# SF Salaries Exercise¶

We will be using the SF Salaries Dataset 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.**

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

**Check the head of the DataFrame.**

In [8]:

Out[8]:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **0** 1 | GENERAL MANAGER-  NATHANIEL FORD METROPOLITAN TRANSIT 167411.18 0.00  AUTHORITY | | | 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 |

**Id EmployeeName JobTitle BasePay OvertimePay OtherPay Benefits TotalPay**

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

In [9]:

<class 'pandas.core.frame.DataFrame'> RangeIndex: 148654 entries, 0 to 148653

Data columns (total 13 columns):

Id 148654 non-null int64

EmployeeName 148654 non-null object

JobTitle 148654 non-null object

This study source was downloaded by 100000779193685 from CourseHero.com on 11-11-2022 11:33:09 GMT -06:00BasePay 148045 non-null float64

OvertimePay 148650 non-null float64 OtherPay 148650 non-null float64

Benefits 112491 non-null float64

TotalPay 148654 non-null float64

TotalPayBenefits 148654 non-null float64

Year 148654 non-null int64

Notes 0 non-null float64

Agency 148654 non-null object Status 0 non-null float64 dtypes: float64(8), int64(2), object(3) memory usage: 14.7+ MB

**What is the average BasePay ?**

In [10]:

Out[10]:

66325.44884050643

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

In [11]:

Out[11]:

245131.88

**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 [12]:

Out[12]:

24 CAPTAIN, FIRE SUPPRESSION

Name: JobTitle, dtype: object

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

In [13]:

Out[13]:

24 270324.91

Name: TotalPayBenefits, dtype: float64

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

In [14]:

Out[14]:

**Id EmployeeName JobTitle BasePay OvertimePay OtherPay Benefits TotalPay**

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

In [15]:

Out[15]:

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

In [16]:

|  |  |
| --- | --- |
| GENERAL MANAGER-  **0** 1 NATHANIEL FORD METROPOLITAN TRANSIT 167411.18 0.0  AUTHORITY | 400184.25 NaN 567595.43 |

|  |  |
| --- | --- |
| **Id EmployeeName** | **JobTitle BasePay OvertimePay OtherPay Benefits TotalPay Tota** |
| **148653** 148654 Joe Lopez | Counselor,  Log Cabin 0.0 0.0 -618.13 0.0 -618.13 -618  Ranch |

Out[16]:

Year

1. 63595.956517
2. 65436.406857
3. 69630.030216
4. 66564.421924

Name: BasePay, dtype: float64

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

In [17]:

Out[17]:

2159

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

In [18]:

Out[18]:

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 occurence in 2013?)**

In [19]:

Out[19]:

202

**How many people have the word Chief in their job title? (This is pretty tricky)** In [21]:

Out[21]: 423

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

In [23]:

Out[23]:

**title\_len TotalPayBenefits**

**title\_len** 1.000000 -0.036878

**TotalPayBenefits** -0.036878 1.000000

[Powered by TCPDF (www.tcpdf.org)](http://www.tcpdf.org/)