01-SF Salaries Exercise

October 25, 2019

————

1 SF Salaries Exercise

Welcome to a quick exercise for you to practice your pandas skills! We will be using the SF Salaries Dataset from Kaggle! Just follow along and complete the tasks outlined in bold below. The tasks will get harder and harder as you go along.

** Import pandas as pd.**

```
In [1]: import pandas as pd
```

** Read Salaries.csv as a dataframe called sal.**

```
In [3]: sal = pd.read_csv('Salaries.csv')
```

** Check the head of the DataFrame. **

```
In [5]: sal.head()
```

```
Out[5]:    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
```

** Use the .info() method to find out how many entries there are.**

```
In [6]: sal.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148654 entries, 0 to 148653
Data columns (total 13 columns):
Id                 148654 non-null int64
EmployeeName       148654 non-null object
JobTitle           148654 non-null object
BasePay            148045 non-null float64
OvertimePay        148650 non-null float64
OtherPay           148650 non-null float64
Benefits           112491 non-null float64
TotalPay           148654 non-null float64
TotalPayBenefits   148654 non-null float64
Year               148654 non-null int64
Notes              0 non-null float64
Agency             148654 non-null object
Status             0 non-null float64
dtypes: float64(8), int64(2), object(3)
memory usage: 14.7+ MB
```

What is the average BasePay ?

```
In [13]: sal['BasePay'].mean()

Out[13]: 66325.44884050643
```

** What is the highest amount of OvertimePay in the dataset ? **

```
In [15]: sal['OvertimePay'].max()

Out[15]: 245131.88
```

** What is the job title of JOSEPH DRISCOLL ? Note: Use all caps, otherwise you may get an answer that doesn't match up (there is also a lowercase Joseph Driscoll). **

```
In [37]: sal[sal['EmployeeName'] == 'JOSEPH DRISCOLL']['JobTitle']

Out[37]: 24    CAPTAIN, FIRE SUPPRESSION
         Name: JobTitle, dtype: object
```

** How much does JOSEPH DRISCOLL make (including benefits)? **

```
In [38]: sal[sal['EmployeeName'] == 'JOSEPH DRISCOLL']['TotalPayBenefits']

Out[38]: 24    270324.91
         Name: TotalPayBenefits, dtype: float64
```

** What is the name of highest paid person (including benefits)?**

```
In [90]: sal[sal['TotalPayBenefits']==sal["TotalPayBenefits"].max()]['EmployeeName']

Out[90]: 0    NATHANIEL FORD
         Name: EmployeeName, dtype: object
```

** What is the name of lowest paid person (including benefits)? Do you notice something strange about how much he or she is paid?**

```
In [91]: sal[sal["TotalPayBenefits"]==sal['TotalPayBenefits'].min()]['EmployeeName']

Out[91]: 148653    Joe Lopez
         Name: EmployeeName, dtype: object
```

** What was the average (mean) BasePay of all employees per year? (2011-2014) ? **

```
In [76]: sal.groupby('Year')["BasePay"].mean()

Out[76]: Year
         2011    63595.956517
         2012    65436.406857
         2013    69630.030216
         2014    66564.421924
         Name: BasePay, dtype: float64
```

** How many unique job titles are there? **

```
In [77]: sal['JobTitle'].nunique()

Out[77]: 2159
```

** What are the top 5 most common jobs? **

```
In [92]: sal['JobTitle'].value_counts().head(5)

Out[92]: Transit Operator           7036
         Special Nurse              4389
         Registered Nurse           3736
         Public Svc Aide-Public Works  2518
         Police Officer 3           2421
         Name: JobTitle, dtype: int64
```

** How many Job Titles were represented by only one person in 2013? (e.g. Job Titles with only one occurence in 2013?) **