

```
#To find the salary of employes using dataset
```

```
#step 1:  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
import pandas as pd  
from google.colab import files  
import io  
  
try:  
    df = pd.read_csv("salary_data.csv")  
    print("File loaded successfully from local environment.")  
  
except FileNotFoundError:  
    print("salary_data.csv not found. Please upload the file.")  
    uploaded = files.upload()  
  
if 'salary_data.csv' in uploaded:  
    df = pd.read_csv(io.BytesIO(uploaded['salary_data.csv']))  
    print("File uploaded and loaded successfully.")  
else:  
    print("Upload failed: salary_data.csv not found in uploaded files.")
```

salary_data.csv not found. Please upload the file.

No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving salary_data.csv to salary_data.csv
file uploaded and loaded successfully

```
#Step3: Display top first five rows of dataset  
df.head()
```

	Years	Experience	Salary
0		1.1	39343
1		1.3	46205
2		1.5	37731
3		2.0	43525
4		2.2	39891

```
#Step4:#Display last five rows of dataset  
df.tail()
```

	Years	Experience	Salary
25		9.0	105582
26		9.5	116969
27		9.6	112635
28		10.3	122391
29		10.5	121872

```
#Check the total size of dataset  
df.shape
```

```
#Get the information about dataset  
df.info
```

```
#check if any missing value present in dataset  
df.isnull().sum()
```

```
#step8  
#get the statistic information  
df.describe
```

```
#step 9  
#split the data into training and test data  
from sklearn.model_selection import train_test_split  
x_train,x_test,y_train,y_test=train_test
```

```
-----  
NameError Traceback (most recent call last)  
/tmp/ipython-input-3205664784.py in <cell line: 0>()  
      2 #split the data into training and test data  
      3 from sklearn.model_selection import train_test_split  
----> 4 x_train,x_test,y_train,y_test=train_test  
  
NameError: name 'train_test' is not defined
```

```
#step 11:  
#visualize whole data set  
sns.pairplot(df)
```