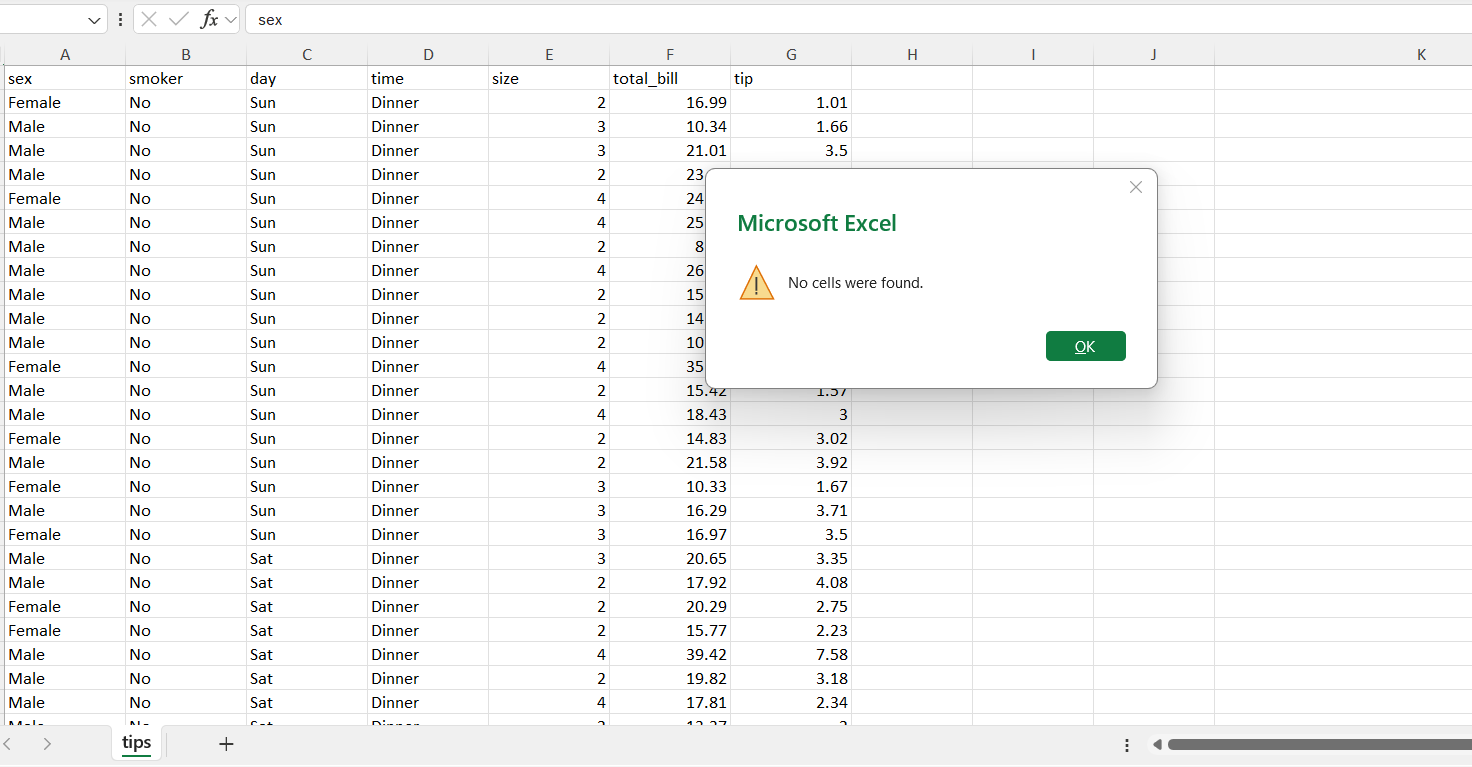
**Tasks to Perform  
1. Use the restaurant tips file for the analytics using Excel  
 2. Find out if there are any missing values and clean the data**  
  
Answer-  
  
I have opened the dataset on excel, to check the missing values I have selected all the data by CTRL+A , then clicked on the find and select tab under Home. Then, selected Go to special->Blanks->Ok

**Predicting Restaurant Tips Using Predictive Analytics on Excel**



To remove the duplicate values , I have selected all the data by CTRL+A , then clicked on Data ->Remove duplicates -> Ok

A screenshot of a computer

AI-generated content may be incorrect.

**3. Find the features that are independent and dependent**   
  
Answer-

As per problem statement, it’s clearly mentioned to predict the tips so in our data set tips filed is dependent upon the other field. Rest all the fields are independent.

**4. Identify which predictive problem is needed**

Answer-

As per the nature of our data, we have dependent variable and independent variables. So, a multiple linear regression model will be best suit in estimating the relationship between a dependent variable and independent variables.

**5. Encode the categorical variables to numeric values using IF conditions • For each independent numeric value, find its correlation coefficient with respect to the tip. The correlation coefficient is derived from the formula CORREL (array1, array2) where array1 is all the independent variables and array2 is the tip variable.**

Answer-

I have used the if conditions for categorical field sex, smoker, day, time to covert text numeric to numeric values.

Inserted new cells contains the numeric values and formula used as-

=IF(A2="Female",1,0)

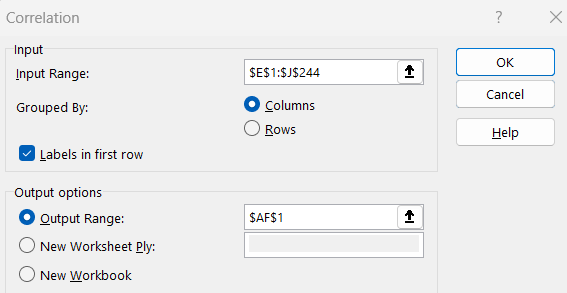
=IF(B2="Yes",1,0)

=IF(C2="Sun",1, IF(C2="Sat",2,IF(C2="Fri",3,4)))

=IF(D2="Dinner",1,0)

Then, I have checked for correlation between the independent variables.

Data -> Data Analysis -> Correlation



In the input range I have selected all the new numeric field variables and selected a blank cell for output range.

A screenshot of a spreadsheet

AI-generated content may be incorrect.

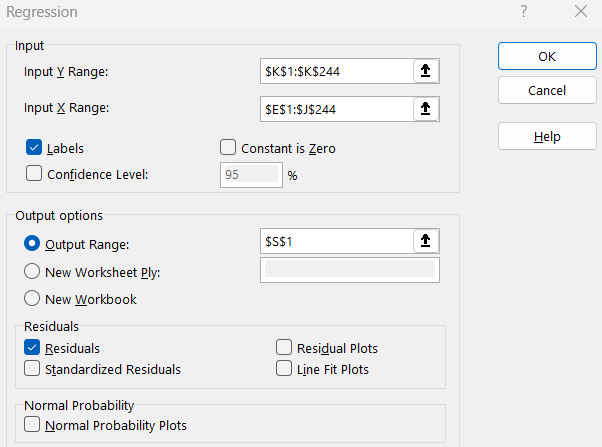
From the above data its clear in between the independent variables there is negative correlation and no correlation. No strong correlation has found.

**6. Build an appropriate model with the dataset**

Answer-

Considering the problem statement, we will use the multiple linear regression here.

Steps followed as Data -> Data analysis -> Regression



Y range- I have selected the range of independent variable i.e. tip.

X range- I have selected the range from sex, smoker, day, time.

I have checked Labels as my data has header and selected a blank cell for output. Also, checked residual for viewing the error.

**7. Calculate the predicted and actual tips values**

Answer-

Regression model already shows the predicted values.

To calculate manually we use the formula-

Y=β0X1+ β0X2+ β0X3+ β0X4+…. +C

Y=Indepenet variable

β0=Dependent variable coefficient

X=Dependent variable

C=Intercept

Formula used in second row for predicted tip is-

=$P$11\*E2+$P$12\*F2+$P$13\*G2+$P$14\*H2+$P$15\*I2+$P$16\*J2+$P$10

Which value is found to be equal to the value present on the regression model.

**8. Calculate the RMSE (Root Mean Square Error) of the model. RMSE is the root of the mean of square errors**

Answer-

Error= Actual – predicted tip

=K2-L2

RMSE= =SQRT (sqrt of Sum of error/COUNT the number of times error occurred)

=SQRT (SUMSQ (M2:M244)/COUNT (M2:M244))

= 1.007959468

--End--