AI-POWERED DATA ANALYSIS REPORT

Generated on: 2025-09-26 10:23:49

EXECUTIVE SUMMARY

1. Data Overview: The dataset contains 50 rows and 11 columns.

2. Data Quality: 3 missing values detected across the dataset.

3. Numeric Analysis: 5 numeric columns found. Age has mean 28.00 and std 4.22.

4. Categorical Analysis: 6 categorical columns identified.

5. Strong Correlations: Found between Age & Experience (0.92), Age & Salary (0.77), Experience & Salary (0.90)

DATASET OVERVIEW

- Total Rows: 50
- Total Columns: 11
- Missing Values: 3
- Duplicate Rows: 0
- Numeric Columns: 5
- Categorical Columns: 6

DATA TYPES SUMMARY

- EmpID: object
- EmpName: object
- Department: object
- Location: object
- Age: int64
- No_of_trainings: int64
- KPIs_met >80%: int64
- Experience: float64
- Salary: int64
- Project_id: object

NUMERIC ANALYSIS

- Age: count=50.0, mean=28.00, std=4.22

- No_of_trainings: count=50.0, mean=1.20, std=0.45

- KPIs_met >80%: count=50.0, mean=0.36, std=0.48

- Experience: count=47.0, mean=3.29, std=1.60

- Salary: count=50.0, mean=61460.58, std=46405.68

CATEGORICAL ANALYSIS

- EmpID: 50 unique values
 - S1001: 1 occurrences
 - S1002: 1 occurrences

- EmpName: 47 unique values
 - Michael Johnson: 2 occurrences
 - David Wilson: 2 occurrences

- Department: 10 unique values
 - Marketing: 7 occurrences
 - Sales: 6 occurrences

- Location: 5 unique values
 - Mumbai: 13 occurrences
 - Bangalore: 12 occurrences

RECOMMENDATIONS & CONCLUSIONS

1. Review columns with high missing values and consider imputation strategies

2. Remove or investigate duplicate records to ensure data quality

3. Analyze strong correlations for feature engineering opportunities

4. Validate data distributions and handle outliers appropriately

5. Consider collecting more data if the dataset is small (<1000 records)

6. Explore categorical variables for potential grouping or encoding strategies

7. Perform additional statistical tests based on the data characteristics

8. Consider time-series analysis if temporal patterns are present

DATA QUALITY ASSESSMENT

- Completeness Score: 99.5%
- Duplicate Rate: 0.0%
- Data Volume: Small dataset
- Variable Diversity: Good mix of numeric and categorical variables

CONCLUSION

This comprehensive analysis of the dataset containing 50 records and 11 variables
provides valuable insights for data-driven decision making. The analysis highlights key patterns,
data quality considerations, and opportunities for further exploration.

The dataset demonstrates 99.5% completeness with 5 numeric
and 6 categorical variables available for analysis. The recommendations
provided offer actionable steps for data improvement and advanced analytical modeling.

For further analysis, consider implementing machine learning models, time-series forecasting,
or advanced statistical techniques based on the specific business objectives.