HR Analytics Project Report: Employee Attrition Analysis

Author: Aanchal Singh
Project Title: Identifying Top Factors Driving Employee Attrition

---

1. Objective

The goal of this project is to identify key factors that contribute to employee attrition within an organization using HR data. This will help HR teams take data-driven decisions to improve employee retention and reduce turnover.

---

2. Dataset Overview

- **Source**: Public HR dataset (CSV file)
- **Records**: [Number of employees]
- **Features**: Age, JobRole, OverTime, MonthlyIncome, etc.
- **Target Variable**: Attrition (Yes/No)

---

3. Tools & Technologies

- Python (PyCharm IDE)
- Pandas, NumPy, Scikit-learn, Matplotlib
- Logistic Regression for model building

---

4. Data Preprocessing

- Encoded categorical variables using LabelEncoder

- Converted 'Attrition' column to binary format (Yes=1, No=0)
- Scaled numeric features using StandardScaler
- Split data into 80% training and 20% testing

---

5. Model Building

Used Logistic Regression to predict the likelihood of an employee leaving.

- **Algorithm**: LogisticRegression (max_iter=2000)
- **Evaluation**: Classification report and feature coefficients

---

6. Key Findings

Top 3 factors influencing attrition based on model coefficients:

1. **OverTime** – Employees working overtime are significantly more likely to leave.
2. **JobRole** – Certain roles have higher turnover rates.
3. **MonthlyIncome** – Lower income levels are associated with higher attrition.

CSV of full feature importance: `feature_importance.csv` PDF of bar chart visualization: `attrition_factors.pdf`

---

7. Recommendations

- Introduce flexible work hours or reduce excessive overtime
- Conduct detailed reviews for high-attrition job roles
- Evaluate and optimize compensation for lower-income roles

- Use exit interviews to gather qualitative insights

---

8. Conclusion

This analysis highlights actionable drivers of employee attrition using a logistic regression model. HR departments can use these insights to implement targeted policies for retention. Future work can involve testing with Random Forest or XGBoost for better accuracy.

---

9. Next Steps

- Share the project on GitHub with README and dataset
- Create a LinkedIn post to showcase the project
- Use this as part of your interview portfolio

GitHub Repository Structure (suggested):

```
hr-analytics-project/
|-- data/
|   |-- hr_data.csv
|-- visuals/
|   |-- attrition_factors.pdf
|-- reports/
|   |-- feature_importance.csv
|   |-- hr_attrition_report.docx
|-- hr_analytics_model.py
|-- README.md
```

LinkedIn Post Template:

> Just completed an HR Analytics project using Python & ML! 📊
>
> I analyzed employee attrition using real HR data and identified key factors like OverTime, JobRole, and MonthlyIncome contributing to employee turnover.

Built with: Python, Pandas, Scikit-learn, Matplotlib

This project helped me build actionable business recommendations that HR teams can use to improve retention.

#HRAnalytics #DataScience #MachineLearning #Python #Attrition #PortfolioProject #AanchalSingh

Let me know when you're ready for the Word and PDF file generation!