

```
In [1]: pip install numpy
```

```
Requirement already satisfied: numpy in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (1.26.3)
Note: you may need to restart the kernel to use updated packages.
```

```
In [2]: pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (2.1.4)
Requirement already satisfied: numpy<2,>=1.26.0 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from pandas) (1.26.3)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from pandas) (2023.4)
Requirement already satisfied: six>=1.5 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
In [4]: pip install matplotlib
```

```
Requirement already satisfied: matplotlib in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (3.8.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (4.47.2)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (1.4.5)
Requirement already satisfied: numpy<2,>=1.21 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (1.26.3)
Requirement already satisfied: packaging>=20.0 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (23.2)
Requirement already satisfied: pillow>=8 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
In [5]: pip install seaborn
```

## Collecting seaborn

```
Using cached seaborn-0.13.1-py3-none-any.whl.metadata (5.4 kB)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from seaborn) (1.26.3)
Requirement already satisfied: pandas>=1.2 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from seaborn) (2.1.4)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from seaborn) (3.8.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.47.2)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.5)
Requirement already satisfied: packaging>=20.0 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (23.2)
Requirement already satisfied: pillow>=8 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from pandas>=1.2->seaborn) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from pandas>=1.2->seaborn) (2023.4)
Requirement already satisfied: six>=1.5 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
```

Downloading seaborn-0.13.1-py3-none-any.whl (294 kB)

```
----- 0.0/294.8 kB ? eta -:--:--
----- 10.2/294.8 kB ? eta -:--:--
----- 10.2/294.8 kB ? eta -:--:--
----- 30.7/294.8 kB 325.1 kB/s eta 0:00:01
----- 81.9/294.8 kB 573.4 kB/s eta 0:00:01
----- 286.7/294.8 kB 1.6 MB/s eta 0:00:01
----- 294.8/294.8 kB 1.5 MB/s eta 0:00:00
```

Installing collected packages: seaborn

Successfully installed seaborn-0.13.1

Note: you may need to restart the kernel to use updated packages.

In [6]: pip install sklearn

```
Collecting sklearn
```

```
  Using cached sklearn-0.0.post12.tar.gz (2.6 kB)
```

```
  Installing build dependencies: started
```

```
  Installing build dependencies: finished with status 'done'
```

```
  Getting requirements to build wheel: started
```

```
  Getting requirements to build wheel: finished with status 'error'
```

```
Note: you may need to restart the kernel to use updated packages.
```

```
error: subprocess-exited-with-error
```

```
Getting requirements to build wheel did not run successfully.
```

```
exit code: 1
```

```
[15 lines of output]
```

```
The 'sklearn' PyPI package is deprecated, use 'scikit-learn'  
rather than 'sklearn' for pip commands.
```

```
Here is how to fix this error in the main use cases:
```

- use 'pip install scikit-learn' rather than 'pip install sklearn'
- replace 'sklearn' by 'scikit-learn' in your pip requirements files  
(requirements.txt, setup.py, setup.cfg, Pipfile, etc ...)
- if the 'sklearn' package is used by one of your dependencies,  
it would be great if you take some time to track which package uses  
'sklearn' instead of 'scikit-learn' and report it to their issue tracker
- as a last resort, set the environment variable  
SKLEARN\_ALLOW\_DEPRECATED\_SKLEARN\_PACKAGE\_INSTALL=True to avoid this error

```
More information is available at
```

```
https://github.com/scikit-learn/scikit-learn-pypi-package
```

```
[end of output]
```

```
note: This error originates from a subprocess, and is likely not a problem with pip.
```

```
error: subprocess-exited-with-error
```

```
Getting requirements to build wheel did not run successfully.
```

```
exit code: 1
```

```
See above for output.
```

```
note: This error originates from a subprocess, and is likely not a problem with pip.
```

```
In [ ]:
```

```
In [8]: pip install scikit-learn
```

```
Requirement already satisfied: scikit-learn in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (1.3.2)
```

```
Requirement already satisfied: numpy<2.0,>=1.17.3 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from scikit-learn) (1.26.3)
```

```
Requirement already satisfied: scipy>=1.5.0 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from scikit-learn) (1.11.4)
```

```
Requirement already satisfied: joblib>=1.1.1 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from scikit-learn) (1.3.2)
```

```
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\vlanj\appdata\local\programs\python\python312\lib\site-packages (from scikit-learn) (3.2.0)
```

```
Note: you may need to restart the kernel to use updated packages.
```

```
In [ ]: # Practicle 1 : Preprocessing of dataset  
# Preprocessing for data sciencentist salary dataset  
# 2021BIT023
```

```
In [1]: import pandas as pd  
import matplotlib.pyplot as plt  
import numpy as np
```

```
In [2]: df=pd.read_csv(ds_salaries1.csv)  
df
```

```
NameError Traceback (most recent call last)  
Cell In[2], line 1  
----> 1 df=pd.read_csv(ds_salaries1.csv)  
      2 df  
  
NameError: name 'ds_salaries1' is not defined
```

```
In [3]: df=pd.read_csv("ds_salaries1.csv")  
df
```

Out[3]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency
<b>0</b>	2023	SE	FT	Principal Data Scientist	80000.0	EUR
<b>1</b>	2023	MI	CT	ML Engineer	30000.0	USD
<b>2</b>	2023	MI	CT	ML Engineer	25500.0	USD
<b>3</b>	2023	SE	FT	Data Scientist	175000.0	USD
<b>4</b>	2023	SE	FT	Data Scientist	120000.0	USD
...	...	...	...	...	...	...
<b>3750</b>	2020	SE	FT	Data Scientist	412000.0	USD
<b>3751</b>	2021	MI	FT	Principal Data Scientist	151000.0	USD
<b>3752</b>	2020	EN	FT	Data Scientist	105000.0	USD
<b>3753</b>	2020	EN	CT	Business Data Analyst	100000.0	USD
<b>3754</b>	2021	SE	FT	Data Science Manager	7000000.0	INR

3755 rows × 11 columns



In [4]: df.head(10)

Out[4]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency	sal
0	2023	SE	FT	Principal Data Scientist	80000.0	EUR	
1	2023	MI	CT	ML Engineer	30000.0	USD	
2	2023	MI	CT	ML Engineer	25500.0	USD	
3	2023	SE	FT	Data Scientist	175000.0	USD	
4	2023	SE	FT	Data Scientist	120000.0	USD	
5	2023	SE	FT	Applied Scientist	222200.0	USD	
6	2023	SE	FT	Applied Scientist	136000.0	USD	
7	2023	SE	FT	Data Scientist	219000.0	USD	
8	2023	SE	FT	Data Scientist	141000.0	USD	
9	2023	SE	FT	Data Scientist	147100.0	USD	

◀ ▶

In [5]: `df.tail(10)`

Out[5]:

	<b>work_year</b>	<b>experience_level</b>	<b>employment_type</b>	<b>job_title</b>	<b>salary</b>	<b>salary_currency</b>
<b>3745</b>	2021	SE	FT	Director of Data Science	168000.0	USD
<b>3746</b>	2021	MI	FT	Data Scientist	160000.0	SGD
<b>3747</b>	2021	MI	FT	Applied Machine Learning Scientist	423000.0	USD
<b>3748</b>	2021	MI	FT	Data Engineer	24000.0	EUR
<b>3749</b>	2021	SE	FT	Data Specialist	165000.0	USD
<b>3750</b>	2020	SE	FT	Data Scientist	412000.0	USD
<b>3751</b>	2021	MI	FT	Principal Data Scientist	151000.0	USD
<b>3752</b>	2020	EN	FT	Data Scientist	105000.0	USD
<b>3753</b>	2020	EN	CT	Business Data Analyst	100000.0	USD
<b>3754</b>	2021	SE	FT	Data Science Manager	7000000.0	INR



In [6]: `df.shape # row, attributes`

Out[6]: (3755, 11)

In [7]: `df1=df.copy()`

In [8]: `df1`

Out[8]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency
<b>0</b>	2023	SE	FT	Principal Data Scientist	80000.0	EUR
<b>1</b>	2023	MI	CT	ML Engineer	30000.0	USD
<b>2</b>	2023	MI	CT	ML Engineer	25500.0	USD
<b>3</b>	2023	SE	FT	Data Scientist	175000.0	USD
<b>4</b>	2023	SE	FT	Data Scientist	120000.0	USD
...	...	...	...	...	...	...
<b>3750</b>	2020	SE	FT	Data Scientist	412000.0	USD
<b>3751</b>	2021	MI	FT	Principal Data Scientist	151000.0	USD
<b>3752</b>	2020	EN	FT	Data Scientist	105000.0	USD
<b>3753</b>	2020	EN	CT	Business Data Analyst	100000.0	USD
<b>3754</b>	2021	SE	FT	Data Science Manager	7000000.0	INR

3755 rows × 11 columns

In [9]: `df1.isnull().sum()`

Out[9]:

work_year	0
experience_level	0
employment_type	0
job_title	0
salary	2
salary_currency	1
salary_in_usd	3
employee_residence	0
remote_ratio	1
company_location	1
company_size	0
dtype: int64	

```
In [28]: # voool series
print(df1.isna()) # false for not null data aani true for NAN value i.e for null va
```

	work_year	experience_level	employment_type	job_title	salary	\
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
3750	False	False	False	False	False	False
3751	False	False	False	False	False	False
3752	False	False	False	False	False	False
3753	False	False	False	False	False	False
3754	False	False	False	False	False	False
	salary_currency	salary_in_usd	employee_residence	remote_ratio	\	
0	False	False	False	False	False	
1	False	False	False	False	False	
2	False	False	False	False	False	
3	False	False	False	False	False	
4	False	False	False	False	False	
...	...	...	...	...	...	
3750	False	False	False	False	False	
3751	False	False	False	False	False	
3752	False	False	False	False	False	
3753	False	False	False	False	False	
3754	False	False	False	False	False	
	company_location	company_size				
0	False	False				
1	False	False				
2	False	False				
3	False	False				
4	False	False				
...	...	...				
3750	False	False				
3751	False	False				
3752	False	False				
3753	False	False				
3754	False	False				

[3751 rows x 11 columns]

```
In [11]: df1.dropna(inplace=True)
```

```
In [12]: df1.isnull().sum()
```

```
Out[12]: work_year      0  
experience_level      0  
employment_type        0  
job_title              0  
salary                 0  
salary_currency         0  
salary_in_usd          0  
employee_residence      0  
remote_ratio            0  
company_location         0  
company_size             0  
dtype: int64
```

```
In [14]: # in salary_currency we replace salary from INR to USD  
df_repl=df1.replace({'INR':'USD'})  
df_repl
```

Out[14]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency
<b>0</b>	2023	SE	FT	Principal Data Scientist	80000.0	EUR
<b>1</b>	2023	MI	CT	ML Engineer	30000.0	USD
<b>2</b>	2023	MI	CT	ML Engineer	25500.0	USD
<b>3</b>	2023	SE	FT	Data Scientist	175000.0	USD
<b>4</b>	2023	SE	FT	Data Scientist	120000.0	USD
...	...	...	...	...	...	...
<b>3750</b>	2020	SE	FT	Data Scientist	412000.0	USD
<b>3751</b>	2021	MI	FT	Principal Data Scientist	151000.0	USD
<b>3752</b>	2020	EN	FT	Data Scientist	105000.0	USD
<b>3753</b>	2020	EN	CT	Business Data Analyst	100000.0	USD
<b>3754</b>	2021	SE	FT	Data Science Manager	7000000.0	USD

3751 rows × 11 columns



In [15]:

```
df_intrplt=df1.interpolate()
df_intrplt
```

C:\Users\vlanj\AppData\Local\Temp\ipykernel\_2636\455564773.py:1: FutureWarning: DataFrame.interpolate with object dtype is deprecated and will raise in a future version. Call obj.infer\_objects(copy=False) before interpolating instead.
df\_intrplt=df1.interpolate()

Out[15]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency
<b>0</b>	2023	SE	FT	Principal Data Scientist	80000.0	EUR
<b>1</b>	2023	MI	CT	ML Engineer	30000.0	USD
<b>2</b>	2023	MI	CT	ML Engineer	25500.0	USD
<b>3</b>	2023	SE	FT	Data Scientist	175000.0	USD
<b>4</b>	2023	SE	FT	Data Scientist	120000.0	USD
...	...	...	...	...	...	...
<b>3750</b>	2020	SE	FT	Data Scientist	412000.0	USD
<b>3751</b>	2021	MI	FT	Principal Data Scientist	151000.0	USD
<b>3752</b>	2020	EN	FT	Data Scientist	105000.0	USD
<b>3753</b>	2020	EN	CT	Business Data Analyst	100000.0	USD
<b>3754</b>	2021	SE	FT	Data Science Manager	7000000.0	INR

3751 rows × 11 columns



In [ ]:

In [18]: df1=df1.infer\_objects()

```
-----  
AttributeError                                 Traceback (most recent call last)  
~\AppData\Local\Temp\ipykernel_2636\2477449844.py in ?()  
----> 1 df1=df1.infer_object()  
  
~\AppData\Local\Programs\Python\Python312\Lib\site-packages\pandas\core\generic.py in ?(self, name)  
    6200             and name not in self._accessors  
    6201             and self._info_axis._can_hold_identifiers_and_holds_name(name)  
    6202         ):  
    6203             return self[name]  
-> 6204         return object.__getattribute__(self, name)  
  
AttributeError: 'DataFrame' object has no attribute 'infer_object'
```

In [19]: df2=df1.infer\_objects()

In [20]: df2

Out[20]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency
<b>0</b>	2023	SE	FT	Principal Data Scientist	80000.0	EUR
<b>1</b>	2023	MI	CT	ML Engineer	30000.0	USD
<b>2</b>	2023	MI	CT	ML Engineer	25500.0	USD
<b>3</b>	2023	SE	FT	Data Scientist	175000.0	USD
<b>4</b>	2023	SE	FT	Data Scientist	120000.0	USD
...	...	...	...	...	...	...
<b>3750</b>	2020	SE	FT	Data Scientist	412000.0	USD
<b>3751</b>	2021	MI	FT	Principal Data Scientist	151000.0	USD
<b>3752</b>	2020	EN	FT	Data Scientist	105000.0	USD
<b>3753</b>	2020	EN	CT	Business Data Analyst	100000.0	USD
<b>3754</b>	2021	SE	FT	Data Science Manager	7000000.0	INR

3751 rows × 11 columns



In [21]:

```
df1=df1.interpolate()
df1
```

C:\Users\vlanj\AppData\Local\Temp\ipykernel\_2636\1327046382.py:1: FutureWarning: DataFrame.interpolate with object dtype is deprecated and will raise in a future version. Call obj.infer\_objects(copy=False) before interpolating instead.
df1=df1.interpolate()

Out[21]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency
<b>0</b>	2023	SE	FT	Principal Data Scientist	80000.0	EUR
<b>1</b>	2023	MI	CT	ML Engineer	30000.0	USD
<b>2</b>	2023	MI	CT	ML Engineer	25500.0	USD
<b>3</b>	2023	SE	FT	Data Scientist	175000.0	USD
<b>4</b>	2023	SE	FT	Data Scientist	120000.0	USD
...	...	...	...	...	...	...
<b>3750</b>	2020	SE	FT	Data Scientist	412000.0	USD
<b>3751</b>	2021	MI	FT	Principal Data Scientist	151000.0	USD
<b>3752</b>	2020	EN	FT	Data Scientist	105000.0	USD
<b>3753</b>	2020	EN	CT	Business Data Analyst	100000.0	USD
<b>3754</b>	2021	SE	FT	Data Science Manager	7000000.0	INR

3751 rows × 11 columns

In [24]: `print(df1.notnull()) # true for not missing value , false for missing value`

```

      work_year  experience_level  employment_type  job_title  salary  \
0           True                True             True       True    True
1           True                True             True       True    True
2           True                True             True       True    True
3           True                True             True       True    True
4           True                True             True       True    True
...
3750        ...                 ...              ...        ...    ...
3751        True                True             True       True    True
3752        True                True             True       True    True
3753        True                True             True       True    True
3754        True                True             True       True    True

      salary_currency  salary_in_usd  employee_residence  remote_ratio  \
0            True          True            True            True
1            True          True            True            True
2            True          True            True            True
3            True          True            True            True
4            True          True            True            True
...
3750        ...                 ...              ...        ...
3751        True          True            True            True
3752        True          True            True            True
3753        True          True            True            True
3754        True          True            True            True

      company_location  company_size
0            True          True
1            True          True
2            True          True
3            True          True
4            True          True
...
3750        ...                 ...
3751        True          True
3752        True          True
3753        True          True
3754        True          True

```

[3751 rows x 11 columns]

In [25]: `df1.notnull().sum()`

Out[25]:

work_year	3751
experience_level	3751
employment_type	3751
job_title	3751
salary	3751
salary_currency	3751
salary_in_usd	3751
employee_residence	3751
remote_ratio	3751
company_location	3751
company_size	3751

dtype: int64

```
In [26]: df2=df.copy()
```

```
In [27]: df2.fillna=df2.fillna(0)
df2.fillna
```

Out[27]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency
<b>0</b>	2023	SE	FT	Principal Data Scientist	80000.0	EUR
<b>1</b>	2023	MI	CT	ML Engineer	30000.0	USD
<b>2</b>	2023	MI	CT	ML Engineer	25500.0	USD
<b>3</b>	2023	SE	FT	Data Scientist	175000.0	USD
<b>4</b>	2023	SE	FT	Data Scientist	120000.0	USD
...	...	...	...	...	...	...
<b>3750</b>	2020	SE	FT	Data Scientist	412000.0	USD
<b>3751</b>	2021	MI	FT	Principal Data Scientist	151000.0	USD
<b>3752</b>	2020	EN	FT	Data Scientist	105000.0	USD
<b>3753</b>	2020	EN	CT	Business Data Analyst	100000.0	USD
<b>3754</b>	2021	SE	FT	Data Science Manager	7000000.0	INR

3755 rows × 11 columns



```
In [33]: #filter data
filter_data=df2[df['salary']>=81000]
filter_data
```

Out[33]:

	<b>work_year</b>	<b>experience_level</b>	<b>employment_type</b>	<b>job_title</b>	<b>salary</b>	<b>salary_currency</b>
<b>3</b>	2023	SE	FT	Data Scientist	175000.0	USD
<b>4</b>	2023	SE	FT	Data Scientist	120000.0	USD
<b>5</b>	2023	SE	FT	Applied Scientist	222200.0	USD
<b>6</b>	2023	SE	FT	Applied Scientist	136000.0	USD
<b>7</b>	2023	SE	FT	Data Scientist	219000.0	USD
...	...	...	...	...	...	...
<b>3750</b>	2020	SE	FT	Data Scientist	412000.0	USD
<b>3751</b>	2021	MI	FT	Principal Data Scientist	151000.0	USD
<b>3752</b>	2020	EN	FT	Data Scientist	105000.0	USD
<b>3753</b>	2020	EN	CT	Business Data Analyst	100000.0	USD
<b>3754</b>	2021	SE	FT	Data Science Manager	7000000.0	INR

3082 rows × 11 columns



In [37]:

```
# dict2lst=df2.to_dict(orient='list')
# print(dict2lst)
```

In [36]:

```
df2
```

Out[36]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency
<b>0</b>	2023	SE	FT	Principal Data Scientist	80000.0	EUR
<b>1</b>	2023	MI	CT	ML Engineer	30000.0	USD
<b>2</b>	2023	MI	CT	ML Engineer	25500.0	USD
<b>3</b>	2023	SE	FT	Data Scientist	175000.0	USD
<b>4</b>	2023	SE	FT	Data Scientist	120000.0	USD
...	...	...	...	...	...	...
<b>3750</b>	2020	SE	FT	Data Scientist	412000.0	USD
<b>3751</b>	2021	MI	FT	Principal Data Scientist	151000.0	USD
<b>3752</b>	2020	EN	FT	Data Scientist	105000.0	USD
<b>3753</b>	2020	EN	CT	Business Data Analyst	100000.0	USD
<b>3754</b>	2021	SE	FT	Data Science Manager	7000000.0	INR

3755 rows × 11 columns



In [ ]: