

Exploratory Data Analysis (EDA) Report (By Yitbarek Tesfaye)

1, Summary Statistics

- Compute key summary statistics for each numerical column, including:
 - Mean, median, and mode
 - Standard deviation
 - Minimum and maximum values

Variables	Mean	Median	Mode	Standard Deviation	Minumum	Maximum
Age	36.92381	36	35	9.132265691	18	60
DailyRate	802.48571	802	691	403.3718286	102	1499
DistanceFromHome	9.192517	7	2	8.10410653	1	29
Education	2.9129252	3	3	1.02381653	1	5
EmployeeCount	1	1	1	0	1	1
EmployeeNumber	1024.8653	1020.5	0	601.8195298	1	2068
EnvironmentSatisfaction	2.7217687	3	3	1.092710355	1	4
HourlyRate	65.891156	66	66	20.32251165	30	100
JobInvolvement	2.729932	3	3	0.711319074	1	4
JobLevel	2.0639456	2	1	1.106563325	1	5
JobSatisfaction	2.7285714	3	4	1.102470942	1	4
MonthlyIncome	6502.9313	4919	2342	4706.355165	1009	19999
MonthlyRate	14313.103	14235.5	9150	7115.364617	2094	26999
NumCompaniesWorked	2.6931973	2	1	2.497159199	0	9
PercentSalaryHike	15.209524	14	11	3.658692628	11	25
PerformanceRating	3.1537415	3	3	0.360700775	3	4
RelationshipSatisfaction	2.7122449	3	3	1.080841066	1	4
StandardHours	80	80	80	0	80	80
StockOptionLevel	0.7938776	1	0	0.851786797	0	3
TotalWorkingYears	11.279592	10	10	7.778134701	0	40
TrainingTimesLastYear	2.7993197	3	2	1.288832019	0	6
WorkLifeBalance	2.7612245	3	3	0.706235491	1	4
YearsAtCompany	7.0081633	5	5	6.124440946	0	40
YearsInCurrentRole	4.2292517	3	2	3.621904465	0	18
YearsSinceLastPromotion	2.1877551	1	0	3.221334028	0	15
YearsWithCurrManager	4.1231293	3	2	3.566922262	0	17

2, Correlation Analysis

- Identify relationships between numerical variables by calculating correlation coefficients.
- Create a **heatmap** to visualize these correlations. Highlight and explain any strong positive or negative correlations.

Variables	Age	DailyRate	DistanceFromHome	Education	EmployeeNumber	Environment	HourlyRate	JobInvolvement	JobLevel	JobSatisfaction	MonthlyIncome	MonthlyRate	NumCompaniesWorked	PercentSalaryHike	PerformanceRating	RelationshipSatisfaction	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrentManager
Age	1	0.010661	-0.00169	0.208034	-0.01015	0.010146	0.024287	0.02982	0.509604	-0.00489	0.497855	0.028051	0.299635	0.003634	0.001904	0.053535	0.03751	0.680381	-0.01962	-0.02149	0.311309	0.212901	0.216513	0.2020886
DailyRate	0.010661	1	-0.00499	-0.01681	-0.05099	0.018355	0.023381	0.046135	0.002966	0.030571	0.007707	-0.03218	0.038153	0.022704	0.000473	0.007846	0.042143	0.014515	0.002453	-0.03785	-0.03405	0.009932	-0.03323	-0.0263632
DistanceFromHome	-0.00169	-0.00499	1	0.021042	0.032916	-0.01608	0.031131	0.008783	0.005303	-0.00367	-0.01701	0.027473	-0.02925	0.040235	0.02711	0.006557	0.044872	0.004628	-0.03684	-0.02656	0.009508	0.018845	0.010029	0.01440605
Education	0.208034	-0.01681	0.021042	1	0.04207	-0.02713	0.016775	0.042438	0.101589	-0.0113	0.094961	-0.02608	0.126317	-0.0111	-0.02454	-0.00912	0.018422	0.14828	-0.0251	0.009819	0.069114	0.060236	0.054254	0.06906538
EmployeeNumber	-0.01015	-0.05099	0.032916	0.04207	1	0.017621	0.035179	-0.00689	-0.01852	-0.04625	-0.01483	0.012648	-0.00125	-0.01294	-0.02036	-0.06986	0.062227	-0.01437	0.023603	0.010309	-0.01124	-0.00842	-0.009902	-0.0091966
EnvironmentSatisfaction	0.010146	0.018355	-0.01608	-0.02713	0.017621	1	-0.04986	-0.00828	0.001212	-0.00678	-0.00626	0.0376	0.012594	-0.0317	-0.02955	0.007665	0.003432	-0.00269	-0.01936	0.027627	0.001458	0.018007	0.016194	-0.0049987
HourlyRate	0.024287	0.023381	0.031131	0.016775	0.035179	-0.04986	1	0.042861	-0.02785	-0.07133	-0.01579	-0.0153	0.022157	-0.00906	-0.00217	0.00133	0.050263	-0.00233	-0.00855	-0.00461	-0.01958	-0.02411	-0.02672	-0.0201232
JobInvolvement	0.02982	0.046135	0.008783	0.042438	-0.00689	-0.00828	0.042861	1	-0.01263	-0.02148	-0.01527	-0.01632	0.015012	-0.0172	-0.02907	0.034297	0.021523	-0.00553	-0.01534	-0.01462	-0.02136	0.008717	-0.02418	0.02597581
JobLevel	0.509604	0.002966	0.005303	0.101589	-0.01852	0.001212	-0.02785	-0.01263	1	-0.00194	0.9501	0.039563	0.142501	-0.03473	-0.02122	0.021642	0.013984	0.782208	-0.01819	0.037818	0.534739	0.389447	0.353885	0.37528061
JobSatisfaction	-0.00489	0.030571	-0.00367	-0.0113	-0.04625	-0.00678	-0.07133	-0.02148	-0.00194	1	-0.00716	0.000644	-0.0557	0.020002	0.002297	-0.01245	0.01069	-0.02019	-0.00578	-0.01946	-0.0038	-0.0023	-0.01821	-0.0276562
MonthlyIncome	0.497855	0.007707	-0.01701	0.094961	-0.01483	-0.00626	-0.01579	-0.01527	0.9501	-0.00716	1	0.034814	0.149515	-0.02727	-0.01712	0.025873	0.005408	0.772393	-0.02174	0.030683	0.514285	0.363818	0.344978	0.34407888
MonthlyRate	0.028051	-0.03218	0.027473	-0.02608	0.012648	0.0376	-0.0153	-0.01632	0.039563	0.000644	0.034814	1	0.017521	-0.00643	-0.00981	-0.00409	-0.03492	0.026442	0.001467	0.007963	-0.02366	-0.01281	0.001567	-0.0367459
NumCompaniesWorked	0.299635	0.038153	-0.02925	0.126317	-0.00125	0.012594	0.022157	0.015012	0.142501	-0.0557	0.149515	0.017521	1	-0.01024	-0.01409	0.052733	0.030075	0.237639	-0.06605	-0.00837	-0.00705	-0.03681	-0.010718	
PercentSalaryHike	0.003634	0.022704	0.040235	-0.01111	-0.01294	-0.0317	-0.00906	-0.0172	-0.03473	0.020002	-0.02727	-0.00643	-0.01024	1	0.77355	-0.04049	0.007528	-0.02061	-0.00522	-0.00328	-0.03599	-0.00152	-0.02215	-0.0119852
PerformanceRating	0.001904	0.000473	0.02711	-0.02454	-0.02036	-0.02955	-0.00217	-0.02907	-0.02122	0.002297	-0.01712	-0.00981	-0.01409	0.77355	1	-0.03135	0.003506	0.006744	-0.01558	0.002572	0.003435	0.034986	0.017896	0.02282717
RelationshipSatisfaction	0.053535	0.007846	0.006557	-0.00912	-0.06986	0.007665	0.00133	0.034297	0.021642	-0.01245	0.025873	-0.00409	0.052733	-0.04049	-0.03135	1	-0.04595	0.024054	0.002497	0.019604	0.019367	-0.01512	0.033493	-0.0008675
StockOptionLevel	0.03751	0.042143	0.044872	0.018422	0.062227	0.003432	0.021523	0.013984	0.01069	0.005408	-0.03432	0.030075	0.007528	0.003506	-0.04595	-0.04595	1	0.010136	0.011274	0.004129	0.015058	0.014352	0.02469823	
TotalWorkingYears	0.680381	0.014515	0.004628	0.14828	-0.01437	-0.00769	-0.00233	-0.00553	0.782208	-0.02019	0.772393	0.026442	0.237639	-0.02061	0.006744	0.024054	0.010136	1	-0.03566	0.001008	0.628133	0.460365	0.404858	0.4591884
TrainingTimesLastYear	-0.01962	0.002453	-0.03684	-0.0251	0.023603	-0.01936	-0.00855	-0.01534	-0.01819	-0.00578	-0.02174	0.001467	-0.06605	-0.00522	-0.01558	0.002497	0.011274	-0.03566	1	0.028072	0.003569	-0.00574	-0.00207	-0.0040955
WorkLifeBalance	-0.02149	-0.03785	-0.02656	0.009819	0.010309	0.027627	-0.00461	-0.01462	0.037818	-0.01946	0.030683	0.007963	-0.00837	-0.00328	0.002572	0.019604	0.004129	0.001008	0.028072	1	0.012089	0.049856	0.008941	0.00275944
YearsAtCompany	0.311309	-0.03405	0.009508	0.069114	-0.01124	0.001458	-0.01958	-0.02136	0.534739	-0.0038	0.514285	-0.02366	-0.01194	-0.03599	0.003435	0.019367	0.015058	0.628133	0.003569	0.012089	1	0.758754	0.618409	0.76921243
YearsInCurrentRole	0.212901	0.009932	0.018845	0.060236	-0.00842	0.018007	-0.02411	0.008717	0.389447	-0.0023	0.363818	-0.01281	-0.09075	-0.00152	0.034986	-0.01512	0.050818	0.460365	-0.00574	0.049856	0.758754	1	0.548056	0.71436476
YearsSinceLastPromotion	0.216513	-0.03323	0.010029	0.054254	-0.00902	0.016194	-0.02672	-0.02418	0.353885	-0.01821	0.344978	0.001567	-0.03681	-0.02215	0.017896	0.033493	0.014352	0.404858	-0.00207	0.008941	0.618409	0.548056	1	0.51022364
YearsWithCurrentManager	0.202089	-0.02636	0.014406	0.069065	-0.0092	-0.005	-0.02012	0.025976	0.375281	-0.02766	0.344079	-0.03675	-0.11038	-0.01199	0.022827	-0.00087	0.024698	0.459188	-0.0041	0.002759	0.769212	0.714365	0.510224	1

Explanation

The heatmap provides a visual representation of the correlation between different numerical variables in the dataset. Each cell in the heatmap shows how strongly two variables are related, with values ranging from **-1 to +1**:

- Positive Correlation (+1)**: Indicates a strong direct relationship (e.g., as one variable increases, the other also increases).
- Negative Correlation (-1)**: Indicates a strong inverse relationship (e.g., as one variable increases, the other decreases).
- No Correlation (0)**: Indicates no relationship between the variables.

Key Notes About This Heatmap

✓ Removed Variables:

- Two variables, **EmployeeCount** and **StandardHours**, were removed from the analysis because their standard deviation is zero. This means these columns contain constant values, which do not contribute to meaningful correlations.

✓ Color Coding:

- Strong Positive Correlations** (closer to +1) are shown in green.
- Strong Negative Correlations** (closer to -1) are shown in red.

- **Weak or No Correlation** (near 0) are in lighter or neutral colors.

✓ **Insights:**

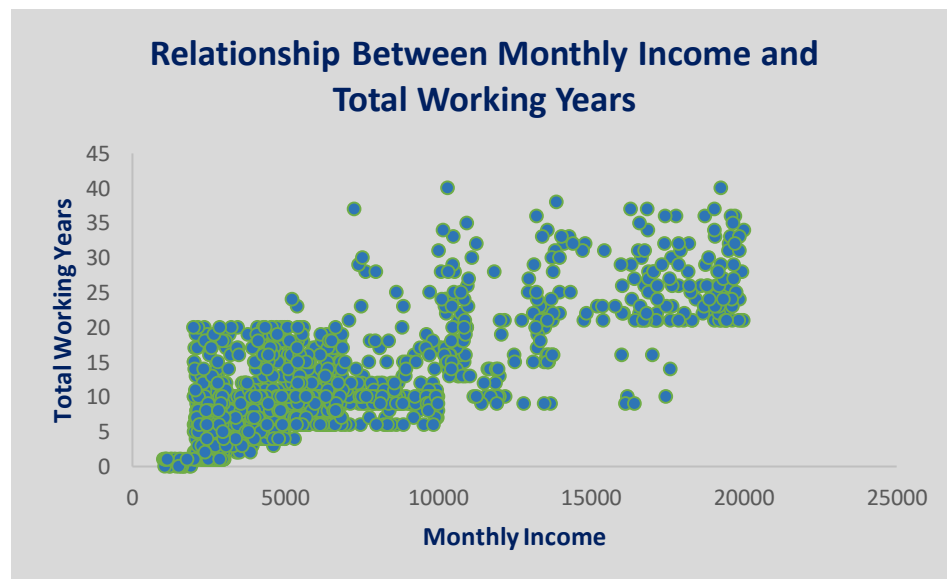
- For example, the heatmap might show a strong positive correlation between **Monthly Income** and **Total Working Years**, indicating that employees with more experience tend to earn higher incomes.
- It might also show weak or no correlation between **Age** and **Distance from Home**, suggesting these factors are not related.

✓ **Diagonal Values:**

- The diagonal values are always **1** because each variable is perfectly correlated with itself.

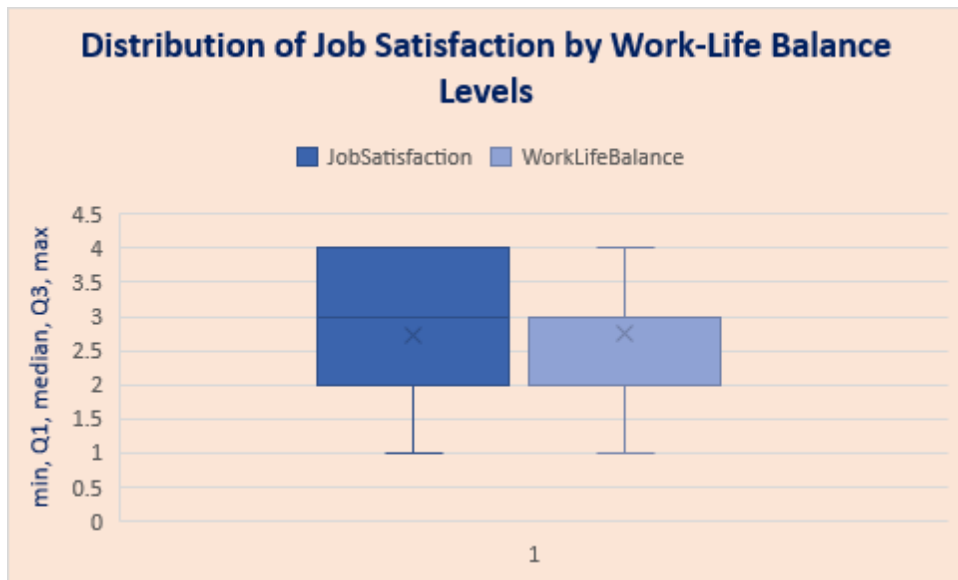
3, Variable Relationships

- Visualize the relationship between at least two pairs of variables. Use plots such as scatter plots, box plots, or pair plots.
 - **Plot 1- Scatter Plot (Monthly Income vs. Total Working Years)**



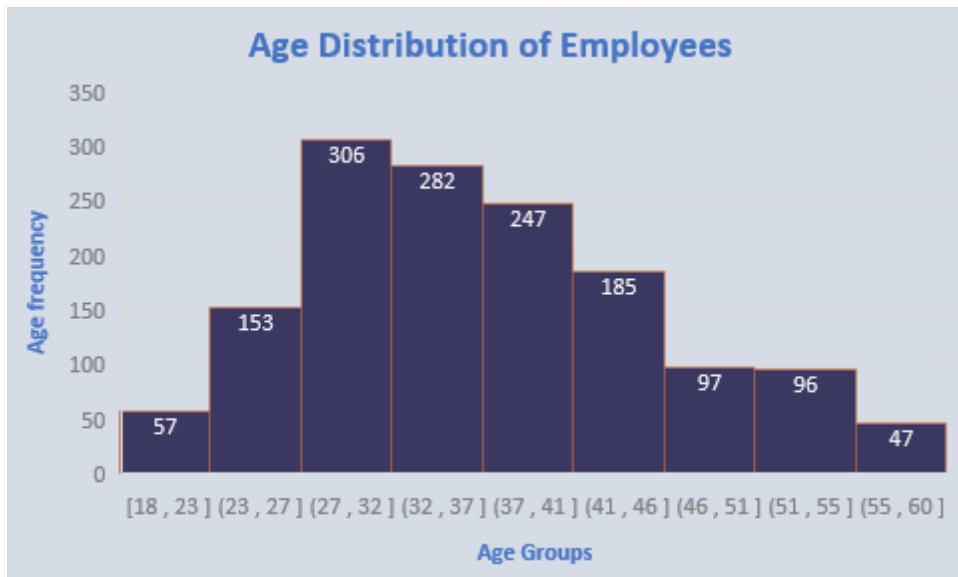
Insight: Employees with more total working years tend to earn higher monthly incomes, indicating that experience significantly impacts income.

- Plot 2- Box Plot (Job Satisfaction vs. Work-Life Balance)



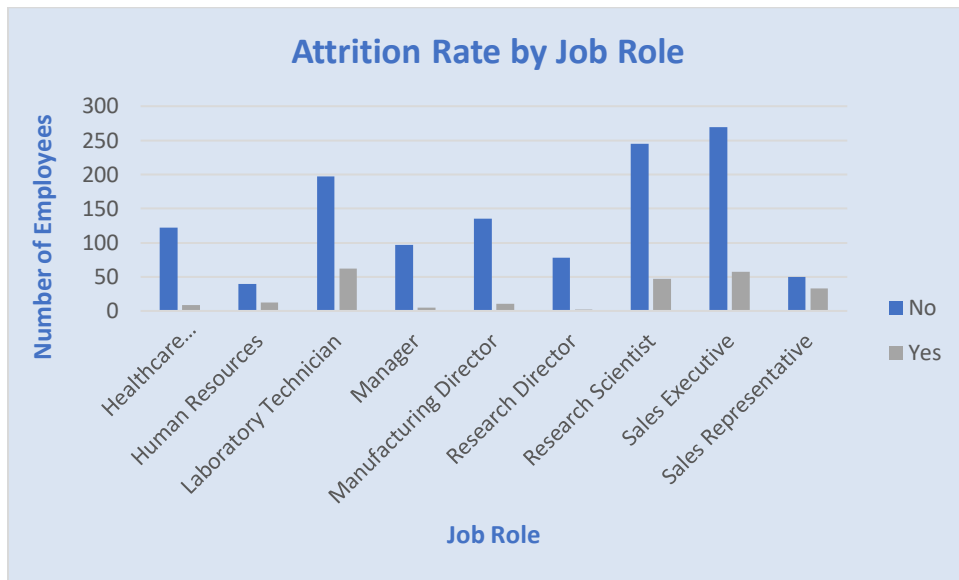
Insight: Higher levels of work-life balance are associated with better job satisfaction, though there is some variation across categories.

- Plot 3 - Histogram (Age Distribution)



Insight: The majority of employees are in the age range of 30–40, which might indicate a mid-career workforce composition.

▪ **Plot 4 - Bar Chart (Attrition vs. Job Role)**



Insight: Attrition is notably higher in sales-related roles, suggesting potential challenges in retaining employees in these positions.

4, Insights Summary

The exploratory data analysis (EDA) of the IBM HR Analytics Employee Attrition Dataset revealed several key insights into the workforce dynamics and employee attrition trends.

1. **Age Distribution:** Most employees fall within the 30–40 age range, suggesting a mid-career demographic. This age group often balances career growth with personal responsibilities, which may influence their engagement and retention.
2. **Monthly Income and Total Working Years:** A strong positive correlation was observed between monthly income and total working years. This indicates that experience is a significant determinant of salary progression, reflecting standard industry practices.
3. **Attrition by Job Role:** Attrition rates vary significantly across job roles, with sales-related roles experiencing the highest turnover. This trend may indicate challenges in retaining employees in sales positions, possibly due to job-specific stressors or compensation concerns.
4. **Job Satisfaction and Work-Life Balance:** Employees reporting higher levels of work-life balance also tend to have better job satisfaction. This finding underscores the

importance of maintaining a healthy work-life balance to improve employee happiness and productivity.

Overall, the analysis highlights critical areas for organizational improvement, such as supporting employees in high-stress roles, promoting work-life balance, and tailoring strategies to engage and retain mid-career professionals. These insights can guide data-driven decisions to reduce attrition and enhance workforce satisfaction.