
(http://www.pieriandata.com)

SF Salaries Exercise

Welcome to a quick exercise for you to practice your pandas skills! We will be using the SF Salaries Dataset (https://www.kaggle.com/kaggle/sf-salaries) from Kaggle! Just follow along and complete the tasks outlined in bold below. The tasks will get harder and harder as you go along.

** Import pandas as pd.**

In [1]:
```
import pandas as pd
```

** Read Salaries.csv as a dataframe called sal.**

In [2]:
```
df=pd.read_csv("salaries.csv")
```

** Check the head of the DataFrame. **

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

Out[3]:

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	To
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	

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

In [*]:
```
df.info()
```

What is the average BasePay ?

In [*]:
```
df["BasePay"].mean()
```

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

In [*]:
```
df["OvertimePay"].max()
```

** 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). **

In [*]:
```
df[df["EmployeeName"]=="JOSEPH DRISCOLL"]["JobTitle"]
```

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

In [*]:
```
df[df["EmployeeName"]=="JOSEPH DRISCOLL"]["TotalPayBenefits"]
```

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

In [*]:
```
df[df["TotalPayBenefits"]==df["TotalPayBenefits"].max()]
```

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

In [*]:
```
df[df["TotalPayBenefits"]==df["TotalPayBenefits"].min()]
```

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

In [*]:
```
df.groupby("Year").mean()["BasePay"]
```

** How many unique job titles are there? **

In [*]:
```
df["JobTitle"].nunique()
```

** What are the top 5 most common jobs? **

In [*]:
```
jobs=df.groupby("JobTitle").count()
top=jobs.sort_values(by="Id", ascending=False)[:5]
top["Id"]
```

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

one occurence in 2013?) **

In [*]:
```
year=df[df["Year"]==2013]
group=year.groupby("JobTitle").count()
count=group[group["Id"]==1]
count.count()["Id"]
```

** How many people have the word Chief in their job title? (This is pretty tricky) **

In [ ]:
```
```

In [*]:
```
def fun(job_title):
    if "chief" in job_title.lower().split():
        return True
    else:
        return False
df=pd.read_csv("salaries1.csv")

sum(df["JobTitle"].apply(lambda x: fun(x)))
```

** Bonus: Is there a correlation between length of the Job Title string and Salary? **

In [ ]:
```
```

In [*]:
```
df["title_len"]=df["JobTitle"].apply(len)
df[["title_len","TotalPayBenefits"]].corr()
```

Great Job!