CSV Profile Report

AI-Driven Insights

Key Findings:

- Data Quality Assessment:

Overall data completeness is 95.3%, with 1,233 missing values across 1 columns.

- Dataset Dimensions and Structure:

The dataset contains 1,470 records with 18 attributes, providing a comprehensive view of the data.

- Primary Key Analysis:

Identified 1 potential primary key(s): Employee Number (1,470 unique values). These columns have unique values for each record and can be used as reliable identifiers.

- Strong positive correlation (0.77) between Percent Salary Hike and Performance Rating:

This indicates that as Percent Salary Hike increases, Performance Rating tends to increase proportionally.

- Trend Analysis for Employee Number:

Shows a increasing trend with normal distribution. Mean value is 1024.87 with a standard deviation of 602.02.
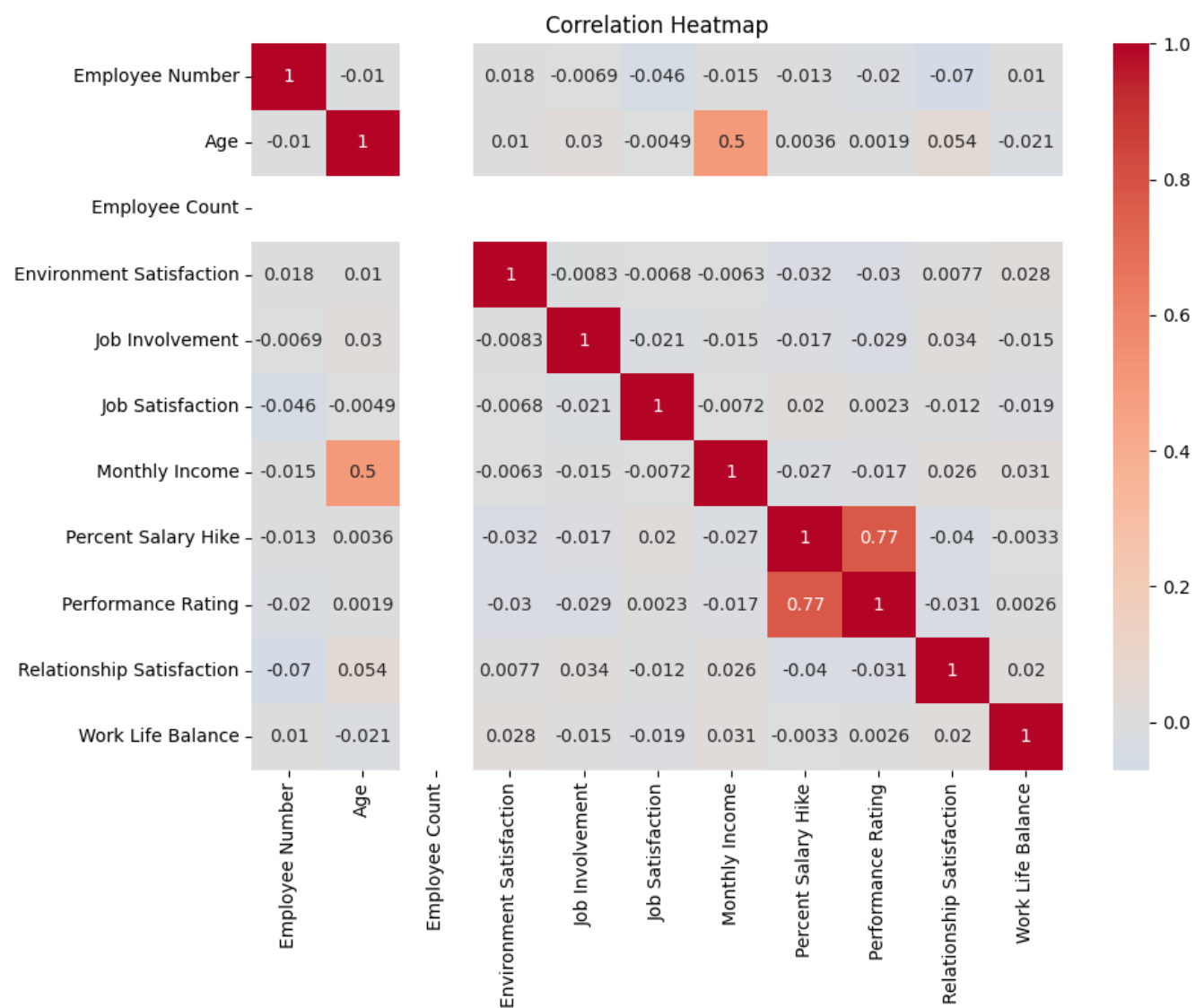
- Data Type Distribution:

The dataset contains 11 numeric columns and 7 categorical columns, suggesting a predominantly numeric dataset.

- Outlier Detection:

Found potential outliers in the following numeric columns: Monthly Income (114 outliers), Performance Rating (226 outliers). These may require further investigation or special handling.

Correlation Analysis:

Correlation Heatmap

	Employee Number	Age	Employee Count	Environment Satisfaction	Job Involvement	Job Satisfaction	Monthly Income	Percent Salary Hike	Performance Rating	Relationship Satisfaction	Work Life Balance
Employee Number	1	-0.01		0.018	-0.0069	-0.046	-0.015	-0.013	-0.02	-0.07	0.01
Age	-0.01	1		0.01	0.03	-0.0049	0.5	0.0036	0.0019	0.054	-0.021
Employee Count											
Environment Satisfaction	0.018	0.01		1	-0.0083	-0.0068	-0.0063	-0.032	-0.03	0.0077	0.028
Job Involvement	-0.0069	0.03		-0.0083	1	-0.021	-0.015	-0.017	-0.029	0.034	-0.015
Job Satisfaction	-0.046	-0.0049		-0.0068	-0.021	1	-0.0072	0.02	0.0023	-0.012	-0.019
Monthly Income	-0.015	0.5		-0.0063	-0.015	-0.0072	1	-0.027	-0.017	0.026	0.031
Percent Salary Hike	-0.013	0.0036		-0.032	-0.017	0.02	-0.027	1	0.77	-0.04	-0.0033
Performance Rating	-0.02	0.0019		-0.03	-0.029	0.0023	-0.017	0.77	1	-0.031	0.0026
Relationship Satisfaction	-0.07	0.054		0.0077	0.034	-0.012	0.026	-0.04	-0.031	1	0.02
Work Life Balance	0.01	-0.021		0.028	-0.015	-0.019	0.031	-0.0033	0.0026	0.02	1

File Overview

Filename: HR_Attrition (Main)_HR_Attrition.csv

File Size: 279.5 kB

Rows: 1470

Columns: 18

Missing Cells: 1233

Duplicate Rows: 0
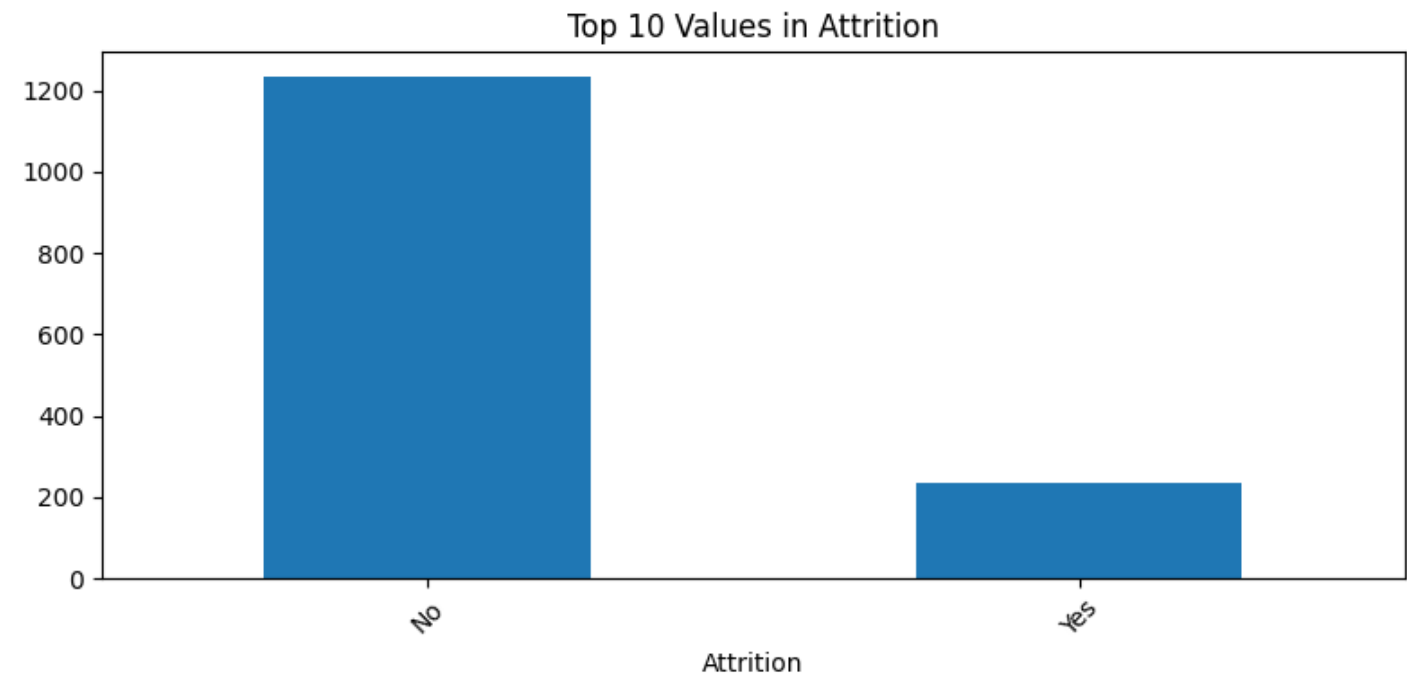
Column Analysis

Column: Attrition

Type: object

Missing Values: 0

Unique Values: 2

Top Values:

No: 1233

Yes: 237

Column: Attrition Date

Type: object

Missing Values: 1233

Unique Values: 116

Top Values:

4/30/22: 8

4/17/22: 7

4/4/22: 7

1/16/22: 6

4/29/22: 6

3/11/22: 5

2/11/22: 5

3/4/22: 5

4/13/22: 5

4/25/22: 5

Column: Department

Type: object

Missing Values: 0

Unique Values: 3

Top Values:

R & D: 961

Sales: 446

HR: 63

Top 10 Values in Department
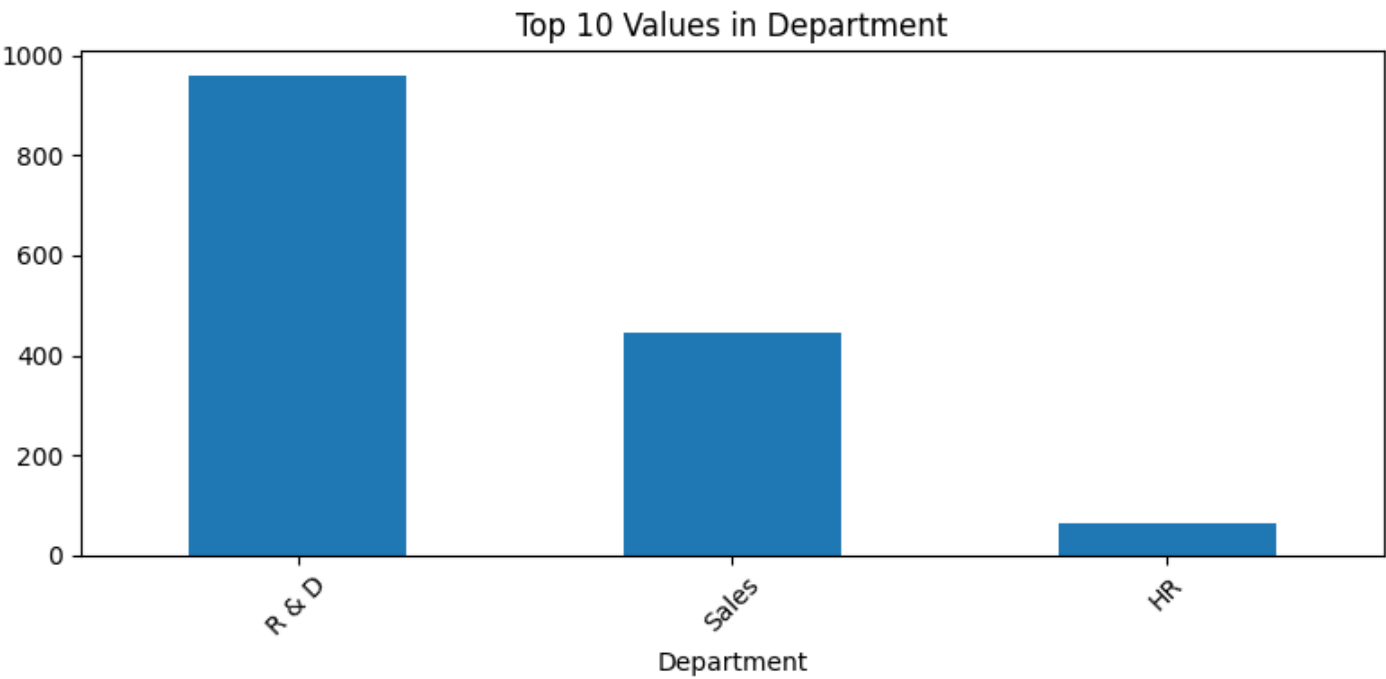
Column: Education

Type: object

Missing Values: 0

Unique Values: 5

Top Values:

Bachelor's Degree: 572

Master's Degree: 398

Associates Degree: 282

High School: 170

Doctoral Degree: 48
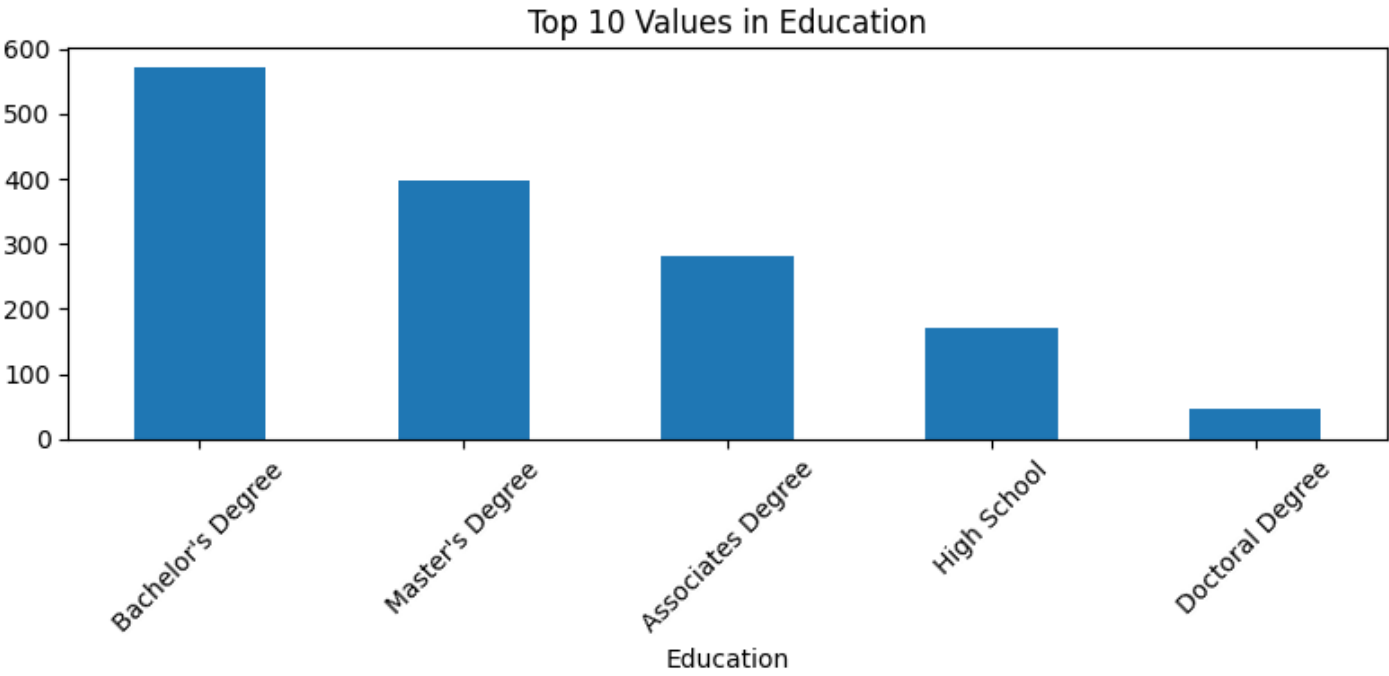
Column: Education Field

Type: object

Missing Values: 0

Unique Values: 6

Top Values:

Life Sciences: 606

Medical: 464

Marketing: 159

Technical Degree: 132

Other: 82

Human Resources: 27

Top 10 Values in Education Field
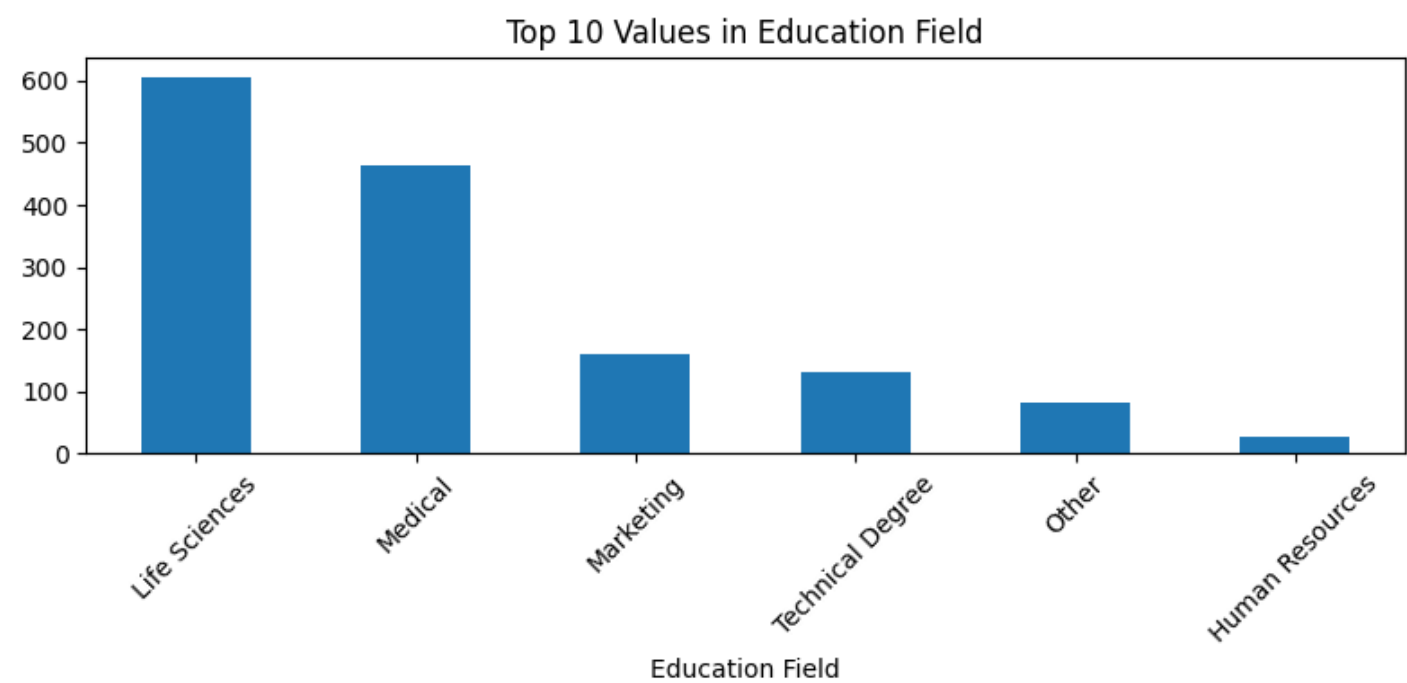
Column: Employee Number

Potential Primary Key

Type: int64

Missing Values: 0

Unique Values: 1470

Mean: 1024.87

Standard Deviation: 602.02

Min: 1

Max: 2068

25th Percentile: 491.25

Median: 1020.50

75th Percentile: 1555.75

Trend Analysis:

Trend: Increasing

Distribution: Normal

Distribution of Employee Number
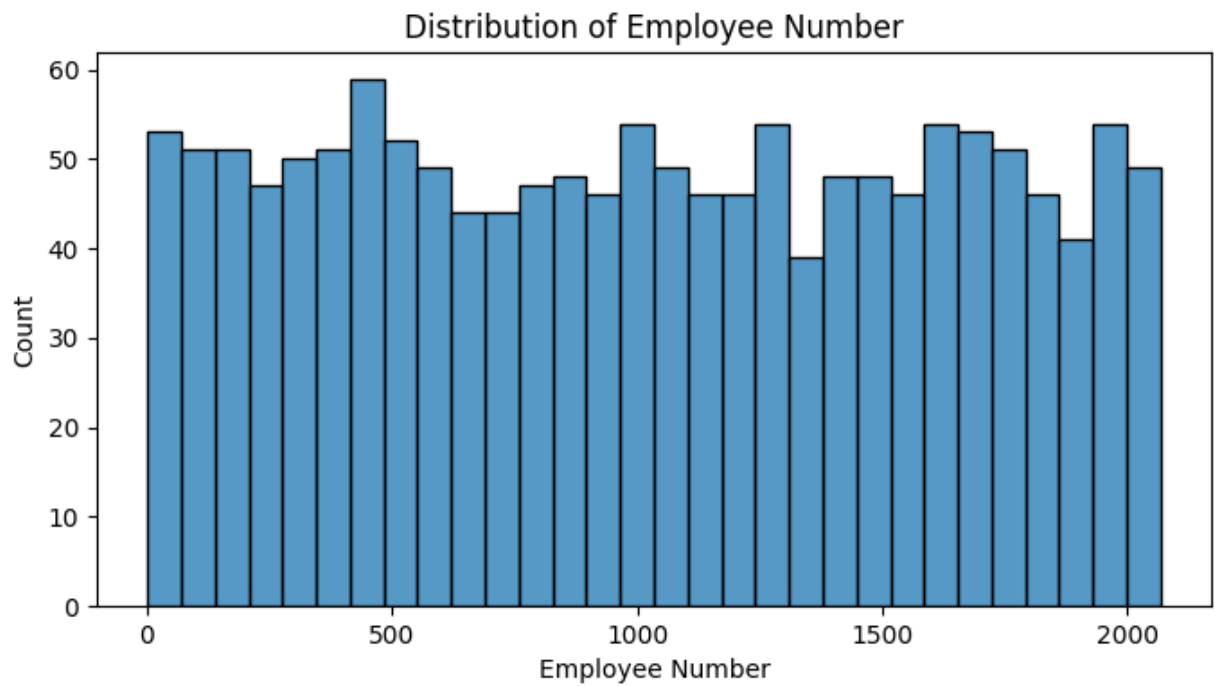
Column: Gender

Type: object

Missing Values: 0

Unique Values: 2

Top Values:

Male: 882

Female: 588

Top 10 Values in Gender
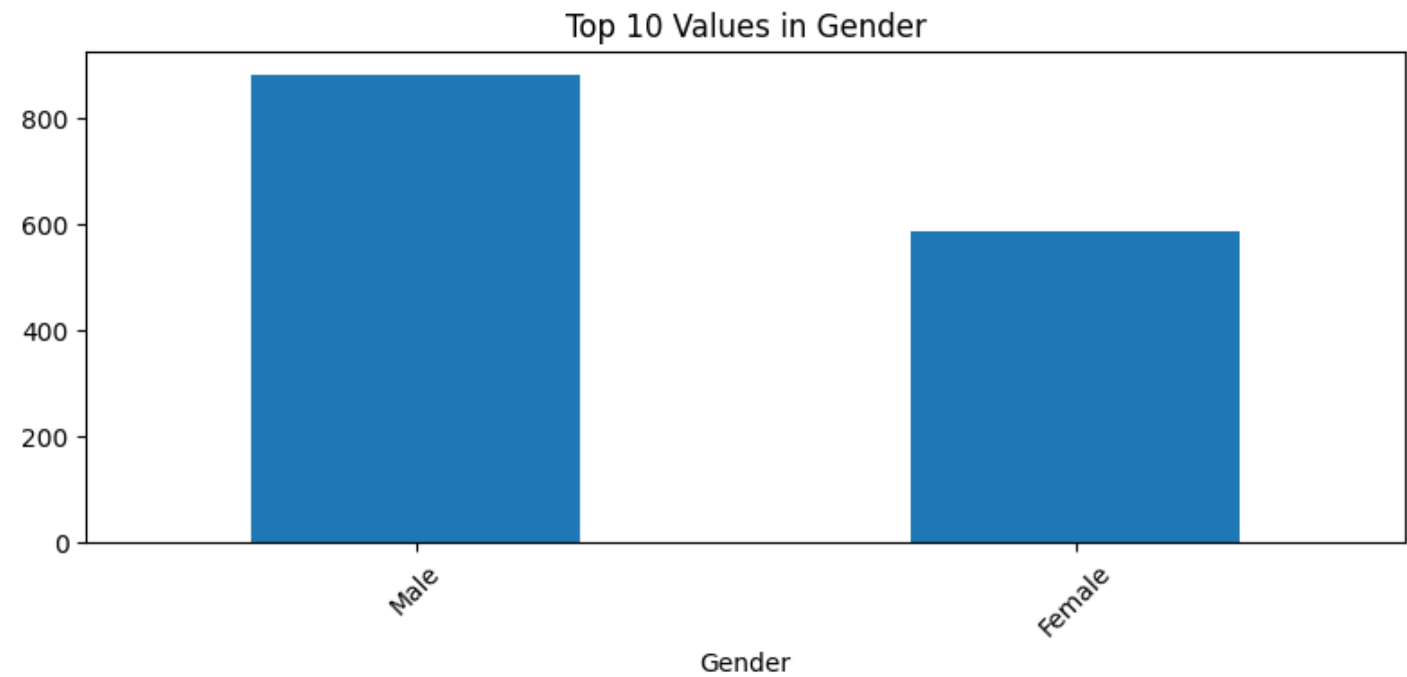
Column: Job Role

Type: object

Missing Values: 0

Unique Values: 9

Top Values:

Sales Executive: 326

Research Scientist: 292

Laboratory Technician: 259

Manufacturing Director: 145

Healthcare Representative: 131

Manager: 102

Sales Representative: 83

Research Director: 80

Human Resources: 52
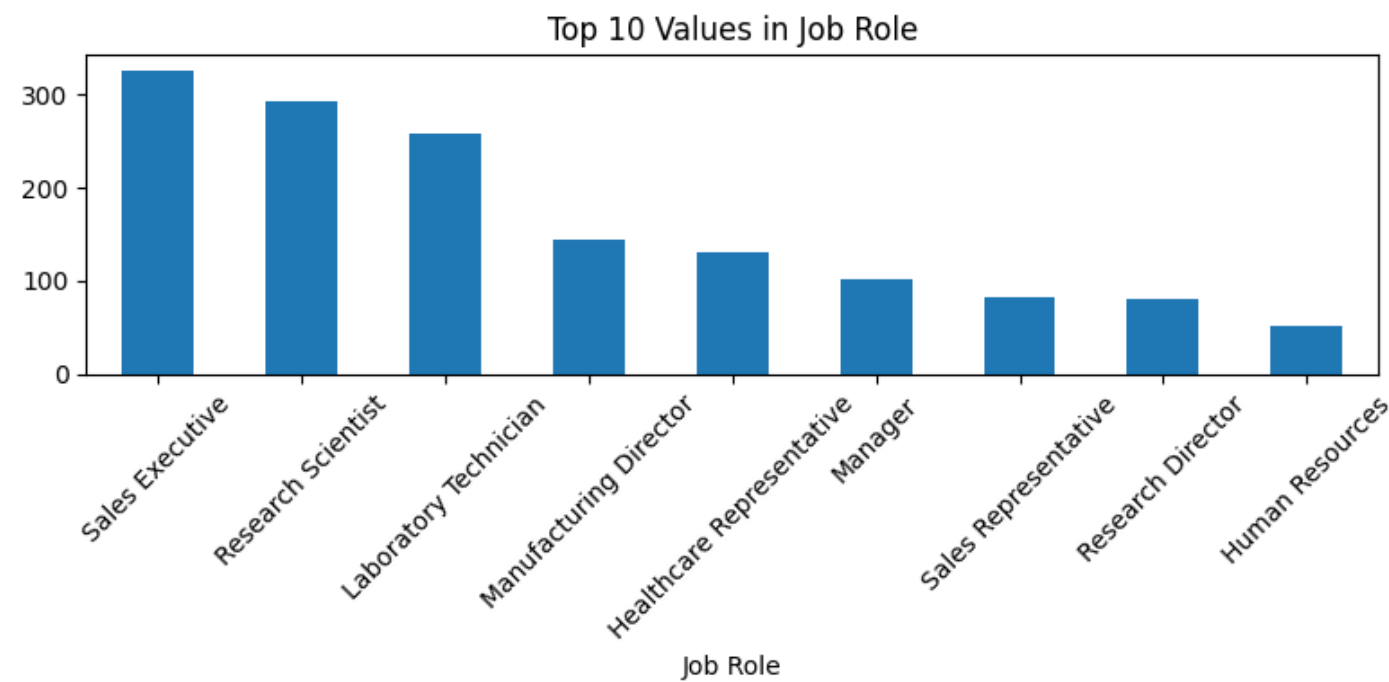
Column: Age

Type: int64

Missing Values: 0

Unique Values: 43

Mean: 36.92

Standard Deviation: 9.14

Min: 18

Max: 60

25th Percentile: 30.00

Median: 36.00

75th Percentile: 43.00

Trend Analysis:

Trend: Stable

Distribution: Normal

Distribution of Age
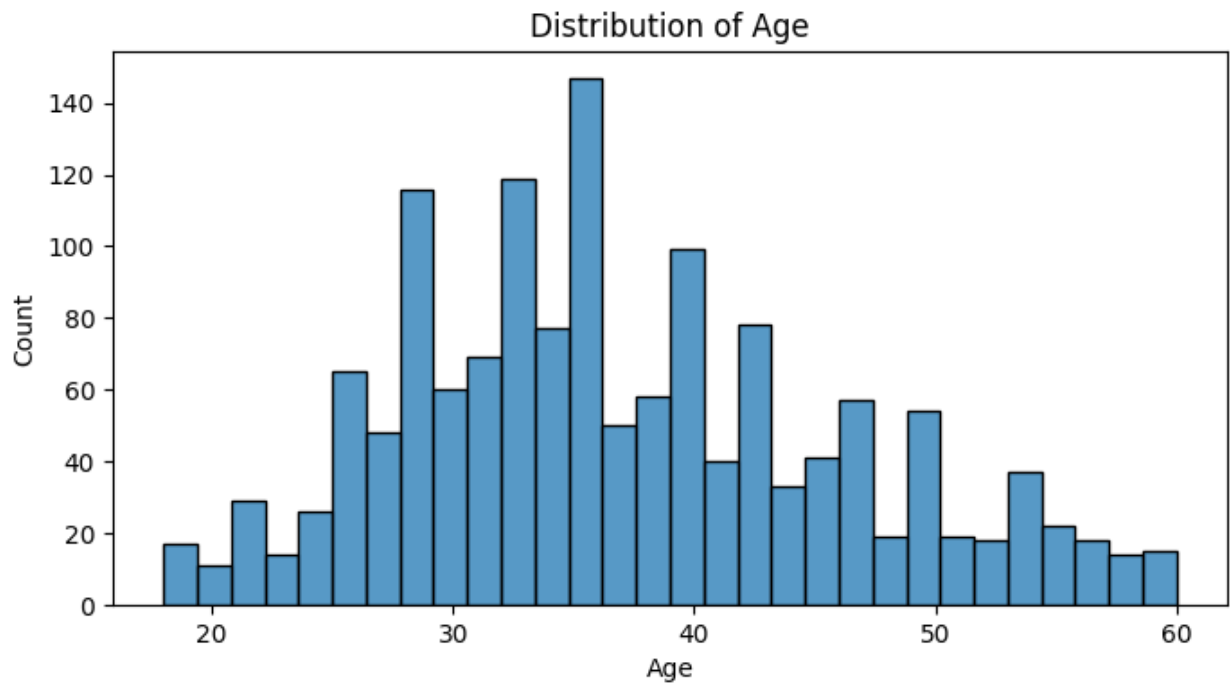
Column: Employee Count

Type: int64

Missing Values: 0

Unique Values: 1

Mean: 1.00

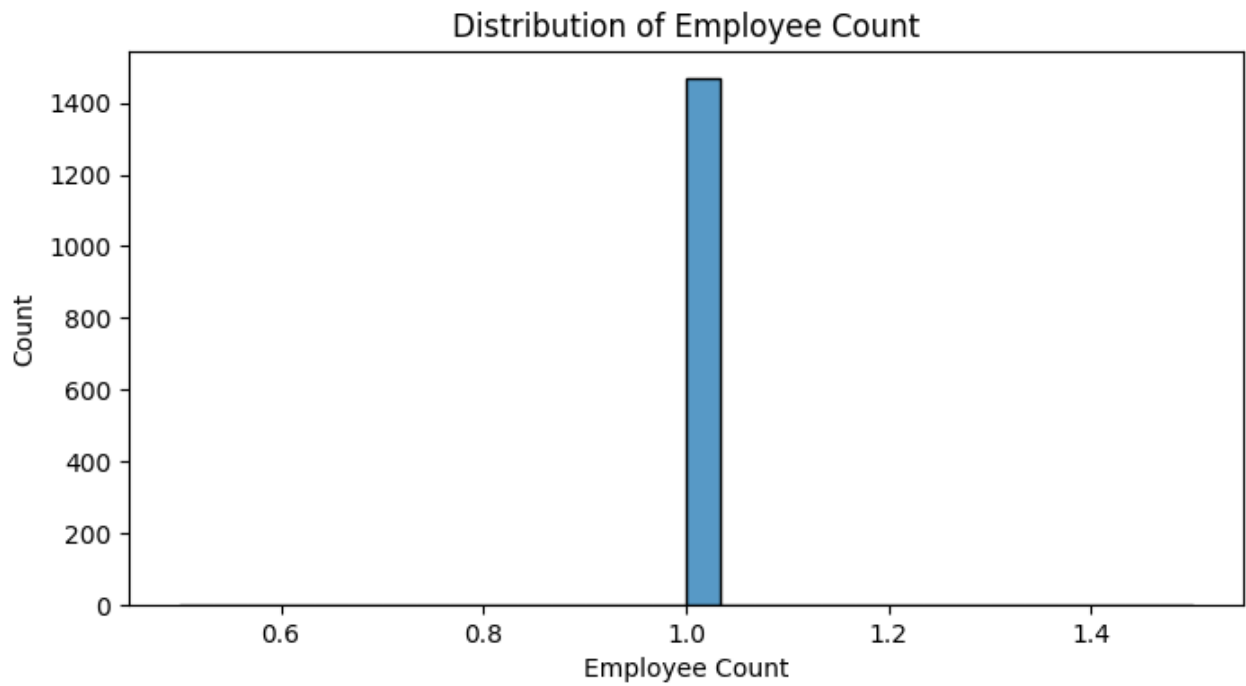Standard Deviation: 0.00

Min: 1

Max: 1

25th Percentile: 1.00

Median: 1.00

75th Percentile: 1.00

Trend Analysis:

Trend: Stable

Distribution: Normal

Distribution of Employee Count

Column: Environment Satisfaction

Type: int64

Missing Values: 0

Unique Values: 4

Mean: 2.72

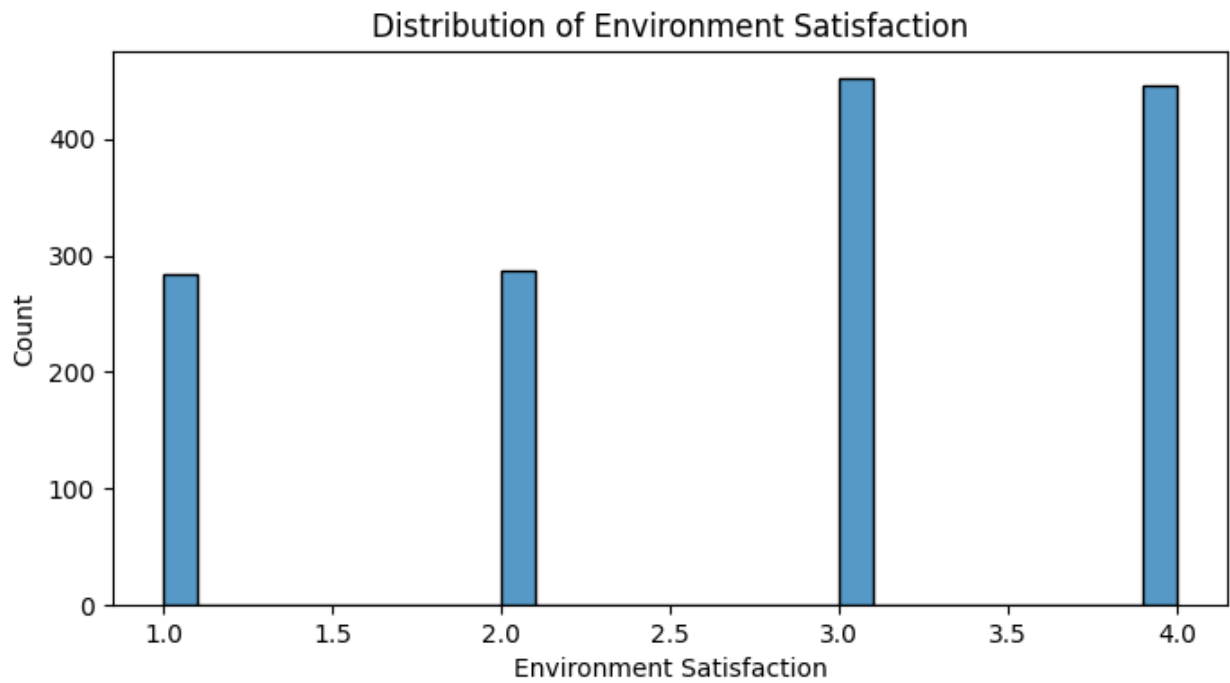Standard Deviation: 1.09

Min: 1

Max: 4

25th Percentile: 2.00

Median: 3.00

75th Percentile: 4.00

Trend Analysis:

Trend: Stable

Distribution: Normal

Distribution of Environment Satisfaction

Column: Job Involvement

Type: int64

Missing Values: 0

Unique Values: 4

Mean: 2.73

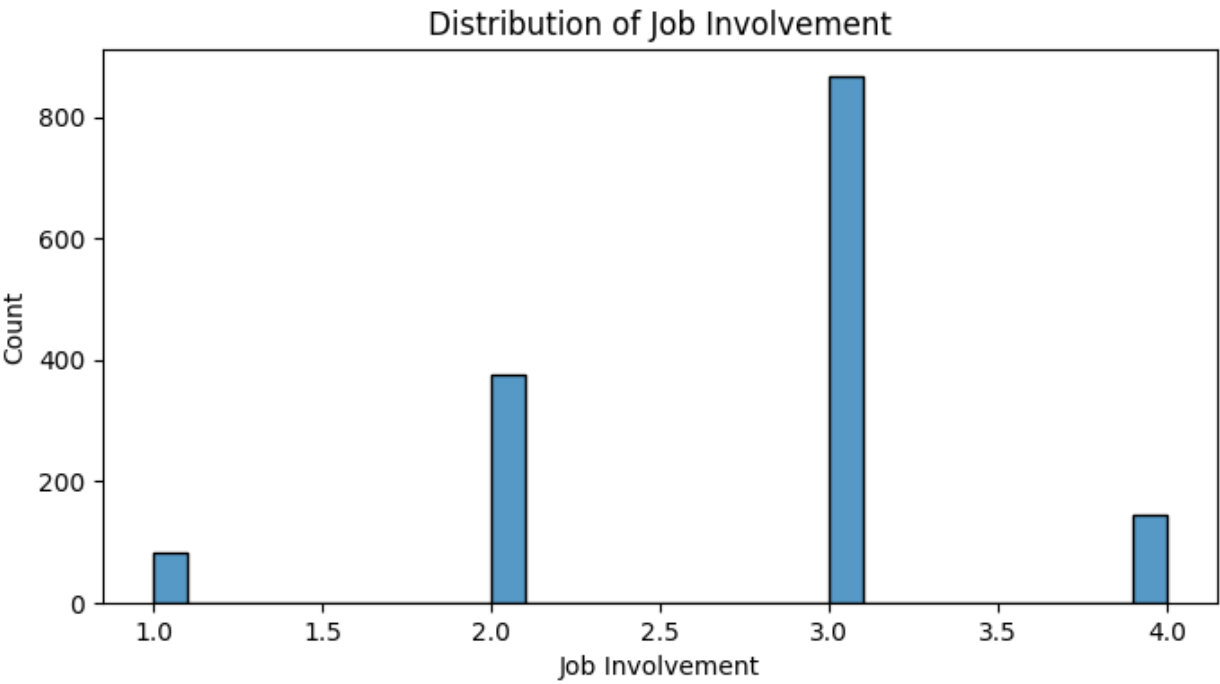Standard Deviation: 0.71

Min: 1

Max: 4

25th Percentile: 2.00

Median: 3.00

75th Percentile: 3.00

Trend Analysis:

Trend: Stable

Distribution: Normal

Distribution of Job Involvement

Column: Job Satisfaction

Type: int64

Missing Values: 0

Unique Values: 4

Mean: 2.73

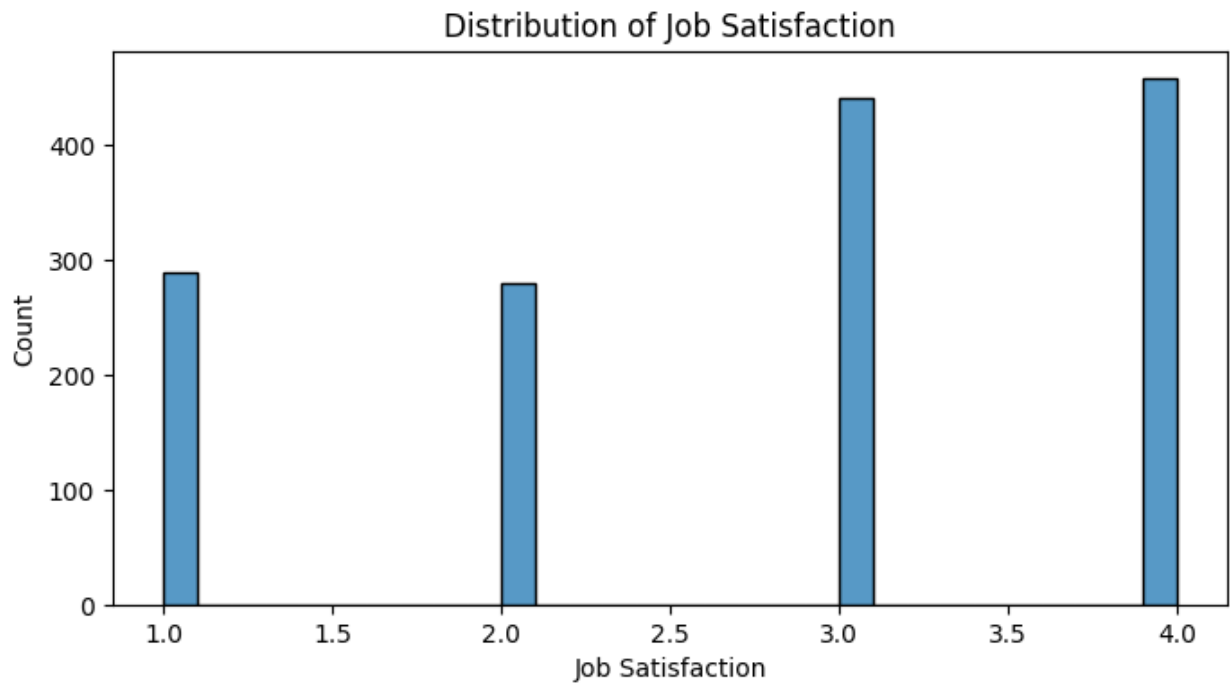Standard Deviation: 1.10

Min: 1

Max: 4

25th Percentile: 2.00

Median: 3.00

75th Percentile: 4.00

Trend Analysis:

Trend: Stable

Distribution: Normal

Distribution of Job Satisfaction

Column: Monthly Income

Type: int64

Missing Values: 0

Unique Values: 1349

Mean: 6502.93

Standard Deviation: 4707.96

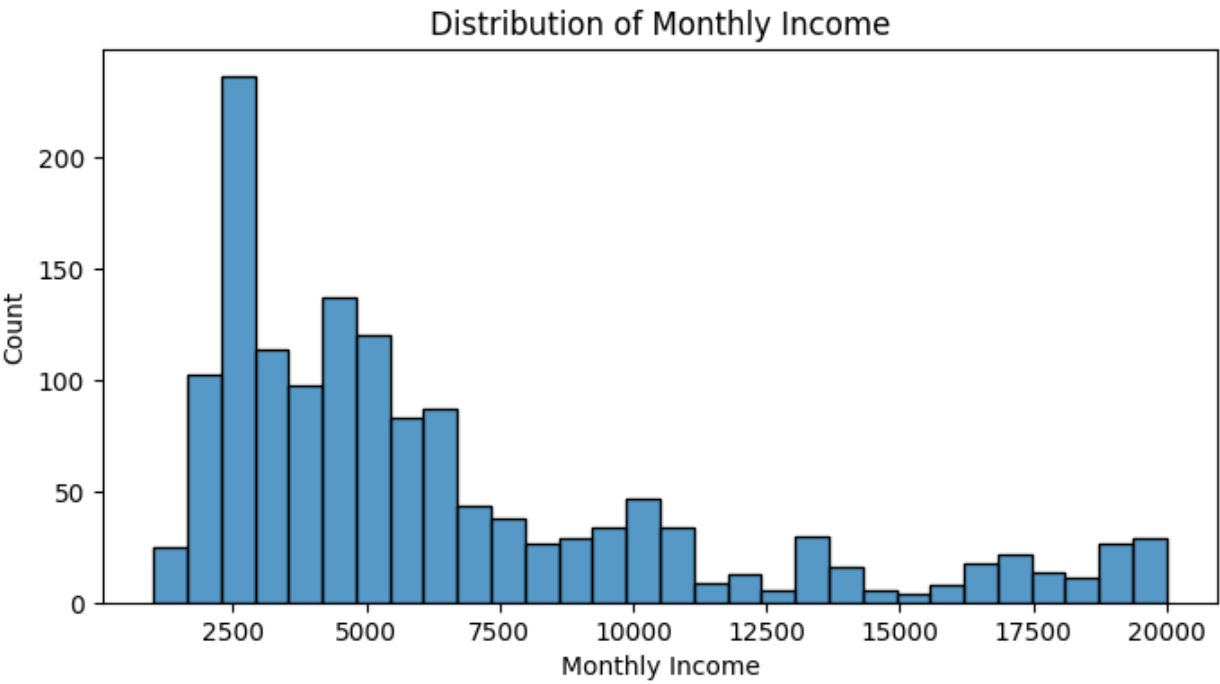Min: 1009

Max: 19999

25th Percentile: 2911.00

Median: 4919.00

75th Percentile: 8379.00

Trend Analysis:

Trend: Stable

Distribution: Right-Skewed

Distribution of Monthly Income

Column: Percent Salary Hike

Type: int64

Missing Values: 0

Unique Values: 15

Mean: 15.21

Standard Deviation: 3.66

Min: 11
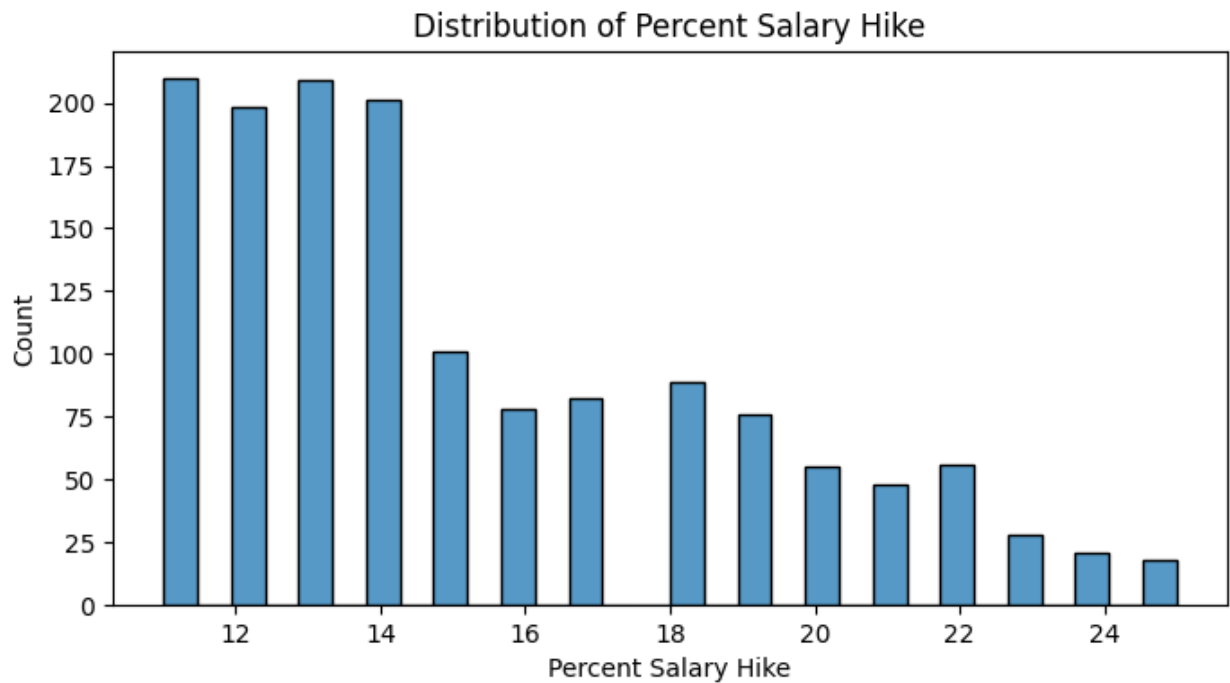
Max: 25

25th Percentile: 12.00

Median: 14.00

75th Percentile: 18.00

Trend Analysis:

Trend: Stable

Distribution: Right-Skewed

Distribution of Percent Salary Hike

Column: Performance Rating

Type: int64

Missing Values: 0

Unique Values: 2

Mean: 3.15

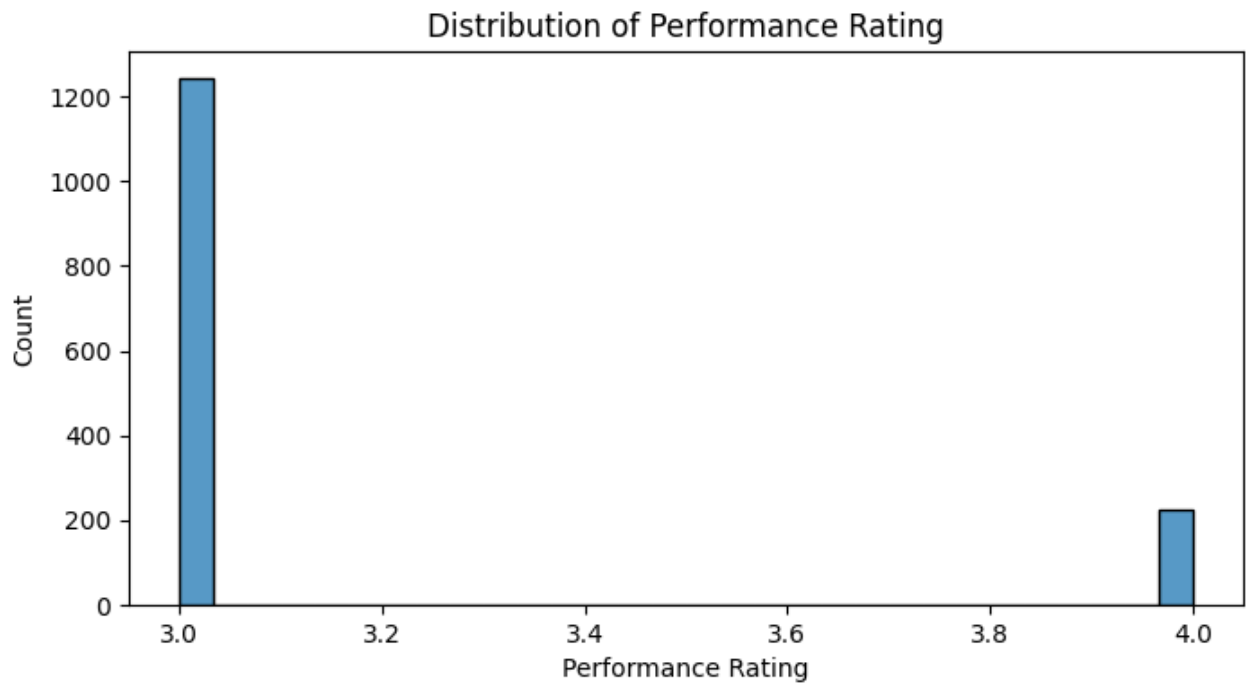Standard Deviation: 0.36

Min: 3

Max: 4

25th Percentile: 3.00

Median: 3.00

75th Percentile: 3.00

Trend Analysis:

Trend: Stable

Distribution: Right-Skewed

Distribution of Performance Rating

Column: Relationship Satisfaction

Type: int64

Missing Values: 0

Unique Values: 4

Mean: 2.71

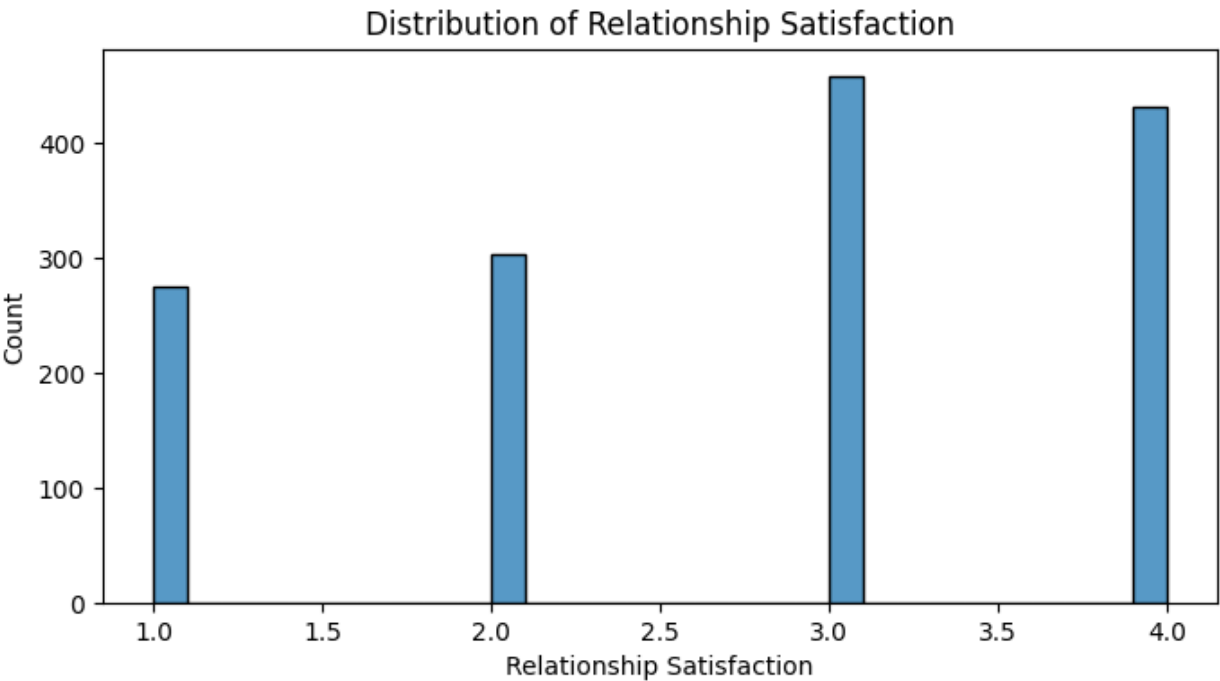Standard Deviation: 1.08

Min: 1

Max: 4

25th Percentile: 2.00

Median: 3.00

75th Percentile: 4.00

Trend Analysis:

Trend: Stable

Distribution: Normal

Distribution of Relationship Satisfaction

Column: Work Life Balance

Type: int64

Missing Values: 0

Unique Values: 4

Mean: 2.76

Standard Deviation: 0.71

Min: 1

Max: 4

25th Percentile: 2.00

Median: 3.00

75th Percentile: 3.00

Trend Analysis:

Trend: Stable

Distribution: Left-Skewed

Distribution of Work Life Balance