```
import pandas as pd

data = pd.read_csv('Salaries.csv')

data
```

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	TotalPayBenefits	Year	Notes	Ag
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.00	400184.25	NaN	567595.43	567595.43	2011.0	NaN	Fran
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	538909.28	538909.28	2011.0	NaN	Fran
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN	335279.91	335279.91	2011.0	NaN	Fran
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.00	56120.71	198306.90	NaN	332343.61	332343.61	2011.0	NaN	Fran
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)	134401.60	9737.00	182234.59	NaN	326373.19	326373.19	2011.0	NaN	Fran
...	...	...	...	...	...	...	...	...	...	...	...	...
40404	40405	David Parry	Police Officer 3	117171.44	15729.27	9444.58	34797.29	142345.29	177142.58	2012.0	NaN	Fran

Next steps: Generate code with data View recommended plots

Display Top 10 Rows of The Dataset

```
data.head(10)
```

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	Total
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.00	400184.25	NaN	56759
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	53890
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN	33527
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.00	56120.71	198306.90	NaN	33234
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT, (FIRE DEPARTMENT)	134401.60	9737.00	182234.59	NaN	32637
5	6	DAVID SULLIVAN	ASSISTANT DEPUTY CHIEF II	118602.00	8601.00	189082.74	NaN	31628

Next steps: Generate code with data View recommended plots

Display last 10 Rows of The Dataset

```
data.tail(10)
```

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	Tot
40399	40400	Irina Tomashevsky	IS Engineer-Principal	130728.00	0.00	153.85	46287.30	1308
40400	40401	Benjamin Ma	Electronic Maintenance Tech	101088.00	34695.30	328.05	41051.70	1361
40401	40402	Richard Look	IS Engineer-Principal	130727.99	0.00	289.77	46145.01	1310
40402	40403	Michael Johnson	Transit Power Line Worker	101660.00	32624.07	0.00	42868.60	1342
40403	40404	Umesh Gupta	IS Project Director	130728.00	0.00	289.77	46125.98	1310
40404	40405	David Barry	Police	117171.44	15729.27	9444.58	34797.29	1423

Find Shape of Our Dataset (Number of Rows And Number of Columns)

```
data.shape
```

```
(40409, 13)
```

```
print("Number of Rows", data.shape[0])
```

```
Number of Rows 40409
```

```
print("Number of Columns", data.shape[1])
```

```
Number of Columns 13
```

Getting Information About Our Dataset Like Total Number Rows, Total Number of Columns, Datatypes of Each Column And Memory Requirement

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40409 entries, 0 to 40408
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   Id                40409 non-null  int64  
 1   EmployeeName      40409 non-null  object 
 2   JobTitle          40409 non-null  object 
 3   BasePay           40408 non-null  float64
 4   OvertimePay       40408 non-null  float64
 5   OtherPay          40408 non-null  float64
 6   Benefits          4249 non-null   float64
 7   TotalPay          40408 non-null  float64
 8   TotalPayBenefits  40408 non-null  float64
 9   Year              40408 non-null  float64
 10  Notes             0 non-null      float64
 11  Agency            40408 non-null  object 
 12  Status            0 non-null      float64
dtypes: float64(9), int64(1), object(3)
memory usage: 4.0+ MB
```

Check Null Values In The Dataset

```
data.isnull().sum()
```

```
Id                    0
EmployeeName          0
JobTitle              0
BasePay               1
```

```
OvertimePay            1
OtherPay               1
Benefits           36160
TotalPay               1
TotalPayBenefits       1
Year                   1
Notes              40409
Agency                 1
Status             40409
dtype: int64
```

Drop ID, Notes, Agency, and Status Columns

```
data = data.drop(['Id', 'Notes', 'Agency', 'Status'], axis = 1)
```

```
data
```

	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	Total
0	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.00	400184.25	NaN	56759
1	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	53890
2	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN	33527
3	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.00	56120.71	198306.90	NaN	33234
	PATRICK	DEPUTY CHIEF OF					

Next steps: Generate code with data View recommended plots

Get Overall Statistics About The Dataframe

```
data.describe()
```

	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	TotalPayE
count	40408.000000	40408.000000	40408.000000	4249.000000	40408.000000	4040
mean	71096.685778	5982.550624	4509.982961	45494.410864	81589.219362	8637
std	45068.269721	13351.669634	9602.085625	8884.249780	54129.447643	6277
min	0.000000	0.000000	0.000000	189.790000	0.000000	
25%	40675.032500	0.000000	0.000000	38649.880000	44544.412500	4454
50%	66431.535000	0.000000	1074.835000	43804.430000	74444.545000	7444
75%	105195.145000	5424.432500	5237.362500	52360.450000	119558.342500	11955
max	302578.000000	245131.880000	400184.250000	84681.820000	567595.430000	56759

Find Occurance of Employee Name of top 5

```
data['EmployeeName'].value_counts().head()
```

```
EmployeeName
KEVIN LEE       9
MICHAEL LEE     8
DAVID WONG      8
RICHARD LEE     8
WILLIAM WONG    7
Name: count, dtype: int64
```

Find The Number of Unique Job Titles

```
data['JobTitle'].nunique()
```

```
1284
```

Total Number Of Job Title Contains Captain

```
len(data[data['JobTitle'].str.contains('Captain', case = False)])
```

```
274
```

Display all the employee name from the fire department

```
data[data['JobTitle'].str.contains('Fire Department', case = False)]["EmployeeName"]
```

```
4           PATRICK GARDNER
6                 ALSON LEE
8             MICHAEL MORRIS
9         JOANNE HAYES-WHITE
10             ARTHUR KENNEY
                   ...
5498           VINCENT PEREZ
8436           JENSEN RHODES
16285       AARON STEVENSON
32623          JAMES BARDEN
36162    Joanne Hayes-White
Name: EmployeeName, Length: 223, dtype: object
```

Find Minimum, Maximum, and Average BasePay

```
data['BasePay'].max()
```

```
302578.0
```

```
data['BasePay'].min()
```

```
0.0
```

```
data['BasePay'].mean()
```

```
71096.6857775688
```

Replace 'Not Provided' in EmployeeName' Column to NaN

```
import numpy as np
```

```
data['EmployeeName'] = data['EmployeeName'].replace('Not provided', np.nan)
```

```
data['EmployeeName']
```

```
0           NATHANIEL FORD
1             GARY JIMENEZ
2           ALBERT PARDINI
3         CHRISTOPHER CHONG
4           PATRICK GARDNER
                  ...
40404          David Parry
40405         Francisco Ho
40406      Manuel Gonzales
40407       Lauro Baca III
40408      Milagros Brosas
Name: EmployeeName, Length: 40409, dtype: object
```

Drop the rows having 5 missing values

```
data.drop(data[data.isnull().sum(axis=1) == 5].index, axis = 0, inplace = True)
```

Find Job Title of ALBERT PARDINI

```
data[data['EmployeeName'] == 'ALBERT PARDINI']['JobTitle']
```

```
2    CAPTAIN III (POLICE DEPARTMENT)
Name: JobTitle, dtype: object
```

How Much ALBERT PARDINI Make (Include Benefits)?

How Much ALBERT PARDINI Make (Include Benefits)?

```
data[data['EmployeeName'] == 'ALBERT PARDINI']['TotalPayBenefits']
```

```
2    335279.91
Name: TotalPayBenefits, dtype: float64
```

Display Name of The Person Having The Highest BasePay

```
data[data['BasePay'].max() == data['BasePay']] ['EmployeeName']
```

```
36160    Gregory Suhr
Name: EmployeeName, dtype: object
```

Find Average BasePay of Employee Having Job Title ACCOUNTANT

```
data[data['JobTitle'] == 'ACCOUNTANT']['BasePay'].mean()
```

```
46643.172
```

Find Top 5 Most Common Jobs

```
data['JobTitle'].value_counts().head()
```

```
JobTitle
TRANSIT OPERATOR    2388
SPECIAL NURSE       1402
REGISTERED NURSE    1219
CUSTODIAN            796
FIREFIGHTER          794
Name: count, dtype: int64
```