SF Salaries Exercise

Welcome to a quick exercise for you to practice your pandas skills! We will be using the SF Salaries Dataset (https://www.kaggle.com/kaggle/sf-salaries) 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 [2]:

```
import pandas as pd
```

Read Salaries.csv as a dataframe called sal.

In [7]:

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

Check the head of the DataFrame.

In [11]:

```
sal.head()
```

Out[11]:

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalP
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.00	400184.25	NaN	567595
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	538909
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN	335279
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.00	56120.71	198306.90	NaN	332343
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)	134401.60	9737.00	182234.59	NaN	326373

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

In [12]:

```
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: 13.0+ MB
```

What is the average BasePay ?

In [13]:

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

Out[13]:

```
66325.4488404877
```

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 [16]:

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

Out[16]:

```
24    CAPTAIN, FIRE SUPPRESSION
Name: JobTitle, dtype: object
```

How much does JOSEPH DRISCOLL make (including benefits)?

In [18]:

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

Out[18]:

```
24    270324.91
Name: TotalPayBenefits, dtype: float64
```

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

In [25]:

```
sal.loc[sal['TotalPayBenefits'].idxmax()]
```

Out[25]:

```
Id                                                          1
EmployeeName                                      NATHANIEL FORD
JobTitle            GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY
BasePay                                                167411
OvertimePay                                                 0
OtherPay                                               400184
Benefits                                                  NaN
TotalPay                                               567595
TotalPayBenefits                                       567595
Year                                                     2011
Notes                                                     NaN
Agency                                          San Francisco
Status                                                    NaN
Name: 0, 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 [32]:

```
sal.loc[sal['TotalPayBenefits'].idxmin()]
```

Out[32]:

```
Id                                  148654
EmployeeName                     Joe Lopez
JobTitle            Counselor, Log Cabin Ranch
BasePay                                  0
OvertimePay                              0
OtherPay                            -618.13
Benefits                                 0
TotalPay                            -618.13
TotalPayBenefits                    -618.13
Year                                   2014
Notes                                   NaN
Agency                        San Francisco
Status                                  NaN
Name: 148653, dtype: object
```

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

In [36]:

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

Out[36]:

```
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 [37]:

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

Out[37]:

```
2159
```

What are the top 5 most common jobs?

In [39]:

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

Out[39]:

```
202
```

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

In [44]:

```
def chief_string(title):
    if 'chief' in title.lower():
        return True
    else:
        return False
```

How many people have the word Chief in their job title? (This is pretty tricky)

In [47]:

```
sum(sal['JobTitle'].apply(lambda x: chief_string(x)))
```

Out[47]:

```
627
```

Great Job!