Pandas Exercise

** Import pandas as pd.**

In [1]:

```
1 import pandas as pdP
```

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

In [10]:

```
1 sal= pd.read_csv("D:\A.S\Working\Material\Machinfy\Sessions\Session 9\Assignments\Copy
2 sal.head()
```

Out[10]:

	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

** Check the head of the DataFrame. **

In [11]:

```
sal= pd.read_csv("D:\A.S\Working\Material\Machinfy\Sessions\Session 9\Assignments\Copy
sal.tail()
```

Out[11]:

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalP
148649	148650	Roy I Tillery	Custodian	0.0	0.0	0.00	0.0	0.(
148650	148651	Not provided	Not provided	NaN	NaN	NaN	NaN	0.(
148651	148652	Not provided	Not provided	NaN	NaN	NaN	NaN	0.(
148652	148653	Not provided	Not provided	NaN	NaN	NaN	NaN	0.(
148653	148654	Joe Lopez	Counselor, Log Cabin Ranch	0.0	0.0	-618.13	0.0	-618.

** 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):
 #   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
```

What is the average BasePay ?

In [14]:

```
1 sal["BasePay"].mean()
```

Out[14]:

```
66325.44884050643
```

In [ ]:

```
1
```

Out[11]:

```
66325.44884050643
```

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

In [15]:

```
1 sal["OvertimePay"].max()
```

Out[15]:

```
245131.88
```

In [ ]:

```
1
```

Out[12]:

```
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 [21]:

```
1 sal[sal['EmployeeName']=='Joseph Driscoll']['JobTitle']
```

Out[21]:

```
36198    Captain, Fire Suppression
Name: JobTitle, dtype: object
```

In [ ]:

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

Out[19]:

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

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

In [43]:

```
s=sal[sal['EmployeeName']=='JOSEPH DRISCOLL']["TotalPayBenefits"]
pd.DataFrame(s)
```

Out[43]:

	TotalPayBenefits
24	270324.91

In [ ]:

```

```

Out[22]:

	TotalPayBenefits
24	270324.91

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

In [44]:

```
sal["TotalPayBenefits"].max()
```

Out[44]:

```
567595.43
```

In [48]:

```
sal[sal['TotalPayBenefits']==567595.43]
```

Out[48]:

	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

In [ ]:

```
1
```

Out[35]:

	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

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

In [50]:

```
1 sal['JobTitle'].nunique()
2
```

Out[50]:

```
2159
```

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

In [61]:

```
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
```

In [64]:

```
1 sal['JobTitle'].value_counts().head()
```

Out[64]:

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

In [ ]:

```
1
2
```

Out[49]:

```
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?) **

In [90]:

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

Out[90]:

```
202
```

In [101]:

```
1 pd.Series(sal[sal['Year']==2013]['JobTitle'].value_counts()==1).tail(202)
```

Out[101]:

```
IS Technician Assistant          True
Real Estate Devt. Mgr, SFMTA     True
Crime Lab Mgr                    True
Trnst Power Line Wrk Sprv 2      True
Airport Communications Officer   True
                                 ...
Farmer                           True
Chf Victim/Witness Invstgtor     True
Senior Offset Machine Operator   True
Asst Superintendent Rec          True
Employee Assistance Counselor    True
Name: JobTitle, Length: 202, dtype: bool
```

Great Job!