# 

# **Asset-3@16x-1024x322**

# PROJECT REPORT

## (Project of Employee)

###### 

###### Written by : Anfal Majed

### Index:

Catalog

[ABSTRACT: 3](#_Toc1924)

[1. Data Cleaning & Analysis data: 3](#_Toc14130)

[1.Clean the data: 3](#_Toc18069)

[2.Change the first 5 rows (input any values),edit head(): 3](#_Toc19928)

[3.Print the row with the largest salary: 4](#_Toc9662)

[4. Save work in new Excel file: 5](#_Toc27436)

[2. MATPLOTLIB: 5](#_Toc12528)

[Age: 5](#_Toc25897)

[We will notice the following: 5](#_Toc8415)

[Country : 5](#_Toc27087)

[Ethnicity: 6](#_Toc19258)

[We will notice the following: 6](#_Toc8436)

[ANNUAL SALARY: 7](#_Toc26050)

[GENDER: 7](#_Toc27030)

[CITY: 8](#_Toc15055)

[JOB TITEL: 9](#_Toc30389)

[DEPARTMENT: 9](#_Toc9029)

[BUSINESS UNIT: 10](#_Toc28667)

[BOUNS%: 10](#_Toc30997)

[Annual Salary & Bonus % 11](#_Toc7365)

[Conclusion: 11](#_Toc3755)

### Figure Table:

[Figure 1 (Age) 6](#_Toc31520)

[Figure2(Country) :............................................................................................................................................7](#_Toc727)

[Figure 3(Ethnicity) 7](#_Toc20959)

[Figure 4(Annual salary) 8](#_Toc23547)

[Figure 5(Gender) 8](#_Toc24835)

[Figure 6 (city) 9](#_Toc23829)

[Figure 7 (Job titel) 10](#_Toc30593)

[Figure 8 (Department) 10](#_Toc22368)

[Figure 9 (Bussnise unit) 11](#_Toc26550)

[Figure 10 (Bouns) 11](#_Toc24677)

[Figure 11 (Annual salary &Bouns) 12](#_Toc22550)

ABSTRACT:

In the third week, we received an Excel file and were tasked with executing several commands, including analyzing and cleaning the file using the Python programming language. We utilized Jupyter Notebook to write the code, leveraging the Pandas library. Subsequently, we were required to transform the information into visual representations using Matplotlib for further comprehension of the file.

### Data Cleaning & Analysis data:

The first thing is to analyze and clean the data using the pandas library:

#### 1.Clean the data:

df.drop\_duplicates(inplace = True) ==> I used this command to remove duplicates between rows and replace them with correct values using the(inplace= True) . It will remove all duplicates from the original DataFrame.

{df["City"].fillna('United States', inplace = True)

df["Ethnicity"].fillna('Caucasian', inplace = True)} ==> I used fillna() to replace empty values I used for City row and Ethnicity row.

#### 2.Change the first 5 rows (input any values),edit head():

df["Job Title"]**.**iloc[:5]**=**"Computer Systems Manager"

df["Department"]**.**iloc[:5]**=**"IT"

df["Business Unit"]**.**iloc[:5]**=**"Manufacturing"

df["Ethnicity"]**.**iloc[:5]**=**"Caucasian"

df["Age"]**.**iloc[:5]**=**50

df["Exit Date"]**.**iloc[:5]**=**"10/16/2021"

df["City"]**.**iloc[:5]**=**"Phoenix"

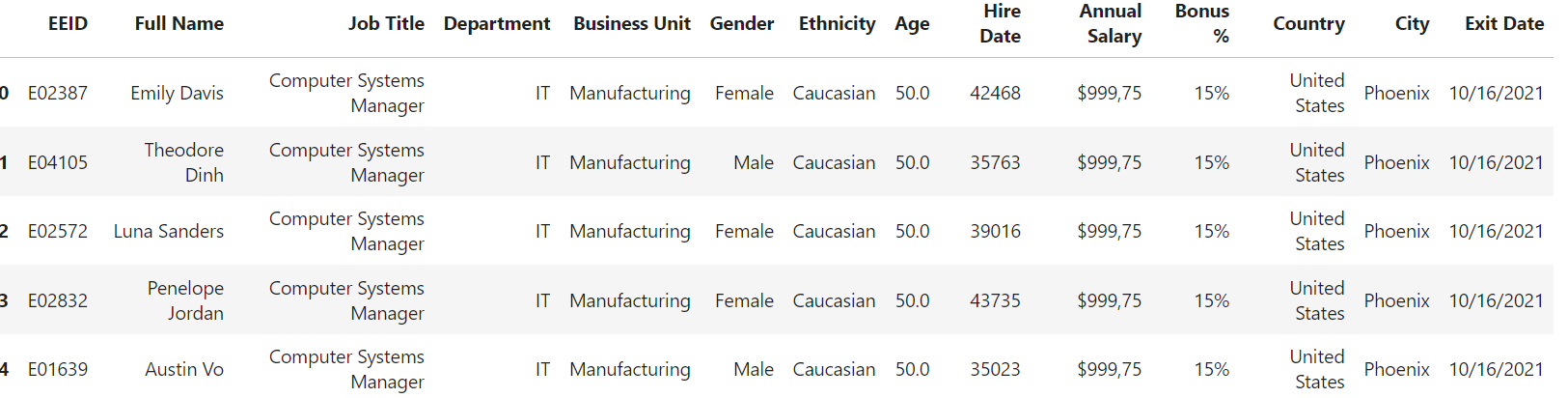
df["Country"]**.**iloc[:5]**=**"United States"

df["Bonus %"]**.**iloc[:5]**=**"15%"

df["Annual Salary"]**.**iloc[:5]**=**"$999,75"

Here I have replaced the first five rows with new values, so that each column has a unified value . The iloc property gets, or sets, the value(s) of the specified indexes.

THIS IS THE RESULT:



#### 3.Print the row with the largest salary:

temp\_column=pd.to\_numeric(df['Annual Salary'], error='coerce')

max\_salary\_index=temp\_column.idxmax()

max\_salary=temp\_column.max()

max\_salary\_row=df.loc[max\_salary\_index]

print(f"The largest salary={max\_salary} \n \n")

print(max\_salary\_row)

Here it will retrieve the value of the entire row that contains the highest salary value.

THIS IS THE 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 new Excel file:

df.to\_excel("Employee Sample Data an.xlsx")

Here I saved the Excel file with the extension(xlsx),Save all the changes that occurred in a new Excel file.

### MATPLOTLIB:

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

Age:

In the Age column, we represented the data as follows:

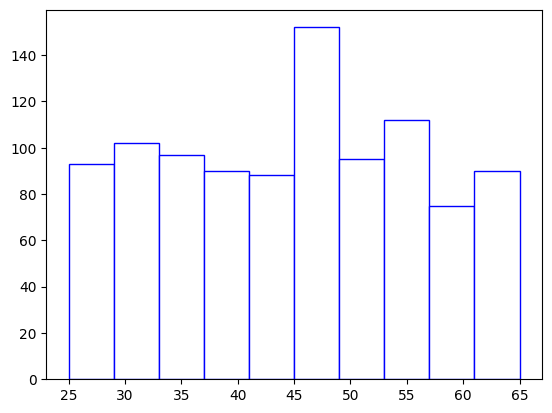


Figure 1

We will notice the following:

1. The highest number of employees is between the ages of 45-48.
2. The smallest number of employees is between the ages of 57-61.
3. From the ages of 25-44, their numbers are similar.

#### Country :

In the Country column, we represented the data as :

Country

United States 641

China 218

Brazil 139

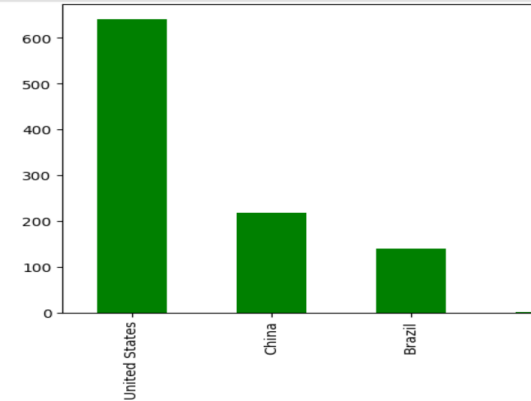
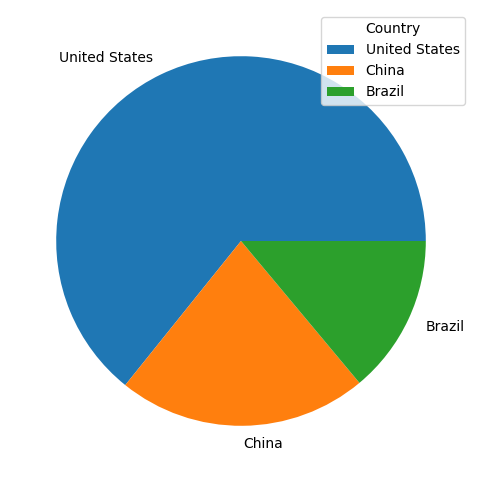


Figure 2

#### Ethnicity:

In the Ethnicity column, we represented the data as :

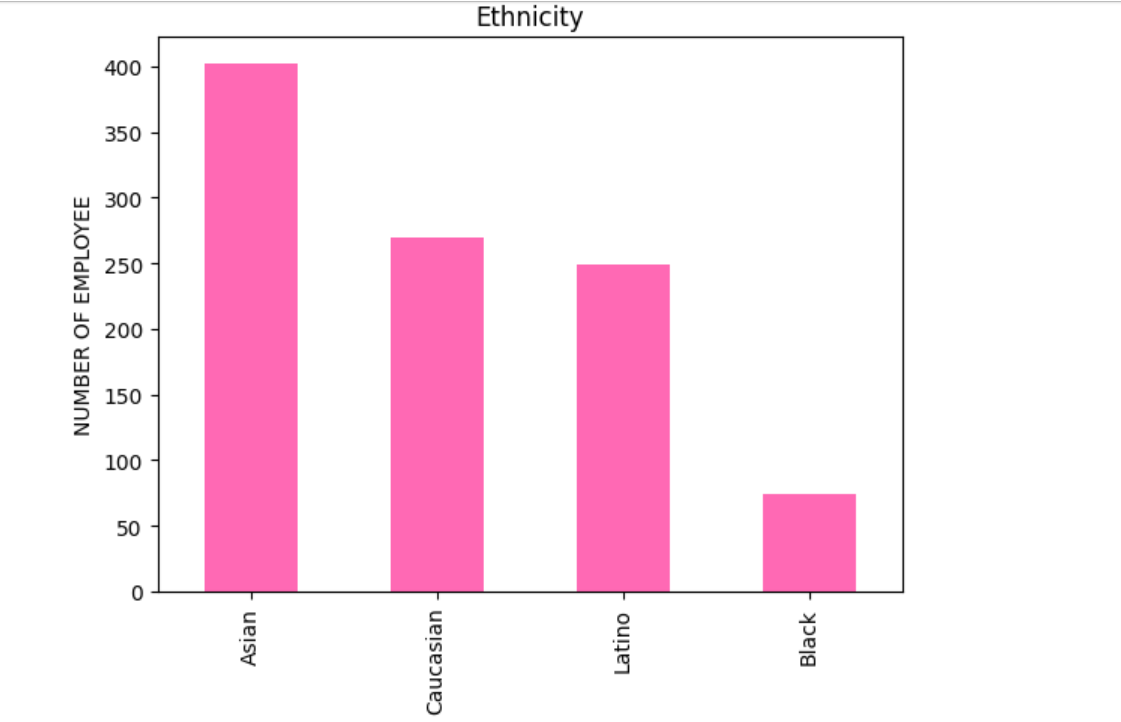


Figure 3

We will notice the following:

1-Asian employees is the highest number of employess in the company.

2-Black employees is the lowest number of employess in the company.

3-Caucasian &latino is have the nearest number of employees.

#### ANNUAL SALARY:

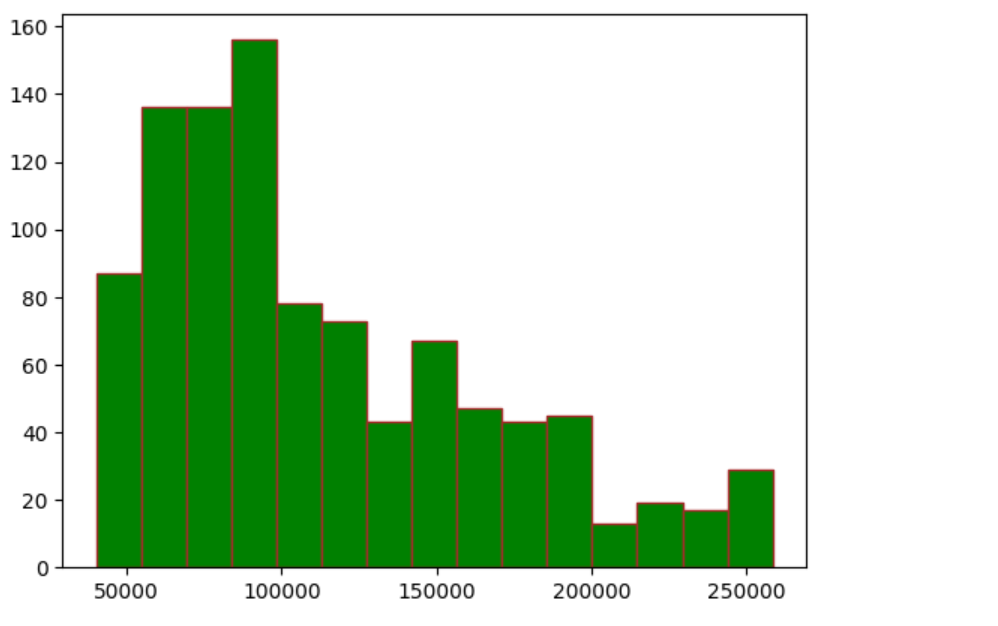


Figure 4

#### GENDER:

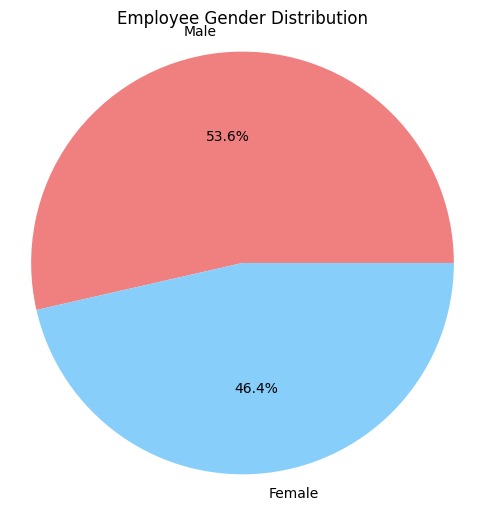


Figure 5

**The number of females in the company is less than the number of males by percent 7.2%.**

#### CITY:

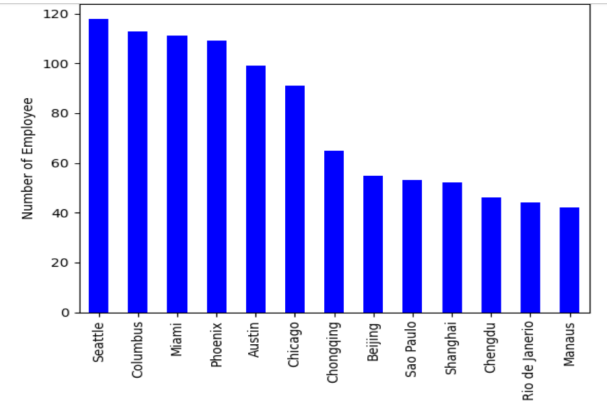
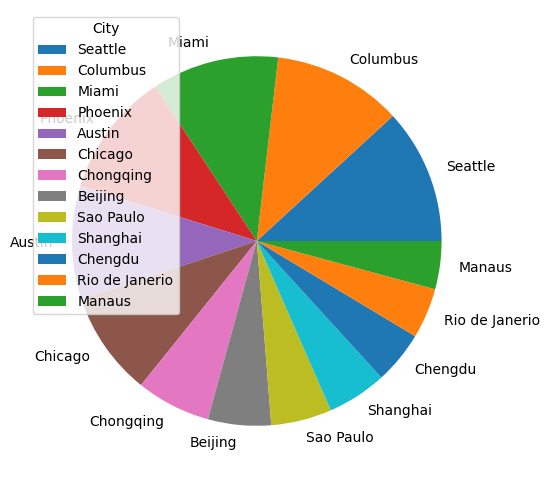


Figure 6

#### JOB TITEL:

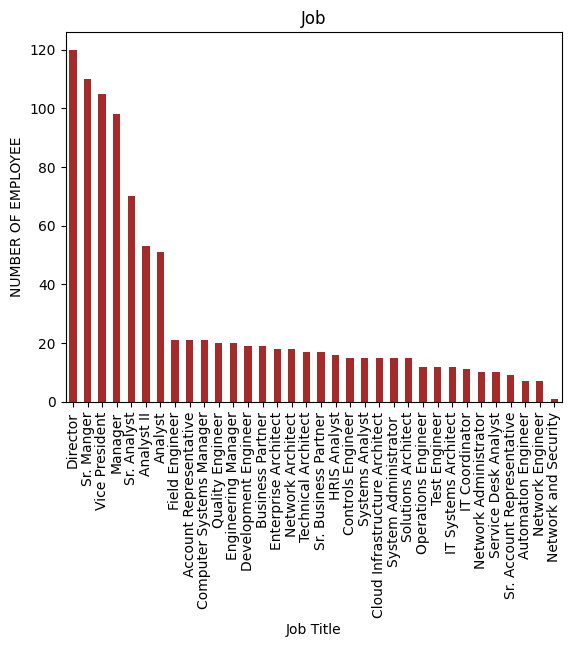
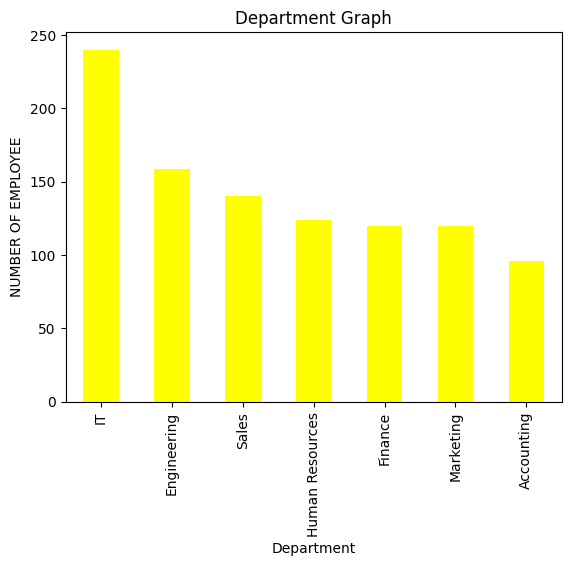


Figure 7

#### DEPARTMENT:

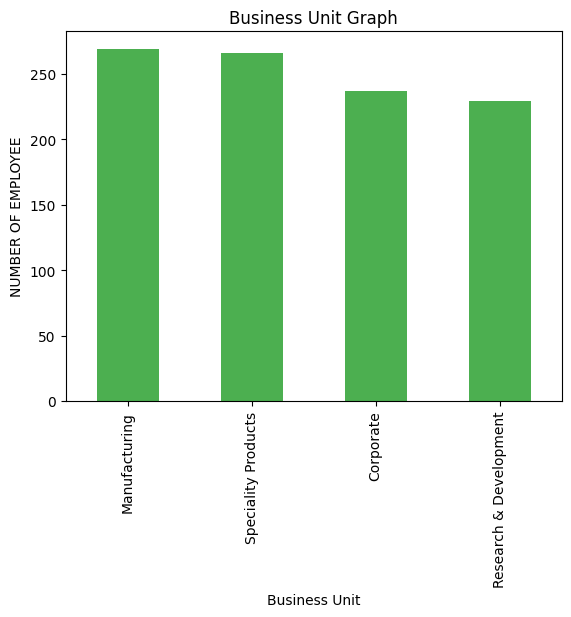


**The largest number of employees is in the IT department and the least in the IT department**

**Accounting, but not by a large percentage. As for the rest of the departments, the number of employees varies by a small amount**

Figure 8

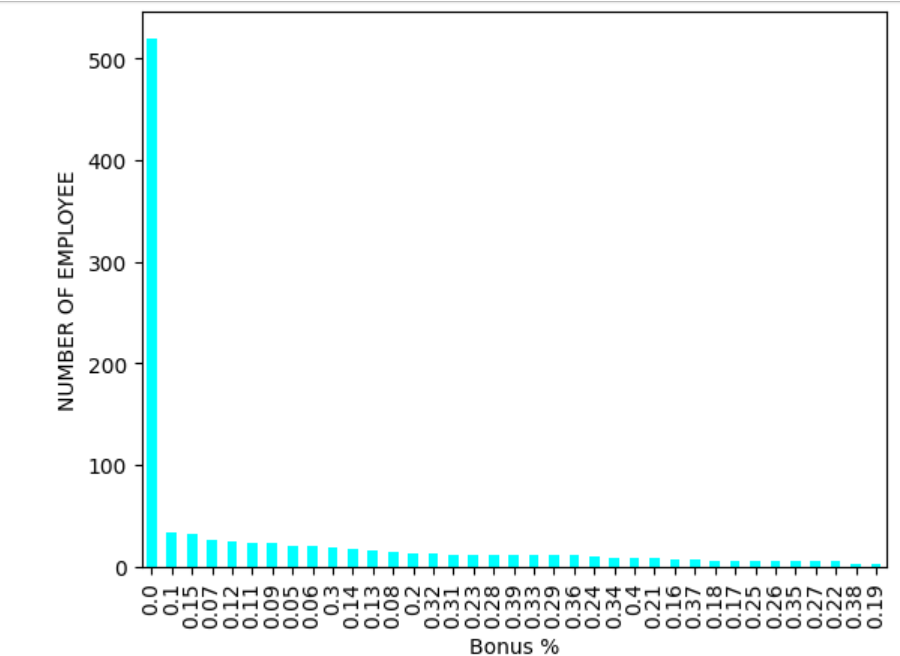
#### BUSINESS UNIT:



**The number of employees in each business unit is similar in number**

Figure 9

#### BOUNS%:



**We note that the largest number of employees do not receive any bonus, and the remaining percentages are close to each other.**

Figure 10

#### Annual Salary & Bonus %

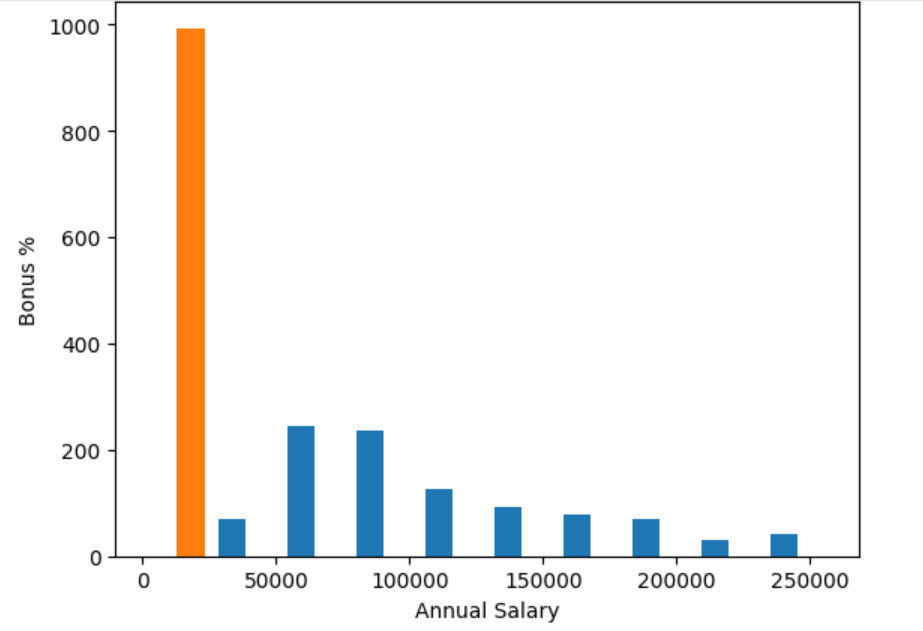


Figure 11

**Here, those with high salaries do not receive a high reward, but rather a small one, and those with somewhat low salaries receive a very high reward, which may reach 50% annually.**

Conclusion**:**

**At the end of the report, we learned how to modify files and how to get rid of incomprehensible or non-existent values ​​using the Pandas library, which analyzes data, and the Matplotlib library, which summarizes all existing information into understandable, clear information and uncomplicated graphics.**

**It was noted, based on the analysis, that black-skinned individuals constitute a smaller percentage in this area compared to populations of European descent, but this does not reflect any racial discrimination. Rather, it stems from complex social and economic reasons.”**

**“The data reveal significant disparities in the distribution of rewards among employees, reflecting growing economic gaps that need to be addressed comprehensively and fairly.”**