

Mitarbeiter-Attrition Bericht

Explorative Datenanalyse & Vorhersagemodelle

HR-Analytics-Datensatz

[GitHub-Repository-Link](#)

Dies ist ein fiktiver Datensatz, der von IBM-Datenwissenschaftler erstellt wurde.



Einleitung

In diesem Bericht fasse ich die Ergebnisse der Explorativen Datenanalyse (EDA) und der Modellierung zur Vorhersage von Mitarbeiter-Attrition anhand des HR-Analytics-Datensatzes zusammen. Ich integriere dazu im Abschnitt EDA passende Visualisierungen in Form von Diagrammen und Heatmaps.

Workflow:

1. Datensatzüberblick
2. Explorative Datenanalyse (EDA) mit Visualisierungen
3. Datenvorverarbeitung
4. Modellauswahl & -bewertung
5. Feature-Wichtigkeit
6. Fazit & Empfehlungen

Wie wirkt sich Mitarbeiter-Attrition auf ein Unternehmen aus?

- Hohe Attrition kann Kosten durch Neuanstellungen, Wissensverlust, geringere Produktivität und negative Auswirkungen auf die Kundenzufriedenheit verursachen. Langfristig schwächt es die Unternehmensstabilität und den Aufbau von Erfahrungswerten.

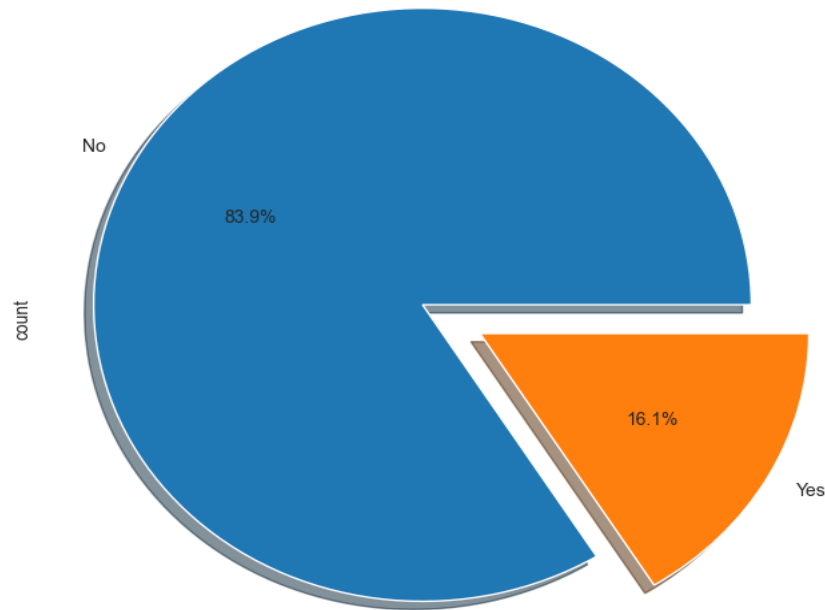
Datensatzüberblick

- - Anzahl Datensätze: 1.470 Mitarbeiter
- - Anteil Attrition: 16,1 % (237/1.470) verlassen das Unternehmen
- - 35 Merkmale (Alter, Gehalt, Abteilung, JobRole, OverTime, Zufriedenheitswerte, etc.)

Explorative Datenanalyse

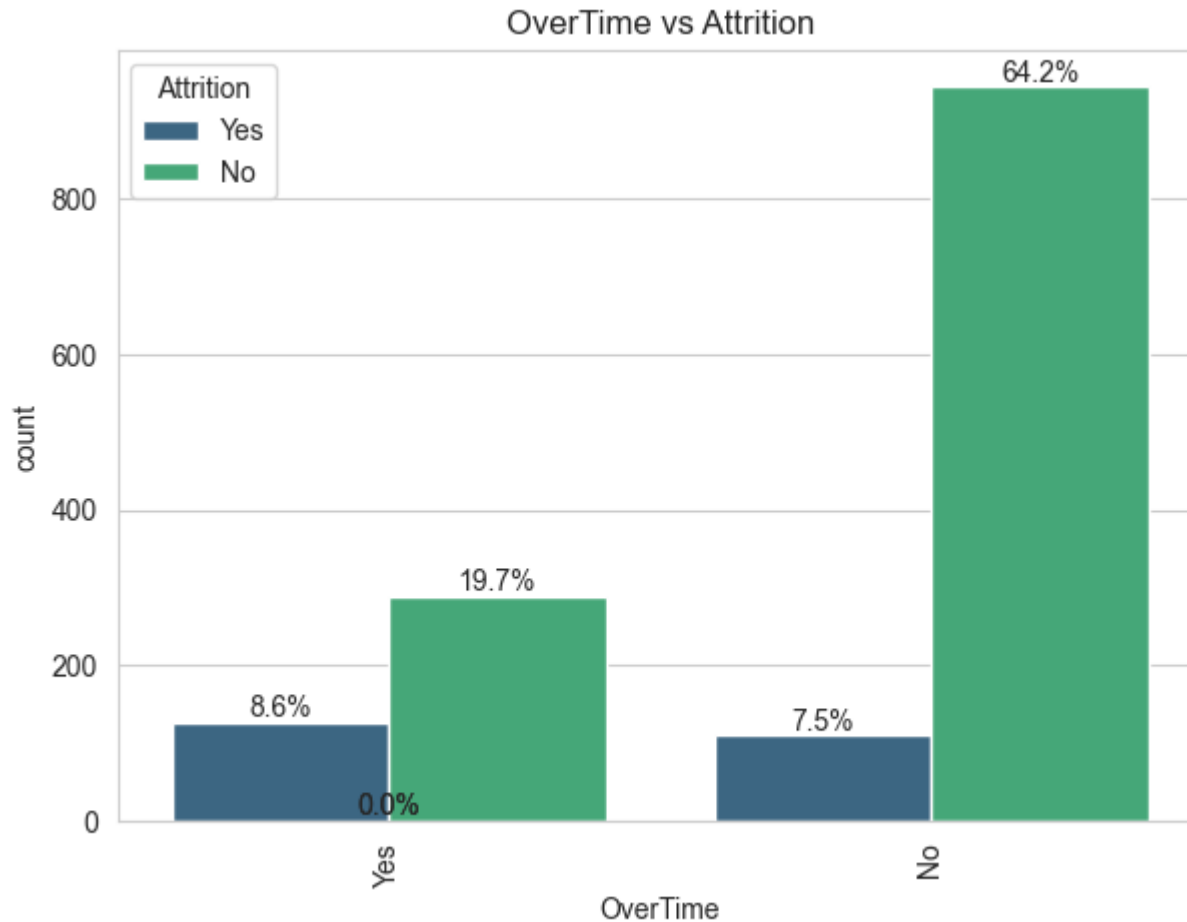
- Verteilung der Attrition

```
Attrition
No      1233
Yes      237
Name: count, dtype: int64
```



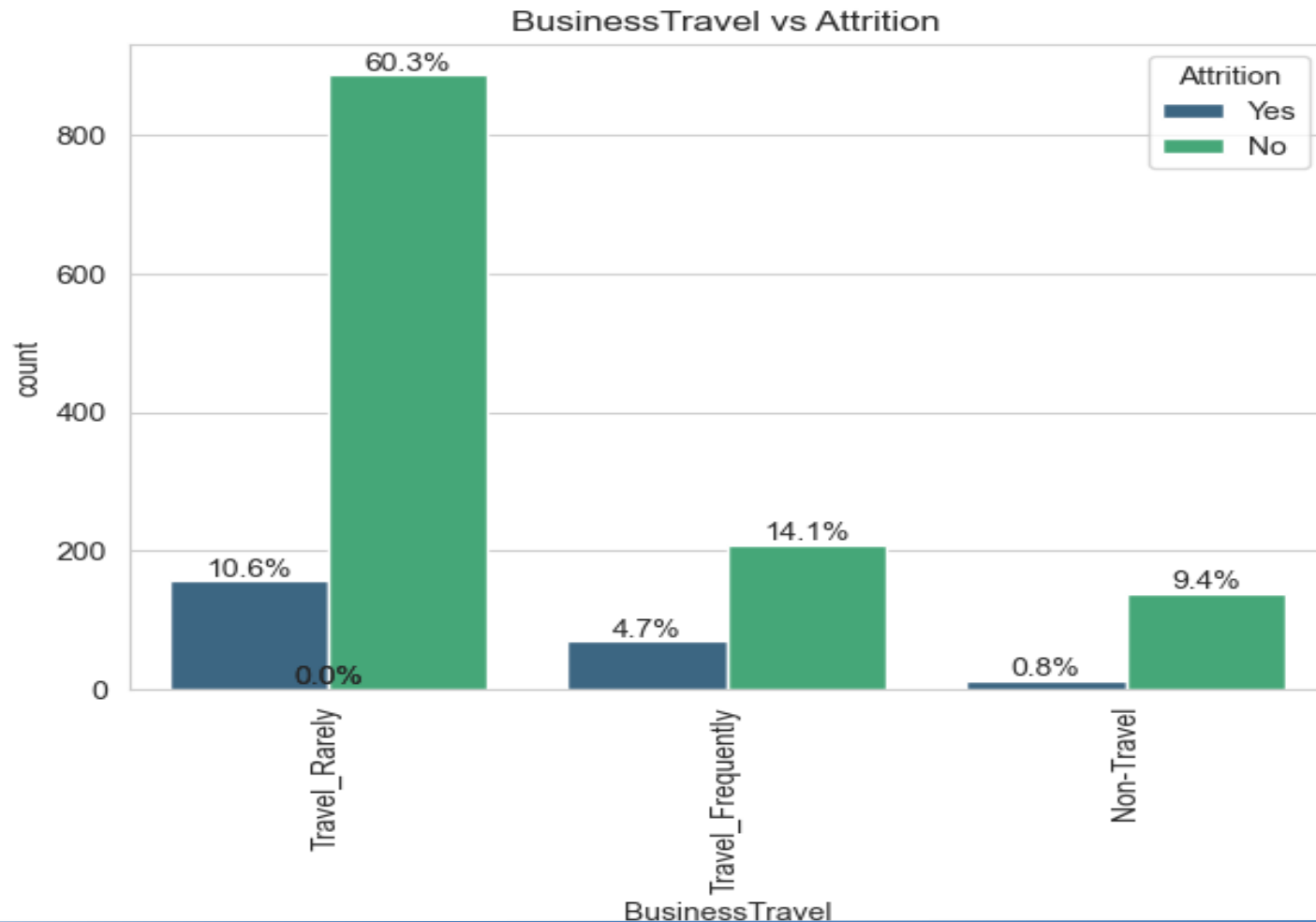
Interpretation: Etwa 16 % der Belegschaft geben an, das Unternehmen verlassen zu haben.

- OverTime vs. Attrition



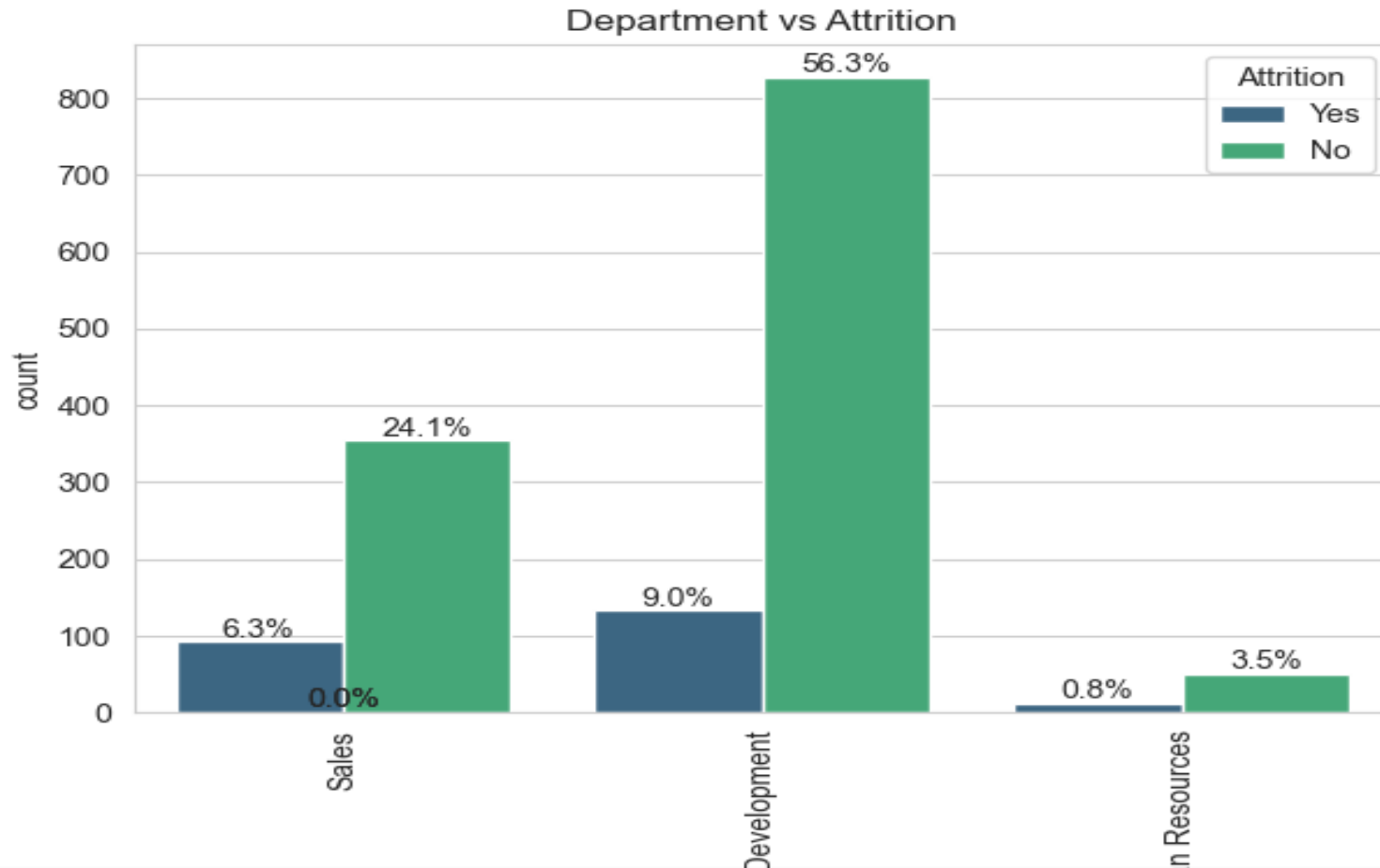
Ergebnis: Mitarbeiter mit Überstunden („Yes“) weisen deutlich höhere Abwanderung auf.

BusinessTravel vs. Attrition



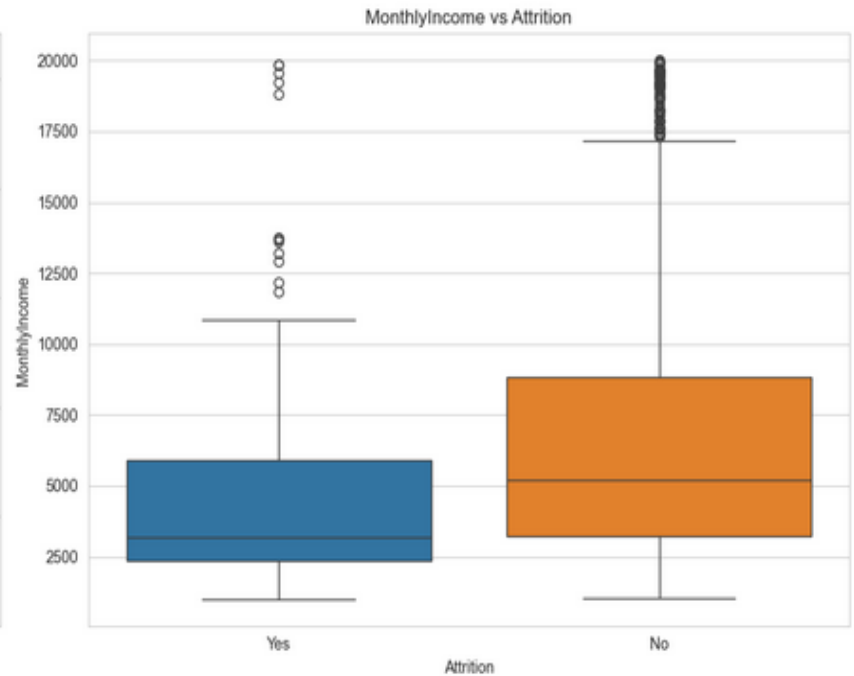
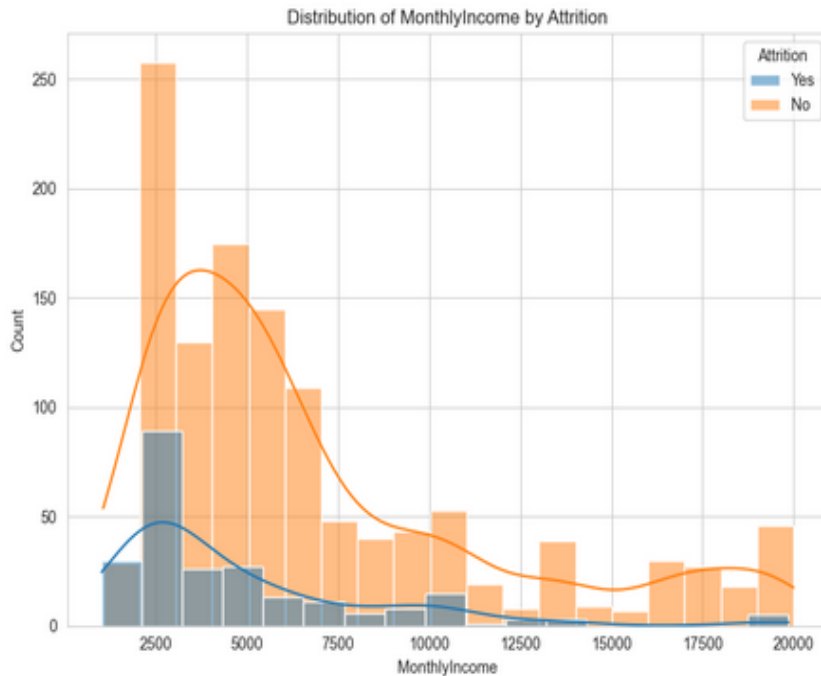
Ergebnis: Häufig reisende Mitarbeiter haben eine erhöhte Wechselbereitschaft.

Department vs. Attrition



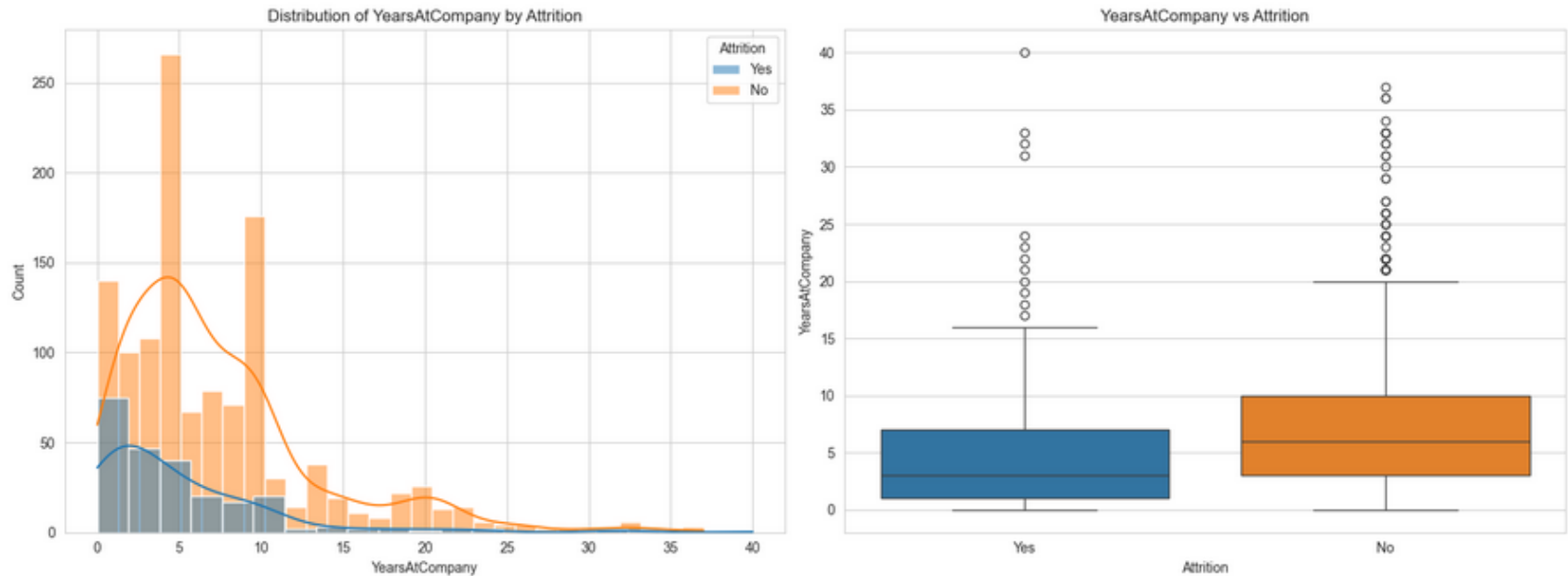
Abteilung: (Research & Development und Sales) besonders betroffen

MonthlyIncome



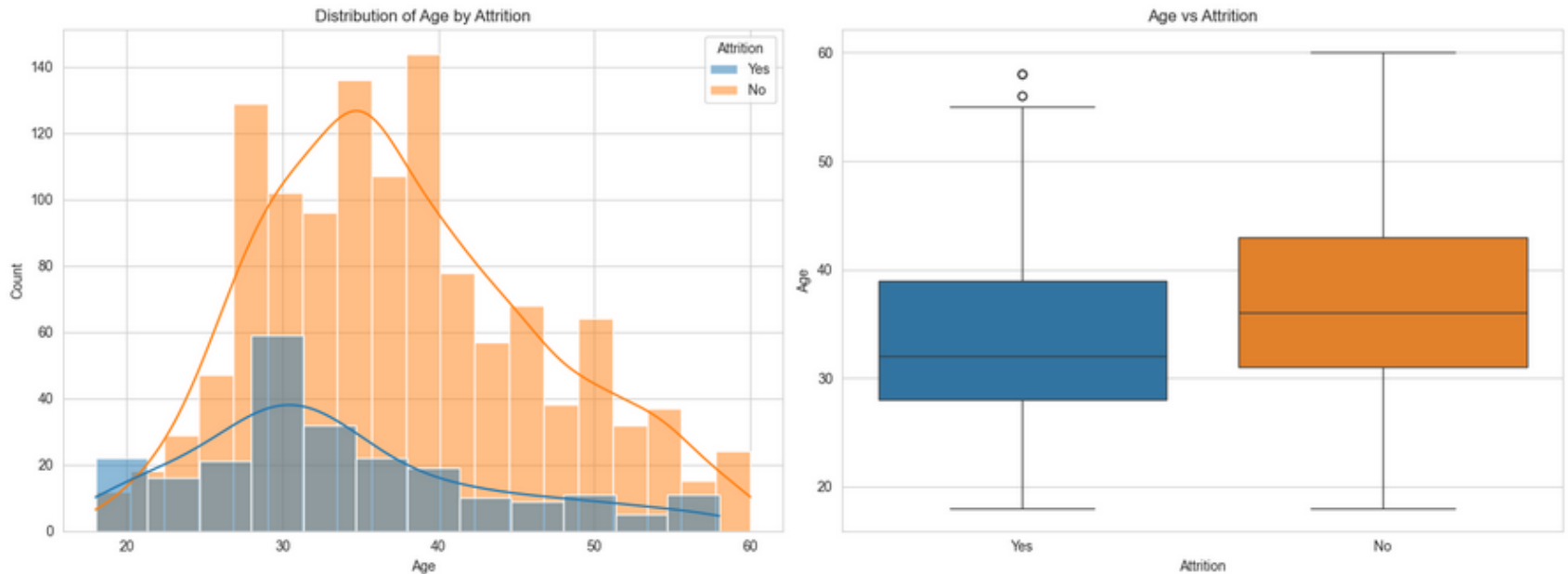
Monatsgehalt: niedriges Gehalt → höhere Abwanderung.

YearsAtCompany



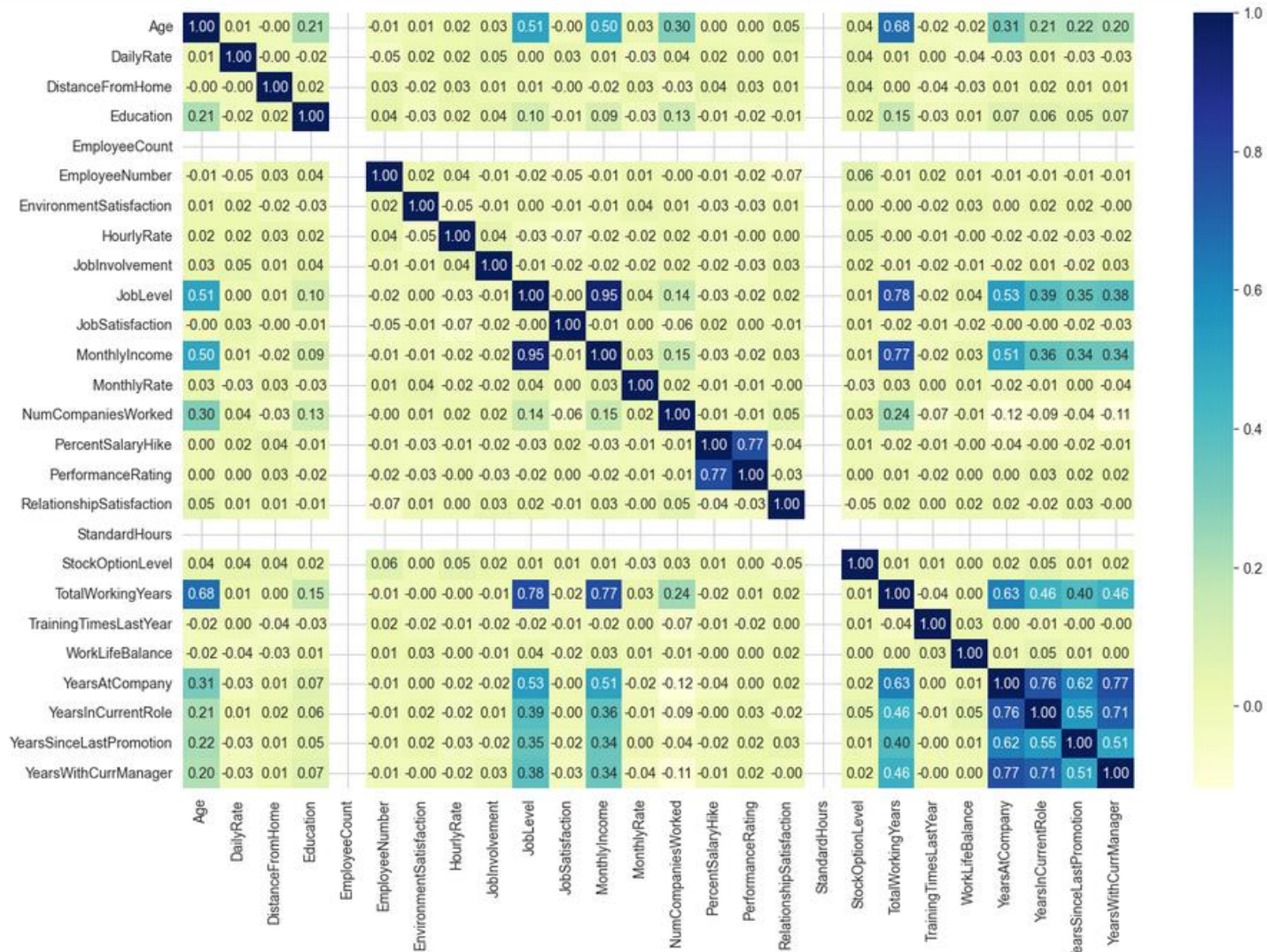
- Betriebszugehörigkeit: kürzere Zeit → Wechselbereitschaft

Age



Alter: Jüngere Mitarbeiter haben höhere Abwanderungsraten.

Korrelationsanalyse



Moderate Zusammenhänge: YearsAtCompany vs. YearsInCurrentRole

Datenvorverarbeitung

- 1. Label-Encoding aller Kategorialen Merkmale (LabelEncoder)
- 2. Skalierung numerischer Merkmale (StandardScaler)
- 3. Train-Test-Split (80/20-Aufteilung, test_size = 0.2)

preprocessing

```
In [60]: df['Attrition'] = df['Attrition'].map({'Yes': 1, 'No': 0})  
df['Attrition'] = df['Attrition'].astype(int)
```

```
In [61]: #encoding categorical columns  
from sklearn.preprocessing import LabelEncoder  
le = LabelEncoder()  
for col in cat.columns:  
    df[col] = le.fit_transform(df[col])
```

target and features

```
In [62]: X = df.drop('Attrition', axis=1)  
y = df['Attrition']
```

```
In [63]: from sklearn.model_selection import train_test_split  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
In [64]: #Scaling the data  
from sklearn.preprocessing import StandardScaler  
scaler = StandardScaler()  
  
X_train = scaler.fit_transform(X_train)  
X_test = scaler.transform(X_test)
```

Modellauswahl & Bewertung

- Logistic Regression:

Accuracy ~89 %

Precision (Yes): 0,68

Recall (Yes): 0,33

F1-Score (Yes): 0,45

Konfusionsmatrix :

[[249 6]

[26 13]]

Interpretation: Hohe Gesamtgenauigkeit, jedoch niedriges Recall auf der positiven Klasse (Attrition)

Feature-Wichtigkeit

- Top-5 Merkmale:
- 1. OverTime
- 2. MonthlyIncome
- 3. YearsAtCompany
- 4. JobSatisfaction
- 5. Age

Fazit & Empfehlungen

- - Reduktion von Überstunden
- - Gehaltsanpassungen
- - Mentoring für neue Mitarbeiter
- - Dashboard zur Früherkennung von Abwanderung