Predicting Employee Attrition

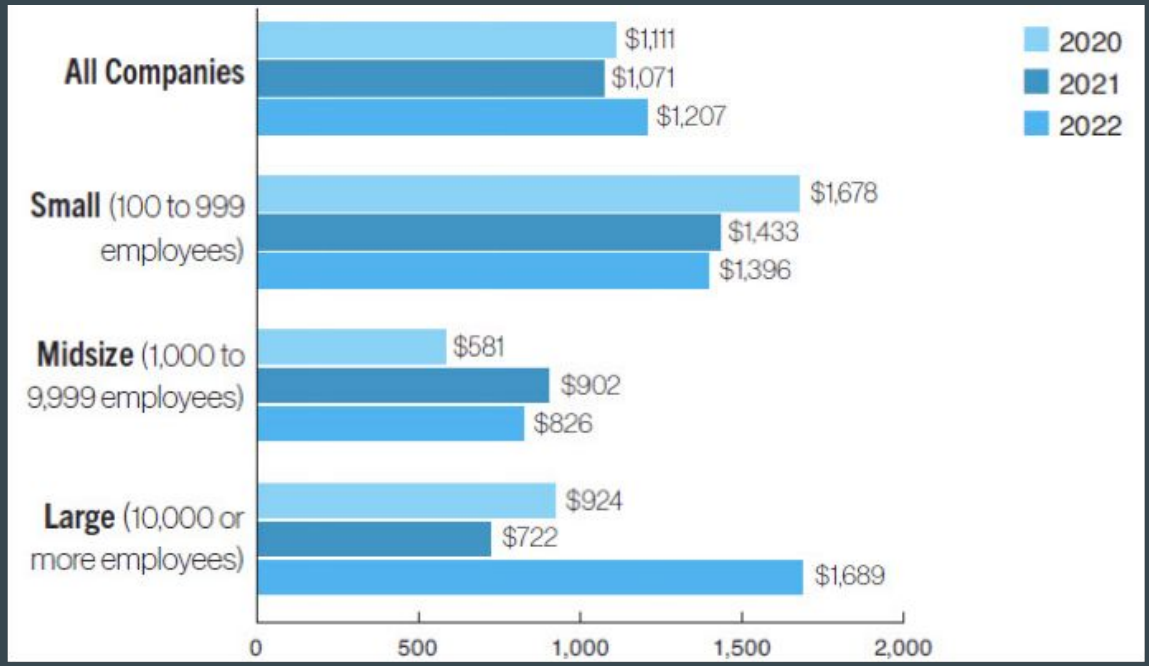
Fatih Demiroz
Capstone 2 Project Presentation

Problem: Attrition

Companies spend 50 - 80% of their operating budgets on employees

Employee attrition costs companies

- money,
- time,
- productivity,
- culture, and
- institutional memory



Context and Purpose

Context:

- CW Research and Analytics LLC is a medical and genetics consulting firm
- CW has experienced 16% employee turnover since last year (industry standard is 4%)

Purpose:

- Find the factors contributing to attrition
- Bring attrition to at or below 4% within 2 years

Data

Data:

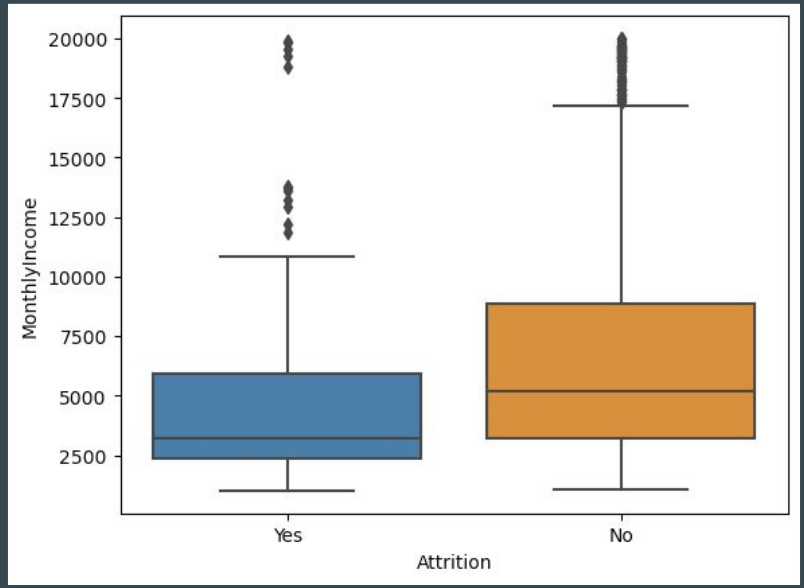
- Single CSV file with 35 columns and 1,470 rows.

Table 1: Feature Properties

Column Name	Data Type	Value Type	Column Name	Data Type	Value Type
Age	Int64	Discrete	MonthlyIncome	Int64	Discrete
Attrition	Object	Binary	MonthlyRate	Int64	Discrete
BusinessTravel	Object	Ordinal	NumCompaniesWorked	Int64	Discrete
DailyRate	Int64	Discrete	Over18	Int64	Discrete
Department	Object	Nominal	OverTime	Int64	Binary
DistanceFromHome	Int64	Ordinal	PercentSalaryHike	Int64	Discrete
Education	Int64	Ordinal	PerformanceRating	Int64	Ordinal
EducationField	Object	Nominal	RelationshipSatisfaction	Int64	Ordinal
EmployeeCount	Int64	Discrete	StandardHours	Int64	Discrete
EmployeeNumber	Int64	Discrete	StockOptionLevel	Int64	Ordinal
EnvironmentSatisfaction	Int64	Ordinal	TotalWorkingYears	Int64	Discrete
Gender	Object	Binary	TrainingTimesLastYear	Int64	Discrete
HourlyRate	Int64	Discrete	WorkLifeBalance	Int64	Ordinal
JobInvolvement	Int64	Ordinal	YearsAtCompany	Int64	Discrete
JobLevel	Int64	Ordinal	YearsInCurrentRole	Int64	Discrete
JobRole	Object	Nominal	YearsSinceLastPromotion	Int64	Discrete
JobSatisfaction	Int64	Ordinal	YearsWithCurrManager	Int64	Discrete
MaritalStatus	Object	Nominal			

Exploratory Data Analysis

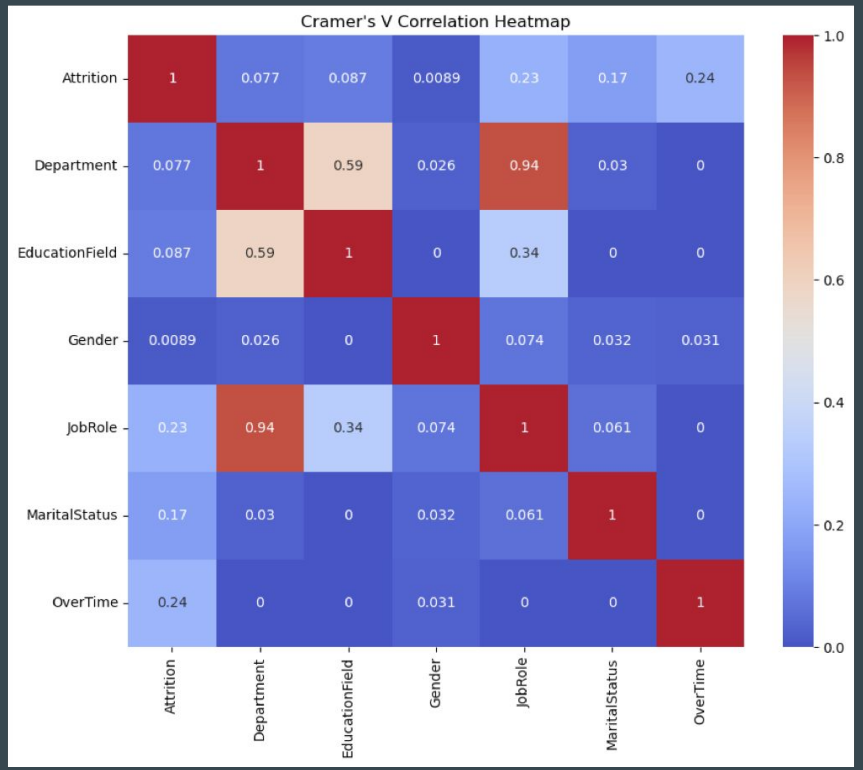
Monthly Income and Attrition



Correlation Heatmap: Nominal Features

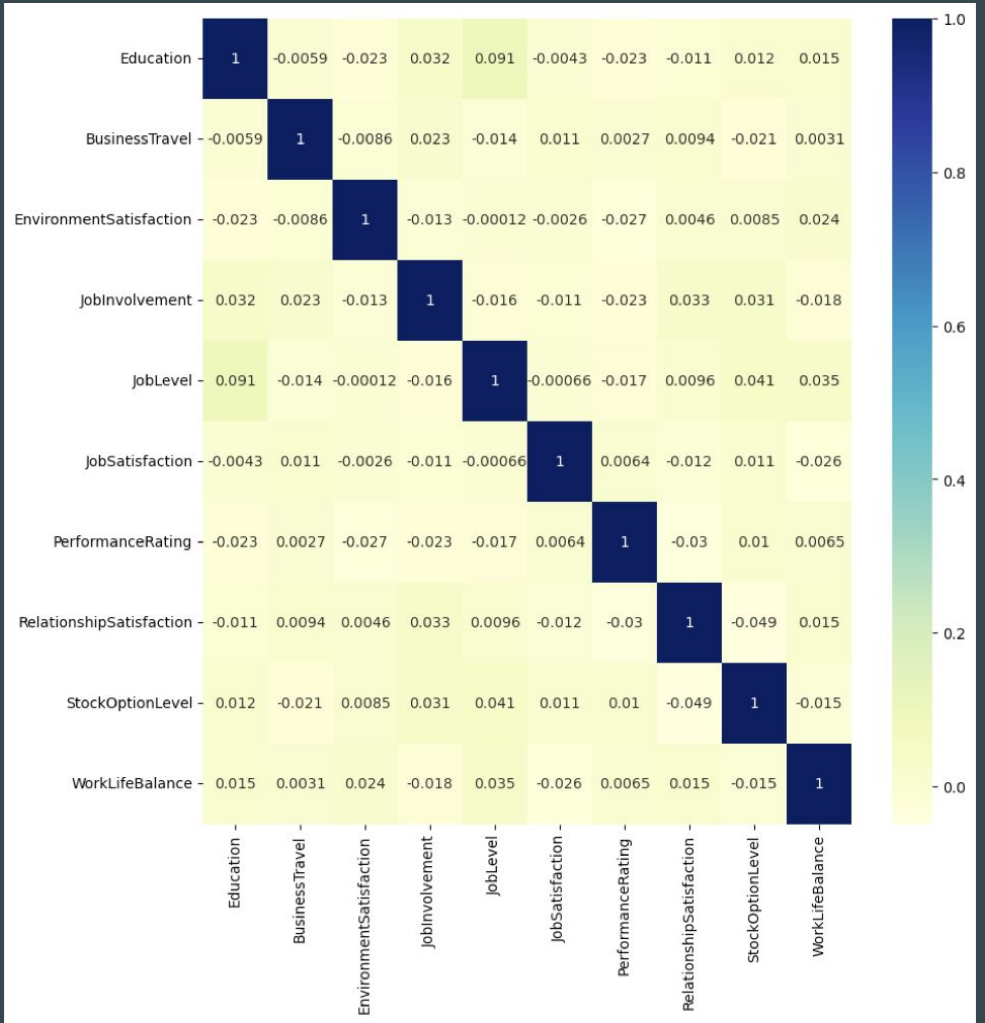
Some association

- Attrition - JobRole
- Attrition - Overtime



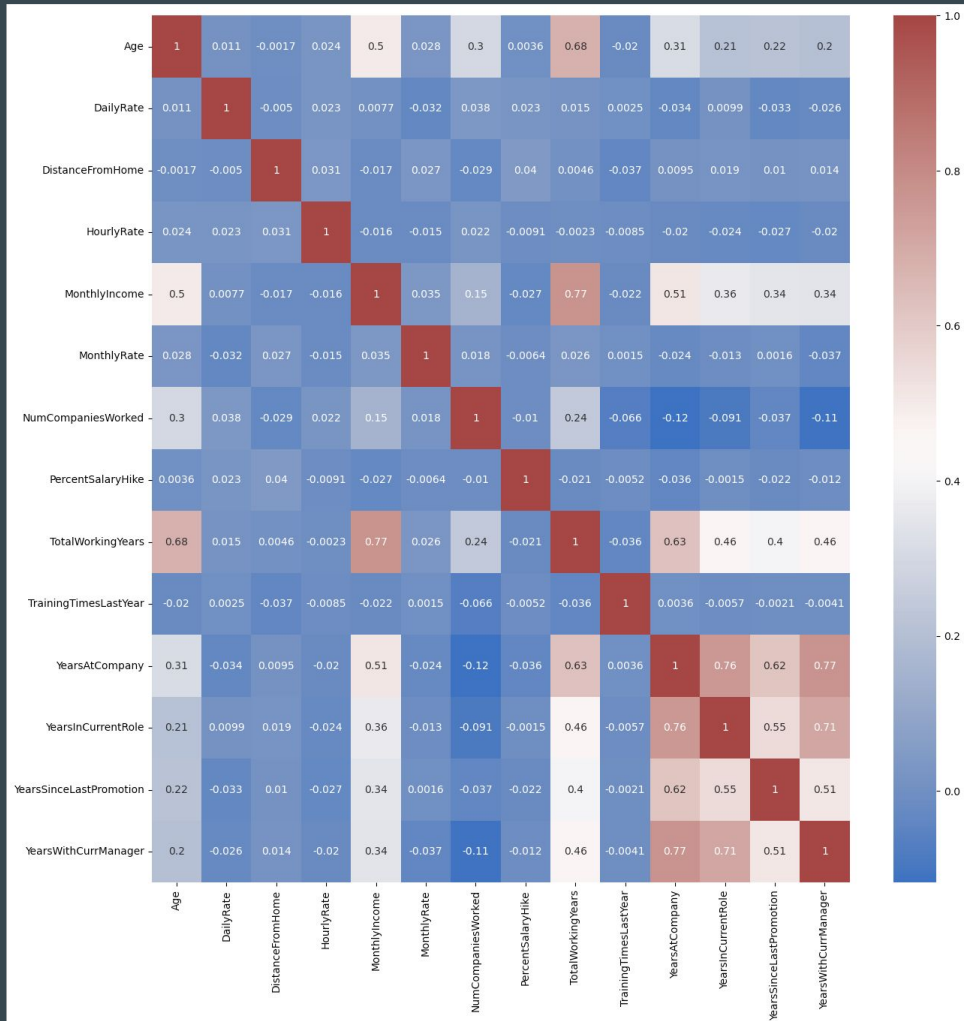
Correlation Heatmap: Ordinal

No visible correlation between ordinal features



Correlation Heatmap: Numeric

Moderate correlation between year related features



Preprocessing

The dataframe split into three smaller dataframes for preprocessing

- Df for nominal features: converted to dummy variables
- Df for ordinal features: values were mapped and converted to numeric
- Df for numeric features: standardized and scaled using MinMax scaler

Then, these three processed dataframes were combined into one df for modeling phase.

Modeling

Five separate classification models:

We have tolerance for False Positives

Best results from Logistic Regression

		Models				
METRIC		**SVC**	**Decision Tree**	**Random Forest**	**XG Boost**	**Logistic Regression**
Accuracy		0.89	0.83	0.87	0.88	0.88
Precision	No Attrition	0.89	0.89	0.87	0.88	0.9
	Attrition	0.77	0.33	0.64	0.71	0.63
Recall	No Attrition	0.99	0.92	0.99	0.99	0.97
	Attrition	0.28	0.26	0.11	0.2	0.36
F1-Score	No Attrition	0.94	0.9	0.93	0.93	0.93
	Attrition	0.41	0.29	0.19	0.31	0.46
False Positive		5	32	4	8	13
False Negative		44	45	54	45	39
True Negative		375	348	376	372	367
True Positive		17	16	7	16	22

Features Contributing to Attrition

Primary causes

- Years since last promotion
- Number of companies worked
- Overtime
- Job Roles
  - Lab Technicians
  - Sales Reps
  - HR staff
- Business Travel

Feature	Coefficient	% Increase in Odds of Attrition
YearsSinceLastPromotion	1.569122	380.24
NumCompaniesWorked	1.543734	368.20
OverTime_Yes	1.299739	266.83
MaritalStatus_Single	1.064923	190.06
JobRole_Laboratory Technician	0.947741	157.99
BusinessTravel	0.855795	135.32
JobRole_Sales Representative	0.80235	123.08
EducationField_Technical Degree	0.758158	113.43
JobRole_Human Resources	0.566221	76.16
Gender_Male	0.560651	75.18
EducationField_Human Resources	0.501213	65.07
Department_Sales	0.480375	61.67
YearsAtCompany	0.388956	47.54

Features Reducing Attrition

Factors that reduce attrition are:

- Years in current role
- Total working years
- Age
- No overtime
- Job roles:
  - Healthcare reps
  - Research director

Feature	Coefficient	% Decrease in Odds of Attrition
YearsInCurrentRole	-1.49803	-77.64
TotalWorkingYears	-1.34368	-73.91
Age	-1.156059	-68.53
JobRole_Healthcare Representative	-0.84335	-56.97
JobRole_Research Director	-0.793108	-54.76
YearsWithCurrManager	-0.752833	-52.9
OverTime_No	-0.609669	-45.65
TrainingTimesLastYear	-0.59165	-44.66
JobInvolvement	-0.464995	-37.19
MaritalStatus_Divorced	-0.423062	-34.5
EnvironmentSatisfaction	-0.3512	-29.62
EducationField_Life Sciences	-0.349265	-29.48

Recommendations

1. Year gap between Promotions: Invest in employee training and development programs. Find ways to promote employees or show other pathos for career advancement
2. Overtime: Minimize overtime
3. Job Roles: Investigate complaints about pay and burnout for Lab technicians, Sales Reps, and HR staff.
4. Number of Companies Worked: HR shall prioritize job applicants with fewer job changes throughout their careers. HR should redflag the job applicants who change too many jobs in a short time period. Also, the C Suite should consider structural changes for accommodating the work habits of the Millennials and GenZ.

Q & A