HR ANALYTICS: PREDICTING EMPLOYEE ATTRITION

- Tools: Python, PostgreSQL, Tableau, ML (Logistic Regression, RandomForest , XGBoost)

BY Prashant Singh

PROBLEM STATEMENT

GOAL:

IDENTIFY EMPLOYEES WHO ARE LIKELY TO LEAVE THE COMPANY USING HISTORICAL HR DATA.

BUSINESS NEED:

- REDUCE ATTRITION COSTS
- IMPROVE EMPLOYEE RETENTION
- BETTER WORKFORCE PLANNING

DATASET OVERVIEW AND DATA CLEANING

- Data source : Kaggle
- Total Rows and column : 1470 and 35 and Target variable is Attrition
- Dataset is clean and no null value is present and no duplicate value.
- Drop irrelevant columns : EmployeeNumbers , StandardHours etc
- Binary encoding : Attrition , OverTime , Gender.
- Key Columns :

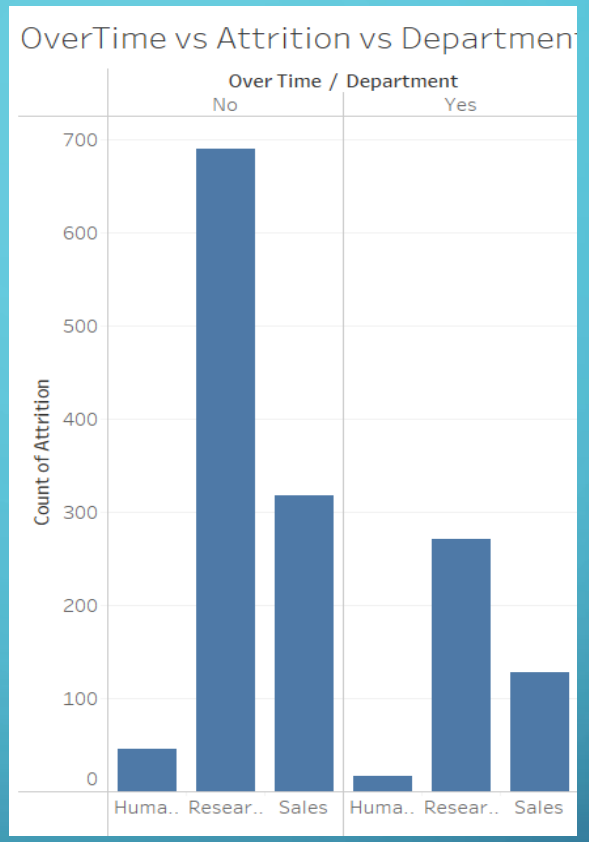
  Age, MonthlyIncome, JobSatisfaction, WorkLifeBalance

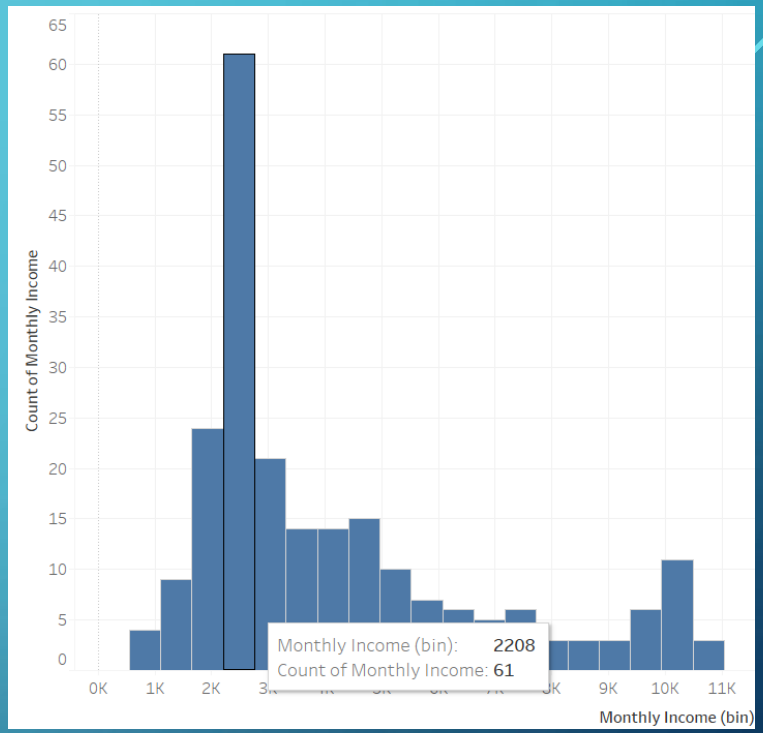
  OverTime, Department, Gender, EducationField, JobRole

	age	attrition	businesstravel	dailyrate	...	yearsatcompany	yearsincurrentrole	yearssincelastpromotion	yearswithcurrmanager
0	41	Yes	Travel_Rarely	1102	...	6	4	0	5
1	49	No	Travel_Frequently	279	...	10	7	1	7
2	37	Yes	Travel_Rarely	1373	...	0	0	0	0
3	33	No	Travel_Frequently	1392	...	8	7	3	0
4	27	No	Travel_Rarely	591	...	2	2	2	2

INSIGHTS FROM TABLEAU DASHBOARD

- OverTime workers had **3x higher attrition**
- Employees in **Sales Executive** role left most
- Employees with **low monthly income** left more
- Attrition rate ≈ **16%**
- One of the reason is work life balance.





"MACHINE LEARNING MODELS USED FOR ATTRITION PREDICTION"

- Model applied : -- Logistic Regression : Baseline model for binary classification

  -- Random Forest : Ensemble model with good accuracy

  -- XGBoost : Best performance + handles imbalance

- **Train/Test Split :** •Total Records: 1470
  •Split: 80% Train – 20% Test
  •Stratified on Attrition to maintain class balance

- **Performance matrix :**
- **87 % accuracy**

	precision	recall	f1-score	support
0	0.88	0.99	0.93	247
1	0.87	0.28	0.42	47
accuracy			0.88	294
macro avg	0.87	0.63	0.68	294
weighted avg	0.88	0.88	0.85	294

"KEY LEARNINGS FROM THE PROJECT"

- **Bullet Points:**
- Hands-on experience with **EDA, encoding, feature engineering**
- Built & compared ML models (Logistic, RF, XGBoost)
- Integrated **PostgreSQL** with **Python** for real-world data pipelines
- Designed **interactive dashboards in Tableau**
- Improved understanding of **model evaluation metrics**
- Learned how to present insights visually & analytically

CONCLUSION

- **Project Outcome:**
- Successfully predicted employee attrition using Machine Learning
- Achieved highest accuracy of **~89.6%** with **XGBoost** model
- Identified key attrition factors: **OverTime, MonthlyIncome, JobRole,** and **YearsAtCompany**
- Created an interactive **Tableau dashboard** for HR decision-makers

- **Business Impact:**
- Helps HR teams **identify high-risk employees early**
- Supports **data-driven retention strategies**
- Can reduce **attrition-related costs & losses**

"This project shows how data analytics can support smarter workforce planning in real life".

Thank You