

Predicting restaurant tips using predictive analytics on Excel

- I have used regression analysis to predict tips.
- The dependent variable to be predicted was tips and independent variables that may impact the tips prediction were sex, smoker, day, time, size, and total bill.
- Data cleaning was done to check missing values, duplicate values, etc. (Luckily the data was already clean).
- I started by plotting graphs to observe any correlation.
- Size and total_bill showed a positive correlation with the tips. Based on the graphs, it was unclear if sex, smoker, day, and time impact tips.
- Getting data ready for the multi-linear regression. The categorical variables were converted to dummy(numeric) variables.
- Multi-linear regression was performed using the data analysis toolpak.
- Sex, smoker, day, and time were excluded from the analysis as their p-value was larger than 5% which could be explained by randomness.
- Multi-linear regression was performed again on the independent variables(size and total_bill) with a p-value less or close to 5%.

- A prediction calculator was built to calculate tips based on size and total_bill.

Predicted Tips= intercept +(coefficient of size * size) + (coefficient of total_bill* total_bill)
 $Y = \text{Constant} + B_1(X_1) + B_2(X_2) + \dots B_n X_n$.

- Root mean square error was calculated which could be used to evaluate the quality of predictions.