

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 [2]:

ให้นำเข้าข้อมูลจากไฟล์ Salaries.csv มาในรูปของ dataframe โดยตั้งชื่อตัวแปรว่า sal

In [3]:

Check the head of the DataFrame.

In [4]:

Out[4]:

	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 [5]:

```
<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
```

ให้หาค่า average ของ BasePay ?

In [6]:

Out[6]: 66325.44884050643

OvertimePay สูงที่สุด ใน dataset เท่ากับเท่าไร?

In [7]:

Out[7]: 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 [8]:

Out[8]: 24 CAPTAIN, FIRE SUPPRESSION
Name: JobTitle, dtype: object

JOSEPH DRISCOLL ได้เงินไปทั้งหมดเท่าไร (รวมทั้ง benefits)?

In [9]:

Out[9]: 24 270324.91
Name: TotalPayBenefits, dtype: float64

ใครคือคนที่ได้รับเงินมากที่สุด (รวมทั้ง benefits)?

In [10]:

Out[10]:

	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

ใครคือคนที่ได้รับเงินน้องที่สุด (รวมทั้ง benefits)?

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

In [15]:

Out[15]:

	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

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

In [12]:

Out[12]:

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

มีชื่อตำแหน่งงานต่างๆ (unique job) อยู่กี่ชื่อ?

In [17]:

Out[17]: 2159

top 5 ตำแหน่งเป็นที่ต้องการในที่ต่างๆ มีอะไรบ้าง ?

In [18]:

```
Out[18]: Transit Operator    7036
Special Nurse              4389
Registered Nurse           3736
Public Svc Aide-Public Works 2518
Police Officer 3           2421
Name: JobTitle, dtype: int64
```

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

In [1]:

#ตอบ 69

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

In [19]:

In [20]:

Out[20]: 627

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