

# ML\_02

March 15, 2022

## 1 Multiple Linear Regression

- More than two Variables
- One Dependent and more than one are independent variables

### Step-1 Import data or data loading

```
[ ]: import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
```

```
[ ]: # Importing Data
df = pd.read_csv("salary1.csv")
df.head()
```

```
[ ]:      age  distance  YearsExperience  Salary
0   31.1      77.75           1.1    39343
1   31.3      78.25           1.3    46205
2   31.5      78.75           1.5    37731
3   32.0      80.00           2.0    43525
4   32.2      80.50           2.2    39891
```

```
[ ]: X = df[['age', 'distance', 'YearsExperience']]
y = df["Salary"]
```

```
[ ]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2,
↳random_state=0)
a = X_train['age']
b = X_train['distance']
c = X_train['YearsExperience']
```

```
[ ]: model = LinearRegression()
model.fit(X_train,y_train)
model
```

```
[ ]: LinearRegression()
```

```
[ ]: model.coef_
```

```
[ ]: array([-7.08794422e+15, -1.35296027e+14,  7.42618429e+15])
```

```
[ ]: model.intercept_
```

```
[ ]: 2.2278552872528128e+17
```

```
[ ]: model.predict([[31,77,1.1]])
```

```
C:\Users\Sartaj\AppData\Local\Programs\Python\Python39\lib\site-  
packages\sklearn\base.py:450: UserWarning: X does not have valid feature names,  
but LinearRegression was fitted with feature names  
warnings.warn(
```

```
[ ]: array([8.10266443e+14])
```

```
[ ]: model.score(X_train,y_train)
```

[ ]: 0.9411807333463256

```
[ ]: model.score(X_test,y_test)
```

[ ]: 0.9882301655447226

```
[ ]: r = model.predict(X_train)
      r
```

```
[ ]: array([116160., 64032., 76160., 100352., 53792., 74240., 56576.,
          68704., 103104., 90112., 38880., 124544., 54720., 47264.,
          81728., 82656., 61248., 56576., 110560., 45408., 37024.,
          92864., 72416., 64032.] )
```

```
[ ]: # Visualization
import matplotlib.pyplot as plt
plt.scatter(a,y_train)
plt.scatter(b,y_train)
plt.scatter(c,y_train)
plt.plot(X_train,r,color='red')
plt.xlabel("Age || Distance || Years of Experience")
plt.ylabel("Salary")
plt.title("Salary visualization ")
```

```
[ ]: Text(0.5, 1.0, 'Salary visualization ')
```



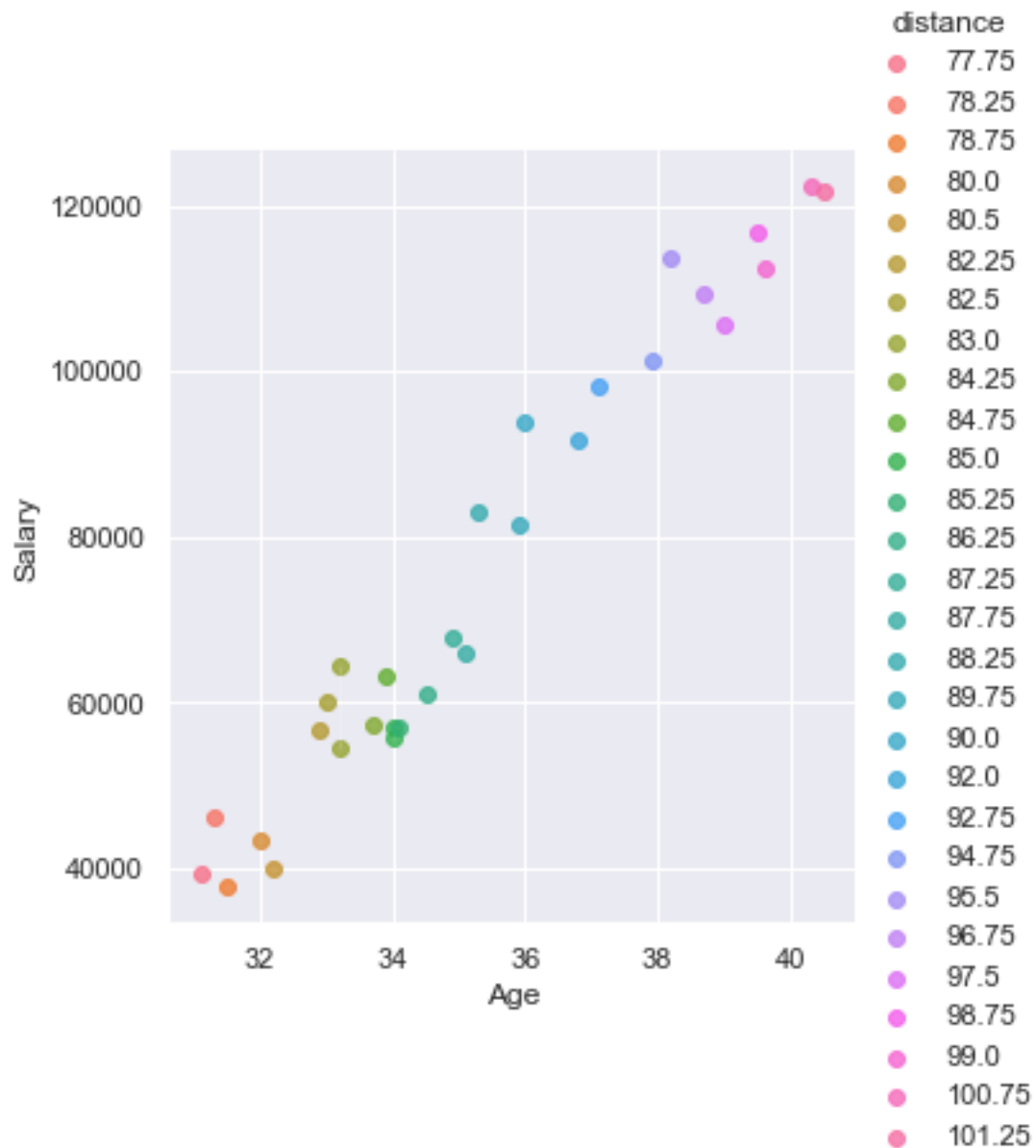
```
[ ]: # Some random Visualization of the data
# Nothing to do with the multiple regression

import seaborn as sns
sns.set_theme()

penguins = pd.read_csv("salary1.csv")

g = sns.lmplot(
    data=penguins,
    x="age", y="Salary", hue="distance",
    height=5
)
g.set_axis_labels("Age", "Salary")
```

```
[ ]: <seaborn.axisgrid.FacetGrid at 0x2288e4edb20>
```



```
[ ]: import seaborn as sns
sns.set_theme()

penguins = pd.read_csv("salary1.csv")

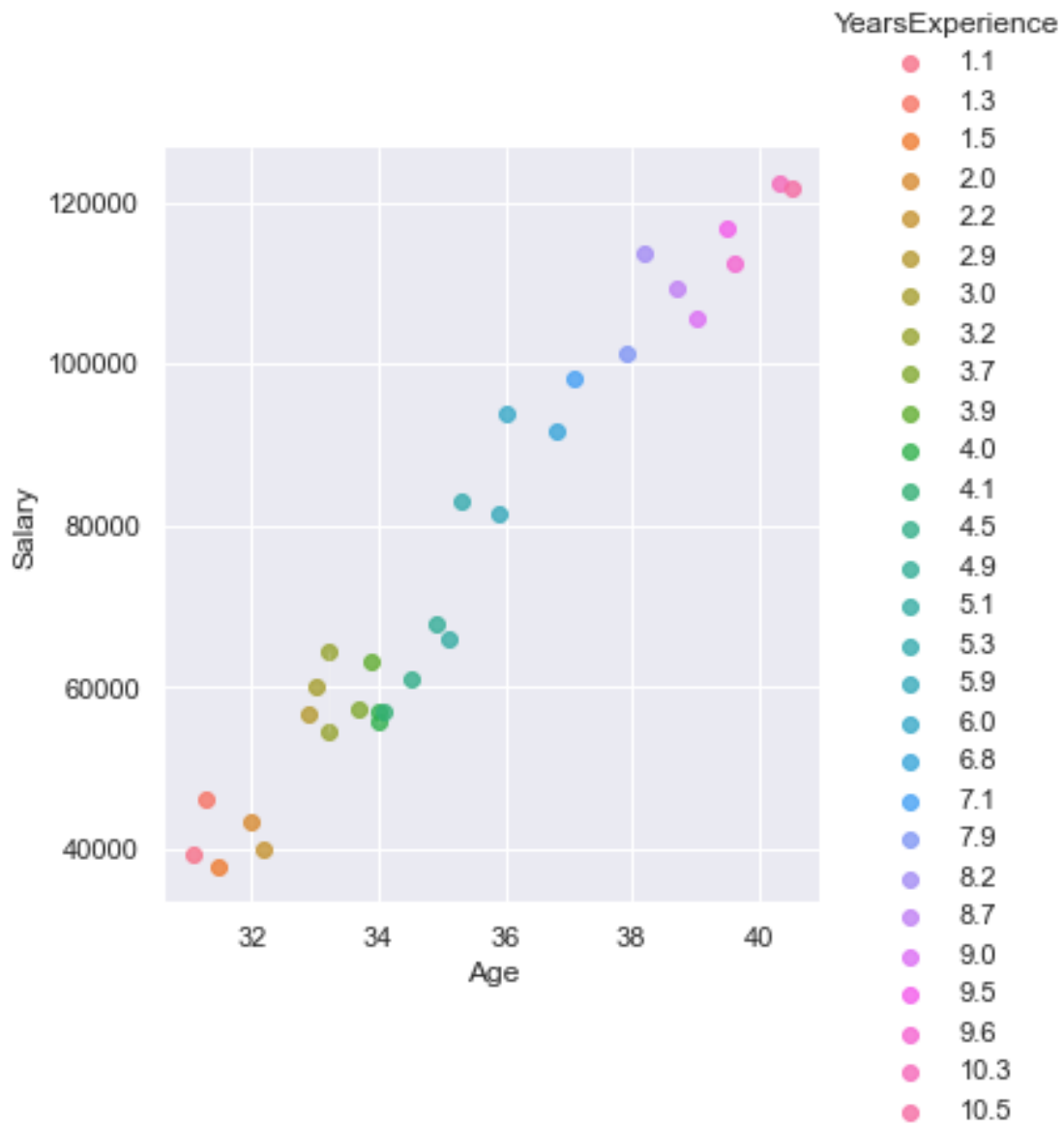
g = sns.lmplot(
    data=penguins,
```

```

    x="age", y="Salary", hue="YearsExperience",
    height=5
)
g.set_axis_labels("Age", "Salary")

```

```
[ ]: <seaborn.axisgrid.FacetGrid at 0x22893f28880>
```



```
[ ]:
```