```
# -*- coding: utf-8 -*-
"""
Created on Thu Aug 20 23:28:37 2020

@author: Taaashu
"""

#importing libraries
import pandas as pd
import matplotlib.pyplot as plt



#reading data from files
data = pd.read_csv("advertising.csv")
data.head()



#to visualizie data/data exploration
fig , axs = plt.subplots(1,3,sharey = True)
data.plot(kind='scatter',x='TV',y='Sales',ax=axs[0], figsize = (14,7))
data.plot(kind='scatter',x='Radio',y='Sales',ax=axs[1])
data.plot(kind='scatter',x='Newspaper',y='Sales',ax=axs[2])


#creating X&Y for linear regression
feature_cols = ['TV']
X = data[feature_cols]
y = data.Sales



#importing linear regression ALGO
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(X, y)

print(lr.intercept_)
print(lr.coef_)


#y=a+bx ie Result=Intercept+coefficent(investment)
result = 6.9748214882298925+0.05546477*50
print(result)


#least squared line
#create a dataframa with min & max value of the table

X_new = pd.DataFrame({'TV':[data.TV.min(),data.TV.max()]})
X_new.head()


preds = lr.predict(X_new)
```

```
preds


data.plot(kind = 'scatter',x='TV',y='Sales')

plt.plot(X_new,preds,c='red',linewidth = 1)



import statsmodels.formula.api as smf
lm = smf.ols(formula = 'Sales ~ TV',data =data).fit()
lm.conf_int()


#probality values?
lm.pvalues




#find the R-Squared values
lm.rsquared



#multi linear regression
feature_cols = ['TV','Radio','Newspaper']
X = data[feature_cols]
y = data.Sales

lr = LinearRegression()
lr.fit(X,y)


print(lr.intercept_)
print(lr.coef_)



#y=a+bx

lm = smf.ols(formula='Sales ~ TV+Radio+Newspaper', data=data).fit()
lm.conf_int()
lm.summary()


lm = smf.ols(formula='Sales ~ TV+Radio', data=data).fit()
lm.conf_int()
lm.summary()
```