Predictive Modeling Using Multiple Regression

AIM: To predict Restaurant Tips from the given dataset.
Tools: Microsoft Excel, Data Analysis Add-in.

Description:

In the dataset in file ***Restaurant tips dataset.xlsx*** contains tips data for different customers. The following are the features in the dataset:

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

Procedure :

1. Data Analysis and Cleaning

- Begin by thoroughly analyzing and cleaning the dataset from the "Restaurant tips dataset.xlsx" file.

2. Regression Analysis with Data Analysis Add-In

- Utilize the Data Analysis add-in in Excel to perform regression analysis on the dataset.

- Identify potential independent and dependent variables based on the features provided, including gender, smoking status, day of the visit, mealtime, number of diners, total bill amount, and tip amount.

3. Feature Selection

- Carefully examine the results of the regression analysis to identify significant independent variables that contribute to predicting tips.

- Consider factors such as p-values to determine the statistical significance of each variable.

4. Encode Categorical Variables

- If necessary, encode categorical variables (such as gender, smoking status, day, and time) into numeric values for compatibility with the regression model.

5. Build the Regression Model

- Based on the selected independent variables (size and total bill) from the analysis, construct an appropriate multiple regression model.

6. Model Evaluation - RMSE Calculation

- Assess the model's accuracy by calculating the Root Mean Square Error (RMSE).

- Subtract the predicted values from the actual values, square these differences, find the mean of the squared differences, and finally, take the square root to obtain the RMSE.

7. Interpretation and Conclusion

- Interpret the results of the RMSE calculation to evaluate how well the predictive model performs in estimating restaurant tips based on the chosen independent variables.

By following these steps, the project aims to establish a robust predictive model using multiple regression for predicting restaurant tips, providing insights into the key factors influencing tip amounts.