Summary

In this task I

- calculate the number of rows and columns
- determine the data types of each column
- Calculate the number of missing values in each column
- calculate basic statistics mean, median, mode, minimum, and maximum salary, determine the range of salaries, and find the standard deviation
- Handle missing data by suitable method
- Create histograms to visualize the distribution of salaries,
- use pie charts to represent the proportion of employees in different departments.
- Group the data by Agency and year and I calculate summary statistics for each group
- Group the data by Year and I calculate summary statistics for each group
- compare the average salaries across different groups.
- Identify any correlation between salary and another numerical column,
- plot a scatter plot to visualize the relationship.

number of rows and columns

number of rows	148654
number of columns	13

the data types of each column

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

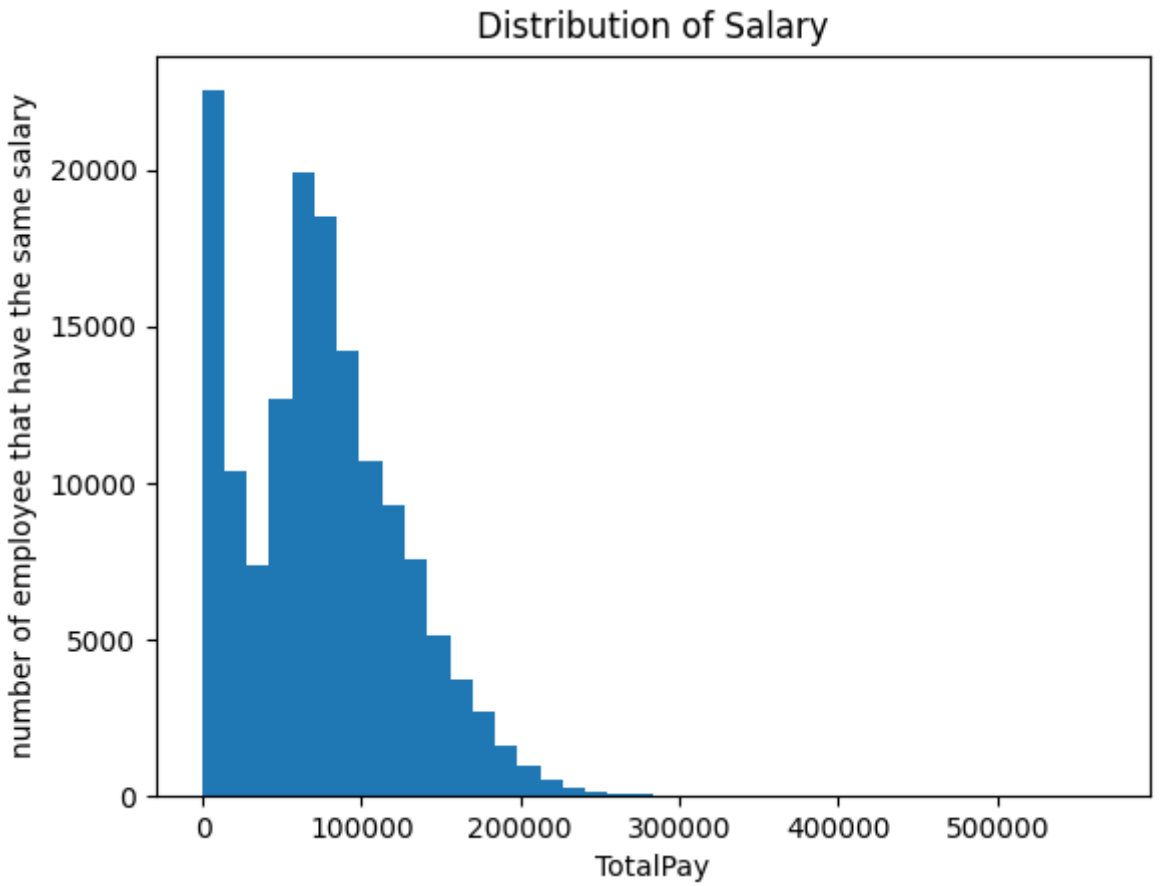
the number of missing values in each column

Id	0
EmployeeName	0
JobTitle	0
BasePay	609
OvertimePay	4
OtherPay	4
Benefits	36163
TotalPay	0
TotalPayBenefits	0
Year	0
Notes	148654
Agency	0
Status	148654

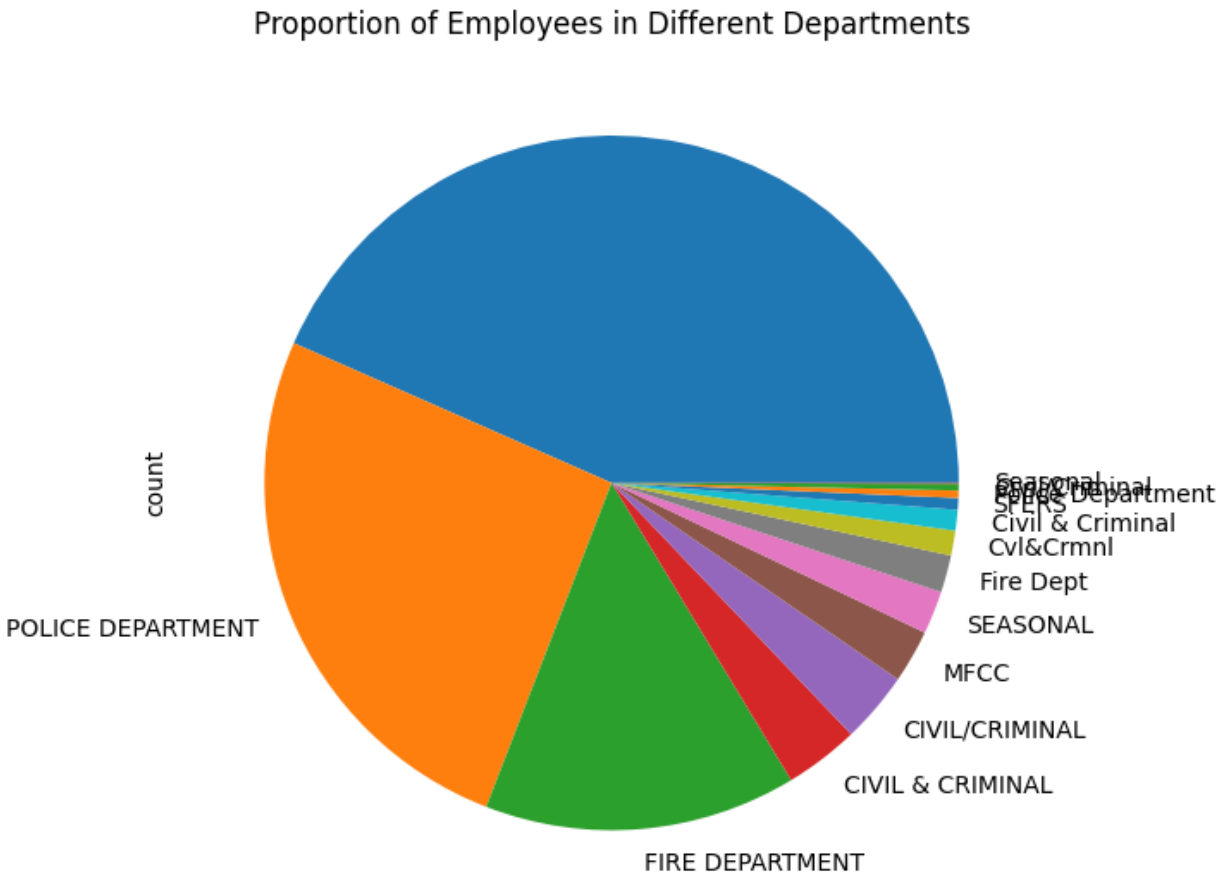
basic statistics

Max	**567595.43**
Min	-618.13
Mean	74768.32197169267
Median	71426.60999999999
Mode	0.0
Range	568213.56
standard deviation	50517.00527394987

histograms of distribution of salaries,



pie charts of the proportion of employees in different departments.



summary statistics for each group by agency and year

```
mean of salary in group by year and agency Year  Agency
2011  San Francisco    71744.103871
2012  San Francisco    74113.262265
2013  San Francisco    77611.443142
2014  San Francisco    75463.918140
```

```
max of salary in group by year and agency Year  Agency
2011  San Francisco    567595.43
2012  San Francisco    362844.66
2013  San Francisco    347102.32
2014  San Francisco    471952.64
```

```
min of salary in group by year and agency Year  Agency
2011  San Francisco      0.00
2012  San Francisco      0.00
2013  San Francisco      0.00
2014  San Francisco   -618.13
```

```
median of salary in group by year and agency Year  Agency
2011  San Francisco    68213.240
2012  San Francisco    70714.680
2013  San Francisco    74500.015
2014  San Francisco    72359.980
```

```
mode of salary in group by year and agency Year
2011    123290.7
2012     17085.0
2013         0.0
2014         0.0
```

summary statistics for each group by year

```
mean of salary in group by year  Year
2011    71744.103871
2012    74113.262265
2013    77611.443142
2014    75463.918140
```

```
mode of salary in group by year  Year
2011    123290.7
2012     17085.0
2013         0.0
2014         0.0
```

```
max of salary in group by year  Year
2011    567595.43
2012    362844.66
2013    347102.32
2014    471952.64
```

```
standard deviation of salary in group by year  Year
2011    47498.706594
2012    49523.904319
2013    52854.774783
2014    51697.713029
```

```
min of salary in group by year  Year
2011      0.00
2012      0.00
2013      0.00
2014   -618.13
```

Mean of sal group by year and agency

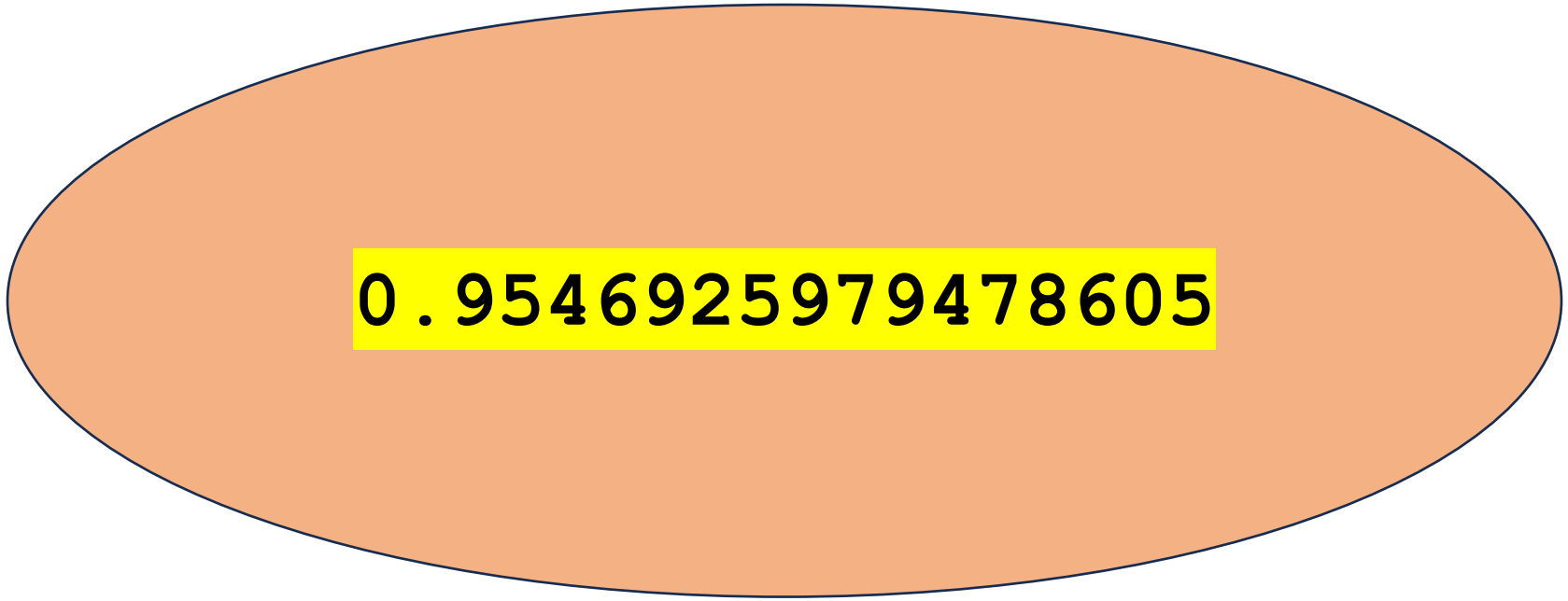
```
mean of salary in group by agenct and year Year  Agency
2011  San Francisco    71744.103871
2012  San Francisco    74113.262265
2013  San Francisco    77611.443142
2014  San Francisco    75463.918140
```

Mean of sal group by year

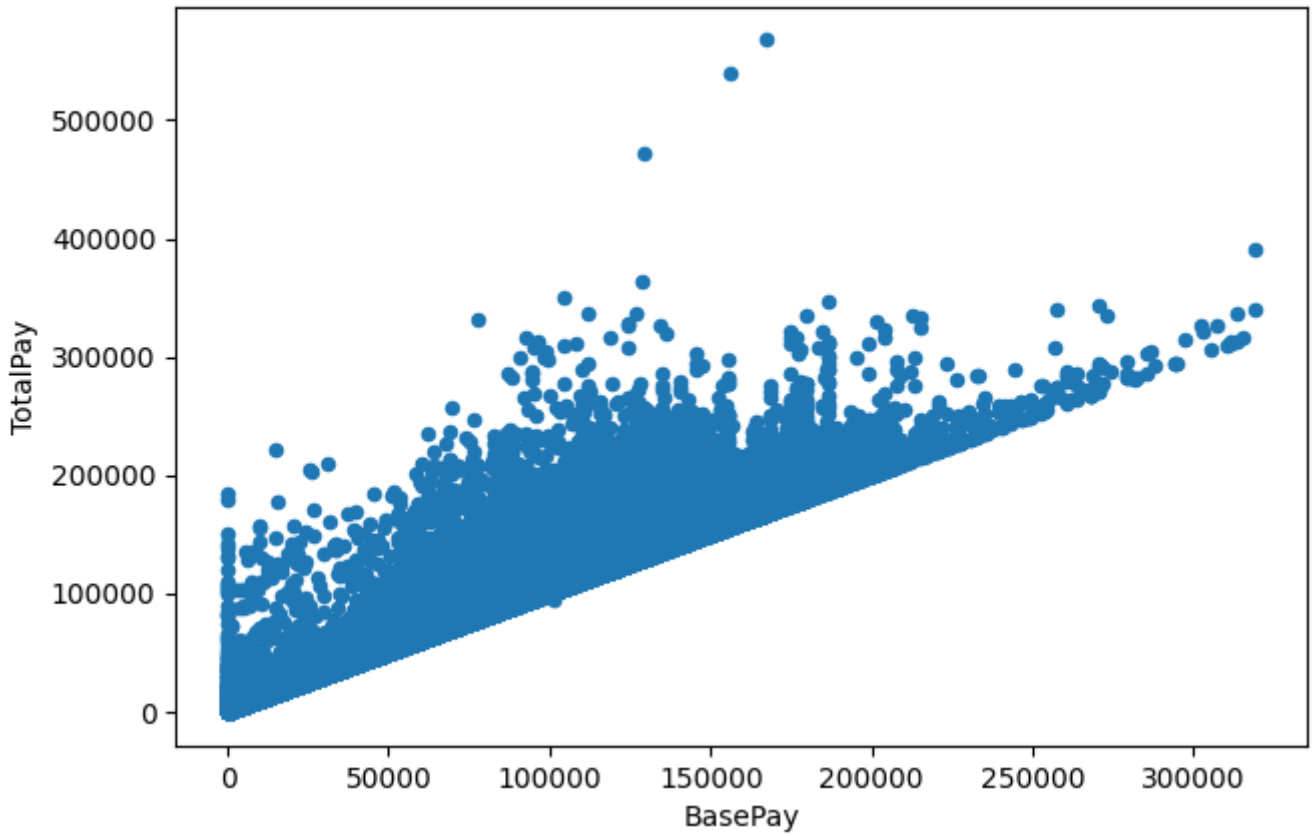
```
mean of salary in group by year Year
2011    71744.103871
2012    74113.262265
2013    77611.443142
2014    75463.918140
```

The same

correlation between salary and BasePay



scatter plot of the relationship



Done by Obieda Khalil