**Index:**

[Overview: 1](#_Toc3790)

[Data Cleaning: 1](#_Toc6168)

[Edit HEAD() : 1](#_Toc32422)

[print the row with the largest salary: 2](#_Toc6189)

[Save work in a new Excel file: 2](#_Toc9908)

[Matplotlib: 2](#_Toc25015)

[Conclusion: 12](#_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:**

EEID Full Name Job Title Department \

0 E012030 John Doe HRIS Analyst sales

1 E12034 Jane Smith Computer Systems Manager Engineering

2 E20987 Michael Johnson Network Architect Marketing

3 E27640 Emily Brown Field Engineer Human Resources

4 E03451 Daniel Williams Manager Accounting

Business Unit Gender Ethnicity Age Hire Date Annual Salary

Manufacturing Male Asian 30 20/1/2000 $119,7

Research & Development Female Black 40 1/1/2$222,130

Corporate Male Asian 32 9/8/1990 $110,230 Speciality Products Male Caucasian 39 9/2/1999 $420,110

Research & Development Female Black 44 17/9/2017 $550,123

Bonus % Country City Exit Date

1% China Austin 10/10/2010

2% Jordan Austin 20/2/2002

5% Bali Austin 1/3/2003

3% London Austin 20/5/2006

10% Lebanon Austin 3/12/2002

[9]:

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

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

**Result:**

**The largest salary = 246589.0**

**EEID E03247**

**Full Name Aaliyah Mai**

**Job Title Vice President**

**Department IT**

**Business Unit Speciality Products**

**Gender Female**

**Ethnicity Asian**

**Age 57.0**

**Hire Date 2016-11-11 00:00:00**

**Annual Salary 246589.0**

**Bonus % 0.33**

**Country United States**

**City Phoenix**

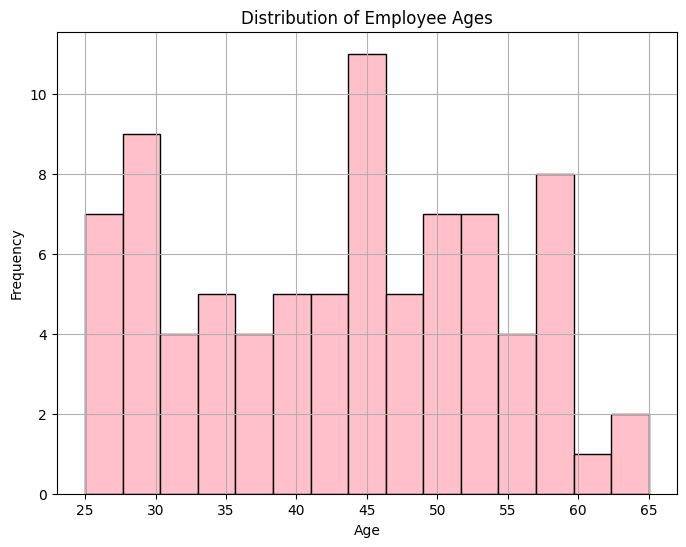
**Exit Date 2017-03-26 00:00:00**

## 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.**

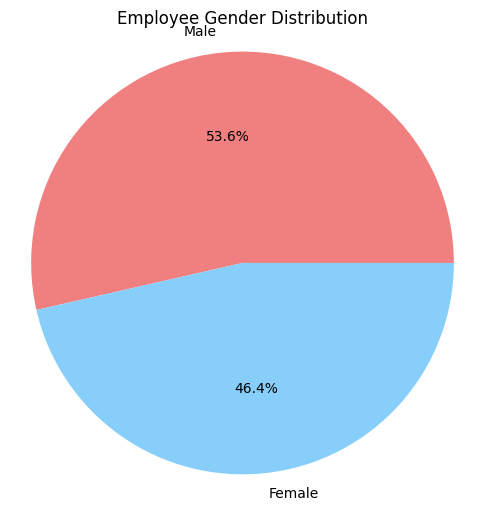


**The youngest employees age in this company is 25**

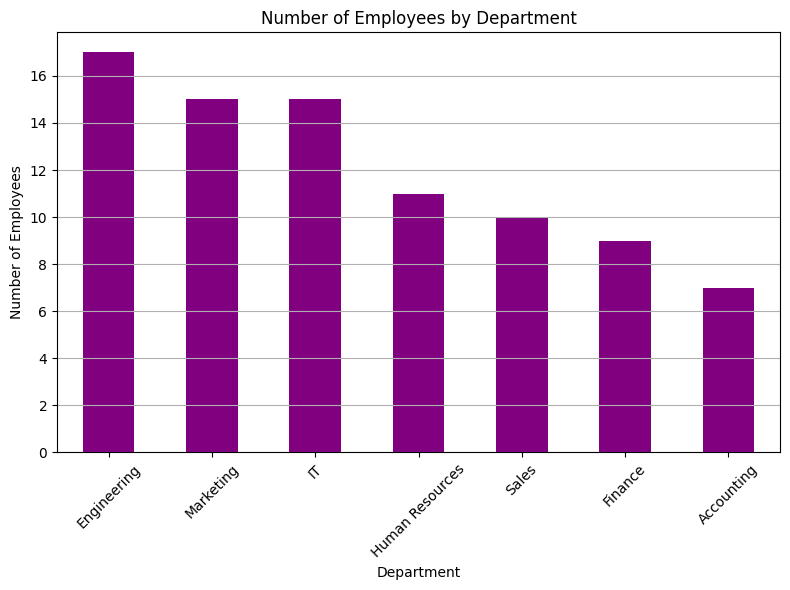
**The oldest employees age in this company is 65**

**The most common employees age in this company is 45**

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

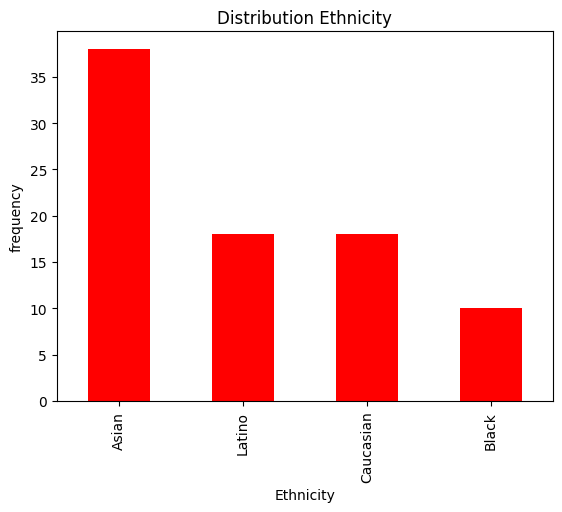


**The percentage of male in company is higher about 7.2% than female**

 **Engineering department have the highest number of employees**

**Marketing & It department have tha same number of employees**

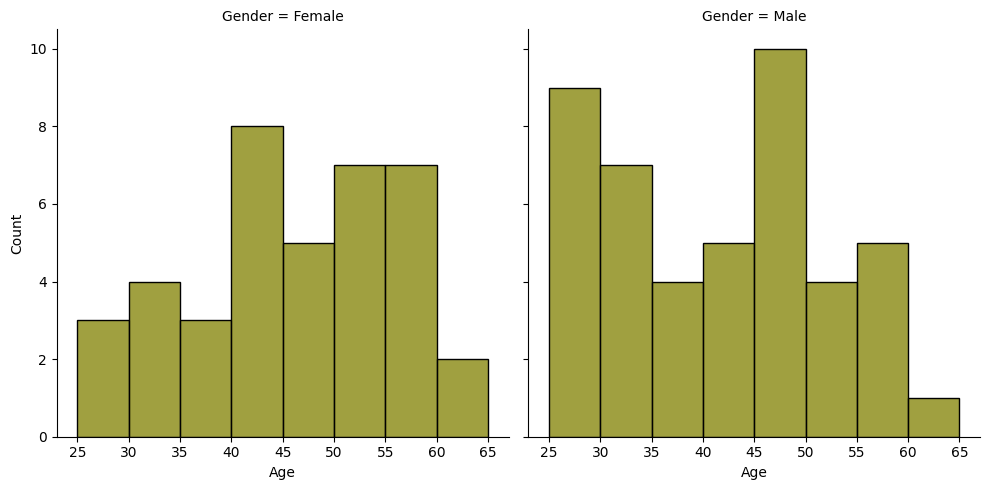
**Accounting department have the least number of employees**



**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**



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

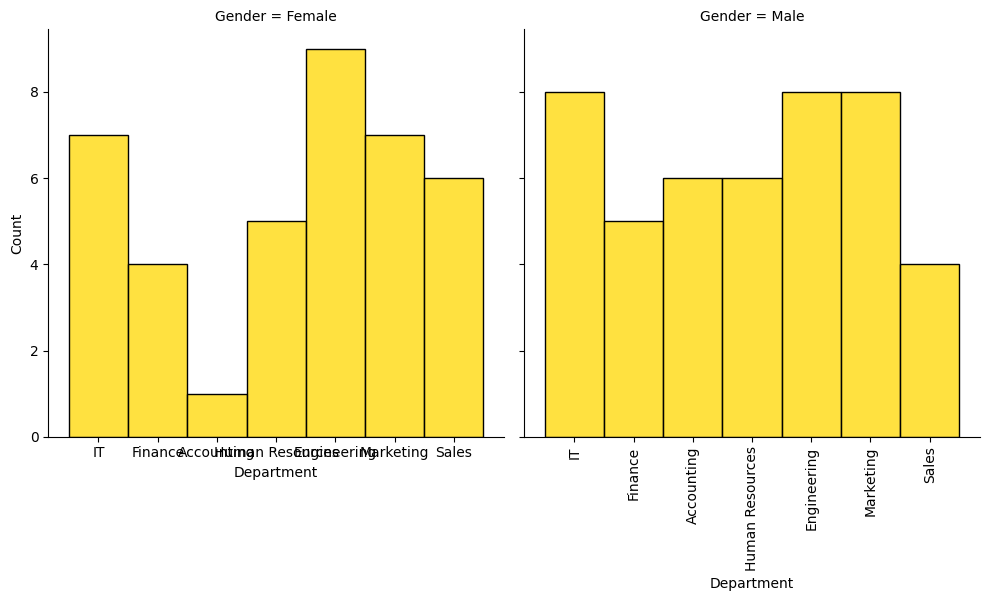
**the least common of female age in this company between 60-65**

**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)**

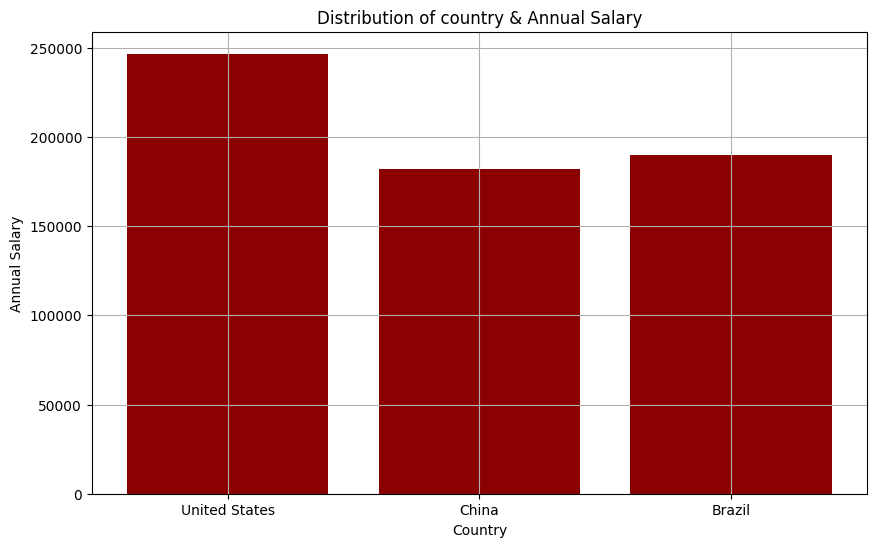
**Male of ages(25-30)is most than female in the same ages(25-30)**

**the highest number of female works on Human Resources**

**The least number of female work on Accounting**

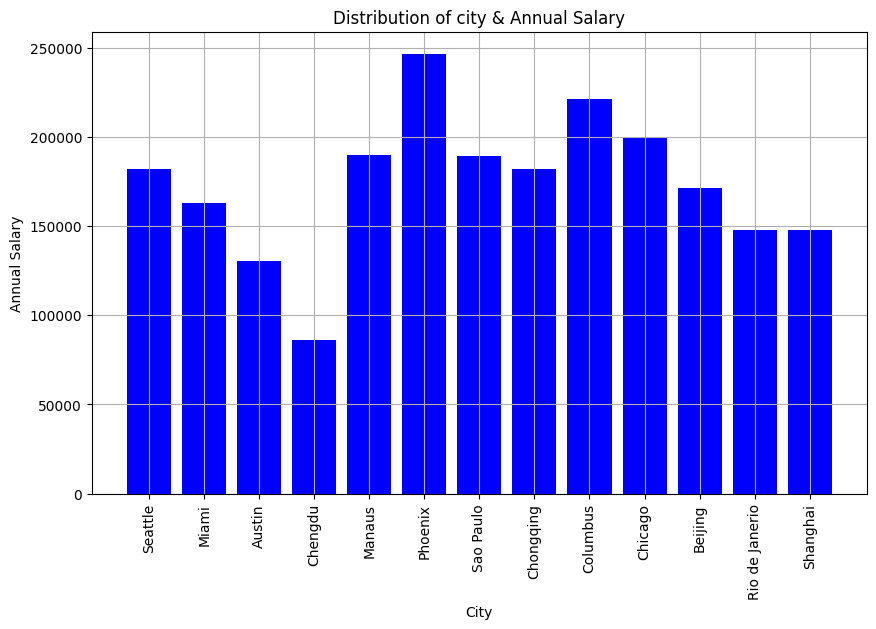
**the highest number of male works on (Engineering,markiting,It)**

**The least number of male work on sales**



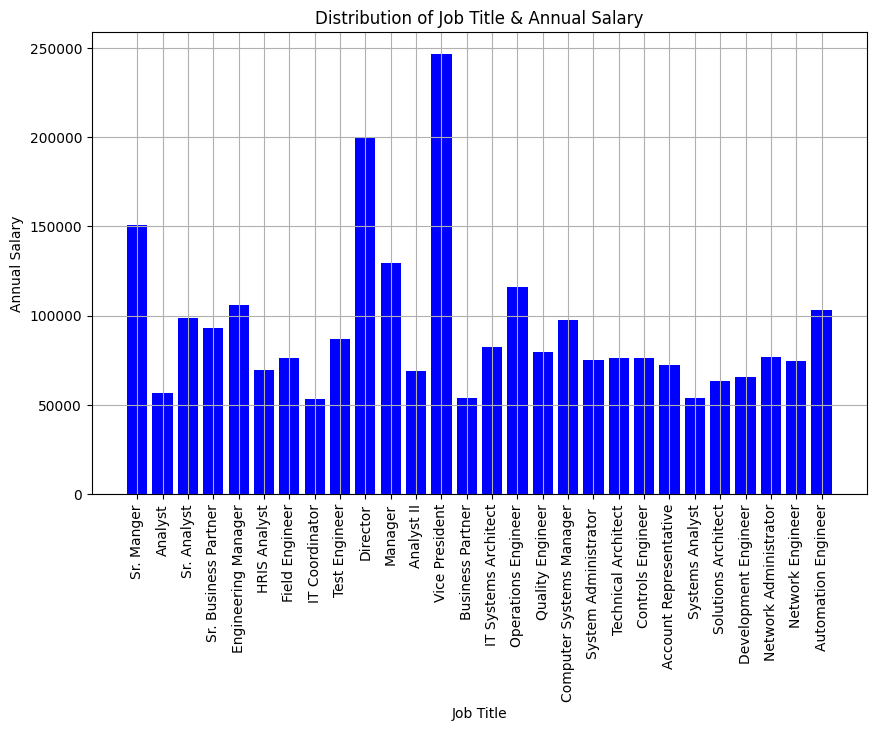
**United states Get the highest annual salary**

**China get the lowest annual salary**

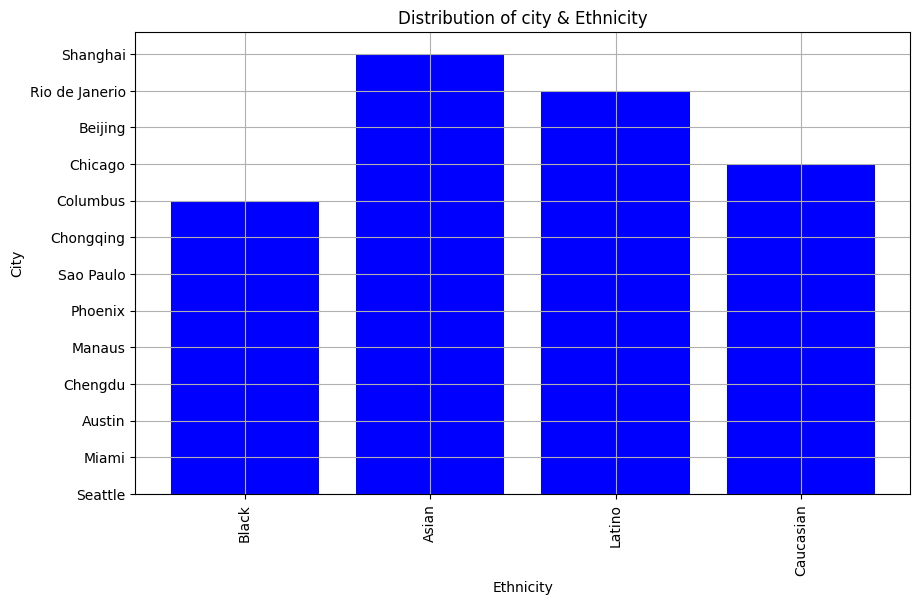
**phoenix city get the highest annual salary**

**Rio de Janerio & shanghai city get the same annual salary**

**Chengdu city get the lowest annual salary**

**vice president job get the highest annual salary**

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

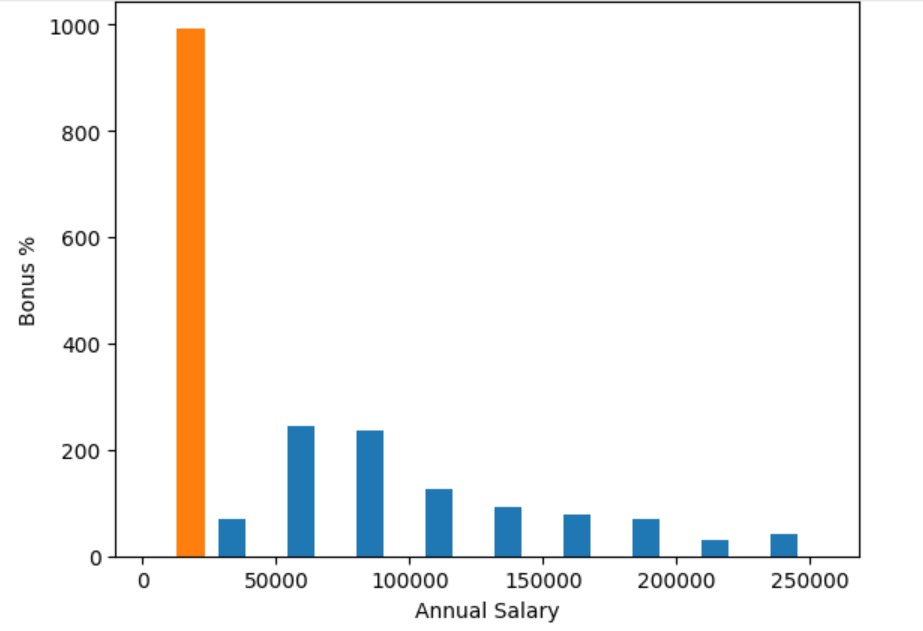


**Columbus city have black ethnicity more than other city**

**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**



**Bonus Percentage for Different Salary Ranges:**

**$0: There's a significant spike in bonus percentage, reaching close to 1000%. This is likely an anomaly or could 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.**

###### 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.**