**Employee Turnover Analytics Outline**

**Data Import & Quality Check** – Checks for missing values and fixes column names.

**EDA** – Plots:

* Correlation heatmap
* Distributions for satisfaction, evaluation, and hours
* Project count bar chart by turnover status

**KMeans Clustering** – Segments employees who left into 3 clusters by satisfaction and evaluation.

**Data Preprocessing** – Encodes categorical variables, performs a stratified train-test split, and applies SMOTE to fix imbalance.

**Model Training & Evaluation** – Trains three models using 5-fold CV:

* Logistic Regression
* Random Forest
* Gradient Boosting  
  Includes classification reports, confusion matrices, ROC curves, and AUC comparisons.

**Best Model Prediction** – Uses best model to predict turnover probabilities on test data.

**Risk Zones & Retention Suggestions**:

* Employees categorized into: Safe, Low-Risk, Medium-Risk, High-Risk
* Includes actionable retention strategies