**SVKM’s NMIMS**

**Mukesh Patel School of Technology Management & Engineering**

Program: B.Tech\MBA.Tech

**Course: Machine Learning**

**Experiment No.05**

PART A

(PART A : TO BE REFFERED BY STUDENTS)

**A.1 Aim:** To implement multiple linear regression

**A.2 Prerequisite:**

Knowledge of multiple linear regression

**A.3 Outcome:**

**After successful completion of this experiment students will be able to:**

1. Implement multiple linear regression by using sklearn package statsmodels.
2. Interpret the results obtained from different models and choose the best model for the given data set.

**A.4 Theory:**

**A.4.1 Linear Regression**

* Multiple regression is an extension of simple linear regression. A linear regression model that contains more than one predictor variable is called a *multiple linear regression model*.

http://reliawiki.org/images/math/1/3/3/1330581a877c4dc631b00c24577f5293.png

* The model is linear because it varies linearly with the change in the parameters β0, β1, β2
* The parameter β0 is the intercept of this plane, parameters β1 and β2 are referred to as partial regression coefficients.
* Parameter β1 represents the change in the mean response corresponding to a unit change x1 when x2 is held constant.
* Parameter β2 represents the change in the mean response corresponding to a unit change x2 when x1 is held constant.

Steps for Multiple linear regression

1. Identify the Independent and dependent variables.
2. Check the relationships between independent variable and dependent variable using scatter plot and correlations.
3. Check the relationships between independent variables using scatter plot and correlations.
4. Conduct simple linear regression for each IV/DV pair.
5. Use the non-redundant independent variables in the analysis to find the best fitting model.
6. Use the best fitting model to make predictions about the dependent variable

**Tasks:**

**Task 1:** **Relationships between the different features.**

* + - 1. Import the relevant libraries.
      2. Load the MLR\_data.csv in your notebooks.
      3. Perform EDA on the dataset by using head, shape and describe functions.
      4. Identify the independent variables (IV) and dependent variable (DV) and plot the scatter plot of IV vs DV. Write your inference for each plot.
      5. Plot the scatter plots for IV vs IV. Write your inference for each plot.

**Task 2: Fitting different simple linear regression models for each IV/DV pair.**

* + - 1. Import LinearRegression from SKlearn.

1. Create model for linear regression.
2. Conduct simple linear regression for each IV/DV pair.
   1. Interest rate vs stock index price
   2. Unemployment vs stock index price
3. Determine and tabulate the values of R2, slope and intercept for each model.
4. Determine the predicted value of stock index price for interest rate of 2.75.
5. Determine the predicted value of stock index price for unemployment rate of 6.
6. Compare both the models based on R2 and the mean square error between predicted and actual stock index price.

**Task 3: Multiple regression model**

* + - 1. Use both interest rate and unemployment rate and create a multiple regression model.
      2. Determine the value of R2, slope and intercept for the model.
      3. Determine the predicted value of stock index price for
  1. interest rate =2.75 and unemployment rate = 5.3
  2. interest rate =2 and unemployment rate = 6
     + 1. Compare the model with both the models from task 2.
       2. Identify and state the best model for the dataset.

PART B

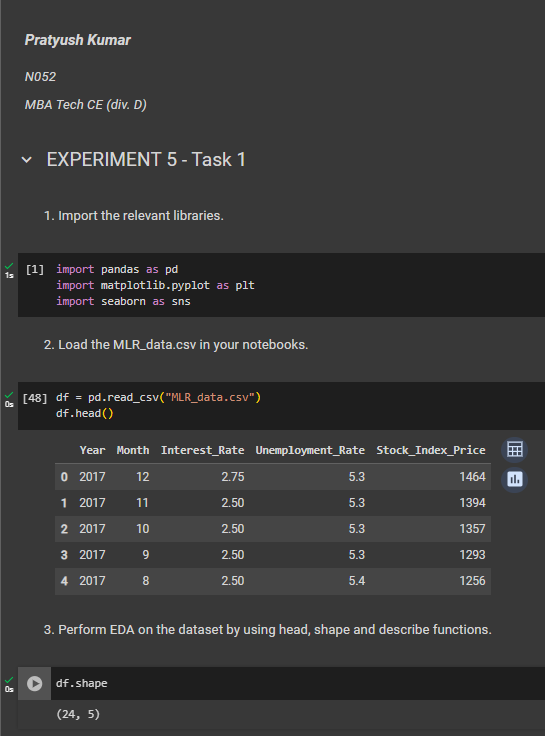
(PART B : TO BE COMPLETED BY STUDENTS)

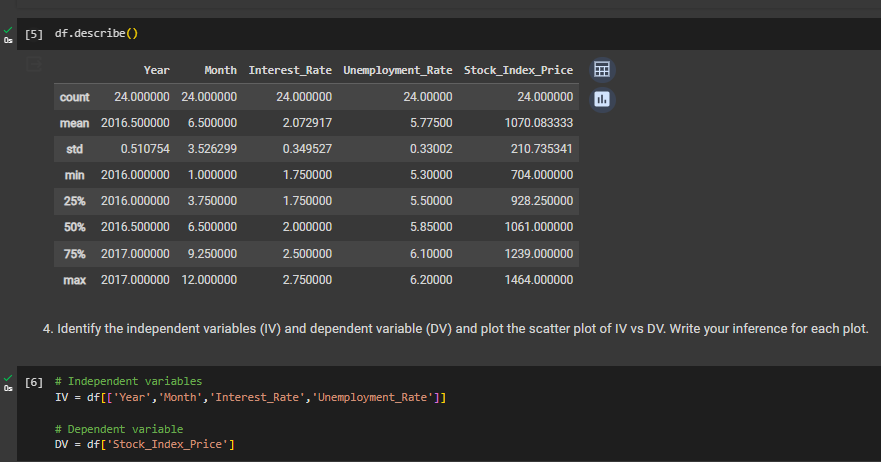
***(Students must submit the soft copy as per following segments within two hours of the practical.)***

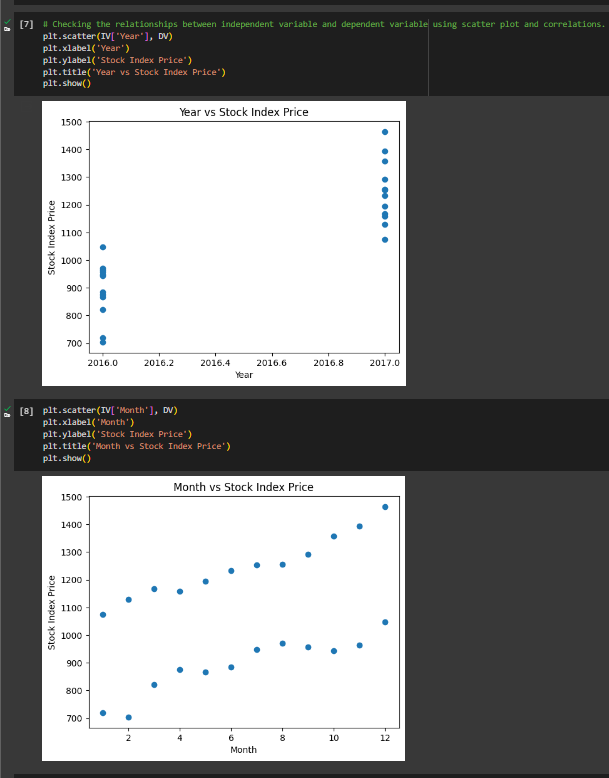
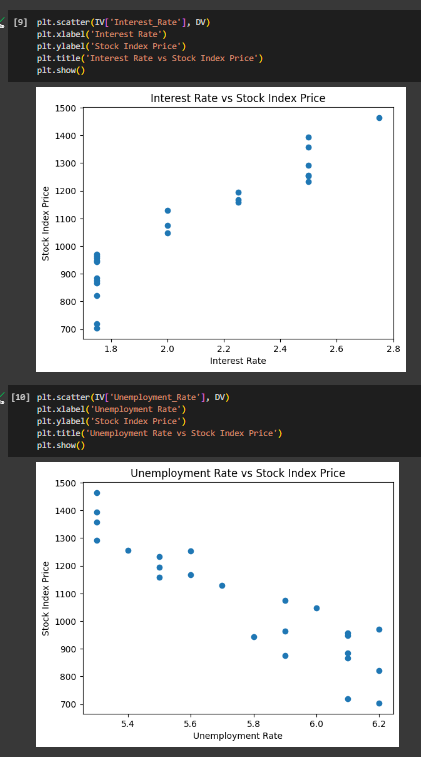
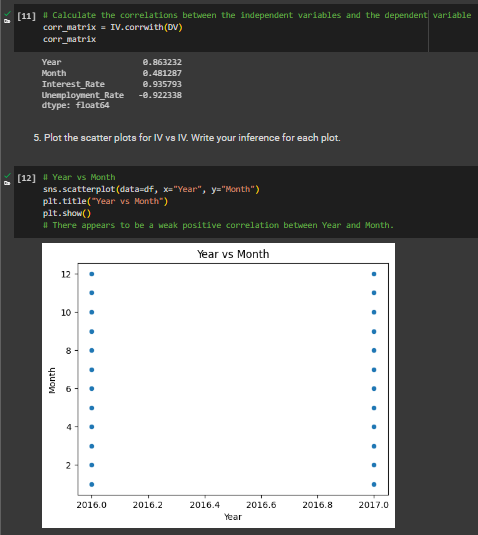
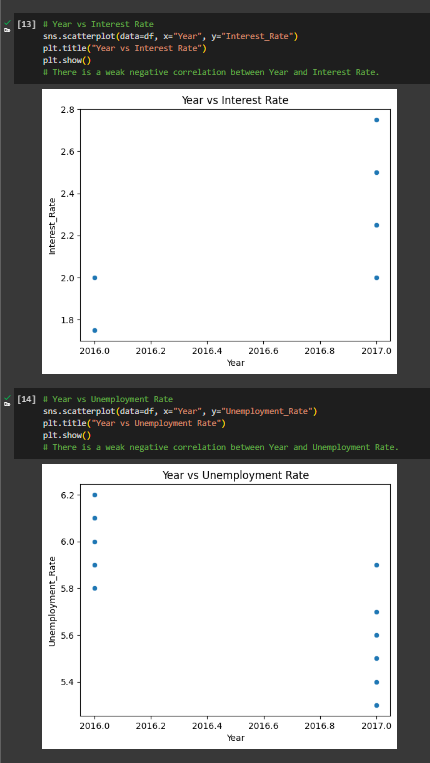
|  |  |
| --- | --- |
| Roll No. N052 | Name: Pratyush Kumar |
| Class : MBA Tech CE (div. D) | Batch : B2 |
| Date of Experiment: 03-02-2024 | Date of Submission: 10-02-2024 |
| Grade : |  |

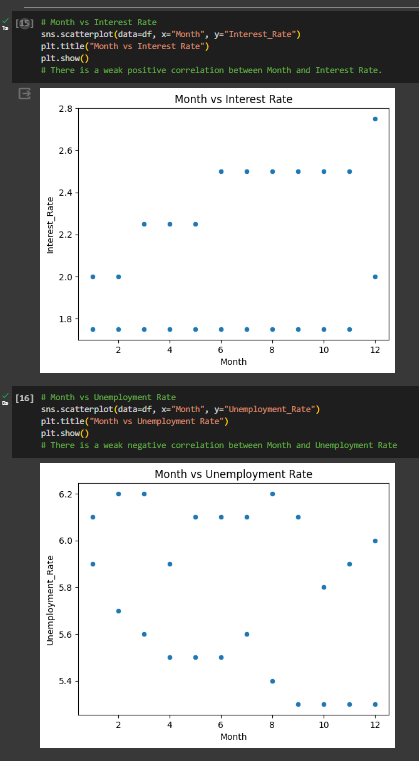
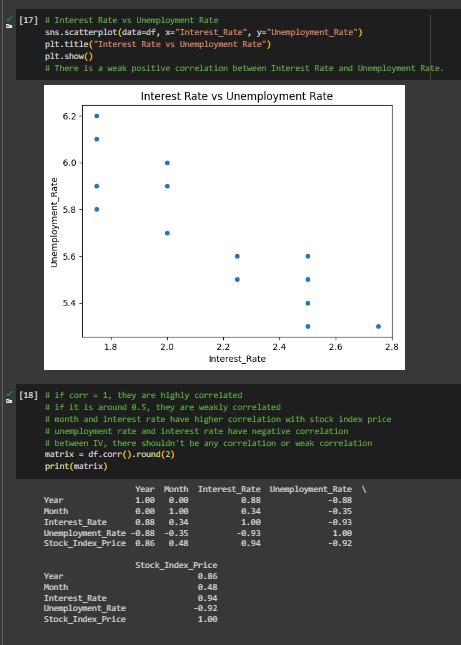
**B.1 Task 1**

**Colab link:** https://colab.research.google.com/drive/18CwN138HpRyw9ZSYfhPWRqWAbKINw7A\_?usp=sharing

* **Source Code**
* *"""  
   \* This file contains code snippets to find Relationships between the different features  
   \* ML-E5-Task1  
   \*  
   \* Original file is located at: https://colab.research.google.com/drive/18CwN138HpRyw9ZSYfhPWRqWAbKINw7A\_?usp=sharing  
   \* @author Pratyush Kumar (github.com/pratyushgta)  
  """*"""  
  ## EXPERIMENT 5 - Task 1  
  """  
    
  """  
  1. Import the relevant libraries.  
  """  
    
  import pandas as pd  
  import matplotlib.pyplot as plt  
  import seaborn as sns  
    
  """2. Load the MLR\_data.csv in your notebooks."""  
    
  df = pd.read\_csv("MLR\_data.csv")  
  df  
    
  """3. Perform EDA on the dataset by using head, shape and describe functions."""  
    
  df.head()  
    
  df.shape  
    
  df.describe()  
    
  """4. Identify the independent variables (IV) and dependent variable (DV) and plot the scatter plot of IV vs DV. Write your inference for each plot."""  
    
  # Independent variables  
  IV = df[['Year','Month','Interest\_Rate','Unemployment\_Rate']]  
    
  # Dependent variable  
  DV = df['Stock\_Index\_Price']  
    
  # Checking the relationships between independent variable and dependent variable using scatter plot and correlations.  
  plt.scatter(IV['Year'], DV)  
  plt.xlabel('Year')  
  plt.ylabel('Stock Index Price')  
  plt.title('Year vs Stock Index Price')  
  plt.show()  
    
  plt.scatter(IV['Month'], DV)  
  plt.xlabel('Month')  
  plt.ylabel('Stock Index Price')  
  plt.title('Month vs Stock Index Price')  
  plt.show()  
    
  plt.scatter(IV['Interest\_Rate'], DV)  
  plt.xlabel('Interest Rate')  
  plt.ylabel('Stock Index Price')  
  plt.title('Interest Rate vs Stock Index Price')  
  plt.show()  
    
  plt.scatter(IV['Unemployment\_Rate'], DV)  
  plt.xlabel('Unemployment Rate')  
  plt.ylabel('Stock Index Price')  
  plt.title('Unemployment Rate vs Stock Index Price')  
  plt.show()  
    
  # Calculate the correlations between the independent variables and the dependent variable  
  corr\_matrix = IV.corrwith(DV)  
  corr\_matrix  
    
  """5. Plot the scatter plots for IV vs IV. Write your inference for each plot."""  
    
  # Year vs Month  
  sns.scatterplot(data=df, x="Year", y="Month")  
  plt.title("Year vs Month")  
  plt.show()  
  # There appears to be a weak positive correlation between Year and Month.  
    
  # Year vs Interest Rate  
  sns.scatterplot(data=df, x="Year", y="Interest\_Rate")  
  plt.title("Year vs Interest Rate")  
  plt.show()  
  # There is a weak negative correlation between Year and Interest Rate.  
    
  # Year vs Unemployment Rate  
  sns.scatterplot(data=df, x="Year", y="Unemployment\_Rate")  
  plt.title("Year vs Unemployment Rate")  
  plt.show()  
  # There is a weak negative correlation between Year and Unemployment Rate.  
    
  # Month vs Interest Rate  
  sns.scatterplot(data=df, x="Month", y="Interest\_Rate")  
  plt.title("Month vs Interest Rate")  
  plt.show()  
  # There is a weak positive correlation between Month and Interest Rate.  
    
  # Month vs Unemployment Rate  
  sns.scatterplot(data=df, x="Month", y="Unemployment\_Rate")  
  plt.title("Month vs Unemployment Rate")  
  plt.show()  
  # There is a weak negative correlation between Month and Unemployment Rate  
    
  # Interest Rate vs Unemployment Rate  
  sns.scatterplot(data=df, x="Interest\_Rate", y="Unemployment\_Rate")  
  plt.title("Interest Rate vs Unemployment Rate")  
  plt.show()  
  # There is a weak positive correlation between Interest Rate and Unemployment Rate.  
    
  # if corr = 1, they are highly correlated  
  # if it is around 0.5, they are weakly correlated  
  # month and interest rate have higher correlation with stock index price  
  # unemployment rate and interest rate have negative correlation  
  # between IV, there shouldn't be any correlation or weak correlation  
  matrix = df.corr().round(2)  
  print(matrix)
* **Input/ Output**

****

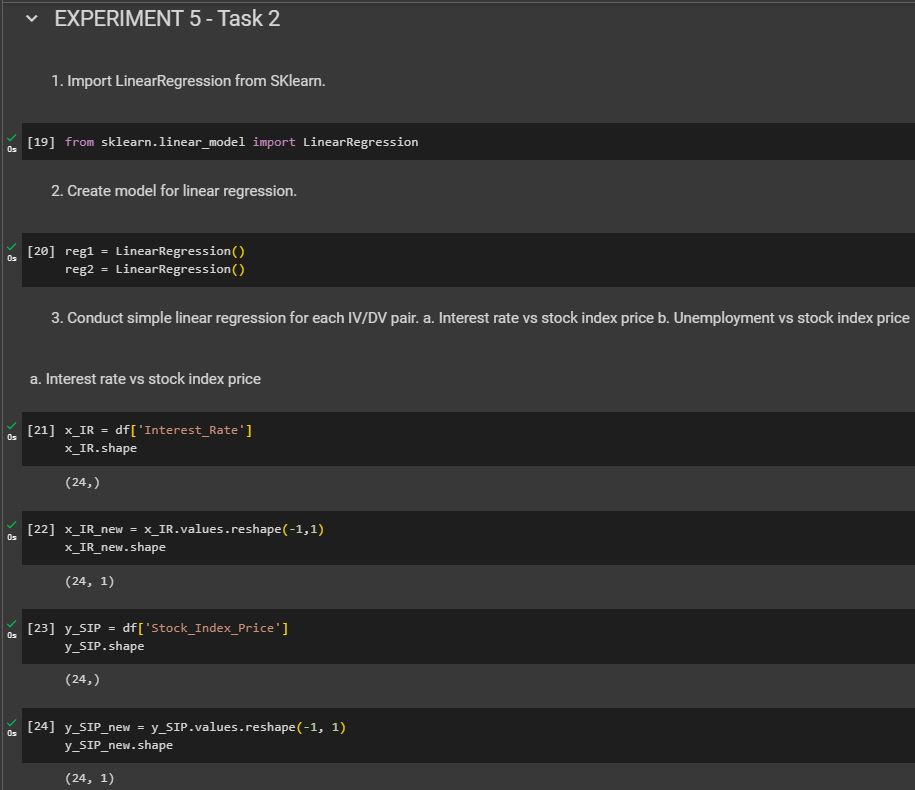
****

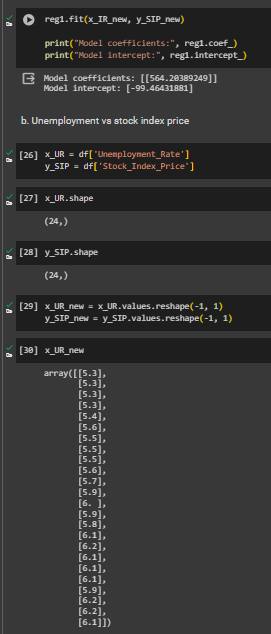
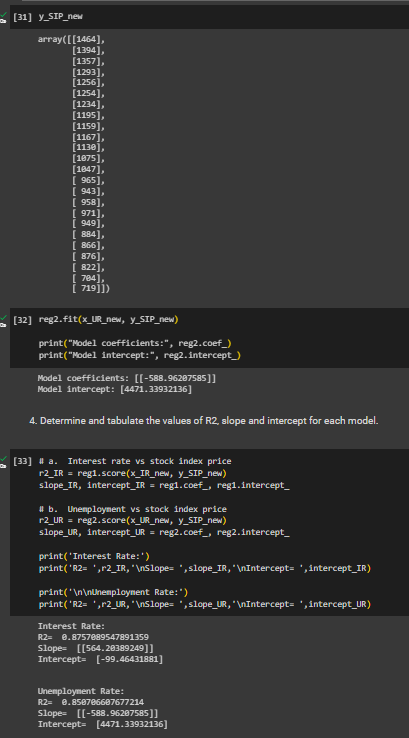
****

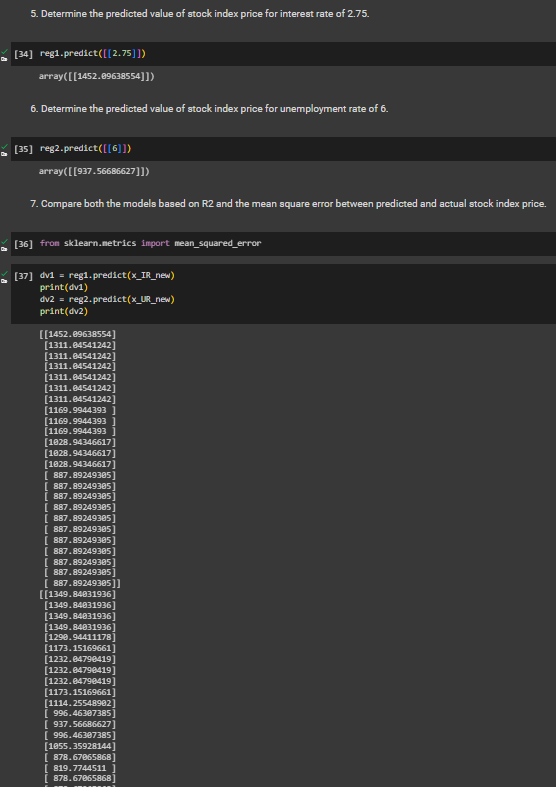
**B.2 Task 2**

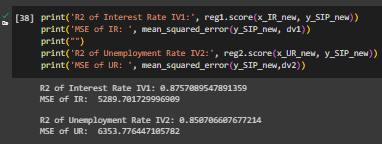
**Colab link:** https://colab.research.google.com/drive/18CwN138HpRyw9ZSYfhPWRqWAbKINw7A\_?usp=sharing

* **Source Code**
* *"""  
   \* This file contains code snippets to fit different simple linear regression models for each IV/DV pair  
   \* ML-E5-Task2  
   \*  
   \* Original file is located at: https://colab.research.google.com/drive/18CwN138HpRyw9ZSYfhPWRqWAbKINw7A\_?usp=sharing  
   \* @author Pratyush Kumar (github.com/pratyushgta)  
  """*"""  
  ## EXPERIMENT 5 - Task 2  
  """  
    
  """  
  1. Import LinearRegression from SKlearn.  
  """  
    
  from sklearn.linear\_model import LinearRegression  
    
  """2. Create model for linear regression."""  
    
  reg1 = LinearRegression()  
  reg2 = LinearRegression()  
    
  """3. Conduct simple linear regression for each IV/DV pair.  
  a. Interest rate vs stock index price  
  b. Unemployment vs stock index price  
    
  a. Interest rate vs stock index price  
  """  
    
  x\_IR = df['Interest\_Rate']  
  x\_IR.shape  
    
  x\_IR\_new = x\_IR.values.reshape(-1,1)  
  x\_IR\_new.shape  
    
  y\_SIP = df['Stock\_Index\_Price']  
  y\_SIP.shape  
    
  y\_SIP\_new = y\_SIP.values.reshape(-1, 1)  
  y\_SIP\_new.shape  
    
  reg1.fit(x\_IR\_new, y\_SIP\_new)  
    
  print("Model coefficients:", reg1.coef\_)  
  print("Model intercept:", reg1.intercept\_)  
    
  """b. Unemployment vs stock index price"""  
    
  x\_UR = df['Unemployment\_Rate']  
  y\_SIP = df['Stock\_Index\_Price']  
    
  x\_UR.shape  
    
  y\_SIP.shape  
    
  x\_UR\_new = x\_UR.values.reshape(-1, 1)  
  y\_SIP\_new = y\_SIP.values.reshape(-1, 1)  
    
  x\_UR\_new  
    
  y\_SIP\_new  
    
  reg2.fit(x\_UR\_new, y\_SIP\_new)  
    
  print("Model coefficients:", reg2.coef\_)  
  print("Model intercept:", reg2.intercept\_)  
    
  """4. Determine and tabulate the values of R2, slope and intercept for each model."""  
    
  # a. Interest rate vs stock index price  
  r2\_IR = reg1.score(x\_IR\_new, y\_SIP\_new)  
  slope\_IR, intercept\_IR = reg1.coef\_, reg1.intercept\_  
    
  # b. Unemployment vs stock index price  
  r2\_UR = reg2.score(x\_UR\_new, y\_SIP\_new)  
  slope\_UR, intercept\_UR = reg2.coef\_, reg2.intercept\_  
    
  print('Interest Rate:')  
  print('R2= ',r2\_IR,'\nSlope= ',slope\_IR,'\nIntercept= ',intercept\_IR)  
    
  print('\n\nUnemployment Rate:')  
  print('R2= ',r2\_UR,'\nSlope= ',slope\_UR,'\nIntercept= ',intercept\_UR)  
    
  """5. Determine the predicted value of stock index price for interest rate of 2.75."""  
    
  reg1.predict([[2.75]])  
    
  """6. Determine the predicted value of stock index price for unemployment rate of 6."""  
    
  reg2.predict([[6]])  
    
  """7. Compare both the models based on R2 and the mean square error between predicted and actual stock index price."""  
    
  from sklearn.metrics import mean\_squared\_error  
    
  dv1 = reg1.predict(x\_IR\_new)  
  print(dv1)  
  dv2 = reg2.predict(x\_UR\_new)  
  print(dv2)  
    
  print('R2 of Interest Rate IV1:', reg1.score(x\_IR\_new, y\_SIP\_new))  
  print('MSE of IR: ', mean\_squared\_error(y\_SIP\_new, dv1))  
  print("")  
  print('R2 of Unemployment Rate IV2:', reg2.score(x\_UR\_new, y\_SIP\_new))  
  print('MSE of UR: ', mean\_squared\_error(y\_SIP\_new,dv2))
* **Input/ Output**

****

****

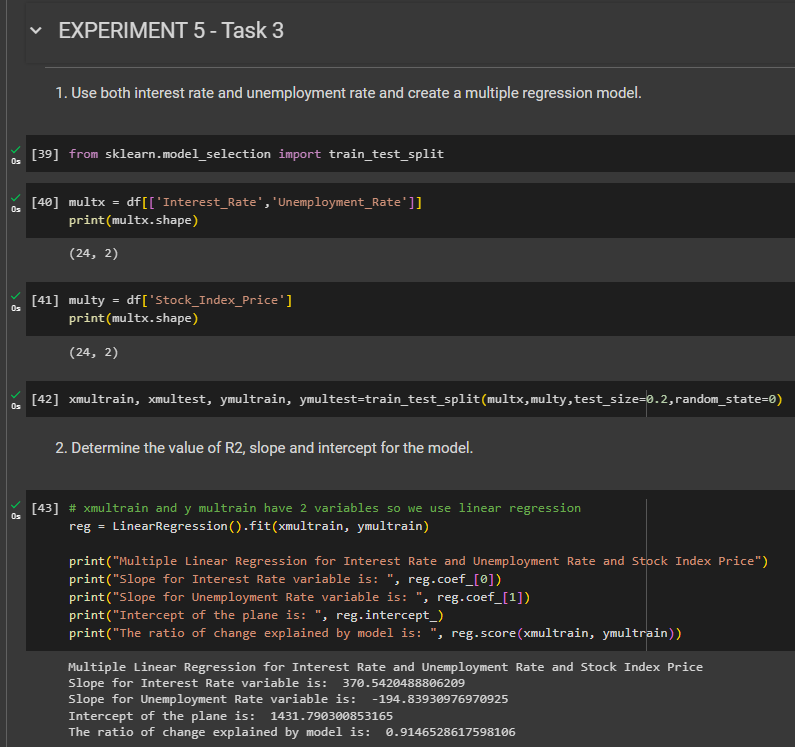
****

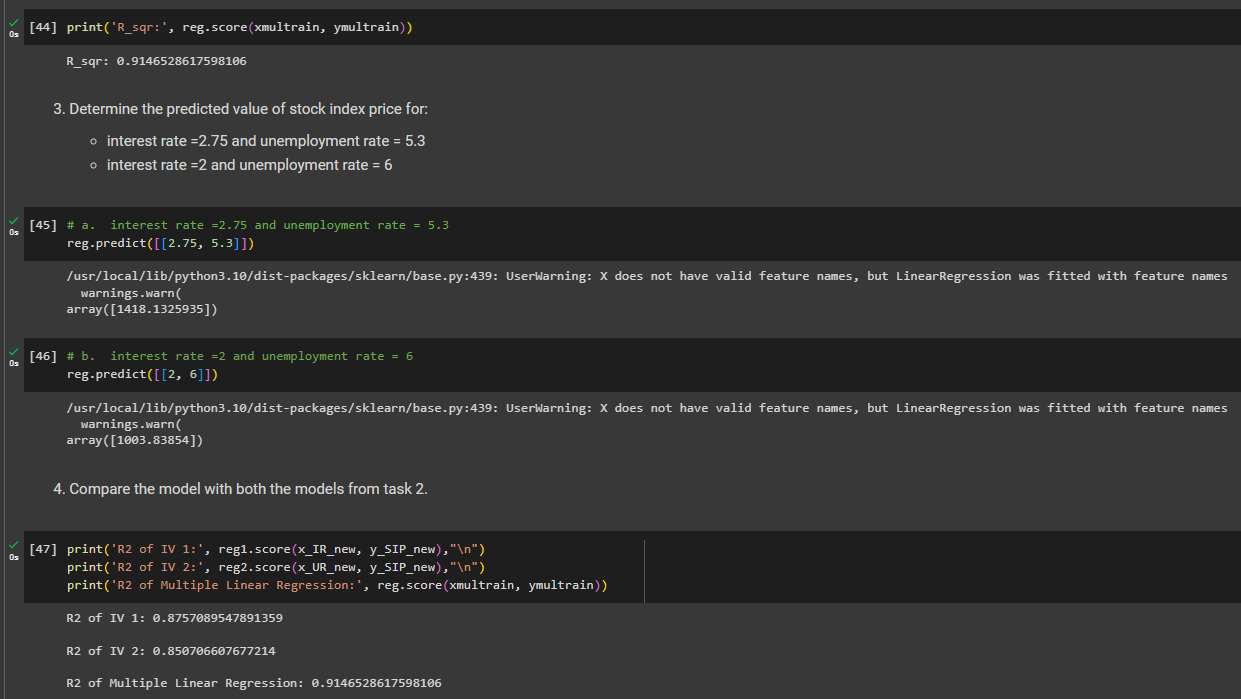
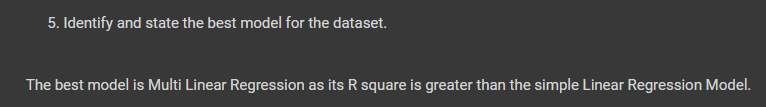
****

**B.3 Task 3**

**Colab link:** https://colab.research.google.com/drive/18CwN138HpRyw9ZSYfhPWRqWAbKINw7A\_?usp=sharing

* **Source Code**
* *"""  
   \* This file contains code snippets to implement a Multiple regression model  
   \* ML-E5-Task3  
   \*  
   \* Original file is located at: https://colab.research.google.com/drive/18CwN138HpRyw9ZSYfhPWRqWAbKINw7A\_?usp=sharing  
   \* @author Pratyush Kumar (github.com/pratyushgta)  
  """*"""  
  ## EXPERIMENT 5 - Task 3  
  """  
    
  *"""  
  1. Use both interest rate and unemployment rate and create a multiple regression model.  
  """*from sklearn.model\_selection import train\_test\_split  
    
  multx = df[['Interest\_Rate','Unemployment\_Rate']]  
  print(multx.shape)  
    
  multy = df['Stock\_Index\_Price']  
  print(multx.shape)  
    
  xmultrain, xmultest, ymultrain, ymultest=train\_test\_split(multx,multy,test\_size=0.2,random\_state=0)  
    
  """2. Determine the value of R2, slope and intercept for the model."""  
    
  # xmultrain and y multrain have 2 variables so we use linear regression  
  reg = LinearRegression().fit(xmultrain, ymultrain)  
    
  print("Multiple Linear Regression for Interest Rate and Unemployment Rate and Stock Index Price")  
  print("Slope for Interest Rate variable is: ", reg.coef\_[0])  
  print("Slope for Unemployment Rate variable is: ", reg.coef\_[1])  
  print("Intercept of the plane is: ", reg.intercept\_)  
  print("The ratio of change explained by model is: ", reg.score(xmultrain, ymultrain))  
    
  print('R\_sqr:', reg.score(xmultrain, ymultrain))  
    
  """3. Determine the predicted value of stock index price for:  
   \* interest rate =2.75 and unemployment rate = 5.3  
   \* interest rate =2 and unemployment rate = 6  
    
    
  """  
    
  # a. interest rate =2.75 and unemployment rate = 5.3  
  reg.predict([[2.75, 5.3]])  
    
  # b. interest rate =2 and unemployment rate = 6  
  reg.predict([[2, 6]])  
    
  """4. Compare the model with both the models from task 2."""  
    
  print('R2 of IV 1:', reg1.score(x\_IR\_new, y\_SIP\_new),"\n")  
  print('R2 of IV 2:', reg2.score(x\_UR\_new, y\_SIP\_new),"\n")  
  print('R2 of Multiple Linear Regression:', reg.score(xmultrain, ymultrain))  
    
  """5. Identify and state the best model for the dataset.  
    
  The best model is Multi Linear Regression as its R square is greater than the simple Linear Regression Model.  
  """
* **Input/ Output**

****

****

**B.4 Conclusion:**

*(Students must write the conclusion in their own words.)*

Implemented multiple linear regression using the sklearn package and statsmodels. Also interpreted the results obtained from different models and chose the best model for the given MLR dataset.