Employee Attrition Analysis: From Data to Retention Strategies

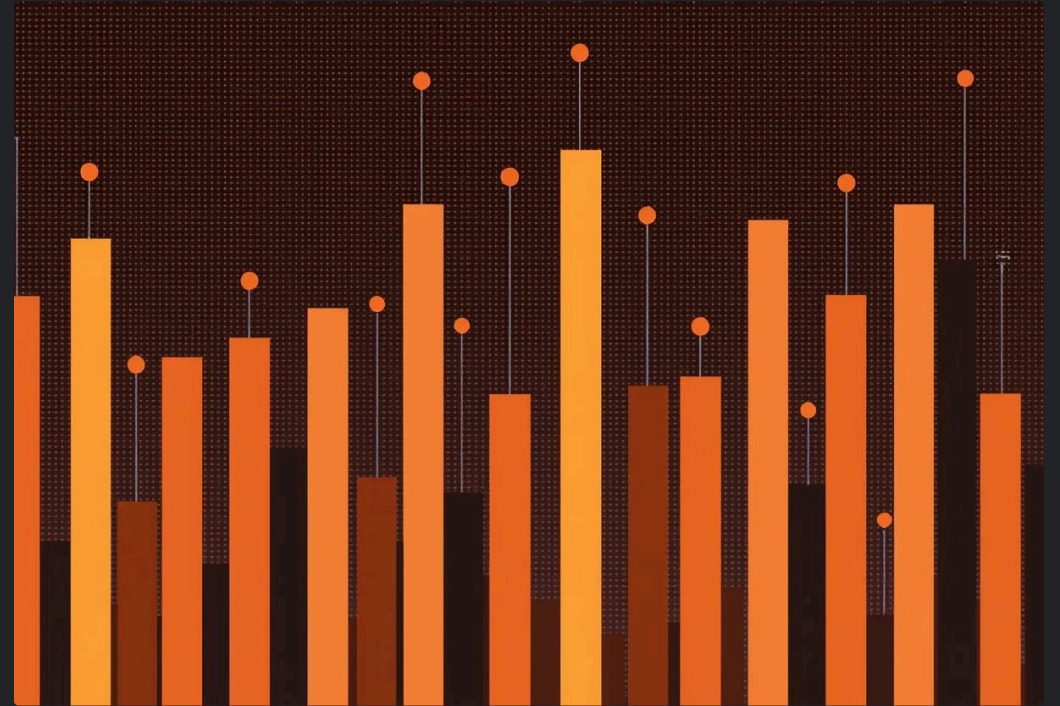Identifying Key Drivers and Mitigation Opportunities

Data Overview and Initial Exploration

Dataset Size

1470 entries, 35 columns.

Data Types

26 integer, 9 object columns.

Initial View

First 5 rows of the dataset.

Categorical Data Transformation

Binary Conversion

Attrition, Gender, Over18, OverTime.

One-Hot Encoding

BusinessTravel, Department, EducationField, JobRole, MaritalStatus.

Numerical Conversion

All dataframe values to numerical format.

Transformed Data Snapshot

After data preprocessing, the dataset now contains 53 columns, all in numerical format, ready for model training.

Age	41
Attrition	1 (Yes)
DailyRate	1102
DistanceFromHome	1
Education	2

Data Distribution Analysis

Histograms visualize the distribution of various features in the dataset, providing insights into data patterns and potential outliers.

Model Training and Performance
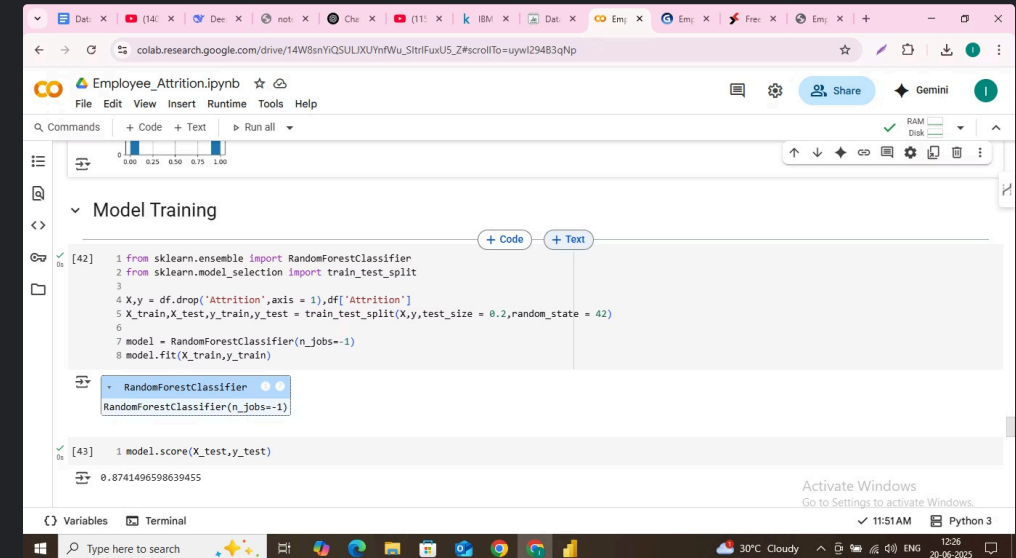
Model Used

RandomForestClassifier for attrition prediction.

Accuracy Score

87.41% correct predictions.

Interpretation

High accuracy indicates strong model performance.

Feature Importance Analysis

Identifying key factors influencing employee attrition.
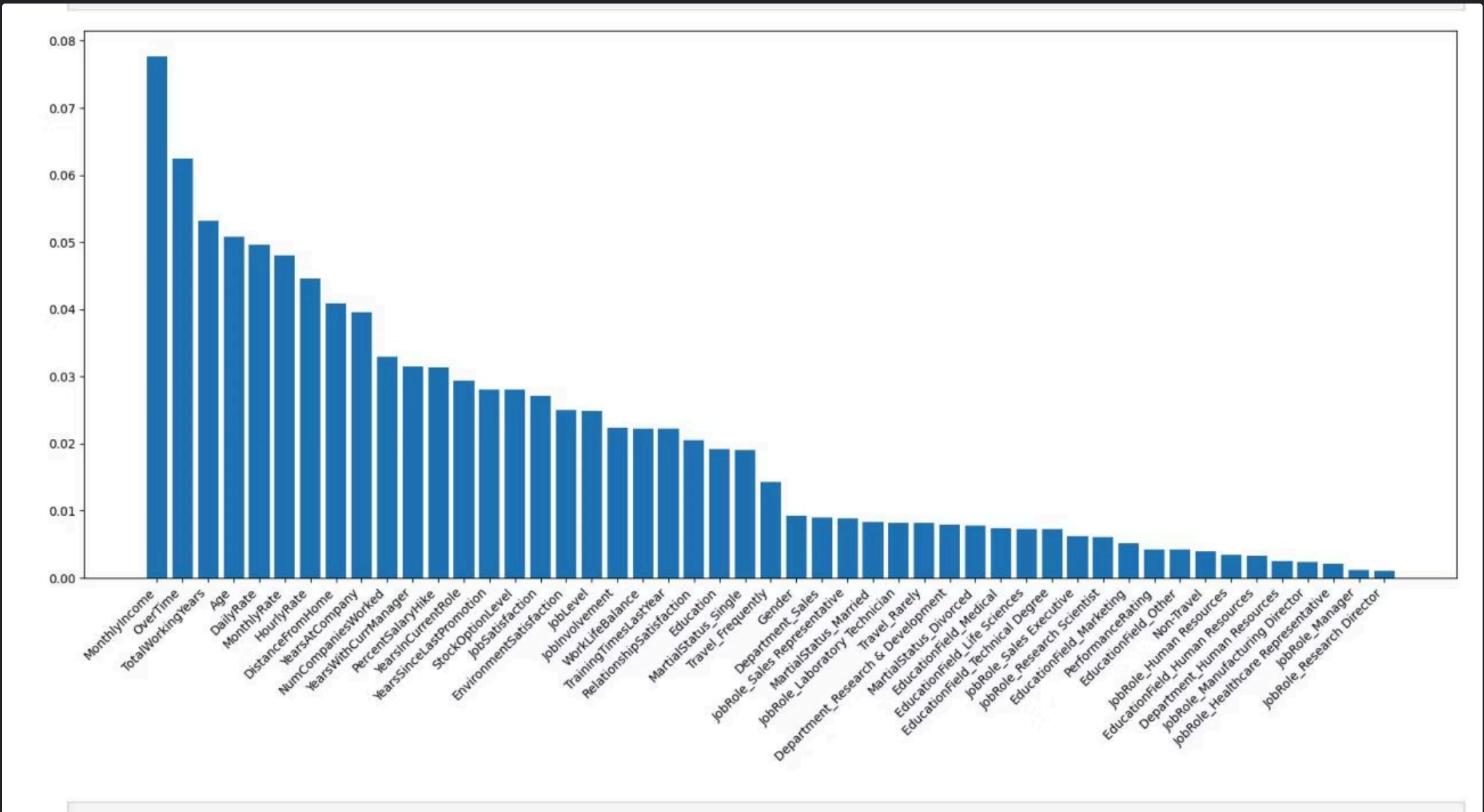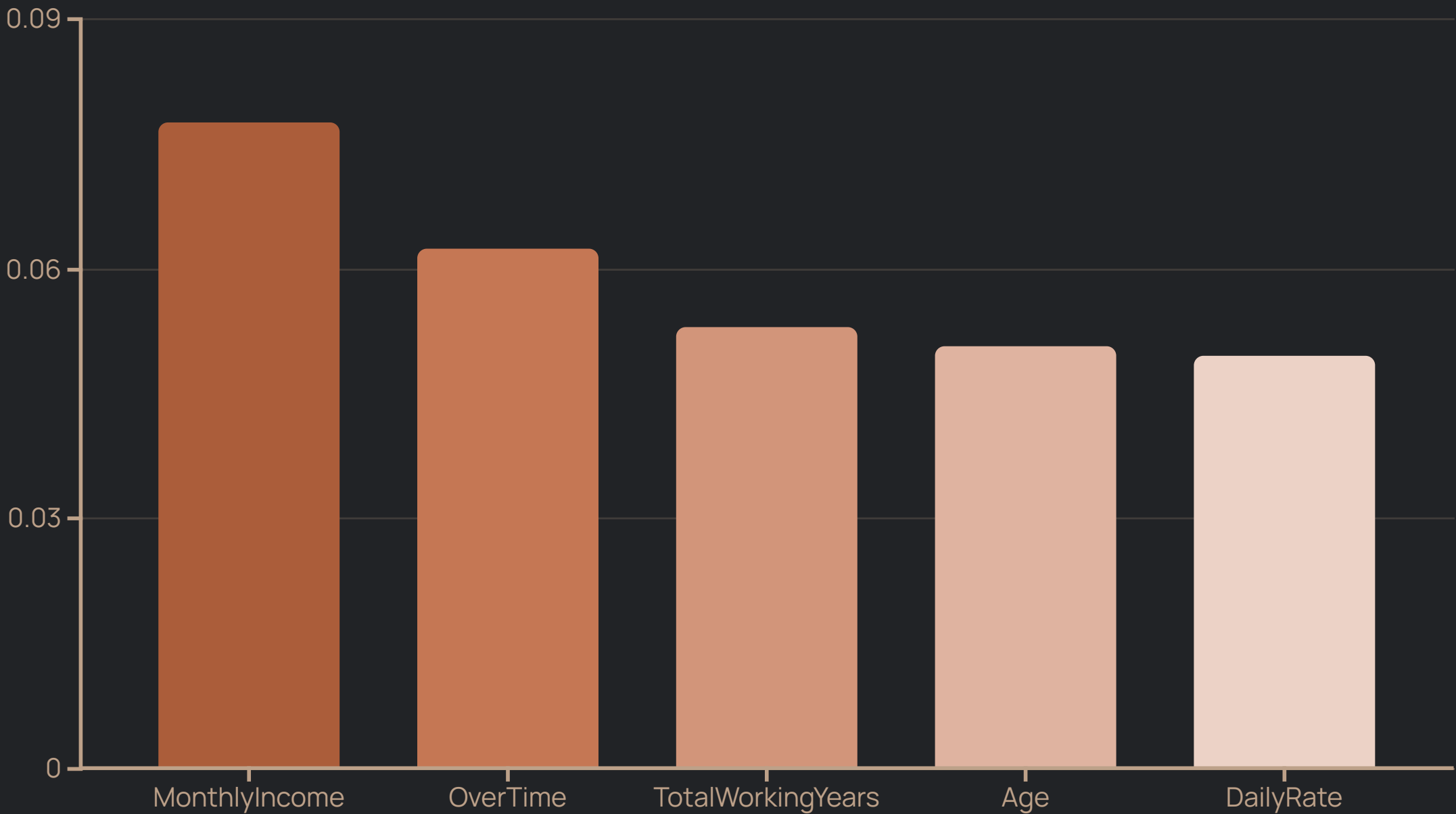
Key Insights and Actionable Strategies

Top Predictors
Monthly income and overtime are strongest.

Significant Factors
Work experience and job satisfaction matter.

Retention Focus
Address these areas to reduce attrition.