



# Employee Attrition Analysis: From Data to Retention Strategies

Identifying Key Drivers and Mitigation Opportunities

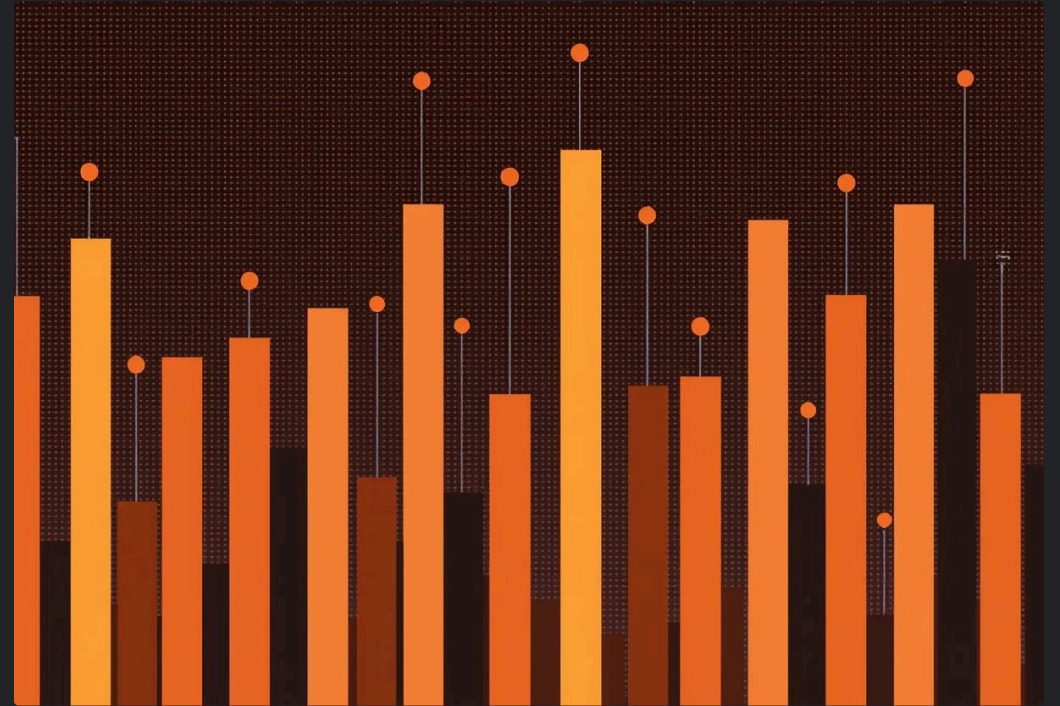
# Data Overview and Initial Exploration

## Dataset Size

1470 entries, 35 columns.

## Data Types

26 integer, 9 object columns.



## Initial View

First 5 rows of the dataset.



11111000000505555  
11111055136559553  
11111051100105515  
11110050156005506  
11111000006606055

# 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



# Model Training and Performance

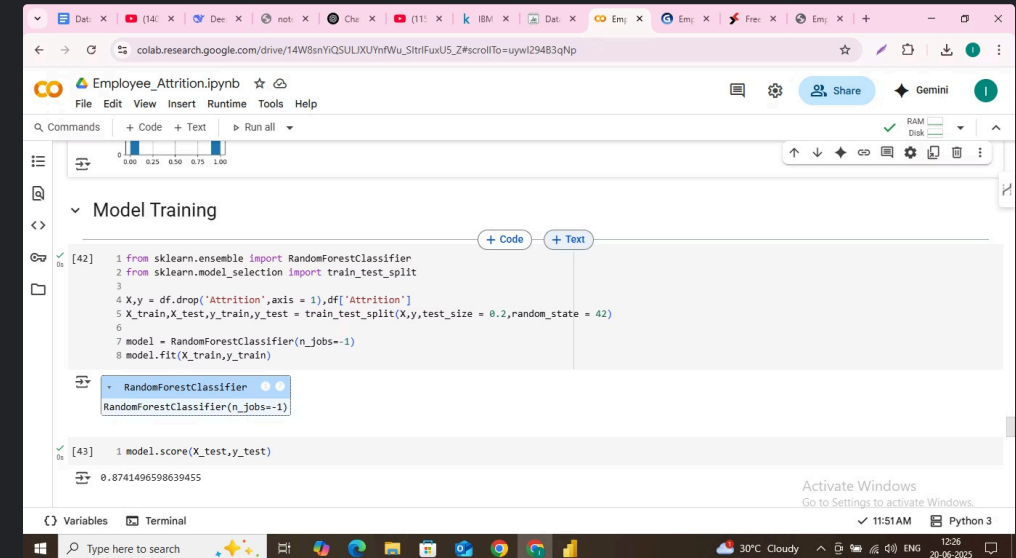
## Model Used

RandomForestClassifier for attrition prediction.

## Accuracy Score

87.41% correct predictions.

## Interpretation



The screenshot displays a Google Colab notebook titled 'Employee\_Attrition.ipynb'. The notebook contains two code cells. The first cell, labeled [42], imports the RandomForestClassifier and train\_test\_split functions from sklearn, drops the 'Attrition' column from the dataset, splits the data into training and testing sets, and trains the RandomForestClassifier model. The second cell, labeled [43], calculates the accuracy score of the trained model on the test set, resulting in a score of 0.8741496598639455. The notebook interface includes a file explorer on the left, a command bar at the top, and a status bar at the bottom showing the current time and system settings.

```
1 from sklearn.ensemble import RandomForestClassifier
2 from sklearn.model_selection import train_test_split
3
4 X,y = df.drop("Attrition",axis = 1),df["Attrition"]
5 X_train,X_test,y_train,y_test = train_test_split(X,y,test_size = 0.2,random_state = 42)
6
7 model = RandomForestClassifier(n_jobs=-1)
8 model.fit(X_train,y_train)

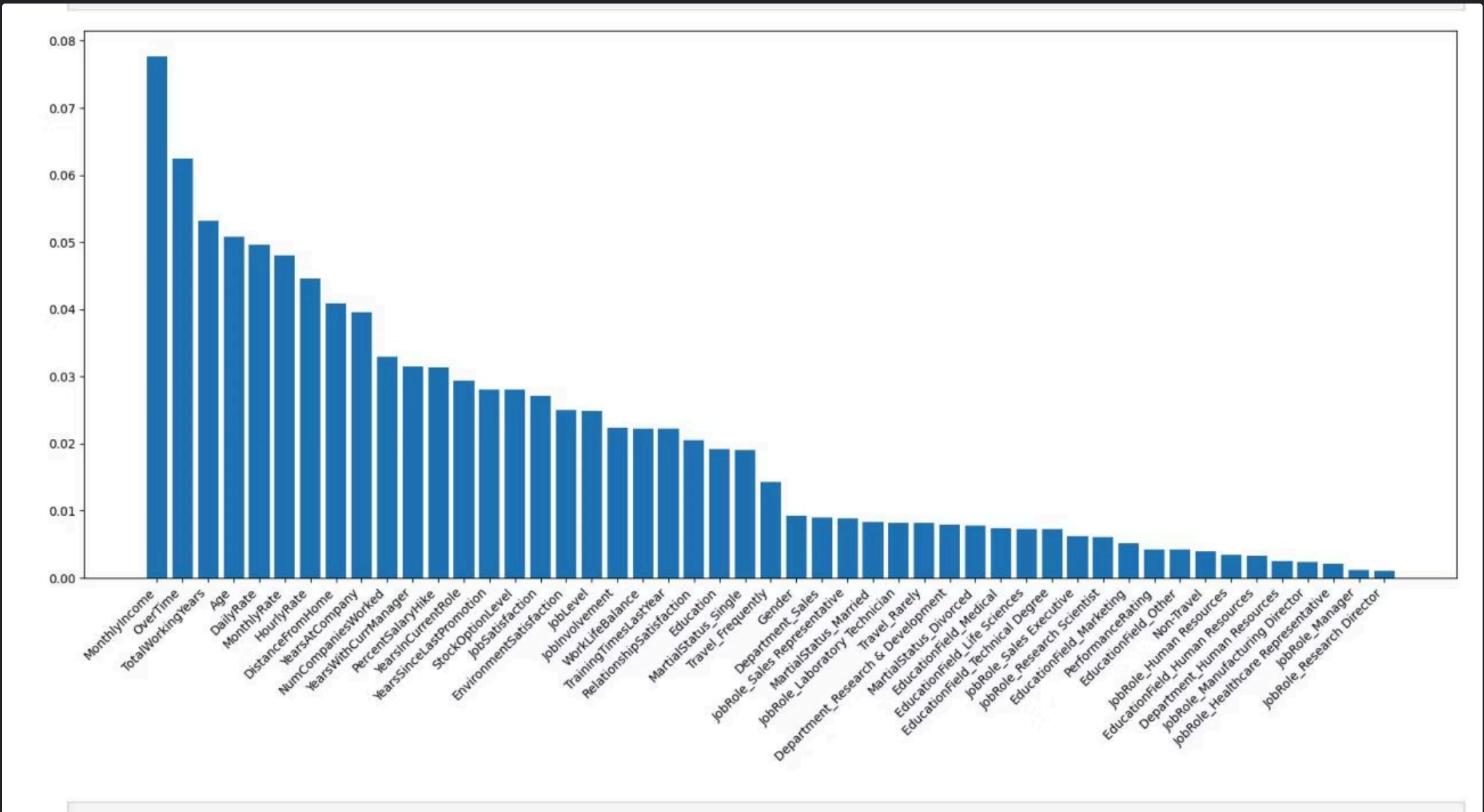
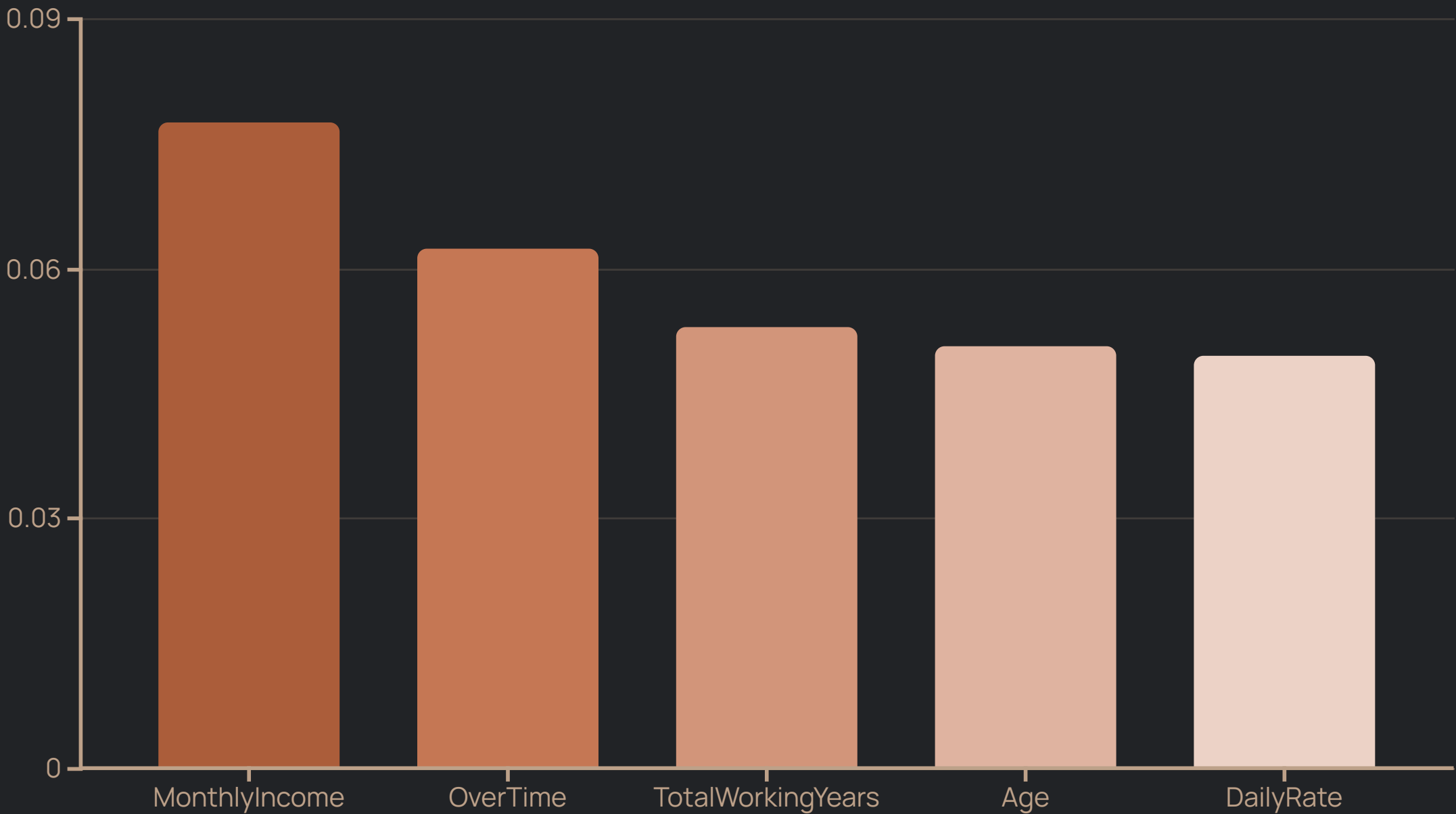
[42]

[43] 1 model.score(X_test,y_test)
      0.8741496598639455
```

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.

