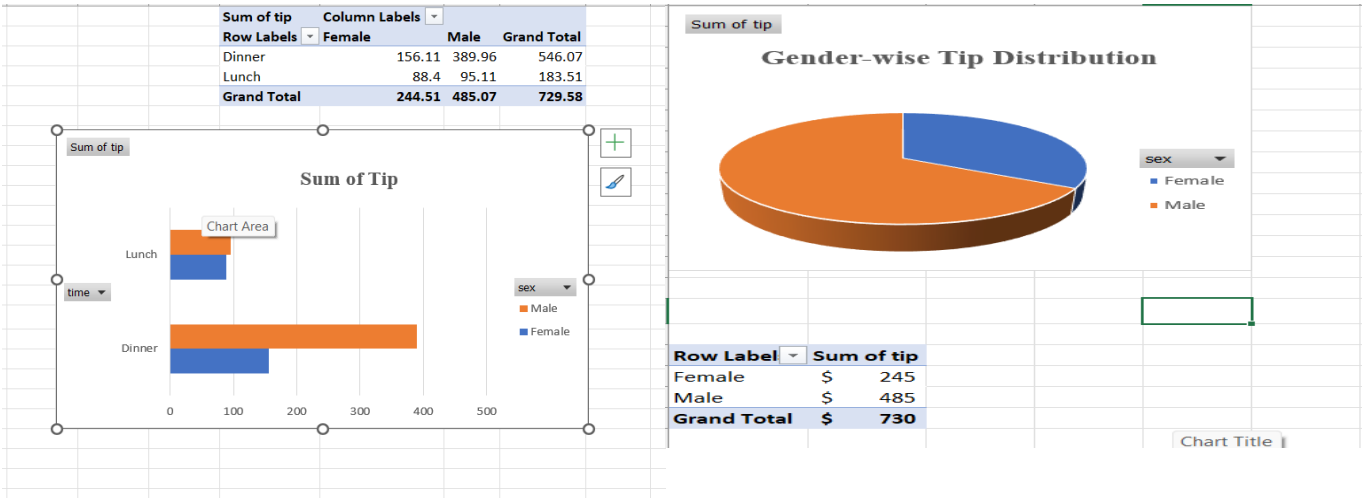
Predicting Restaurant Tips

Figure 1: Data Preparation

sex_num	smoker_num	time_num	day_num	size	total_bill	tip
2	1	1	1	2	$ 16.99	$ 1.01
1	1	1	1	3	$ 10.34	$ 1.66
1	1	1	1	3	$ 21.01	$ 3.50
1	1	1	1	2	$ 23.68	$ 3.31
2	1	1	1	4	$ 24.59	$ 3.61
1	1	1	1	4	$ 25.29	$ 4.71
1	1	1	1	2	$ 8.77	$ 2.00
1	1	1	1	4	$ 26.88	$ 3.12
1	1	1	1	2	$ 15.04	$ 1.96
1	1	1	1	2	$ 14.78	$ 3.23
1	1	1	1	2	$ 10.27	$ 1.71
2	1	1	1	4	$ 35.26	$ 5.00
1	1	1	1	2	$ 15.42	$ 1.57
1	1	1	1	4	$ 18.43	$ 3.00
2	1	1	1	2	$ 14.83	$ 3.02
1	1	1	1	2	$ 21.58	$ 3.92
2	1	1	1	3	$ 10.33	$ 1.67
1	1	1	1	3	$ 16.29	$ 3.71
2	1	1	1	3	$ 16.97	$ 3.50
1	1	1	3	3	$ 20.65	$ 3.35
1	1	1	3	2	$ 17.92	$ 4.08
2	1	1	3	2	$ 20.29	$ 2.75
2	1	1	3	2	$ 15.77	$ 2.23
1	1	1	3	4	$ 39.42	$ 7.58
1	1	1	3	2	$ 19.82	$ 3.18
1	1	1	3	4	$ 17.81	$ 2.34
1	1	1	3	2	$ 13.37	$ 2.00
1	1	1	3	2	$ 12.69	$ 2.00
1	1	1	3	2	$ 21.70	$ 4.30
2	1	1	3	2	$ 19.65	$ 3.00
1	1	1	3	2	$ 9.55	$ 1.45
1	1	1	3	4	$ 18.35	$ 2.50
2	1	1	3	2	$ 15.06	$ 3.00
2	1	1	3	4	$ 20.69	$ 2.45
1	1	1	3	2	$ 17.78	$ 3.27
1	1	1	3	3	$ 24.06	$ 3.60
1	1	1	3	3	$ 16.31	$ 2.00
2	1	1	3	3	$ 16.93	$ 3.07

Categorical features (sex, smoker, day, and time) converted to numeric values using IF statements.

Figure 2: Pivot Table & Graphs



Pivot table was prepared to summarize tips for males and females across different time periods, and a bar chart was created to compare these values.

A pivot table was created to calculate the total tips for males and females, and a pie chart was generated to visualize the share of each gender.

Figure 3: Correlation between the tips and the other variable

Correlation value	
sex_num	-0.08527398
smoker_num	0.00976275
time_num	-0.11759639
day_num	-0.1184303
size	0.488400395
total_bill	0.674997857

Correlation ranges from –1 to +1. A negative correlation means the variable has an inverse relationship with the tips, while a positive correlation means the variable has a direct relationship with the tips.

Figure 4: multiple regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.68474709
R Square	0.468878578
Adjusted R Square	0.455375491
Standard Error	1.022113214
Observations	243

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	217.6589759	36.27649598	34.7238063	5.98352E-30
Residual	236	246.5528398	1.044715423		
Total	242	464.2118156			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.759443397	0.400840472	1.89462754	0.059364253	-0.03023913	1.549125924	-0.03023913	1.549125924
sex_num	0.037183538	0.141110588	0.263506364	0.792390247	-0.240813755	0.315180831	-0.240813755	0.315180831
smoker_num	-0.073687759	0.140070613	-0.526075791	0.599329557	-0.34963623	0.202260713	-0.34963623	0.202260713
time_num	0.088727763	0.197945596	0.448243177	0.654388943	-0.301238294	0.47869382	-0.301238294	0.47869382
day_num	-0.053498843	0.074899968	-0.714270572	0.475765749	-0.201056787	0.094059101	-0.201056787	0.094059101
size	0.173933116	0.0889047	1.956399563	0.051597599	-0.001215086	0.349081318	-0.001215086	0.349081318
total_bill	0.09437403	0.009550949	9.881115107	1.7175E-19	0.075558021	0.113190039	0.075558021	0.113190039

The model is statistically significant overall. Total bill is the strongest predictor. Other variables (sex, smoker, time, day) are not statistically significant.

Figure 5: predicted tips

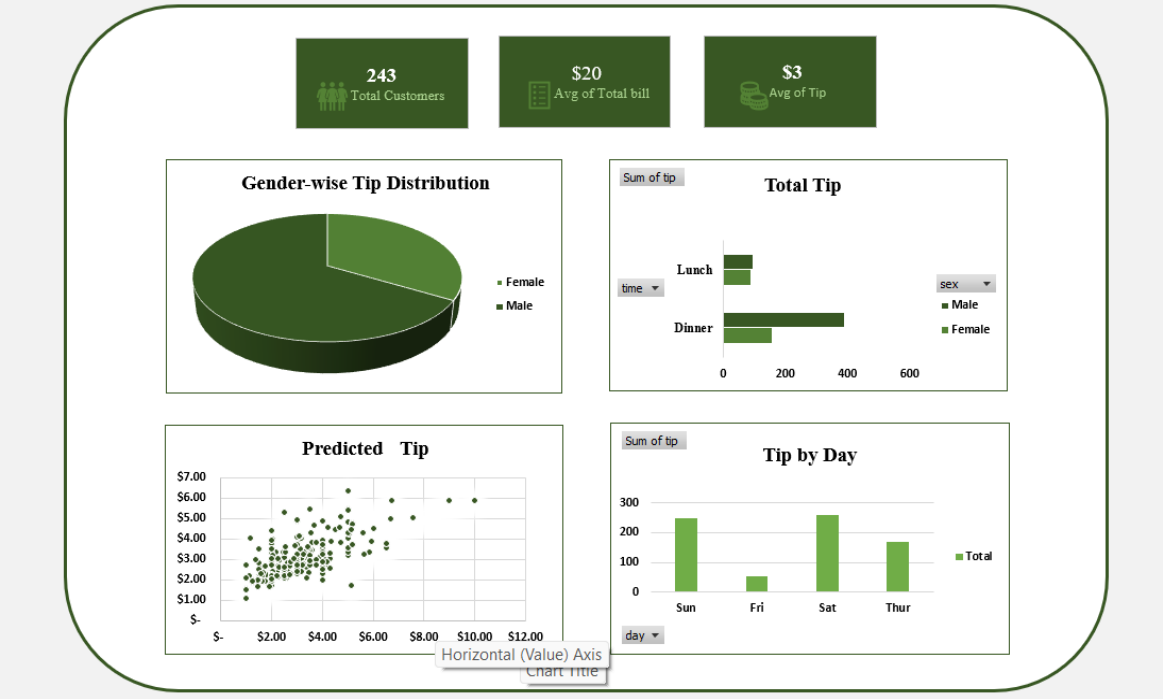
V	W	X	Y	Z	AA	AB	AC
Intercept							**Predicted tip**
0.759443397	0.074367076	0.088727763	-0.073687759	-0.053498843	0.347866232	1.603414773	$ 2.75
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.521799348	0.975827472	$ 2.26
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.521799348	1.982798374	$ 3.26
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.347866232	2.234777035	$ 3.34
0.759443397	0.074367076	0.088727763	-0.073687759	-0.053498843	0.695732464	2.320657402	$ 3.81
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.695732464	2.386719223	$ 3.84
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.347866232	0.827660245	$ 1.93
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.695732464	2.536773931	$ 3.99
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.347866232	1.419385414	$ 2.53
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.347866232	1.394848166	$ 2.50
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.347866232	0.96922129	$ 2.08
0.759443397	0.074367076	0.088727763	-0.073687759	-0.053498843	0.695732464	3.327628304	$ 4.82
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.347866232	1.455247545	$ 2.56
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.695732464	1.739313376	$ 3.19
0.759443397	0.074367076	0.088727763	-0.073687759	-0.053498843	0.347866232	1.399566868	$ 2.54
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.347866232	2.036591571	$ 3.14
0.759443397	0.074367076	0.088727763	-0.073687759	-0.053498843	0.521799348	0.974883732	$ 2.29
0.759443397	0.037183538	0.088727763	-0.073687759	-0.053498843	0.521799348	1.537352952	$ 2.82
0.759443397	0.074367076	0.088727763	-0.073687759	-0.053498843	0.521799348	1.601527292	$ 2.92
0.759443397	0.037183538	0.088727763	-0.073687759	-0.160496528	0.521799348	1.948823723	$ 3.12
0.759443397	0.037183538	0.088727763	-0.073687759	-0.160496528	0.347866232	1.691182621	$ 2.69
0.759443397	0.074367076	0.088727763	-0.073687759	-0.160496528	0.347866232	1.914849072	$ 2.95
0.759443397	0.074367076	0.088727763	-0.073687759	-0.160496528	0.347866232	1.488278456	$ 2.52
0.759443397	0.037183538	0.088727763	-0.073687759	-0.160496528	0.695732464	3.72022427	$ 5.07
0.759443397	0.037183538	0.088727763	-0.073687759	-0.160496528	0.347866232	1.870493278	$ 2.87
0.759443397	0.037183538	0.088727763	-0.073687759	-0.160496	5732464	1.680801478	$ 3.03
0.759443397	0.037183538	0.088727763	-0.073687759	-0.160496528	0.347866232	1.261780784	$ 2.26
0.759443397	0.037183538	0.088727763	-0.073687759	-0.160496528	0.347866232	1.197606443	$ 2.20

Chart Title

ard +

The coefficients and intercept were used to construct the regression equation, and the equation was applied to all input values to generate the predicted tips.

Figure 6: Dashboard



Present data in a clear, visual format.