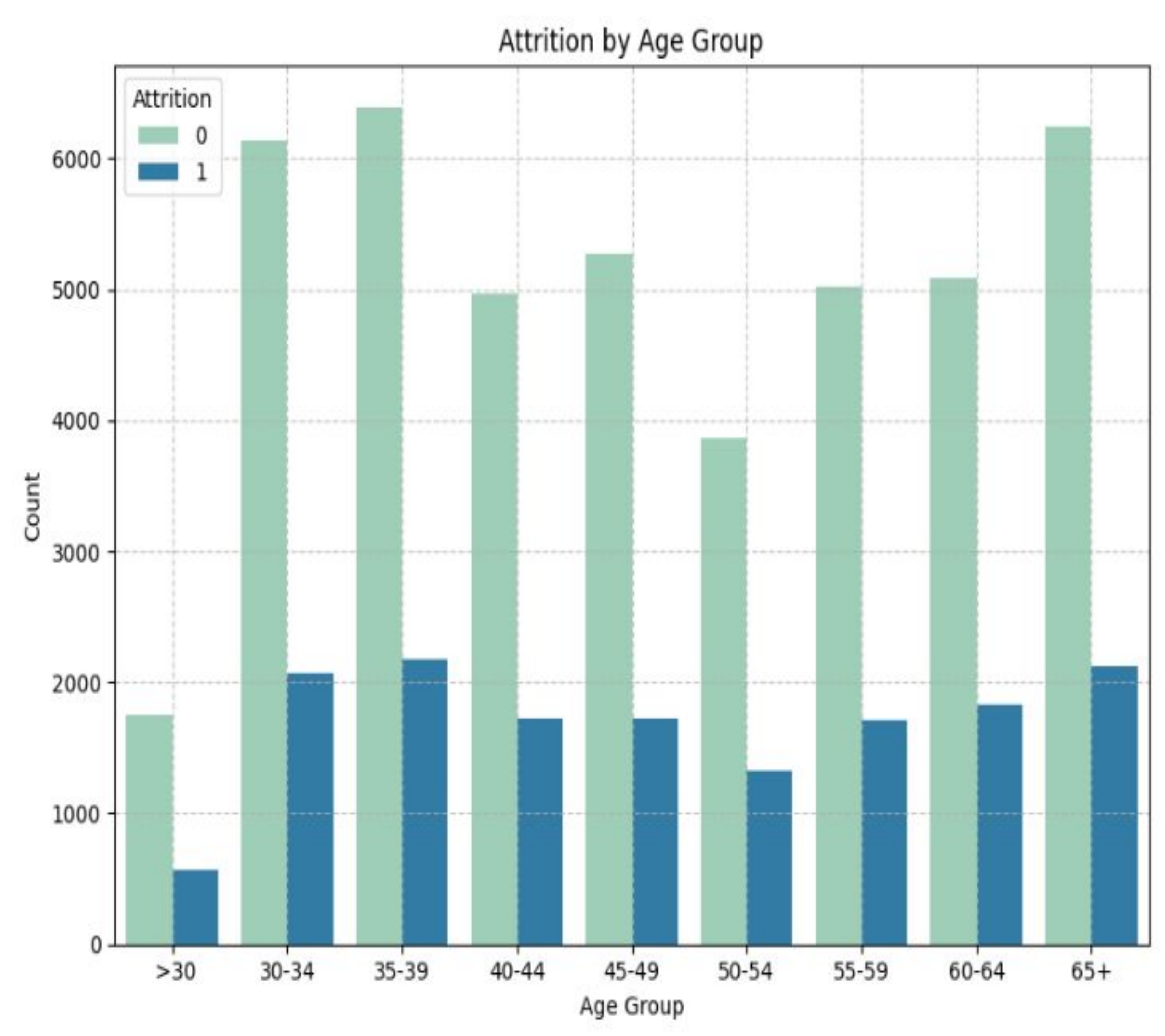
Observations:

- 36% of the workforce is 55+ yrs old.
- 25.4% are 60+

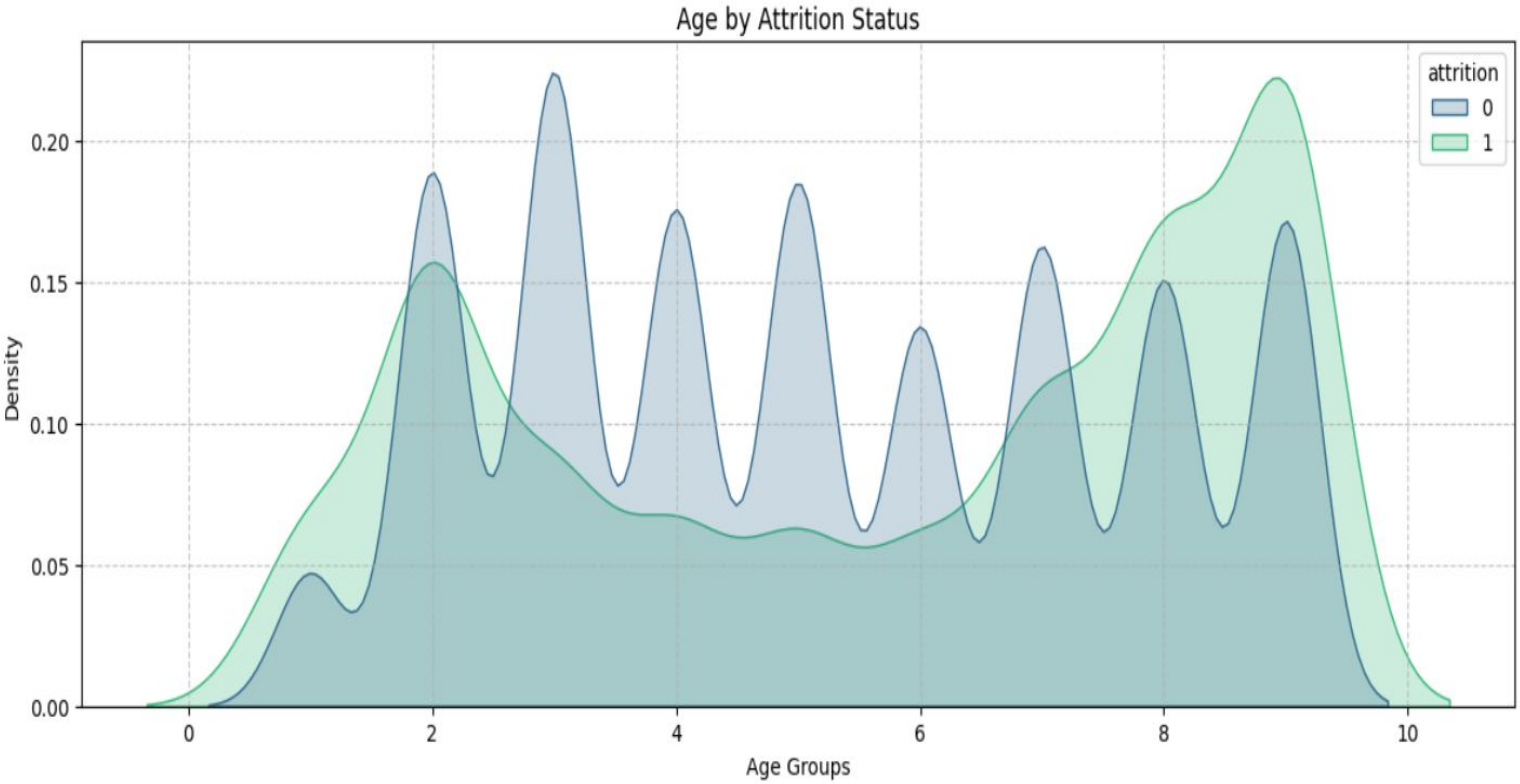
MedicalMatch



attrition age	0	1
1	69.542709	30.457291
2	78.909578	21.090422
3	89.809735	10.190265
4	89.907435	10.092565
5	90.810347	9.189653
6	88.514813	11.485187
7	82.909793	17.090207
8	74.710816	25.289184
9	70.502392	29.497608

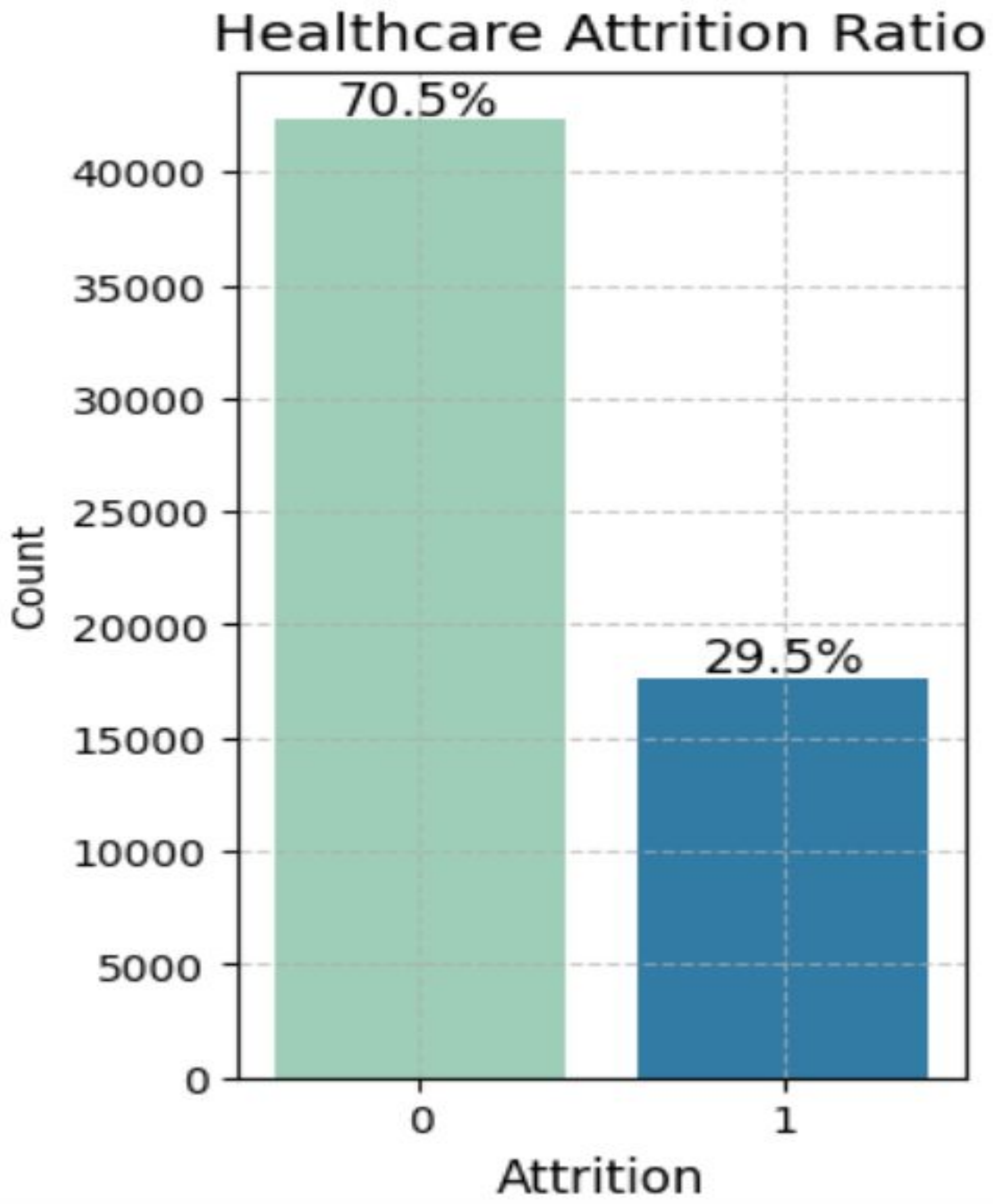


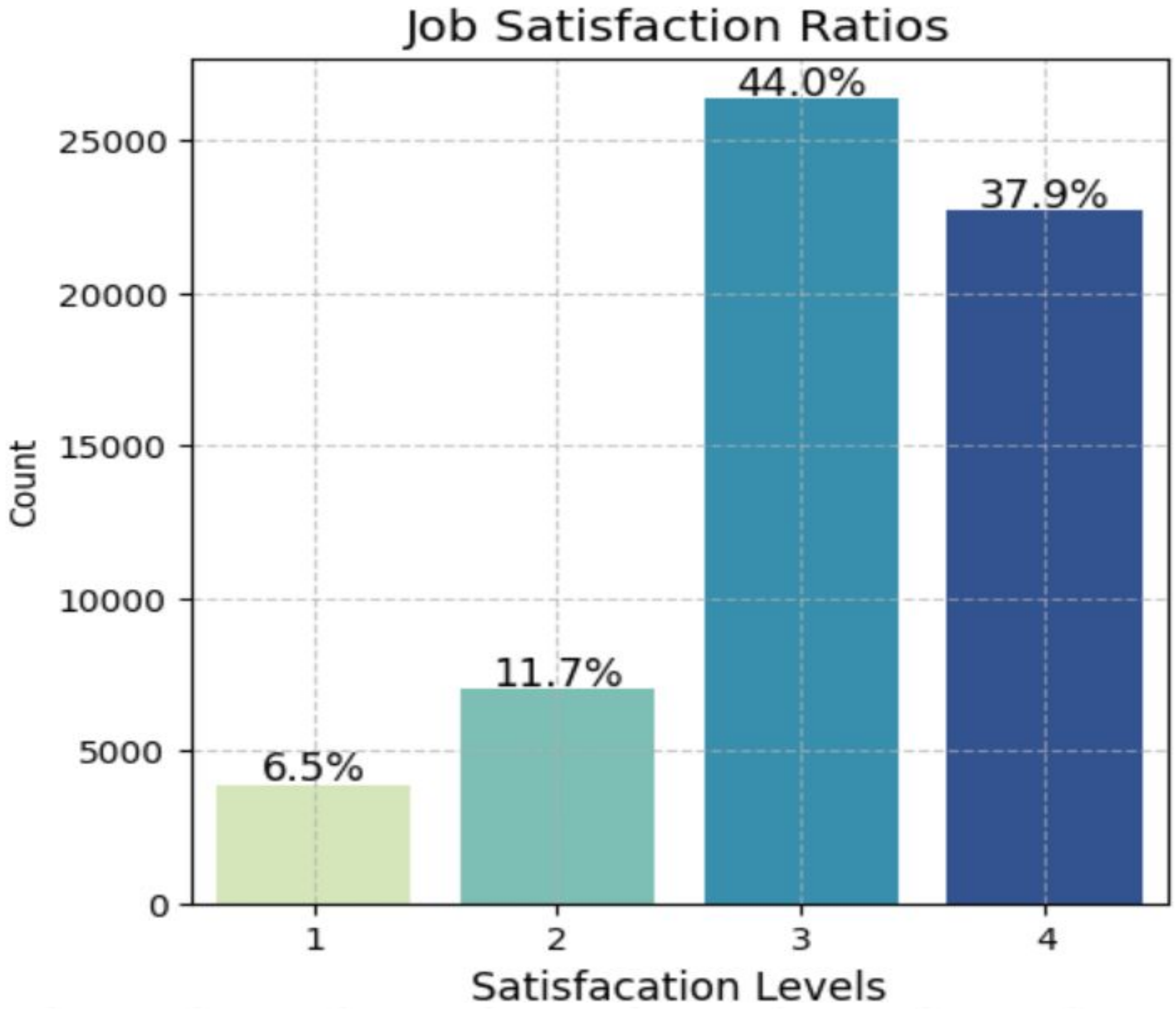
Age Category



Welcome To The Great Healthcare Paradox

MedicalMatch





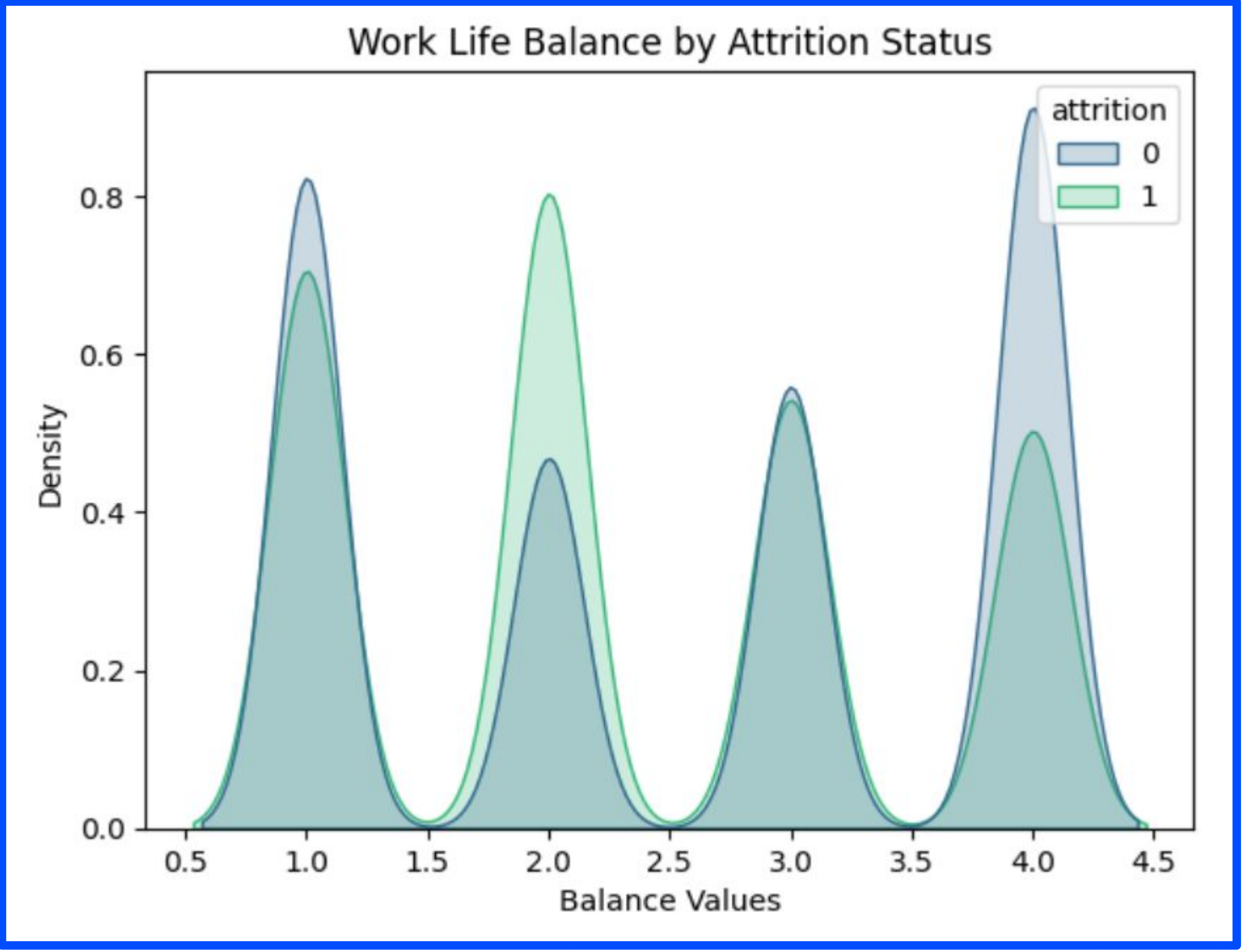
Work Life Balance Impact

WLB Values

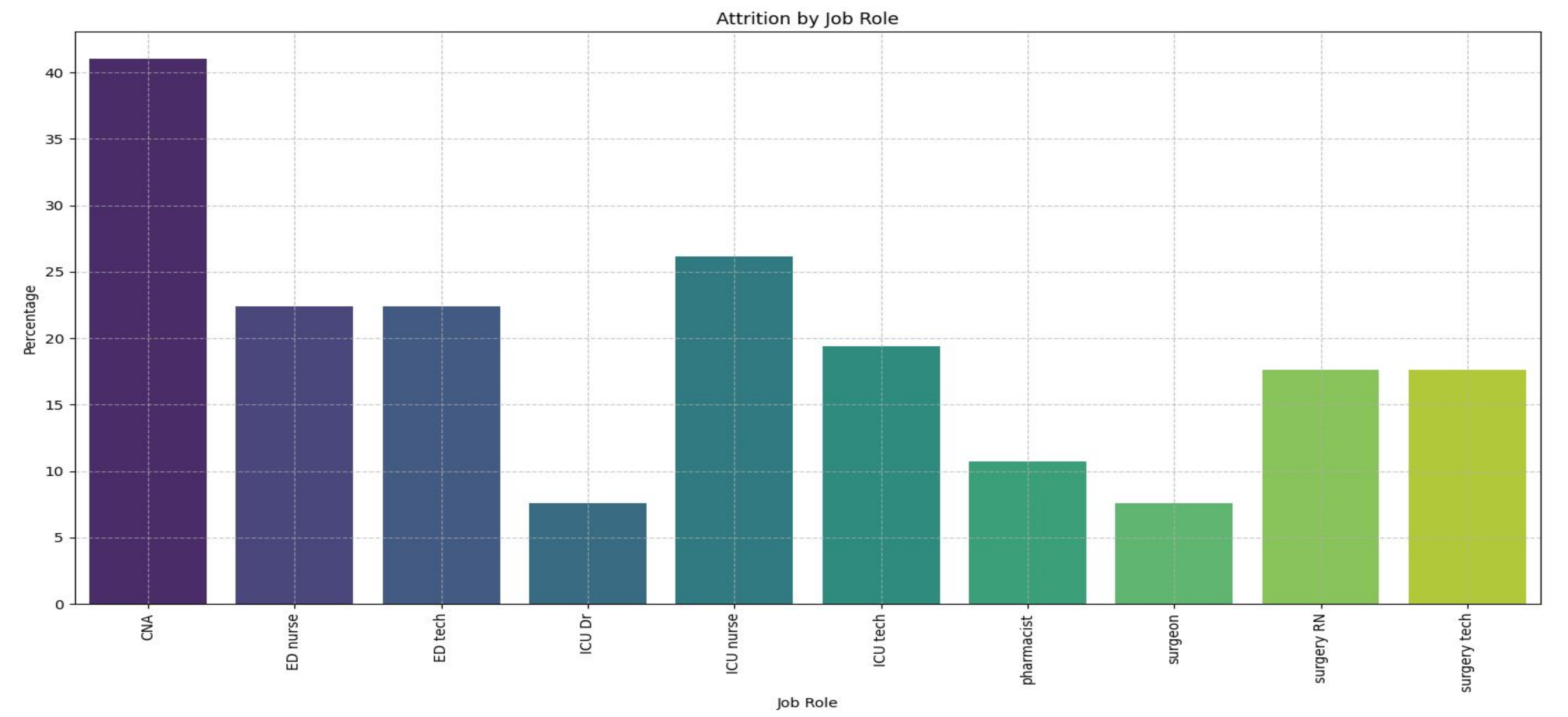
1: Poor
2: Moderate
3: Good
4: Excellent

Attrition Impact

1: 8% more likely
2: 12% more likely



Attrition by Job role



Attrition Cost by Job Role

- Reducing MD attrition by .5% would save US hospitals $250 - 500 billion dollars.

- For every 20 Travel Nurses eliminated, the average hospital could save $2 Million Dollars per year.

Doctors - 7%

$500k - $1Million

RNs - 23%

$56,000

CNAs - 55-100%

$5,000 - $16,000

Work-Life Balance

Recommendations:

- Nursing homes would benefit greatly from MedicalMatch’s staffing solutions as they have the highest ratio of CNA’s

- West, North-East and North-Central Regions have the highest Recruitment Difficulty Index. Those states would benefit greatly from the flexibility and work-life balance that MedicalMatch’s platform facilitates

MedicalMatch

source: NSi 2024 National Healthcare Retention & RN Staffing Report

Classification Models Tested and Performance

Models tested: Bagging, RF, GBC, Adaboost, Dtree, XGBoost, Logistic Regression, ANN

Round 1:
I ran each model w/out tuning to get baseline performance

Round 2:
Oversampling with SMOTE

Round 3:
Random Undersampling

Round: 4
Multiple ANNs w/dropout, Relu/Sig Adam

As expected with an imbalance dataset and no tuning

Dtree and RF overfit on training and tanked on testing.

Tree based models had Recall Scores of 72-75%

ANNs performed better out the gate with train/test performance 80-82%/

MedicalMatch

Tanya Seegmiller