In [3]:
```python
import pandas as pd
import numpy as np
```

In [6]:
```python
df = pd.read_csv('./Dataset.csv')
print(df)
```

```
      work_year                 job_title                     job_category  \
0          2023      Data DevOps Engineer                 Data Engineering
1          2023            Data Architect   Data Architecture and Modeling
2          2023            Data Architect   Data Architecture and Modeling
3          2023            Data Scientist         Data Science and Research
4          2023            Data Scientist         Data Science and Research
...         ...                       ...                              ...
9350       2021          Data Specialist    Data Management and Strategy
9351       2020            Data Scientist         Data Science and Research
9352       2021  Principal Data Scientist         Data Science and Research
9353       2020            Data Scientist         Data Science and Research
9354       2020      Business Data Analyst                    Data Analysis

     salary_currency  salary  salary_in_usd employee_residence  \
0                EUR   88000          95012            Germany
1                USD  186000         186000      United States
2                USD   81800          81800      United States
3                USD  212000         212000      United States
4                USD   93300          93300      United States
...              ...     ...            ...                ...
9350             USD  165000         165000      United States
9351             USD  412000         412000      United States
9352             USD  151000         151000      United States
9353             USD  105000         105000      United States
9354             USD  100000         100000      United States

     experience_level employment_type work_setting company_location  \
0           Mid-level       Full-time       Hybrid          Germany
1              Senior       Full-time    In-person    United States
2              Senior       Full-time    In-person    United States
3              Senior       Full-time    In-person    United States
4              Senior       Full-time    In-person    United States
...               ...             ...          ...              ...
9350           Senior       Full-time       Remote    United States
9351           Senior       Full-time       Remote    United States
9352        Mid-level       Full-time       Remote    United States
9353        Entry-level      Full-time       Remote    United States
9354        Entry-level       Contract       Remote    United States

     company_size
0               L
1               M
2               M
3               M
4               M
...           ...
9350            L
9351            L
9352            L
9353            S
9354            L

[9355 rows x 12 columns]
```

```
In [7]: print(df.head(5))
```

```
   work_year             job_title                        job_category  \
0        2023  Data DevOps Engineer                     Data Engineering
1        2023        Data Architect   Data Architecture and Modeling
2        2023        Data Architect   Data Architecture and Modeling
3        2023        Data Scientist        Data Science and Research
4        2023        Data Scientist        Data Science and Research

   salary_currency   salary   salary_in_usd employee_residence experience_level  \
0              EUR    88000           95012            Germany        Mid-level
1              USD   186000          186000      United States           Senior
2              USD    81800           81800      United States           Senior
3              USD   212000          212000      United States           Senior
4              USD    93300           93300      United States           Senior

   employment_type work_setting company_location company_size
0        Full-time       Hybrid          Germany            L
1        Full-time    In-person    United States            M
2        Full-time    In-person    United States            M
3        Full-time    In-person    United States            M
4        Full-time    In-person    United States            M
```

```
In [9]: print(df.isnull())
```

	work_year	job_title	job_category	salary_currency	salary \
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
...	...	...	...	...	...
9350	False	False	False	False	False
9351	False	False	False	False	False
9352	False	False	False	False	False
9353	False	False	False	False	False
9354	False	False	False	False	False

	salary_in_usd	employee_residence	experience_level	employment_type \
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...	...	...	...	...
9350	False	False	False	False
9351	False	False	False	False
9352	False	False	False	False
9353	False	False	False	False
9354	False	False	False	False

	work_setting	company_location	company_size
0	False	False	False
1	False	False	False
2	False	False	False
3	False	False	False
4	False	False	False
...	...	...	...
9350	False	False	False
9351	False	False	False
9352	False	False	False
9353	False	False	False
9354	False	False	False

[9355 rows x 12 columns]

In [10]:
```python
print(df.describe())
```

	work_year	salary	salary_in_usd
count	9355.000000	9355.000000	9355.000000
mean	2022.760449	149927.981293	150299.495564
std	0.519470	63608.835387	63177.372024
min	2020.000000	14000.000000	15000.000000
25%	2023.000000	105200.000000	105700.000000
50%	2023.000000	143860.000000	143000.000000
75%	2023.000000	187000.000000	186723.000000
max	2023.000000	450000.000000	450000.000000

In [28]:
```python
size = df.size
shape = df.shape
ndim = df.ndim


print("Size = {} \nShape = {}".format(size, shape) )

print("Ndim of Dataframe = {}  ".format(ndim))
```

```
            Size = 112260
            Shape = (9355, 12)
            Ndim of Dataframe = 2
```
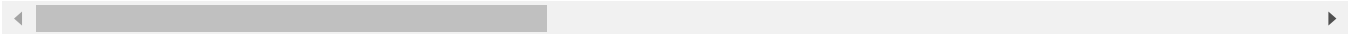
In [29]:
```python
dts =df.dtypes
print("Type of Dataframe = {}  ".format(dts))
```

```
Type of Dataframe = work_year                    int64
job_title              object
job_category           object
salary_currency        object
salary                  int64
salary_in_usd           int64
employee_residence     object
experience_level       object
employment_type        object
work_setting           object
company_location       object
company_size           object
dtype: object
```

In [32]:
```python
dummies = pd.get_dummies(df.job_title)

merged = pd.concat([df, dummies] , axis = 'columns')

merged.drop(['job_title'] , axis = 'columns')
```

Out[32]:

	work_year	job_category	salary_currency	salary	salary_in_usd	employee_residence	experi
0	2023	Data Engineering	EUR	88000	95012	Germany	
1	2023	Data Architecture and Modeling	USD	186000	186000	United States	
2	2023	Data Architecture and Modeling	USD	81800	81800	United States	
3	2023	Data Science and Research	USD	212000	212000	United States	
4	2023	Data Science and Research	USD	93300	93300	United States	
...	...	...	...	...	...	...	
9350	2021	Data Management and Strategy	USD	165000	165000	United States	
9351	2020	Data Science and Research	USD	412000	412000	United States	
9352	2021	Data Science and Research	USD	151000	151000	United States	
9353	2020	Data Science and Research	USD	105000	105000	United States	
9354	2020	Data Analysis	USD	100000	100000	United States	

9355 rows × 136 columns

In [33]: `print(merged)`

```
       work_year                 job_title                      job_category  \
0           2023         Data DevOps Engineer                  Data Engineering
1           2023               Data Architect   Data Architecture and Modeling
2           2023               Data Architect   Data Architecture and Modeling
3           2023               Data Scientist        Data Science and Research
4           2023               Data Scientist        Data Science and Research
...          ...                        ...                               ...
9350        2021              Data Specialist    Data Management and Strategy
9351        2020               Data Scientist        Data Science and Research
9352        2021     Principal Data Scientist        Data Science and Research
9353        2020               Data Scientist        Data Science and Research
9354        2020          Business Data Analyst                    Data Analysis

     salary_currency   salary   salary_in_usd employee_residence  \
0                EUR    88000            95012            Germany
1                USD   186000           186000      United States
2                USD    81800            81800      United States
3                USD   212000           212000      United States
4                USD    93300            93300      United States
...              ...      ...              ...                ...
9350             USD   165000           165000      United States
9351             USD   412000           412000      United States
9352             USD   151000           151000      United States
9353             USD   105000           105000      United States
9354             USD   100000           100000      United States

     experience_level employment_type work_setting  ...  \
0           Mid-level       Full-time       Hybrid  ...
1              Senior       Full-time    In-person  ...
2              Senior       Full-time    In-person  ...
3              Senior       Full-time    In-person  ...
4              Senior       Full-time    In-person  ...
...               ...             ...          ...  ...
9350           Senior       Full-time       Remote  ...
9351           Senior       Full-time       Remote  ...
9352        Mid-level       Full-time       Remote  ...
9353      Entry-level       Full-time       Remote  ...
9354      Entry-level        Contract       Remote  ...

     Principal Machine Learning Engineer Product Data Analyst  \
0                                      0                     0
1                                      0                     0
2                                      0                     0
3                                      0                     0
4                                      0                     0
...                                  ...                   ...
9350                                   0                     0
9351                                   0                     0
9352                                   0                     0
9353                                   0                     0
9354                                   0                     0

     Research Analyst  Research Engineer  Research Scientist  \
0                   0                  0                   0
1                   0                  0                   0
2                   0                  0                   0
3                   0                  0                   0
4                   0                  0                   0
...               ...                ...                 ...
9350                0                  0                   0
9351                0                  0                   0
9352                0                  0                   0
9353                0                  0                   0
9354                0                  0                   0
```