Salary Prediction with Machine Learning

For salary prediction, we need to find relationships in the data on how the salary is determined. For this task, we need to have a dataset based on salaries. I found a dataset that contains data about how job experience affects salary.

```
import pandas as pd
import numpy as np
import plotly.express as px
import plotly.graph_objects as go

data = pd.read_csv("/content/Salary_Data.csv")
print(data.head())
```

```
   YearsExperience   Salary
0              1.1  39343.0
1              1.3  46205.0
2              1.5  37731.0
3              2.0  43525.0
4              2.2  39891.0
```

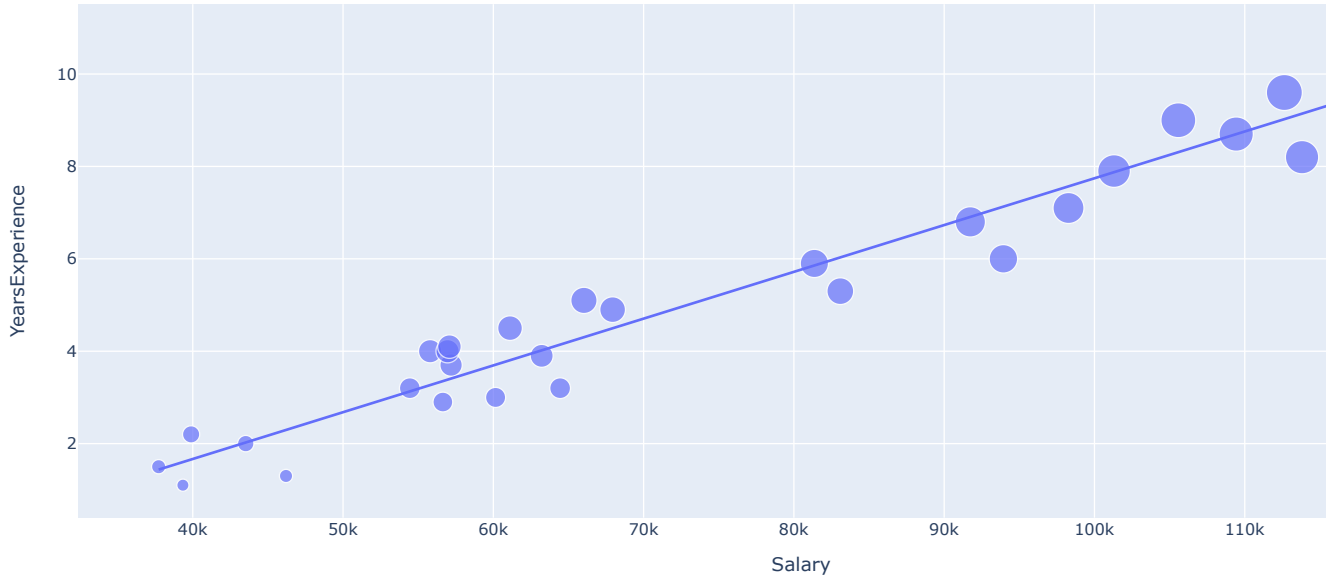
Let's check if the dataset has any null values or not:

```
print(data.isnull().sum())
```

```
YearsExperience    0
Salary             0
dtype: int64
```

The dataset doesn't have any null values. Let's have a look at the relationship between the salary and job experience of the people:

```
figure = px.scatter(data_frame = data,
                    x="Salary",
                    y="YearsExperience",
                    size="YearsExperience",
                    trendline="ols")
figure.show()
```



There is a perfect linear relationship between the salary and the job experience of the people. It means more job experience results in a higher salary.

Training a Machine Learning Model

As this is a regression analysis problem, we will train a regression model to predict salary with Machine Learning. Here's how we can split the data into training and test sets before training the model:

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression

x = np.asanyarray(data[["YearsExperience"]])
y = np.asanyarray(data[["Salary"]])
xtrain, xtest, ytrain, ytest = train_test_split(x, y,
                                                test_size=0.2,
                                                random_state=42)
```

Now here's how we can train the Machine Learning model:

```
model = LinearRegression()
model.fit(xtrain, ytrain)
```

```
▾ LinearRegression
LinearRegression()
```

Now let's predict the salary of a person using the trained Machine Learning model:

```
a = float(input("Years of Experience : "))
features = np.array([[a]])
print("Predicted Salary = ", model.predict(features))
```

```
Years of Experience : 6
Predicted Salary =  [[81864.47494996]]
```

Summary

Salary prediction is a popular problem among the Data Science community for complete beginners. Through this regression analysis, we found a perfect linear relationship between the salary and the job experience of the people. It means more job experience results in a higher salary. I hope you liked this article on the task of salary prediction with Machine Learning using Python.