

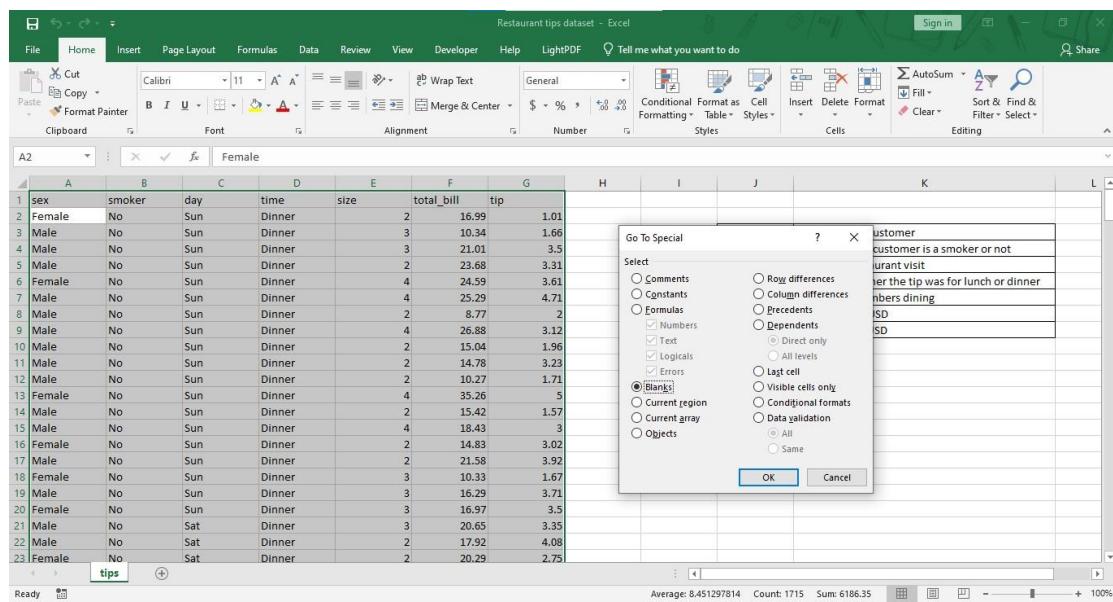
Predicting Restaurant Tips Using Predictive Analytics on Excel

1. Use the restaurant tips file for the analytics using Excel

Open ‘Restaurant tips dataset.xlsx’ file in your computer.

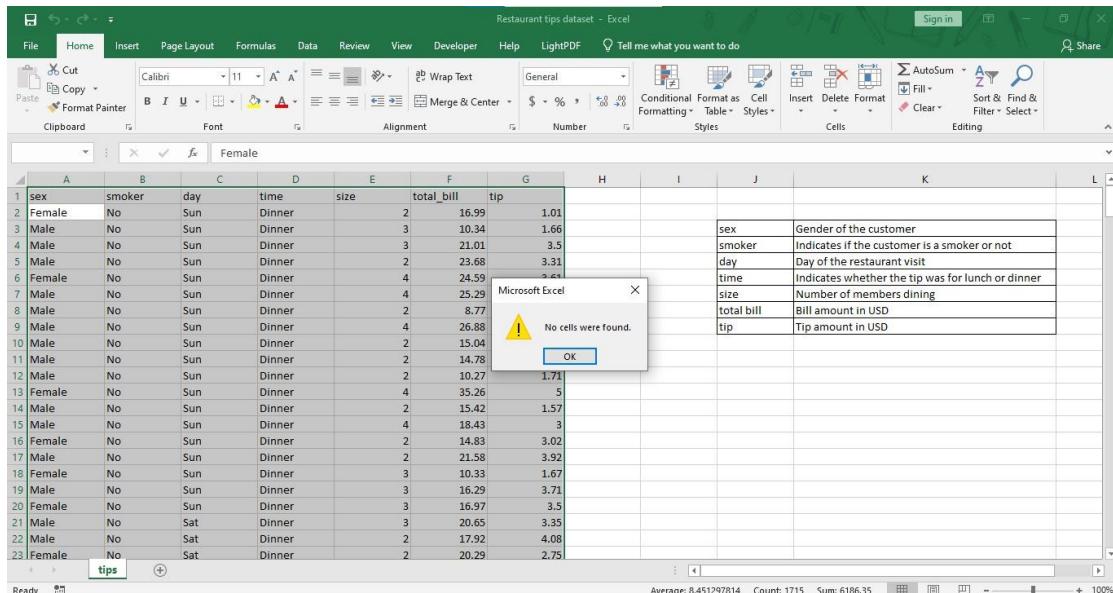
2. Find out if there are any missing values and clean the data

- Remove rows with at least one empty cell; choose all data using Control+A and in the home tab, click on Find & Select and go to Special.



The screenshot shows the 'Go To Special' dialog box in Excel. The 'Blanks' option is selected in the 'Select' section. The 'Customer' column (Column H) is highlighted in blue, indicating it contains no data. The main Excel interface shows a dataset with columns: sex, smoker, day, time, size, total_bill, tip.

There is no missing value.



The screenshot shows a 'Microsoft Excel' message box with the text 'No cells were found.' and an 'OK' button. The background shows the same dataset as the previous screenshot, with the 'Customer' column still highlighted in blue.

- b. Check for duplicates: Choose the dataset using Control+A; In the Data tab, click on Remove duplicates.

To check for duplicate values create helper column using CONCAT() function.

	A	B	C	D	E	F	G	H	I	J	K
1	sex	smoker	day	time	size	total_bill	tip	Helper column for duplicate data			
2	Female	No	Sun	Dinner	2	16.99	1.01	=CONCAT(A2:G2)			
3	Male	No	Sun	Dinner	3	10.34	1.66	N CONCAT(text1,...) 10.341.66	sex	Gender of the customer	
4	Male	No	Sun	Dinner	3	21.01	3.5	MaleNoSunDinner321.013.5	smoker	Indicates if the customer is a smoker or no	
5	Male	No	Sun	Dinner	2	23.68	3.31	MaleNoSunDinner23.683.31	day	Day of the restaurant visit	
6	Female	No	Sun	Dinner	4	24.59	3.61	FemaleNoSunDinner424.593.61	time	Indicates whether the tip was for lunch or dinner	
7	Male	No	Sun	Dinner	4	25.29	4.71	MaleNoSunDinner425.294.71	size	Number of members dining	
8	Male	No	Sun	Dinner	2	8.77	2	MaleNoSunDinner28.772	total_bill	Bill amount in USD	
9	Male	No	Sun	Dinner	4	26.88	3.12	MaleNoSunDinner426.883.12	tip	Tip amount in USD	
10	Male	No	Sun	Dinner	2	15.04	1.96	MaleNoSunDinner215.041.96			
11	Male	No	Sun	Dinner	2	14.78	3.23	MaleNoSunDinner214.783.23			
12	Male	No	Sun	Dinner	2	10.27	1.71	MaleNoSunDinner210.271.71			
13	Female	No	Sun	Dinner	4	35.26	5	FemaleNoSunDinner435.265			
14	Male	No	Sun	Dinner	2	15.42	1.57	MaleNoSunDinner215.421.57			
15	Male	No	Sun	Dinner	4	18.43	3	MaleNoSunDinner418.433			
16	Female	No	Sun	Dinner	2	14.83	3.02	FemaleNoSunDinner214.833.02			
17	Male	No	Sun	Dinner	2	21.58	3.92	MaleNoSunDinner221.583.92			
18	Female	No	Sun	Dinner	3	10.33	1.67	FemaleNoSunDinner310.331.67			
19	Male	No	Sun	Dinner	3	16.29	3.71	MaleNoSunDinner316.293.71			
20	Female	No	Sun	Dinner	3	16.97	3.5	FemaleNoSunDinner316.973.5			
21	Male	No	Sat	Dinner	3	20.65	3.35	MaleNoSatDinner320.653.35			
22	Male	No	Sat	Dinner	2	17.92	4.08	MaleNoSatDinner217.924.08			
23	Female	No	Sat	Dinner	2	20.29	2.75	FemaleNoSatDinner220.292.75			

Check for similar data in helper column using conditional formatting.

	A	B	C	D	E	F	G	H	I	J	K
1	sex	smoker	day	time	size	total_bill	tip	Helper column			
2	Female	No	Sun	Dinner	2	16.99	1.01	FemaleNoSun			
3	Male	No	Sun	Dinner	3	10.34	1.66	MaleNoSun			
4	Male	No	Sun	Dinner	3	21.01	3.5	MaleNoSun			
5	Male	No	Sun	Dinner	2	23.68	3.31	MaleNoSun			
6	Female	No	Sun	Dinner	4	24.59	3.61	FemaleNoSun			
7	Male	No	Sun	Dinner	4	25.29	4.71	MaleNoSun			
8	Male	No	Sun	Dinner	2	8.77	2	MaleNoSun			
9	Male	No	Sun	Dinner	4	26.88	3.12	MaleNoSun			
10	Male	No	Sun	Dinner	2	15.04	1.96	MaleNoSun			
11	Male	No	Sun	Dinner	2	14.78	3.23	MaleNoSun			
12	Male	No	Sun	Dinner	2	10.27	1.71	MaleNoSun			
13	Female	No	Sun	Dinner	4	35.26	5	FemaleNoSun			
14	Male	No	Sun	Dinner	2	15.42	1.57	MaleNoSun			
15	Male	No	Sun	Dinner	4	18.43	3	MaleNoSun			
16	Female	No	Sun	Dinner	2	14.83	3.02	FemaleNoSun			
17	Male	No	Sun	Dinner	2	21.58	3.92	MaleNoSun			
18	Female	No	Sun	Dinner	3	10.33	1.67	FemaleNoSun			
19	Male	No	Sun	Dinner	3	16.29	3.71	MaleNoSun			
20	Female	No	Sun	Dinner	3	16.97	3.5	FemaleNoSun			
21	Male	No	Sat	Dinner	3	20.65	3.35	MaleNoSat			
22	Male	No	Sat	Dinner	2	17.92	4.08	MaleNoSat			
23	Female	No	Sat	Dinner	2	20.29	2.75	FemaleNoSat			

There are only two rows with similar data i.e. only one duplicate data.

	A	B	C	D	E	F	G	H	I	J	K
1	sex	smoker	day	time	size	total_bill	tip	Helper column for duplicate data			
197	Male	No	Thur	Lunch	2	7.56	1.44	1 MaleNoThurLunch72.561.44			
198	Male	Yes	Thur	Lunch	2	10.34	2.04	2 MaleYesThurLunch210.342			
199	Female	Yes	Thur	Lunch	4	43.11	5.67	5 FemaleYesThurLunch43.115			
200	Female	Yes	Thur	Lunch	2	13	2.00	2 FemaleYesThurLunch43.112			
201	Male	Yes	Thur	Lunch	2	13.81	2.01	2 MaleYesThurLunch213.812			
202	Male	Yes	Thur	Lunch	3	18.71	3.40	4 MaleYesThurLunch318.714			
203	Female	Yes	Thur	Lunch	2	12.74	2.01	2.01 FemaleYesThurLunch212.742.01			
204	Female	Yes	Thur	Lunch	2	13	2.00	2 FemaleYesThurLunch213.002			
205	Female	Yes	Thur	Lunch	2	16.4	2.56	2.5 FemaleYesThurLunch216.42.5			
206	Male	Yes	Thur	Lunch	4	20.53	4.09	4 MaleYesThurLunch420.534			
207	Female	Yes	Thur	Lunch	3	16.47	3.23	3.23 FemaleYesThurLunch316.473.23			
208	Male	Yes	Sat	Dinner	3	26.59	5.41	5.41 MaleYesSatDinner326.593.41			
209	Male	Yes	Sat	Dinner	4	38.73	7.74	3 MaleYesSatDinner438.733			
210	Male	Yes	Sat	Dinner	2	24.27	4.05	2.01 MaleYesSatDinner224.272.03			
211	Female	Yes	Sat	Dinner	2	12.76	2.23	2.21 FemaleYesSatDinner212.762.23			
212	Male	Yes	Sat	Dinner	3	30.06	5.61	2 MaleYesSatDinner330.062			
213	Male	Yes	Sat	Dinner	4	25.89	5.16	5.16 MaleYesSatDinner425.895.16			
214	Male	No	Sat	Dinner	4	48.33	9.67	9 MaleNoSatDinner448.339			
215	Female	Yes	Sat	Dinner	2	13.27	2.56	2.5 FemaleYesSatDinner213.272.5			
216	Female	Yes	Sat	Dinner	3	28.17	5.63	6.5 FemaleYesSatDinner328.176.5			
217	Female	Yes	Sat	Dinner	2	12.9	2.11	1.1 FemaleYesSatDinner212.91.1			
218	Male	Yes	Sat	Dinner	5	28.15	3.36	3.36 MaleYesSatDinner528.153			

Remove duplicate data.

	A	B	C	D	E	F	G	H	I	J	K
1	sex	smoker	day	time	size	total_bill	tip	sex	Gender of the customer		
2	Female	No	Sun	Dinner	2	16.99	1.01	smoker	Indicates if the customer is a smoker or not		
3	Male	No	Sun	Dinner	3	10.34	1.66				
4	Male	No	Sun	Dinner	3	21.01	3.5				
5	Male	No	Sun	Dinner	2	23.68	3.31				
6	Female	No	Sun	Dinner	4	24.59	3.61				
7	Male	No	Sun	Dinner	4	25.29	4.09				
8	Male	No	Sun	Dinner	2	8.77	1.75				
9	Male	No	Sun	Dinner	4	26.88	3.12				
10	Male	No	Sun	Dinner	2	15.04	1.96				
11	Male	No	Sun	Dinner	2	14.78	3.23				
12	Male	No	Sun	Dinner	2	10.27	1.71				
13	Female	No	Sun	Dinner	4	35.26	3.31				
14	Male	No	Sun	Dinner	2	15.42	1.57				
15	Male	No	Sun	Dinner	4	18.43	3				
16	Female	No	Sun	Dinner	2	14.83	3.02				
17	Male	No	Sun	Dinner	2	21.58	3.92				
18	Female	No	Sun	Dinner	3	10.33	1.67				
19	Male	No	Sun	Dinner	3	15.29	3.71				
20	Female	No	Sun	Dinner	3	15.97	3.5				
21	Male	No	Sat	Dinner	3	20.65	3.35				
22	Male	No	Sat	Dinner	2	17.92	4.08				
23	Female	No	Sat	Dinner	2	20.29	2.75				

One record is deleted.

	A	B	C	D	E	F	G	H	I	J	K
1	sex	smoker	day	time	size	total_bill	tip	sex	Gender of the customer		
2	Female	No	Sun	Dinner	2	16.99	1.01	smoker	Indicates if the customer is a smoker or not		
3	Male	No	Sun	Dinner	3	10.34	1.66				
4	Male	No	Sun	Dinner	3	21.01	3.5				
5	Male	No	Sun	Dinner	2	23.68	3.31				
6	Female	No	Sun	Dinner	4	24.59	3.61				
7	Male	No	Sun	Dinner	4	25.29	4.09				
8	Male	No	Sun	Dinner	2	8.77	1.75				
9	Male	No	Sun	Dinner	4	26.88	3.12				
10	Male	No	Sun	Dinner	2	15.04	1.96				
11	Male	No	Sun	Dinner	2	14.78	3.23				
12	Male	No	Sun	Dinner	2	10.27	1.71				
13	Female	No	Sun	Dinner	4	35.26	3.31				
14	Male	No	Sun	Dinner	2	15.42	1.57				
15	Male	No	Sun	Dinner	4	18.43	3				
16	Female	No	Sun	Dinner	2	14.83	3.02				
17	Male	No	Sun	Dinner	2	21.58	3.92				
18	Female	No	Sun	Dinner	3	10.33	1.67				
19	Male	No	Sun	Dinner	3	16.29	3.71				
20	Female	No	Sun	Dinner	3	16.97	3.5				
21	Male	No	Sat	Dinner	3	20.65	3.35				
22	Male	No	Sat	Dinner	2	17.92	4.08				
23	Female	No	Sat	Dinner	2	20.29	2.75				

3. Encode the categorical variables to numeric values using IF conditions

There are four categorical columns – Sex, Smoker, Day and Time.

- New column ‘Sex_numerical’ is added to covert ‘Sex’ into numerical values.

Male = 1

Female=0

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	sex	smoker	day	time	size	total_bill	tip	sex numerical						
2	Female	No	Sun	Dinner		2	16.99	1.01	=IF(A2="Female",0,1)					1.01
3	Male	No	Sun	Dinner	3	10.34	1.66	1						1.66
4	Male	No	Sun	Dinner	3	21.01	3.5	1						3.5
5	Male	No	Sun	Dinner	2	23.68	3.31	1						3.31
6	Female	No	Sun	Dinner	4	24.59	3.61	0						3.61
7	Male	No	Sun	Dinner	4	25.29	4.71	1						4.71
8	Male	No	Sun	Dinner	2	8.77	2	1						2
9	Male	No	Sun	Dinner	4	26.88	3.12	1						3.12
10	Male	No	Sun	Dinner	2	15.04	1.96	1						1.96
11	Male	No	Sun	Dinner	2	14.78	3.23	1						3.23
12	Male	No	Sun	Dinner	2	10.27	1.71	1						1.71
13	Female	No	Sun	Dinner	4	35.26	5	0						5
14	Male	No	Sun	Dinner	2	15.42	1.57	1						1.57
15	Male	No	Sun	Dinner	4	18.43	3	1						3
16	Female	No	Sun	Dinner	2	14.83	3.02	0						3.02
17	Male	No	Sun	Dinner	2	21.58	3.92	1						3.92
18	Female	No	Sun	Dinner	3	10.33	1.67	0						1.67
19	Male	No	Sun	Dinner	3	16.29	3.71	1						3.71
20	Female	No	Sun	Dinner	3	16.97	3.5	0						3.5
21	Male	No	Sat	Dinner	3	20.65	3.35	1						3.35
22	Male	No	Sat	Dinner	2	17.92	4.08	1						4.08
23	Female	No	Sat	Dinner	2	20.29	2.75	0						2.75

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	sex	smoker	day	time	size	total_bill	tip	sex numerical						
2	Female	No	Sun	Dinner	2	16.99	1.01	0						1.01
3	Male	No	Sun	Dinner	3	10.34	1.66	1						1.66
4	Male	No	Sun	Dinner	3	21.01	3.5	1						3.5
5	Male	No	Sun	Dinner	2	23.68	3.31	1						3.31
6	Female	No	Sun	Dinner	4	24.59	3.61	0						3.61
7	Male	No	Sun	Dinner	4	25.29	4.71	1						4.71
8	Male	No	Sun	Dinner	2	8.77	2	1						2
9	Male	No	Sun	Dinner	4	26.88	3.12	1						3.12
10	Male	No	Sun	Dinner	2	15.04	1.96	1						1.96
11	Male	No	Sun	Dinner	2	14.78	3.23	1						3.23
12	Male	No	Sun	Dinner	2	10.27	1.71	1						1.71
13	Female	No	Sun	Dinner	4	35.26	5	0						5
14	Male	No	Sun	Dinner	2	15.42	1.57	1						1.57
15	Male	No	Sun	Dinner	4	18.43	3	1						3
16	Female	No	Sun	Dinner	2	14.83	3.02	0						3.02
17	Male	No	Sun	Dinner	2	21.58	3.92	1						3.92
18	Female	No	Sun	Dinner	3	10.33	1.67	0						1.67
19	Male	No	Sun	Dinner	3	16.29	3.71	1						3.71
20	Female	No	Sun	Dinner	3	16.97	3.5	0						3.5
21	Male	No	Sat	Dinner	3	20.65	3.35	1						3.35
22	Male	No	Sat	Dinner	2	17.92	4.08	1						4.08
23	Female	No	Sat	Dinner	2	20.29	2.75	0						2.75

- b. New column 'Smoker_numerical' is added to convert 'Smoker' into numerical values.

Yes = 1

No = 0

Restaurant tips dataset - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	sex	smoker	day	time	size	total_bill	tip	sex_numerical	smoker_numerical					
2	Female	No	Sun	Dinner	2	16.99	1.01	0	=IF(B2="No",0,1)					1.01
3	Male	No	Sun	Dinner	3	10.34	1.66	1						1.66
4	Male	No	Sun	Dinner	3	21.01	3.5	1						3.5
5	Male	No	Sun	Dinner	2	23.68	3.31	1						3.31
6	Female	No	Sun	Dinner	4	24.59	3.61	0						3.61
7	Male	No	Sun	Dinner	4	25.29	4.71	1						4.71
8	Male	No	Sun	Dinner	2	8.77	2	1						2
9	Male	No	Sun	Dinner	4	26.88	3.12	1						3.12
10	Male	No	Sun	Dinner	2	15.04	1.96	1						1.96
11	Male	No	Sun	Dinner	2	14.78	3.23	1						3.23
12	Male	No	Sun	Dinner	2	10.27	1.71	1						1.71
13	Female	No	Sun	Dinner	4	35.26	5	0						5
14	Male	No	Sun	Dinner	2	15.42	1.57	1						1.57
15	Male	No	Sun	Dinner	4	18.43	3	1						3
16	Female	No	Sun	Dinner	2	14.83	3.02	0						3.02
17	Male	No	Sun	Dinner	2	21.58	3.92	1						3.92
18	Female	No	Sun	Dinner	3	10.33	1.67	0						1.67
19	Male	No	Sun	Dinner	3	16.29	3.71	1						3.71
20	Female	No	Sun	Dinner	3	16.97	3.5	0						3.5
21	Male	No	Sat	Dinner	3	20.65	3.35	1						3.35
22	Male	No	Sat	Dinner	2	17.92	4.08	1						4.08
23	Female	No	Sat	Dinner	2	20.29	2.75	0						2.75

Restaurant tips dataset - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	sex	smoker	day	time	size	total_bill	tip	sex_numerical	smoker_numerical					
47	Male	No	Sun	Dinner	2	18.29	3	1	0					3
48	Male	No	Sun	Dinner	2	22.23	5	1	0					5
49	Male	No	Sun	Dinner	4	32.4	6	1	0					6
50	Male	No	Sun	Dinner	3	28.55	2.05	1	0					2.05
51	Male	No	Sun	Dinner	2	18.04	3	1	0					3
52	Male	No	Sun	Dinner	2	12.54	2.5	1	0					2.5
53	Female	No	Sun	Dinner	2	10.29	2.6	0	0					2.6
54	Female	No	Sun	Dinner	4	34.81	5.2	0	0					5.2
55	Male	No	Sun	Dinner	2	9.94	1.56	1	0					1.56
56	Male	No	Sun	Dinner	4	25.56	4.34	1	0					4.34
57	Male	No	Sun	Dinner	2	19.49	3.51	1	0					3.51
58	Male	Yes	Sat	Dinner	4	38.01	3	1	1					3
59	Female	No	Sat	Dinner	2	26.41	1.5	0	0					1.5
60	Male	Yes	Sat	Dinner	2	11.24	1.76	1	1					1.76
61	Male	No	Sat	Dinner	4	48.27	6.73	1	0					6.73
62	Male	Yes	Sat	Dinner	2	20.29	3.21	1	1					3.21
63	Male	Yes	Sat	Dinner	2	13.81	2	1	1					2
64	Male	Yes	Sat	Dinner	2	11.02	1.98	1	1					1.98
65	Male	Yes	Sat	Dinner	4	18.29	3.76	1	1					3.76
66	Male	No	Sat	Dinner	3	17.59	2.64	1	0					2.64
67	Male	No	Sat	Dinner	3	20.08	3.15	1	0					3.15
68	Female	No	Sat	Dinner	2	16.45	2.47	0	0					2.47

- c. New column 'Day_numerical' is added to convert 'Day' into numerical values using label encoding.

Thur = 1

Fri = 2

Sat = 3

Sun = 4

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	sex	smoker	day	time	size	total_bill	tip	sex_numerical	smoker_numerical	day_numerical	tip			
1	Female	No	Sun	Dinner	2	16.99	1.01	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))				
2	Female	No	Sun	Dinner	3	10.34	1.66	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	66	sex		
3	Male	No	Sun	Dinner	3	21.01	3.5	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.5	smoker	day	
4	Male	No	Sun	Dinner	2	23.68	3.31	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.31			
5	Male	No	Sun	Dinner	4	24.59	3.61	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.61	time		
6	Female	No	Sun	Dinner	4	25.29	4.71	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	4.71	size		
7	Male	No	Sun	Dinner	2	8.77	2	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	2		total bill	
8	Male	No	Sun	Dinner	4	26.88	3.12	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.12	tip		
9	Male	No	Sun	Dinner	2	15.04	1.96	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	1.96			
10	Male	No	Sun	Dinner	2	14.78	3.23	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.23			
11	Male	No	Sun	Dinner	2	10.27	1.71	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	1.71			
12	Male	No	Sun	Dinner	4	35.26	5	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	5			
13	Female	No	Sun	Dinner	2	15.42	1.57	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	1.57			
14	Male	No	Sun	Dinner	4	18.43	3	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3			
15	Female	No	Sun	Dinner	2	14.83	3.02	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.02			
16	Male	No	Sun	Dinner	2	21.58	3.92	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.92			
17	Female	No	Sun	Dinner	3	10.33	1.67	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	1.67			
18	Male	No	Sun	Dinner	3	16.29	3.71	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.71			
19	Female	No	Sun	Dinner	3	16.97	3.5	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.5			
20	Male	No	Sat	Dinner	3	20.65	3.35	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.35			
21	Male	No	Sat	Dinner	2	17.92	4.08	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	4.08			
22	Female	No	Sat	Dinner	2	20.29	2.75	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	2.75			

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	sex	smoker	day	time	size	total_bill	tip	sex_numerical	smoker_numerical	day_numerical	tip			
1	Female	No	Sun	Dinner	2	16.99	1.01	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	1.01			
2	Female	No	Sun	Dinner	3	10.34	1.66	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	1.66	sex		
3	Male	No	Sun	Dinner	3	21.01	3.5	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.5	smoker	day	
4	Male	No	Sun	Dinner	2	23.68	3.31	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.31			
5	Male	No	Sun	Dinner	4	24.59	3.61	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.61	time		
6	Female	No	Sun	Dinner	4	25.29	4.71	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	4.71	size		
7	Male	No	Sun	Dinner	2	8.77	2	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	2		total bill	
8	Male	No	Sun	Dinner	4	26.88	3.12	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.12	tip		
9	Male	No	Sun	Dinner	2	15.04	1.96	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	1.96			
10	Male	No	Sun	Dinner	2	14.78	3.23	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.23			
11	Male	No	Sun	Dinner	2	10.27	1.71	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	1.71			
12	Male	No	Sun	Dinner	4	35.26	5	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	5			
13	Female	No	Sun	Dinner	2	15.42	1.57	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	1.57			
14	Male	No	Sun	Dinner	4	18.43	3	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3			
15	Female	No	Sun	Dinner	2	14.83	3.02	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.02			
16	Male	No	Sun	Dinner	2	21.58	3.92	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.92			
17	Female	No	Sun	Dinner	3	10.33	1.67	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	1.67			
18	Male	No	Sun	Dinner	3	16.29	3.71	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.71			
19	Female	No	Sun	Dinner	3	16.97	3.5	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.5			
20	Male	No	Sat	Dinner	3	20.65	3.35	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	3.35			
21	Male	No	Sat	Dinner	2	17.92	4.08	1	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	4.08			
22	Female	No	Sat	Dinner	2	20.29	2.75	0	0	=IF(C2="Thur",1,IF(C2="Fri",2,IF(C2="Sat",3,IF(C2="Sun",4,0))))	2.75			

Average: 2.732510288 Count: 243 Sum: 664

- d. New column 'Time_numerical' is added to convert 'Time' into numerical values.

Dinner = 1

Lunch = 0

Restaurant tips dataset - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	sex	smoker	day	time	size	total_bill	tip	sex_numerical	smoker_numerical	day_numerical	time_numerical	tip		
2	Female	No	Sun	Dinner	2	16.99	1.01	0	0	0	=IF(D2="Dinner",1,0)	1.66		
3	Male	No	Sun	Dinner	3	10.34	1.66	1	0	4		3.5	sex	
4	Male	No	Sun	Dinner	3	21.01	3.5	1	0	4		3.31	smoker	
5	Male	No	Sun	Dinner	2	23.68	3.31	1	0	4		3.61	day	
6	Female	No	Sun	Dinner	4	24.59	3.61	0	0	4		4.71	time	
7	Male	No	Sun	Dinner	4	25.29	4.71	1	0	4		2	size	
8	Male	No	Sun	Dinner	2	8.77	2	1	0	4		3.12	total_bill	
9	Male	No	Sun	Dinner	4	26.88	3.12	1	0	4		1.96	tip	
10	Male	No	Sun	Dinner	2	15.04	1.96	1	0	4		3.23		
11	Male	No	Sun	Dinner	2	14.78	3.23	1	0	4		1.71		
12	Male	No	Sun	Dinner	2	10.27	1.71	1	0	4		5		
13	Female	No	Sun	Dinner	4	35.26	5	0	0	4		1.57		
14	Male	No	Sun	Dinner	2	15.42	1.57	1	0	4		3		
15	Male	No	Sun	Dinner	4	18.43	3	1	0	4		3.02		
16	Female	No	Sun	Dinner	2	14.83	3.02	0	0	4		3.92		
17	Male	No	Sun	Dinner	2	21.58	3.92	1	0	4		1.67		
18	Female	No	Sun	Dinner	3	10.33	1.67	0	0	4		3.71		
19	Male	No	Sun	Dinner	3	16.29	3.71	1	0	4		3.35		
20	Female	No	Sun	Dinner	3	16.97	3.5	0	0	4		4.08		
21	Male	No	Sat	Dinner	3	20.65	3.35	1	0	3		2.75		
22	Male	No	Sat	Dinner	2	17.92	4.08	1	0	3				
23	Female	No	Sat	Dinner	2	20.29	2.75	0	0	3				

Restaurant tips dataset - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	sex	smoker	day	time	size	total_bill	tip	sex_numerical	smoker_numerical	day_numerical	time_numerical	tip		
69	Female	Yes	Sat	Dinner	1	3.07	1	0	1	3	1	1	1	
70	Male	No	Sat	Dinner	2	20.23	2.01	1	0	3	1	2.01		
71	Male	Yes	Sat	Dinner	2	15.01	2.09	1	1	3	1	2.09		
72	Male	No	Sat	Dinner	2	12.02	1.97	1	0	3	1	1.97		
73	Female	No	Sat	Dinner	3	17.07	3	0	0	3	1	3		
74	Female	Yes	Sat	Dinner	2	26.86	3.14	0	1	3	1	3.14		
75	Female	Yes	Sat	Dinner	2	25.28	5	0	1	3	1	5		
76	Female	No	Sat	Dinner	2	14.73	2.2	0	0	3	1	2.2		
77	Male	No	Sat	Dinner	2	10.51	1.25	1	0	3	1	1.25		
78	Male	Yes	Sat	Dinner	2	17.92	3.08	1	1	3	1	3.08		
79	Male	No	Thur	Lunch	4	27.2	4	1	0	1	0	4		
80	Male	No	Thur	Lunch	2	22.76	3	1	0	1	0	3		
81	Male	No	Thur	Lunch	2	17.29	2.71	1	0	1	0	2.71		
82	Male	Yes	Thur	Lunch	2	19.44	3	1	1	1	0	3		
83	Male	No	Thur	Lunch	2	16.66	3.4	1	0	1	0	3.4		
84	Female	No	Thur	Lunch	1	10.07	1.83	0	0	1	0	1.83		
85	Male	Yes	Thur	Lunch	2	32.68	5	1	1	1	0	5		
86	Male	No	Thur	Lunch	2	15.98	2.03	1	0	1	0	2.03		
87	Female	No	Thur	Lunch	4	34.83	5.17	0	0	1	0	5.17		
88	Male	No	Thur	Lunch	2	13.03	2	1	0	1	0	2		
89	Male	No	Thur	Lunch	2	18.28	4	1	0	1	0	4		
90	Male	No	Thur	Lunch	2	24.71	5.85	1	0	1	0	5.85		

4. For each independent numeric value, find its correlation coefficient with respect to the tip.

The screenshot shows the Microsoft Excel ribbon with the 'Data' tab selected. In the 'Data Tools' group, the 'Data Analysis' button is highlighted. A 'Data Analysis' dialog box is open, listing various statistical tools. The 'Descriptive Statistics' option is selected. The data range is set to \$G\$2:\$K\$23, and the output range is \$N\$2. The 'Labels in first row' checkbox is checked.

The screenshot shows the Microsoft Excel ribbon with the 'Data' tab selected. In the 'Data Tools' group, the 'Data Analysis' button is highlighted. A 'Correlation' dialog box is open, showing the input range \$E\$2:\$K\$245, grouped by columns, and the output range \$N\$3. The 'Labels in first row' checkbox is checked.

The screenshot shows the Microsoft Excel ribbon with the 'Home' tab selected. The data range \$K\$2:\$P\$9 has been copied and pasted into the cells \$N\$2:\$U\$9. The resulting table is a Correlation Matrix:

	size	total_bill	sex_numerical	smoker_numerical	day_numerical	time_numerical	tip
size	1						
total_bill	0.597588931	1					
sex_numerical	0.083248017	0.141349744	1				
smoker_numerical	-0.130564411	0.090136102	0.009930188	1			
day_numerical	0.162524711	0.169978112	0.224387598	-0.025007759	1		
time_numerical	0.100045303	0.179231854	0.198128623	0.063911231	0.873133015	1	
tip	0.488400395	0.674997857	0.085273975	0.00976275	0.131797526	0.11759639	1

5. Find the features that are independent and dependent.

As per objective of the problem, we need to prepare a model to predict restaurant tips given input values with the mathematical equation for predicting the tips value.

So, Our **Dependent Variable (Target)** is 'tip', which we want to predict.

And rest of the features including 'size', 'total_bill', 'sex_numerical', 'smoker_numerical', 'day_numerical' and 'time_numerical' are our **Independent Variables**.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1					X(Independent features)						Y(Dependent variable)		
2	sex	smoker	day	time	size	total_bill	sex_numerical	smoker_numerical	day_numerical	time_numerical	tip		
3	Female	No	Sun	Dinner	2	16.99	0	0	4	1	1.01		
4	Male	No	Sun	Dinner	3	10.34	1	0	4	1	1.66	sex	G
5	Male	No	Sun	Dinner	3	21.01	1	0	4	1	3.5	smoker	Ir
6	Male	No	Sun	Dinner	2	23.68	1	0	4	1	3.31	day	D
7	Female	No	Sun	Dinner	4	24.59	0	0	4	1	3.61	time	If
8	Male	No	Sun	Dinner	4	25.29	1	0	4	1	4.71	size	N
9	Male	No	Sun	Dinner	2	8.77	1	0	4	1	2	total bill	B
10	Male	No	Sun	Dinner	4	26.88	1	0	4	1	3.12	tip	T
11	Male	No	Sun	Dinner	2	15.04	1	0	4	1	1.96		
12	Male	No	Sun	Dinner	2	14.78	1	0	4	1	3.23		
13	Male	No	Sun	Dinner	2	10.27	1	0	4	1	1.71		
14	Female	No	Sun	Dinner	4	35.26	0	0	4	1	5		
15	Male	No	Sun	Dinner	2	15.42	1	0	4	1	1.57		
16	Male	No	Sun	Dinner	4	18.43	1	0	4	1	3		
17	Female	No	Sun	Dinner	2	14.83	0	0	4	1	3.02		
18	Male	No	Sun	Dinner	2	21.58	1	0	4	1	3.92		
19	Female	No	Sun	Dinner	3	10.33	0	0	4	1	1.67		
20	Male	No	Sun	Dinner	3	16.29	1	0	4	1	3.71		
21	Female	No	Sun	Dinner	3	16.97	0	0	4	1	3.5		
22	Male	No	Sat	Dinner	3	20.65	1	0	3	1	3.35		
23	Male	No	Sat	Dinner	2	17.92	1	0	3	1	4.08		

6. Identify which predictive problem is needed.

This is a **regression problem** because we are predicting a continuous numerical value ('tip' amount).

7. Build an appropriate model with the dataset.

Building the Regression Model

Restaurant tips dataset - Excel

Data Analysis Tools

- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram
- Mean, Average
- Random Number Generation
- Rank and Percentile
- Sampling
- t-test: Paired Two Sample for Means
- t-test: Two-Sample Assuming Equal Variances

Regression

Input

Input Y Range: \$K\$2:\$K\$244

Input X Range: \$E\$2:\$J\$244

Labels

Constant is Zero

Output options

Output Range: \$N\$13

New Worksheet By:

New Workbook

Residuals

Residual Plots

Standardized Residuals

Normal Probability

Normal Probability Plots

Font

Clipboard

Calibri

Number

Alignment

Format Painter

Wrap Text

Merge & Center

Conditional Format as

Formatting

Cells

Insert

Delete

Format

AutoSum

Find & Select

Editing

SUMMARY OUTPUT

	Regression				
Multiple R	0.684226461				
R Square	0.46816585				
Adjusted R Square	0.454644643				
Standard Error	1.022798786				
Observations	243				

ANOVA

	df	SS	MS	F	Significance F
Regression	6	217.3281192	36.22115321	34.62455999	6.98847E-30
Residual	236	246.8836964	1.046117358		
Total	242	464.2118156			

Regression Statistics

Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.675080945	0.254241143	2.655278129	0.008463999	0.174208898	0.175952993	0.175952993
size	0.174984371	0.089388965	1.957775507	0.051454839	-0.00109846	0.351027206	0.351027206
total_bill	0.094263519	0.009561371	9.858786532	2.01057E-19	0.075426978	0.113100059	0.113100059
sex_numerical	-0.03663251	0.141648177	-0.25861621	0.796156785	-0.3156889	0.242423864	0.242423864
smoker_numerical	-0.07284483	0.141051329	-0.51643846	0.060603781	-0.35072488	0.205036218	0.205036218
day_numerical	0.053005305	0.120582605	0.439576112	0.660645986	-0.18455049	0.290561098	0.290561098
time_numerical	-0.11600626	0.308553216	-0.37596841	0.0707278124	-0.72387672	0.491864201	0.491864201

8. Calculate the predicted tips values.

Create a new column named predicted_tip, which is calculated using the coefficients obtained from the regression model.

```
= { Intercept_coeff(P29) +
    size_coeff(P30) * size(E3) +
    total_bill_coeff(P31) * total_bill(F3) +
    Sex_numerical_coeff(P32) * Sex_numerical(G3) +
    Smoker_numerical_coeff(P33) * Smoker_numerical(H3) +
    Day_numerical_coeff(P34) * Day_numerical(I3) +
    Time_numerical_coeff(P35) * Time_numerical(J3) }
```

	D	E	F	G	H	I	J	K	L	M	N	O
	X(Independent features)							Y(Dependent variable)	OUTPUT			
1								tip				
2	time	size	total_bill	sex_numerical	smoker_numerical	day_numerical	time_numerical					
3	Dinner	2	16.99	0	0	4	1	1.01	=(\$P\$29+\$P\$30*\$E3+\$P\$31*\$F3+\$P\$32*\$G3+\$P\$33*\$H3+\$P\$34*\$I3+\$P\$35*\$J3)			
4	Dinner	3	10.34	1	0	4	1	1.66	total_bill			
5	Dinner	3	21.01	1	0	4	1	3.5	sex_numerical			
6	Dinner	2	23.68	1	0	4	1	3.31	smoker_numerical			
7	Dinner	4	24.59	0	0	4	1	3.61	day_numerical			
8	Dinner	4	25.29	1	0	4	1	4.71	time_numerical			
9	Dinner	2	8.77	1	0	4	1	2	tip			
10	Dinner	4	26.88	1	0	4	1	3.12				
11	Dinner	2	15.04	1	0	4	1	1.96				
12	Dinner	2	14.78	1	0	4	1	3.23				
13	Dinner	2	10.27	1	0	4	1	1.71	SUMMARY OUTPUT			
14	Dinner	4	35.26	0	0	4	1	5				
15	Dinner	2	15.42	1	0	4	1	1.57	Regression Statistics			
16	Dinner	4	18.43	1	0	4	1	3	Multiple R			
17	Dinner	2	14.83	0	0	4	1	3.02	R Square			
18	Dinner	2	21.58	1	0	4	1	3.92	Adjusted R Square			
19	Dinner	3	10.33	0	0	4	1	1.67	Standard Error			
20	Dinner	3	16.29	1	0	4	1	3.71	Observations			
21	Dinner	3	16.97	0	0	4	1	3.5	ANOVA			
22	Dinner	3	20.65	1	0	3	1	3.35				
23	Dinner	2	17.92	1	0	3	1	4.08	Regression			
24	Dinner	2	20.29	0	0	3	1	2.75	Residual			
25	Dinner	2	15.77	0	0	3	1	2.23				

	D	E	F	G	H	I	J	K	L	M	N	O
	X(Independent features)							Y(Dependent variable)	OUTPUT			
1								tip				
2	time	size	total_bill	sex_numerical	smoker_numerical	day_numerical	time_numerical					
3	Dinner	2	16.99	0	0	4	1	1.01	2.732363723			
4	Dinner	3	10.34	1	0	4	1	1.66	2.23827711			
5	Dinner	3	21.01	1	0	4	1	3.5	3.243549556			
6	Dinner	2	23.68	1	0	4	1	3.31	3.319367416			
7	Dinner	4	24.59	0	0	4	1	3.61	3.790154669			
8	Dinner	4	25.29	1	0	4	1	4.71	3.822218156			
9	Dinner	2	8.77	1	0	4	1	2	1.914930752			
10	Dinner	4	26.88	1	0	4	1	3.12	3.972019767			
11	Dinner	2	15.04	1	0	4	1	1.96	2.505657859			
12	Dinner	2	14.78	1	0	4	1	3.23	2.481163998			
13	Dinner	2	10.27	1	0	4	1	1.71	2.056253026			
14	Dinner	4	35.26	0	0	4	1	5	4.795827115			
15	Dinner	2	15.42	1	0	4	1	1.57	2.541459502			
16	Dinner	4	18.43	1	0	4	1	3	3.175904288			
17	Dinner	2	14.83	0	0	4	1	3.02	2.519759648			
18	Dinner	2	21.58	1	0	4	1	3.92	3.121822976			
19	Dinner	3	10.33	0	0	4	1	1.67	2.27121869			
20	Dinner	3	16.29	1	0	4	1	3.71	2.798855465			
21	Dinner	3	16.97	0	0	4	1	3.5	2.89860847			
22	Dinner	3	20.65	1	0	3	1	3.35	3.147310982			
23	Dinner	2	17.92	1	0	3	1	4.08	2.714675398			
24	Dinner	2	20.29	0	0	3	1	2.75	2.971851499			
25	Dinner	2	15.77	0	0	3	1	2.23	2.546000379			

Average: 3.001172612 Count: 243 Sum: 729.2849446

90%

9. Calculate the RMSE (Root Mean Square Error) of the model.

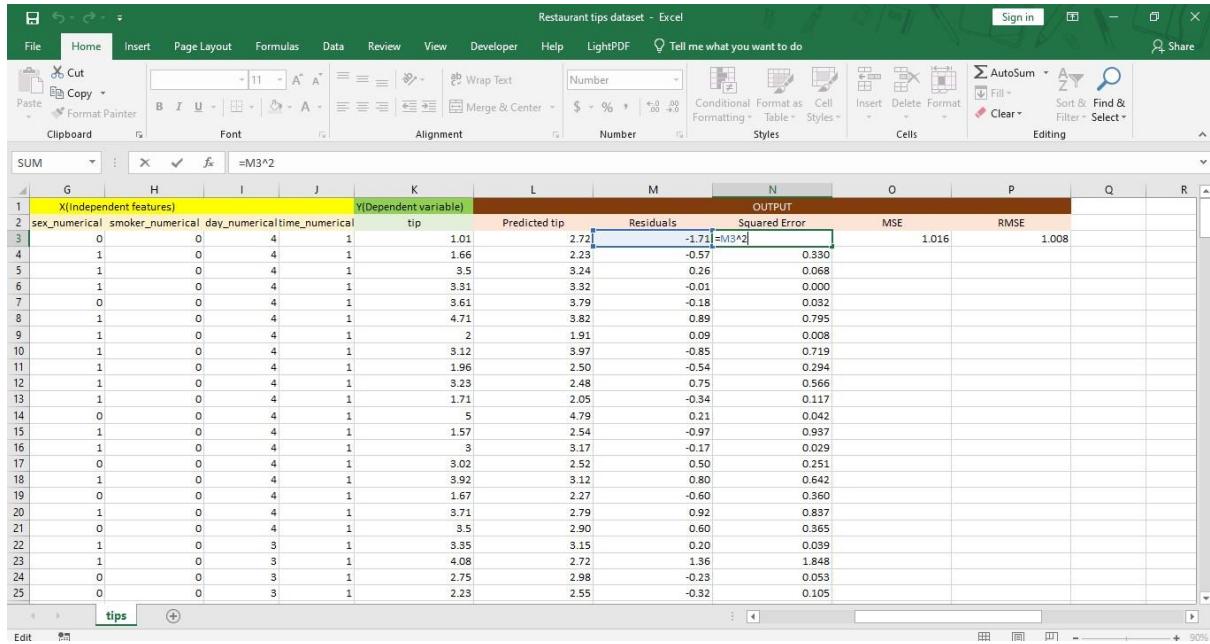
To calculate RMSE, first we need to calculate errors (Residuals).

M3	D	E	F	G	H	I	J	K	L	M	N	O
1		X(Independent features)						Y(Dependent variable)	OUTPUT			
2	time	size	total_bill	sex_numerical	smoker_numerical	day_numerical	time_numerical	tip	Predicted tip	Residuals		
3	Dinner	2	16.99	0	0	4	1	1.01	2.723263723	=\\$K3-\$L3		
4	Dinner	3	10.34	1	0	4	1	1.66	2.23827711			
5	Dinner	3	21.01	1	0	4	1	3.5	3.243549556			
6	Dinner	2	23.68	1	0	4	1	3.31	3.31967416			
7	Dinner	4	24.59	0	0	4	1	3.61	3.790154669			
8	Dinner	4	25.29	1	0	4	1	4.71	3.822218156			
9	Dinner	2	8.77	1	0	4	1	2	1.914930752			
10	Dinner	4	26.88	1	0	4	1	3.12	3.972019767			
11	Dinner	2	15.04	1	0	4	1	1.96	2.505657859			
12	Dinner	2	14.78	1	0	4	1	3.23	2.481161998			
13	Dinner	2	10.27	1	0	4	1	1.71	2.056253026			
14	Dinner	4	35.26	0	0	4	1	5	4.795427115			
15	Dinner	2	15.42	1	0	4	1	1.57	2.541459502			
16	Dinner	4	18.43	1	0	4	1	3	3.175904288			
17	Dinner	2	14.83	0	0	4	1	3.02	2.519759648			
18	Dinner	2	21.58	1	0	4	1	3.92	3.121822976			
19	Dinner	3	10.33	0	0	4	1	1.67	2.271221869			
20	Dinner	3	16.29	1	0	4	1	3.71	2.798855465			
21	Dinner	3	16.97	0	0	4	1	3.5	2.89680847			
22	Dinner	3	20.65	1	0	3	1	3.35	3.147310982			
23	Dinner	2	17.92	1	0	3	1	4.08	2.714675398			
24	Dinner	2	20.29	0	0	3	1	2.75	2.971851499			
25	Dinner	2	15.77	0	0	3	1	2.23	2.546000379			

M3	D	E	F	G	H	I	J	K	L	M	N	O
1		X(Independent features)						Y(Dependent variable)	OUTPUT			
2	time	size	total_bill	sex_numerical	smoker_numerical	day_numerical	time_numerical	tip	Predicted tip	Residuals		
3	Dinner	2	16.99	0	0	4	1	1.01	2.723263723	-1.713263723		
4	Dinner	3	10.34	1	0	4	1	1.66	2.23827711	-0.57827711		
5	Dinner	3	21.01	1	0	4	1	3.5	3.243549556	0.256405444		
6	Dinner	2	23.68	1	0	4	1	3.31	3.31967416	-0.00967416		
7	Dinner	4	24.59	0	0	4	1	3.61	3.790154669	-0.180154669		
8	Dinner	4	25.29	1	0	4	1	4.71	3.822218156	0.887781844		
9	Dinner	2	8.77	1	0	4	1	2	1.914930752	0.085069248		
10	Dinner	4	26.88	1	0	4	1	3.12	3.972019767	-0.852019767		
11	Dinner	2	15.04	1	0	4	1	1.96	2.505657859	-0.545657859		
12	Dinner	2	14.78	1	0	4	1	3.23	2.481161998	0.748838002		
13	Dinner	2	10.27	1	0	4	1	1.71	2.056253026	-0.346253026		
14	Dinner	4	35.26	0	0	4	1	5	4.795427115	0.204572885		
15	Dinner	2	15.42	1	0	4	1	1.57	2.541459502	-0.971459502		
16	Dinner	4	18.43	1	0	4	1	3	3.175904288	-0.175904288		
17	Dinner	2	14.83	0	0	4	1	3.02	2.519759648	0.500240352		
18	Dinner	2	21.58	1	0	4	1	3.92	3.121822976	0.798177024		
19	Dinner	3	10.33	0	0	4	1	1.67	2.271221869	-0.601221869		
20	Dinner	3	16.29	1	0	4	1	3.71	2.798855465	0.911144535		
21	Dinner	3	16.97	0	0	4	1	3.5	2.89680847	0.60319153		
22	Dinner	3	20.65	1	0	3	1	3.35	3.147310982	0.202689018		
23	Dinner	2	17.92	1	0	3	1	4.08	2.714675398	1.365324602		
24	Dinner	2	20.29	0	0	3	1	2.75	2.971851499	-0.221851499		
25	Dinner	2	15.77	0	0	3	1	2.23	2.546000379	-0.316000379		

Average: 0.0012422 Count: 243 Sum: 0.295055379

Now, we need to find square of these errors.

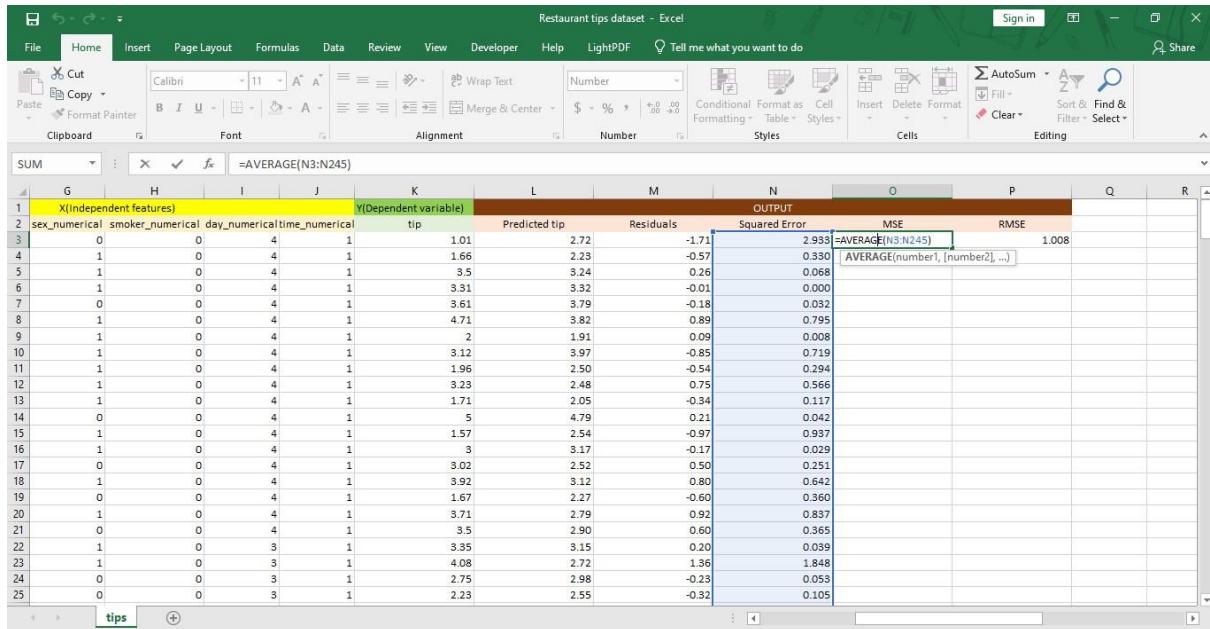


The screenshot shows a Microsoft Excel spreadsheet titled "Restaurant tips dataset - Excel". The data is organized into several columns:

- Columns G through K are labeled "X(Independent features)" and contain numerical values.
- Column L is labeled "Y(Dependent variable)" and contains the value "tip".
- Column M is labeled "Predicted tip".
- Column N is labeled "Residuals".
- Column O is labeled "OUTPUT".
- Column P is labeled "Squared Error".
- Column Q is labeled "MSE".
- Column R is labeled "RMSE".

In the formula bar, the formula `=M3^2` is entered into cell P3. The cell P3 contains the value `-1.71`. The formula `=M3^2` is also highlighted in the formula bar.

Now, we need to find average of these squared error to find Mean Squared Error (MSE).



The screenshot shows the same Microsoft Excel spreadsheet as before, but now the formula `=AVERAGE(N3:N245)` is entered into cell P3. The cell P3 now contains the value `2.933`. The formula `=AVERAGE(N3:N245)` is highlighted in the formula bar.

Ultimately, RMSE is found taking square root of MSE.

The screenshot shows a Microsoft Excel spreadsheet titled "Restaurant tips dataset - Excel". The data is organized into several columns:

	X(Independent features)				Y(Dependent variable)		OUTPUT			
	sex_numerical	smoker_numerical	day_numerical	time_numerical	tip	Predicted tip	Residuals	Squared Error	MSE	RMSE
3	0	0	4	1	1.01	2.72	-1.71	2.993	1.016	=SQRT(\$O\$3)
4	1	0	4	1	1.66	2.23	-0.57	0.330		SQRT(number)
5	1	0	4	1	3.5	3.24	0.26	0.068		
6	1	0	4	1	3.31	3.32	-0.01	0.000		
7	0	0	4	1	3.61	3.79	-0.18	0.032		
8	1	0	4	1	4.71	3.82	0.89	0.795		
9	1	0	4	1	2	1.91	0.09	0.008		
10	1	0	4	1	3.12	3.97	-0.85	0.719		
11	1	0	4	1	1.96	2.50	-0.54	0.294		
12	1	0	4	1	3.23	2.48	0.75	0.566		
13	1	0	4	1	1.71	2.05	-0.34	0.117		
14	0	0	4	1	5	4.79	0.21	0.042		
15	1	0	4	1	1.57	2.94	-0.97	0.997		
16	1	0	4	1	3	3.17	-0.17	0.029		
17	0	0	4	1	3.02	2.52	0.50	0.251		
18	1	0	4	1	3.92	3.12	0.80	0.642		
19	0	0	4	1	1.67	2.27	-0.60	0.360		
20	1	0	4	1	3.71	2.79	0.92	0.837		
21	0	0	4	1	3.5	2.90	0.60	0.365		
22	1	0	3	1	3.35	3.15	0.20	0.059		
23	1	0	3	1	4.08	2.72	1.36	1.848		
24	0	0	3	1	2.75	2.98	-0.23	0.053		
25	0	0	3	1	2.23	2.55	-0.32	0.105		