**Steps Performed:**

1. Selected the data and filtered if there are any blank cells or not

2. Checked and removed duplicates

3. Found Independent and dependent variables

Independent Variables = Sex, Smoker, Day, Time, size, total\_bill

Dependent Variable = Tips

4. Converted categorical values into numerical values using IFS condition

5. Calculated Correlation Coefficients with the numerical values obtained.

6. Used regression analysis to obtain data model and observed the following:

* P-value of Sex, Smoker, day, time are greater than 0.05

7. Used regression analysis with size and total bill as the only independent variables

8. Then using predictive analysis formula, found the Predicted tips along alongside actual tips.

* Predicted Tip=Intercept+(Coefficient for Size×Size)+(Coefficient for Total Bill×Total Bill)

9. Finally calculated RMSE after calculating Squared error and Mean of the Squared error using the following equations:

* Squared Error = (Actual Tip - Predicted Tip)^2
* mean of the squared errors = AVERAGE([range of cells containing squared errors])
* RMSE = square root of the mean of the squared errors