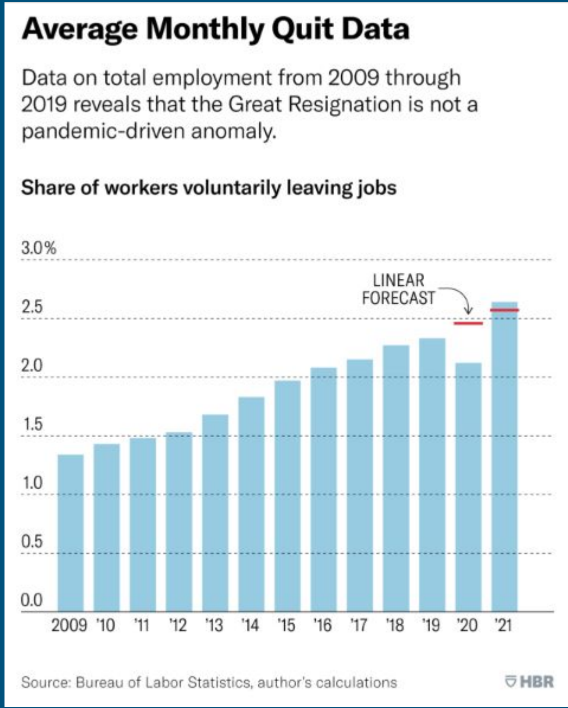
Predicting Voluntary Attrition

Eric Graves

Springboard Data Science Capstone Project

Voluntary Turnover is Rising & Costly

- Trend goes beyond the “Great Resignation”
  - Rising before, still rising after
- FTE turnover can cost up to 2x salary
  - Recruiting & interviewing time/costs
  - Onboarding, lost experience/knowledge
  - Lost business for customer-facing roles
- Identifying flight risks & responding can save $



https://hbr.org/2022/03/the-great-resignation-didnt-start-with-the-pandemic

Leveraging Data Science to Help

- Attrition cannot be “solved”
  - Some attrition inevitable
  - Many drivers cannot be known by the company or captured well in a dataset
- However, data science techniques can help
  - Reduce manual effort
  - Find risks not identified by humans

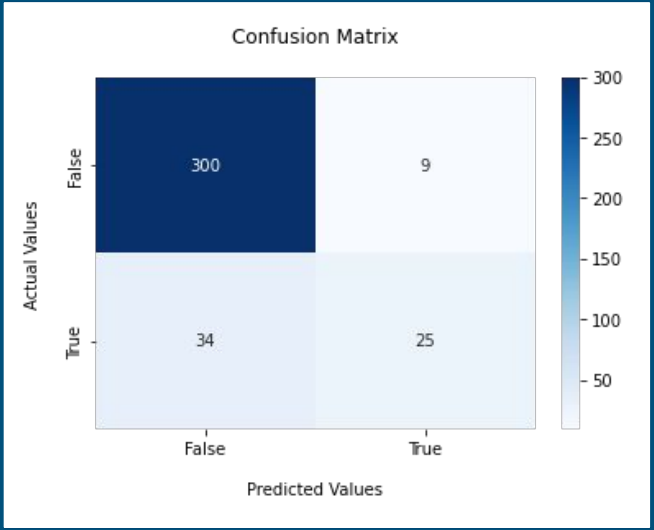
This Project: Predictive Model

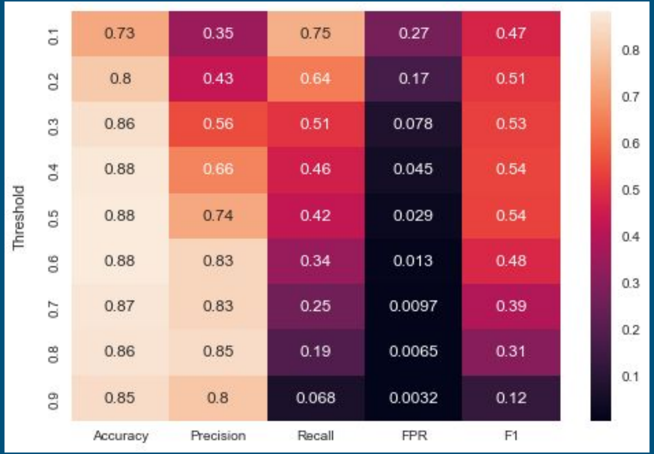
- Statistical model to predict voluntary attrition
- Based on a real-world dataset of ~1,500 anonymized employees
- Dataset contains wide array of info: demography, pay, career, survey, etc.
- Achieved ~75% precision
- Lower recall (could not improve recall without losing material precision)
  - To be expected due to nature of problem (many drivers we can't know or capture in a database)

Dataset: https://data.world/aaizemberg/hr-employee-attrition. Project not completed with reference to any other work analyzing/using this dataset

Model Details

- Logistic regression
  - ~75% precision from both manually created model and Pycaret-generated model
  - Lower recall – tried to increase by adjusting probability threshold, but always lost precision
  - Other Pycaret models did not perform better



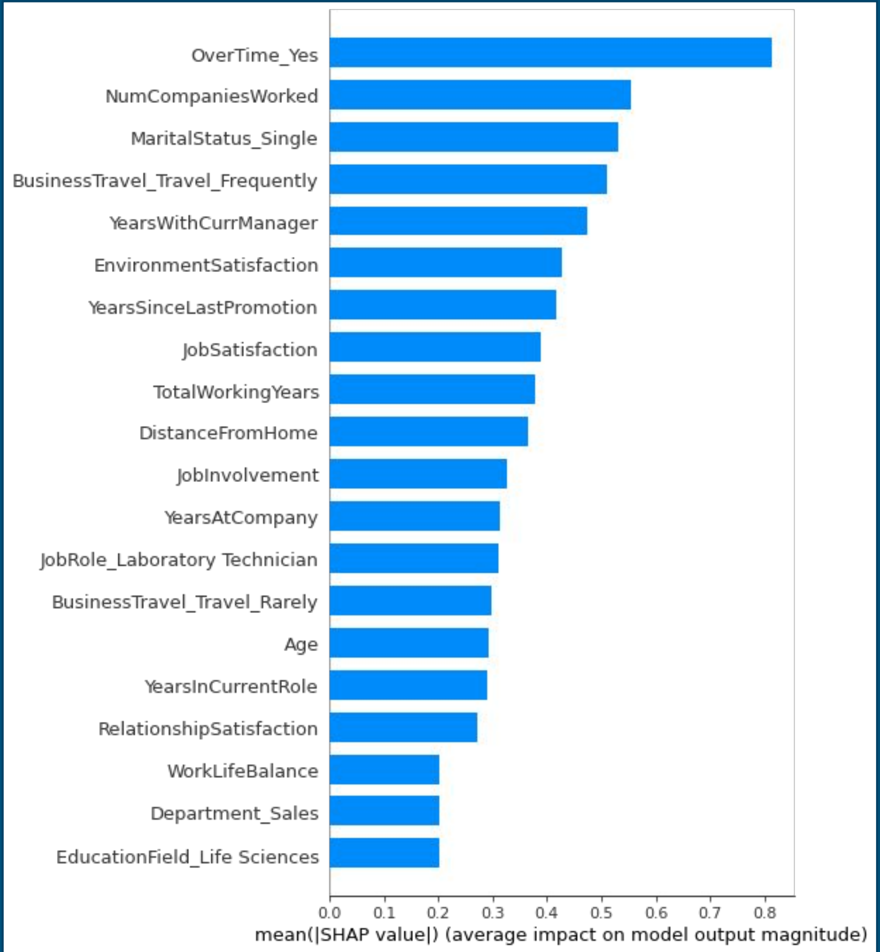


Model Value

- Precision is main value-add
  - If model identifies a flight risk, it is probably real
- Not a crystal ball
  - Supplement to human knowledge/intuition
  - Can reduce manual efforts
- Opportunities to improve include finding additional relevant features and perhaps exploring ensemble modeling

Model Findings

- Chart shows features with highest feature importance (greatest impact on predictions)
- Following pages visually explore relationship between these features and voluntary attrition



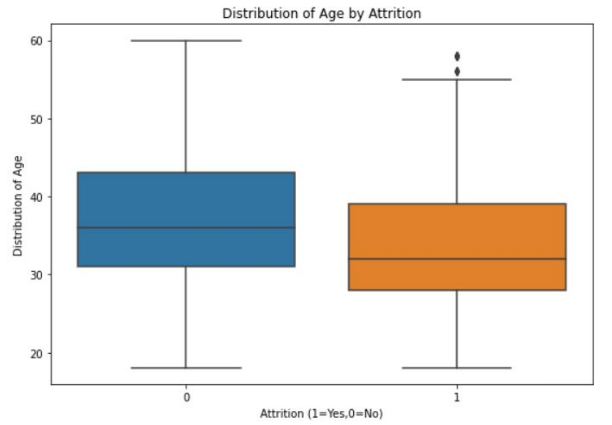
The Data

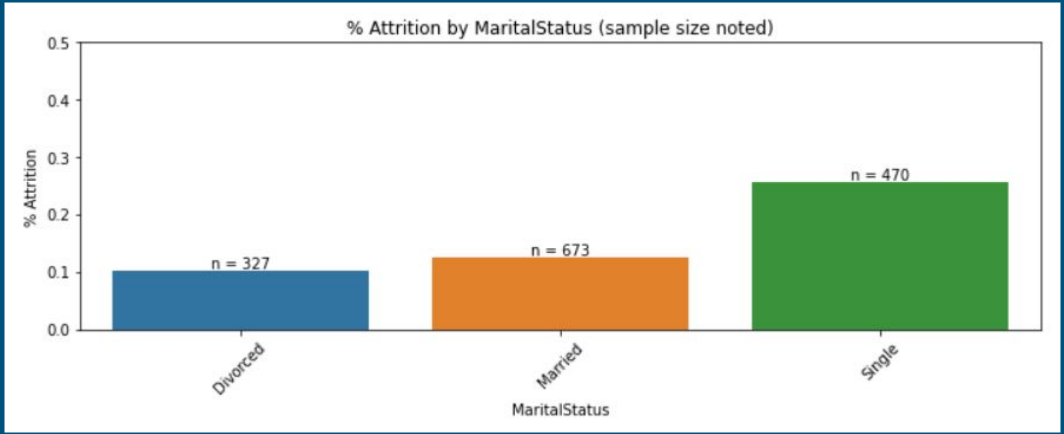
- Information-rich – six categories below
- Clean – no missing values

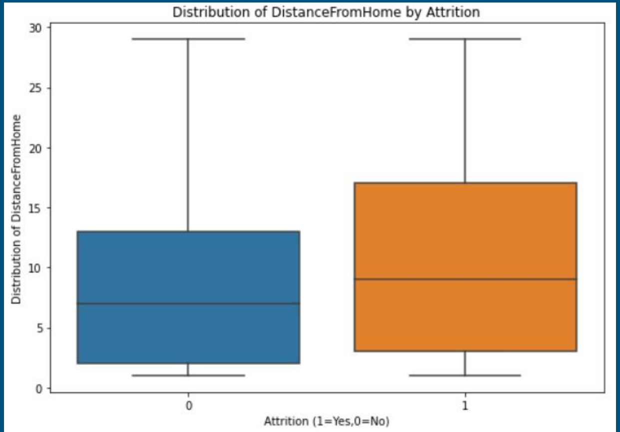
Compensation	Demographic	Education	Role	Survey	Work History
Monthly income	Age	Education field	Department	Environment satisfaction	Number of companies worked
Most recent percentage raise	Distance from home	Education level	Job level	Job involvement	Number of trainings last year
Stock option level	Gender		Job role	Job satisfaction	Performance rating
	Marital status		Overtime-eligible	Relationship satisfaction	Total years at company
			Travel frequency	Work-life balance	Total years working
					Years in current role
					Years since last promotion
					Years with current manager

Key Findings: Demographic Features

- Predictive demographic features included:
  - Age
  - Workplace distance from home
  - Marital status

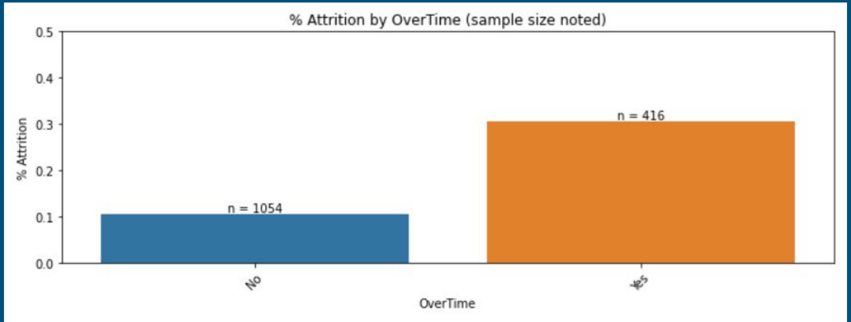


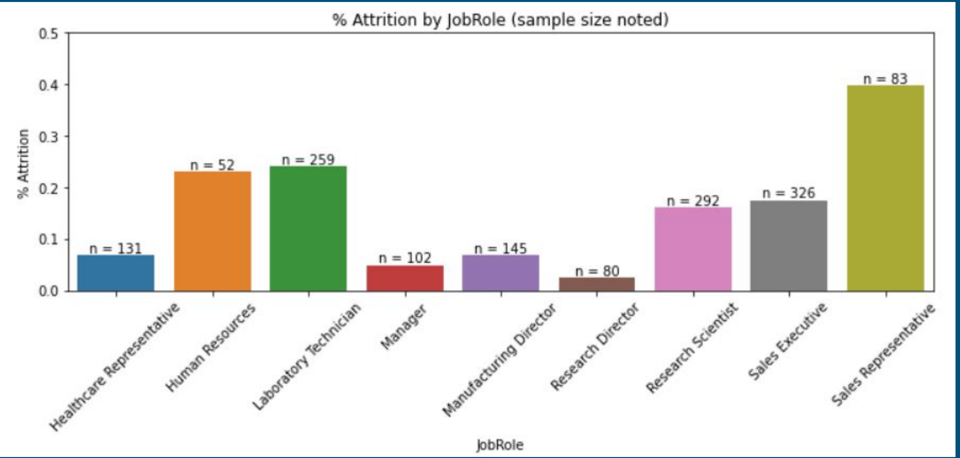


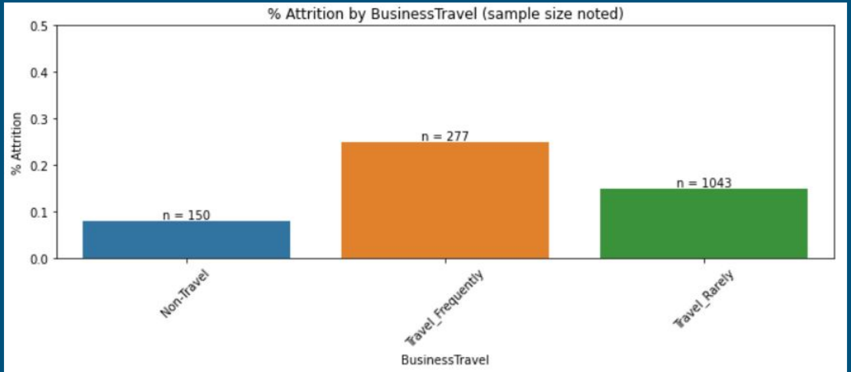


Key Findings: Role-Related Features

- Predictive role-related features included:
    - Overtime-eligible role vs. not
    - Amount of business travel
    - The role function itself

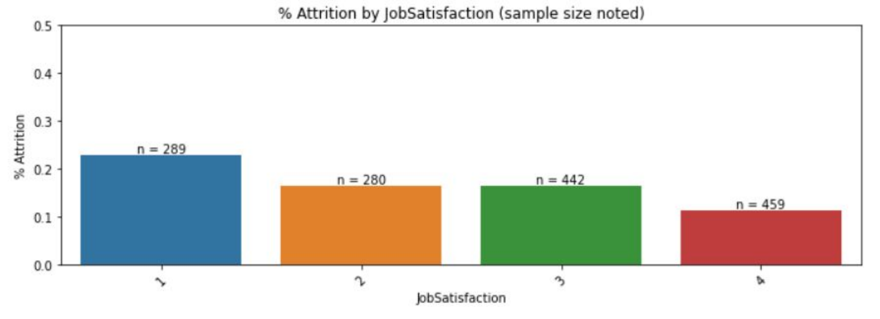


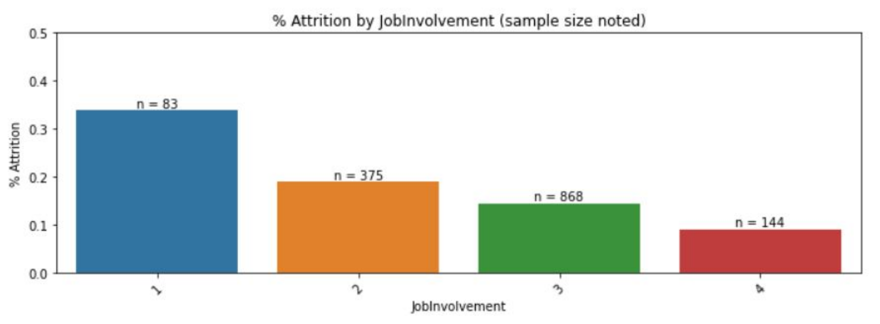


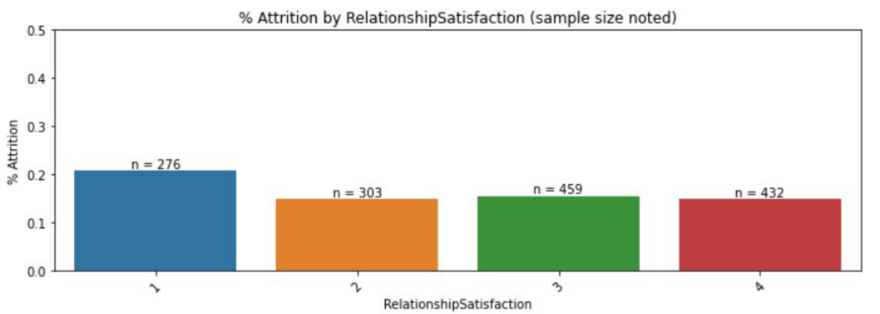


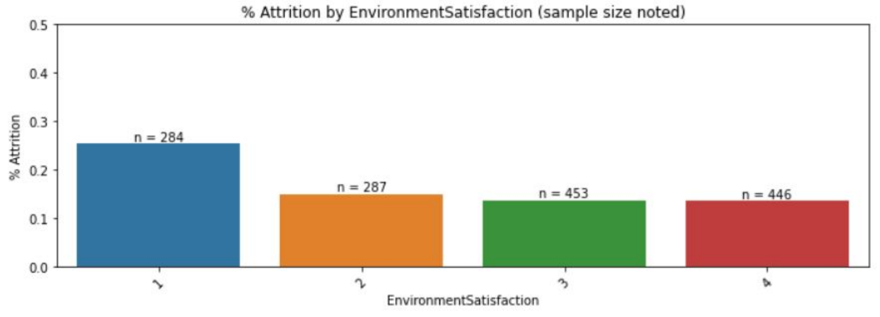
Key Findings: Survey Responses

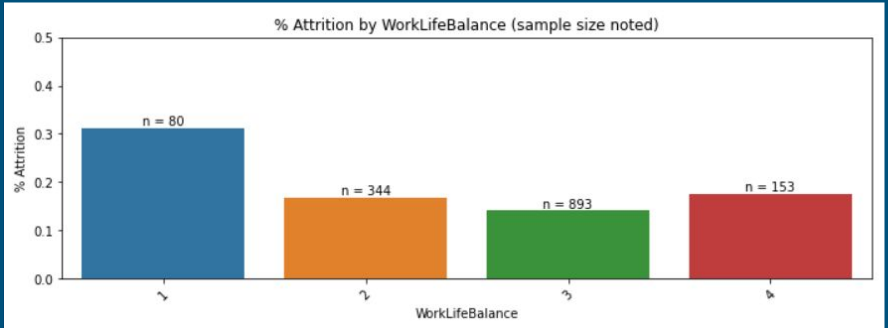
- All survey features were helpful predictors











Key Findings: Work History Features

- Predictive work history features included:
    - Total number of companies worked at
    - Total number of years working (at any company)
    - Years working with current manager