**Description:**

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 |

The following project tasks are required to be performed in excel:

* Use the restaurant tips file for the analytics using Excel
* Find out if there are any missing values and clean the data
* Find the features that are independent and dependent
* Identify which predictive problem is needed.
* Encode the categorical variables to numeric values using IF conditions
* Build an appropriate model with the dataset.
* Calculate the predicted and actual tips values.
* Calculate the RMSE(Root Mean Square Error) of the model. RMSE is root of mean of square errors.

**Tools required:** Microsoft Excel, Data Analysis Add-in.

**Expected Deliverables:**Model to predict restaurant tips given input values with the mathematical equation for predicting the tips value.

You can download the Datasets from [**HERE**](https://github.com/Simplilearn-Edu/BA-With-Excel-Revamp-2021)

Solution: Steps to follow in excel:

1. Use the restaurant tips file for the analytics using Excel.
2. Find out if there are any missing values and clean the data.

* Step1: Download the restaurant Tip.xlxs
* Step2: Use the restaurant tips file for the analytics using Excel. Find out any missing values and clean the data. Click on home and select find and select option and click on go to special, in the dialogue box click on blank and click ok.
* Step3: if you see “No cells were found” hence there are no missing values.

1. Find the features that are independent and dependent.

Step1: click on data analysis under the data tab click on column size, total bills, and tips to check correlation between dependent and independent variable:

Using correlation function, we can detect the correlation matrix between the columns.

Using the regression, we can also check the p value and determine which column we need to proceed for prediction.

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

Use the if () clause for converting categorical variables into numerical variables.

For day code:

=IFS(C2="Thur",1,C2="Fri",2,C2="Sat",3,C2="Sun",4)

For time:

=IFS(D2="Lunch",0,D2="Dinner",1)

For Smoker:

=IFS(B2="No",1, B2="Yes",0)

1. Identify which predictive problem is needed.

Using multiple regression we can check the predictive problem .

1. Identify which predictive problem is needed.

Using the regression function determine the output residuals.

1. Build an appropriate model with the dataset.

Using the regression, we can determine the model prediction and residuals.

1. Calculate the predicted and actual tips values.

Using moving average determines the tip for the 245th row.

1. Calculate the RMSE (Root Mean Square Error) of the model. RMSE is root of mean of square errors.

Square =SUMSQ(columns)

Count=COUNTNA(column range)

RMSE=SQRT(SUMSQ(Columns range)/COUNTNA(column range))