DATASET ASSESSMENT DOCUMENT (DAD)

A. DATASET OVERVIEW

Item	Details
Dataset Title	HR Analytics – Employee Demographics, Compensation, Attrition & Performance
Domain / Theme	Human Resources, Workforce Analytics
Source	Provided Excel file: *HR Data.xlsx*
File Type	Excel (XLSX)
No. of Rows	~1,470 records
No. of Columns	~30–35 columns (across 4 tables: Attrition, Compensation, Employee, Job, Satisfaction)
Data Structure Summary	Multiple tables combined using Employee Number as primary key

Column-Level Summary

Column Name	Data Type	Dimension / Measure	Description
Employee Number	Integer	Dimension	Unique ID for each employee
Attrition	String (Yes/No)	Dimension	Whether the employee left the company

Over Time	String (Yes/No)	Dimension	Employee overtime status
Performance Rating	Integer	Measure	Performance score
Total Working Years	Integer	Measure	Total years worked
Training Times Last Year	Integer	Measure	Training frequency
Years Since Last Promotion	Integer	Measure	Time since last promotion
Years at Company	Integer	Measure	Tenure at current company
Years in Current Role	Integer	Measure	Time in current job role
Gender	String	Dimension	Employee gender
Age	Integer	Measure	Employee age
Job Role	String	Dimension	Designation of the employee
Department	String	Dimension	Department type
Monthly Income	Integer	Measure	Employee's salary
Percent Salary Hike	Integer	Measure	Yearly salary increase percentage

Stock Option Level	Integer	Dimension	Employee stock plan level
Education Field	String	Dimension	Educational background
Job Satisfaction	Integer (14)	Measure	Satisfaction score
Environment Satisfaction	Integer	Measure	Satisfaction with work environment
Work-Life Balance	Integer	Measure	Employee's WLB rating

B. DATA QUALITY ASSESSMENT

Aspect Checked	Observations	Action Needed
Missing Data	Very few or no nulls in key fields (based on Tableau charts working smoothly).	No action needed unless deep cleaning required.
Duplicates	Employee Number appears unique.	None required.
Outliers	Salary, Age, Years at Company have a few extreme values (seen in histogram bins like 55–60, 21+ years).	Treat outliers as insights, not errors. Keep them.

Incorrect Data Types	Tableau automatically converted measures/dimensions correctly.	No action needed.
Consistency Issues	All categorical fields (Department, Job Role, Gender)	None required.
Data Integration Issues	Multiple tables merged correctly (Attrition, Compensation, Job, Satisfaction).	No action required.

C. Key Metrics & KPIs Identified

KPI	Why It Matters
Total Employees	Helps understand workforce size.
Attrition Rate	Measures employee turnover, critical HR metric.
Average Salary	Helps compare compensation across job roles.
Average Salary Hike %	Shows reward trends and fairness in compensation.
Promotion Delay (Avg Years Since Last Promotion)	Identifies slow career progression.
Average Years in Current Role	Indicates employee growth & stagnation.

Job Role-Wise Attrition	Highlights departments needing retention strategies.
Age Distribution	Helps with workforce planning and hiring strategy.
Employee Experience Heatmap (Age × Tenure)	Shows clusters of experienced staff.
Gender Distribution	Useful for diversity & inclusion analysis.

D. Assumptions / Risks

Assumption / Risk	Explanation
Data represents a single organization	Insights may not generalize to other firms.
Performance Rating Scale is consistent	Assumes values 1–4 or 1–5 across the dataset.
Self-reported satisfaction scores	Subjective; may not reflect true satisfaction.
No hidden missing values	Assumes no "0" or "Unknown" fake data entries.
Historical data not provided	Analysis is point-in-time, not time-series.