Data_Analysis_SF_Salaries

February 21, 2020

1 SF _ Salaries Data Analysis

```
[1]: import numpy as np
      import pandas as pd
 []: sal = pd.read_csv('Salaries.csv')
[18]:
      sal
[18]:
                  Ιd
                            EmployeeName
      0
                   1
                          NATHANIEL FORD
                   2
      1
                            GARY JIMENEZ
      2
                   3
                          ALBERT PARDINI
      3
                   4
                      CHRISTOPHER CHONG
                   5
                         PATRICK GARDNER
      148649
              148650
                           Roy I Tillery
      148650
              148651
                            Not provided
      148651
              148652
                            Not provided
      148652
              148653
                            Not provided
      148653
              148654
                               Joe Lopez
                                                      JobTitle
                                                                      BasePay \
      0
              GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY
                                                                       167411
      1
                              CAPTAIN III (POLICE DEPARTMENT)
                                                                       155966
                              CAPTAIN III (POLICE DEPARTMENT)
      2
                                                                       212739
      3
                         WIRE ROPE CABLE MAINTENANCE MECHANIC
                                                                        77916
      4
                DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)
                                                                       134402
      148649
                                                     Custodian
                                                                         0.00
      148650
                                                  Not provided
                                                                Not Provided
      148651
                                                  Not provided
                                                                Not Provided
      148652
                                                  Not provided
                                                                Not Provided
      148653
                                   Counselor, Log Cabin Ranch
                                                                         0.00
                                                           TotalPay TotalPayBenefits \
               OvertimePay
                                 OtherPay
                                                Benefits
      0
                                   400184
                                                          567595.43
                                                                             567595.43
```

									F0000	00
1		245132		13781	L	NaN	538909	. 28	538909	. ۷۵
2		106088		16452.6	3	NaN	335279	.91	335279	.91
3		56120.7		198307	7	NaN	332343	. 61	332343	.61
4		9737		18223	5	NaN	326373	. 19	326373	. 19
		•••		•••		•••		•••		
148649		0.00		0.00)	0.00	0	.00	0	.00
148650	Not	Provided	Not	Provided	l Not F	Provided	0	.00	0	.00
148651	Not	Provided	Not	Provided	l Not F	Provided	0	.00	0	.00
148652	Not	Provided	Not	Provided	l Not F	Provided	0	.00	0	.00
148653		0.00		-618.13	3	0.00	-618	. 13	-618	. 13
	Year	Notes		Agency	/ Status	3				
0	2011	NaN	San I	Francisco						
1	2011	NaN	San I	Francisco	n NaN	V				
2	2011			Francisco						
3	2011			Francisco						
4	2011	NaN		Francisco						
	•••	•••								
148649	2014	NaN	San I	Francisco) P1	Γ				
148650	2014	NaN	San I	Francisco	o NaN	V				
148650 148651	2014 2014			Francisco Francisco						
148651	2014	NaN	San I		na Na	Ŋ				
		NaN NaN	San I	Francisco	o NaN o NaN	1 1				
148651 148652 148653	2014 2014 2014 1 rows	NaN NaN	San I San I San I	Francisco Francisco Francisco	o NaN o NaN	1 1				
148651 148652 148653 [148654	2014 2014 2014 1 rows	NaN NaN NaN x 13 co	San I San I San I	Francisco Francisco Francisco	o NaN o NaN	1 1		Te	hTi+la	
148651 148652 148653 [148654 sal.hea	2014 2014 2014 1 rows	NaN NaN NaN x 13 co	San I San I San I lumns	Francisco Francisco Francisco	o NaN	V Г	I.TTAN TF			\
148651 148652 148653 [148654 sal.hea Id 0 1	2014 2014 2014 1 rows ad()	NaN NaN NaN x 13 co	San I San I San I Iumns	Francisco Francisco Francisco	o NaN	V V C R-METROPO		RANSIT AUT	THORITY	\
148651 148652 148653 [148654 sal.hea Id 0 1 1 2	2014 2014 2014 1 rows ad()	NaN NaN NaN x 13 col Employeel THANIEL I	San I San I San I lumns Name FORD ENEZ	Francisco Francisco Francisco	o NaN	N T R-METROPO CAPTAIN	III (POI	RANSIT AUT LICE DEPAF	THORITY RTMENT)	\
148651 148652 148653 [148654 sal.hea Id 0 1 1 2 2 3	2014 2014 2014 1 rows ad() NA	NaN NaN NaN x 13 co. Employeel THANIEL 1 GARY JIMI BERT PARI	San I San I San I lumns Name FORD ENEZ DINI	Francisco Francisco Francisco	o NaN o PT	N T R-METROPO CAPTAIN CAPTAIN	III (POI III (POI	RANSIT AUT LICE DEPAF LICE DEPAF	THORITY RTMENT) RTMENT)	\
148651 148652 148653 [148654 sal.hea Id 0 1 1 2 2 3 3 4	2014 2014 2014 1 rows ad() NA AL CHRIS	NaN NaN NaN x 13 co. Employeel THANIEL 1 GARY JIM BERT PARI	San I San I San I San I Lumns Name FORD ENEZ DINI HONG	Francisco Francisco Francisco	MANAGEF	N N C R-METROPO CAPTAIN CAPTAIN ROPE CAB	III (POI III (POI LE MAINT	RANSIT AUT LICE DEPAF LICE DEPAF FENANCE ME	THORITY RTMENT) RTMENT) ECHANIC	\
148651 148652 148653 [148654 sal.hea Id 0 1 1 2 2 3 3 4 4 5	2014 2014 2014 1 rows ad() NA AL CHRIS	NaN NaN NaN x 13 co. Employeel THANIEL I GARY JIMI BERT PARI TOPHER CI RICK GARI	San I San I San I lumns Name FORD ENEZ DINI HONG DNER	Francisco Francisco Francisco GENERAL DEPUTY	MANAGEF WIRE	N N C CAPTAIN CAPTAIN ROPE CAB OF DEPAR	III (POI III (POI LE MAINT TMENT,(I	RANSIT AUT LICE DEPAF LICE DEPAF FENANCE ME FIRE DEPAF	THORITY RTMENT) RTMENT) ECHANIC RTMENT)	
148651 148652 148653 [148654 sal.hea Id 0 1 1 2 2 3 3 4 4 5	2014 2014 2014 f rows ad() NA AL CHRIS PAT	NaN NaN NaN x 13 co. Employeel THANIEL 1 GARY JIMI BERT PARI TOPHER CI RICK GARI	San I San I San I San I Iumns Name FORD ENEZ DINI HONG DNER	Francisco Francisco Francisco GENERAL DEPUTY erPay Ber	MANAGEF WIRE CHIEF	N N T CAPTAIN CAPTAIN ROPE CAB OF DEPAR	III (POI III (POI LE MAINT TMENT,(I y Total	RANSIT AUT LICE DEPAF LICE DEPAF TENANCE ME FIRE DEPAF LPayBenefi	THORITY RTMENT) RTMENT) ECHANIC RTMENT)	`
148651 148652 148653 [148654 sal.hea Id 0 1 1 2 2 3 3 4 4 5 BaseF 0 1674	2014 2014 2014 1 rows ad() NA AL CHRIS PAT	NaN NaN NaN x 13 co. Employeel THANIEL I GARY JIMI BERT PARI TOPHER CI RICK GARI ertimePa	San I San I San I San I Iumns Name FORD ENEZ DINI HONG DNER	Francisco Francisco Francisco GENERAL DEPUTY erPay Ber 00184	MANAGER WIRE CHIEF CHIEF NaN	R-METROPO CAPTAIN CAPTAIN ROPE CAB OF DEPAR TotalPa 567595.4	III (POI III (POI LE MAINT TMENT,(I y Total	RANSIT AUT LICE DEPAF LICE DEPAF FENANCE ME FIRE DEPAF LPayBenefi 567595.	THORITY RTMENT) RTMENT) ECHANIC RTMENT) its Year 43 2011	\
148651 148652 148653 [148654 sal.hea Id 0 1 1 2 2 3 3 4 4 5 BaseF 0 1674 1 1559	2014 2014 2014 1 rows ad() NA AL CHRIS PAT	NaN NaN NaN x 13 co. Employeel THANIEL I GARY JIMI BERT PARI TOPHER CI RICK GARI ertimePay	San I San I San I San I Iumns Name FORD ENEZ DINI HONG DNER y Othe 0 40	Francisco Francisco Francisco GENERAL DEPUTY erPay Ber 00184 37811	MANAGER WIRE CHIEF CHIEF NaN NaN	R-METROPO CAPTAIN ROPE CAB OF DEPAR TotalPa 567595.4 538909.2	III (POI III (POI LE MAINT TMENT,(I y Total 3	RANSIT AUT LICE DEPAR LICE DEPAR TENANCE ME FIRE DEPAR LPayBenefi 567595.	THORITY RTMENT) RTMENT) ECHANIC RTMENT) its Year 143 2011	`
148651 148652 148653 [148654 sal.hea Id 0 1 1 2 2 3 3 4 4 5 BaseF 0 1674 1 1559 2 2127	2014 2014 2014 1 rows ad() NA' AL. CHRIS' PAT' Pay Ov 111 966 739	NaN NaN NaN x 13 co. Employeel THANIEL 1 GARY JIM BERT PARI TOPHER CI RICK GARI ertimePay (245133	San I	Francisco Francisco Francisco GENERAL DEPUTY erPay Ben 00184 37811 452.6	MANAGER WIRE CHIEF CHIEF NaN NaN NaN	R-METROPO CAPTAIN CAPTAIN ROPE CAB OF DEPAR TotalPa 567595.4 538909.2 335279.9	III (POI III (POI LE MAINT TMENT,(I y Total 3 8 1	RANSIT AUT LICE DEPAR LICE DEPAR TENANCE ME FIRE DEPAR LPayBenefi 567595. 538909. 335279.	THORITY RTMENT) RTMENT) ECHANIC RTMENT) its Year .43 2011 .28 2011	\
148651 148652 148653 [148654 sal.hea Id 0 1 1 2 2 3 3 4 4 5 BaseF 0 1674 1 1559	2014 2014 2014 1 rows ad() NA AL CHRIS PAT Pay Ov 111 966 739	NaN NaN NaN x 13 co. Employeel THANIEL I GARY JIMI BERT PARI TOPHER CI RICK GARI ertimePay	San I San I San I San I Iumns Name FORD ENEZ DINI HONG DNER y Othe 0 40 2 13 8 164 7 19	Francisco Francisco Francisco GENERAL DEPUTY erPay Ber 00184 37811	MANAGER WIRE CHIEF CHIEF NaN NaN	R-METROPO CAPTAIN ROPE CAB OF DEPAR TotalPa 567595.4 538909.2	III (POI III (POI LE MAINT TMENT,(I y Total 3 8 1	RANSIT AUT LICE DEPAR LICE DEPAR TENANCE ME FIRE DEPAR LPayBenefi 567595.	THORITY RTMENT) RTMENT) ECHANIC RTMENT) its Year 43 2011 28 2011 91 2011	\

Agency Status

 ${\tt NaN}$

 ${\tt NaN}$

 ${\tt NaN}$

 ${\tt NaN}$

Notes

 ${\tt NaN}$

 ${\tt NaN}$

 ${\tt NaN}$

0

1

2

3

NaN San Francisco

San Francisco

San Francisco

San Francisco

4 NaN San Francisco NaN

[76]:	sal.tai	1(10)					
[76]:		Id	EmployeeName		JobTitle	BasePay	\
	148644	148645	Randy D Winn	Stationary Eng,	Sewage Plant	0.00	
	148645	148646 Car	olyn A Wilson	Human Servic	0.00		
	148646	148647	Not provided		Not provided	Not Provided	
	148647	148648 J	oann Anderson	Communications Dispatcher 2		0.00	
	148648	148649	Leon Walker		0.00		
	148649	148650	Roy I Tillery		Custodian	0.00	
	148650	148651	Not provided		Not Provided		
	148651	148652	Not provided		Not Provided		
	148652	148653	Not provided		Not Provided		
	148653	148654	Joe Lopez	Counselor, Lo	og Cabin Ranch	0.00	
		OvertimePa	y OtherPay	Benefits	TotalPay Tot	alPayBenefits	\
	148644	0.0	0.00	0.00	0.00	0.00	
	148645	0.0	0.00	0.00	0.00	0.00	
	148646	Not Provide	d Not Provided	Not Provided	0.00	0.00	
	148647	0.0	0.00	0.00	0.00	0.00	
	148648	0.0	0.00	0.00	0.00	0.00	
	148649	0.0	0.00	0.00	0.00	0.00	
	148650	Not Provide	d Not Provided	Not Provided	0.00	0.00	
	148651	Not Provide	d Not Provided	Not Provided	0.00	0.00	
	148652	Not Provide	d Not Provided	Not Provided	0.00	0.00	
	148653	0.0	0 -618.13	0.00	-618.13	-618.13	
		Year Notes	Agency	Status title_	len		
	148644	2014 NaN	San Francisco	PT	28		
	148645	2014 NaN	San Francisco	PT	25		
	148646	2014 NaN	San Francisco	NaN	12		
	148647	2014 NaN	San Francisco	PT	27		
	148648	2014 NaN	San Francisco	PT	9		
	148649	2014 NaN	San Francisco	PT	9		
	148650	2014 NaN	San Francisco	NaN	12		
	148651	2014 NaN	San Francisco	NaN	12		
	148652	2014 NaN	San Francisco	NaN	12		
	148653	2014 NaN	San Francisco	PT	26		

[6]: sal.info() # no. of column and no. of entry rows

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148654 entries, 0 to 148653
Data columns (total 13 columns):

Column Non-Null Count Dtype

```
EmployeeName
      1
                            148654 non-null object
      2
          JobTitle
                            148654 non-null object
      3
          BasePay
                            148049 non-null object
          OvertimePay
                            148654 non-null object
      4
      5
          OtherPay
                            148654 non-null object
      6
          Benefits
                            112495 non-null object
      7
          TotalPay
                            148654 non-null float64
          TotalPayBenefits 148654 non-null float64
      9
                            148654 non-null int64
          Year
                                             float64
      10 Notes
                            0 non-null
                            148654 non-null object
      11
         Agency
      12
          Status
                            38119 non-null
                                             object
     dtypes: float64(3), int64(2), object(8)
     memory usage: 14.7+ MB
[20]: sal['BasePay']
[20]: 0
                      167411
      1
                      155966
      2
                      212739
                       77916
      3
      4
                      134402
      148649
                        0.00
                Not Provided
      148650
      148651
                Not Provided
      148652
                Not Provided
      148653
                        0.00
      Name: BasePay, Length: 148654, dtype: object
[33]: sal['TotalPay'].max()
[33]: 567595.43
     sal[sal['EmployeeName'] == 'ALBERT PARDINI']
[44]:
         Ιd
               EmployeeName
                                                    JobTitle BasePay OvertimePay \
           ALBERT PARDINI CAPTAIN III (POLICE DEPARTMENT)
                                                              212739
                                                                          106088
                            TotalPay TotalPayBenefits Year Notes
        OtherPay Benefits
                                                                            Agency \
      2 16452.6
                      NaN 335279.91
                                             335279.91 2011
                                                                NaN San Francisco
        Status
      2
           NaN
[46]: sal['TotalPayBenefits'] == sal['TotalPayBenefits'].max()
```

148654 non-null

int64

0

Ιd

```
[46]: 0
                 True
                False
      2
                False
      3
                False
      4
                False
      148649
                False
      148650
                False
      148651
                False
      148652
                False
      148653
                False
      Name: TotalPayBenefits, Length: 148654, dtype: bool
[47]: sal.loc[sal['TotalPayBenefits'].idxmax()]
      # sal.iloc[sal['TotalPayBenefits'].argmax()]
[47]: Id
      EmployeeName
                                                            NATHANIEL FORD
      JobTitle
                          GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY
      BasePay
                                                                    167411
      OvertimePay
                                                                         0
                                                                    400184
      OtherPay
      Benefits
                                                                       NaN
      TotalPay
                                                                    567595
      TotalPayBenefits
                                                                    567595
      Year
                                                                      2011
      Notes
                                                                       NaN
                                                             San Francisco
      Agency
      Status
                                                                       NaN
      Name: 0, dtype: object
[48]: sal.iloc[sal['TotalPayBenefits'].argmin()]
      # sal[sal['TotalPayBenefits'] == sal['TotalPayBenefits'].min()]
      #lowest paid person including benefits
[48]: Id
                                               148654
      EmployeeName
                                            Joe Lopez
      JobTitle
                          Counselor, Log Cabin Ranch
      BasePay
                                                  0.00
      OvertimePay
                                                  0.00
      OtherPay
                                              -618.13
      Benefits
                                                 0.00
      TotalPay
                                              -618.13
      TotalPayBenefits
                                              -618.13
      Year
                                                  2014
      Notes
                                                  NaN
```

```
Status
                                                  PT
      Name: 148653, dtype: object
[55]: sal.groupby('Year').mean()
      # average mean all employees per year
[55]:
                  Ιd
                          TotalPay TotalPayBenefits Notes
     Year
      2011
             18080.0 71744.103871
                                        71744.103871
                                                        NaN
      2012
             54542.5 74113.262265
                                       100553.229232
                                                        NaN
      2013
             91728.5 77611.443142
                                       101440.519714
                                                        NaN
      2014 129593.0 75463.918140
                                       100250.918884
                                                        NaN
[57]: sal['JobTitle'].nunique() #unique job title
[57]: 2159
     sal['JobTitle'].value_counts().head(5) # top 5 most common jobs
[58]: Transit Operator
                                      7036
      Special Nurse
                                      4389
     Registered Nurse
                                      3736
     Public Svc Aide-Public Works
                                      2518
     Police Officer 3
                                      2421
     Name: JobTitle, dtype: int64
[61]: sum(sal[sal['Year']==2013]['JobTitle'].value_counts() ==1)
      #How many job titles were represented by only one person in 2013
[61]: 202
[62]: def chief_string(title): #upper and lower case that's why used functions
          if 'chief' in title.lower().split():
              return True
          else:
              return False
 []:
[67]: sal['JobTitle'].iloc[0]
[67]: 'GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY'
[72]:
      sal['title_len'] = sal['JobTitle'].apply(len)
[73]: sal[['JobTitle','title_len']]
```

San Francisco

Agency

```
[73]:
                                                     JobTitle title_len
      0
              GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY
                                                                      46
                             CAPTAIN III (POLICE DEPARTMENT)
                                                                      31
      1
      2
                             CAPTAIN III (POLICE DEPARTMENT)
                                                                      31
      3
                        WIRE ROPE CABLE MAINTENANCE MECHANIC
                                                                      36
                DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)
      4
                                                                      44
                                                    Custodian
                                                                       9
      148649
      148650
                                                 Not provided
                                                                      12
      148651
                                                 Not provided
                                                                      12
                                                 Not provided
      148652
                                                                      12
      148653
                                  Counselor, Log Cabin Ranch
                                                                      26
      [148654 rows x 2 columns]
[74]: sal[['JobTitle','title_len']].corr() #correlation
[74]:
                 title_len
                       1.0
      title_len
[75]: sal[['TotalPayBenefits','title_len']].corr() # no-correlation
[75]:
                        TotalPayBenefits title_len
      TotalPayBenefits
                                1.000000
                                          -0.036878
     title_len
                               -0.036878
                                            1.000000
 []:
```