

# HR Analytics

Job Satisfaction Analysis

# Problem Identification

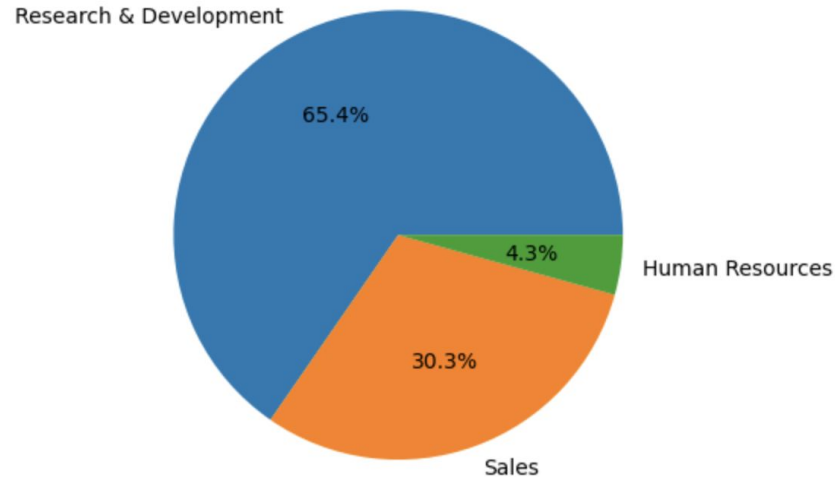
- ❖ Understand key factors influencing employee job satisfaction
  - Use a data driven approach to manage human resources
  - Build a machine learning model to predict employee job satisfaction
  - With the help of model understand which employees are not satisfied and take preventive action to reduce attrition

# Target feature

- ❖ Job Satisfaction is chosen as target feature
  - Contains four values: 1, 2, 3, 4
  - 1 and 2 classified as employee not satisfied
  - 3 and 4 classified as employee is satisfied

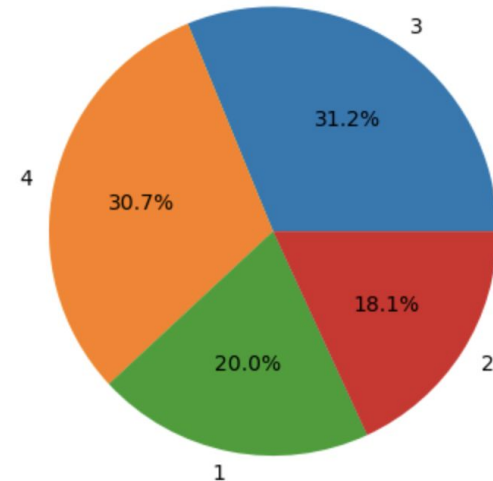
# Key findings

- ❖ Dataset include majority of employees from Research and development department



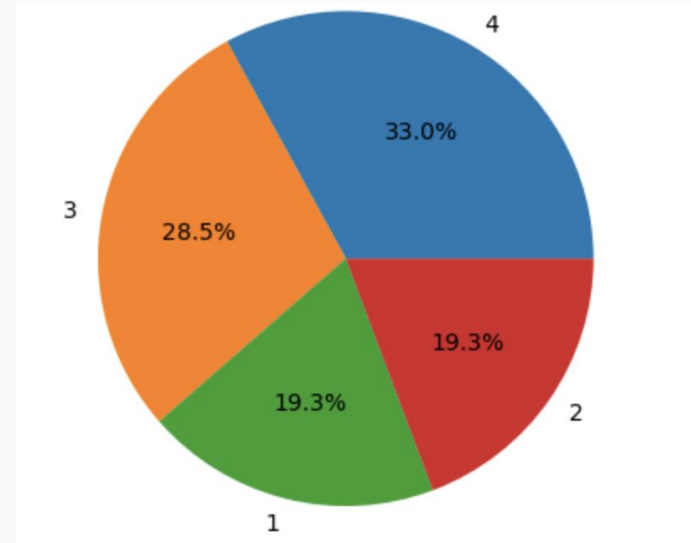
# Research and Development satisfaction rate

- ❖ 62% employees choose job satisfaction level 3 or 4, means they are satisfied
- ❖ 38% choose 1 or 2 level, which means they are not satisfied



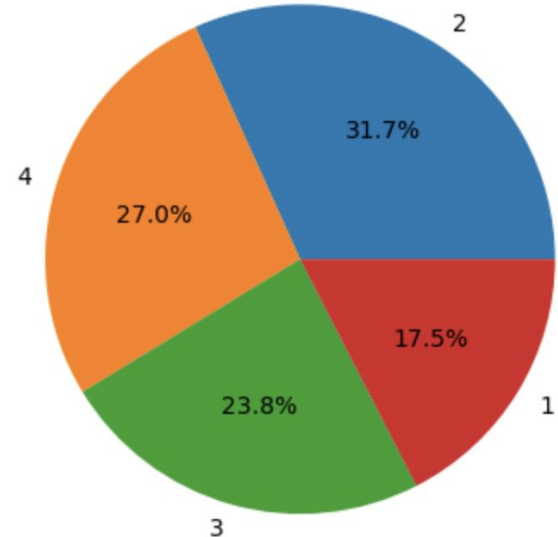
# Sales satisfaction rate

- ❖ 61% employees choose job satisfaction level 3 or 4, means they are satisfied
- ❖ 39% choose 1 or 2 level, which means they are not satisfied



# HR satisfaction rate

- ❖ 51% employees choose job satisfaction level 3 or 4, means they are satisfied
- ❖ 49% choose 1 or 2 level, which means they are not satisfied



# Modeling

- ❖ Approached problem in 2 ways
  - Linear regression, Random Regressor
    - Converted job satisfaction column to percentage to get continuous results
  - Logistic regression, KNN, Decision Tree, Random Forest
    - By Converting job satisfaction column to 2 classes 0 and 1



# Modeling Results

Model	Target Value	Precision	Recall	F1 Score	Support	Model Accuracy
Logistic Regression	0	0.45	0.18	0.25	142	0.60
	1	0.62	0.86	0.72	226	
KNN	0	0.42	0.28	0.34	142	0.57
	1	0.62	0.75	0.68	226	
Decision Tree	0	0.36	0.33	0.34	142	0.51
	1	0.60	0.62	0.61	226	
Random Forest	0	0.34	0.11	0.17	142	0.57
	1	0.61	0.86	0.71	226	

# Modeling Results

- ❖ Linear regression and random regressor performance not good
  - MAPE around 50%, means models results are not promising
- ❖ Logistic regression, KNN, Decision Tree, and Random Forest
  - show better accuracy score but still performance of models is not very good
  - Recall is very low
  - Means employee who are not satisfied are showing satisfied

# Modeling with resampling

- ❖ Imbalanced dataset with 0 values are 39% and 1 values are 61%
- ❖ Resampled with oversampling and undersampling techniques
- ❖ Logistic regression with random under sampling and XGBoost with over sampling showed better results
- ❖ Recall score improved to 50% to 51%

# Summary and Conclusion

- ❖ Logistic regression with random under sampling model has better accuracy score (52%) with highest recall score (51%)
- ❖ It showed good balance of false positives and false negatives

# Future steps

- ❖ Model results can be improved by
  - Choosing optimal threshold for the ROC Curve and Precision-Recall Curve directly
  - Or manually choosing threshold value for recall and precision based on decision better recall or precision is required
- ❖ Create UI interface for model and handover to HR team