

SFO Public Department, referred to as SFO, has captured the salary data of all its employees for the years 2011-2014. Now in 2018, the organization is facing some financial crisis. As a first step, HR wants to rationalize employee cost to save payroll budget. You have to perform data manipulation and answer the questions below:

1. How much total salary cost has increased from year 2011 to 2014?
2. Who was the top-earning employee across all the years? Objective: Perform data manipulation and visualization techniques.

```
In [1]: import numpy,pandas
import seaborn
%matplotlib inline
import matplotlib.pyplot as plt
```

```
In [4]: df = pandas.read_csv(r'C:\Users\ctoqu\Desktop\SALARIES.csv')
```

```
C:\Users\ctoqu\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3
071: DtypeWarning: Columns (12) have mixed types.Specify dtype option on impo
rt or set low_memory=False.
has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

```
In [5]: df.head()
```

Out[5]:

	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

```
In [8]: #Checking the mean salary cost per year and see how has increase per year (answer 1)  
mean_year = df.groupby('Year').mean()['TotalPayBenefits']  
print(mean_year)
```

```
Year  
2011    71743.819645  
2012   100551.886807  
2013   101440.519714  
2014   100261.438668  
Name: TotalPayBenefits, dtype: float64
```

```
In [10]: #Group the total salary with respect to employee name  
top_sal = df.groupby('EmployeeName').sum()['TotalPayBenefits']  
print((top_sal.sort_values(axis=0)))
```

```
EmployeeName  
Joe Lopez          -618.13  
David P Kucia      -33.89  
Mark E Laherty     -8.20  
Timothy E Gibson   -2.73  
Mark W McClure      0.00  
...  
Richard Lee       1268668.84  
Stanley Lee       1457279.68  
William Wong      1459954.17  
Steven Lee        1706550.76  
Kevin Lee         1778487.17  
Name: TotalPayBenefits, Length: 110809, dtype: float64
```

```
In [ ]:
```