Employee Data Analysis Report

Done by: Dareen Fawwaz

Table of Contents

[Table of Figures 3](#_Toc211941987)

[Introduction: 4](#_Toc211941988)

[1. Data Exploration: 4](#_Toc211941989)

[2. Data Cleaning: 5](#_Toc211941990)

[3. Modify first 5 rows: 5](#_Toc211941991)

[4. Largest Salary Employee: 5](#_Toc211941992)

[5. Department Statistics: 6](#_Toc211941993)

[6. Department + Ethnicity Statistics: 6](#_Toc211941994)

[7. Saving Work into file: 7](#_Toc211941995)

[Graphs Analysis: 8](#_Toc211941996)

[I. Graph(1): a box plot graph that represent the age distribution of employees by the department 8](#_Toc211941997)

[II. Graph(2): a scatter plot the annual salary and age relation grouped by department 9](#_Toc211941998)

[III. Graph(3): a stacked bar graph that show the no. of employees from each gender in each department 9](#_Toc211941999)

[IV. Graph(4): a histogram plot that shows the bonuses distribution 10](#_Toc211942000)

# Table of Figures

[Figure 1 8](file:///C:\Users\daree\Desktop\task%202\Employee%20Data%20Analysis%20Report.docx#_Toc211941937)

[Figure 2 9](file:///C:\Users\daree\Desktop\task%202\Employee%20Data%20Analysis%20Report.docx#_Toc211941938)

[Figure 3 10](file:///C:\Users\daree\Desktop\task%202\Employee%20Data%20Analysis%20Report.docx#_Toc211941939)

[Figure 4 11](file:///C:\Users\daree\Desktop\task%202\Employee%20Data%20Analysis%20Report.docx#_Toc211941940)

# Introduction:

this analysis purpose to explore the employee data and clean them from any wrong data and fill blank spots with average value of the column and after that analyse according to some columns

## Data Exploration:

Columns and data types:

* Data columns (total 14 columns):
* # Column Non-Null Count Dtype
* --- ------ -------------- -----
* 0 EEID 1000 non-null object
* 1 Full Name 998 non-null object
* 2 Job Title 999 non-null object
* 3 Department 998 non-null object
* 4 Business Unit 1000 non-null object
* 5 Gender 999 non-null object
* 6 Ethnicity 993 non-null object
* 7 Age 994 non-null float64
* 8 Hire Date 993 non-null datetime64[ns]
* 9 Annual Salary 989 non-null float64
* 10 Bonus % 992 non-null float64
* 11 Country 998 non-null object
* 12 City 998 non-null object
* 13 Exit Date 85 non-null datetime64[ns]
* dtypes: datetime64[ns](2), float64(3), object(9)
* memory usage: 109.5+ KB
* None

Missing values:

EEID 0

Full Name 2

Job Title 1

Department 2

Business Unit 0

Gender 1

Ethnicity 7

Age 6

Hire Date 7

Annual Salary 11

Bonus % 8

Country 2

City 2

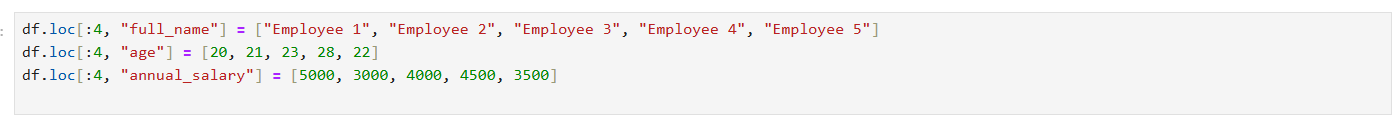
Exit Date 915

dtype: int64

## Data Cleaning:



## Modify first 5 rows:



## Largest Salary Employee:

eeid E04354

full\_name Raelynn Rios

job\_title Vice President

department Sales

business\_unit Manufacturing

gender Female

ethnicity Latino

age 43.0

hire\_date 2016-08-21 00:00:00

annual\_salary 258498.0

bonus\_% 0.35

country United States

city Columbus

exit\_date NaT

Name: 989, dtype: object

## Department Statistics:

average\_age average\_salary

department

Accounting 43.656250 123146.947917

Engineering 45.670886 109053.101266

Finance 44.708333 120894.616667

Human Resources 44.443548 117907.088710

IT 44.070833 96634.050000

Marketing 43.216667 129663.033333

Sales 43.614286 111120.635714

## Department + Ethnicity Statistics:

maxAge minAge medianSalary

department ethnicity

Accounting Asian 64.0 25.0 114893.0

Black 51.0 30.0 91853.0

Caucasian 62.0 26.0 121159.5

Latino 62.0 26.0 92317.0

Engineering Asian 65.0 25.0 91230.0

Black 56.0 27.0 74412.0

Caucasian 64.0 25.0 96057.0

Latino 64.0 27.0 96818.0

Finance Asian 64.0 22.0 124928.0

Black 65.0 25.0 131652.0

Caucasian 64.0 23.0 95045.0

Latino 65.0 25.0 124553.0

Human Resources Asian 64.0 25.0 125871.5

Black 55.0 25.0 142318.0

Caucasian 64.0 26.0 100364.5

Latino 65.0 29.0 101985.0

IT Asian 65.0 21.0 89494.0

Black 61.0 20.0 72901.5

Caucasian 65.0 26.0 81218.0

Latino 64.0 25.0 83934.0

Marketing Asian 65.0 25.0 147752.0

Black 65.0 27.0 86075.5

Caucasian 65.0 25.0 113135.0

Latino 65.0 29.0 97063.5

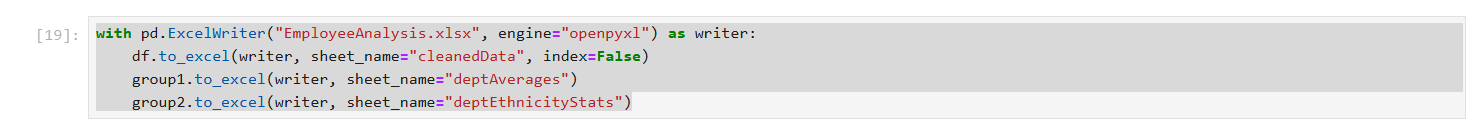
Sales Asian 64.0 25.0 89528.5

Black 65.0 26.0 94246.0

Caucasian 63.0 26.0 96092.0

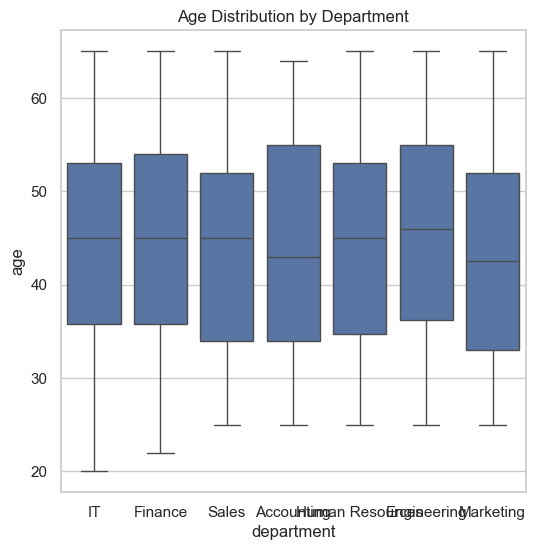
Latino 61.0 26.0 113269.0

## Saving Work into file:



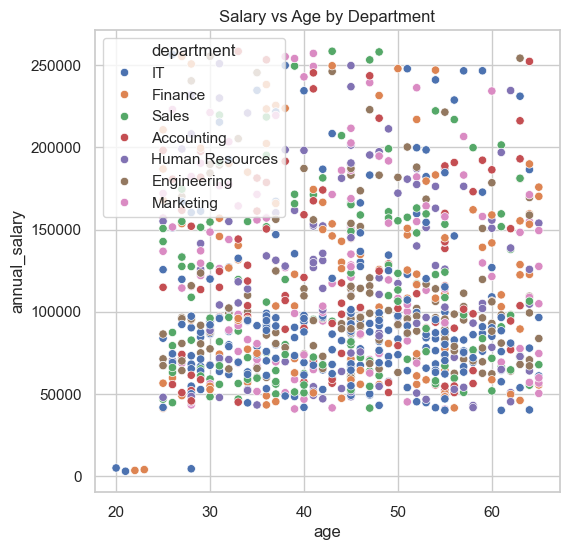
# Graphs Analysis:

## Graph(1): a box plot graph that represent the age distribution of employees by the department



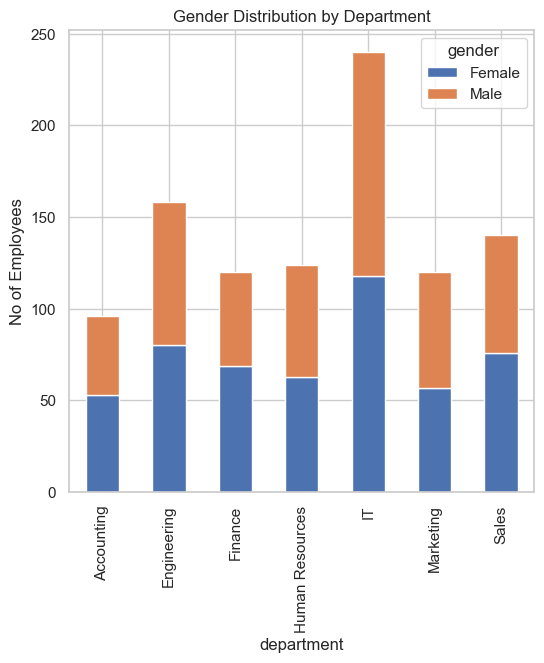
Figure

## Graph(2): a scatter plot the annual salary and age relation grouped by department



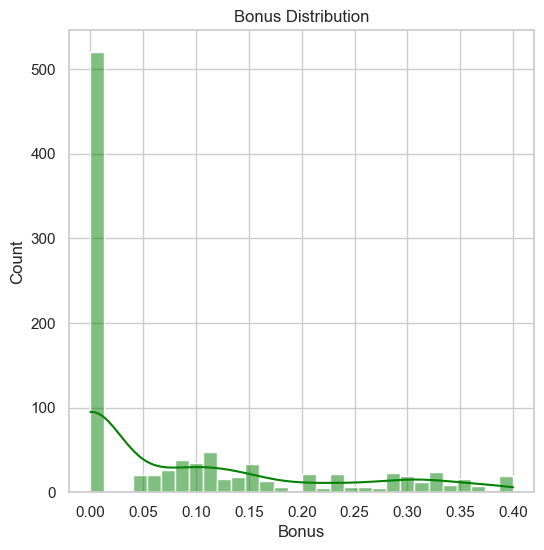
Figure

## Graph(3): a stacked bar graph that show the no. of employees from each gender in each department



Figure

## Graph(4): a histogram plot that shows the bonuses distribution



Figure