In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_csv("Salaries.csv", low_memory=False)
df.head()
```

Out[2]:

	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 [3]:
```
# 1.Compute how much total salary cost has increased from year 2011 to 2014
grouped_df = df.groupby(["Year"]).agg("sum").filter(["Year", "TotalPay"])
grouped_df

total_pay_2011 = grouped_df.query("Year == 2011")["TotalPay"]
total_pay_2014 = grouped_df.query("Year == 2014")["TotalPay"]
total_change_2011_2014 = float(total_pay_2014) - float(total_pay_2011)

print("Total change from 2011 to 2014 is>> " + str(total_change_2011_2014))
```

```
Total change from 2011 to 2014 is>> 282797920.53994226
```

In [4]:
```
# 2.Which Job Title in Year 2014 has highest mean salary?
grouped_df_jobtitle = df.query("Year == 2014").groupby(["JobTitle"]).agg('mea
n').filter(["JobTitle", "TotalPay"])
grouped_df_jobtitle
```

Out[4]:

JobTitle	TotalPay
ACPO,JuvP, Juv Prob (SFERS)	62290.780000
ASR Senior Office Specialist	62386.654091
ASR-Office Assistant	44687.796000
Account Clerk	37681.233226
Accountant I	45404.635000
...	...
Wire Rope Cable Maint Sprv	190575.900000
Worker's Comp Supervisor 1	56689.446667
Worker's Compensation Adjuster	70636.455385
X-Ray Laboratory Aide	51211.566857
Youth Comm Advisor	36465.910000

996 rows × 1 columns

In [5]:
```
max_min_salary = grouped_df_jobtitle[grouped_df_jobtitle["TotalPay"] == groupe
d_df_jobtitle["TotalPay"].max()]
print("Job Title with highest mean salary is >>\n" + str(max_min_salary))
```

```
Job Title with highest mean salary is >>
                          TotalPay
JobTitle
Chief Investment Officer  339653.7
```

In [6]:
```
# 3.How much money could have been saved in Year 2014 by stopping OverTimePay?
grouped_df_overtime = df.groupby(["Year"]).agg("sum").filter(["OvertimePay"]).
query("Year == 2014")
print(str(float(grouped_df_overtime["OvertimePay"])) + " money can be saved in
2014 by stopping overtime pay")
```

```
205918599.27000067 money can be saved in 2014 by stopping overtime pay
```

In [7]:
```
# 4.Which are the top 5 common job in Year 2014 and how much do they cost SFO
 ?
grouped_df_common_jobs = df.groupby(["Year", "JobTitle"]).agg("count").filter(
    ["Year", "TotalPay", "JobTitle"]).query("Year == 2014").sort_values("Total
Pay", ascending=False)
grouped_df_common_jobs
```

Out[7]:

		TotalPay
Year	**JobTitle**	
2014	**Transit Operator**	2479
	Special Nurse	1478
	Registered Nurse	1234
	Public Svc Aide-Public Works	916
	Firefighter	815
	...	...
	Marina Associate Manager	1
	Marina Assistant Manager	1
	Light Rail Vehicle Equip Eng	1
	Lieutenant (Police Department)	1
	Youth Comm Advisor	1

996 rows × 1 columns

In [10]:
```
# 5.Who was the top earning employee across all the years?
grouped_df_top_employee = df.groupby(["Year", "TotalPay"]).agg("max").filter([
"Year", "TotalPay", "EmployeeName"])
grouped_df_top_employee
```

Out[10]:

		EmployeeName
Year	**TotalPay**	
2011	**0.00**	PAULETTE ADAMS
	0.30	JOE BROWN JR
	4.17	SERENA HUGHES
	5.80	ROCK CRAWFORD
	8.86	MARCELLA TUCKER
...	**...**	...
2014	**335484.96**	Samson Lai
	339653.70	William J Coaker Jr.
	344187.46	Ellen G Moffatt
	390111.98	Amy P Hart
	471952.64	David Shinn

139333 rows × 1 columns

In [ ]: