Salaries

August 17, 2021

[1]:
```
# importing pandas

import pandas as pd
```

[2]:
```
# to read csv

sal = pd.read_csv('salaries.csv')
```

[3]:
```
sal
```

[3]:
```
             Id       EmployeeName  \
0             1     NATHANIEL FORD
1             2       GARY JIMENEZ
2             3     ALBERT PARDINI
3             4  CHRISTOPHER CHONG
4             5    PATRICK GARDNER
…          …                 …
148649   148650      Roy I Tillery
148650   148651       Not provided
148651   148652       Not provided
148652   148653       Not provided
148653   148654          Joe Lopez

                                              JobTitle    BasePay  \
0       GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY  167411.18
1                    CAPTAIN III (POLICE DEPARTMENT)  155966.02
2                    CAPTAIN III (POLICE DEPARTMENT)  212739.13
3                WIRE ROPE CABLE MAINTENANCE MECHANIC   77916.00
4         DEPUTY CHIEF OF DEPARTMENT,(FIRE DEPARTMENT)  134401.60
…                                                  …        …
148649                                     Custodian       0.00
148650                                Not provided        NaN
148651                                Not provided        NaN
148652                                Not provided        NaN
148653                   Counselor, Log Cabin Ranch       0.00

        OvertimePay   OtherPay  Benefits   TotalPay  TotalPayBenefits  Year  \
0              0.00  400184.25       NaN  567595.43         567595.43  2011
```

```
1         245131.88  137811.38       NaN  538909.28         538909.28  2011
2         106088.18   16452.60       NaN  335279.91         335279.91  2011
3          56120.71  198306.90       NaN  332343.61         332343.61  2011
4           9737.00  182234.59       NaN  326373.19         326373.19  2011
…               …        …         …          …                 …    …
148649         0.00       0.00       0.0       0.00              0.00  2014
148650          NaN        NaN       NaN       0.00              0.00  2014
148651          NaN        NaN       NaN       0.00              0.00  2014
148652          NaN        NaN       NaN       0.00              0.00  2014
148653         0.00    -618.13       0.0    -618.13           -618.13  2014

        Notes         Agency  Status
0         NaN  San Francisco     NaN
1         NaN  San Francisco     NaN
2         NaN  San Francisco     NaN
3         NaN  San Francisco     NaN
4         NaN  San Francisco     NaN
…       …             …      …
148649    NaN  San Francisco     NaN
148650    NaN  San Francisco     NaN
148651    NaN  San Francisco     NaN
148652    NaN  San Francisco     NaN
148653    NaN  San Francisco     NaN

[148654 rows x 13 columns]
```

[4]:
```
# cheack the head dataframe

sal.head()
```

[4]:
```
   Id       EmployeeName                                      JobTitle  \
0   1     NATHANIEL FORD  GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY
1   2       GARY JIMENEZ                 CAPTAIN III (POLICE DEPARTMENT)
2   3     ALBERT PARDINI                 CAPTAIN III (POLICE DEPARTMENT)
3   4  CHRISTOPHER CHONG            WIRE ROPE CABLE MAINTENANCE MECHANIC
4   5    PATRICK GARDNER    DEPUTY CHIEF OF DEPARTMENT,(FIRE DEPARTMENT)

     BasePay  OvertimePay   OtherPay  Benefits   TotalPay  TotalPayBenefits  \
0  167411.18         0.00  400184.25       NaN  567595.43         567595.43
1  155966.02    245131.88  137811.38       NaN  538909.28         538909.28
2  212739.13    106088.18   16452.60       NaN  335279.91         335279.91
3   77916.00     56120.71  198306.90       NaN  332343.61         332343.61
4  134401.60      9737.00  182234.59       NaN  326373.19         326373.19

   Year  Notes         Agency  Status
0  2011    NaN  San Francisco     NaN
1  2011    NaN  San Francisco     NaN
```

```
 2  2011    NaN  San Francisco     NaN
 3  2011    NaN  San Francisco     NaN
 4  2011    NaN  San Francisco     NaN
```

[5]:
```
# find out how many entries are there

sal.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148654 entries, 0 to 148653
Data columns (total 13 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   Id                148654 non-null  int64
 1   EmployeeName      148654 non-null  object
 2   JobTitle          148654 non-null  object
 3   BasePay           148045 non-null  float64
 4   OvertimePay       148650 non-null  float64
 5   OtherPay          148650 non-null  float64
 6   Benefits          112491 non-null  float64
 7   TotalPay          148654 non-null  float64
 8   TotalPayBenefits  148654 non-null  float64
 9   Year              148654 non-null  int64
 10  Notes             0 non-null       float64
 11  Agency            148654 non-null  object
 12  Status            0 non-null       float64
dtypes: float64(8), int64(2), object(3)
memory usage: 14.7+ MB
```

[7]:
```
# what is average basepay

sal['BasePay'].mean()
```

[7]:
```
66325.44884050643
```

[10]:
```
# what is higest amount of overtime in database

sal['OvertimePay'].max()
```

[10]:
```
245131.88
```

[21]:
```
# what is the job title of  JOSEPH DRISCOLL

sal[sal['EmployeeName'] == 'JOSEPH DRISCOLL']
```

[21]:
```
     Id     EmployeeName                  JobTitle    BasePay  OvertimePay  \
24   25  JOSEPH DRISCOLL  CAPTAIN, FIRE SUPPRESSION  140546.86     97868.77
```

```
    OtherPay  Benefits   TotalPay  TotalPayBenefits  Year  Notes  \
24  31909.28       NaN  270324.91         270324.91  2011    NaN

           Agency  Status
24  San Francisco     NaN
```

[22]:
```
sal[sal['EmployeeName'] == 'JOSEPH DRISCOLL']['JobTitle']
```

[22]:
```
24    CAPTAIN, FIRE SUPPRESSION
Name: JobTitle, dtype: object
```

[23]:
```
# how much dose JOSEPH DRISCOLL make (including benefits)

sal[sal['EmployeeName'] == 'JOSEPH DRISCOLL']['TotalPayBenefits']
```

[23]:
```
24    270324.91
Name: TotalPayBenefits, dtype: float64
```

[27]:
```
# What is the name of highest paid person (including benefits)

sal['TotalPayBenefits'].max()
```

[27]:
```
567595.43
```

[30]:
```
sal[sal['TotalPayBenefits'] == sal['TotalPayBenefits'].max()]
```

[30]:
```
   Id    EmployeeName                                      JobTitle  \
0   1  NATHANIEL FORD  GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY

     BasePay  OvertimePay   OtherPay  Benefits   TotalPay  TotalPayBenefits  \
0  167411.18          0.0  400184.25       NaN  567595.43         567595.43

   Year  Notes         Agency  Status
0  2011    NaN  San Francisco     NaN
```

[31]:
```
sal[sal['TotalPayBenefits'] == sal['TotalPayBenefits'].max()]['EmployeeName']
```

[31]:
```
0    NATHANIEL FORD
Name: EmployeeName, dtype: object
```

[32]:
```
# What is the name of lowest paid person (including benefits)


sal['TotalPayBenefits'].min()
```

[32]:
```
-618.13
```

[33]:
```
sal['TotalPayBenefits'] == sal['TotalPayBenefits'].min()
```

[33]:
```
0         False
1         False
2         False
3         False
4         False
          …
148649    False
148650    False
148651    False
148652    False
148653     True
Name: TotalPayBenefits, Length: 148654, dtype: bool
```

[36]:
```
sal[sal['TotalPayBenefits'] == sal['TotalPayBenefits'].min()]
```

[36]:
```
            Id EmployeeName                    JobTitle  BasePay  OvertimePay  \
148653  148654    Joe Lopez  Counselor, Log Cabin Ranch      0.0          0.0

        OtherPay  Benefits  TotalPay  TotalPayBenefits  Year  Notes  \
148653   -618.13       0.0   -618.13           -618.13  2014    NaN

               Agency  Status
148653  San Francisco     NaN
```

[37]:
```
sal[sal['TotalPayBenefits'] == sal['TotalPayBenefits'].min()]['EmployeeName']
```

[37]:
```
148653    Joe Lopez
Name: EmployeeName, dtype: object
```

[39]:
```
# What was the average (mean) BasePay of all employees per year? (2011-2014)

sal.groupby('Year').mean()
```

[39]:
```
             Id       BasePay  OvertimePay     OtherPay      Benefits  \
Year
2011   18080.0  63595.956517  4531.065429  3617.081926           NaN
2012   54542.5  65436.406857  5023.417824  3653.437583  26439.966967
2013   91728.5  69630.030216  5281.641980  3819.969007  23829.076572
2014  129593.0  66564.421924  5401.993737  3505.421251  24789.601756

          TotalPay  TotalPayBenefits  Notes  Status
Year
2011  71744.103871      71744.103871    NaN     NaN
2012  74113.262265     100553.229232    NaN     NaN
2013  77611.443142     101440.519714    NaN     NaN
2014  75463.918140     100250.918884    NaN     NaN
```

[40]:
```
sal.groupby('Year').mean()['BasePay']
```

[40]:
```
Year
2011    63595.956517
2012    65436.406857
2013    69630.030216
2014    66564.421924
Name: BasePay, dtype: float64
```

[48]:
```
#  How many unique job titles are there

sal['JobTitle'].nunique()
```

[48]:
```
2159
```

[52]:
```
# What are the top 5 most common jobs

sal['JobTitle'].value_counts()
```

[52]:
```
Transit Operator                  7036
Special Nurse                     4389
Registered Nurse                  3736
Public Svc Aide-Public Works      2518
Police Officer 3                  2421
                                  …
PRINCIPAL RECREATION SUPERVISOR      1
ORTHOPEDIC TECHNICIAN II             1
IS TECHNICIAN - ASSISTANT            1
DIRECTOR, JUVENILE HALL              1
Public Service Aide-Technical        1
Name: JobTitle, Length: 2159, dtype: int64
```

[53]:
```
sal['JobTitle'].value_counts().head()
```

[53]:
```
Transit Operator                7036
Special Nurse                   4389
Registered Nurse                3736
Public Svc Aide-Public Works    2518
Police Officer 3                2421
Name: JobTitle, dtype: int64
```

[59]:
```
# How many Job Titles were represented by only one person in 2013?

sum(sal[sal['Year'] == 2013]['JobTitle'].value_counts() == 1 )
```

[59]:
```
202
```

[ ]:
```

```