IBM HR ANALYTICS

Jasper Hu

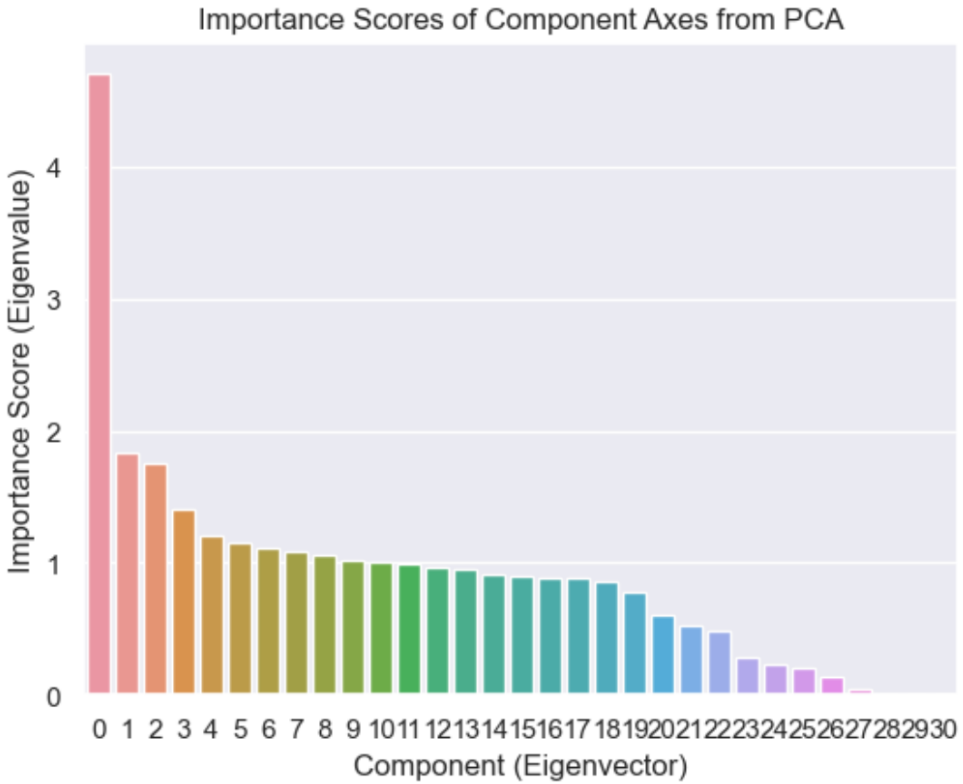
wh169

Central Questions: Reducing attrition & notable relationships between variables

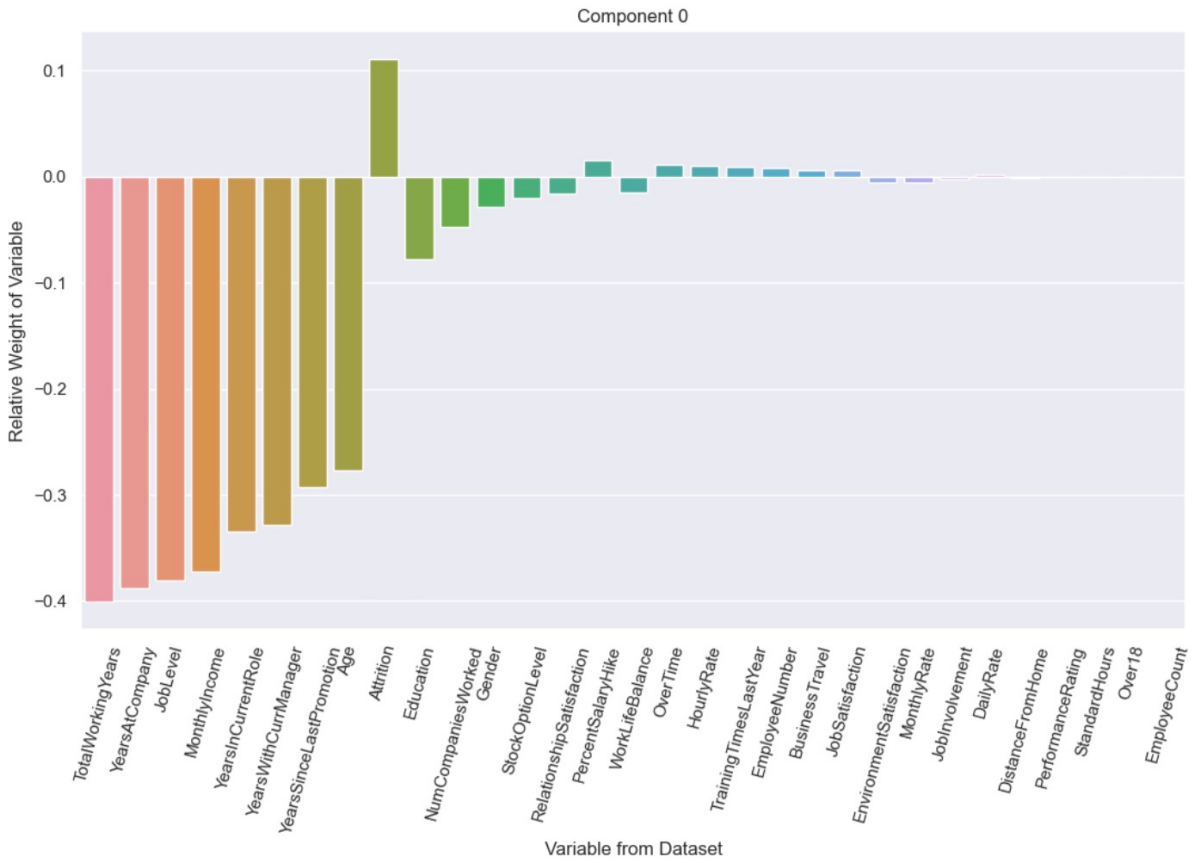
Class of 2026

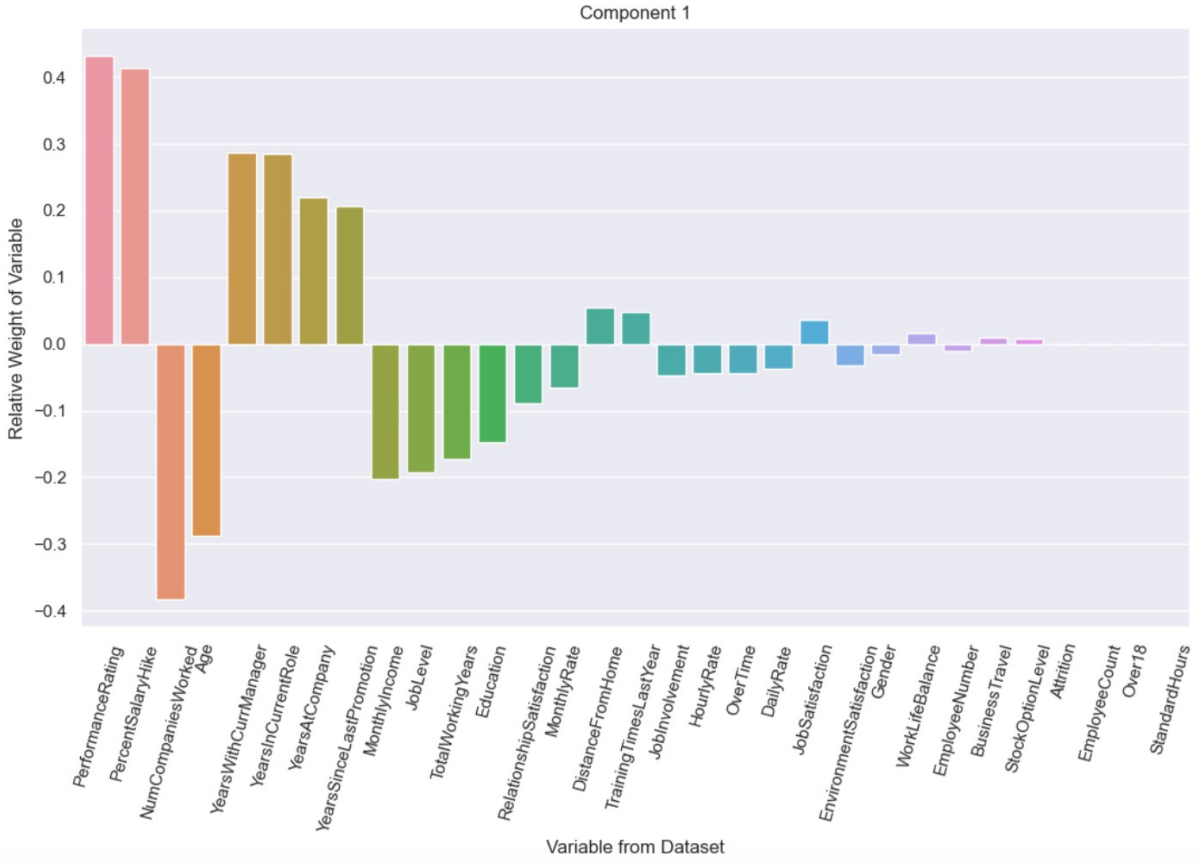
PCA

- Cleaned dataset, normalized, and performed Principle Component Analysis
  - Centered around 0, Standard Deviation at 1
- Illustrates significant covariances/relationships between variables on component axes



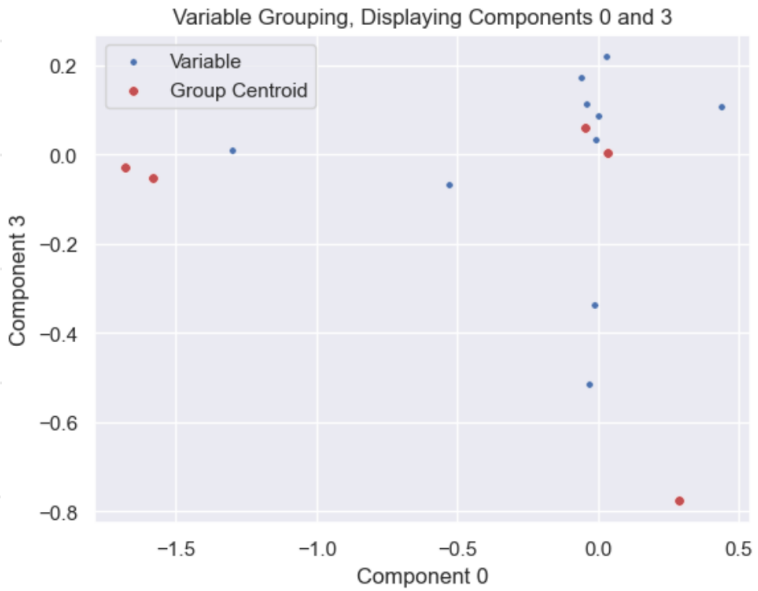
Some Notable Components from PCA

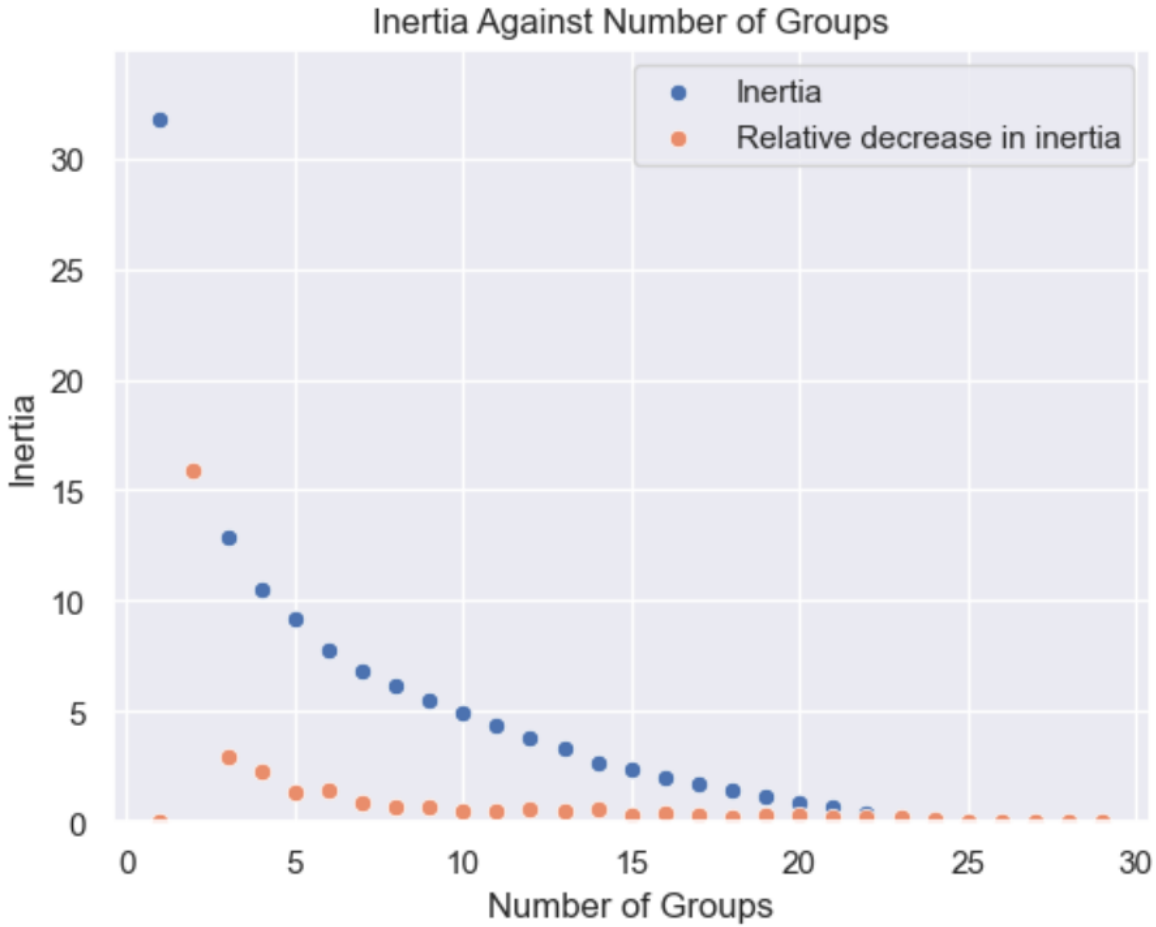




Grouping Variables

- Apply K-Means algorithm
- Group variables with similar representations on the 10 most significant components
- Number of groups?
  - Minimize “inertia”
  - Not too fragmented
- Number of Groups = 5

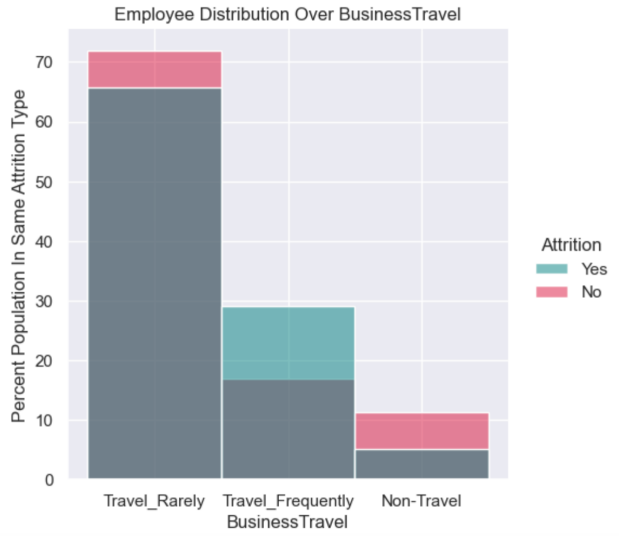


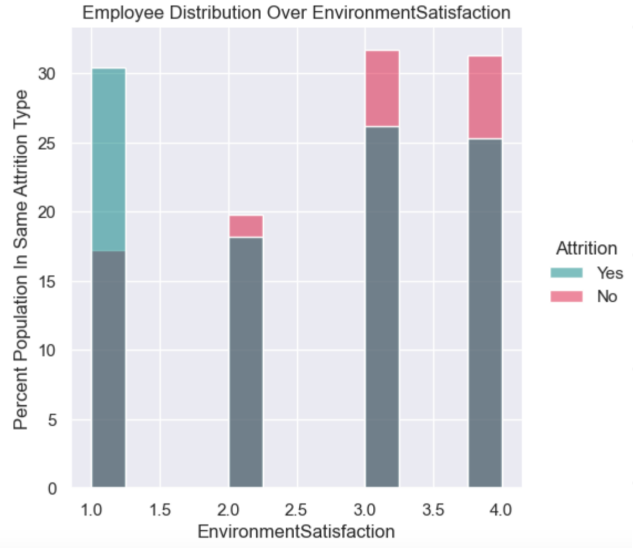


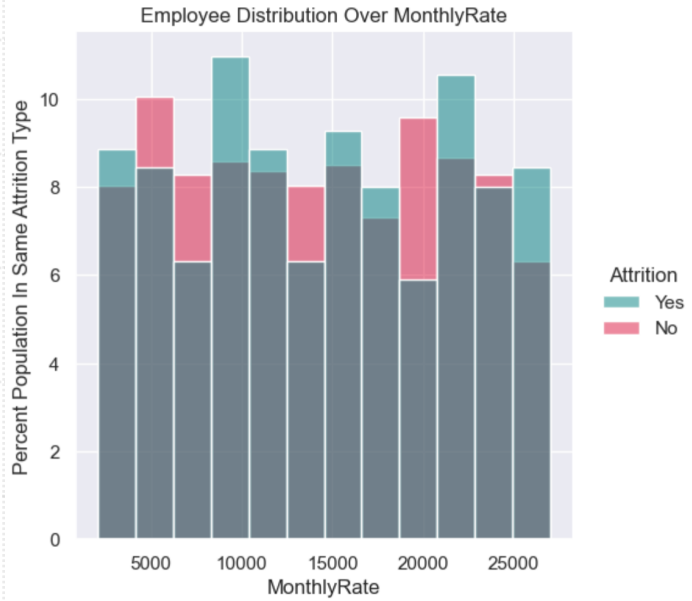
Grouping Variables

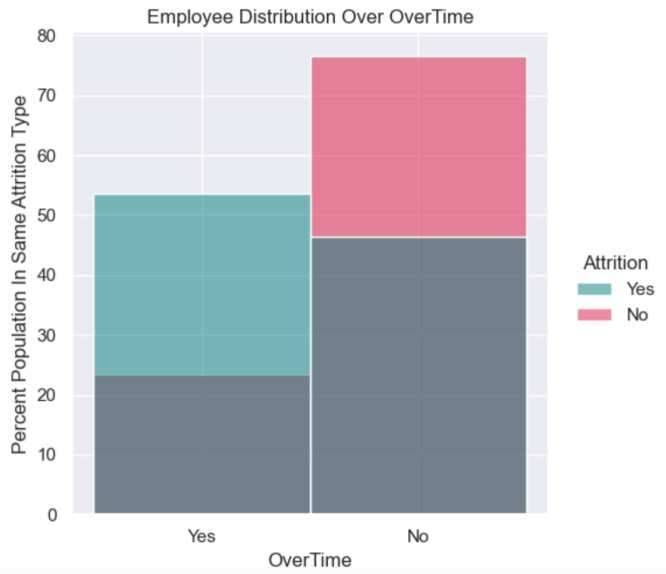
Group Number	1	2	3	4	5
Variables	Attrition BusinessTravel DistanceFromHome EnvironmentSatis. MonthlyRate OverTime	Age JobLevel MonthlyIncome TotalWorkingYears	PercentSalaryHike PerformanceRating	DailyRate Education EmployeeCount EmployeeNumber Gender HourlyRate JobInvolvement JobSatisfaction NumCompaniesWorked Over18 RelationshipSatisfaction StandardHours StockOptionLevel TrainingTimesLastYear WorkLifeBalance	YearsAtCompany YearsInCurrentRole YearsSinceLastPromotion

Attrition V.S. Variables Within Group









Findings & Suggestions for Reducing Attrition

- Less business travel
- Set up offices closer to employees' homes
  - Or hire employees closer to offices
- Improve working environment
- Less overtime

Additional Analysis

- Employee grouping
- Incorporating more categorical data
  - EducationField, Department, etc