DATA_MANIPULATION

April 21, 2021

[ ]:
```
import pandas as pd
import seaborn as sns
import matplotlib as plt
%matplotlib inline
```

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

# Total salary cost has increased from year 2011 to 2014

df.info()
df.head()
#df['Year'].unique()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148648 entries, 0 to 148647
Data columns (total 13 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   Id                148648 non-null  int64
 1   EmployeeName      148648 non-null  object
 2   JobTitle          148648 non-null  object
 3   BasePay           148043 non-null  float64
 4   OvertimePay       148648 non-null  float64
 5   OtherPay          148648 non-null  float64
 6   Benefits          112490 non-null  float64
 7   TotalPay          148648 non-null  float64
 8   TotalPayBenefits  148648 non-null  float64
 9   Year              148648 non-null  int64
 10  Notes             0 non-null       float64
 11  Agency            148648 non-null  object
 12  Status            38119 non-null   object
dtypes: float64(7), int64(2), object(4)
memory usage: 14.7+ MB
```

[2]:

	Id	EmployeeName	…	Agency	Status
0	1	NATHANIEL FORD	…	San Francisco	NaN
1	2	GARY JIMENEZ	…	San Francisco	NaN

```
2   3     ALBERT PARDINI  …  San Francisco     NaN
3   4  CHRISTOPHER CHONG  …  San Francisco     NaN
4   5    PATRICK GARDNER  …  San Francisco     NaN

[5 rows x 13 columns]
```

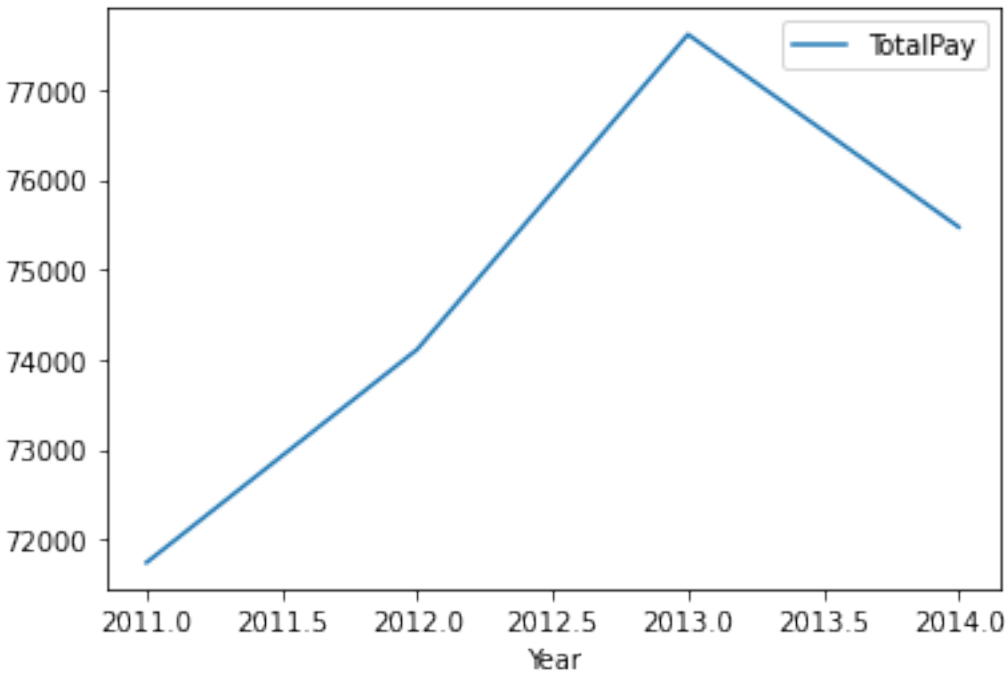
[6]:
```
feature=df[['Year','TotalPay']]
feature

salary_mean = df.groupby('Year').mean()[['TotalPay']]
print(salary_mean)

salary_dif = salary_mean.loc[2014]-salary_mean.loc[2011]
salary_dif
```

```
          TotalPay
Year
2011  71743.819645
2012  74112.234931
2013  77611.443142
2014  75471.836912
```

[6]:
```
TotalPay    3728.017267
dtype: float64
```

[4]:
```
sns.lineplot(data=salary_mean)
```

[4]:
```
<AxesSubplot:xlabel='Year'>
```



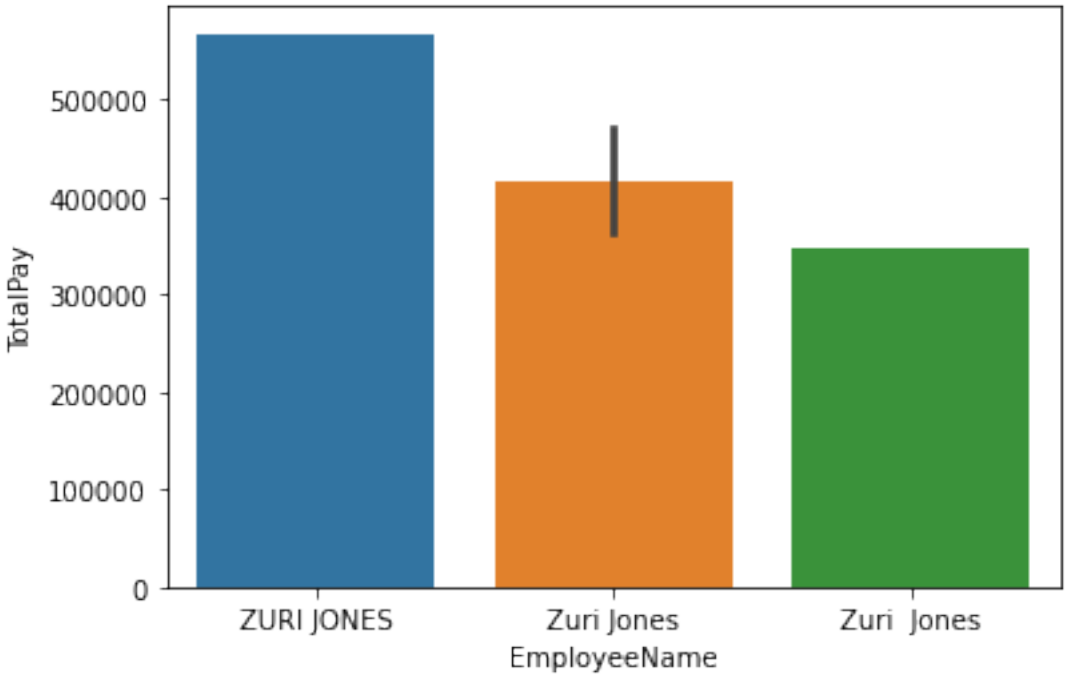
[28]:
```
emp_mean = df.groupby('Year').max()[['TotalPay','EmployeeName']].reset_index()
emp_mean
```

[28]:

	Year	TotalPay	EmployeeName
0	2011	567595.43	ZURI JONES
1	2012	362844.66	Zuri Jones
2	2013	347102.32	Zuri Jones
3	2014	471952.64	Zuri Jones

[29]:
```
sns.barplot(data=emp_mean,x='EmployeeName', y='TotalPay')
```

[29]:
```
<AxesSubplot:xlabel='EmployeeName', ylabel='TotalPay'>
```


500000
400000
300000
200000
100000
0
TotalPay
ZURI JONES
Zuri Jones
Zuri Jones
EmployeeName

[ ]: