👋Project: Employee Attrition Analysis and Prediction

Goal:

- Understand factors driving employee attrition.
- Develop a model to identify at-risk employees.
- Inform HR policies for better retention.

Data:

- Dataset: "HR-Employee-Attrition.csv"
- Includes demographics, job details, compensation, work history, satisfaction, travel, and other factors.
- Target: 'Attrition' (Yes/No)

Methodology:

- 📊 Exploratory Data Analysis
- ⚙️ Feature Engineering
- 🧠 Model Building & Evaluation
- 📈 Feature Importance Analysis
- 💡 Business Insights & Recommendations

📉 Attrition Distribution

⚖️ **Class Imbalance:**

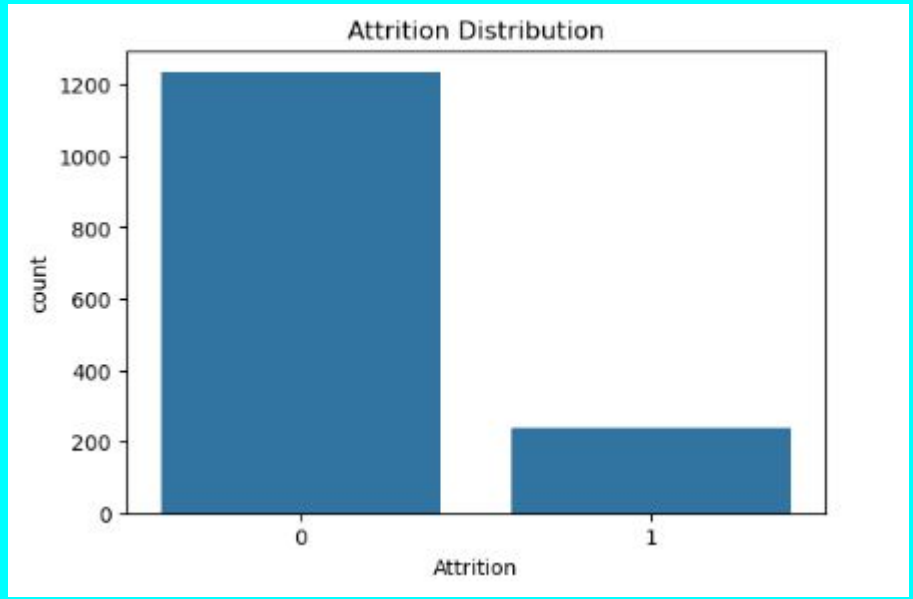
- Class 0 (No Attrition): ~1200 employees (majority class)
- Class 1 (Attrition): ~200 employees (minority class)
- Imbalance Ratio: ~6:1 (No Attrition : Attrition)

❗ **Interpretation:**

- Dataset is highly imbalanced.
- Models may be biased towards "No Attrition."
- Use metrics beyond accuracy (Precision, Recall, F1, ROC-AUC).

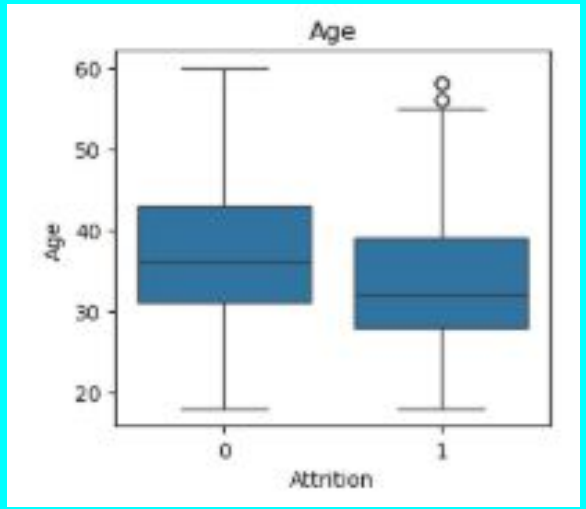
💸 **Business Impact:**

- Attrition rate is ~14% (200/1400) - a key HR concern.
- Addressing imbalance is crucial for:
  - 🎯 Retention strategies (identifying at-risk employees)
  - 💰 Cost reduction (high costs of hiring & training)



🎂 Age vs. Attrition

- **Median Age:**
  - Stayed: ~35-36 years old
  - Left: ~32-33 years old
- **Interquartile Range (IQR):**
  - Stayed: 32-43 years old
  - Left: 28-40 years old (narrower)
- **Age Range:**
  - Stayed: 18 - 60 years old
  - Left: 18 - 55 years old
- **Outliers:**
  - Stayed: None
  - Left: Few older outliers (around 56-58 years)

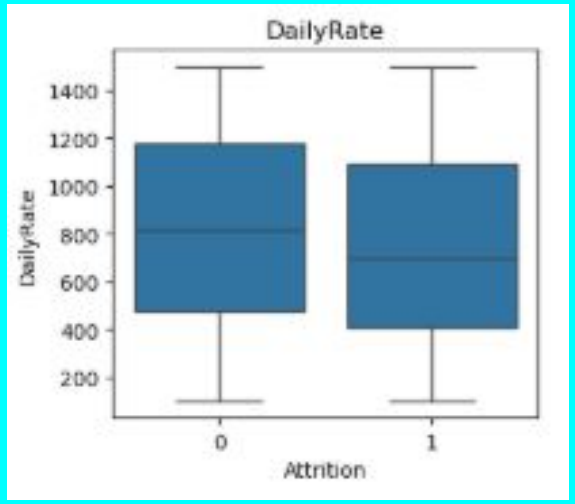


Implications:

- 👶 Younger employees may leave slightly more frequently.
- 📉 Attrition group has a more concentrated age range.
- 👴 Older employees leaving are less common, but present.

💰 Daily Rate vs. Attrition

- **Median Daily Rate:**
  - Stayed: ~800-850
  - Left: ~650-700
- **Interquartile Range (IQR):**
  - Stayed: 500 - 1200
  - Left: 400 - 1050
- **Range:**
  - Stayed: 100 - 1500
  - Left: 100 - 1450-1500
- **Outliers:**
  - Both groups: None significant

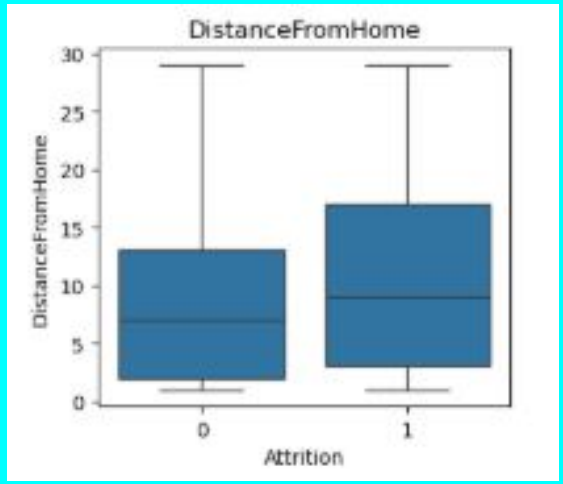


Implications:

- 👊 Slightly lower daily rates for employees who left.
- 📊 Similar distribution of daily rates for both groups.
- ❌ Daily Rate alone is not a strong predictor of attrition.

🏠 Distance from Home vs. Attrition

- **Median Distance:**
  - Stayed: ~7-8
  - Left: ~9-10
- **Interquartile Range (IQR):**
  - Stayed: 2 - 14
  - Left: 3 - 17 (slightly wider)
- **Range:**
  - Stayed: 1 - 29
  - Left: 1 - 29
- **Outliers:**
  - Both groups: None significant

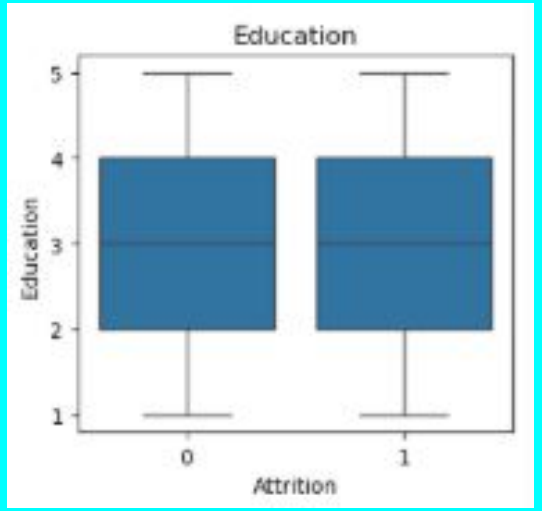


Implications:

- 🚶‍♂️🚶‍♀️ Employees who left tend to live slightly further away.
- 📊 More variability in commute distance for those who left.
- 👨‍👩‍👧 Distance from home might be a contributing, but not a primary, factor in attrition.

🎓 Education vs. Attrition

- **Median Education:**
  - Stayed: ~3 (Bachelor's)
  - Left: ~3 (Bachelor's)
- **Interquartile Range (IQR):**
  - Stayed: 2 (College) - 4 (Master's)
  - Left: 2 (College) - 4 (Master's)
- **Range:**
  - Stayed: 1 (Below College) - 5 (Doctor's)
  - Left: 1 (Below College) - 5 (Doctor's)
- **Outliers:**
  - Both groups: None

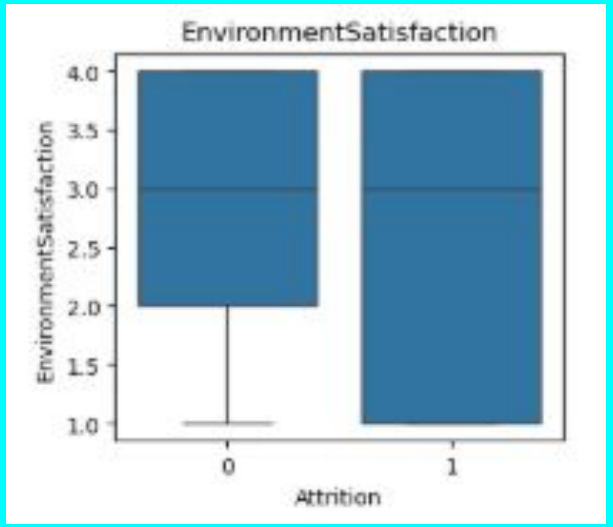


Implications:

- ✅ Similar education levels for both groups.
- 👩‍🎓👨‍🎓 Education level is not a key factor in attrition.

🌳 Environment Satisfaction vs. Attrition

- **Median Environment Satisfaction:**
  - Stayed: ~3.0
  - Left: ~3.0
- **Interquartile Range (IQR):**
  - Stayed: 2.0 - 4.0
  - Left: 2.0 - 4.0
- **Range:**
  - Stayed: 1.0 - 4.0
  - Left: 1.0 - 4.0
- **Outliers:**
  - Both groups: None

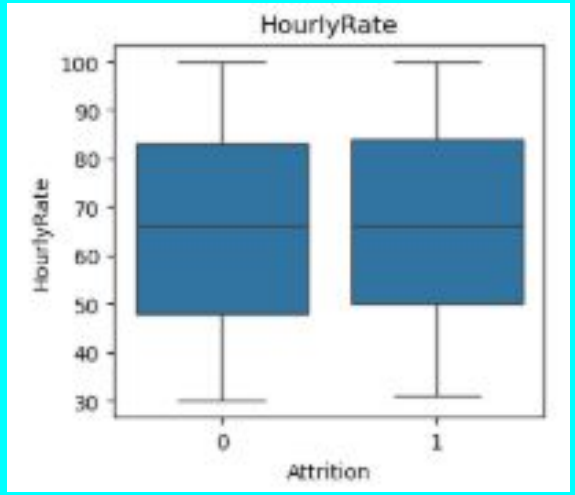


Implications:

- ✅ Similar environment satisfaction levels for both groups.
- 👩‍🤝‍👨 Environment satisfaction, in isolation, is not a strong predictor of attrition.

💰 Hourly Rate vs. Attrition

- **Median Hourly Rate:**
  - Stayed: ~65-67
  - Left: ~65-67
- **Interquartile Range (IQR):**
  - Stayed: 48-50 to 82-84
  - Left: 50-52 to 84-86
- **Range:**
  - Stayed: 30 - 100
  - Left: 31 - 100
- **Outliers:**
  - Both groups: None significant

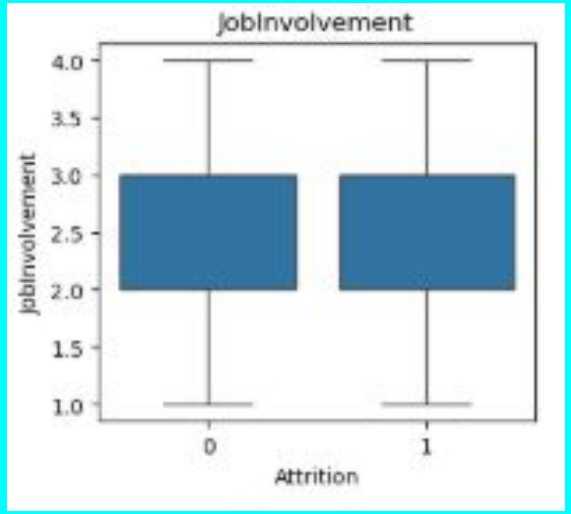


Implications:

- ✅ Hourly rate is virtually the same for both groups.
- 🤷‍♀️🤷‍♂️ Hourly rate, in isolation, does not influence employee attrition.

💼 Job Involvement vs. Attrition

- **Median Job Involvement:**
  - Stayed: ~2.75 - 3.0
  - Left: ~2.75 - 3.0
- **Interquartile Range (IQR):**
  - Stayed: 2.0 - 3.0
  - Left: 2.0 - 3.0
- **Range:** 1.0 - 4.0 for both groups
- **Outliers:** None significant in either group

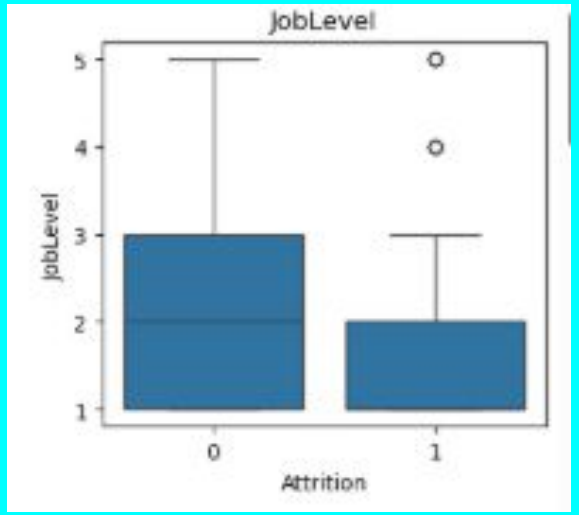


Implications:

- ✅ Similar job involvement levels for both groups.
- 🤷‍♀️🤷‍♂️ Job involvement, in isolation, does not appear to be a strong differentiator in employee attrition.

🪜 Job Level vs. Attrition

- **Median Job Level:**
  - Stayed: ~2
  - Left: ~1.5-2
- **Interquartile Range (IQR):**
  - Stayed: 1 - 3
  - Left: 1 - 2 (smaller, lower)
- **Range:**
  - Stayed: 1 - 5
  - Left: 1 - 3
- **Outliers:**
  - Both groups: Some higher-level outliers, but fewer and lower for those who left.

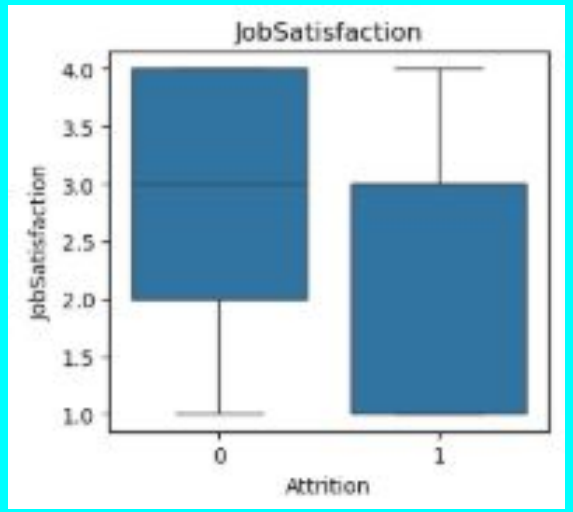


Implications:

- 📉 Lower job levels are associated with higher attrition.
- 😟 Attrition is concentrated among lower-level employees.

😊 Job Satisfaction vs. Attrition

- **Median Job Satisfaction:**
  - Stayed: ~3.0
  - Left: ~2.0 - 3.0
- **Interquartile Range (IQR):**
  - Stayed: 2.0 - 4.0
  - Left: 1.0 - 3.0 (smaller, lower)
- **Range:** 1.0 - 4.0 for both groups
- **Outliers:** None significant in either group

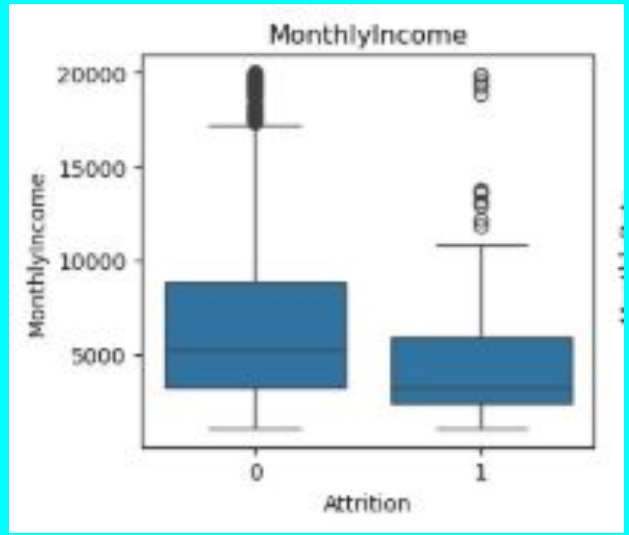


Implications:

- 😕 Lower job satisfaction is associated with higher attrition.
- 😟 Attrition is concentrated among employees with lower job satisfaction.

💰 Monthly Income vs. Attrition

- **Median Monthly Income:**
  - Stayed: ~5000 - 6000
  - Left: ~3000
- **Interquartile Range (IQR):**
  - Stayed: 3000 to 8000 - 9000
  - Left: 2000 to 5000 - 6000
- **Range:**
  - Stayed: 1000/2000 - 17000/18000
  - Left: 1000 - 11000
- **Outliers:**
  - Stayed: Many high outliers (up to 20000)
  - Left: Fewer, lower outliers (up to 20000)

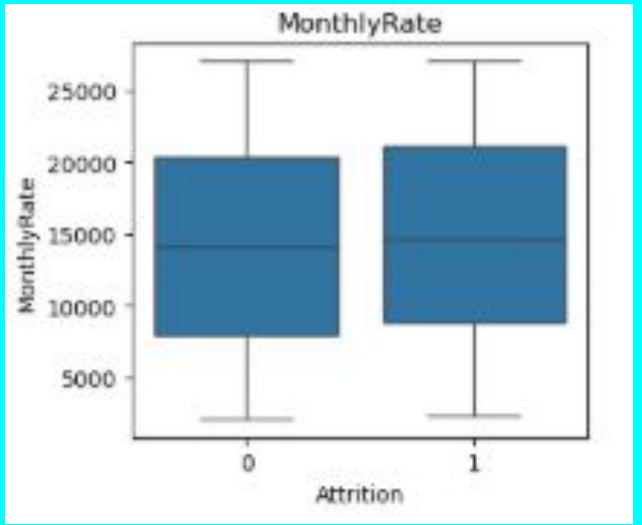


Implications:

- 💸 Lower monthly income = higher attrition risk.
- 😟 Attrition is concentrated among lower-income earners.

💰 Monthly Rate vs. Attrition

- **Median Monthly Rate:**
  - Stayed: ~14000
  - Left: ~14000
- **Interquartile Range (IQR):**
  - Stayed: ~8000 to 20000
  - Left: ~9000 to 21000
- **Range:** ~2000 to 27000 for both groups
- **Outliers:** None apparent in either group

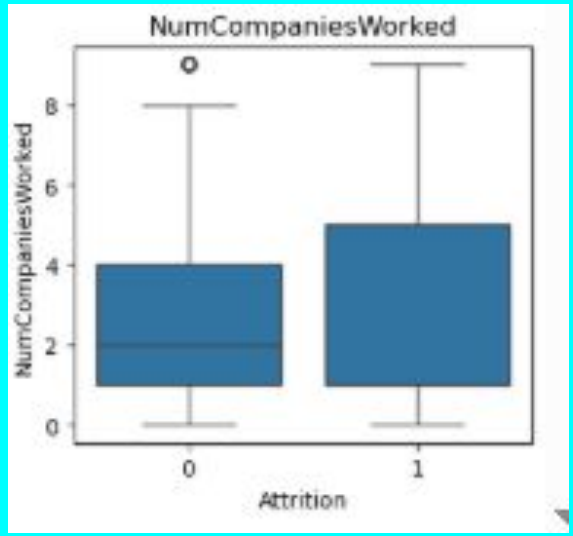


Implications:

- ✅ Similar monthly rates for both groups.
- 👩‍🤝‍👨 Monthly Rate, in isolation, does not appear to be a strong predictor of attrition.

💼 # of Companies Worked vs. Attrition

- **Median # of Companies:**
  - Stayed: ~2
  - Left: ~4
- **Interquartile Range (IQR):**
  - Stayed: 1 to 4
  - Left: 1 to 5
- **Range:** 0 to 9 for both groups
- **Outliers:**
  - Stayed: One outlier at 9
  - Left: None



Implications:

- ➡️ Employees who left have worked for more companies on average.
- ⚠️ Job-hopping history is associated with higher attrition.

✈️ Business Travel vs. Attrition

- **Travel Frequency:**
  - Most employees: Travel Rarely
  - Least employees: Don't Travel
- **Attrition:**
  - Highest attrition *proportionally*: Travel Frequently
  - Lowest attrition: Non-Travel

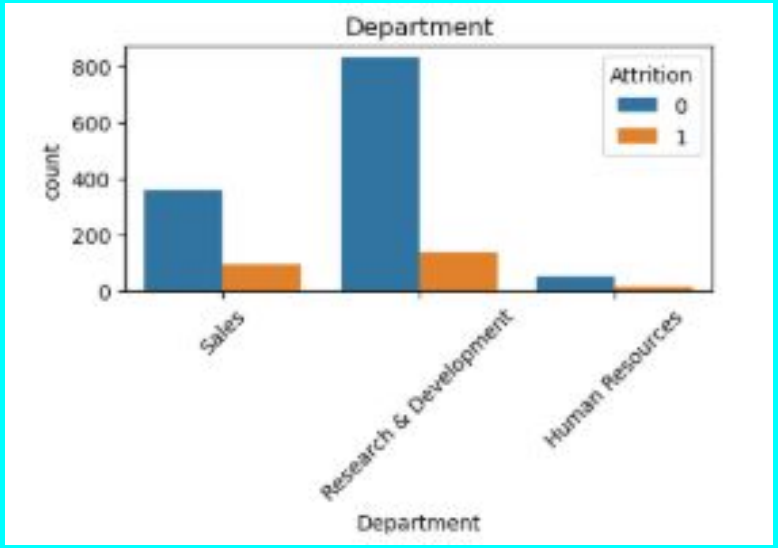
Implications:

- Frequent business travel is associated with higher employee turnover.

🏢 Department vs. Attrition

Key Observations:

- **Largest Dept:** Research & Development (highest overall employee count)
- **Highest Attrition Rate:** Sales (proportionally more employees leaving)
- **Lowest Attrition Rate:** Human Resources



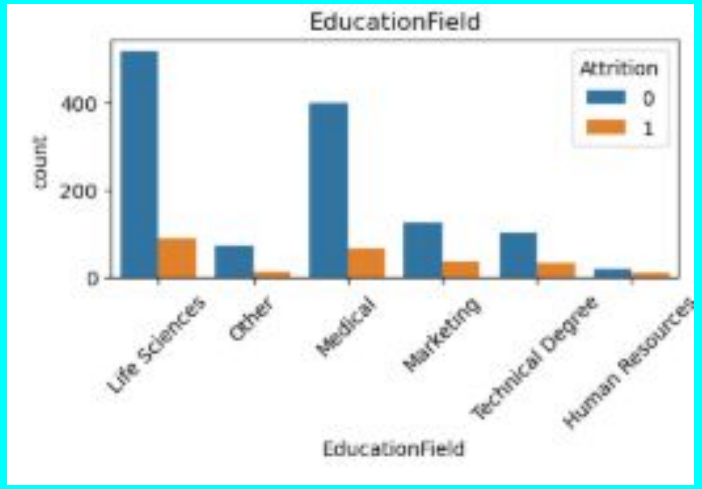
Implications:

- Attrition varies by department.
- Sales has a higher turnover rate.
- HR has the most stable workforce.

🎓 Education Field vs. Attrition

Key Observations:

- **Most Common Fields**: Life Sciences & Medical
- **Highest Retention**: Life Sciences
- **Relatively Higher Attrition**: Marketing, Technical Degree
- **Small Sample Sizes**: Other, Human Resources



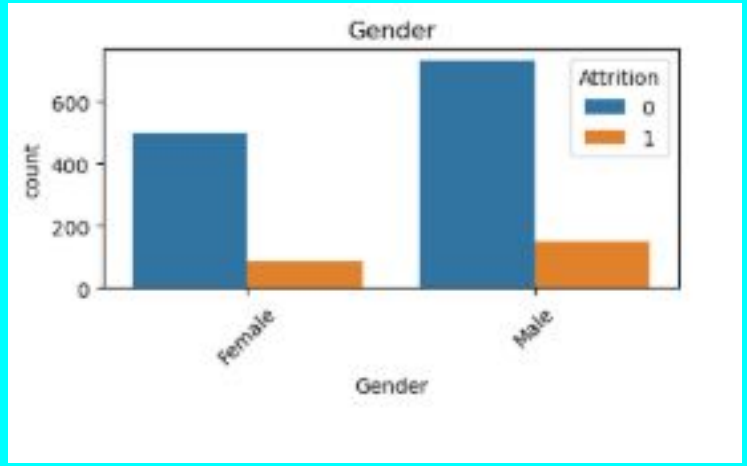
Implications:

- Attrition varies across education fields.
- Life Sciences tend to stay more.
- Marketing & Technical Degree fields tend to leave more frequently.

🚻 Gender vs. Attrition

Key Observations:

- More male employees overall.
- Higher number of both male and female employees stayed.
- Potentially slightly higher attrition *proportion* for female employees (needs further statistical confirmation).



Implications:

- Dataset has more male employees.
- Both genders experience attrition.
- Further analysis (percentage calculation & statistical testing) needed to confirm if there's a significant difference in attrition rates between genders.

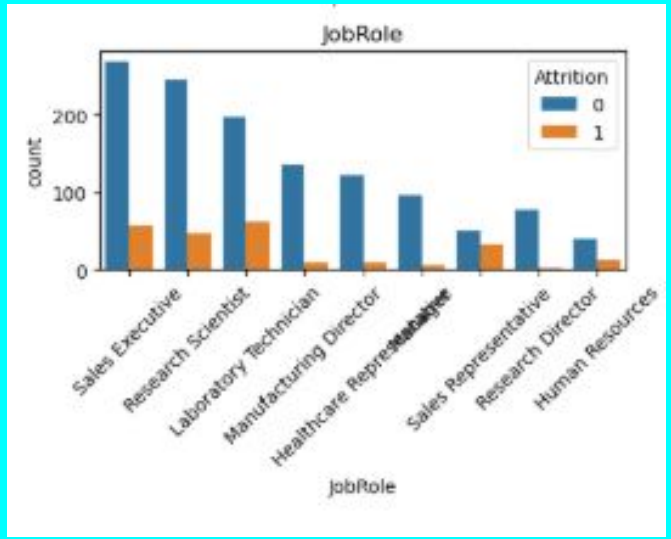
👨‍💼 Job Role vs. Attrition

Key Observations:

- **High Employee Count Roles:** Sales Executive, Research Scientist
- **High Attrition Roles:** Sales Representative, Laboratory Technician
- Attrition varies significantly across job roles.

Implications:

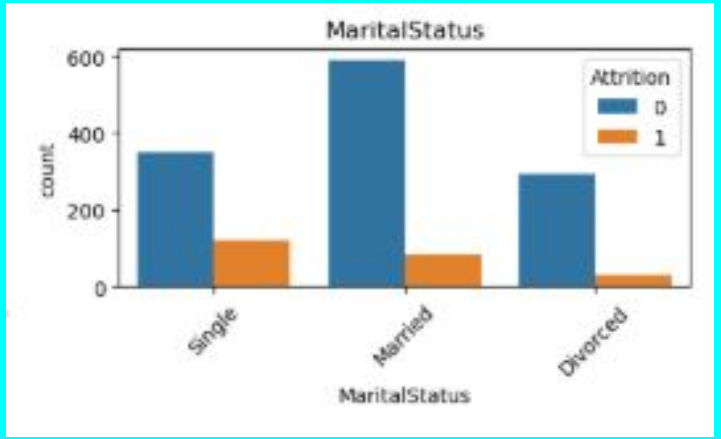
- Sales and Research roles have the highest employee numbers.
- Sales Reps and Lab Technicians have a higher proportion of employees leaving.



💍 Marital Status vs. Attrition

Key Observations:

- **Employee Distribution**:
  - Most employees are Married.
- **Attrition**:
  - Single employees have the highest *relative* attrition.
  - Married employees have the lowest *relative* attrition.



Implications:

- Marital status influences attrition.
- Single employees are more likely to leave compared to married employees.

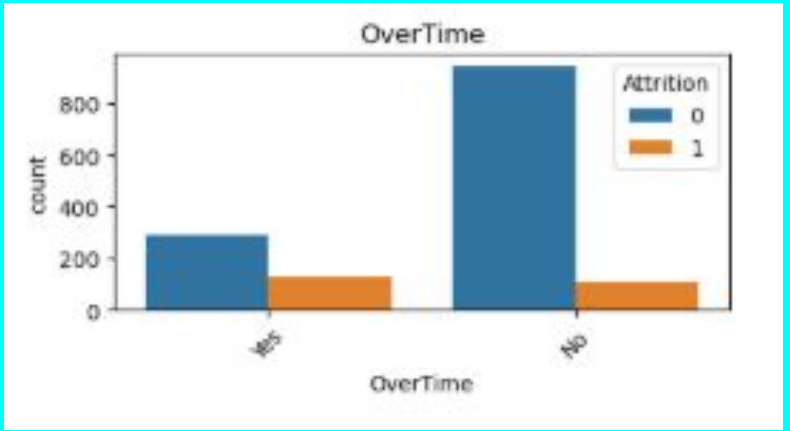
⏰ Overtime vs. Attrition

Key Observations:

- **Overtime Workload**:
  - Most employees: Don't work overtime
- **Attrition**:
  - Employees who work overtime are more likely to leave.

Implications:

- Employees who work overtime are more likely to leave.



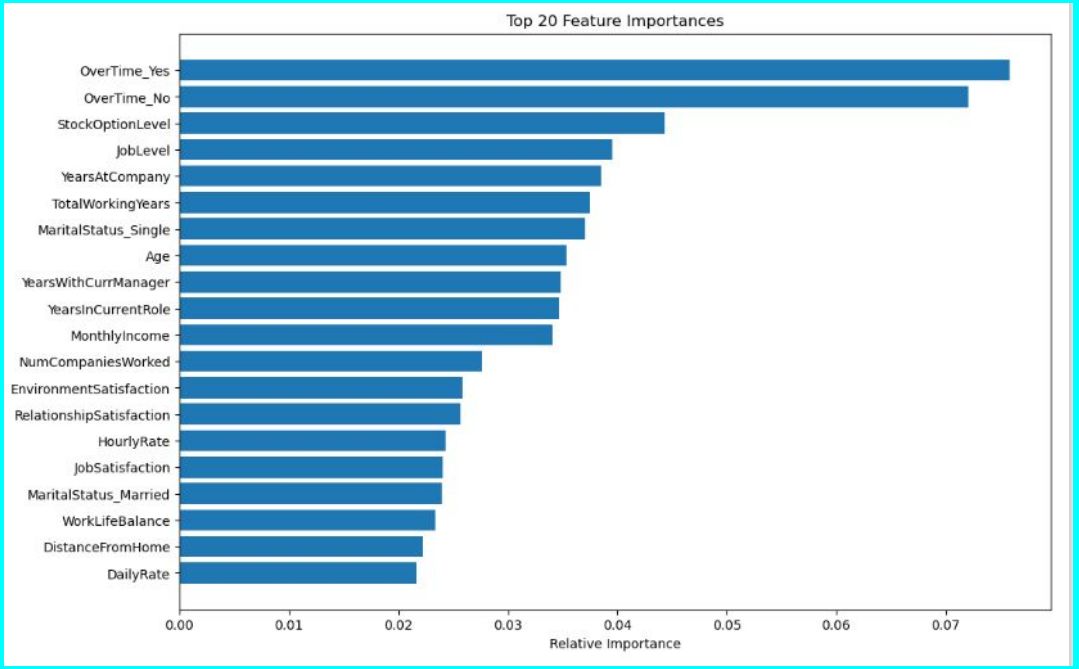
📊 Key Factors Influencing Attrition

Top 5 Most Important Features:

- ⏰ **OverTime:** Most influential factor.
- 💰 **StockOptionLevel:** Significant impact of stock options.
- 💼 **JobLevel:** Employee's position matters.
- 📅 **YearsAtCompany:** Tenure is a key factor.
- 📈 **TotalWorkingYears:** Overall experience plays a role.

Other Influential Features:

- 💔 **MaritalStatus_Single:** Single employees are more likely to leave.
- 🎂 **Age:** Younger employees tend to leave slightly more
- 🏢 **YearsWithCurrManager & YearsInCurrentRole**: Stability within the company is relevant.
- 💸 **MonthlyIncome:** Income is a moderate predictor.
- 🏢 **NumCompaniesWorked**: More previous jobs = higher attrition



Less Influential Features:

- 🌍 **EnvironmentSatisfaction**
- ❤️‍🩹 **RelationshipSatisfaction**
- 💵 **HourlyRate, DailyRate**
- 😊 **JobSatisfaction**
- ⚖️ **WorkLifeBalance**
- 📍 **DistanceFromHome**

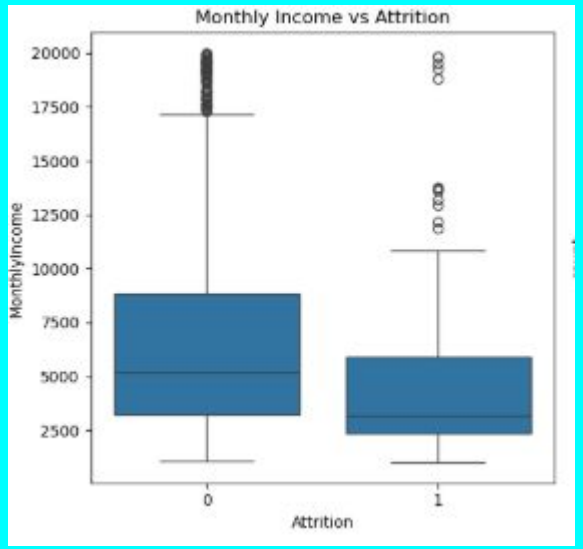
Key Takeaway:

- Overtime, compensation, career progression, and tenure are the strongest drivers of attrition.

💰 Monthly Income vs. Attrition

Key Observations:

- **Median Monthly Income:**
  - Stayed: ~5000 - 6000
  - Left: ~3000
- **Interquartile Range (IQR):**
  - Stayed: 3000 to 8000 - 9000
  - Left: 2000 to 5000 - 6000
- **Range:**
  - Stayed: 1000/2000 - 17000/18000
  - Left: 1000 - 11000
- **Outliers:**
  - Stayed: Many high outliers (up to 20000)
  - Left: Fewer, lower outliers (up to 20000)



Implications:

- 💸 Lower monthly income = higher attrition risk.
- 😟 Attrition is concentrated among lower-income earners.

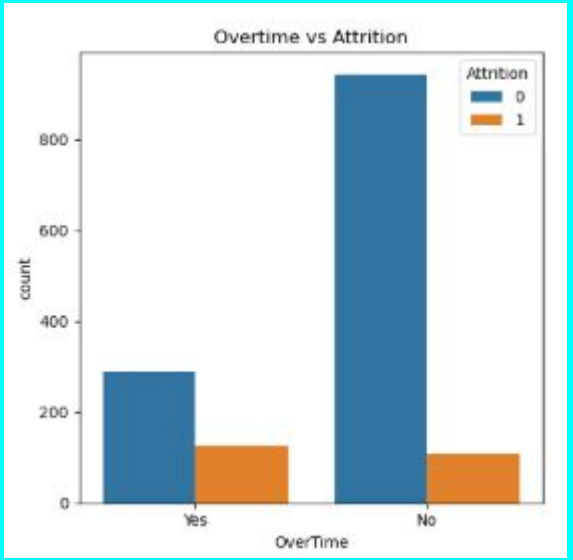
⏰ Overtime vs. Attrition

Key Observations:

- **Overtime Workload**:
  - Most employees: Don't work overtime
- **Attrition**:
  - Employees who work overtime are more likely to leave.

Implications:

- Employees who work overtime are more likely to leave.

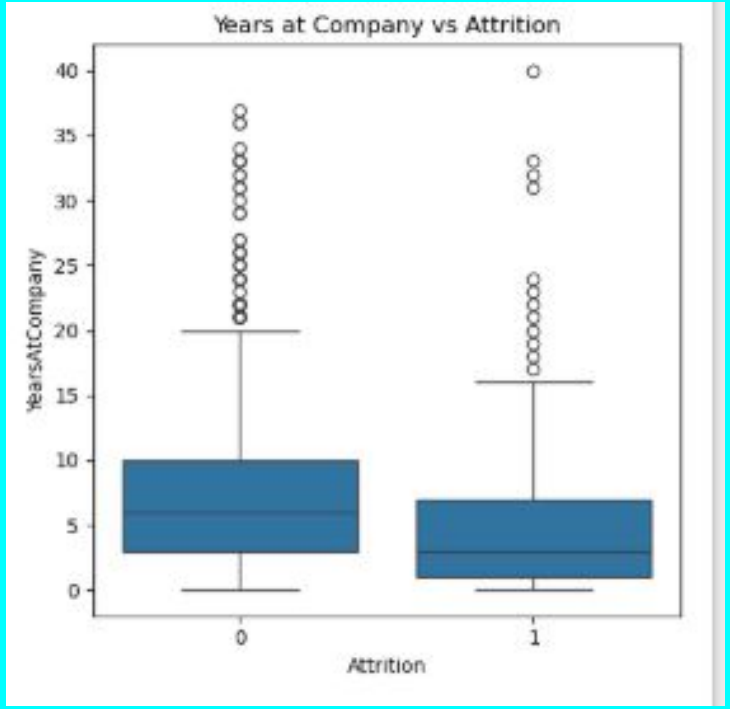


📅 Years at Company vs. Attrition

- **Median Tenure:**
  - Stayed: ~6-7 years
  - Left: ~2-3 years
- **Interquartile Range (IQR):**
  - Stayed: 3 to 10 years
  - Left: 1 to 7 years
- **Range:** 0 to 40 years (Stayed), 0 to 20 years (Left)
- **Outliers:** More high-value outliers for those who stayed.

Implications:

- ⏳ Shorter tenure = higher attrition risk.
- Loyalty increases with time at the company.



📊 Logistic Regression Performance

Key Metrics:

- **Accuracy:** 77%
- **ROC AUC:** 0.808

Precision:

- Class 0 (No Attrition): 92% 👍
- Class 1 (Attrition): 37% 👎

Recall:

- Class 0 (No Attrition): 79% 👍
- Class 1 (Attrition): 65% ⚠️

F1-Score:

- Class 0 (No Attrition): 85% 👍
- Class 1 (Attrition): 47% ⚠️

Interpretation:

- Good at predicting employees who stay.
- Less effective at identifying employees who leave.
- Overall, a balanced performance.

🎯 Project Summary

Project Goal:

- Understand & predict employee attrition to inform HR strategies.

Key Findings:

- ❗ **Strongest Drivers:** Overtime, Compensation, Career Progression, Tenure.
- 📉 **Higher Attrition:** Lower income, lower job level, shorter tenure, job hopping.
- ✈️ **Business Travel:** Frequent travel increases attrition.
- 💔 **Marital Status:** Single employees more likely to leave.

Recommendations:

- Reduce overtime, offer competitive pay & stock options.
- Provide growth opportunities & recognize tenure.
- Address travel-related challenges.


THANK YOU!