

03 - Salaries

August 3, 2022

1 SF Salaries

```
# Import pandas as pd
import pandas as pd
```

```
# Read Salaries.csv
df=pd.read_csv("/home/aceec/Downloads/kris/Salaries.csv")
```

```
# set working directory
```

```
# Check the head of the DataFrame
df.head()
```

	Id	EmployeeName	
JobTitle \			
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)

	BasePay	OvertimePay	OtherPay	Benefits	TotalPay
0	167411.18	0.00	400184.25	NaN	567595.43
1	155966.02	245131.88	137811.38	NaN	538909.28
2	212739.13	106088.18	16452.60	NaN	335279.91
3	77916.00	56120.71	198306.90	NaN	332343.61
4	134401.60	9737.00	182234.59	NaN	326373.19

	Year	Notes	Agency	Status
0	2011	NaN	San Francisco	NaN
1	2011	NaN	San Francisco	NaN
2	2011	NaN	San Francisco	NaN
3	2011	NaN	San Francisco	NaN
4	2011	NaN	San Francisco	NaN

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148654 entries, 0 to 148653
Data columns (total 13 columns):
#   Column              Non-Null Count  Dtype

```

```

----
0   Id          148654 non-null  int64
1   EmployeeName 148654 non-null  object
2   JobTitle     148654 non-null  object
3   BasePay      148045 non-null  float64
4   OvertimePay  148650 non-null  float64
5   OtherPay     148650 non-null  float64
6   Benefits     112491 non-null  float64
7   TotalPay     148654 non-null  float64
8   TotalPayBenefits 148654 non-null float64
9   Year         148654 non-null  int64
10  Notes        0 non-null    float64
11  Agency       148654 non-null  object
12  Status       0 non-null    float64

```

dtypes: float64(8), int64(2), object(3)

memory usage: 14.7+ MB

What is the average BasePay?

```
df1= df[["BasePay"]].mean()
df1
```

```
BasePay    66325.448841
dtype: float64
```

What is the highest amount of OvertimePay in the dataset?

```
df2= df[["OvertimePay"]].max()
df2
```

```
OvertimePay    245131.88
dtype: float64
```

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)

```
df3=df[df['EmployeeName']=='JOSEPH DRISCOLL']['JobTitle']
df3
```

```
24    CAPTAIN, FIRE SUPPRESSION
Name: JobTitle, dtype: object
```

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

```
df4=df[df['EmployeeName']=='JOSEPH DRISCOLL']['TotalPayBenefits']
df4
```

```
24    270324.91
Name: TotalPayBenefits, dtype: float64
```

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

```
ind1=df[['TotalPayBenefits']].idxmax()
df5=df.loc[ind1]['EmployeeName']
df5
```

```

0    NATHANIEL FORD
Name: EmployeeName, dtype: object

# What is the name of lowest paid person (including benefits)? Do you
notice something strange
# about how much he or she is paid?
ind2=df[['TotalPayBenefits']].idxmin()
df6=df.iloc[ind2]
df6

      Id EmployeeName      JobTitle BasePay
OvertimePay \
148653  148654      Joe Lopez  Counselor, Log Cabin Ranch      0.0
0.0

      OtherPay  Benefits  TotalPay  TotalPayBenefits  Year  Notes  \
148653    -618.13      0.0    -618.13          -618.13  2014   NaN

      Agency Status
148653  San Francisco      NaN

#What was the average (mean) BasePay of all employees per year? (2011-
2014) ?
df7=df.groupby('Year').mean()['BasePay']
df7

Year
2011    63595.956517
2012    65436.406857
2013    69630.030216
2014    66564.421924
Name: BasePay, dtype: float64

df8=df['JobTitle'].nunique()
df8

2159

#What are the top 5 most common jobs?
df9=df['JobTitle'].value_counts().head()
df9

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?
(e.g. Job Titles with only
#one occurrence in 2013?)

```

```
df10=(df[df['Year']==2013]['JobTitle'].value_counts()==1).sum()
df10
```

```
202
```

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

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

```
df[['JobTitle','title_len']].head()
```

	JobTitle	title_len
0	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	46
1	CAPTAIN III (POLICE DEPARTMENT)	31
2	CAPTAIN III (POLICE DEPARTMENT)	31
3	WIRE ROPE CABLE MAINTENANCE MECHANIC	36
4	DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)	44

```
df11=df[['JobTitle','title_len']].corr()
df11
```

	title_len
title_len	1.0