1. Handling Missing Data

- **Objective**: Identify missing values and decide how to handle them.

  <mark>1 missing value in income column and we replace it with average = 49878.349</mark>

2. Outlier Detection and Removal

- **Objective**: Identify data points significantly different from the rest.

Average without removing outliers	49983.03
Average with outliers	49878.346
Median	49625
Mode	72843
Variance	803817175
Standard Deviation	28353.615
Quartile-0	100
Quartile-1	25454
Quartile-2	49625
Quartile-3	74391
Quartile-4	99967
IQR	48937
Lower Limit	-47951.5
Higher Limit	147796.5

<mark>As there is very less difference between average with outliers and average without outliers we can remove rows with outliers.</mark>
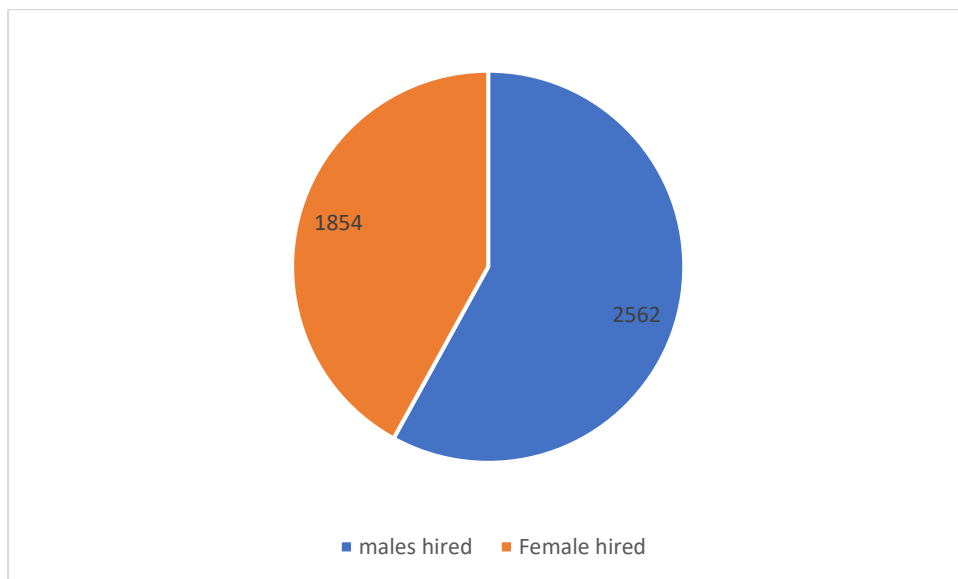
<mark>There are 3 outliers and we can remove them.</mark>

Summarizing Data

- **Objective**: Calculate and present key metrics.

- **Steps in Excel**:

  o Use:

    ▪ =AVERAGE for average salary.

    ▪ =COUNTIF to calculate male/female hires.

A. Hiring Analysis: Gender Distribution

- **Objective: Determine how many males and females have been hired.**

Males hired   =COUNTIFS(C2:C7166,"Hired",D2:D7166,"Male")   2562
Females hired   =COUNTIFS(C2:C7166,"Hired",D2:D7166,"Female")   1854

B. Salary Analysis: Average Salary

Objective: Calculate the average salary.

=Average((SalaryRange)

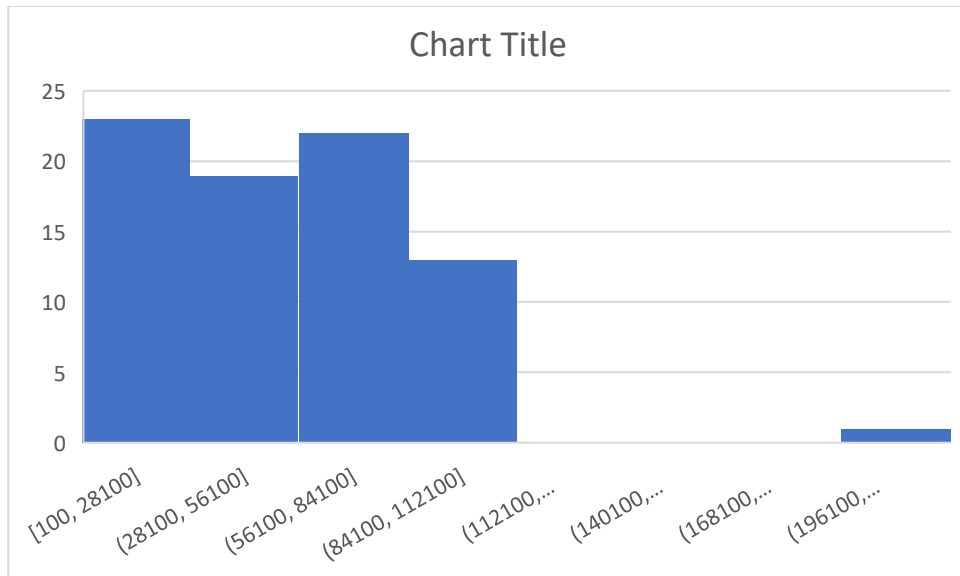|	
Average without removing outliers	49983.03
Average with outliers	49878.346

C. Salary Distribution: Class Intervals

Objective: Create class intervals to understand salary distribution.

Use the MAX and MIN functions to determine the salary range:
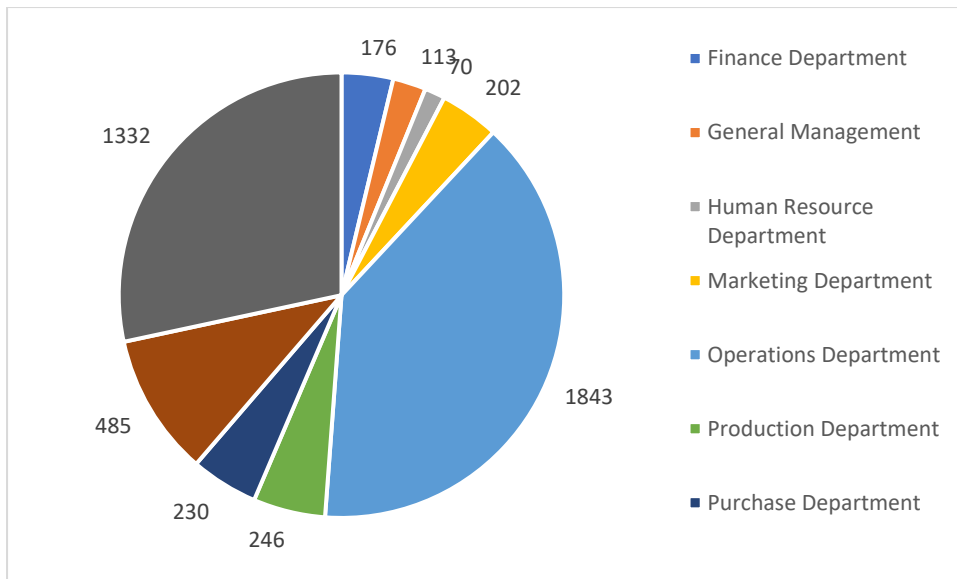
Max = 400000

Min = 100

Chart Title

Bucket	Value
[100, 28100]	23
(28100, 56100]	19
(56100, 84100]	22
(84100, 112100]	13
(112100,...	0
(140100,...	0
(168100,...	0
(196100,...	1

D. Departmental Analysis

Objective: Visualize the proportion of people working in different departments.

Group Data:

- Use a pivot table:
    - Rows: Department.
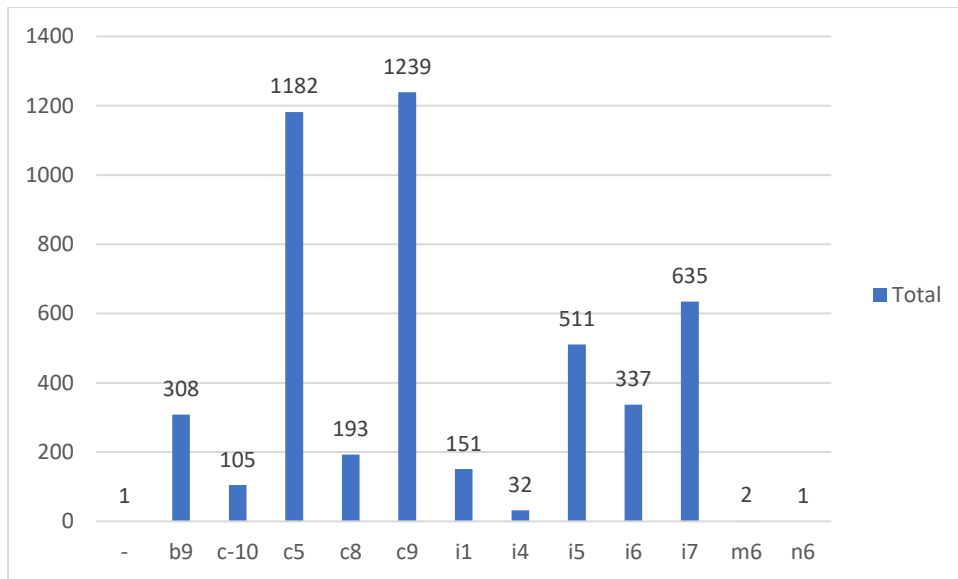   - Values: Count of Employees.
    - Filter: Status

E. Position Tier Analysis

Objective: Represent the distribution of positions across tiers.

Steps:

1. **Group Data**:

    o   Create a pivot table:

        ▪   Rows: Position Tier.

        ▪   Values: Count of Employees.

        ▪   Filter: Status

Related Excel File