HR Attrition Analytics

LetsGraduate.LLC

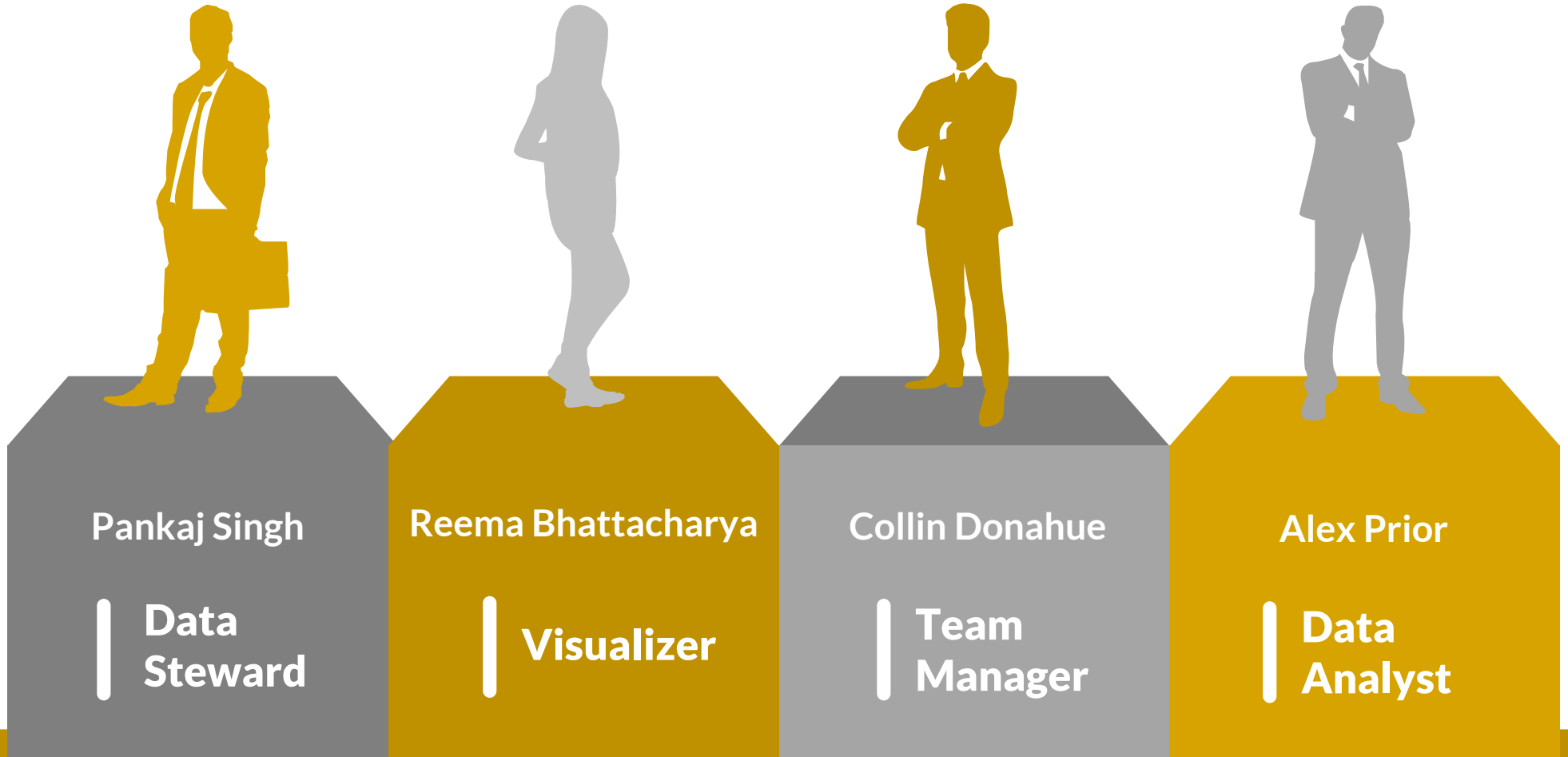
Let’sGraduate.LLC

Project Team

MBA 551

Spring 2021

Final Team Project



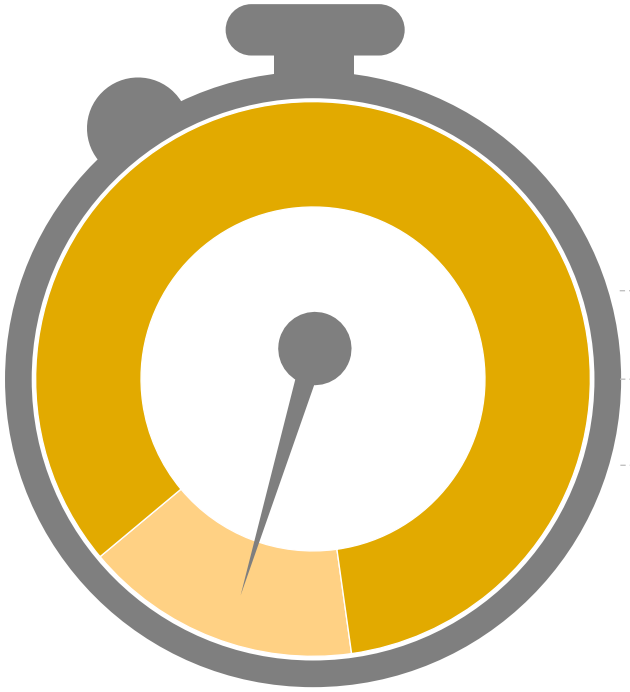
Executive Summary

- The current employee attrition rate is 16.66%
- Build predictive models using five modeling techniques
- The Boosted Tree model had the highest accuracy rate (89.11%) and was used for data analysis
- Overtime, business travel, job level, work life balance, stock options are the top five contributors to employee attrition
- The company has functional attrition of older, higher paid employees
- Recommend reducing overtime, reducing business travel, increasing compensation, and earlier vesting of benefits
- Limitations include no benchmarking
- Next steps further departmental analysis and building risk models

Introduction

Summary of the problem

The HR department provided a dataset with 1470 employee records which shows a 16.66% attrition rate for the organization. The organization needs to comprehend what variables are high contributors to employee attrition in the firm and make a model that can predict if a specific worker will leave the organization or not. The objective is to make or improve distinctive maintenance procedures and help executives make better dynamic activities. Developed three prediction models – Decision Tree, Bootstrap Forest, Boosted Tree to discover best fit model and with the best precision.



Data shows 16.66% attrition rate

Data shows 83.88% employee retention rate

How can we increase employee retention?

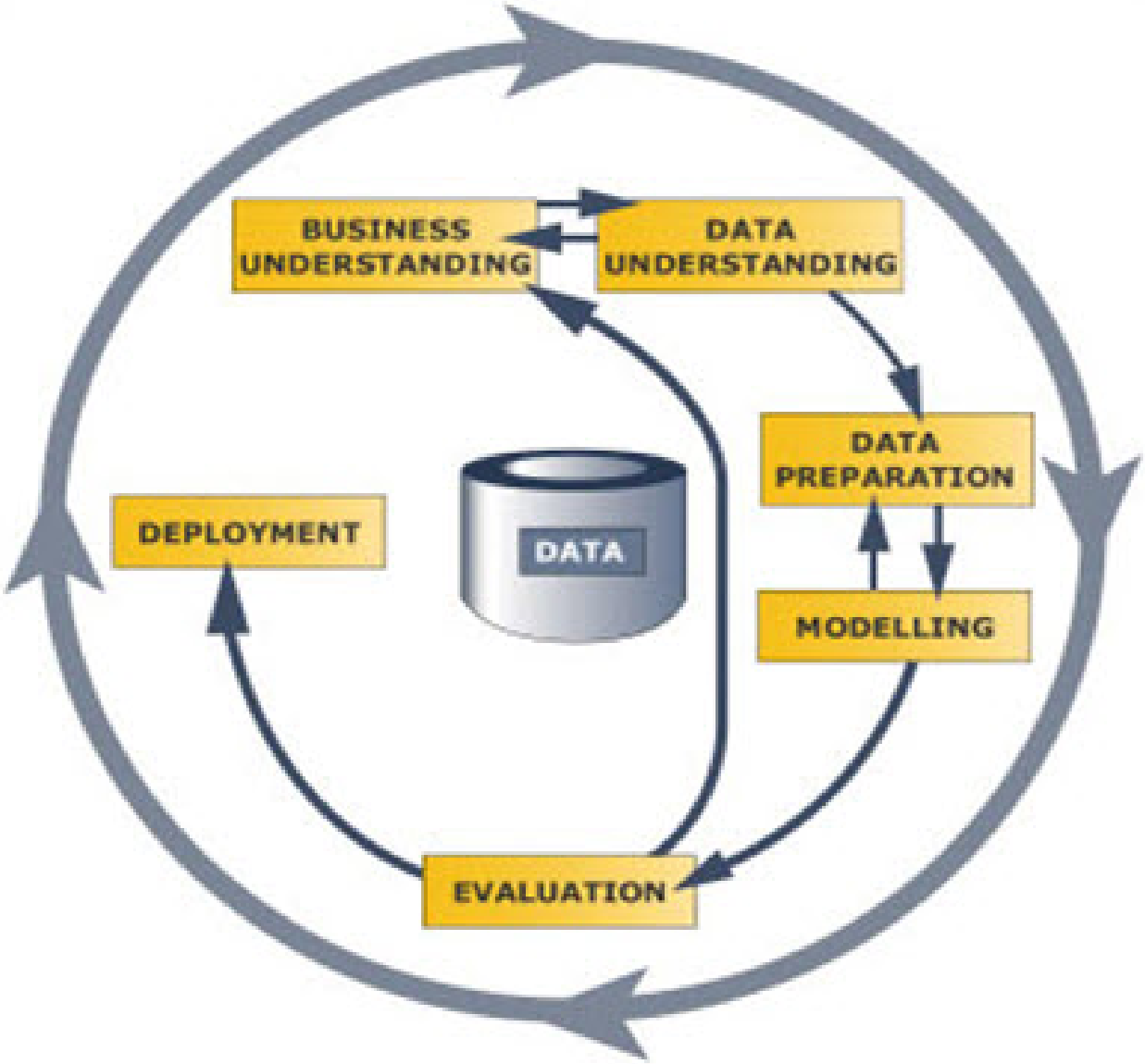
Business Goals

Problem Statement

Employees Build Business

"You don't build a business. You build people, and people build the business." - Zig Ziglar.

1. Find the factors contributing to current attrition rate of the company
2. To develop models to predict if an employee is likely to leave
3. Develop strategies reduce future turnover for better employee retention
4. To drive the right kind of value from the analytic process



Modeling Process

CRISP_DM

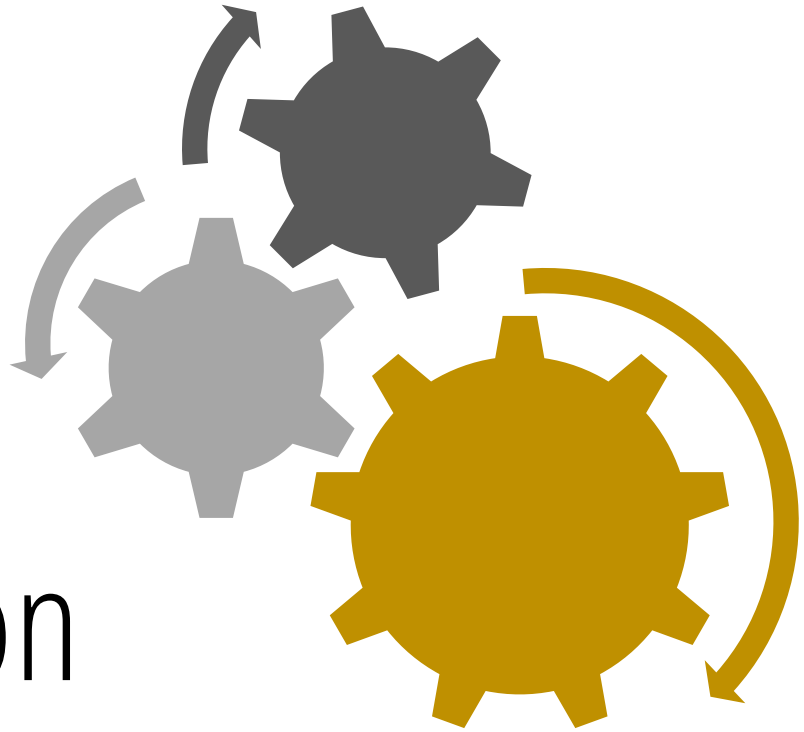
DATA UNDERSTANDING

DATA EXPLORATION

Using various exploratory techniques, we develop data insights

DATA CLEANUP

Using analytic processes, we identify dummy variables and missing values



Data Exploration

Identifying the indicators that will help predict if an employee will leave the company or not

EMPLOYEE DEMOGRAPHICS

Summary of key findings for employee demographic distribution

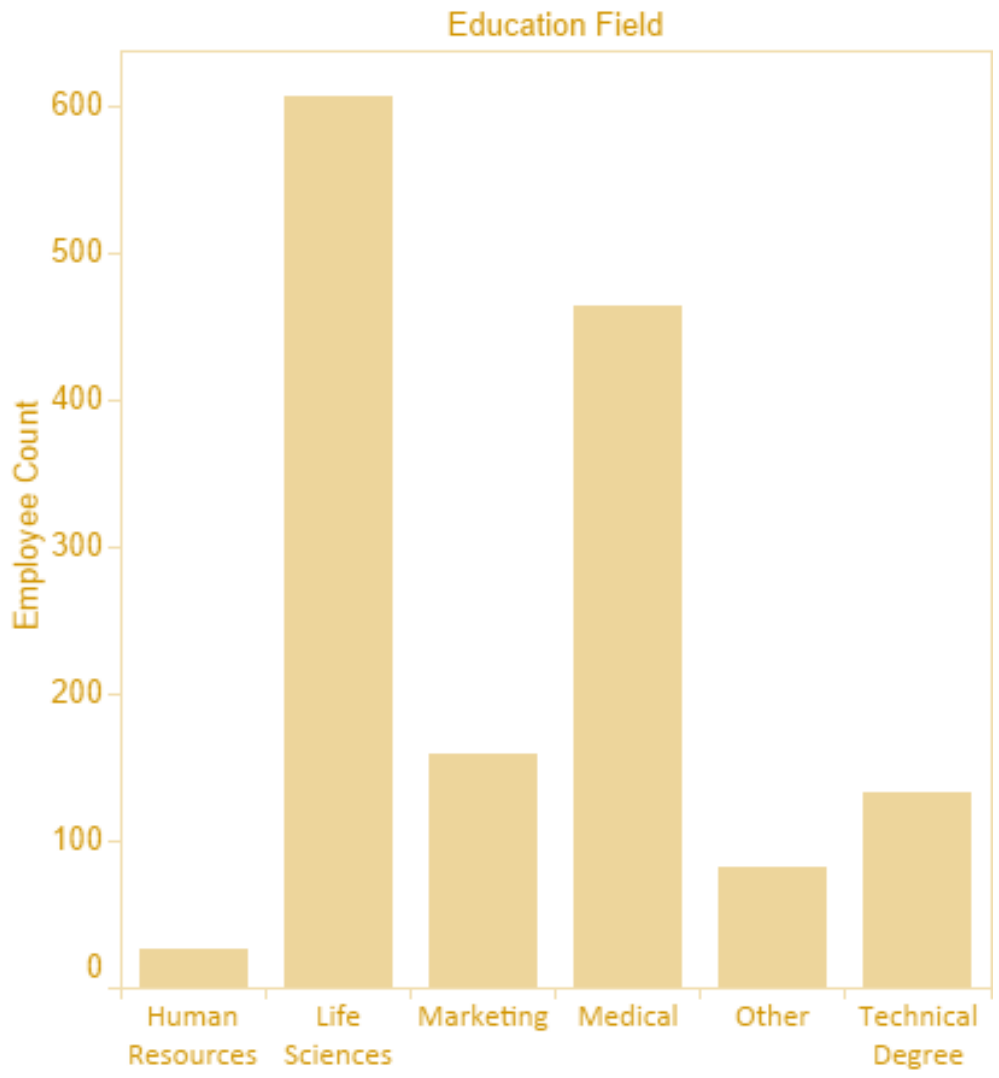
46%
Married

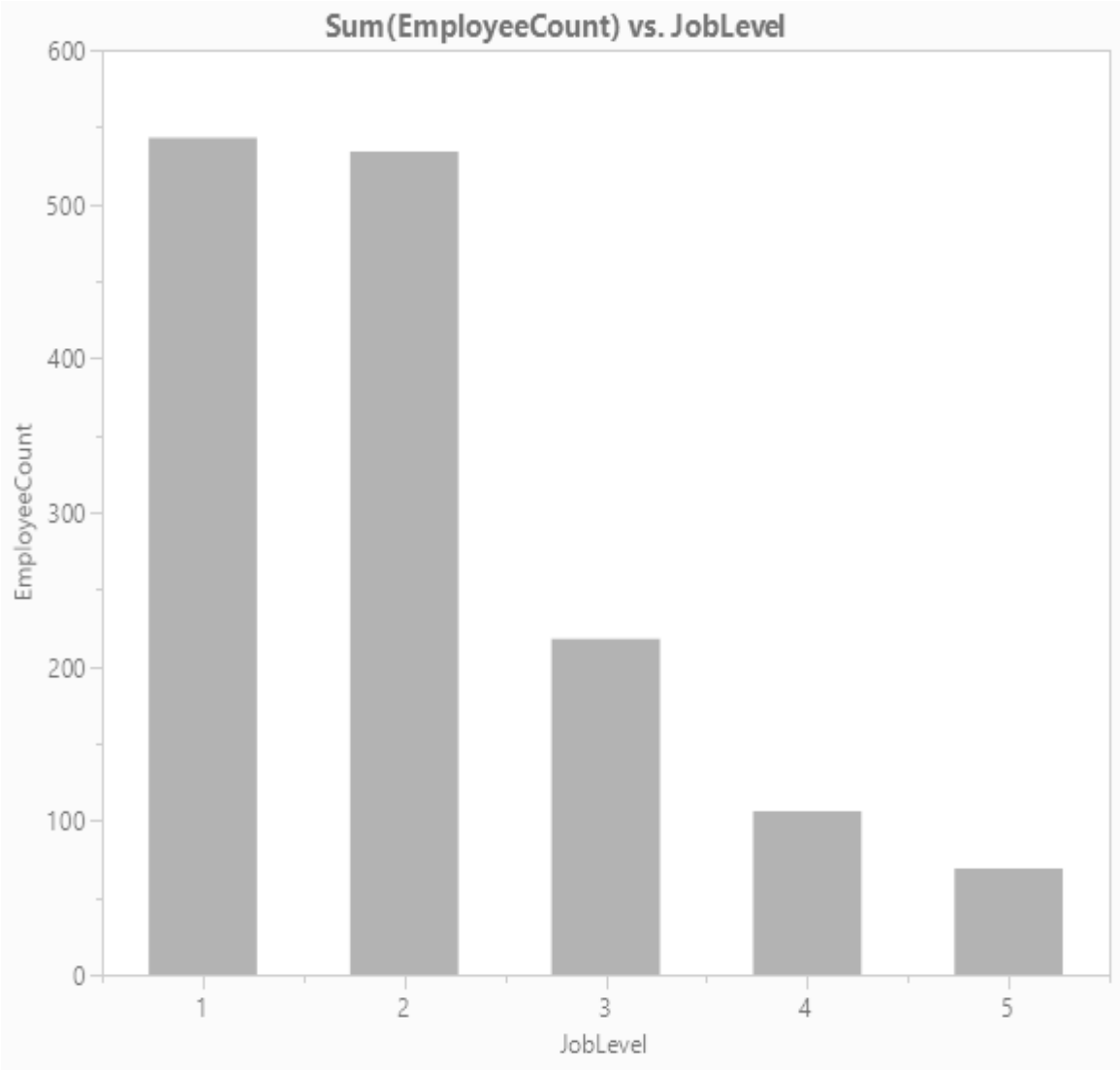
62%
Over 10 years of Experience

63%
Lives within 10 miles

62%
Below the age of 40

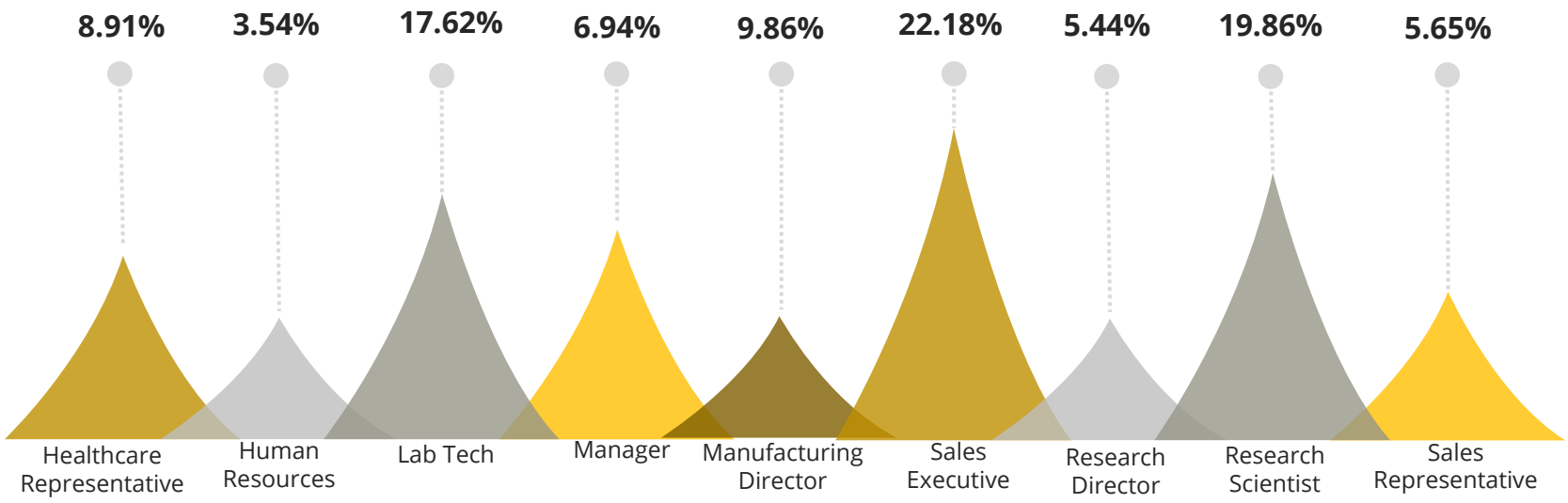
BACKGROUND REPORT

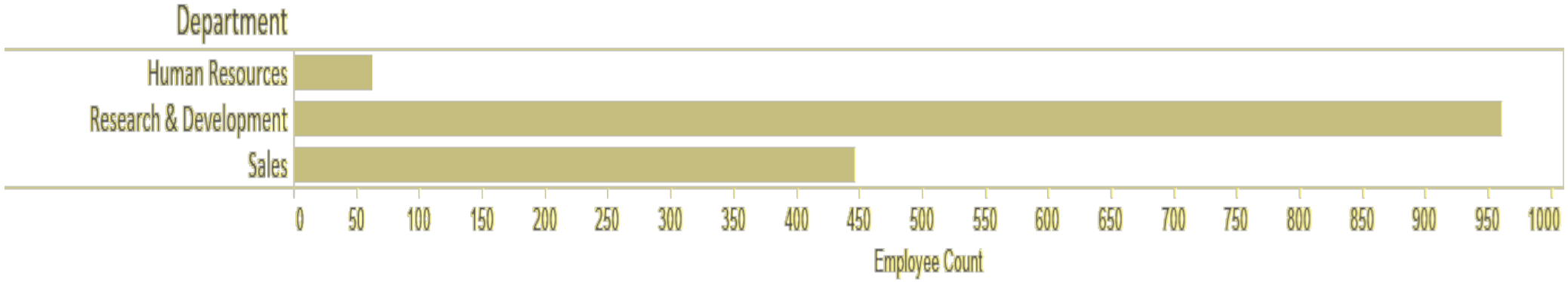


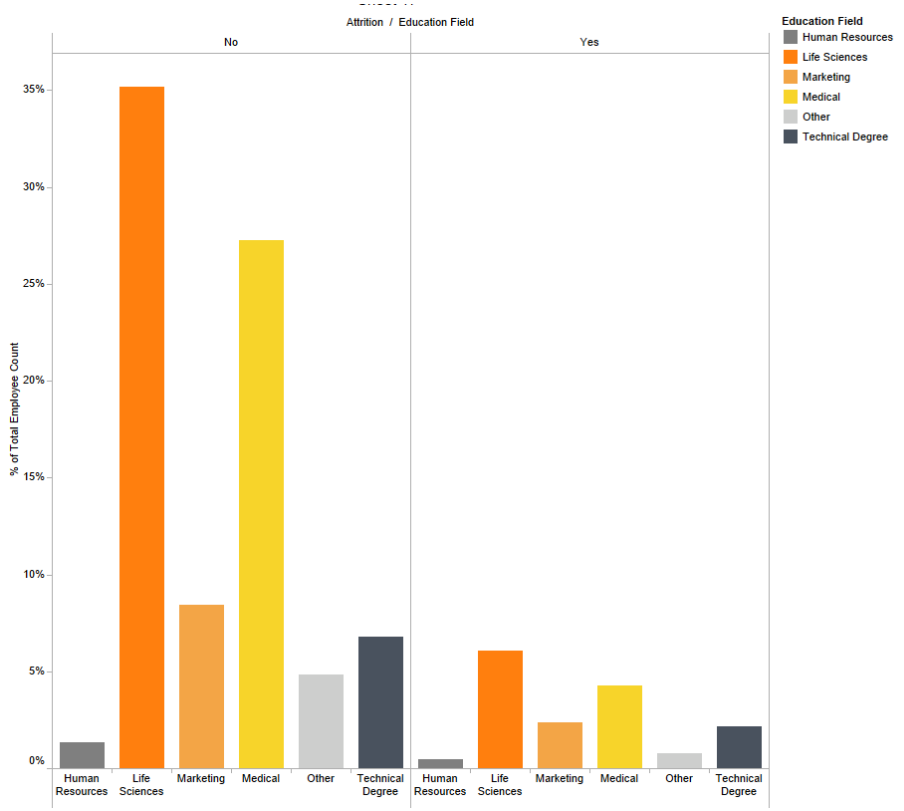


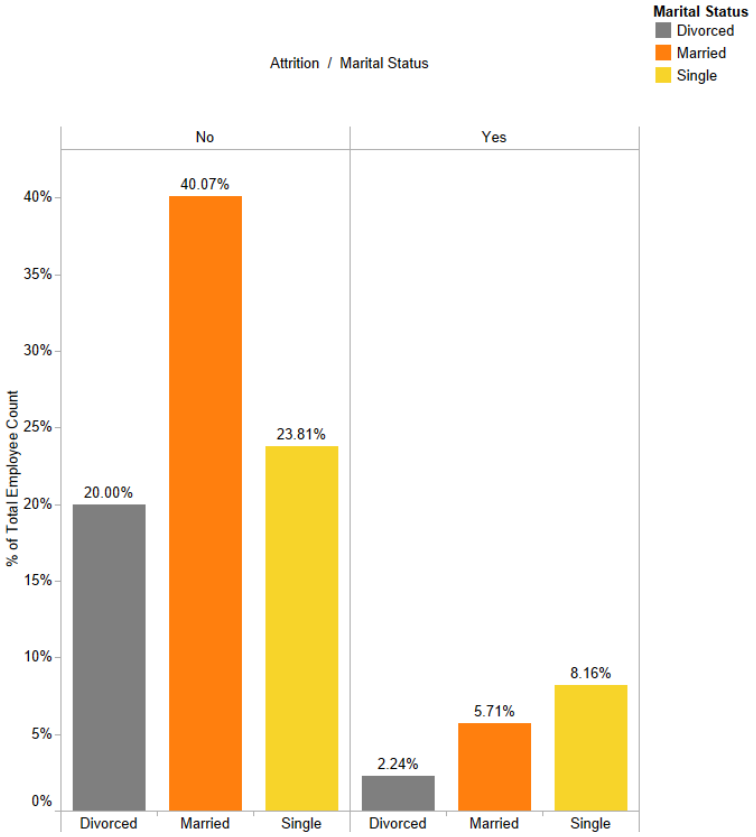
JOB REPORT

Roles and Departments





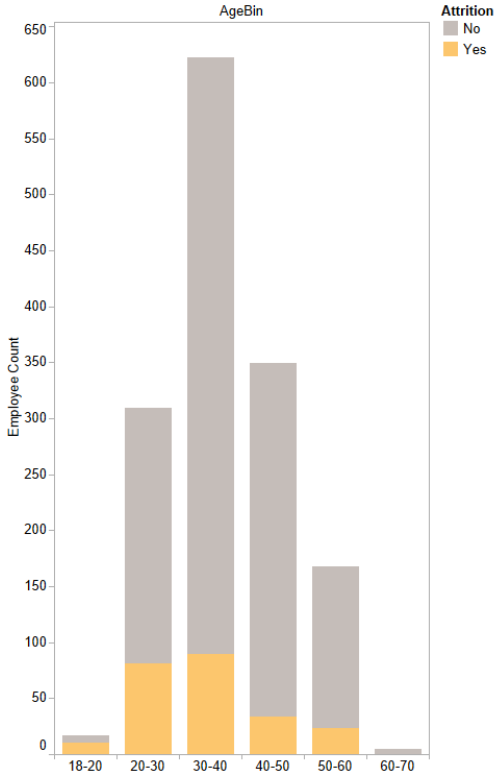




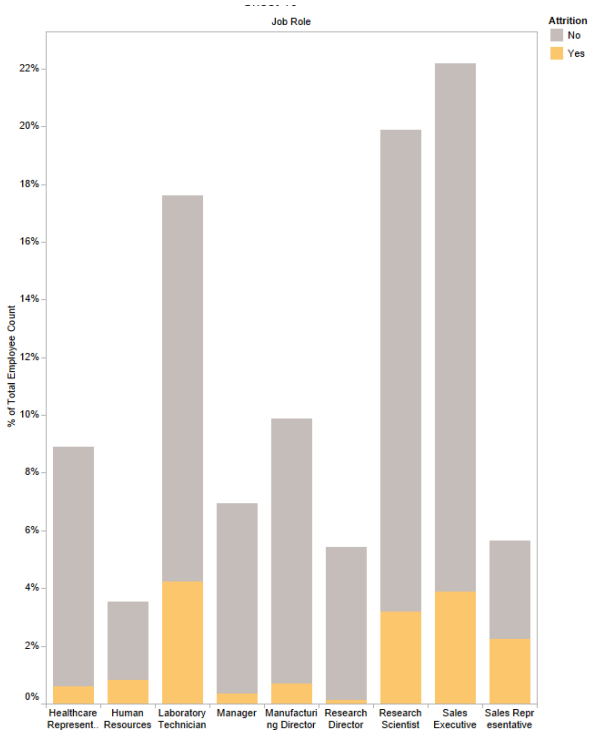
Marital Status & Education Level

Effects on Attrition

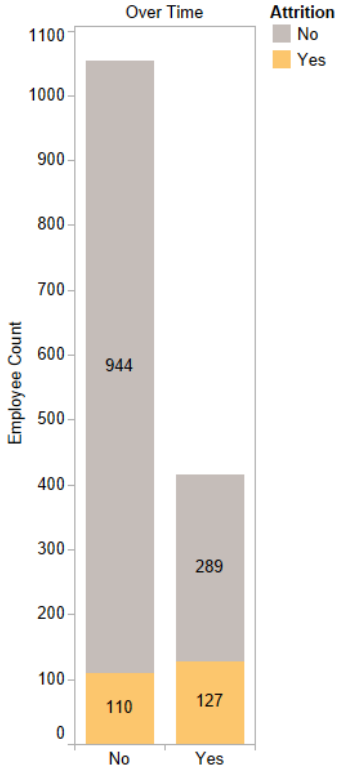
Employee Age and Attrition



Job Role and Attrition



Overtime and Attrition

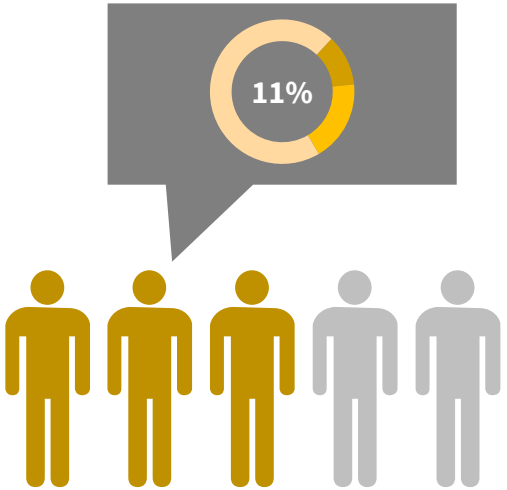


Job Role and Overtime

Effects on Attrition

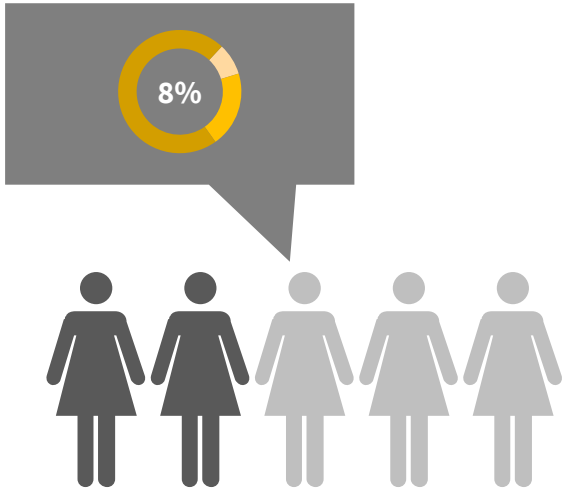
BUSINESS TRAVEL AND ATTRITION

How travelling for business likely to make employees leave the firm



Travel Frequently

		Gender	
Attrition	Business Travel	Female	Male
No	Non-Travel	3.13%	6.26%
	Travel_Frequently	5.92%	8.23%
	Travel_Rarely	25.03%	35.31%
Yes	Non-Travel	0.20%	0.61%
	Travel_Frequently	2.04%	2.65%
	Travel_Rarely	3.67%	6.94%

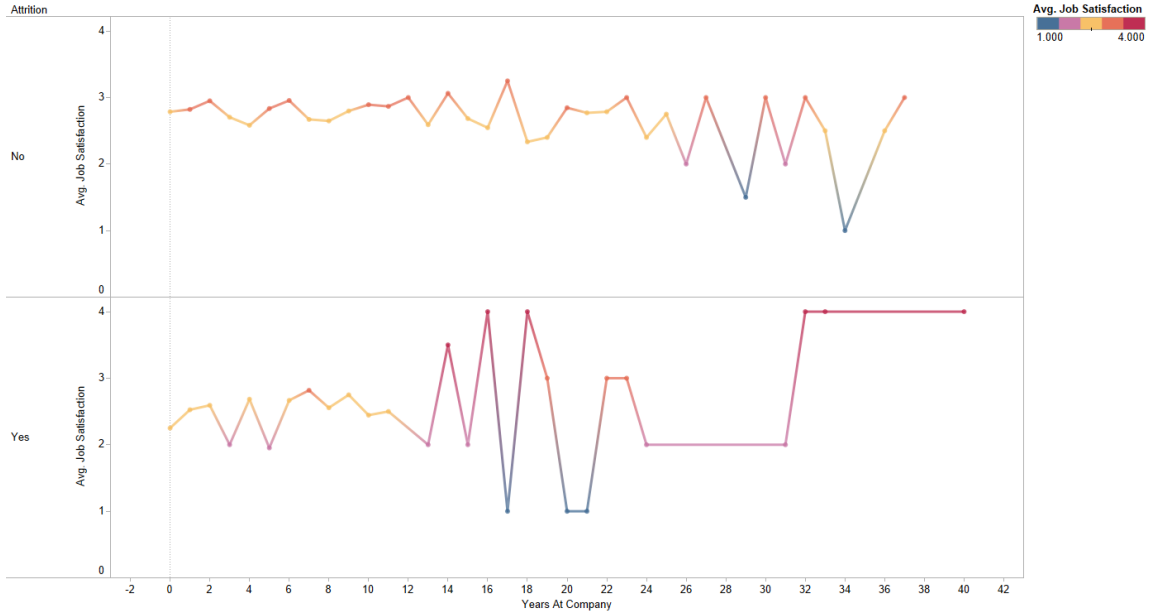


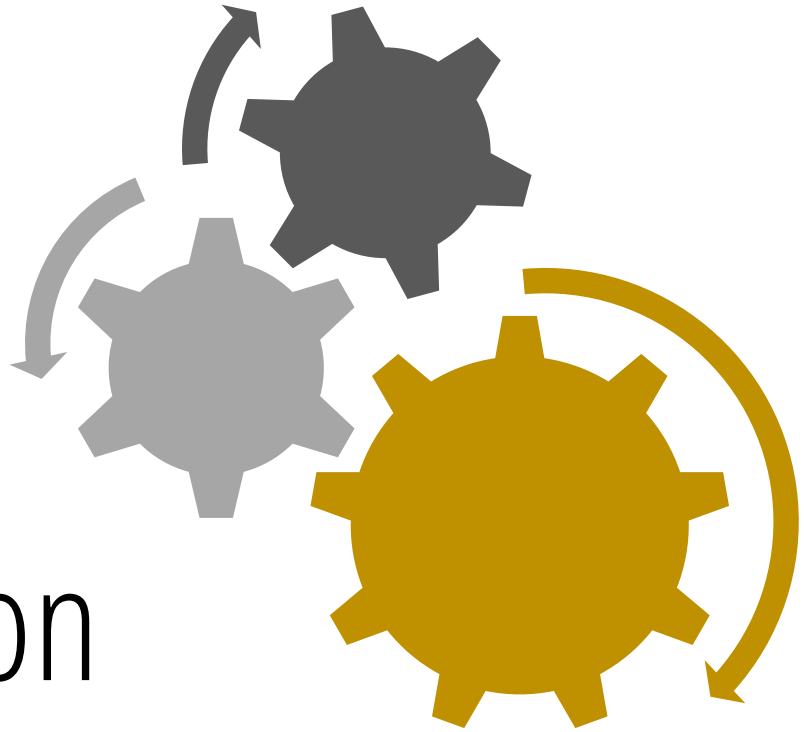
Travel Frequently

Satisfaction and Attrition

Job Satisfaction

Environment Satisfaction





Data Preparation

Identifying the variables and their correlations to determine if we want to use them for our model building.

Data Cleaning

The dataset was analyzed to see if there was any missing data or if there were any outliers that we needed to manage. We found this data set to be very clean with no missing data and no outliers.

Explore Outliers

Quantile Range Outliers

Column	10% Quantile	90% Quantile	Low Threshold	High Threshold	Number of Outliers	Outliers (Count)
EnvironmentSatisfaction	1	4	-8	13	0	
HourlyRate	38	94	-130	262	0	
JobInvolvement	2	4	-4	10	0	
JobLevel	1	4	-8	13	0	
JobSatisfaction	1	4	-8	13	0	
MonthlyIncome	2314.4	13820.4	-32204	48338.4	0	
MonthlyRate	4587	24007.3	-53674	82268.2	0	
NumCompaniesWorked	0	7	-21	28	0	
PercentSalaryHike	11	21	-19	51	0	
PerformanceRating	3	4	0	7	0	
RelationshipSatisfaction	1	4	-8	13	0	
StandardHours	40	40	40	40	0	
StockOptionLevel	0	2	-6	8	0	
TotalWorkingYears	3	23	-57	83	0	
TrainingTimesLastYear	2	5	-7	14	0	
WorkLifeBalance	2	4	-4	10	0	
YearsAtCompany	1	15	-41	57	0	
YearsInCurrentRole	0	9	-27	36	0	
YearsSinceLastPromotion	0	7	-21	28	0	
YearsWithCurrManager	0	9	-27	36	0	

Explore Missing Values

Commands

Missing Value Report — Number of missing values for each column

Missing Value Clustering — Hierarchical clustering of rows and columns missingness

Missing Value Snapshot — Patterns of missing values with graphical map

Imputation disabled because some columns are nominal or ordinal

Missing Columns

☐ Show only columns with missing

Close

Select columns and choose an action.

Select Rows | Color Cells | Exclude Rows | Color Rows

Column	Number Missing
Age	0
DistanceFromHome	0
Education	0
EnvironmentSatisfaction	0
JobInvolvement	0
JobLevel	0
JobSatisfaction	0
MonthlyIncome	0
NumCompaniesWorked	0
PercentSalaryHike	0
PerformanceRating	0
RelationshipSatisfaction	0
StockOptionLevel	0
TotalWorkingYears	0
TrainingTimesLastYear	0
WorkLifeBalance	0
YearsAtCompany	0
YearsInCurrentRole	0
YearsWithCurrManager	0

VARIABLES EXCLUDED

Using statistical process and data exploration various attributes were excluded

EMPLOYEE COUNT

A single incremental value, Excluded from model.

OVER18

Have a single value (Y) for all entries. Excluded from model.

STANDARDHOURS

Have a single value (40). Excluded from model.

EMPLOYEENUMBER

Only an ID, useless for predicting significance. Excluded from model.

SALARY CIRCLES

Mismatch salary rates for monthly, daily and hourly. Only Monthly salary was used in the calculation

MONTHLY SALARY

Has some correlations with attrition and was included in the model building.

DAILY RATE

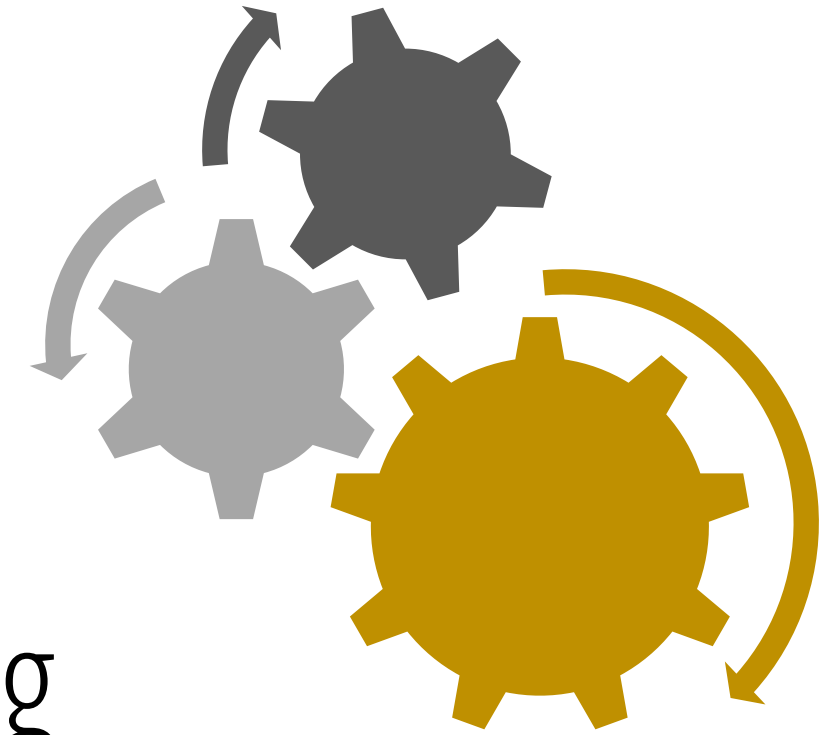
Not significant, excluded from final model building.

MONTHLY RATE

Not significant, excluded from final model building.

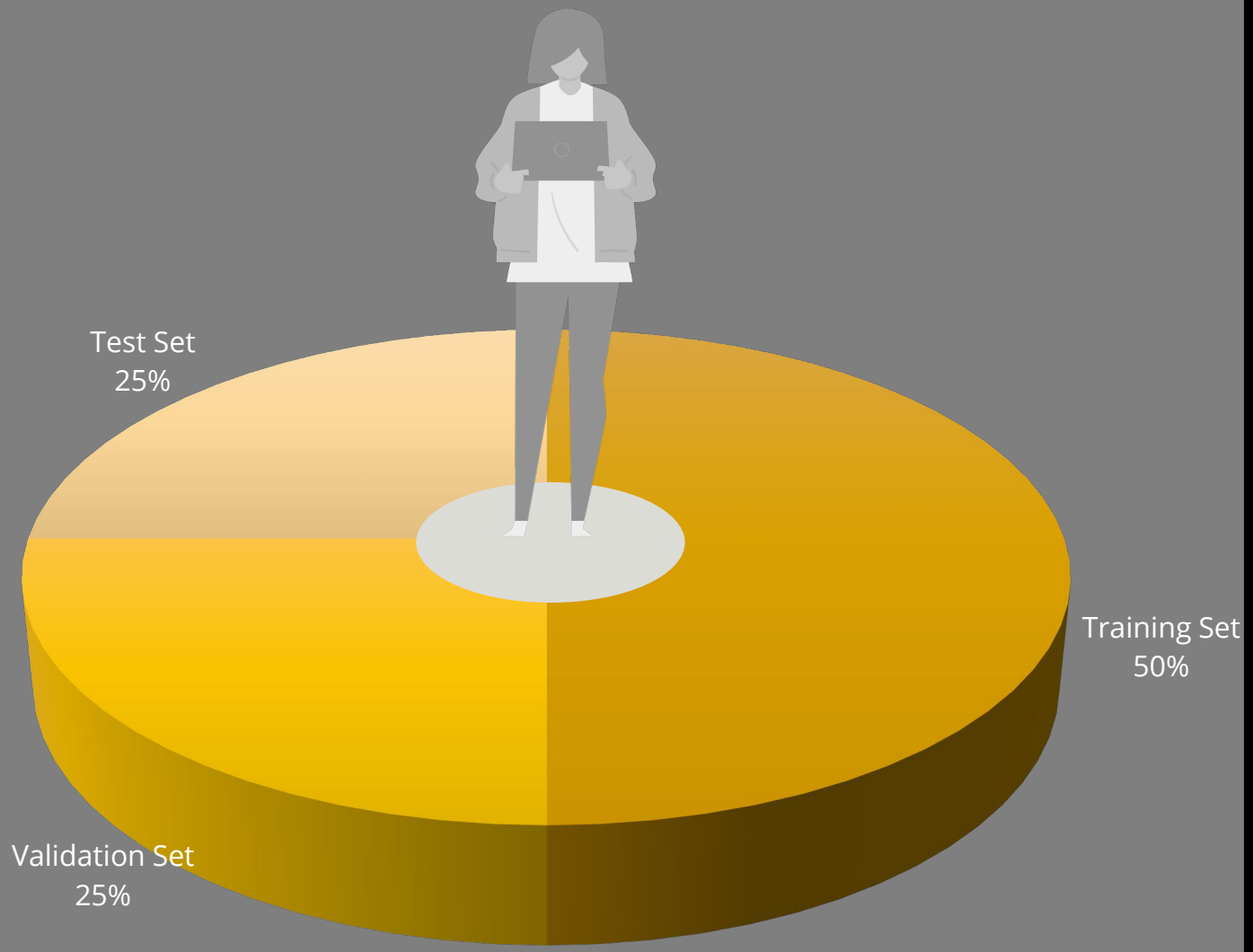
HOURLY RATE

Not significant, excluded from final model building.



Data Modeling

Build Models for prediction analytics with the dependent variable Attrition and other independent variables in the dataset.



PREPARE THE DATASET

Training Set Validation Set Test Set

BUILD MODELS

- Decision Tree
- Bootstrap Forest
- Boosted Tree
- Neural Network
- Logistic Regression

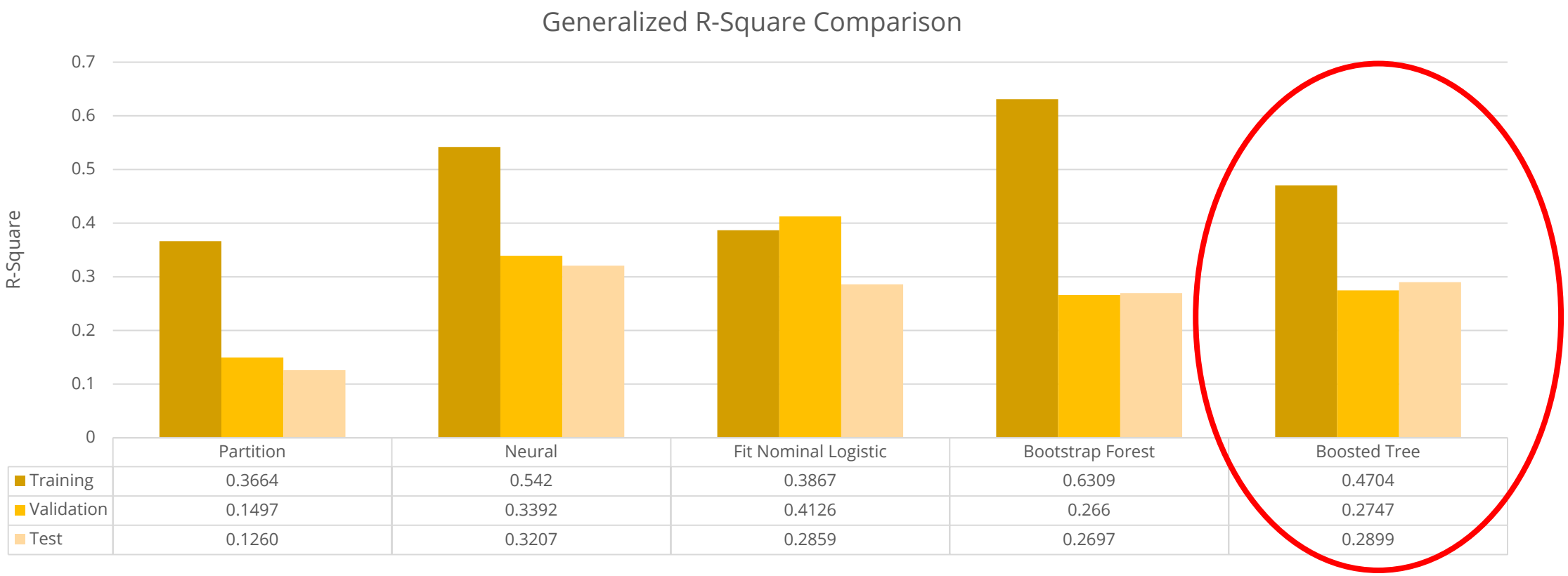
Model Comparison

Measures of Fit for Attrition

Validation	Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RMSE	Mean Abs Dev	Misclassification Rate	N	AUC
Training	Boosted Tree		0.3652	0.4704	0.2813	0.2842	0.1921	0.1145	882	0.9194
Validation	Boosted Tree		0.1939	0.2747	0.3844	0.3431	0.2369	0.1599	294	0.7919
Test	Boosted Tree		0.2178	0.2899	0.3111	0.2991	0.2027	0.1088	294	0.8133
Training	Bootstrap Forest		0.5228	0.6309	0.2115	0.2432	0.1617	0.1066	882	0.9966
Validation	Bootstrap Forest		0.1872	0.2660	0.3876	0.3458	0.2391	0.1735	294	0.8066
Test	Bootstrap Forest		0.2013	0.2697	0.3177	0.3072	0.2085	0.1259	294	0.8301
Training	Fit Nominal Logistic		0.2909	0.3867	0.3142	0.3039	0.1887	0.1190	882	0.8434
Validation	Fit Nominal Logistic		0.3067	0.4126	0.3306	0.3207	0.2070	0.1429	294	0.8590
Test	Fit Nominal Logistic		0.2145	0.2859	0.3124	0.2954	0.1818	0.1156	294	0.8181
Training	Neural		0.4328	0.5420	0.2514	0.2669	0.1504	0.0828	882	0.9003
Validation	Neural		0.2452	0.3392	0.36	0.3254	0.1918	0.1395	294	0.8262
Test	Neural		0.2433	0.3207	0.301	0.2910	0.1668	0.1020	294	0.8304
Training	Partition		0.2736	0.3664	0.3219	0.3095	0.1923	0.1293	882	0.8281
Validation	Partition		0.1012	0.1497	0.4286	0.3611	0.2299	0.1769	294	0.7135
Test	Partition		0.0901	0.1260	0.3619	0.3342	0.2102	0.1701	294	0.7475

R-Square comparison
Boosted Tree

The best comparison is using generalized R-square performance on the test data. But, in this case, there is a lot of degradation in R-Square between the build and test for Neural Network Model and Bootstrap Model. Therefore, overfit and performance degradation are likely an issue for these models. The best model is Boosted tree which has comparable performance but much less degradation between Training and Test.



	Partition	Neural	Fit Nominal Logistic	Bootstrap Forest	Boosted Tree
Training	0.3664	0.542	0.3867	0.6309	0.4704
Validation	0.1497	0.3392	0.4126	0.266	0.2747
Test	0.1260	0.3207	0.2859	0.2697	0.2899

Model Selected

Boosted Tree

Less overfitting

Works well with default parameters

Best accuracy (89.11%)

Confusion Matrix

Training

Actual	Predicted Count	
Attrition	No	Yes
No	739	0
Yes	101	42

Validation

Actual	Predicted Count	
Attrition	No	Yes
No	237	3
Yes	44	10

Test

Actual	Predicted Count	
Attrition	No	Yes
No	253	1
Yes	31	9

Model Findings

Key factors driving the attrition rate based on model findings

Key drivers

Based off the column contributions in the Boosted Tree Model

Column Contributions

Term	Number of Splits	G^2	Portion
OverTime	16	11505.6897	0.1699
BusinessTravel	11	7176.48388	0.1060
JobLevel	10	6035.43859	0.0891
WorkLifeBalance	9	5996.38462	0.0886
StockOptionLevel	10	5114.61821	0.0755
Department	8	5075.97903	0.0750
YearsWithCurrManager	5	4214.03963	0.0622
JobSatisfaction	13	4155.32247	0.0614
EnvironmentSatisfaction	8	3114.19456	0.0460
JobInvolvement	12	2335.67325	0.0345
YearsAtCompany	4	1825.38505	0.0270
DistanceFromHome	8	1725.57534	0.0255
MonthlyIncome	7	1509.15889	0.0223
EducationField	2	1441.46461	0.0213
TotalWorkingYears	1	1356.72093	0.0200
RelationshipSatisfaction	4	1179.83455	0.0174
NumCompaniesWorked	4	939.614507	0.0139
YearsInCurrentRole	5	839.749912	0.0124
Age	4	744.555319	0.0110
JobRole	1	574.650806	0.0085
PercentSalaryHike	2	396.985292	0.0059
YearsSinceLastPromotion	1	255.350203	0.0038
Education	2	150.839568	0.0022
Gender	1	17.878653	0.0003
PerformanceRating	1	14.076325	0.0002
TrainingTimesLastYear	1	6.4645108	0.0001
MaritalStatus	0	0	0.0000

Top 5 contributors to employee attrition

- Overtime
- Business travel
- Job level
- Work life balance
- Stock options

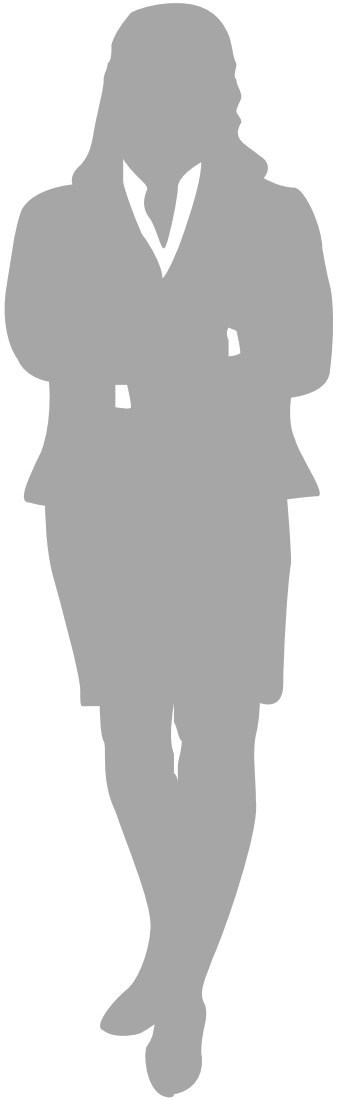
Functional Turnover

- Turnover in some organizations can be viewed as functional and beneficial to the organization. There are two functional turnover metrics in this data set: Attrition of total working years and older employees.
- An additional metric of functional turnover is of higher income earners. This gives an opportunity to add lower-paid, potentially younger employees to work force, reducing overall costs for the company.

Recommendations

Strategies to help reduce attrition and retain your most valuable talent

Actions to reduce turnover in the future

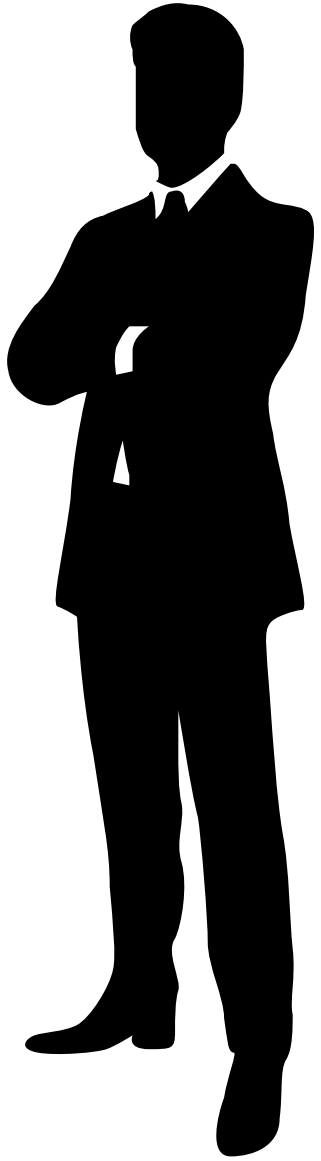


REDUCE OVERTIME
Optimize the value chain to reduce overtime.

BUSINESS TRAVEL
Ensure travel is business critical. Manage employee expectations with travel.

PROVIDE INCENTIVES
Tailor incentive packages to employees' unique needs

MORE VESTED BENEFITS
Implement vested benefits to decrease attrition of newer employees



Limitation and Next Steps

Limitations

- No benchmarking with similar companies
- Outsides factors like state of economy unknown when data was taken
- Some data appears to be random: daily rate, monthly rate and weekly rate.

Next steps

- Further statistical analysis on individual departments (specifically Sales, Lab Techs, and HR)
- Analysis of overtime reduction and value chain optimization
- Build risk models using clustering algorithms to identify key employees to retained

The goal: be proactive and manage attrition before it negatively impacts the business