

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

dataset = pd.read_csv(r'C:\Users\ammyg\Desktop\DATA SCIENCE\clasroom material\16 dec\17th- Slr wit

x = dataset.iloc[:, :-1]
y = dataset.iloc[:, -1]

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)

# random state ke baare mein baad mein batayenge

from sklearn.linear_model import LinearRegression

model = LinearRegression()
model.fit(x_train,y_train)
#prediction
y_pred = model.predict(x_test)
#comparison
comparison = pd.DataFrame({'actual':y_test,'predicted':y_pred})
print(comparison)

plt.scatter(x_test, y_test,color='red')
plt.plot(x_train,model.predict(x_train),color='blue')
plt.title('Salary of employee based on experience')
plt.xlabel('Experience')
plt.ylabel('salary')
#-----
#similarly main kartahu
plt.scatter(x_test, y_test, color = 'green')
plt.plot(x_test,0.5*model.predict(x_test),color = 'black')
#done
# validation and future data

c_inter = model.intercept_
print(f'intercept is: {c_inter}')

slope = model.coef_
print(f'slope is:{slope}')

y_12 = slope*12 + c_inter
print('salary predicton for 12 years exp:',y_12)

#lets implement stats to this model

dataset.mean()
dataset['Salary'].mean()
dataset['YearsExperience'].mean()
dataset.median()
dataset['Salary'].median()
dataset['YearsExperience'].median()
#variance
dataset.var()
dataset['Salary'].var()

```

```

dataset['YearsExperience'].var()
# standard deviation
dataset.std()
dataset['Salary'].std()
dataset['YearsExperience'].std()

print(pd.DataFrame(dataset.values))
from scipy.stats import variation
variation(dataset.values)
variation(dataset['Salary'])
variation(dataset['YearsExperience'])

dataset.corr()

dataset['Salary'].corr(dataset['YearsExperience'])
dataset['Salary'].corr(dataset['Salary'])

dataset.skew()

dataset.sem()

from scipy.stats import stats
dataset.apply(stats.zscore)

from scipy.stats import zscore
print(stats.zscore(dataset))
#stats.zscore(dataset['Salary'])
#stats.zscore(dataset['YearsExperience'])
#stats.zscore(dataset)

#-----
#ANOVA

# ANOVA

y_mean=np.mean(y)
SSR = np.sum((y_pred-y_mean)**2)
print('ssr:',SSR)

y=y[0:6]
SSE=np.sum((y-y_pred)**2)
print('sse:',SSE)

mean_total = np.mean(dataset.values)
# here df.to_numpy()will convert pandas Dataframe to Nump
SST=np.sum((dataset.values-mean_total)**2)
print('sst:',SST)

r_square = 1 - (SSR / SST)
r_square

print('rsquare:',r_square)
#print(bias_training)
#print(variance_testing)

#ML DEVELOPER

```

```
y_20 = slope*20 + c_inter  
print(y_20)
```

```
#-----  
plt.show()
```