SF Salaries Solutions

```
[128]  import pandas as pd
       ✓ 0.2s
```

Read Salaries.csv as a dataframe called sal.

```
[129]  sal = pd.read_csv(r'Data/Salaries.csv')
       ✓ 0.7s
```

```
...  C:\Users\hp\AppData\Roaming\Python\Python310\site-packages\IPython\core\interactiveshell.py:3251: DtypeWarning: Columns
     (3,4,5,6,12) have mixed types.Specify dtype option on import or set low_memory=False.
       exec(code_obj, self.user_global_ns, self.user_ns)
```

Check the head of the Dataframe.

```
[130]  sal.head()
       ✓ 0.8s
```

```
sal.head()
```

[130] ✓ 0.8s Python

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	TotalPayBenefits	Year	Notes	Agency	St
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.0	400184.25	NaN	567595.43	567595.43	2011	NaN	San Francisco	
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	538909.28	538909.28	2011	NaN	San Francisco	
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.6	NaN	335279.91	335279.91	2011	NaN	San Francisco	
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.0	56120.71	198306.9	NaN	332343.61	332343.61	2011	NaN	San Francisco	
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT, (FIRE	134401.6	9737.0	182234.59	NaN	326373.19	326373.19	2011	NaN	San Francisco	

```
sal.set_index("Id", inplace=True)
sal.head()
```

[131] ✓ 0.1s Python

Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	TotalPayBenefits	Year	Notes	Agency	Status
1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.0	400184.25	NaN	567595.43	567595.43	2011	NaN	San Francisco	NaN
2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	538909.28	538909.28	2011	NaN	San Francisco	NaN
3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.6	NaN	335279.91	335279.91	2011	NaN	San Francisco	NaN
4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.0	56120.71	198306.9	NaN	332343.61	332343.61	2011	NaN	San Francisco	NaN
		DEPUTY CHIEF OF										

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

```
[132]  sal.info()
       ✓ 0.1s
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 148654 entries, 1 to 148654
Data columns (total 12 columns):
 #   Column            Non-Null Count   Dtype  
---  ------            --------------   -----  
 0   EmployeeName      148654 non-null  object 
 1   JobTitle          148654 non-null  object 
 2   BasePay           148049 non-null  object 
 3   OvertimePay       148654 non-null  object 
 4   OtherPay          148654 non-null  object 
 5   Benefits          112495 non-null  object 
 6   TotalPay          148654 non-null  float64
 7   TotalPayBenefits  148654 non-null  float64
 8   Year              148654 non-null  int64  
 9   Notes             0 non-null       float64
 10  Agency            148654 non-null  object 
```

What is the avarage TotalPay?

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

[133] ✓ 0.7s

```
74768.32197169267
```

What is the highest amount of TotalPay in the dataset

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

[134] ✓ 0.9s

```
567595.43
```

What is the title of JOSEPH DRISCOLL?

```
sal[sal['EmployeeName'] == 'ALBERT PARDINI']['JobTitle']
```

[135] ✓ 0.1s

What is the title of JOSEPH DRISCOLL?

```
[135]  sal[sal['EmployeeName'] == 'ALBERT PARDINI']['JobTitle']
       ✓ 0.1s
```

```
Id
3    CAPTAIN III (POLICE DEPARTMENT)
Name: JobTitle, dtype: object
```

How Much does ALBERT PARDINI make(including benefits)?

```
[136]  sal[sal['EmployeeName'] == "ALBERT PARDINI"]["TotalPayBenefits"]
       ✓ 0.1s
```

```
Id
3    335279.91
Name: TotalPayBenefits, dtype: float64
```

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

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

[137] ✓ 0.9s Python

Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	TotalPayBenefits	Year	Notes	Agency	Status
1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.0	400184.25	NaN	567595.43	567595.43	2011	NaN	San Francisco	NaN

```
sal.loc[sal["TotalPayBenefits"].idxmax()]
```

[138] ✓ 0.9s Python

```
EmployeeName                                  NATHANIEL FORD
JobTitle           GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY
BasePay                                              167411.18
OvertimePay                                                0.0
```

```
sal.iloc[sal["TotalPayBenefits"].argmax()]
```

[139] ✓ 0.8s

```
EmployeeName                                       NATHANIEL FORD
JobTitle           GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY
BasePay                                               167411.18
OvertimePay                                                0.0
OtherPay                                              400184.25
Benefits                                                   NaN
TotalPay                                              567595.43
TotalPayBenefits                                      567595.43
Year                                                       2011
Notes                                                      NaN
Agency                                           San Francisco
Status                                                     NaN
Name: 1, dtype: object
```

Low Paid Employee with benifits

Low Paid Employee with benifits

```
sal[sal["TotalPayBenefits"] == sal["TotalPayBenefits"].min()]
```

[140] ✓ 0.1s Python

Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	TotalPayBenefits	Year	Notes	Agency	Status
148654	Joe Lopez	Counselor, Log Cabin Ranch	0.00	0.00	-618.13	0.00	-618.13	-618.13	2014	NaN	San Francisco	PT

```
sal.loc[sal["TotalPayBenefits"].idxmin()]
```

[141] ✓ 0.9s Python

```
EmployeeName                     Joe Lopez
JobTitle           Counselor, Log Cabin Ranch
BasePay                               0.00
OvertimePay                           0.00
OtherPay                           -618.13
Benefits                              0.00
```

```
sal.iloc[sal["TotalPayBenefits"].argmin()]
```

[142] ✓ 0.9s

```
EmployeeName                    Joe Lopez
JobTitle          Counselor, Log Cabin Ranch
BasePay                             0.00
OvertimePay                         0.00
OtherPay                         -618.13
Benefits                            0.00
TotalPay                         -618.13
TotalPayBenefits                 -618.13
Year                                2014
Notes                                NaN
Agency                     San Francisco
Status                                PT
Name: 148654, dtype: object
```

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

```
[143]  sal.groupby('Year').mean()['TotalPay']
```

✓ 0.1s

```
Year
2011    71744.103871
2012    74113.262265
2013    77611.443142
2014    75463.918140
Name: TotalPay, dtype: float64
```

How many Unique Job titles here

```
[144]  sal["JobTitle"].nunique()
```

✓ 0.1s

```
2159
```

What are the top 5 most common jobs?

```
[145]  sal["JobTitle"].value_counts().head(5)
```

✓ 0.1s

```
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?

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

✓ 0.1s

```
202
```

How Many people have the word chief in their job title? (This is pretty tricky)

```
def chief_string(title):
    if 'chief' in title.lower().split():
        return True
    else:
        return False
```

[147] ✓ 0.9s

```
chief_string('CHIEF MANAGER-METROPOLITAN TRANSIT AUTHORITY')
```

[148] ✓ 0.8s

```
True
```

```
sum(sal['JobTitle'].apply(lambda x: chief_string(x)))
```

[149] ✓ 0.3s

```
477
```

Bouns: Is there a correlation between length of the job title string and Salary?

```
[150]  sal['title_len'] = sal['JobTitle'].apply(len)
```

✓ 0.2s

```
[155]  sal[['TotalPayBenefits', 'title_len']].corr()
```

✓ 0.1s

	TotalPayBenefits	title_len
TotalPayBenefits	1.000000	-0.036878
title_len	-0.036878	1.000000