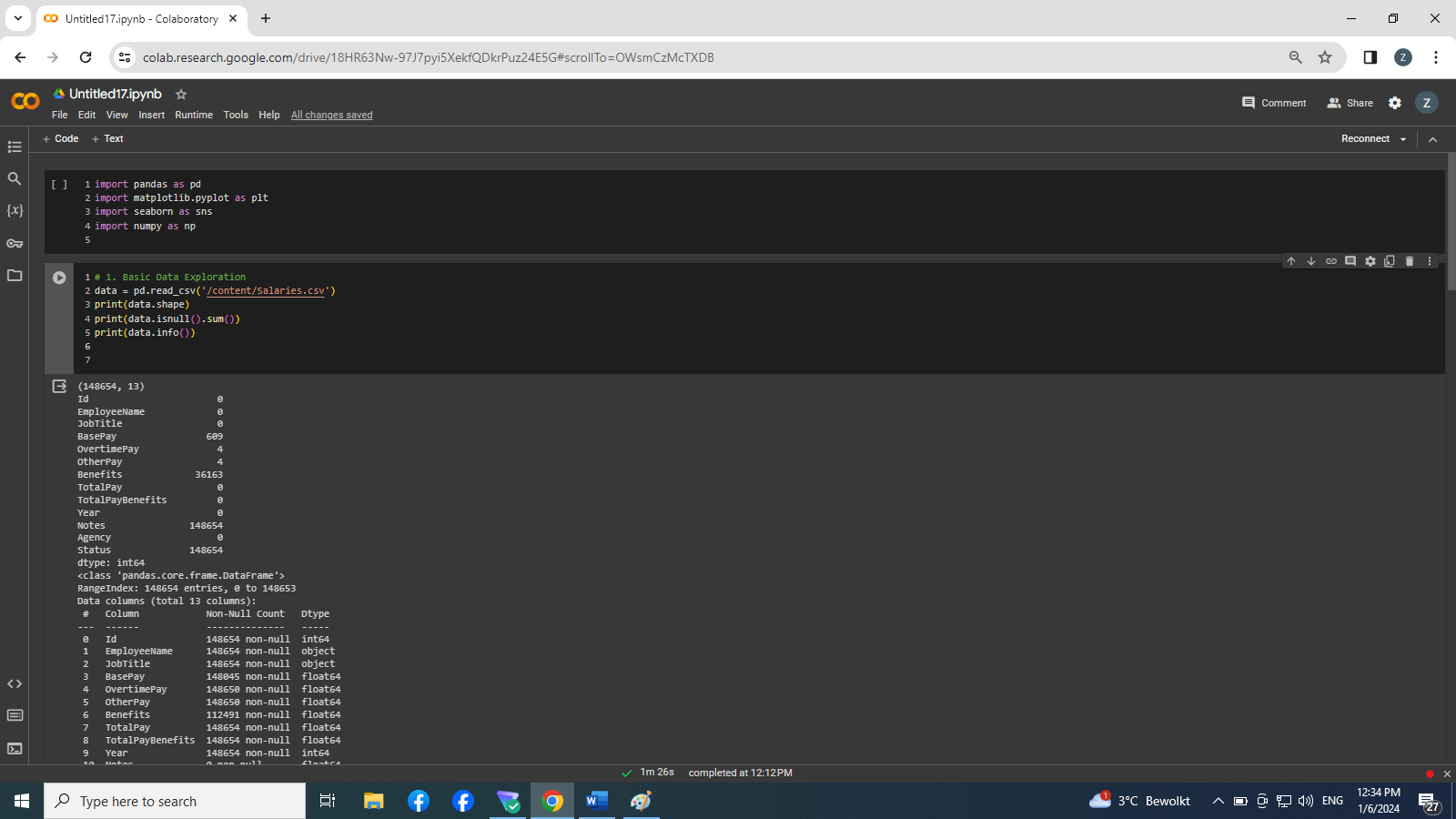
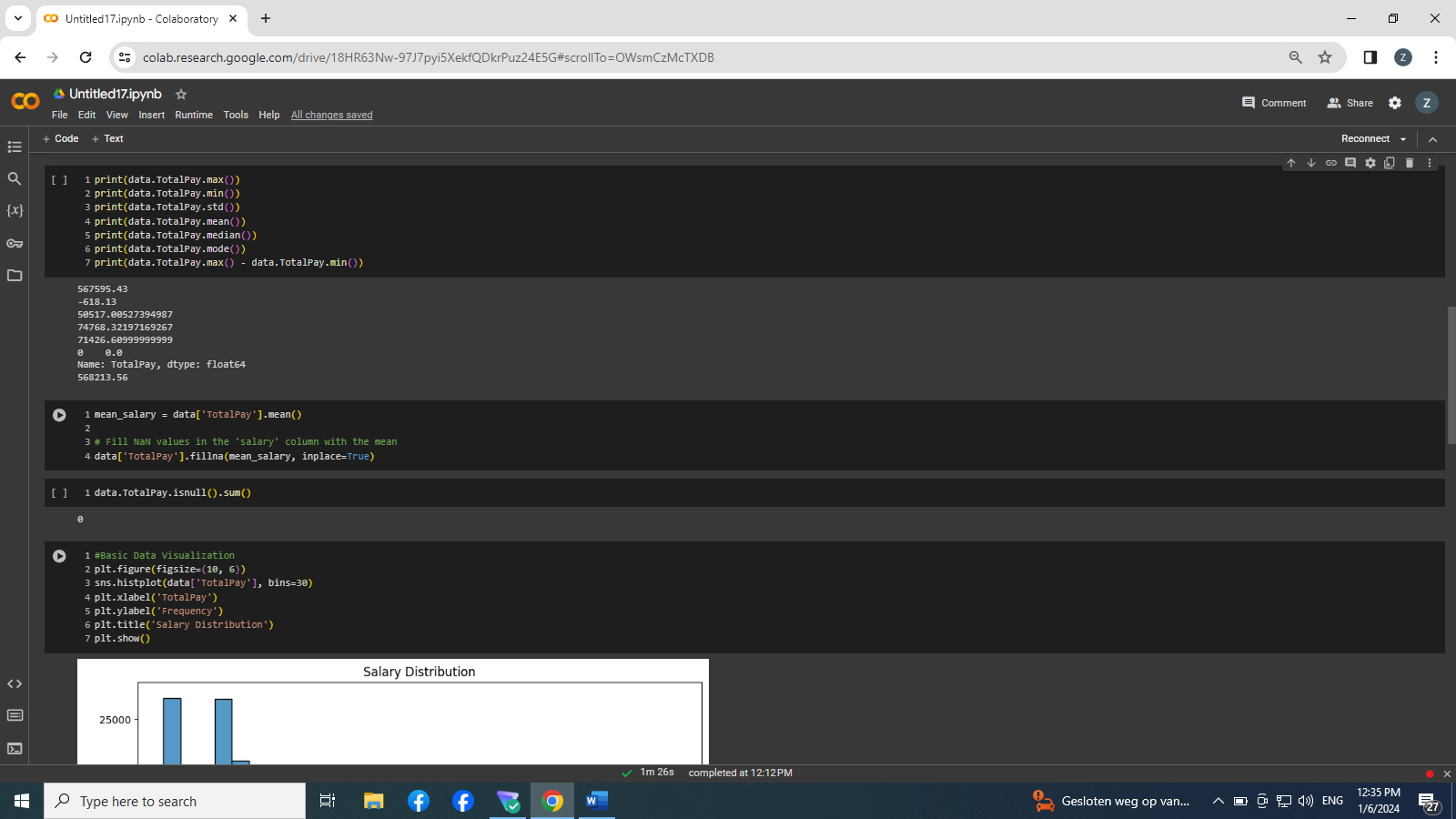
SHAII ASSAIGNEMENT

NAME : ZAIN EBRAHEM



\_Firstly:

We include the library , for example: numpy ,matplotlib , panadas ,seaborn

secondly :

We work in the Google Colab environment. We read the Sarary.csv file by use the function read\_csv() in pandas library .

We used the attribute shape to identify the number of columns and the number of lines(row = 148654 , columns = 13 ).

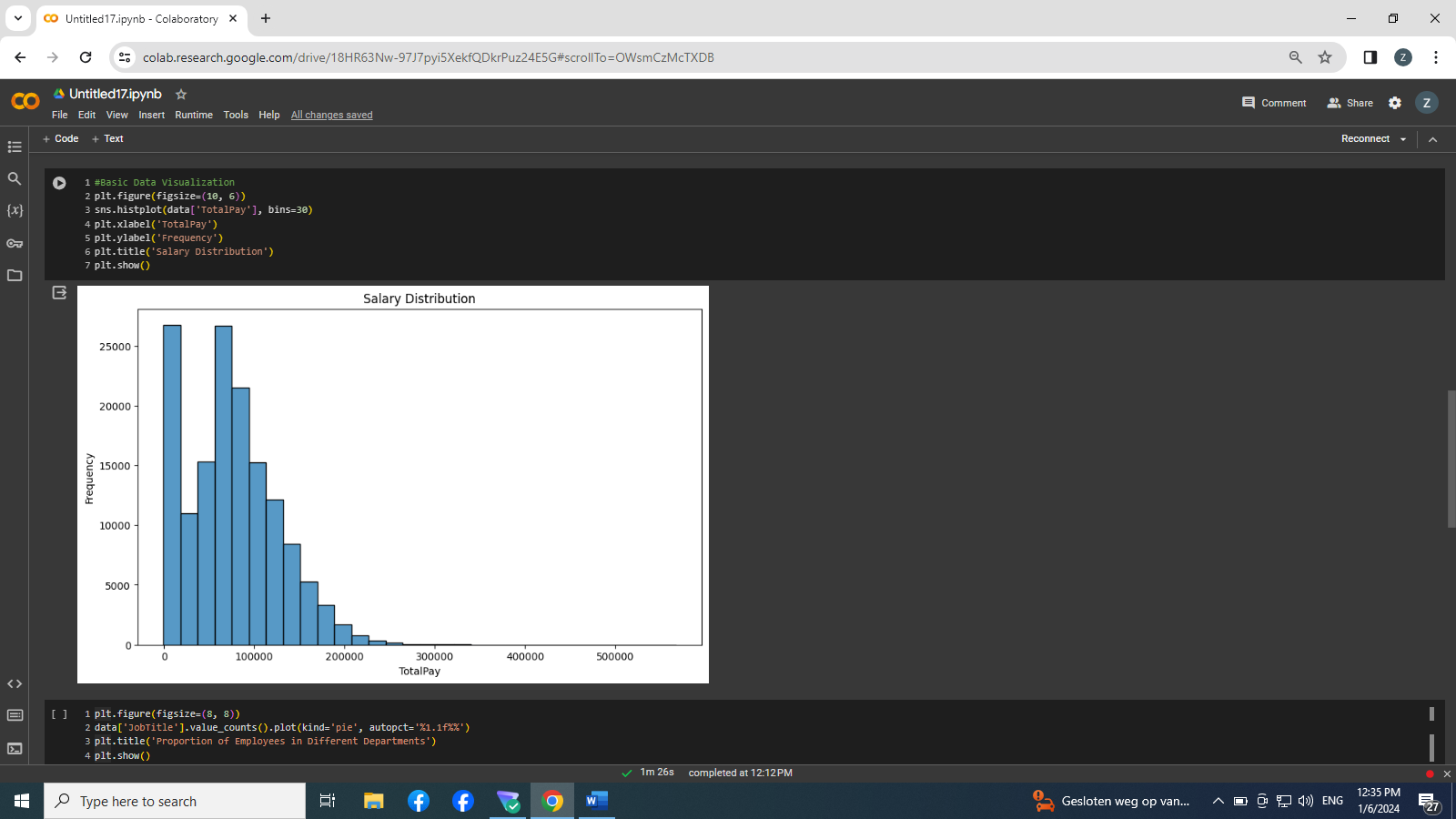
We used the function isnull() to examine the NAN VALUES and we use the function sum() to collect the NAN VALULES for each column in the dataset. Then we used the info() function to give us the type for each column. In the dataset (Id = int64 ,EmployeeName = object , JobTitle = object , BasePay = float64 ,Overtimepay = float64 , OtherPay = float64 , Benefits = float64 , TotalPay = float64 , TotalPayBenefits = float64 , Year = int64 , Notes = float64 ,Agency = object , status = float64)

Thirdly, we use the function max() to get the max value in the columns TotalPay , and the function min() to get the min value , the function std to get the std value , the function mode to get the value that the most frequency in the TotalPay columns and we have the range of the values in the TotalPay by minus the max value from the min value

Fourthly :

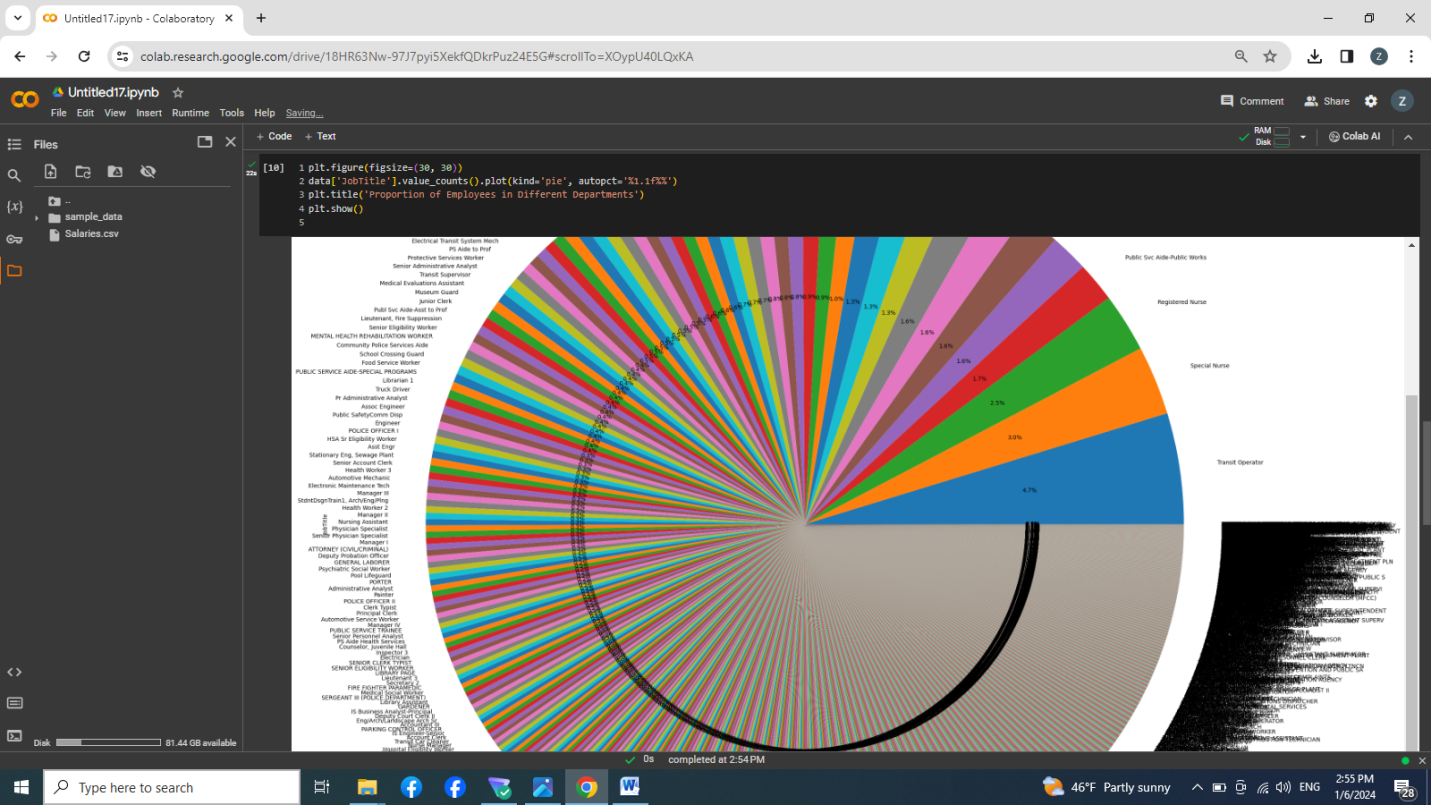
we used the Fill equation to fill the unknown values with the mean salaries, so that the unknown values are filled with the average because the employees whose salaries are unknown belong to different job grades, so it is better to use the average.

Then we checked the number of unknown values to confirm



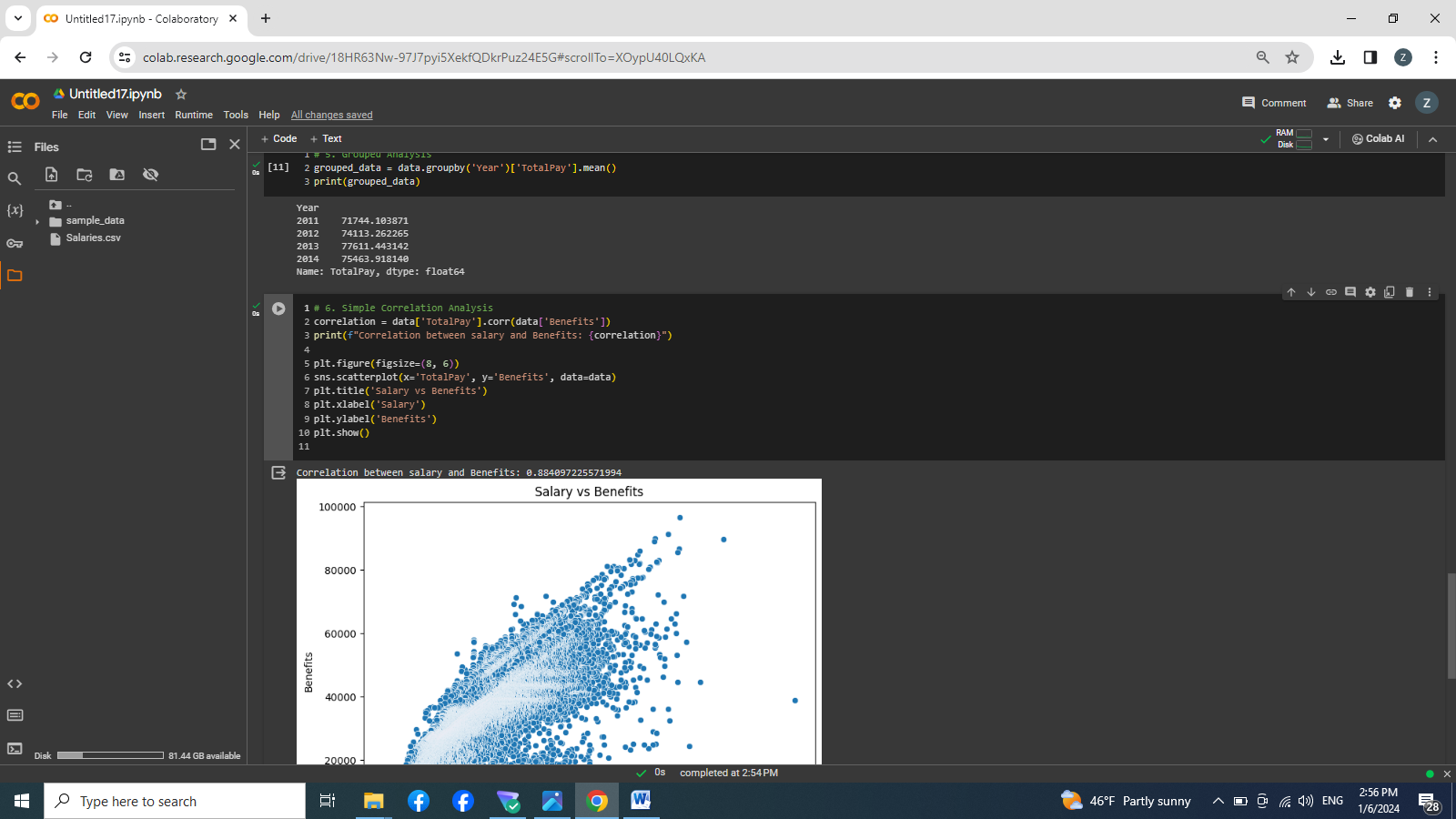
Fifth:

We have plot histogram to show the frequency of different values in the Employee Salaries column (TotalPay)

Sixthly :

I assumed that the Department column was the same as the JobTitle column because I did not find Department column

We drew a pie chart to show the percentage of different values in the Department column



Seventh:

We grouped salaries by year column and found the average salaries of employees in the following years (2011,2012 , 2013 , 2014 )  
we find the max mean average that to the employee in the 2013

And finally:

We have plot the relationship between the Benefits column and the Salaray column