Nurse Attrition in the U.S.

Nidhish Nerur, Michael Crosson, Yue Taira, Megha Sengupta, Erin Kim

OUR TEAM

Erin Kim

Megha Sengupta

Nidhish Nerur

Michael Crosson

Yue Taira

TABLE OF CONTENTS

01.
PROJECT GOALS

DATASET: EMPLOYEE ATTRITION FOR HEALTHCARE

- This dataset has Employee and Hospital Data to predict attrition of nurses in the U.S. Healthcare System

- Contains data related to:
  - Nurses' personal characteristics
  - Satisfaction in the workplace
  - Potential for job growth

Features

Attrition, Age, BusinessTravel, DailyRate, Department, DistanceFromHome, Education, EducationField, EmployeeCount, EnvironmentSatisfaction, Gender, HourlyRate, JobInvolvement, JobLevel, JobRole , JobSatisfaction, MaritalStatus , MonthlyIncome, MonthlyRate, NumCompaniesWorked, Over18, OverTime, PercentSalaryHike, PerformanceRating, RelationshipSatisfaction, StandardHours, Shift, TotalWorkingYears, TrainingTimesLastYear, WorkLifeBalance, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager

QUESTIONS TO ANSWER

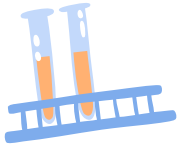
01. — Are there patterns to better understand which nurses are likely to leave a healthcare facility?

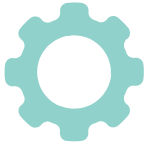
02. — What features are the most important in predicting whether a nurse is likely to leave a healthcare facility?

WHY THIS PROBLEM

Our goal: By determining the features with the greatest predictive power, we can **minimize rehiring costs** while **improving brand reputation** of hospitals

Cost of Nurse Attrition: **$50,000** per nurse





02. EXPLORATORY ANALYSIS

OVERVIEW OF OUR METHODOLOGY

Standardized numeric variables and **one-hot encoded** categorical features

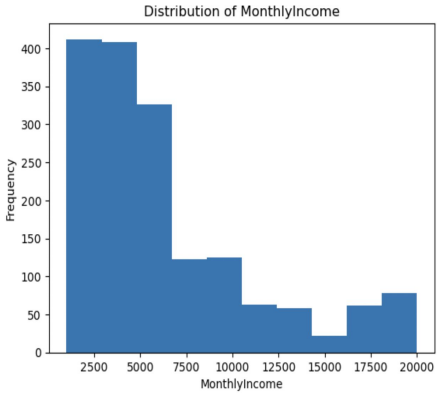
Analyzed **nurse demographics** and work-related factors

Addressed **imbalance nurse attrition outcomes** (12% attrition rate)

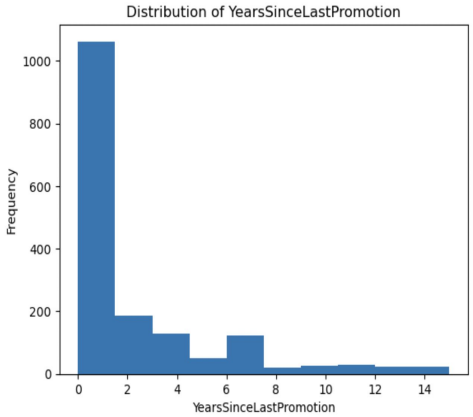
Applied SMOTE for balancing minority classes, improving model performance

DISTRIBUTIONS

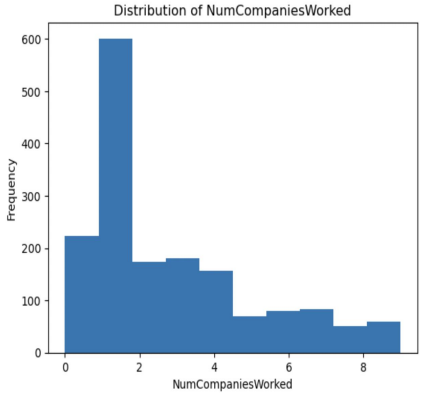
Monthly Income



Years since Last Promotion



Number of Companies Worked

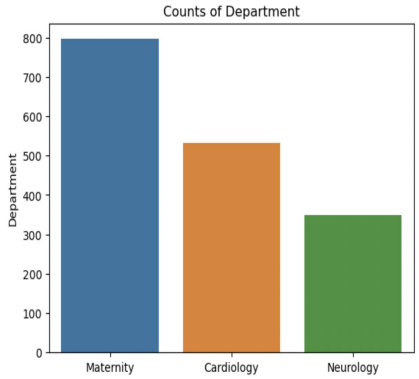


DESCRIPTIVE STATISTICS

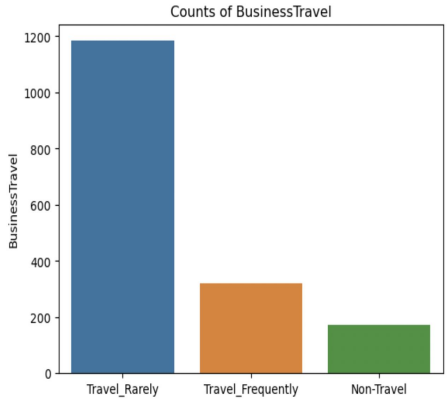
	MonthlyIncome	YearsSinceLastPromotion	NumCompaniesWorked	DistanceFromHome	Age	TotalWorkingYears
count	1676.000000	1676.000000	1676.000000	1676.000000	1676.000000	1676.000000
mean	6516.512530	2.200477	2.662291	9.221957	36.866348	11.338902
std	4728.456618	3.229587	2.477704	8.158118	9.129126	7.834996
min	1009.000000	0.000000	0.000000	1.000000	18.000000	0.000000
25%	2928.250000	0.000000	1.000000	2.000000	30.000000	6.000000
50%	4899.000000	1.000000	2.000000	7.000000	36.000000	10.000000
75%	8380.250000	3.000000	4.000000	14.000000	43.000000	15.000000
max	19999.000000	15.000000	9.000000	29.000000	60.000000	40.000000

CATEGORICAL FEATURES

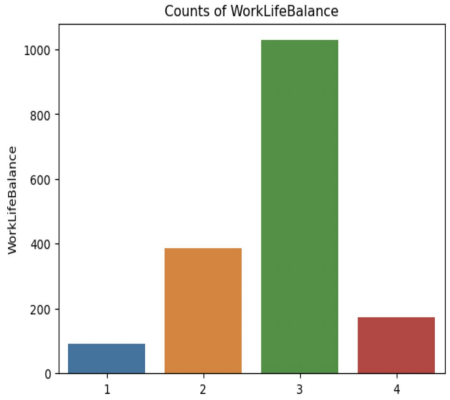
Departments



Business Travel

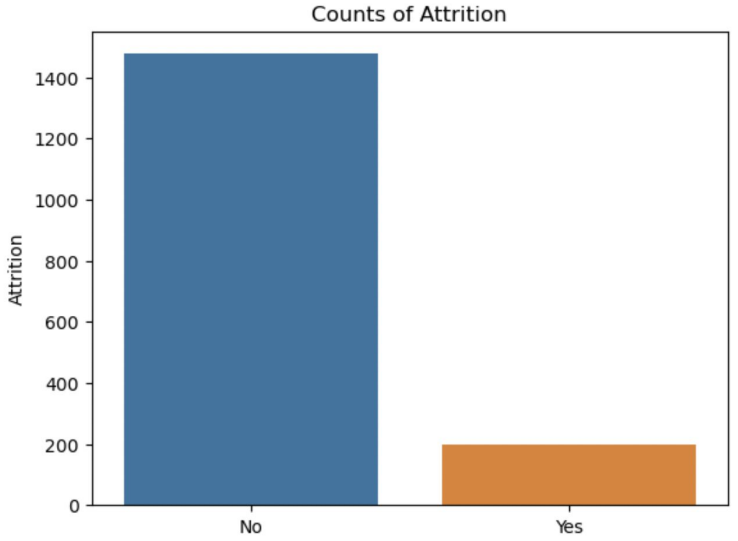


Work-Life Balance



ATTRITION

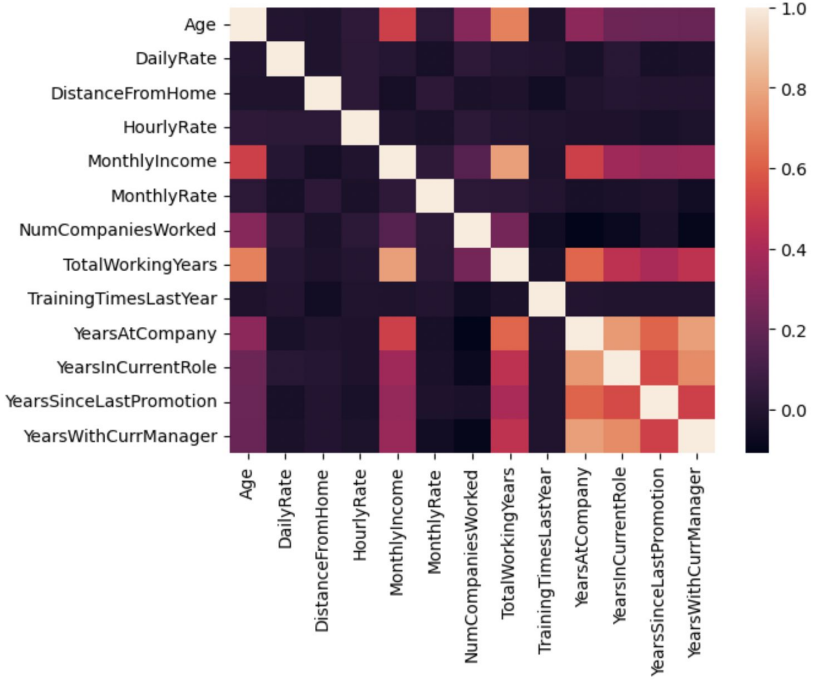
- Nurse Attrition is imbalanced
- 199/1676 leaving
- **12% attrition rate**



CORRELATION HEATMAP

High Correlations

- Monthly Income & Total Working Years
- Years at Company & Years in Current Role
- Years at Company & Years with Current Manager




03.
SOLUTIONS

FEATURES RATIONALE & METHODOLOGY

Removed non-informative features

EmployeCount, Over18, StandardHours

Dropped unique identifier

EmployeeID

Utilized remaining features related to personal life, work, and education

Standardized features, applied one-hot encoding, and addressed class imbalance with SMOTE

CLASSIFIERS TESTED

K-Nearest Neighbors

Logistic Regression

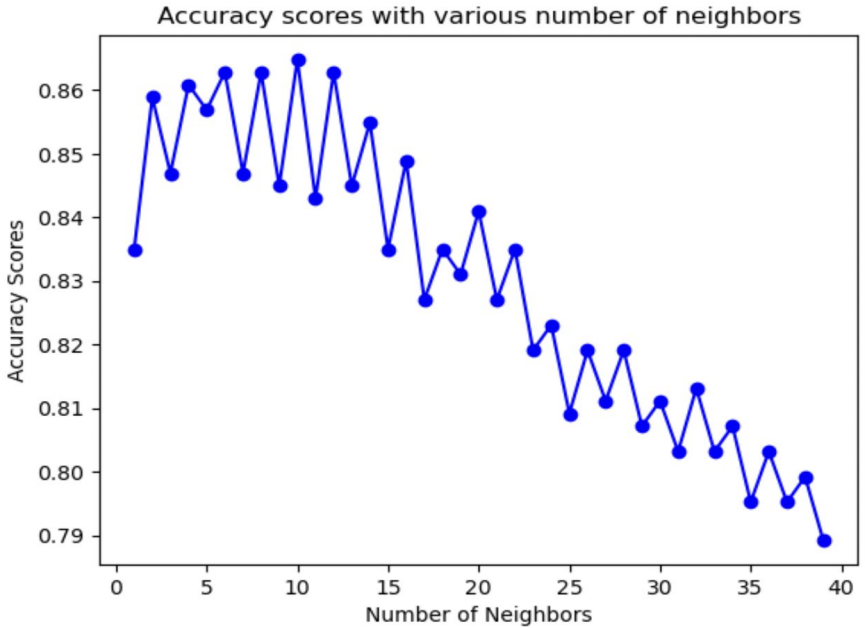
Decision Tree

Random Forest

XGBoost Classifiers

Baseline Accuracy - 87%

K-NEAREST NEIGHBORS



9 nearest neighbors: 87% test accuracy (equal to baseline)

LOGISTIC REGRESSION

	Precision Score	Recall Score
0 (Nurse stays in hospital)	0.96	0.97
1 (Nurse leaves hospital)	0.76	0.73

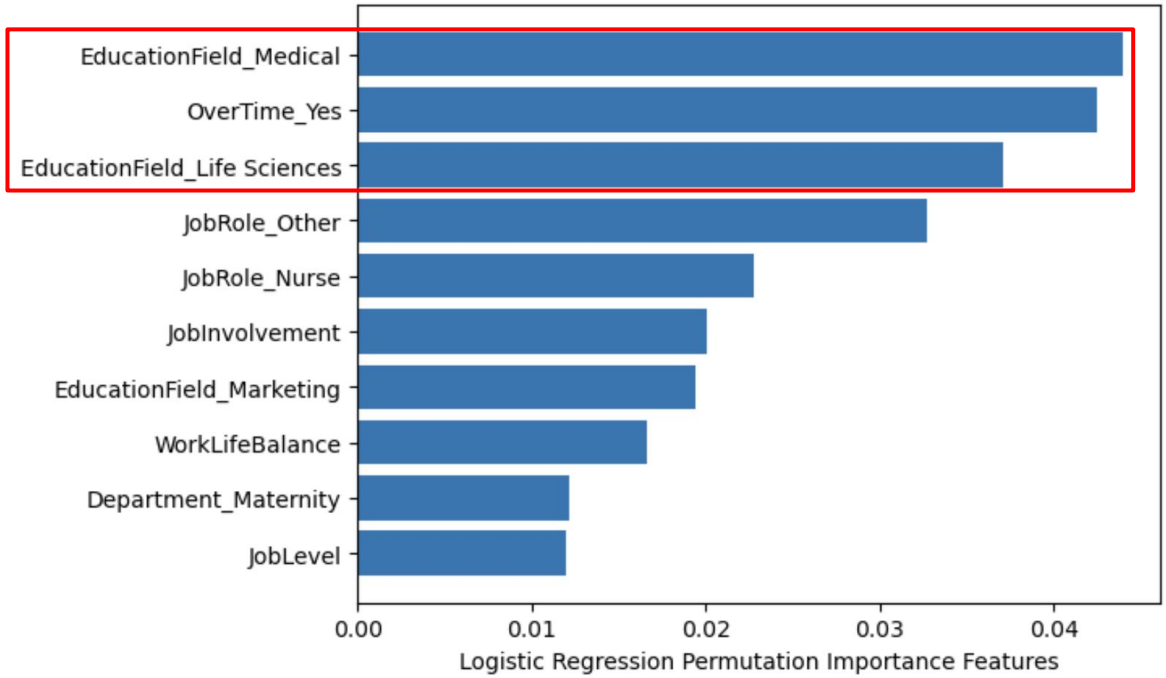
LOGISTIC REGRESSION

Confusion Matrix for Logistic Regression	Predicted 0	Predicted 1
0 (Nurse stays in hospital)	429	14
1 (Nurse leaves hospital)	16	44

04.

INSIGHTS & CONCLUSIONS

CONCLUSIONS



THANKS!

Questions?