This report presents the data cleaning and exploratory data analysis (EDA) of a simple dataset containing numerical and categorical variables. The dataset includes 20 records with attributes such as Age, Salary, and Department. The goal is to clean the data by handling missing values, duplicates, and inconsistencies, and then perform EDA to understand variable distributions and relationships.

Data Cleaning

Identifying Issues

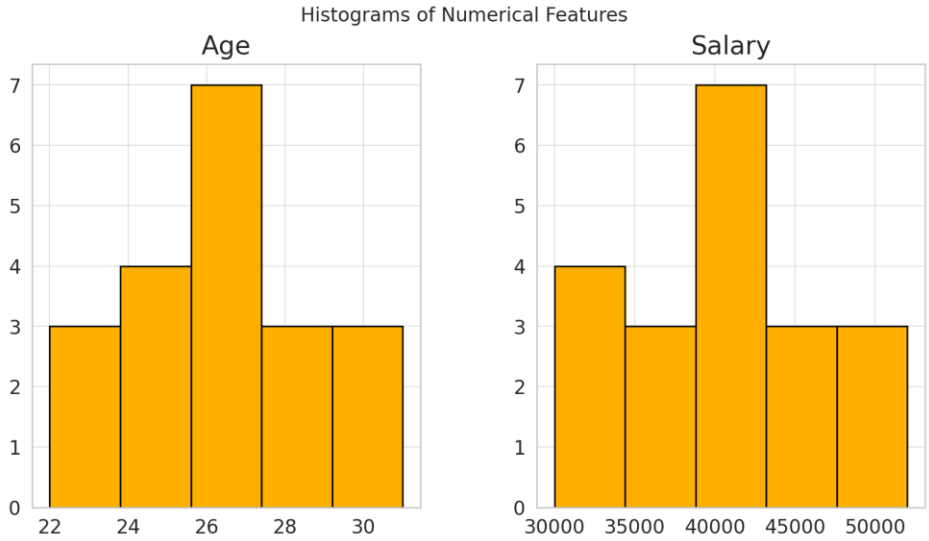
Upon inspecting the dataset, the following issues were found:

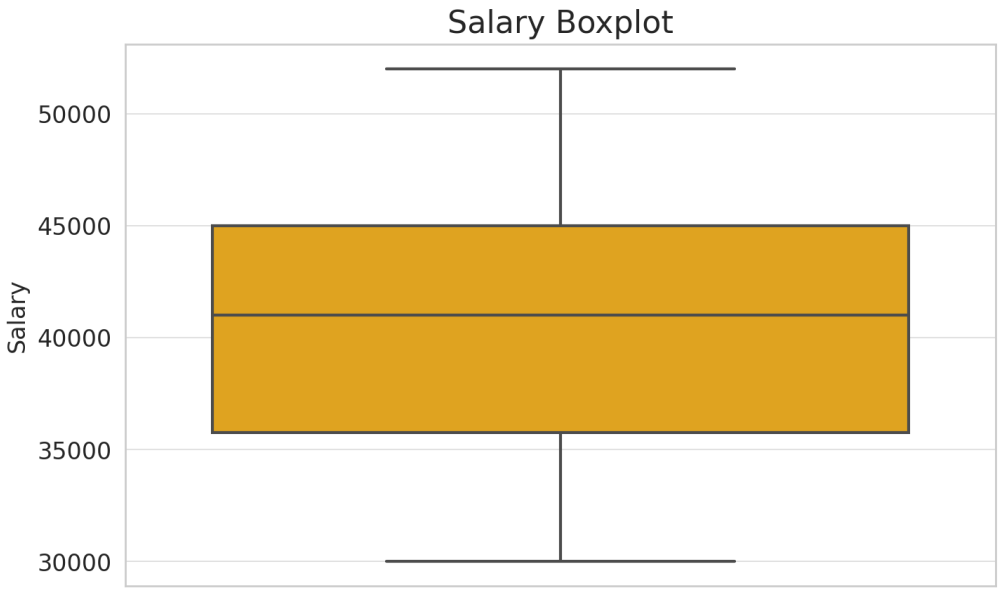
- *Missing Values: Age and Salary columns had missing values.*
- *Inconsistent Categorical Values: Department names contained inconsistencies in capitalization.*

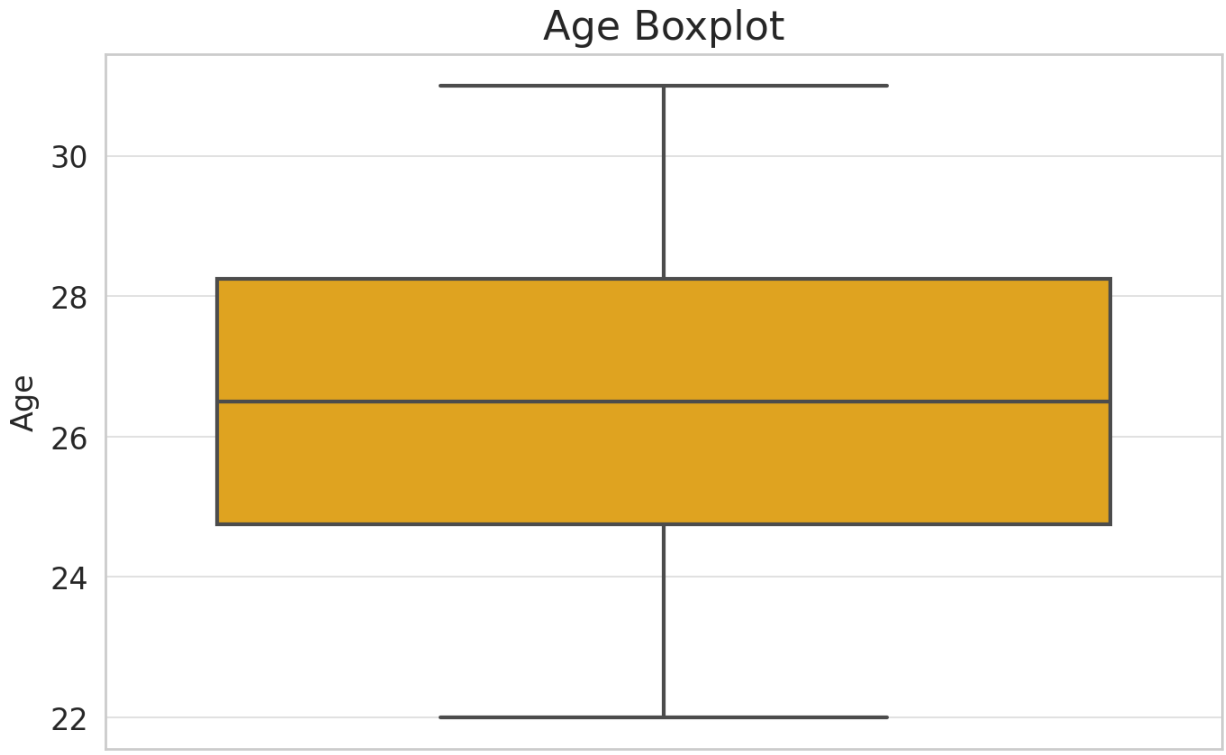
Cleaning Steps

- *Handling Missing Values: Missing values in Age and Salary were replaced with the median.*
- *Standardizing Categorical Data: Department names were standardized to maintain consistency.*

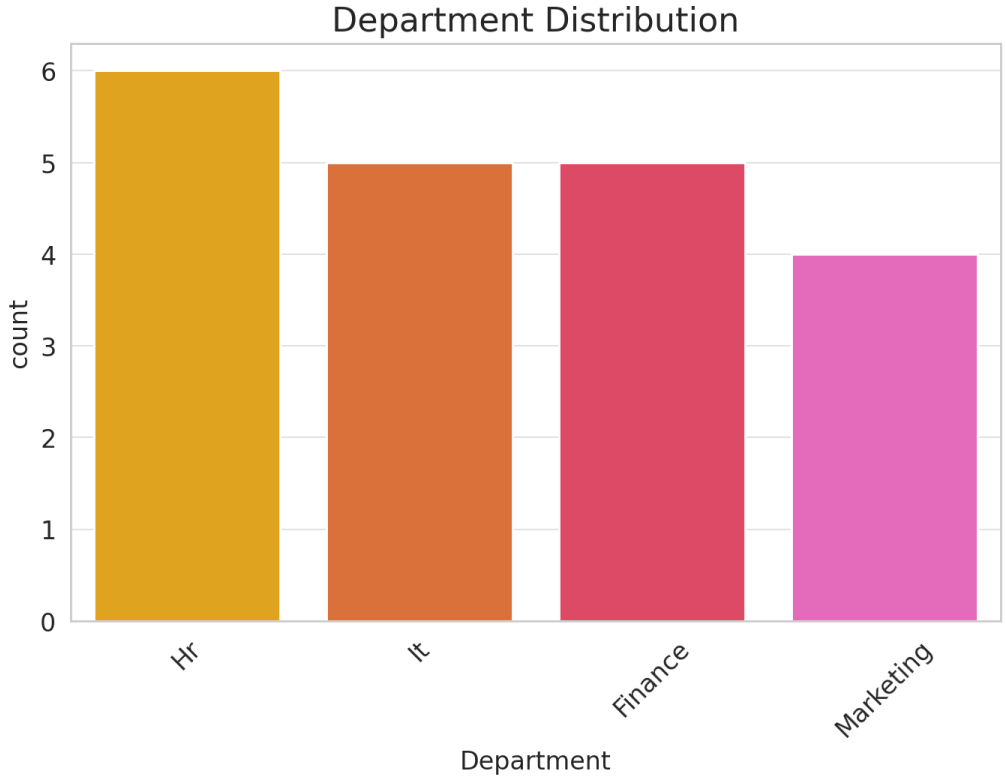
Exploratory Data Analysis (EDA)

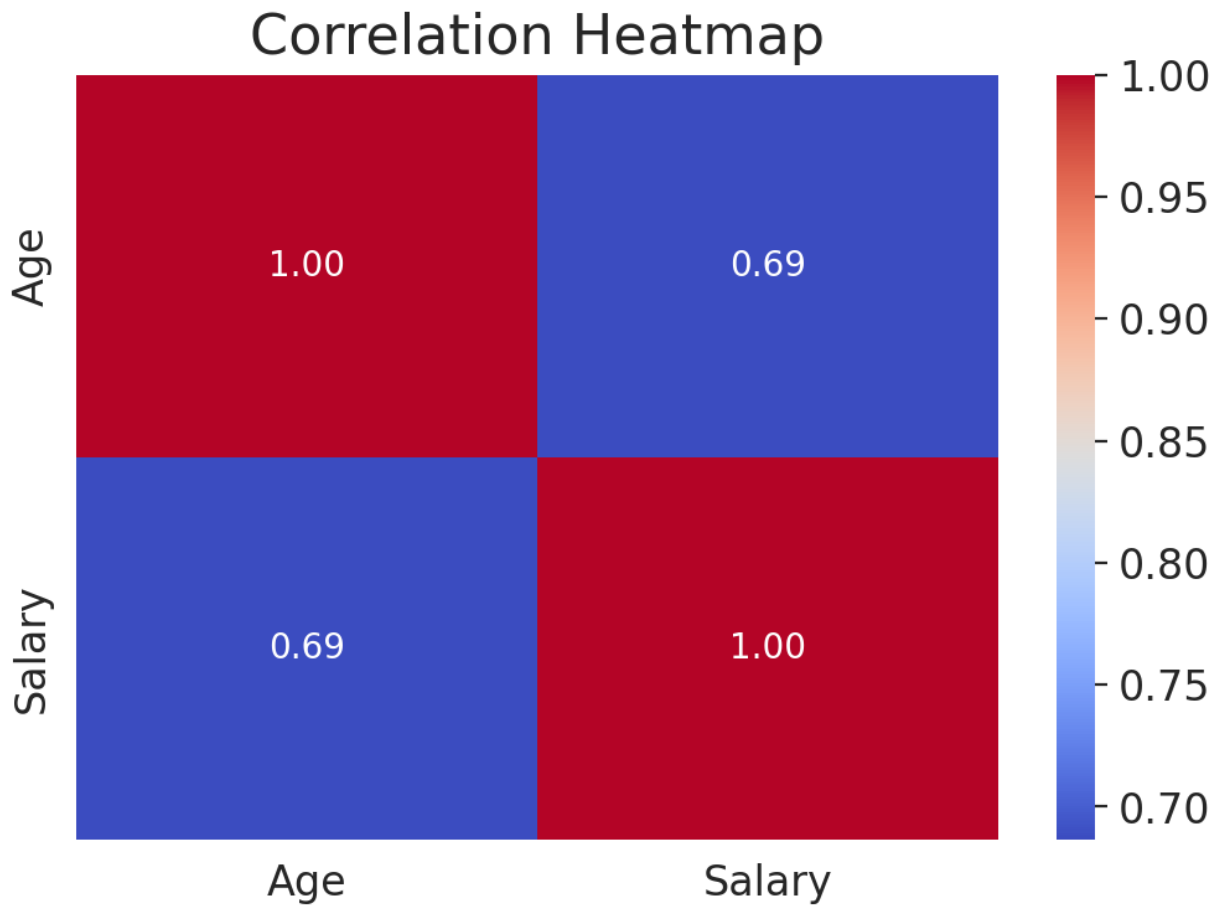



Salary Boxplot
Salary
50000
45000
40000
35000
30000


Age Boxplot
Age
30
28
26
24
22


Age vs Salary
Salary
50000
45000
40000
35000
30000
22
24
26
28
30
Age


Department Distribution
count
6
5
4
3
2
1
0
Hr
It
Finance
Marketing
Department



Univariate Analysis

- *Age Distribution: Histogram and summary statistics showed that ages were fairly evenly distributed.*
- *Salary Distribution: A boxplot indicated a normal salary range after outlier removal.*
- *Department Distribution: A bar plot showed the frequency of employees in each department.*

Bivariate Analysis

- *Age vs. Salary: A scatter plot suggested a weak positive correlation between age and salary.*
- *Correlation Analysis: A heatmap of numerical variables showed a mild correlation between age and salary.*

Multivariate Analysis

- *Pair Plots: A pair plot was generated to visualize relationships between multiple numerical variables simultaneously.*
- *Heatmap Analysis: A correlation heatmap was created to observe interactions across all numerical features.*
- *Grouped Comparisons: Salary distributions were analyzed within different departments to understand how salary varies across categories.*