

Data Analysis Tools with Pandas - SF Salaries Exercise

แบบฝึกหัดนี้เป็นแบบฝึกหัดทดสอบทักษะการใช้งาน library pandas ด้วย [SF Salaries Dataset](https://www.kaggle.com/kaggle/sf-salaries) (<https://www.kaggle.com/kaggle/sf-salaries>) จากเว็บไซต์ Kaggle ให้ทำตามคำสั่ง ต่อไปนี้

Import pandas as pd.

```
In [8]: 1 import pandas as pd
```

ให้นำเข้าข้อมูลจากไฟล์ Salaries.csv มาในรูปของ dataframe โดยตั้งชื่อตัวแปรว่า sal

```
In [15]: 1 sal = pd.read_csv("Salaries.csv")
```

Check the head of the DataFrame.

```
In [16]: 1 sal.head()
```

Out[16]:

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.00	400184.25	NaN	567595.43
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	538909.28
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN	335279.91
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.00	56120.71	198306.90	NaN	332343.61
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)	134401.60	9737.00	182234.59	NaN	326373.19

ใช้คำสั่ง .info() method to ในการดูภาพรวมของข้อมูลทั้งหมด

In [17]: 1 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 [19]: 1 sal["BasePay"].mean()

Out[19]: 66325.4488404877

OvertimePay สูงที่สุด ใน dataset เท่ากับเท่าไร?

In [21]: 1 sal["OvertimePay"].max()

Out[21]: 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 [48]: 1 sal[sal["EmployeeName"] == "JOSEPH DRISCOLL"]["JobTitle"]

Out[48]: 24 CAPTAIN, FIRE SUPPRESSION
Name: JobTitle, dtype: object

1 **JOSEPH DRISCOLL ได้เงินไปทั้งหมดเท่าไร (รวมทั้ง benefits)?**

In [45]: 1 sal[sal["EmployeeName"] == "JOSEPH DRISCOLL"]["TotalPayBenefits"]

Out[45]: 24 270324.91
Name: TotalPayBenefits, dtype: float64

ใครคือคนที่ได้รับเงินมากที่สุด (รวมทั้ง benefits)?

```
In [71]: 1 sal[sal["TotalPayBenefits"] == sal["TotalPayBenefits"].max()]
```

Out[71]:

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.0	400184.25	NaN	567595.43

ใครคือคนที่ได้รับเงินน้อยที่สุด (รวมทั้ง benefits)?

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

```
In [72]: 1 sal[sal["TotalPayBenefits"] == sal["TotalPayBenefits"].min()]
```

Out[72]:

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

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

```
In [79]: 1 sal.groupby("Year")["BasePay"].mean()
```

Out[79]: Year
 2011 63595.956517
 2012 65436.406857
 2013 69630.030216
 2014 66564.421924
 Name: BasePay, dtype: float64

มีชื่อตำแหน่งงานต่างๆ (unique job) อยู่กี่ชื่อ?

```
In [82]: 1 sal["JobTitle"].nunique()
```

Out[82]: 2159

top 5 ตำแหน่งเป็นที่ต้องการในหลายๆ มีอะไรบ้าง ?

```
In [84]: 1 sal["JobTitle"].value_counts().head()
```

```
Out[84]: JobTitle
Transit Operator          7036
Special Nurse             4389
Registered Nurse          3736
Public Svc Aide-Public Works 2518
Police Officer 3          2421
Name: count, dtype: int64
```

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

```
In [100]: 1 tmp1 = sal[sal["Year"] == 2013]["JobTitle"].value_counts()
          2 tmp2 = sal[sal["Year"] == 2013]["JobTitle"].value_counts() == 2
          3 tmp1[tmp2].count()
          4 #ตอบ 69
```

```
Out[100]: 69
```

```
In [101]: 1 (sal[sal["Year"] == 2013]["JobTitle"].value_counts() == 2).sum()
```

```
Out[101]: 69
```

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

```
In [103]: 1 fn = lambda s : "chief" in s.lower()
```

```
In [105]: 1 sal["JobTitle"].apply(fn).sum()
```

```
Out[105]: 627
```

```
In [20]:
```

```
Out[20]: 627
```

----- ภาพนามยปัญญา ปัญญาที่เกิดจากการลงมือทำ! -----