

A decorative border at the top of the slide consists of a grid of squares in various shades of blue and cyan. The squares are arranged in a pattern that is roughly 10 squares wide and 4 squares high, though some squares are missing, creating a fragmented look. The colors range from a deep navy blue to a bright cyan.

HR Department:

Employee Attrition and its Causes

TABLE OF CONTENTS

01

Introduction

Libraries,
Dataset and
Data processing

02

Visualization

Suggested
Plots and EDA

03

Statistics

Summary
Statistics and
Correlation
Analysis

04

Predictive

Challenges and
Opportunities

A decorative pattern of blue squares of various sizes is located in the top corners of the slide, set against a dark blue background.

01

Introduction

Dataset and Data processing

A series of horizontal bars in different shades of blue and cyan are positioned at the bottom of the slide, creating a modern, geometric footer design.

Libraries



- Data storage and processing



- Visualization and Plotting



- Model generation and evaluation



- Data transformation and modelling

Dataset

[Workshop-Shopee-machine-learning...](#)

Name	Description
AGE	Numerical Value - Age of the employee
ATTRITION	Employee leaving the company (NO, YES)
BUSINESS TRAVEL	Business travel frequency
DAILY RATE	Numerical Value - Salary Level
DEPARTMENT	Department of the employee
DISTANCE FROM HOME	Numerical Value - THE DISTANCE FROM WORK TO HOME
EDUCATION	Numerical Value - 1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor'
EDUCATION FIELD	Education filed of the employee
EMPLOYEE COUNT	Numerical Value
EMPLOYEE NUMBER	Numerical Value - EMPLOYEE ID
ENVIROMENT SATISFACTION	Numerical Value - SATISFACTION WITH THE ENVIROMENT (1 'Low' 2 'Medium' 3 'High' 4 'Very High')
GENDER	Gender of the employee
HOURLY RATE	Numerical Value - HOURLY SALARY
JOB INVOLVEMENT	Numerical Value - JOB INVOLVEMENT (1 'Low' 2 'Medium' 3 'High' 4 'Very High')
JOB LEVEL	Numerical Value - LEVEL OF JOB (1 'Junior', 2 'Senior', 3 'Manager', 4 'Senior Manager', 5 'Director')
JOB ROLE	Job role
JOB SATISFACTION	Numerical Value - SATISFACTION WITH THE JOB (1 'Low' 2 'Medium' 3 'High' 4 'Very High')
MARITAL STATUS	Marital status
MONTHLY INCOME	Numerical Value - MONTHLY SALARY
MONTHY RATE	Numerical Value - MONTHY RATE
NUMCOMPANIES WORKED	Numerical Value - NO. OF COMPANIES WORKED AT

- Features 1370 HR files with 35 descriptive and categorial variables of individual anonymous employees of a single undisclosed company
- Includes metadata for each column in a separate documentation
- Is generally well structured, exhibiting no signs of artifacts or inconsistencies
- Can be found on kaggle.com together with the original uploaders more ML driven EDA

Processing

Validation and processing steps included:

- NA Handling: None found
- Empty Values: None found
- Duplicate Entries, especially UUIDs: None found
- Elimination of columns with no variance: 3 Variables
- Transformation of all categorical variables into descriptive dummies

A decorative pattern of light blue squares of various sizes is scattered across the top corners of the slide, set against a dark blue background.

02

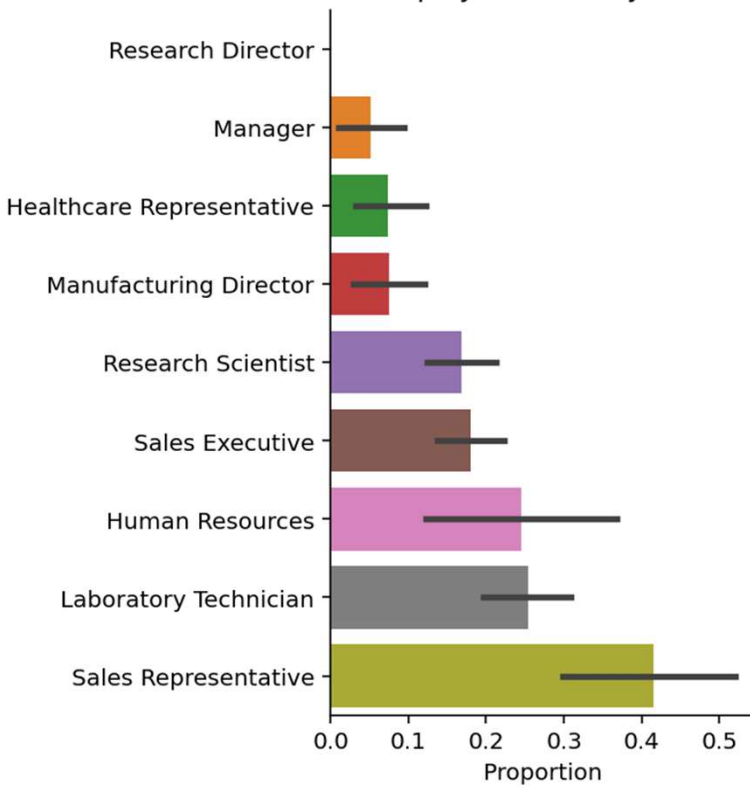
Visualization

Suggested Plots and EDA

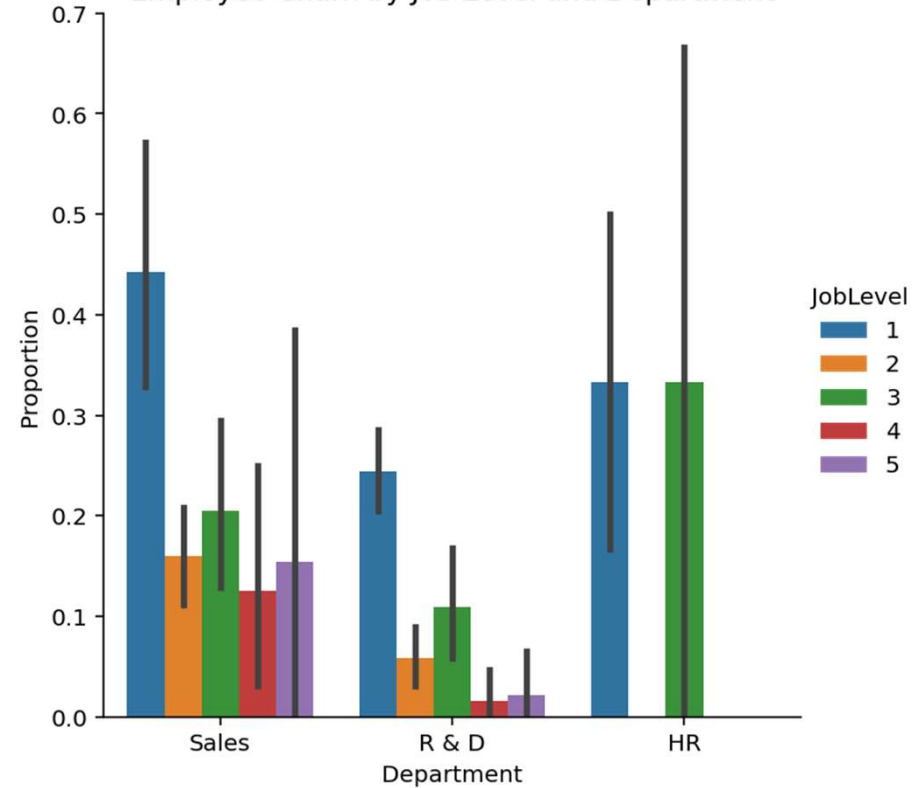
A series of horizontal bars in different shades of blue and cyan are positioned along the bottom edge of the slide, creating a modern, abstract footer design.

Role and Level

Employee Churn by Role

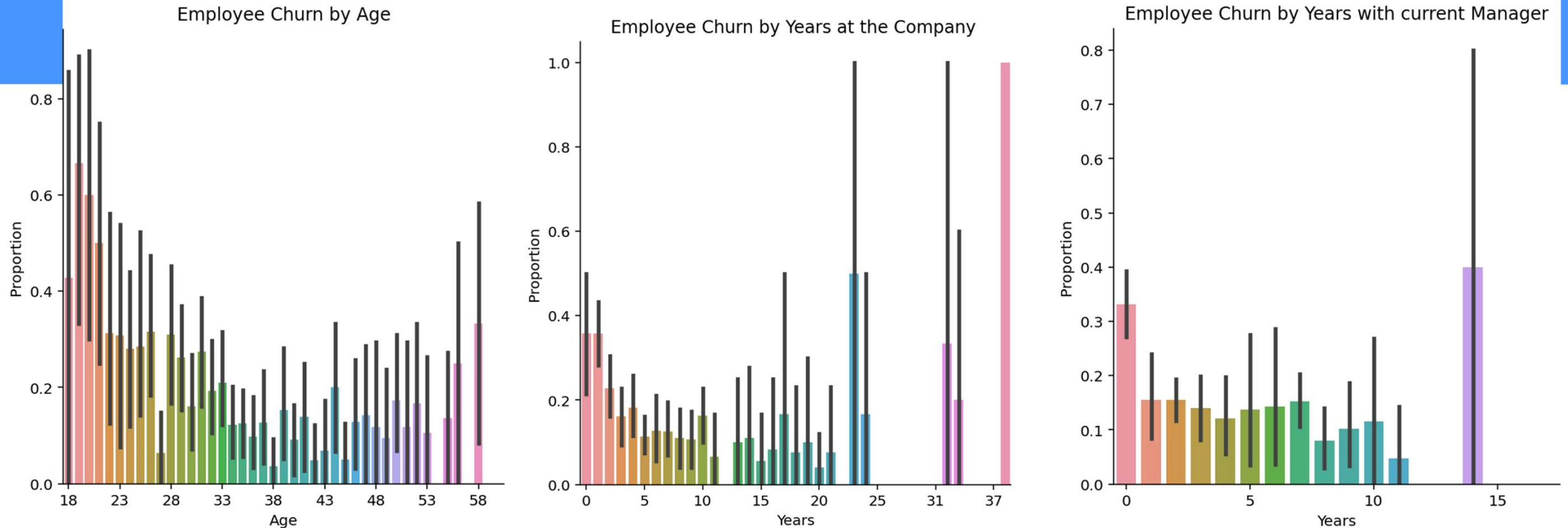


Employee Churn by Job Level and Department



Junior employees have the highest churn rates but outward facing departments generally face increased attrition

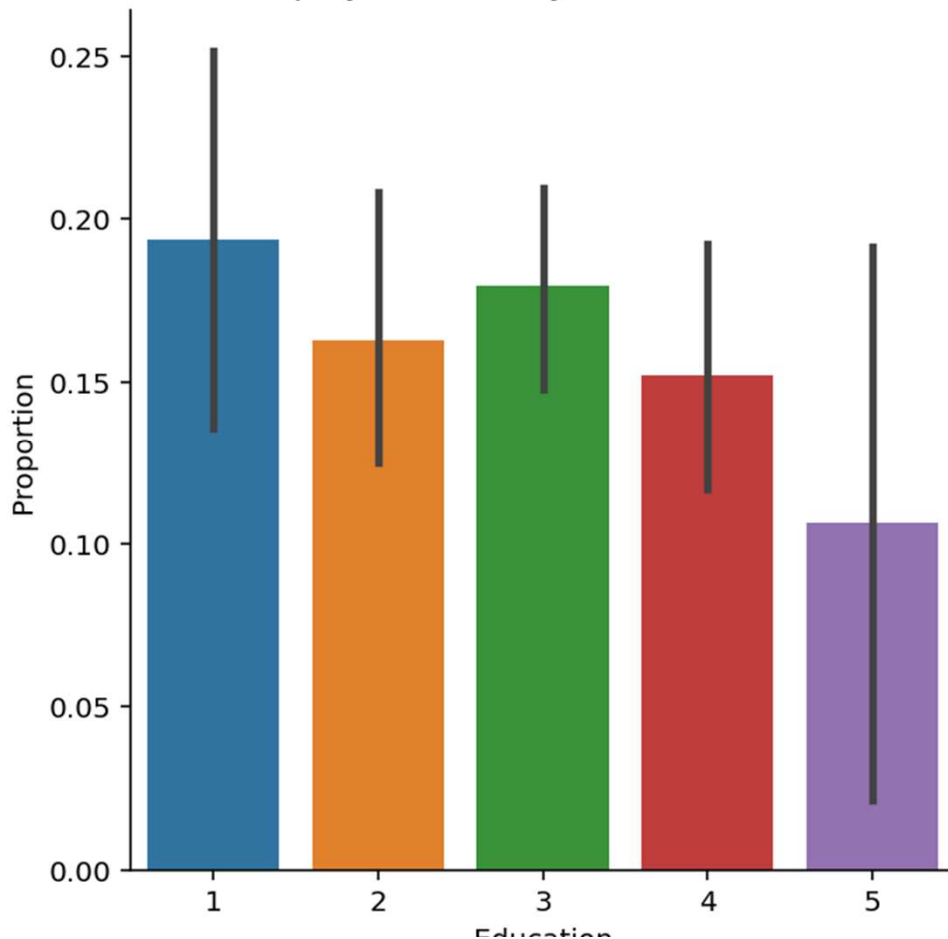
Age and Structures



- Plot 1 shows a relative decrease in churn by age until early retirement age
- But plot 2 strongly indicates a company pension scheme
- Lastly plot 3 highlights the effect of consistent team and management structures to reduce attrition

Education

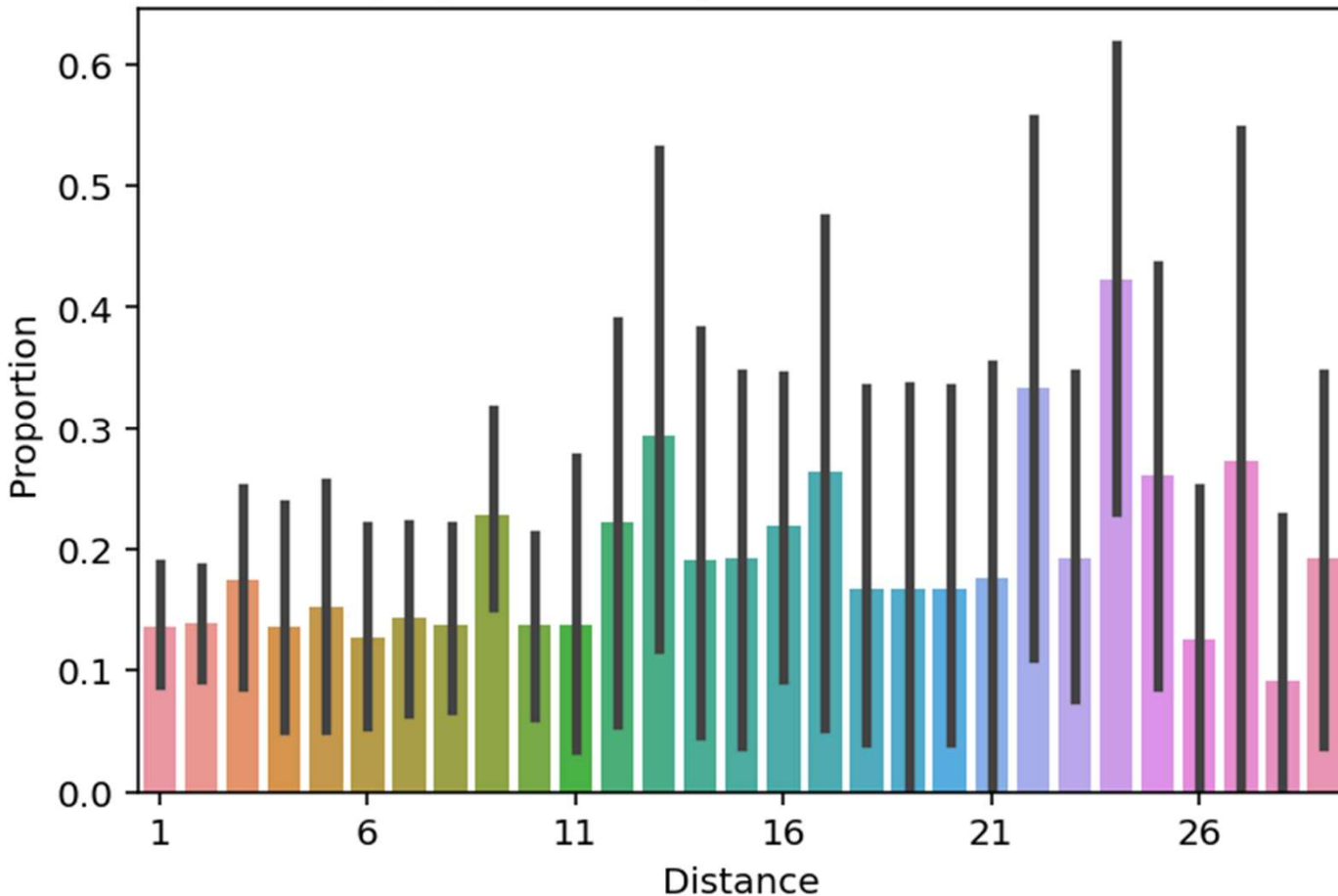
Employee Churn by Education Level



- While the churn rate decreases in education level
- Our confidence intervals increase as the sample sizes for each additional degree grow smaller

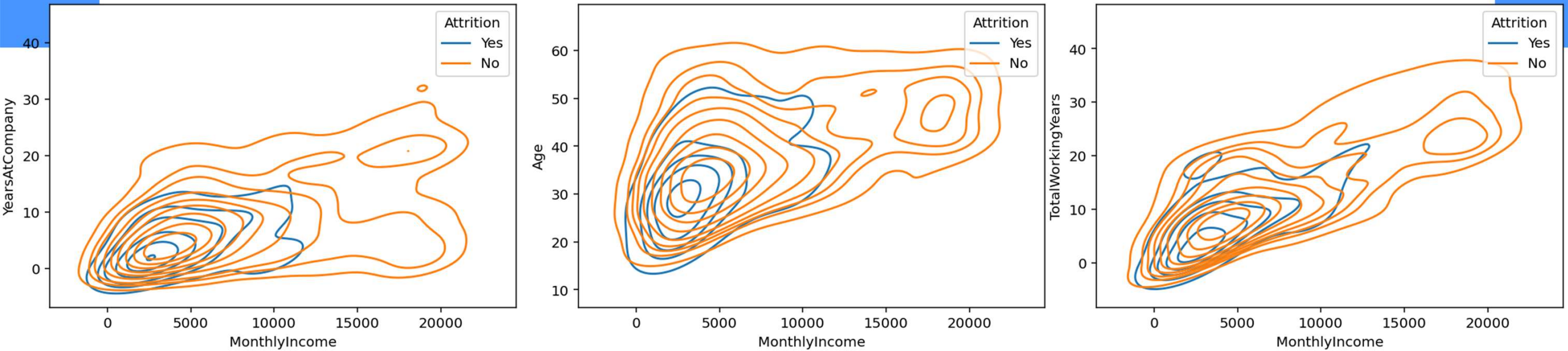
Commute

Employee Churn by Distance from Home



- Churn increases slightly in distance
- Regular peaks indicate changes in mode of travel, type of road and geographical features
- But variance explodes as sample sizes decrease

Income



- Employees are less likely or willing to churn as age or time at the company increases
- But attrition is still far more likely if income is on the lower end of the age~pay scale



03

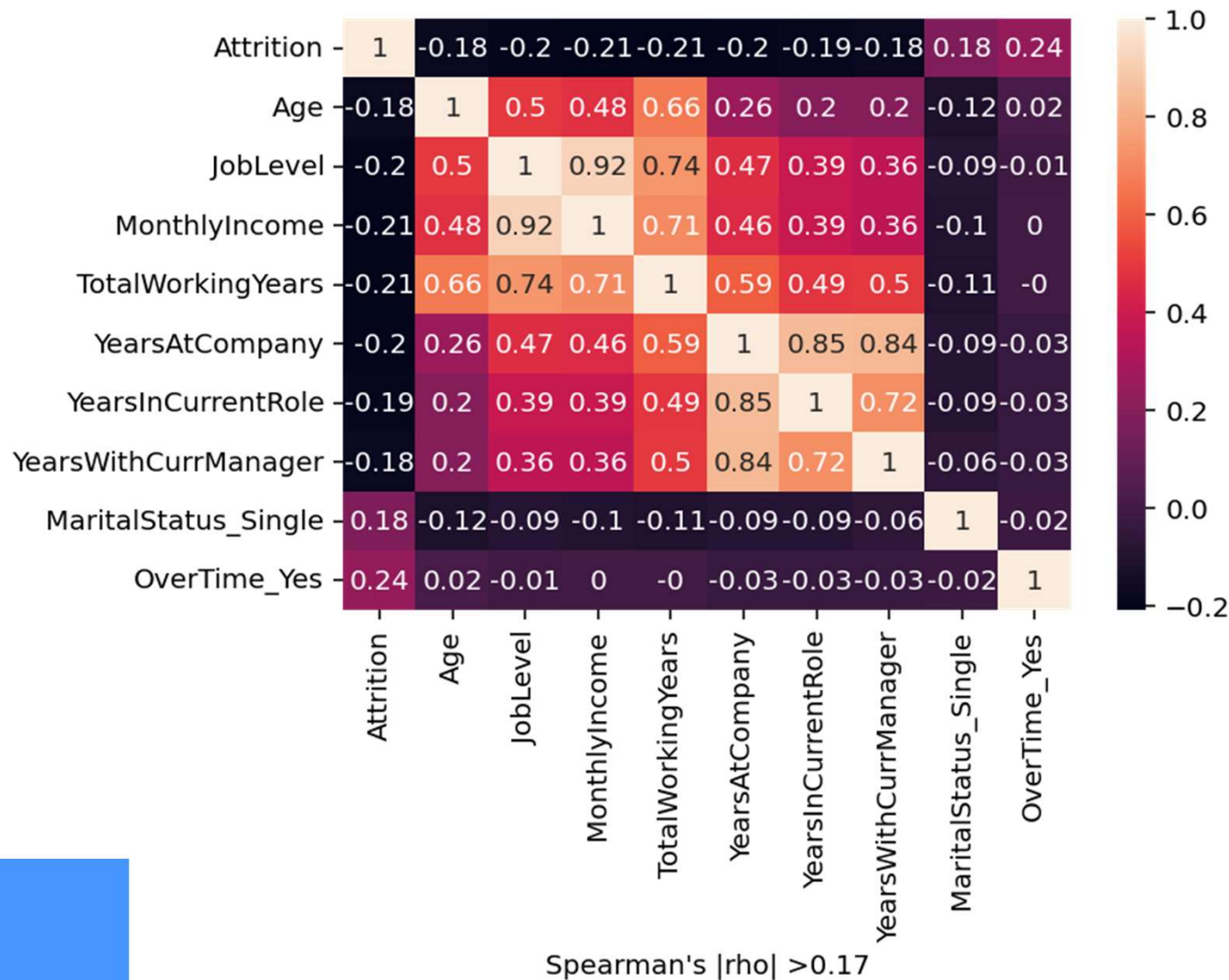
Statistics

Summary Statistics and Correlation Analysis

Summary Statistics

	mean	std	min	50%	max
Age	36.9	9.2	18.0	36.0	60.0
DailyRate	803.9	402.8	102.0	806.5	1499.0
DistanceFromHome	9.3	8.1	1.0	7.0	29.0
Education	2.9	1.0	1.0	3.0	5.0
EmployeeNumber	1022.7	603.0	1.0	1016.5	2068.0
EnvironmentSatisfaction	2.7	1.1	1.0	3.0	4.0
HourlyRate	65.8	20.3	30.0	66.0	100.0
JobInvolvement	2.7	0.7	1.0	3.0	4.0
JobLevel	2.1	1.1	1.0	2.0	5.0
JobSatisfaction	2.7	1.1	1.0	3.0	4.0
MonthlyIncome	6521.1	4716.0	1009.0	4933.0	19999.0
MonthlyRate	14284.8	7115.6	2094.0	14225.5	26999.0
NumCompaniesWorked	2.7	2.5	0.0	2.0	9.0
PercentSalaryHike	15.2	3.7	11.0	14.0	25.0
PerformanceRating	3.2	0.4	3.0	3.0	4.0
RelationshipSatisfaction	2.7	1.1	1.0	3.0	4.0
StockOptionLevel	0.8	0.8	0.0	1.0	3.0
TotalWorkingYears	11.3	7.8	0.0	10.0	40.0
TrainingTimesLastYear	2.8	1.3	0.0	3.0	6.0

Correlation Analysis



- Overtime has the strongest impact
- But generally, factors associated with youth and mobility, like unmarried, low job level etc. also noteworthy
- However, both soft, like consistency and hard factors, like payment in the top 10 causes

The background is a solid dark blue. It is decorated with several squares of different sizes and shades of blue and cyan. Some squares are light cyan, some are medium blue, and some are a darker blue. They are scattered across the slide, with a higher concentration on the left and right sides. A large white square is centered in the upper half of the slide, containing the text '04'.

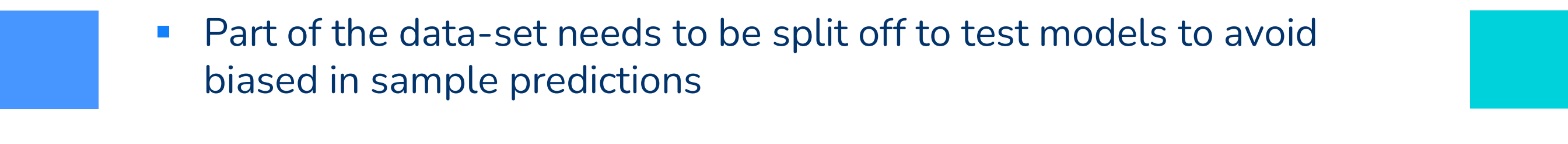
04

Predictive

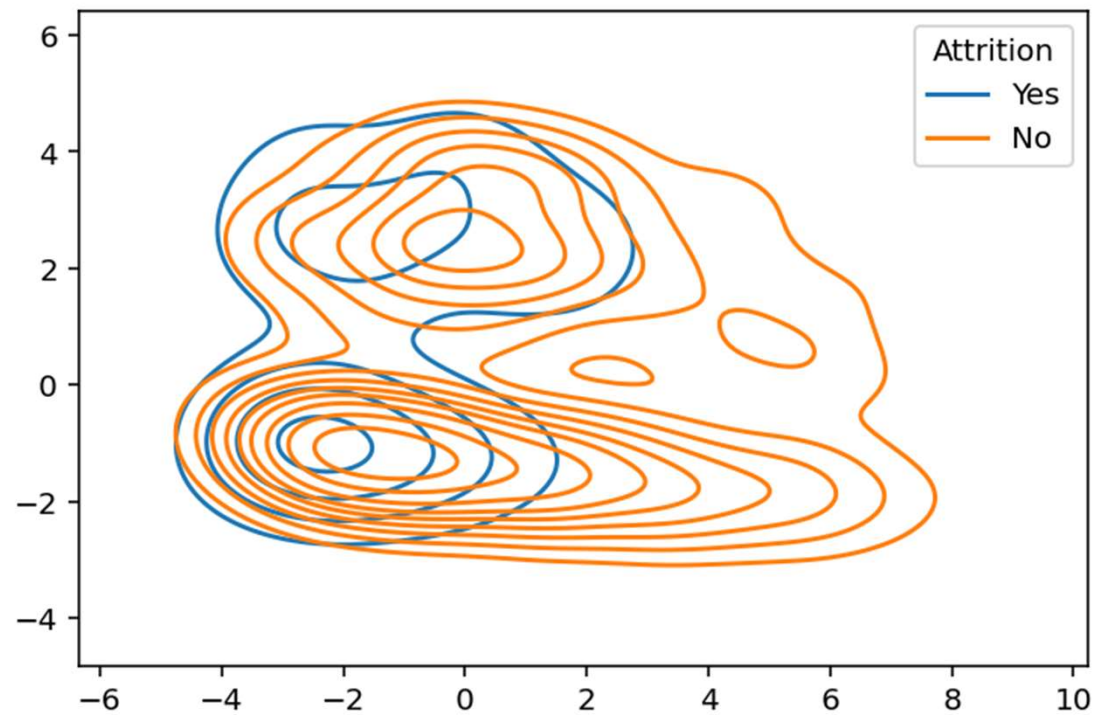
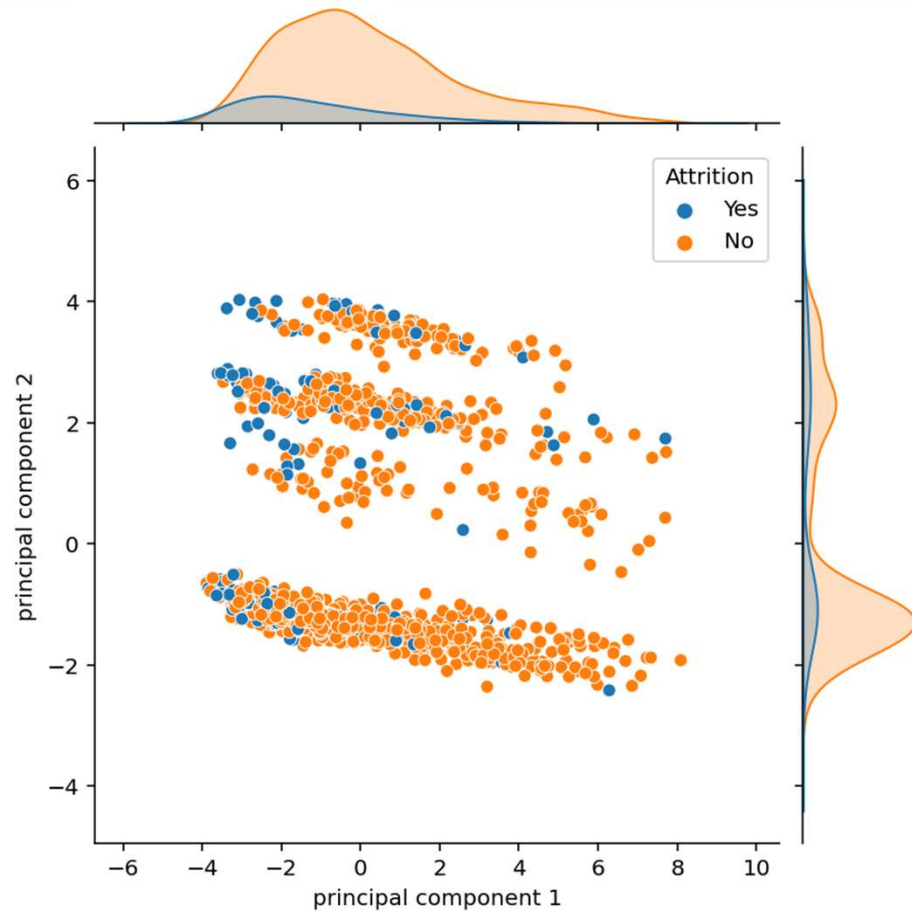
Challenges and Opportunities



Challenges

- High dimensionality hampers intuitive and more hands-on approaches
 - Endogeneity, cross-interactions and causality likely to cause issues and require additional care in analysis
 - The dataset is heavily imbalanced with only 16.8% of observations churning
 - Some more advanced ML algorithms require additional transformation of the data
 - Part of the data-set needs to be split off to test models to avoid biased in sample predictions
- 

Opportunities



- PCA: Enables us to visualize high dimensional data while maintaining a significant share of the original variance

Opportunities

	Coef.	Std.Err.
const	0.630008	0.076038
DistanceFromHome	0.003640	0.001086
EnvironmentSatisfaction	-0.043793	0.008052
JobInvolvement	-0.061574	0.012326
JobSatisfaction	-0.039277	0.008025
NumCompaniesWorked	0.016831	0.003894
RelationshipSatisfaction	-0.023538	0.008203
TotalWorkingYears	-0.007325	0.001630
TrainingTimesLastYear	-0.015804	0.006848
WorkLifeBalance	-0.028556	0.012642
YearsAtCompany	0.006581	0.003043
YearsInCurrentRole	-0.010343	0.003972
YearsSinceLastPromotion	0.009773	0.003523
YearsWithCurrManager	-0.009163	0.004085
BusinessTravel_Travel_Frequently	0.164672	0.034457
BusinessTravel_Travel_Rarely	0.072633	0.029854
EducationField_Human Resources	0.162593	0.065499
EducationField_Technical Degree	0.087606	0.030579
Gender_Male	0.041157	0.018127
JobRole_Laboratory Technician	0.114158	0.025394
JobRole_Sales Executive	0.067513	0.022609
JobRole_Sales Representative	0.228171	0.040819
MaritalStatus_Divorced	-0.150824	0.024570
MaritalStatus_Married	-0.124815	0.020562
OverTime_Yes	0.212079	0.019684

- Linear probabilities enable intuitive interpretation and explanation
- But consistent variance estimation is not possible
- P-value driven exploration is therefore of doubtful benefit
- Can however still perform well if benchmarked

A decorative border composed of a grid of squares in various shades of blue and cyan, framing the central text area.

THANK YOU

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Stories**