PANDAS:

A) IMPORTING PANDAS & READING DATASET:

```
import pandas as pd
```

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

```
sal.head()
```

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	TotalPayBenefits	Year	Notes	Agency	Status
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.00	400184.25	NaN	567595.43	567595.43	2011	NaN	San Francisco	NaN
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	538909.28	538909.28	2011	NaN	San Francisco	NaN
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN	335279.91	335279.91	2011	NaN	San Francisco	NaN
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.00	56120.71	198306.90	NaN	332343.61	332343.61	2011	NaN	San Francisco	NaN
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT,(FIRE DEPARTMENT)	134401.60	9737.00	182234.59	NaN	326373.19	326373.19	2011	NaN	San Francisco	NaN

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

```
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
```

C) **What is the average BasePay ?**

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

```
66325.4488404877
```

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

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

```
245131.88
```

E) **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).**

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

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

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

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

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

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

```
ind = sal['TotalPayBenefits'].idxmax()
sal.loc[ind]['EmployeeName']
```

```
'NATHANIEL FORD'
```

H) **What is the name of lowest paid person (including benefits)? Do you notice something strange about how much he or she is paid?**

```
ind = sal['TotalPayBenefits'].idxmin()
sal.iloc[ind]
```

```
Id                                   148654
EmployeeName                      Joe Lopez
JobTitle           Counselor, Log Cabin Ranch
BasePay                                 0.0
OvertimePay                             0.0
OtherPay                             -618.13
Benefits                                0.0
TotalPay                             -618.13
TotalPayBenefits                     -618.13
Year                                    2014
Notes                                    NaN
Agency                         San Francisco
Status                                   NaN
Name: 148653, dtype: object
```

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

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

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

J) How many unique job titles are there?

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

```
2159
```

K) What are the top 5 most common jobs?

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

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

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

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

```
202
```

M) How many people have the word Chief in their job title?

```
def chief_string(title):
    if 'chief' in title.lower().split():
        return True
    else:
        return False
sum(sal['JobTitle'].apply(lambda x:chief_string(x)))
```

```
477
```

N) Is there a correlation between length of the Job Title string and Salary?

```
sal['title_len']=sal['JobTitle'].apply(len)
sal[['TotalPayBenefits','title_len']].corr()
```

	TotalPayBenefits	title_len
TotalPayBenefits	1.000000	-0.036878
title_len	-0.036878	1.000000

CONCLUSION:

From this practical, I have successfully learned about pandas library in python.