Predicting Restaurant Tips Using Predictive Analytics on Excel.

Name – Sanjeev Pandey

Date – 29/07/2023

Problem Statement

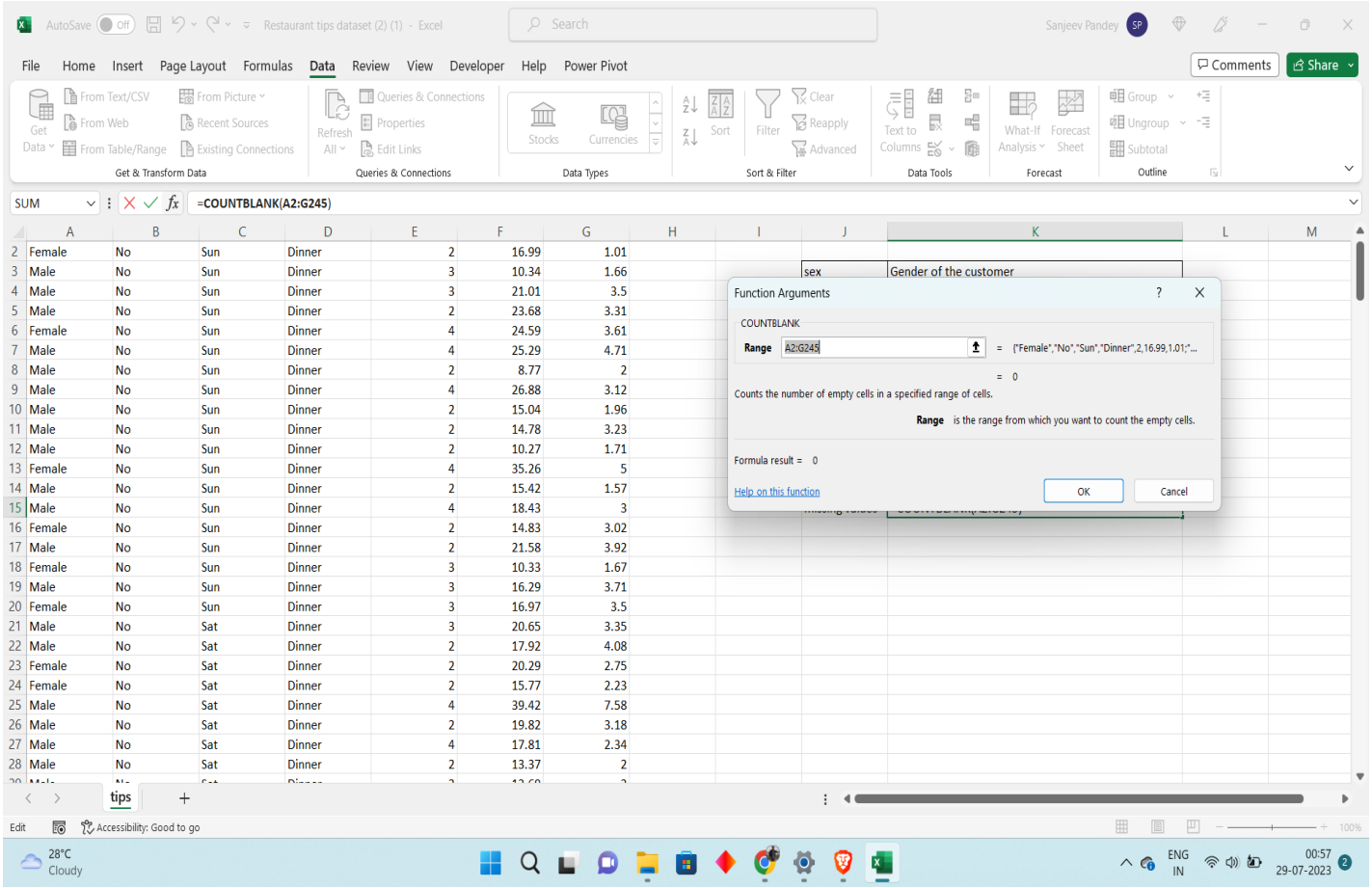
Use Excel to build a model to predict restaurant tips given input values with the mathematical equation for predicting the value of the tips.

Dataset Description

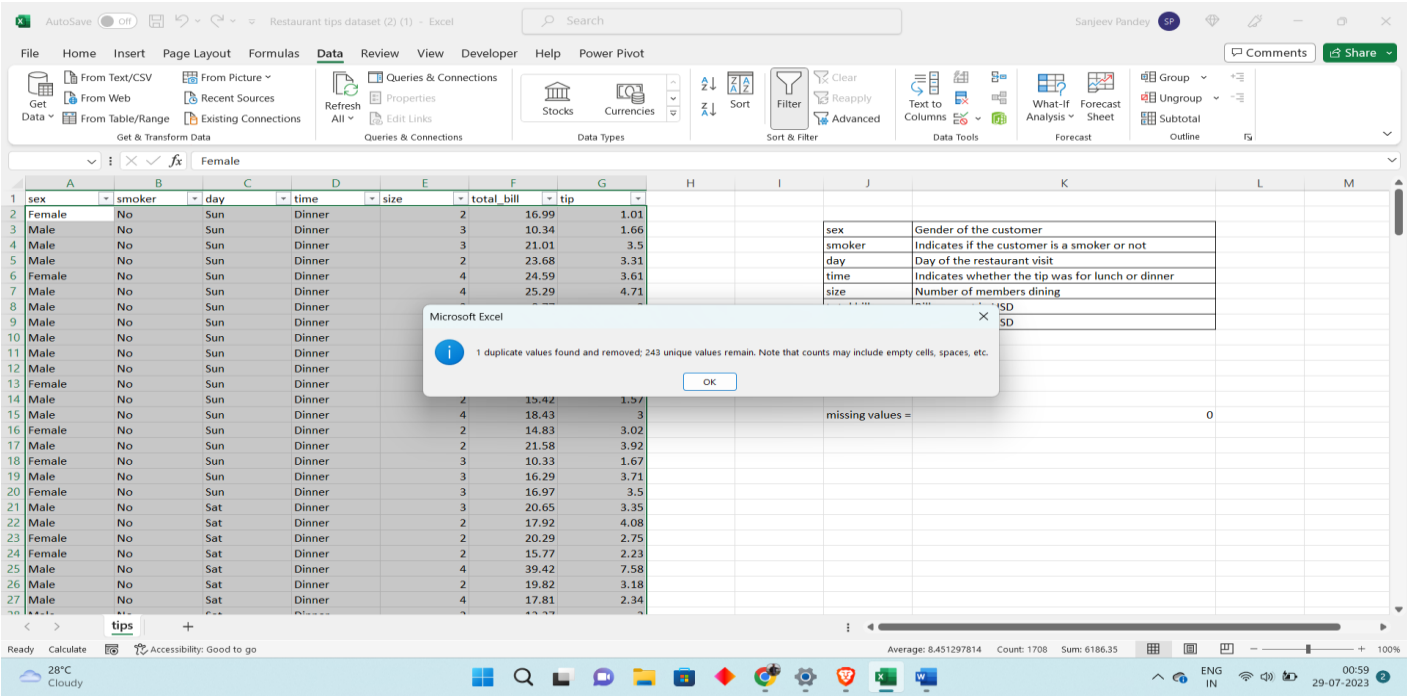
- sex - Gender of the customer
- smoker - Indicates if the customer is a smoker or not
- day - Day of the restaurant visit
- Time - Indicates whether the tip was for lunch or dinner
- size - Number of members dining
- total bill - Bill amount in USD
- tip - Tip amount in USD

Tasks I Performed -

1. I checked for null values using Isblank() function.There were no null values in my dataset.



1. I Checked for duplicate values ,I only found one duplicate values which I removed successfully.

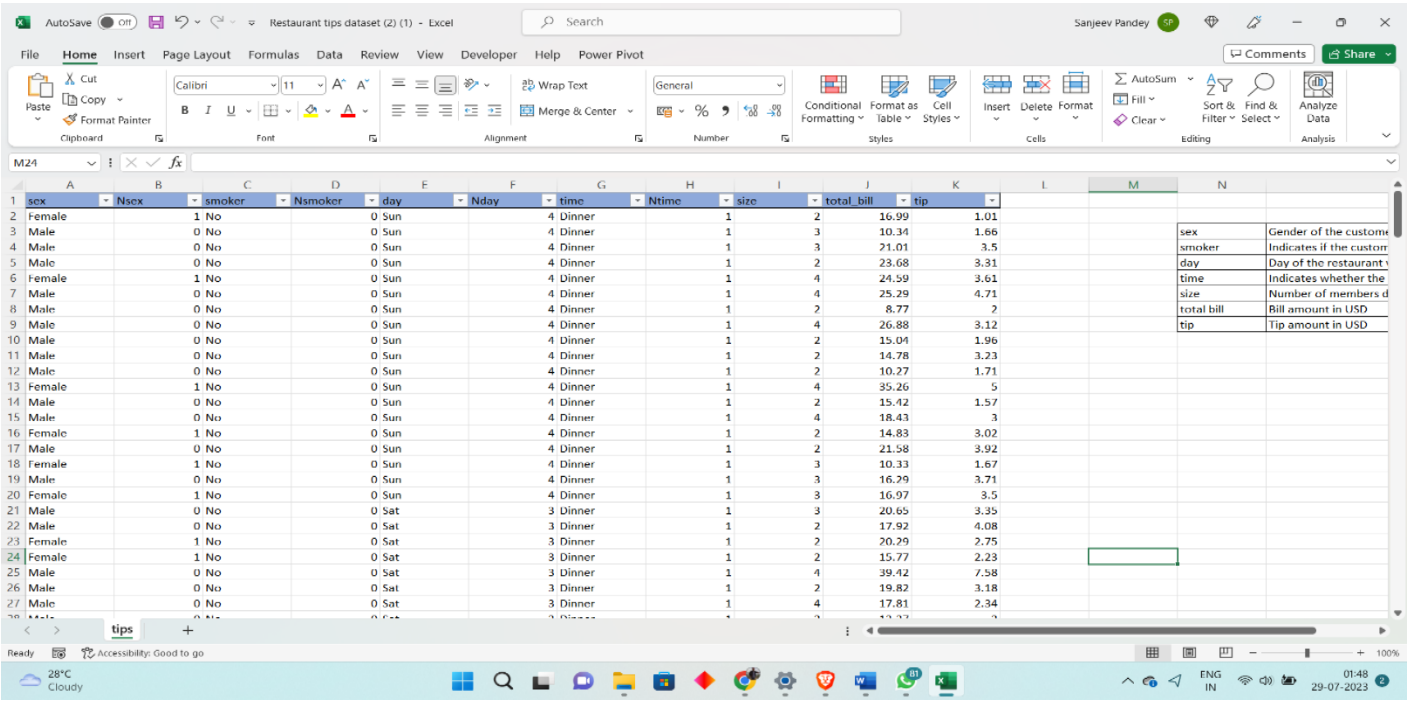


2. I categorised all the columns as independent and dependent values.
   Independent features are = sex, smoker, day, time, size, total bill
   Dependent feature = tip.

3. After understanding the data, I decided to use regression model for model prediction.

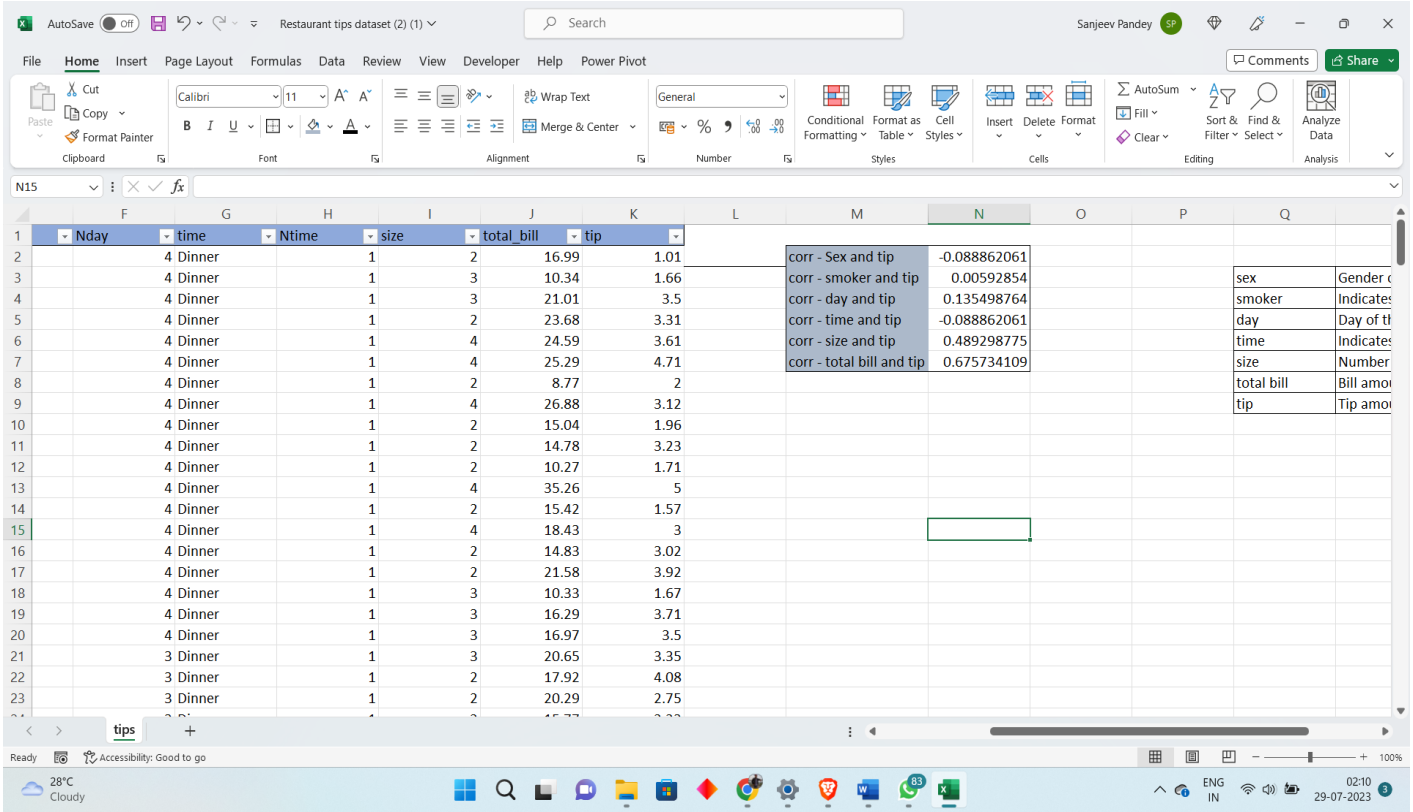
4. In order to use regression, I had to convert all the categorical values to numeric values. I used these codes to convert all the column to numeric.

```
=IF(A2="Male", 0, IF(A2="Female", 1, ""))
=IF(A2="Yes", 1, IF(A2="No", 0, ""))
=IF(A2="thur", 1, IF(A2="fri", 2, IF(A2="sat", 3, IF(A2="sun", 4, ""))))
=IF(A2="dinner", 1, IF(A2="lunch", 2, ""))
```



5. Correlation -

This image is showing the correlation between each independent variable to the dependent variable.



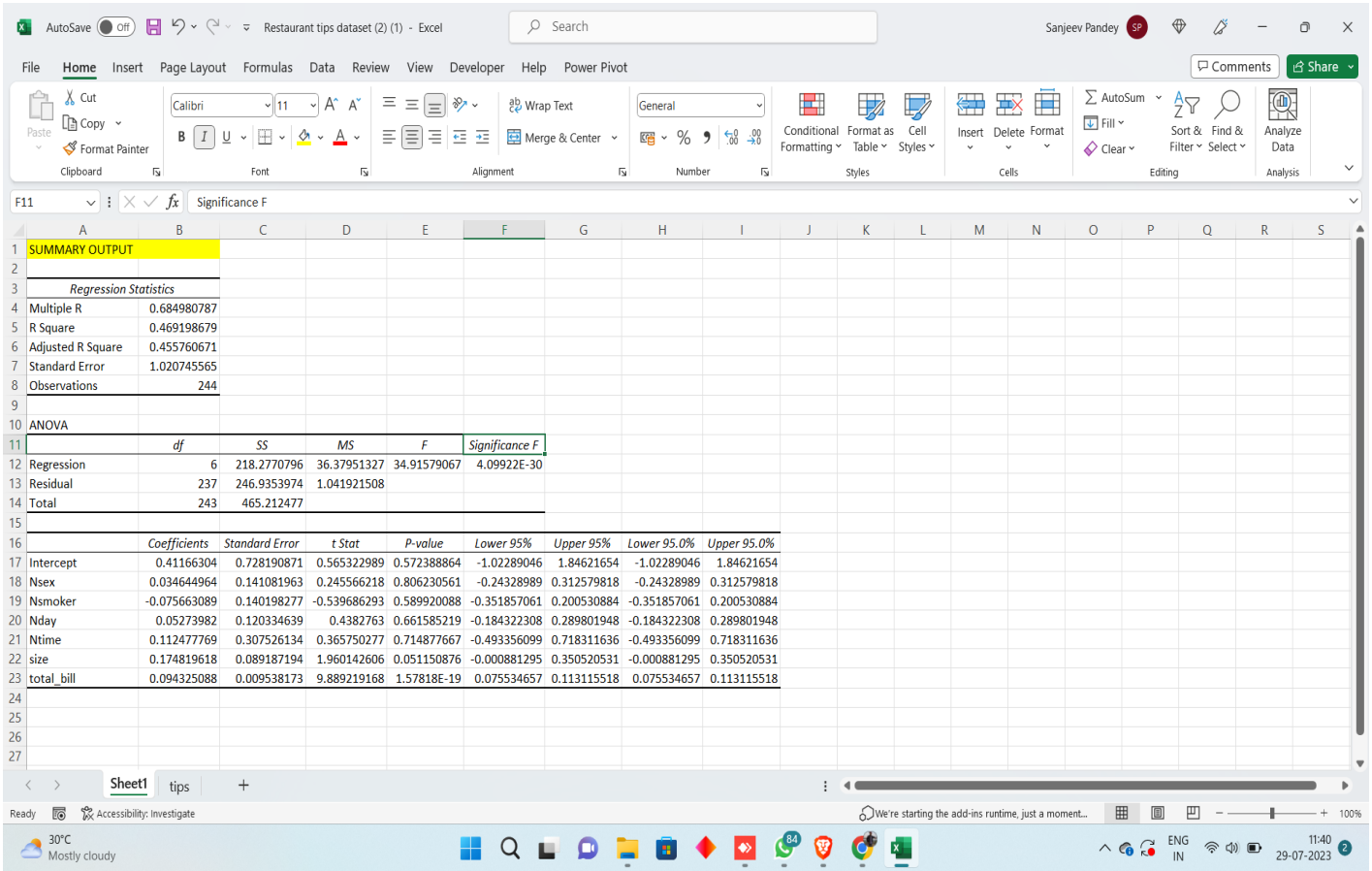
Nday	time	Ntime	size	total_bill	tip
4	Dinner	1	2	16.99	1.01
4	Dinner	1	3	10.34	1.66
4	Dinner	1	3	21.01	3.5
4	Dinner	1	2	23.68	3.31
4	Dinner	1	4	24.59	3.61
4	Dinner	1	4	25.29	4.71
4	Dinner	1	2	8.77	2
4	Dinner	1	4	26.88	3.12
4	Dinner	1	2	15.04	1.96
4	Dinner	1	2	14.78	3.23
4	Dinner	1	2	10.27	1.71
4	Dinner	1	4	35.26	5
4	Dinner	1	2	15.42	1.57
4	Dinner	1	4	18.43	3
4	Dinner	1	2	14.83	3.02
4	Dinner	1	2	21.58	3.92
4	Dinner	1	3	10.33	1.67
4	Dinner	1	3	16.29	3.71
4	Dinner	1	3	16.97	3.5
3	Dinner	1	3	20.65	3.35
3	Dinner	1	2	17.92	4.08
3	Dinner	1	2	20.29	2.75

corr - Sex and tip	-0.088862061
corr - smoker and tip	0.00592854
corr - day and tip	0.135498764
corr - time and tip	-0.088862061
corr - size and tip	0.489298775
corr - total bill and tip	0.675734109

sex	Gender
smoker	Indicates
day	Day of t
time	Indicates
size	Number
total bill	Bill amo
tip	Tip amo

6. I have created a regression model by using excel add-ins.

All the p values are in appropriate limit so we will go with this model.



SUMMARY OUTPUT

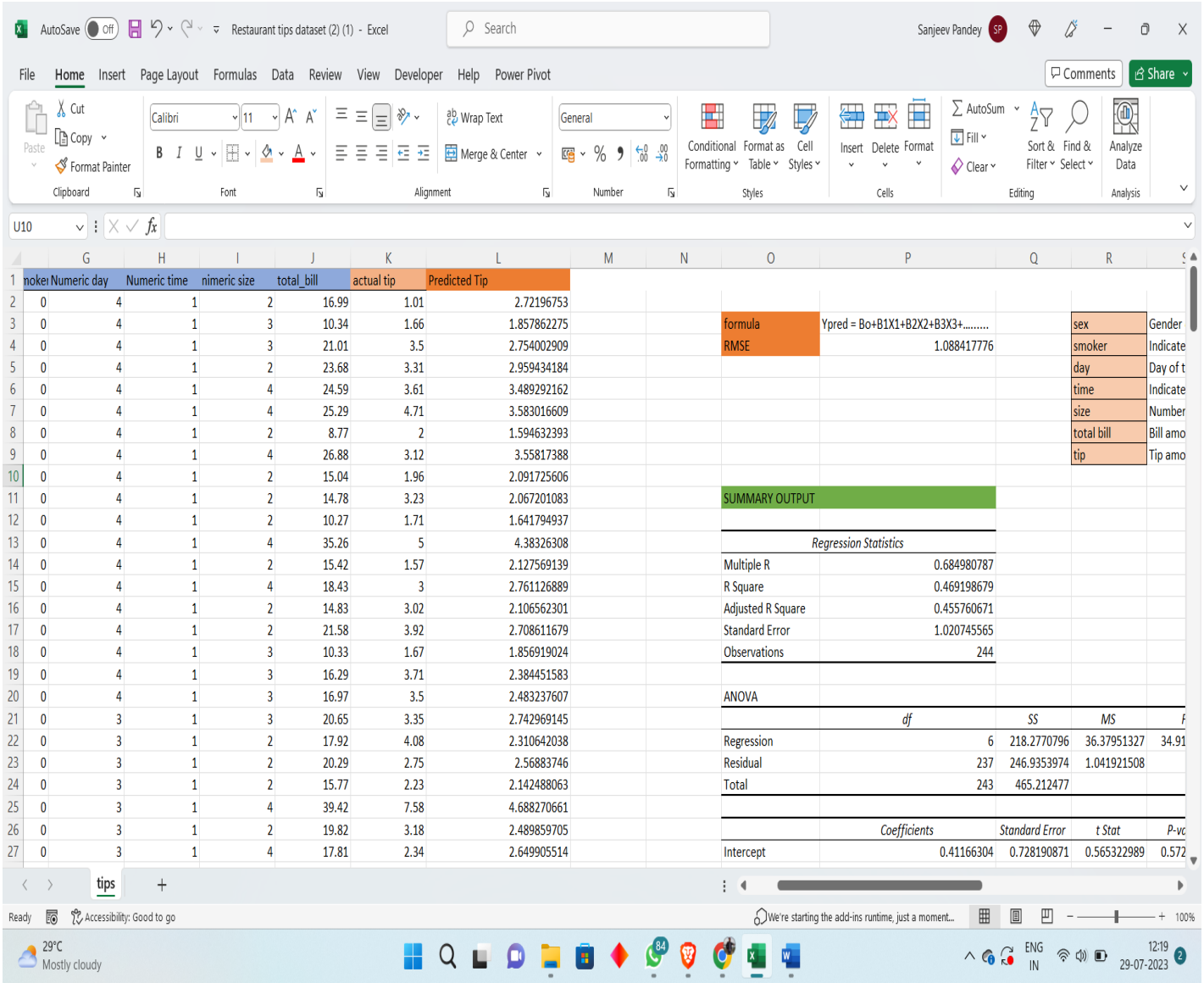
Regression Statistics	
Multiple R	0.684980787
R Square	0.469198679
Adjusted R Square	0.455760671
Standard Error	1.020745565
Observations	244

ANOVA

	df	SS	MS	F	Significance F
Regression	6	218.2770796	36.37951327	34.91579067	4.09922E-30
Residual	237	246.9353974	1.041921508		
Total	243	465.212477			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.41166304	0.728190871	0.565322989	0.572388864	-1.02289046	1.84621654	-1.02289046	1.84621654
Nsex	0.034644964	0.141081963	0.245566218	0.806230561	-0.24328989	0.312579818	-0.24328989	0.312579818
Nsmoker	-0.075663089	0.140198277	-0.539686293	0.589920088	-0.351857061	0.200530884	-0.351857061	0.200530884
Nday	0.05273982	0.120334639	0.4382763	0.661585219	-0.184322308	0.289801948	-0.184322308	0.289801948
Ntime	0.112477769	0.307526134	0.365750277	0.714877667	-0.493356099	0.718311636	-0.493356099	0.718311636
size	0.174819618	0.089187194	1.960142606	0.051150876	-0.000881295	0.350520531	-0.000881295	0.350520531
total_bill	0.094325088	0.009538173	9.889219168	1.57818E-19	0.075534657	0.113115518	0.075534657	0.113115518

7. Now, the regression model gave me the information about by coefficients and intercepts, using that I calculated the predicted values.
   The root mean square error between the actual and predicted values is – 1.08841 which is well within range.



nokei	Numeric day	Numeric time	nimeric size	total_bill	actual tip	Predicted Tip
0	4	1	2	16.99	1.01	2.72196753
0	4	1	3	10.34	1.66	1.857862275
0	4	1	3	21.01	3.5	2.754002909
0	4	1	2	23.68	3.31	2.959434184
0	4	1	4	24.59	3.61	3.489292162
0	4	1	4	25.29	4.71	3.583016609
0	4	1	2	8.77	2	1.594632393
0	4	1	4	26.88	3.12	3.55817388
0	4	1	2	15.04	1.96	2.091725606
0	4	1	2	14.78	3.23	2.067201083
0	4	1	2	10.27	1.71	1.641794937
0	4	1	4	35.26	5	4.38326308
0	4	1	2	15.42	1.57	2.127569139
0	4	1	4	18.43	3	2.761126889
0	4	1	2	14.83	3.02	2.106562301
0	4	1	2	21.58	3.92	2.708611679
0	4	1	3	10.33	1.67	1.856919024
0	4	1	3	16.29	3.71	2.384451583
0	4	1	3	16.97	3.5	2.483237607
0	3	1	3	20.65	3.35	2.742969145
0	3	1	2	17.92	4.08	2.310642038
0	3	1	2	20.29	2.75	2.56883746
0	3	1	2	15.77	2.23	2.142488063
0	3	1	4	39.42	7.58	4.688270661
0	3	1	2	19.82	3.18	2.489859705
0	3	1	4	17.81	2.34	2.649905514

formula	Ypred = Bo+B1X1+B2X2+B3X3+.....			
RMSE	1.088417776			
			sex	Gender
			smoker	Indicate
			day	Day of t
			time	Indicate
			size	Number
			total bill	Bill amo
			tip	Tip amo

SUMMARY OUTPUT				
	Regression Statistics			
Multiple R	0.684980787			
R Square	0.469198679			
Adjusted R Square	0.455760671			
Standard Error	1.020745565			
Observations	244			
ANOVA				
	df	*SS*	*MS*	*F*
Regression	6	218.2770796	36.37951327	34.91
Residual	237	246.9353974	1.041921508	
Total	243	465.212477		
	Coefficients	*Standard Error*	*t Stat*	*P-va*
Intercept	0.41166304	0.728190871	0.565322989	0.572

So lastly, coming to the prediction part – If we enter any new entries in the column, it will automatically convert the values to numeric and give us the predicted tip.