# HR Churn - Eindwerk Data Science 2023-24 - Raf Ledeganck

# Setting the stage

### ## Imports

### # Program switches

### ## Functions

# ## Data 🡪 Data loading

# ## Data Exploration

#### Distinct values per (non-numeric) column

## Data preparation

## ## Data cleaning

## ## Data Transformation (1)

# Feature reduction

## ## Check data redundancy 🡪 Feature reduction - Filter

### ### Check for quasi-constant features

### ### Check feature variance

### ### Check feature correlation with target

### ### Check correlation between bonus details and total

### ### Check mutual information

## Feature reduction - Embedded

### # Data transformation

### ### Principal Components 🡪 **???**

### ### Random Forest

### ### Gradient boosting

### ## Logistic Regression

# # Predictions

## ## Data preparations

### ### Check if dataset is balanced

### ### Train-test split

### ### Scaling & transformation

## Choice of models 🡪 Reduced dataset

### ### Random forest

### ### Support Vector Machines

### ### Logistic Regression

## Full dataset

### ### Neural Network

## ## ~~Retry with~~ balanced dataset

### ### Random forest

### ### Support Vector Machines

### ### Logistic Regression

# Further data analysis

## ## Check where churn is higher than average

# 2nd dataset

## # Data cleaning ~~on 2nd dataset~~