

# **EMPLOYEE DATA ANALYSIS USING EXCEL**

**NAME: M.SANDHIYA  
NMI'D:ASUNM110312201391  
DEPARTMENT:COMMERCE  
COLLEGE: DRBCCC HINDU  
COLLEGE**

# AGENDA

1. Problem Statement
2. Overview
3. End Users
4. Our Solution and Proposition
5. Dataset Description
6. Modeling Approach
7. Result And Discussion
8. Conclusion

# PROBLEM STATEMENT

---

1. **Descriptive Statistics:** Calculate and summarize basic statistics for both age and salary, including mean, median, mode, range, variance, and standard deviation
2. **Correlation Analysis:** Determine if there is a significant correlation between age and salary. Use appropriate statistical methods to measure the strength and direction of this relationship
3. **Salary Distribution:** Analyze the distribution of salaries across different age groups. Create histograms or other relevant visualizations to illustrate how salaries are distributed within each age group
4. **Outlier Detection:** Identify any outliers in the salary data. Use statistical techniques such as Z-scores or IQR to detect and analyze these outlier
5. **Predictive Modeling:** Develop a simple predictive model to estimate salary based on age. Evaluate the model's performance using appropriate metrics.
6. **Insight Generation:** Provide insights and conclusions based on your analysis. Discuss any patterns or trends observed and their potential implications.

# OVERVIEW

- 1. Data Columns:** Age: Represents the age of individuals. Salary: Represents the annual salary of individuals.
- 2. Sample Size:** The dataset contains 50 entries. **Age Distribution:** Ages range from 22 to 40. There are multiple entries for several ages, indicating certain age groups are more represented.
- 3. Salary Distribution:** Salaries range from 11,000 to 69,000. The dataset shows a wide range in salaries, with some high salaries potentially skewing the distribution.
- 4. Key Statistics:** Mean Salary: Likely to be higher due to the presence of high salaries. Median Salary: Provides a better central tendency measure, less affected by extreme values. Mode of Age and Salary: Certain ages and salaries may repeat, providing insight into common values.
- 5. Patterns and Trends:** There may be observable trends or patterns between age and salary. The dataset could reveal if older individuals tend to have higher salaries or if there is significant variability in salaries within certain age groups. **Outliers:** Extreme salary values (e.g., 11,000 and 69,000) suggest potential outliers that could influence the analysis.
- 6. Correlation Insight:** Analyzing the correlation between age and salary might reveal if there's a positive, negative, or no relationship between the two variables.

# END USERS

1. **Descriptive Statistics:** Calculate mean, median, mode, range, and standard deviation for both age and salary.

2. **Correlation Analysis:** Assess the strength and direction of the relationship between age and salary.



3. **Salary Distribution:** Visualize salary distribution across different age groups using histograms or box plots.



4. **Outlier Detection:** Identify and analyze any outliers in the salary data.

5. **Predictive Modeling:** Develop a model to predict salary based on age and evaluate its performance.



# Solution

[Photo](#)[About Us](#)[Contact](#)

- 1.CALCULATE SUMMARY STATISTICS:MEAN AGE: AVERAGE AGE OF THE EMPLOYEES.MEDIAN AGE: MIDDLE VALUE OF THE SORTED AGES.MEAN SALARY: AVERAGE SALARY OF THE EMPLOYEES.**
  
- 2.MEDIAN SALARYMIDDLE VALUE OF THE SORTED SALARIES.IDENTIFY OUTLIERS:FOR AGE AND SALARY, DETECT VALUES SIGNIFICANTLY HIGHER OR LOWER THAN THE TYPICAL RANGE.**
  
- 3.FIND CORRELATION:CALCULATE THE CORRELATION COEFFICIENT BETWEEN AGE AND SALARY TO SEE IF THERE'S A RELATIONSHIP BETWEEN THEM.**

# PROPOSITION

- 1.Age:** Mean Age: Approximately 31.3 years. Median Age: 30 years.
- 2.Salary:** Mean Salary: Approximately \$37,000. Median Salary: \$35,000.
- 3.Outliers:** The highest salaries (e.g., \$69,000) and the lowest salaries (e.g., \$11,000) might be considered outliers.
- 4.Correlation:** The correlation between Age and Salary needs to be calculated to determine if there's a significant relationship between the two variables.

# DATASET DESCRIPTION

**The dataset contains 50 entries with two variables: Age and Salary.**

**Age ranges from 22 to 40 years. Salary ranges from \$11,000 to \$69,000.**

**Key observations:**

**Mean Age: Approximately 30.6 years.**

**Mean Salary: Approximately \$37,200.**

**The dataset includes several outliers with extremely high or low salaries compared to the central values.1**

# MODELLING APPROACH

1. Data preparation
2. Exploratory Data Analysis (EDA)
3. Feature Engineering
4. Model Selection
5. Model Validation
6. Refinement
7. Interpret Results
8. Report Finding

# RESULT

## 1. Descriptive Statistics:

**Age:** Mean = 30.1 years, Median = 30 years, Range = 22 to 40 years.  
**Salary:** Mean = \$39,672, Median = \$45,000, Range = \$11,000 to \$69,000.  
**Correlation:** Person correlation coefficient = 0.15, indicating a weak positive correlation between age and salary.

## 2. Modeling:

**Linear Regression:** R-squared  $\approx$  0.02, suggesting that age alone explains only about 2% of the salary variability.  
**Polynomial Regression (Quadratic):** R-squared  $\approx$  0.30, indicating a moderate improvement but still showing that age is not a strong sole predictor of salary.

# Line chart [for salary]



# Discussion

The weak correlation and low R-squared values in both linear and polynomial models suggest that age alone is not a strong predictor of salary. The polynomial model improves the fit slightly but indicates that other factors likely play a significant role in determining salary. The presence of outliers with extreme salaries also impacts model accuracy. To better understand salary determination, additional variables such as experience, education, and job role should be considered.

# Conclusion

The analysis reveals that there is a weak positive correlation between age and salary, with a Pearson correlation coefficient of 0.15. Linear regression and polynomial regression models show that age alone explains only a small portion of the variability in salary, with R-squared values of 0.02 and 0.30, respectively. The presence of significant outliers also affects the accuracy of these models. Overall, age is not a strong predictor of salary by itself; other factors such as experience, education, and job role likely play a more significant role in determining salary.



# THANK YOU