```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt

saldf = pd.read_csv('/content/Salary Data.csv')
saldf.head()
```

```
{"summary":"{\n  \"name\": \"saldf\",\n  \"rows\": 375,\n  \"fields\":
[\n    {\n      \"column\": \"Age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 7.069072938567496,\n
\"min\": 23.0,\n        \"max\": 53.0,\n        \"num_unique_values\":
31,\n        \"samples\": [\n          46.0,\n          33.0,\n
37.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Gender\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          \"Female\",\n          \"Male\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Education Level\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 3,\n        \"samples\": [\n
\"Bachelor's\",\n          \"Master's\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Job Title\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 174,\n        \"samples\": [\n
\"Junior Advertising Coordinator\",\n          \"Junior Product
Manager\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Years of Experience\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 6.557007136414243,\n        \"min\":
0.0,\n        \"max\": 25.0,\n        \"num_unique_values\": 28,\n
\"samples\": [\n          10.0,\n          24.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Salary\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": 48240.0134818827,\
n        \"min\": 350.0,\n        \"max\": 250000.0,\n
\"num_unique_values\": 36,\n        \"samples\": [\n          350.0,\n
40000.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"saldf"}
```

```
saldf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 375 entries, 0 to 374
Data columns (total 6 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
```

```
 0   Age                  373 non-null    float64
 1   Gender               373 non-null    object
 2   Education Level      373 non-null    object
 3   Job Title            373 non-null    object
 4   Years of Experience  373 non-null    float64
 5   Salary               373 non-null    float64
dtypes: float64(3), object(3)
memory usage: 17.7+ KB
```

```
saldf.isnull().sum()
```

```
Age                    2
Gender                 2
Education Level        2
Job Title              2
Years of Experience    2
Salary                 2
dtype: int64
```

```
inp = saldf[['Years of Experience']]
out = saldf['Salary']

LR = LinearRegression()

df_cleaned = saldf.dropna(subset=['Years of Experience', 'Salary'])

inp_cleaned = df_cleaned[['Years of Experience']]
out_cleaned = df_cleaned['Salary']

LR.fit(inp_cleaned, out_cleaned)
```

```
LinearRegression()
```

```
LR.predict([[5]])
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/utils/
validation.py:2739: UserWarning: X does not have valid feature names,
but LinearRegression was fitted with feature names
  warnings.warn(

array([66143.76948947])
```