

```
import numpy as np
import pandas as pd
```

```
from google.colab import files
uploaded = files.upload(.,)
```

Wybierz pliki Salaries.csv

- **Salaries.csv**(application/vnd.ms-excel) - 16239776 bytes, last modified: 8.01.2019 - 100% done  
Saving Salaries.csv to Salaries.csv

```
#Read Salaries.csv as a dataframe called sal.
sal = pd.read_csv('Salaries.csv')
```

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

```
↳
```

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.00	400184.25	NaN
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE	77916.00	56120.71	198306.90	NaN

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

```
↳
```

```
#What is the average BasePay?
```

```
#What is the average salary?
sal.BasePay.mean()
```

```
↳ 66325.4488404877
      JobTitle      148654 non-null object
```

```
#What is the highest amount of OvertimePay in the dataset ?
sal.OvertimePay.max()
```

```
↳ 245131.88
      TotalPayBenefits      148654 non-null float64
```

```
#What is the job title of JOSEPH DRISCOLL ?
(sal.loc[sal['EmployeeName'] == 'JOSEPH DRISCOLL']).JobTitle
```

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

```
#How much does JOSEPH DRISCOLL make (including benefits)?
(sal.loc[sal['EmployeeName'] == 'JOSEPH DRISCOLL']).TotalPayBenefits
```

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

```
#What is the name of highest paid person (including benefits)?
sal.loc[(sal.TotalPayBenefits)==max(sal.TotalPayBenefits)]
```

```
↳
```

		Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPayBenefits
0	1		NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT	167411.18	0.0	400184.25	NaN	567595.43

```
#What is the name of lowest paid person (including benefits)? Do you notice something strange?
sal.loc[(sal.TotalPayBenefits)==min(sal.TotalPayBenefits)]
#Yes, her salary is negative
```

```
↳
```

		Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPayBenefits
148653	148654	Joe Lopez	Counselor, Log Cabin Ranch	0.0	0.0	-618.13	0.0	-618.13	

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

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

```
#How many unique job titles are there?
sal['JobTitle'].nunique()
```

```
↳ 3150
#What are the top 5 most common job titles?
```

```
#what are the top 5 most common jobs?
sal['JobTitle'].value_counts().head()
```

```
↳ 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 on
sum(sal[sal['Year']==2013]['JobTitle'].value_counts()==1).
```

```
↳ 202
```

```
#How many people have the word Chief in their job title? (This is pretty tricky)
sal['JobTitle'].apply(lambda str:('chief' in str.lower())).sum()
```

```
↳ 627
```

```
#Is there a correlation between length of the Job Title string and Salary?
sal['title_len'] = sal["JobTitle"].str.len()
print(sal["title_len"].corr(sal["TotalPayBenefits"])).
```

```
↳ -0.036878445932606675
```

```
#Generate a histogram plot of base salary with 20 bins?
import matplotlib.pyplot as plt
sal_to_hist=sal['BasePay'].dropna()
sal_to_hist
plt.hist(sal_to_hist, bins=20, rwidth=0.8, alpha=0.8)
```

```
↳
```

```
(array([2.5240e+04, 1.0758e+04, 1.0664e+04, 2.4549e+04, 2.4469e+04,
        1.6672e+04, 1.2000e+04, 1.1510e+04, 5.0020e+03, 1.0260e+03,
```

```
1.0075e+04, 1.2990e+04, 1.1519e+04, 3.9020e+03, 1.9200e+03,  
1.3770e+03, 1.1630e+03, 5.0700e+02, 1.4400e+02, 6.8000e+01,  
3.1000e+01, 3.4000e+01, 1.4000e+01, 7.0000e+00, 1.0000e+01]),  
array([-1.66010000e+02,  1.58060410e+04,  3.17780920e+04,  4.77501430e+04,  
        6.37221940e+04,  7.96942450e+04,  9.56662960e+04,  1.11638347e+05,  
        1.27610398e+05,  1.43582449e+05,  1.59554500e+05,  1.75526551e+05,  
        1.91498602e+05,  2.07470653e+05,  2.23442704e+05,  2.39414755e+05,  
        2.55386806e+05,  2.71358857e+05,  2.87330908e+05,  3.03302959e+05,  
        3.19275010e+05]),  
<a list of 20 Patch objects>)
```

