Employee Attrition Analysis Report

Project Team: 22BEC025, 22BEC030, 22BEC049

Abstract:

This report presents a comprehensive analysis of employee attrition within a company, focusing on identifying key factors influencing turnover and providing actionable insights to improve employee retention. Using a dataset comprising demographic, job-related, and performance metrics for 14,900 employees, the study explores patterns in age, tenure, income, promotions, commute distance, job role, and department-specific factors. Data preprocessing included minimal cleaning, as the dataset was well-maintained with no missing values. To streamline the analysis, similar job roles were merged, resulting in broader categories like Healthcare, Education, Media, Technology, and Finance, enabling a more focused examination of retention trends.

Key findings reveal that younger employees, those with shorter tenure, lower income, and fewer promotions, as well as those in entry-level positions and certain departments (e.g., Education), exhibited higher attrition rates. Insights from the data indicate that commute distance and work-life balance also impact turnover. Visualizations provided deeper understanding, highlighting trends across departments and job levels.

Based on these insights, the report recommends implementing targeted retention strategies, including defined career advancement opportunities, flexible work options, and department-specific initiatives. This data-driven approach aims to enhance employee

satisfaction, reduce turnover, and foster a more stable and engaged workforce.

1. Project Overview

This report provides an extensive examination of employee attrition within a corporate setting, targeting the key variables that contribute to employee turnover. The study uses a dataset of 14,900 employees, covering various demographics, job attributes, and performance-related metrics.

Objective: The primary objective of the analysis is to uncover trends and identify specific factors that influence whether employees stay with or leave the company. The ultimate goal is to generate actionable insights that can guide the company in developing more effective retention strategies, improving employee satisfaction, and reducing turnover rates.

2. Data Summary:

The dataset is comprehensive, including several categories of attributes:

Demographic Information: This includes age, gender, and marital status. Such data can help determine if certain demographic groups are more prone to leaving, highlighting potential areas for demographic-specific retention initiatives.

Job-Related Information: This covers each employee's job role, level, monthly income, years with the company, and overall tenure. These variables are essential in identifying patterns like the impact of role and experience on attrition.

Performance and Satisfaction Metrics: This set of data points includes work-life balance, job satisfaction, and performance ratings. Satisfaction and performance metrics are often

closely tied to turnover, as employees with low satisfaction and performance ratings may be more likely to leave.Company Information: Includes company size, remote work options, leadership and innovation opportunities, and attrition status (lef

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14900 entries, 0 to 14899
Data columns (total 24 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Employee ID               14900 non-null  int64
 1   Age                       14900 non-null  int64
 2   Gender                    14900 non-null  object
 3   Years at Company          14900 non-null  int64
 4   Job Role                  14900 non-null  object
 5   Monthly Income            14900 non-null  int64
 6   Work-Life Balance         14900 non-null  object
 7   Job Satisfaction          14900 non-null  object
 8   Performance Rating        14900 non-null  object
 9   Number of Promotions      14900 non-null  int64
 10  Overtime                  14900 non-null  object
 11  Distance from Home        14900 non-null  int64
 12  Education Level           14900 non-null  object
 13  Marital Status            14900 non-null  object
 14  Number of Dependents      14900 non-null  int64
 15  Job Level                 14900 non-null  object
 16  Company Size              14900 non-null  object
 17  Company Tenure            14900 non-null  int64
 18  Remote Work               14900 non-null  object
 19  Leadership Opportunities  14900 non-null  object
 20  Innovation Opportunities  14900 non-null  object
 21  Company Reputation        14900 non-null  object
 22  Employee Recognition      14900 non-null  object
 23  Attrition                 14900 non-null  object
dtypes: int64(8), object(16)
memory usage: 2.7+ MB
None
```

t/stayed). These variables highlight organizational characteristics that may play a role in employee decisions to remain or exit.

3. Data Preprocessing and Cleaning

- Data Preprocessing and Cleaning
- Overview:
  In the process of preparing the data for analysis, only minimal data cleaning was required due to the high quality and maintenance of the dataset. No missing values were detected, meaning that each employee's data record was complete across all measured variables. This allowed for smoother and more reliable analysis, with no need for imputation or additional data validation. The cleanliness and completeness of the dataset reflect a well-maintained data collection and storage system, enhancing the accuracy and confidence in the analytical results.
- Job Role Merging:
  To streamline the analysis, specific job roles with overlapping skills or functions were merged into broader categories. This step was taken to simplify comparisons and focus on broader trends rather than getting caught in excessive granularity. It was found that certain roles, particularly those in support and technical services, had responsibilities and skill requirements that aligned closely. Merging these roles under a unified category allowed for more concise and actionable insights.
- For example:
- The "Support" and "IT" job roles, which both involved similar technical support functions, were grouped under a new "Technical" category. By combining these, the analysis could consider technical support functions as a whole without splitting insights across several narrower roles. This change also makes the findings more relevant to managers who oversee technical departments broadly and need aggregated insights for department-wide planning.

- This kind of role consolidation was specifically chosen to retain the distinct operational areas (like technology and support services) while simplifying the dataset for a more cohesive analysis. Each merged category still accurately reflects the underlying roles, with only minimal detail sacrificed.
- Post-Transformation Job Roles:
  After the role-merging process, the dataset was organized into the following unique job categories, which allowed for a more focused analysis:

```
#Cleaning Data

#Check for Missing Values
df.isnull().any()
```

```
Employee ID                False
Age                        False
Gender                     False
Years at Company           False
Job Role                   False
Monthly Income             False
Work-Life Balance          False
Job Satisfaction           False
Performance Rating         False
Number of Promotions       False
Overtime                   False
Distance from Home         False
Education Level            False
Marital Status             False
Number of Dependents       False
Job Level                  False
Company Size               False
Company Tenure             False
Remote Work                False
Leadership Opportunities   False
Innovation Opportunities   False
Company Reputation         False
Employee Recognition       False
Attrition                  False
dtype: bool
```

```
#Combine "technical","support" and "IT" into one department
df['Job Role']=np.where(df['Job Role'] =='support', 'technical', df['Job Role'])
df['Job Role']=np.where(df['Job Role'] =='IT', 'technical', df['Job Role'])

#After combining the department looks like
df['Job Role'].unique()
```

```
array(['Healthcare', 'Education', 'Media', 'Technology', 'Finance'],
      dtype=object)
```

4. Key Observations on Employee Attrition

To understand the factors impacting employee attrition, we compared several metrics between employees who left and those who stayed. Below are the key findings:

```
#Now Let us find out how many employees left and stayed
df['Attrition'].value_counts()
```

```
Attrition
Stayed    7868
Left      7032
Name: count, dtype: int64
```

To investigate the reasons behind employee attrition, several factors were compared between employees who left and those who stayed. Here are the primary findings:

- **Age**: Employees who left were, on average, younger (37.6 years) than those who stayed (39.1 years). This difference may indicate that younger employees seek more dynamic or growth-oriented environments and may leave if their expectations are unmet.
- **Years at Company**: Tenure appeared to impact attrition, with those staying having an average tenure of 16.3 years versus 14.8 years for those who left. Longer tenure could correlate with job stability and satisfaction, suggesting that employees feel more invested in the company as they accumulate experience.
- **Monthly Income**: Employees who stayed earned a slightly higher average income ($7303) than those who left ($7269). Though the difference is minor, this factor could contribute

to satisfaction and loyalty, with financial stability playing a role in retention.

- **Number of Promotions**: Employees who stayed received more promotions on average (0.91 promotions) than those who left (0.75 promotions). This suggests that opportunities for advancement could be influential in motivating employees to remain within the company.
- **Distance from Home**: Those who left lived farther away from the company, with an average commute distance of 52.57 units compared to 47.57 units for those who stayed. Commute time and distance may have a psychological impact, affecting work-life balance and convenience, thus influencing attrition.
- **Dependents**: Employees who stayed had, on average, more dependents (1.75) than those who left (1.56), hinting that individuals with family responsibilities might value job stability and are less inclined to leave.

```python
#We will see the various observation of left employee by mean
def mean_by_attrition(df):
    grouped = df.groupby('Attrition')
    mean_values = {}

    for col in df.columns:
        try:
            mean_values[col] = grouped[col].mean()
        except TypeError:
            pass
            #print(f"Skipping non-numeric column: {col}")
```

)ownloads/Untitled (1).html


```python
    return pd.DataFrame(mean_values)

mean_values_df = mean_by_attrition(df)
print(mean_values_df)
#We Have used here try and except method as it was necessary because there are f
```

```
           Employee ID        Age  Years at Company  Monthly Income  \
Attrition
Left       37316.284556  37.636803         14.798493     7269.116468
Stayed     37359.343671  39.054143         16.301983     7303.562913

           Number of Promotions  Distance from Home  Number of Dependents  \
Attrition
Left                   0.746303           52.566837              1.562002
Stayed                 0.913066           47.568251              1.746314

           Company Tenure
Attrition
Left            54.968003
Stayed          56.171708
```

5. Department and Job Level Analysis

Analyzing attrition by department and job level provided additional insights:

Departmental Insights:

The Education department had the highest employee count but also faced higher attrition rates. This department generally had lower average income, which may correlate with higher turnover if employees are dissatisfied with compensation.

The Finance and Technology departments had higher average incomes and comparatively lower attrition rates. This could indicate that competitive pay and career growth opportunities in these sectors contribute to higher retention.

Job Level Insights:

Entry-level employees exhibited slightly lower retention rates than those in mid- or senior-level roles, likely due to fewer career advancement opportunities and the possibility of employees in junior roles exploring different career paths early on.

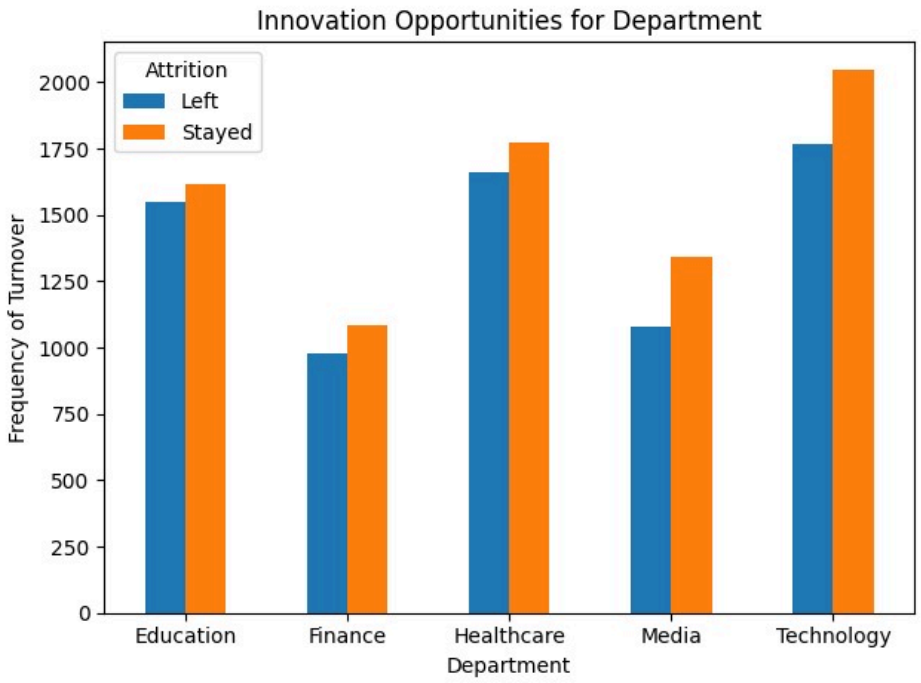
```python
#We will see the various Job level observation by mean
def mean_by_department(df):
    grouped = df.groupby('Job Level')
    mean_values_l = {}

    for col in df.columns:
        try:
            mean_values_l[col] = grouped[col].mean()
        except TypeError:
            pass
            #print(f"Skipping non-numeric column: {col}")

    return pd.DataFrame(mean_values_l)

mean_values_df_l = mean_by_department(df)
print(mean_values_df_l)
#We Have used here try and except method as it was necessary because there are f
```

```
             Employee ID        Age  Years at Company  Monthly Income  \
Job Level
Entry       37109.669034  38.202604         15.478099     7266.856249
Mid         37750.901097  38.494852         15.797806     7282.992068
Senior      36984.933703  38.525800         15.415741     7335.144024

            Number of Promotions  Distance from Home  Number of Dependents  \
Job Level
Entry                   0.831896           49.820058              1.651784
Mid                     0.844726           50.056709              1.660928
Senior                  0.819073           49.884063              1.670803

            Company Tenure
Job Level
Entry           55.470658
Mid             55.973165
Senior          55.145330
```

6. Visualization Insights

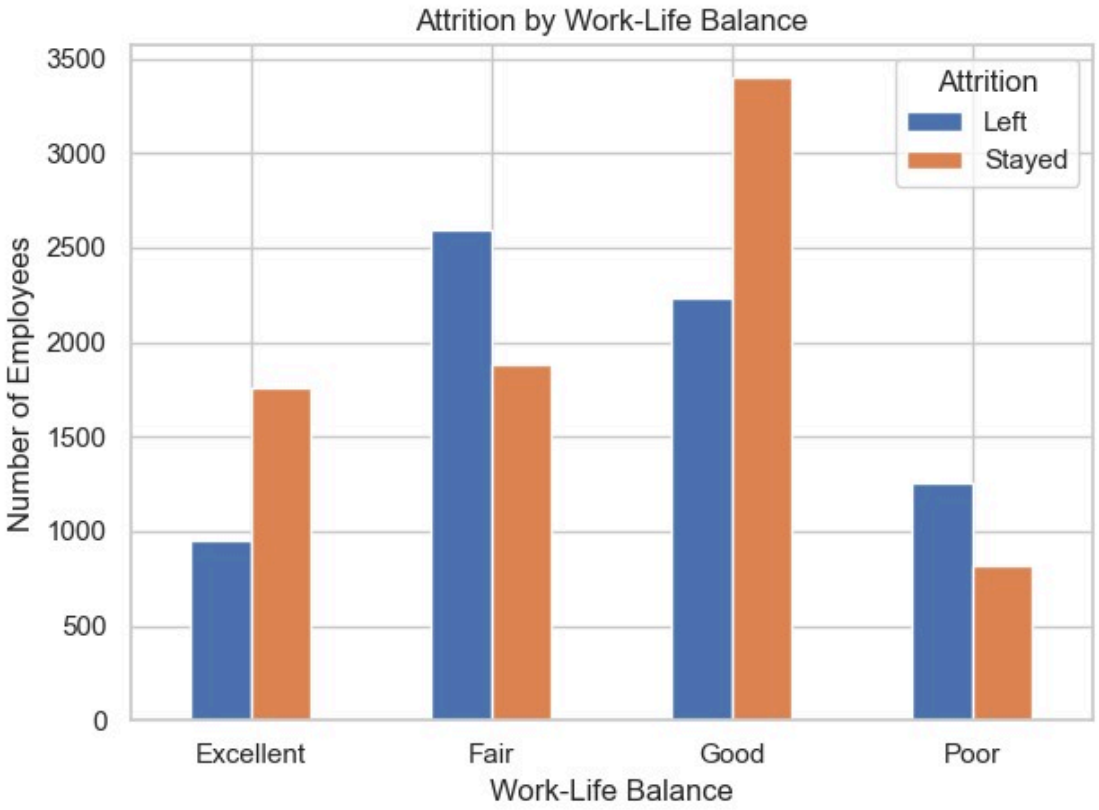
The following visualizations were created to aid in understanding attrition trends:

- **Department vs. Attrition**: A bar chart shows turnover rates across departments, indicating higher attrition in certain departments such as Education.

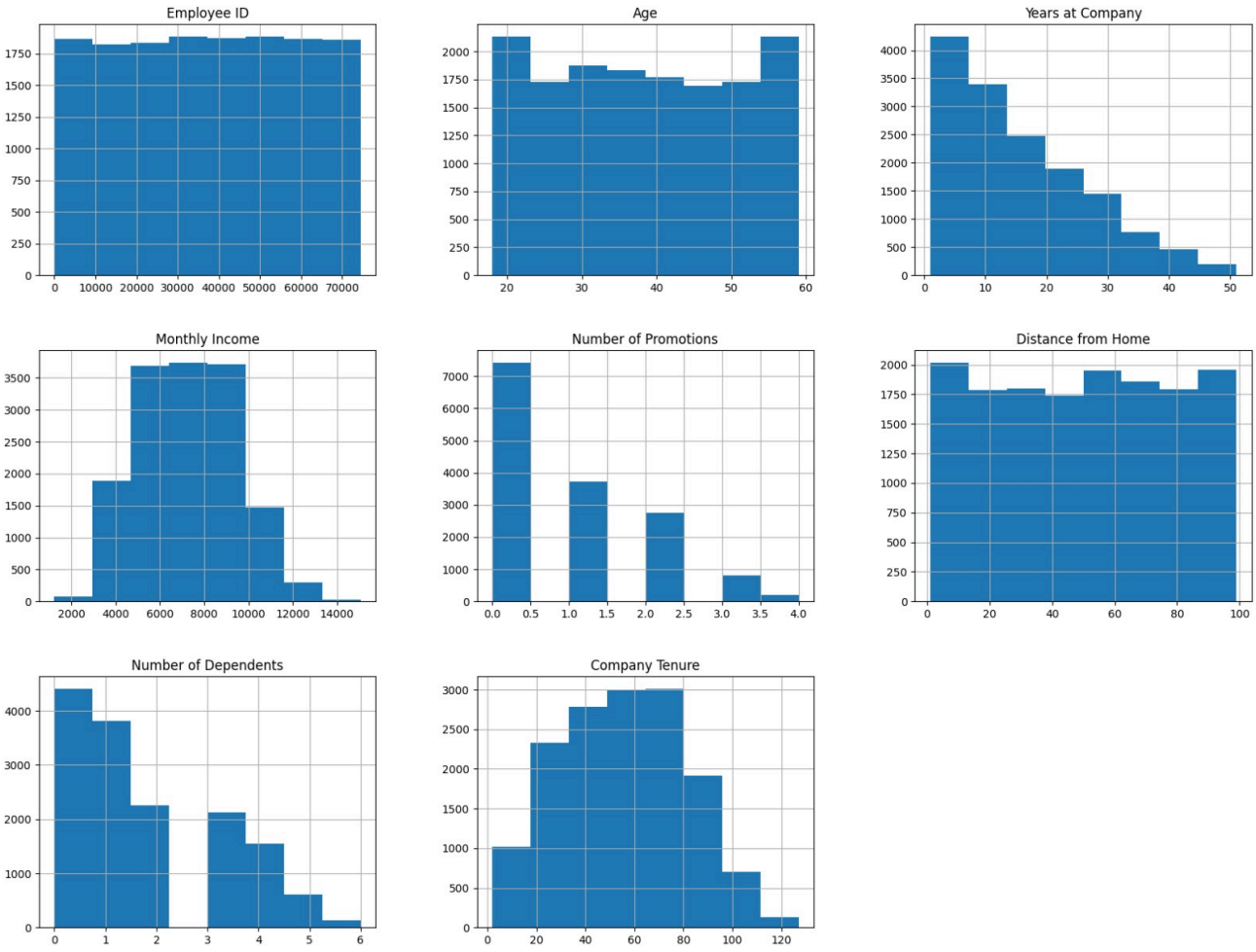


- **Income vs. Attrition**: A stacked bar chart compares turnover rates across income levels, revealing that higher income correlates with slightly lower attrition.

- **Attrition Trends by Work-Life Balance**: The visualization allows us to see if there's a higher proportion of employees who left among those with lower work-life balance ratings (e.g., Poor or Fair), compared to those with higher ratings (e.g., Good or Excellent).
- **Retention Factors**: If the chart shows a significantly higher attrition rate in the Poor or Fair categories, it suggests that work-life balance may be a significant factor in retention. Conversely, if attrition remains low in the Good and Excellent categories, it supports the idea that good work-life balance enhances retention.



- **Histograms**: Numeric variables such as age, years at the company, monthly income, and distance from home are plotted in histograms to display their distributions and assist in identifying trends.



7. Recommendations

- Based on the analysis, several strategies are recommended to improve employee retention:
    - **Promotion Opportunities**: Create and communicate clear career advancement pathways, especially for younger employees, to foster a sense of growth and purpose within the company.
    - **Flexible Work Options**: Provide options for remote work or flexible hours, particularly for employees who live farther from the office, to support work-life balance and reduce turnover linked to commuting difficulties.
    - **Department-Specific Retention Programs**: Develop tailored retention strategies for departments with higher attrition rates. For example, in the Education department, which has lower average income and higher turnover, introducing targeted incentives or career development programs might improve retention.
    - **Salary Adjustments**: Given the slight correlation between income and retention, reviewing and possibly adjusting salaries to be more competitive, especially in departments or roles with higher turnover, could improve job satisfaction and reduce attrition.

8. Conclusion

This analysis reveals various factors influencing employee attrition. Age, tenure, income, and opportunities for promotion appear to have significant impacts on whether employees stay with or leave the company. Additionally, factors such as work-life balance, commute distance, and department-specific challenges contribute to turnover.

Implementing data-driven retention strategies that address these factors could significantly enhance employee satisfaction and reduce attrition rates over time, leading to a more stable and engaged workforce.

This elaboration provides a deeper understanding of the report’s contents and the actionable insights derived from the analysis. Let me know if there's a particular section you'd like to explore in even greater detail.