Data Analysis Tools with Pandas - SF Salaries Exercise

แบบฝึกหัดนี้เป็นแบบฝึกหัดทดสอบทักษะการใช้งาน library pandas ด้วย SF Salaries Dataset จาก เว็ปไซต์ Kaggle ให้ทำตามคำสั่ง ต่อไปนี้

Import pandas as pd.

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

ให้นำเข้าข้อมูลจากไฟล์ Salaries.csv มาในรูปของ dataframe โดยตั้งชื่อตัวแปรว่า sal

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

Check the head of the DataFrame.

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

Out[ ]:

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	Tota
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.00	400184.25	NaN	56759
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	53890
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN	33527
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.00	56120.71	198306.90	NaN	33234
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)	134401.60	9737.00	182234.59	NaN	32637

ใช้คำสั่ง .info() method to ในการดูภาพรวมของข้อมูลทั้งหมด

```
In [ ]:  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
```

ให้หาค่า average ของ BasePay ?

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

```
Out[ ]:  66325.4488404877
```

OvertimePay สูงที่สุด ใน dataset เท่ากับเท่าไหร่?

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

```
Out[ ]:  245131.88
```

JOSEPH DRISCOLL ทำงานอะไร (jobTitle)?

Note: Use all caps, otherwise you may get an answer that doesn't match up (there is also a lowercase Joseph Driscoll).

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

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

JOSEPH DRISCOLL ได้เงินไปทั้งหมดเท่าไหร่ (รวมทั้ง benefits)?

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

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

ใครคือคนที่ได้รับเงินมากที่สุด (รวมทั้ง benefits)?

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

Out[ ]:

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	Tota
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.0	400184.25	NaN	56759

◄ ▬▬▬▬▬▬▬▬▬▬▬▬ ►

ใครคือคนที่ได้รับเงินน้อยที่สุด (รวมทั้ง benefits)?

Do you notice something strange about how much he or she is paid?

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

Out[ ]:

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	To
148653	148654	Joe Lopez	Counselor, Log Cabin Ranch	0.0	0.0	-618.13	0.0	

◄ ▬▬▬▬▬▬▬▬▬▬▬▬ ►

จงหาค่า average (mean) ของ BasePay ของ employees ทั้งหมดในแต่ละปี (2011-2014)

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

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

มีชื่อตำแหน่งงานต่างๆ (unique job) อยู่กี่ชื่อ?

```
In [ ]:  len(sal['JobTitle'].unique())
```

```
Out[ ]:  2159
```

top 5 ตำแหน่งเป็นที่ต้องการในที่ต่างๆ มีอะไรบ้าง ?

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

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

มีจำนวนกี่ตำแหน่งที่ต้องการเพียง 2 คน ในปี 2013? (e.g. Job Titles with only one occurence in 2013?)

```
In [ ]:  df1 = sal[sal['Year'] == 2013]['JobTitle'].value_counts()
```

```
In [ ]:  df2 = sal[sal['Year'] == 2013]['JobTitle'].value_counts() == 2
```

```
In [ ]:  (sal[sal['Year'] == 2013]['JobTitle'].value_counts() == 2).sum()
```

```
Out[ ]:  69
```

```
In [ ]:  df1[df2].count()
         #ตอบ 69
```

```
Out[ ]:  69
```

มีคนกี่คนที่มีคำว่า Chief อยู่ในชื่อตำแหน่ง job title ของเค้า (This is pretty tricky)

```
In [ ]:  ch = lambda s : "chief" in s.lower()
```

```
In [ ]:  sal['JobTitle'].apply(ch).sum()
```

```
Out[ ]:  627
```

```
In [ ]:  (sal[sal['JobTitle'].apply(ch)])
```

Out[ ]:

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benef
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)	134401.60	9737.00	182234.59	Na
5	6	DAVID SULLIVAN	ASSISTANT DEPUTY CHIEF II	118602.00	8601.00	189082.74	Na
6	7	ALSON LEE	BATTALION CHIEF, (FIRE DEPARTMENT)	92492.01	89062.90	134426.14	Na
8	9	MICHAEL MORRIS	BATTALION CHIEF, (FIRE DEPARTMENT)	176932.64	86362.68	40132.23	Na
9	10	JOANNE HAYES-WHITE	CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)	285262.00	0.00	17115.73	Na
...	...	...	...	...	...	...	
138011	138012	Sharon C Jackson	Chief Deputy Adlt Probation Of	56660.18	0.00	0.00	566.
139620	139621	Anita S Lee	Chief Clerk	38598.86	0.00	0.00	386.
139709	139710	George D Vaughan	Chief Nursery Specialist	27319.47	483.90	0.00	10334.
140080	140081	Patrick W Gillespie	Chief Preparator	24091.00	0.00	0.00	10176.
140156	140157	Christopher L Ochoa	Chief Stationary Engineer	24392.00	418.71	0.00	8918.

627 rows × 13 columns

In [ ]:
```python
print (sal[sal['JobTitle'].apply(ch)] ['JobTitle'].to_string)
```

```
<bound method Series.to_string of 4        DEPUTY CHIEF OF DEPARTMENT,(FIRE DEPARTM
ENT)
5                            ASSISTANT DEPUTY CHIEF II
6              BATTALION CHIEF, (FIRE DEPARTMENT)
8              BATTALION CHIEF, (FIRE DEPARTMENT)
9          CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)
                           ...
138011            Chief Deputy Adlt Probation Of
139620                                Chief Clerk
139709                    Chief Nursery Specialist
140080                            Chief Preparator
140156                  Chief Stationary Engineer
Name: JobTitle, Length: 627, dtype: object>
```

----- ภาวนามยปัญญา ปัญญาที่เกิดจากการลงมือทำ! -----