****

**Employee Data Analysis**

**Name：Dana Allan**

**Index:**

[Overview:](#_Toc3790)

[Data Cleaning:](#_Toc6168)

[Edit HEAD() :](#_Toc32422)

[print the row with the largest salary:](#_Toc6189)

[Save work in a new Excel file:](#_Toc9908)

[Matplotlib:](#_Toc25015)

[Conclusion:](#_Toc21783)

# 

# Overview:

We have sample data of employees in a certain company , we have to clean data,edit head(),print the row with the largest salary ,Save work in a new Excel file ,create graphs and plots using matplotlib

## Data Cleaning:

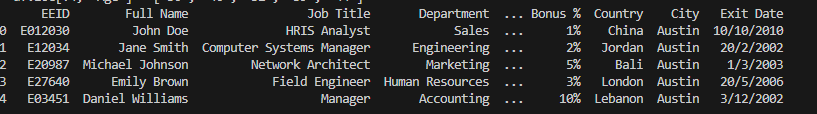
I use dropna(inplace=True) , will remove all rows containing NULL values from the original dataframe.

### Edit HEAD() :

I use df.loc[:4,’name of column’] to edit the values of the first 5 rows by inputting different values to each column in row

Eg:df.loc[:4, 'Full Name'] = ['John Doe', 'Jane Smith', 'Michael Johnson', 'Emily Brown', 'Daniel Williams']

Result:



#### print the row with the largest salary:

Return the whole row with the largest annual salary using the function( .max())

Result:

The largest salary = 550123.0

EEID E03451

Full Name Daniel Williams

Job Title Manager

Department Accounting

Business Unit Research & Development

Gender Female

Ethnicity Black

Age 44

Hire Date 17/9/2017

Annual Salary $550,123

Bonus % 10%

Country Lebanon

City Austin

Exit Date 3/12/2002

## Save work in a new Excel file:

I Use this command

( df.to\_excel('C:/Users/Admin/Downloads/Employee Sample Data - Copy.xlsx') )

##### Matplotlib:

Here we will analyze the data and represent the data on graphs.

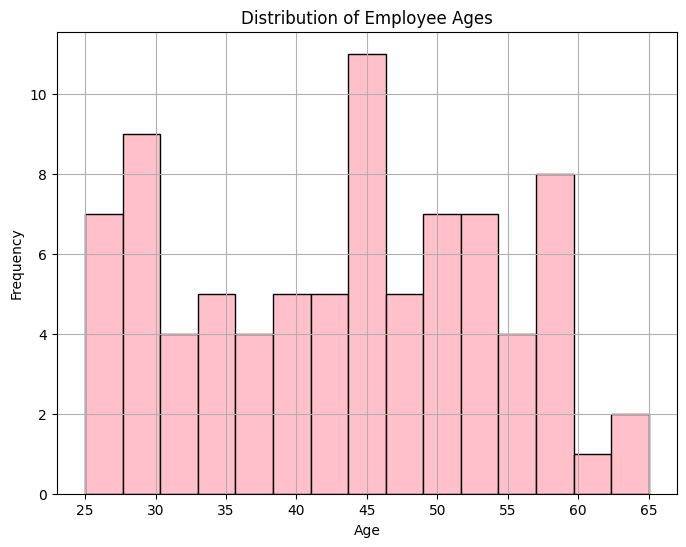


Figure 1.1

The youngest employees age in this company is 25

The oldest employees age in this company is 65

The most common employees age in this company is 45

The least common employees age in this company is 60-63

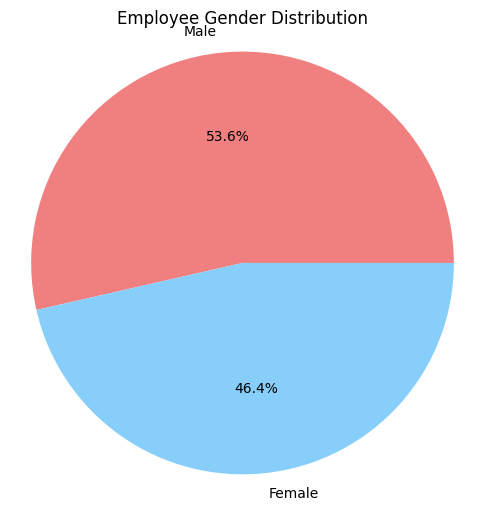


figure1.2

The percentage of male in company is higher about 7.2% than female

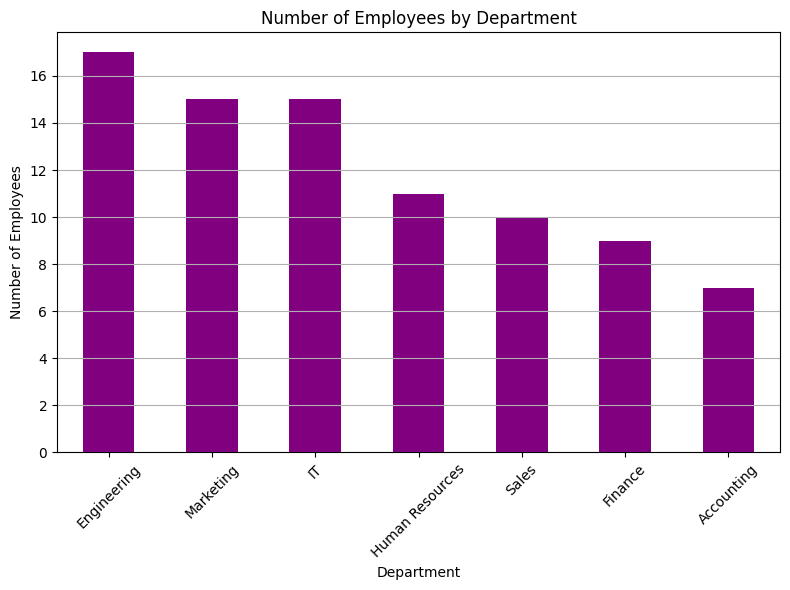


figure1.3

Engineering department have the highest number of employees

Marketing & It department have tha same number of employees

Accounting department have the least number of employees

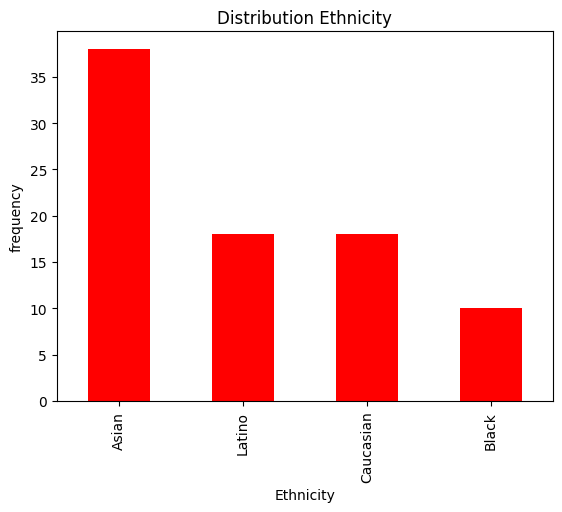
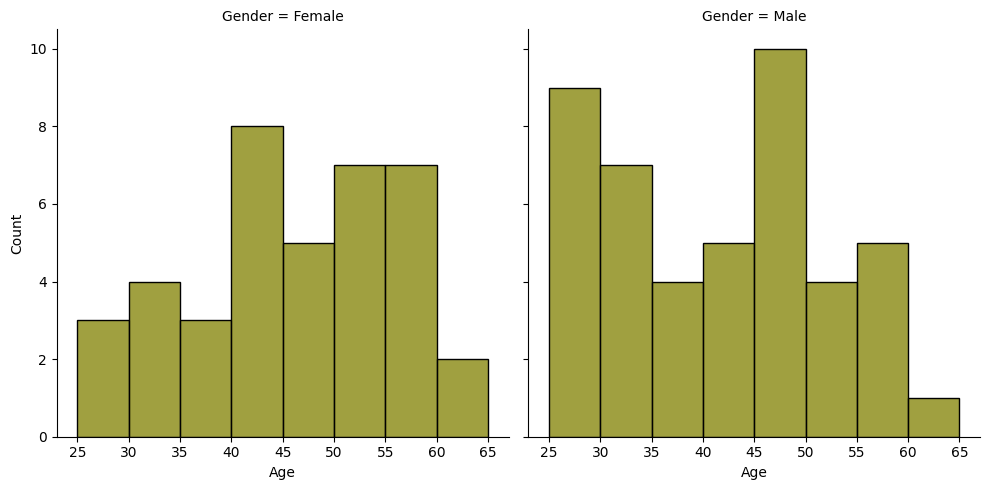


figure1.4

Asian employees is the highest number of employess in the company

Black employees is the lowest number of employess in the company

Caucasian &latino is have the nearest number of employees



Distrubution of age&gender

The most common of female age in this company between 40-45

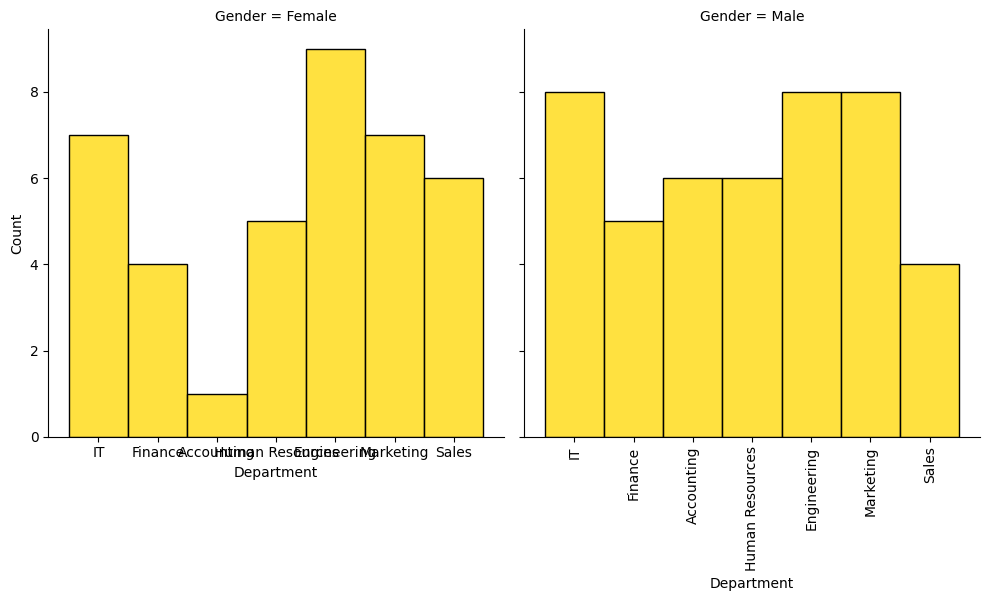
The least common of female age in this company between 60-65

The most common of male age in this company between 45-50

The least common of male age in this company between 60-65

Female of ages (60-65) is most than male in the same ages (60-65)

Male of ages(25-30) is most than female in the same ages(25-30)



distribution of department&gender

figure1.6

The highest number of female works on Human Resources

The least number of female work on Accounting

The highest number of male works on (Engineering , markiting , It)

The least number of male work on sales

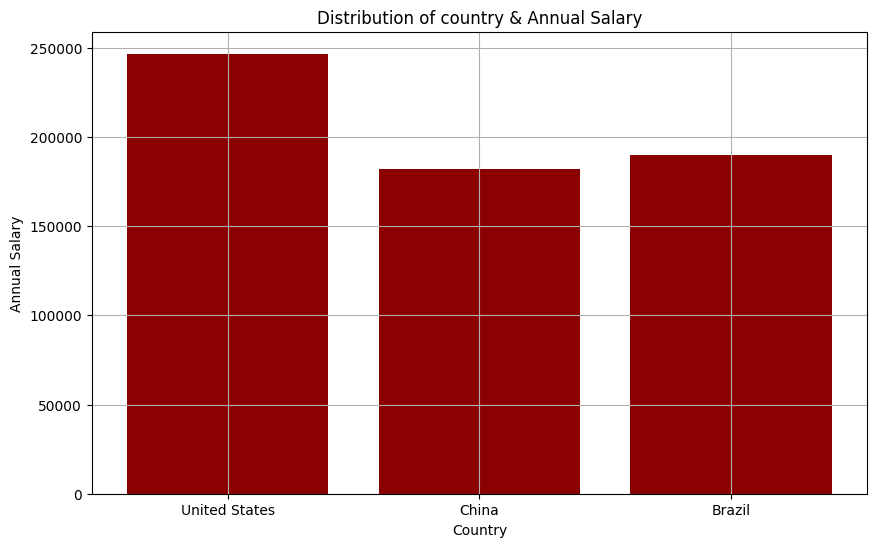


figure1.7

United states Get the highest annual salary

China get the lowest annual salary

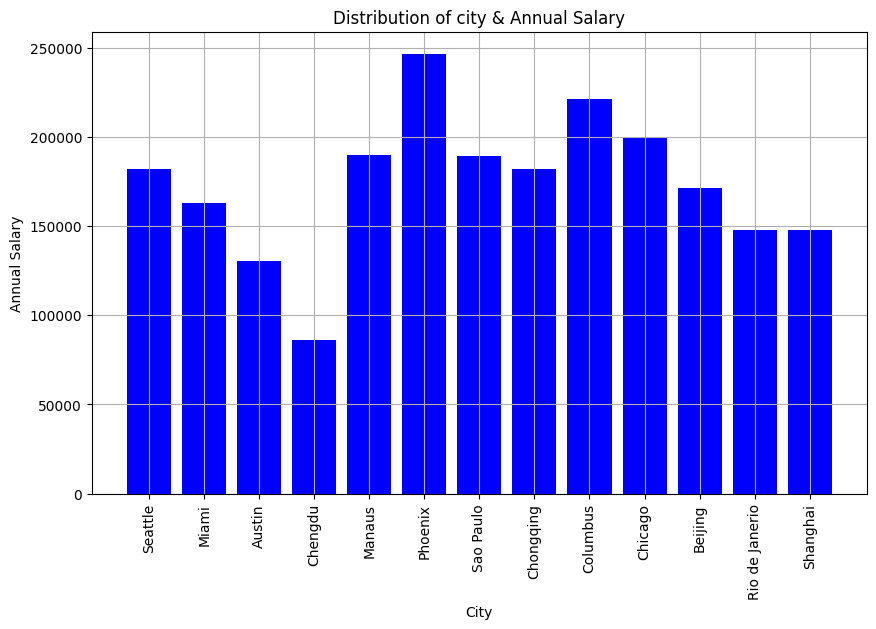


figure1.8

phoenix city get the highest annual salary

Rio de Janerio & shanghai city get the same annual salary

Chengdu city get the lowest annual salary

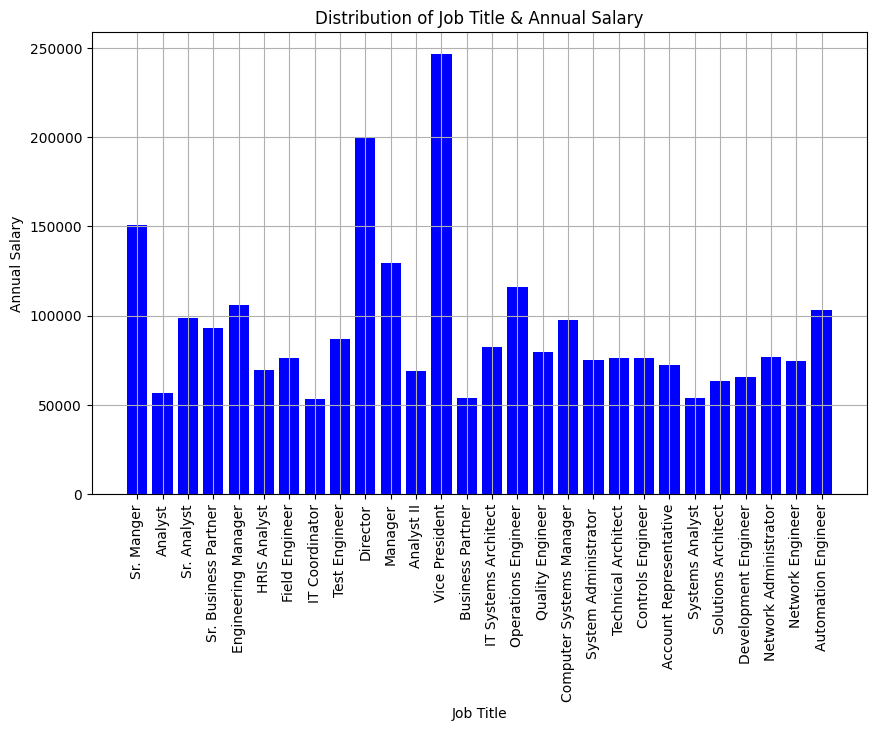


figure1.9

vice president job get the highest annual salary

System analyst &Business Partner & It coordinator get the lowest annual salary

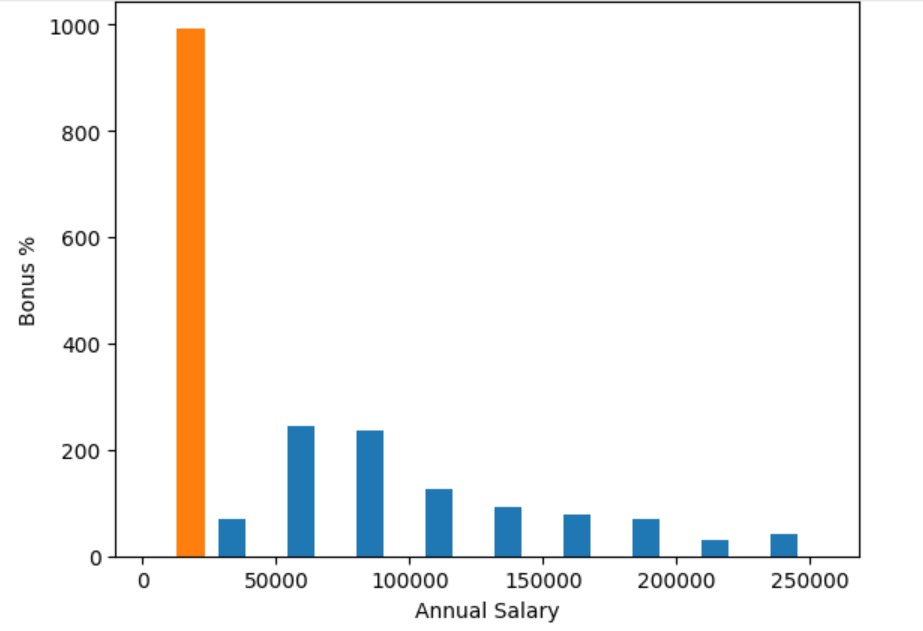


figure1.10

Bonus Percentage for Different Salary Ranges:

$0: There's a significant spike in bonus percentage, reaching close to 1000%. This is likely an anomaly orcould represent non-cash bonuses such as stock options, equity, or other forms of compensation.

$0 - $50,000: Bonus percentages are lower, but there's a noticeable increase as salaries approach $50,000.

$50,000 - $100,000: This range sees a substantial increase in bonus percentages, peaking around $100,000.

$100,000 - $150,000: Bonus percentages remain relatively high but start to decrease as salaries increase within this range.

$150,000 - $200,000: There is a further decline in bonus percentages.

$200,000 - $250,000: Bonus percentages are relatively low but remain present.

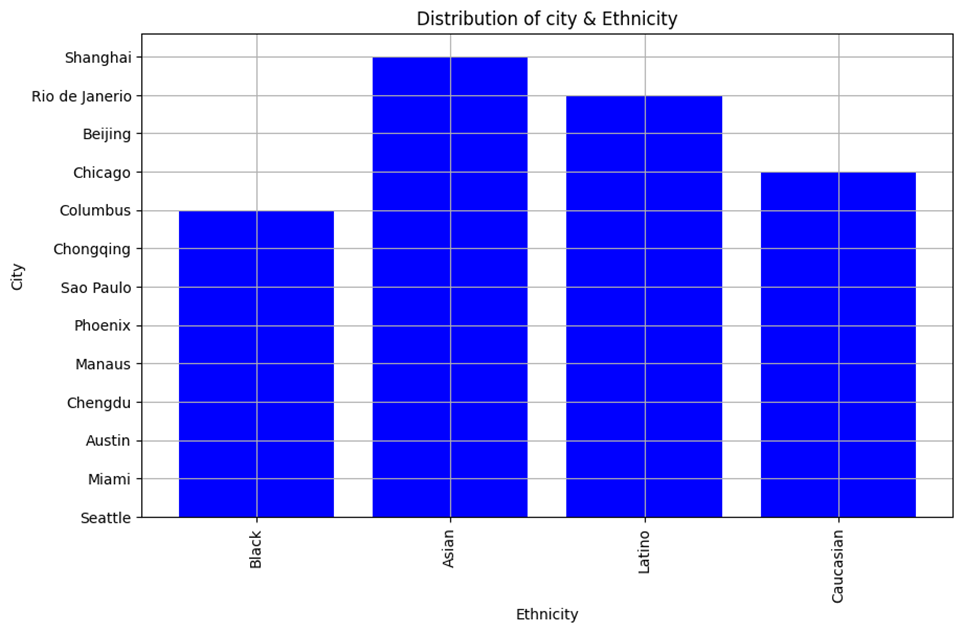


figure1.11

###### 

Columbus city have black ethnicity more than other city

Shanghai city have Asian ethnicity more than other city

Rio de janerio city have latino ethnicity more than other city

Chicago city have caucasian ethnicity more than other city

###### Conclusion:

I noticed that the number of black ethnicity employees in the company is small compared to other ethnicities, so my observation on this point is whether this could be bias and racism or a coincidence?

**Female Employees**:

Predominantly represented in Engineering and Human Resources departments.

Least represented in Accounting and IT departments.

**Male Employees**:

Predominantly represented in IT and Marketing departments.

Least represented in Sales and Accounting departments.

There is an overall inverse relationship between annual salary and bonus percentage (excluding the anomaly at $0 salary). Lower salaries correspond to higher bonus percentages, while higher salaries correspond to lower bonus percentages.