**Predicting restaurant tips using predictive analytics on Excel.**

**Course-end Project 2**

**DESCRIPTION**

Use excel to predict restaurant tips.

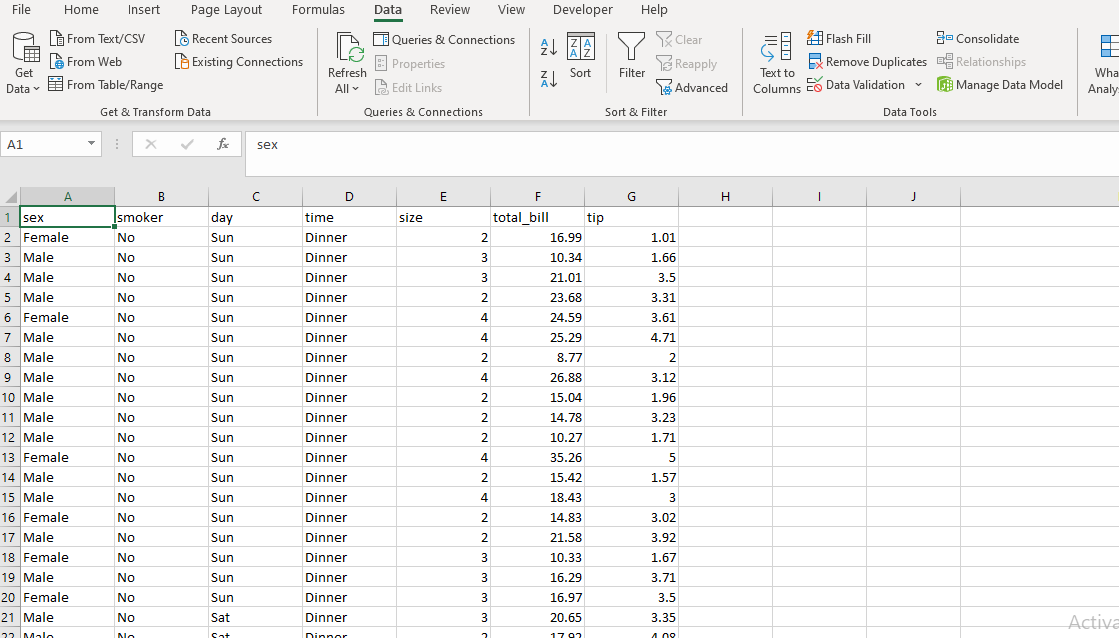
**Description:**

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

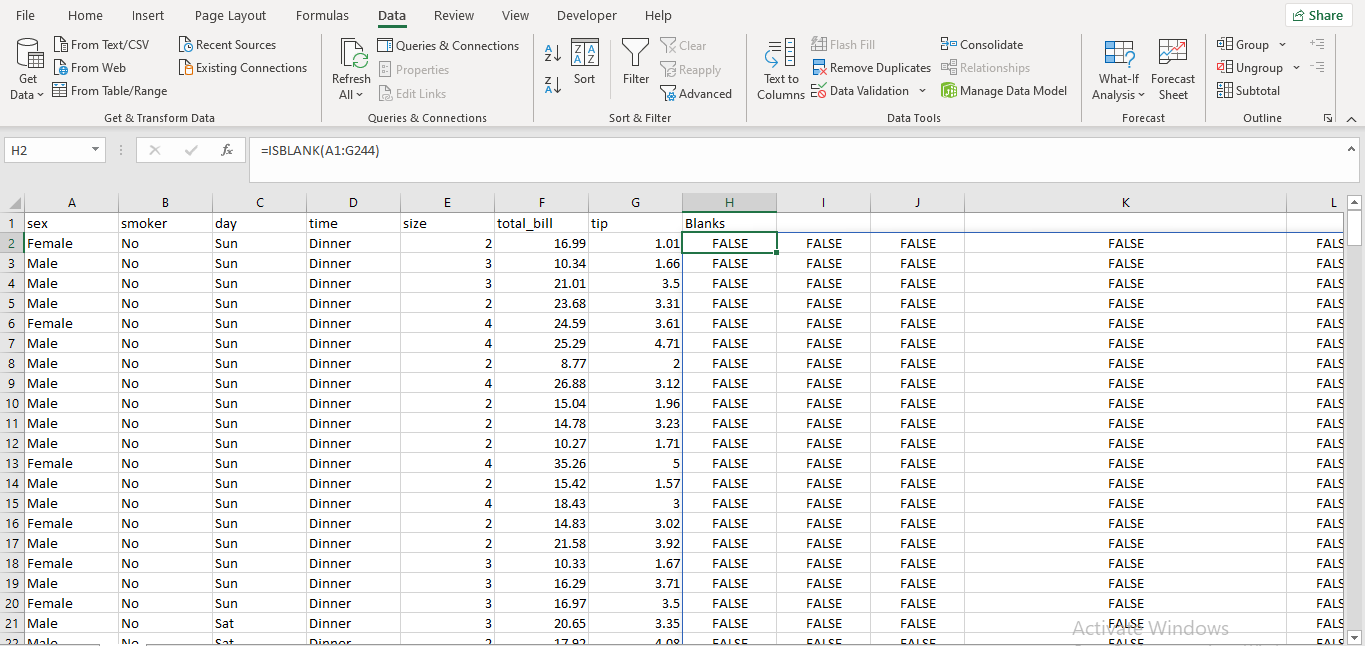
|  |  |
| --- | --- |
| sex | Gender of the customer |
| smoker | Indicates if the customer is a smoker or not |
| day | Day of the restaurant visit |
| time | Indicates whether the tip was for lunch or dinner |
| size | Number of members dining |
| total bill | Bill amount in USD |
| tip | Tip amount 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



Here 1 duplicate value is found and I removed it.



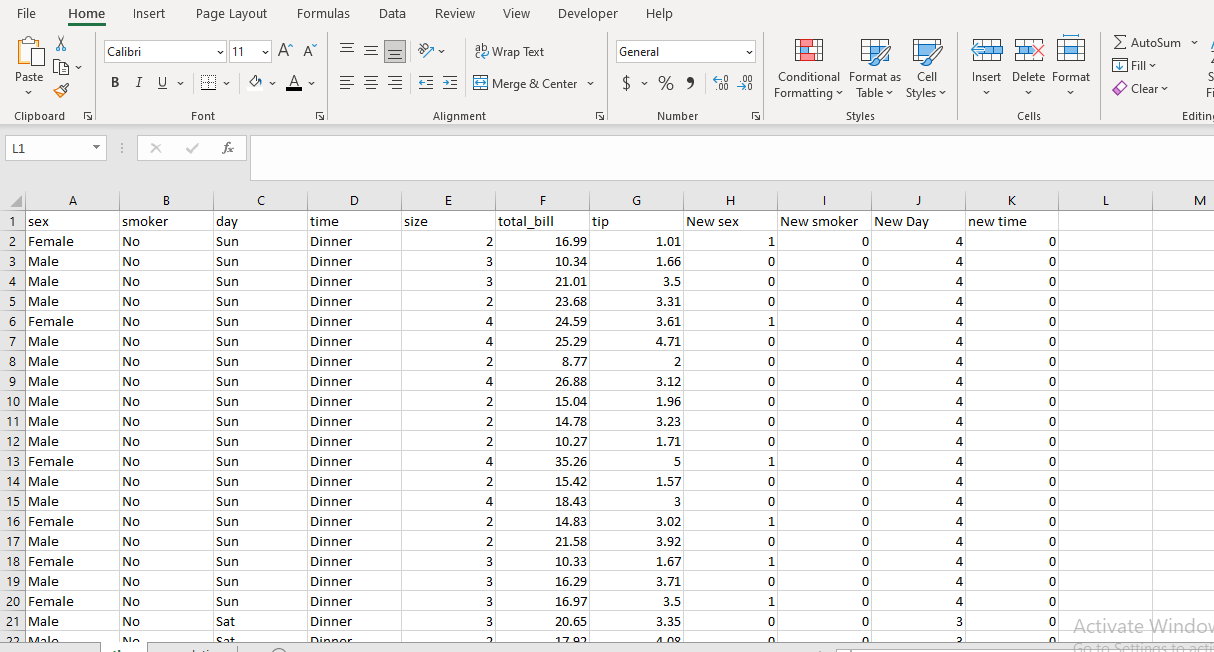
Checked for null values there is no null values.

* Find the features that are independent and dependent

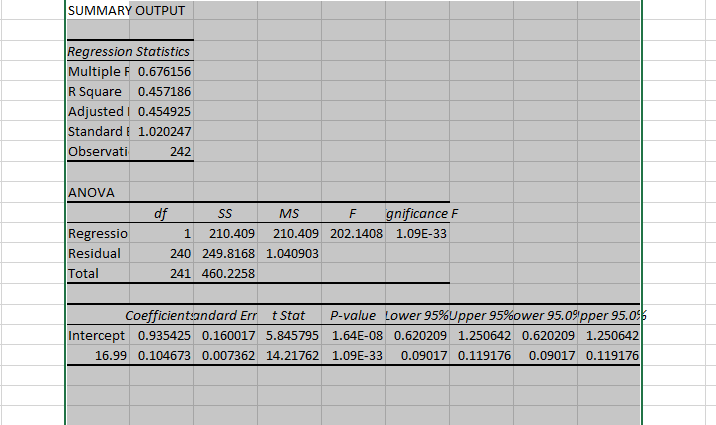
Here tip is dependent feature and remaining all are independent features.

Here find the dependent and independent (input) features by correlation here depenedent feature is tip and we need to find best input feature with high correlation coefficient with tips (output)to predict future tips

* Identify which predictive problem is needed.
* Encode the categorical variables to numeric values using IF conditions

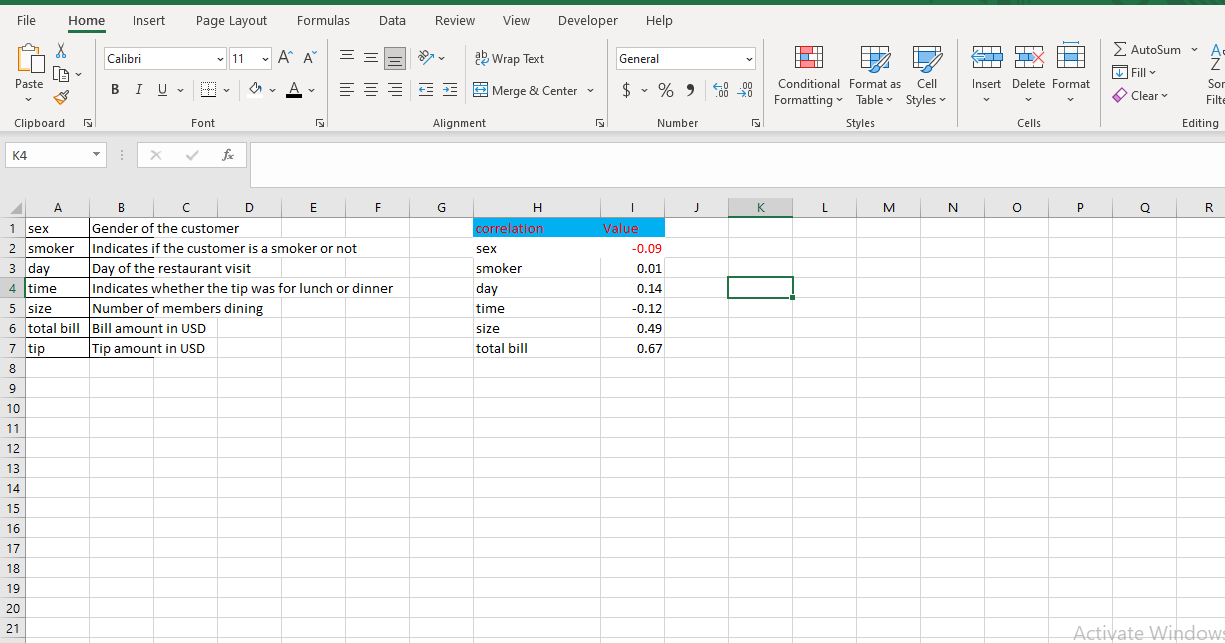


* Build an appropriate model with the dataset.



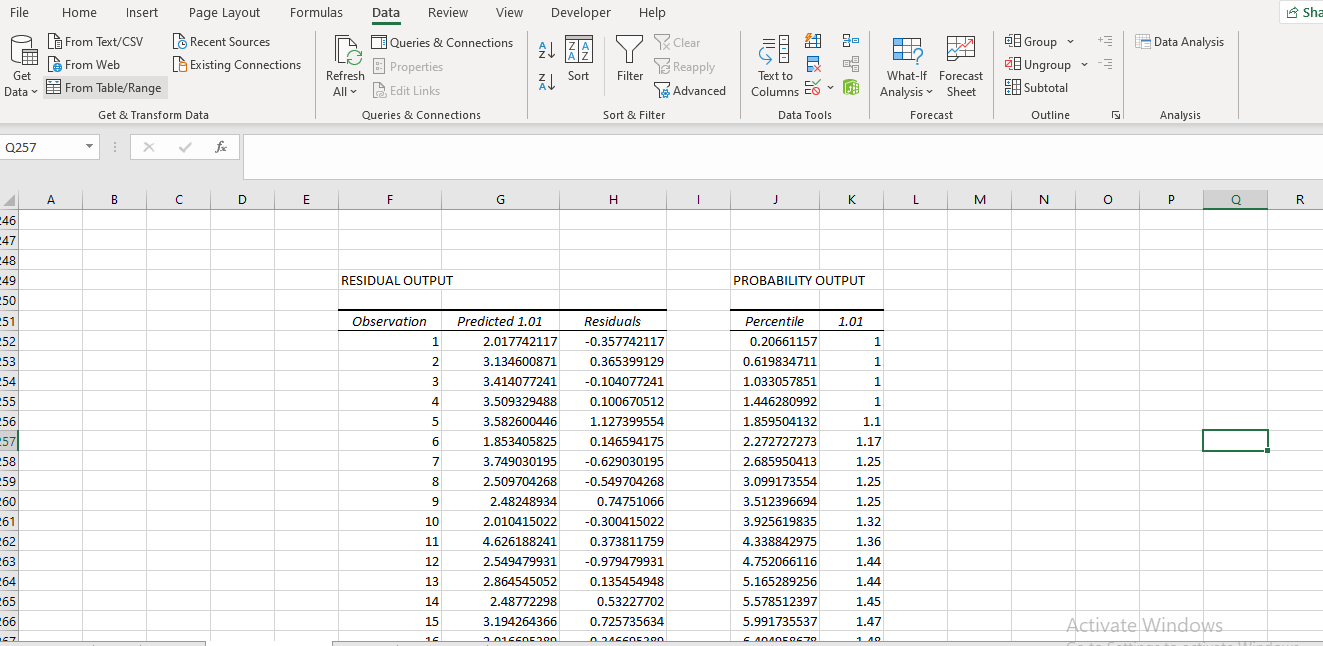
To find future tips values we build model of regression because here we are predicting values not classes.

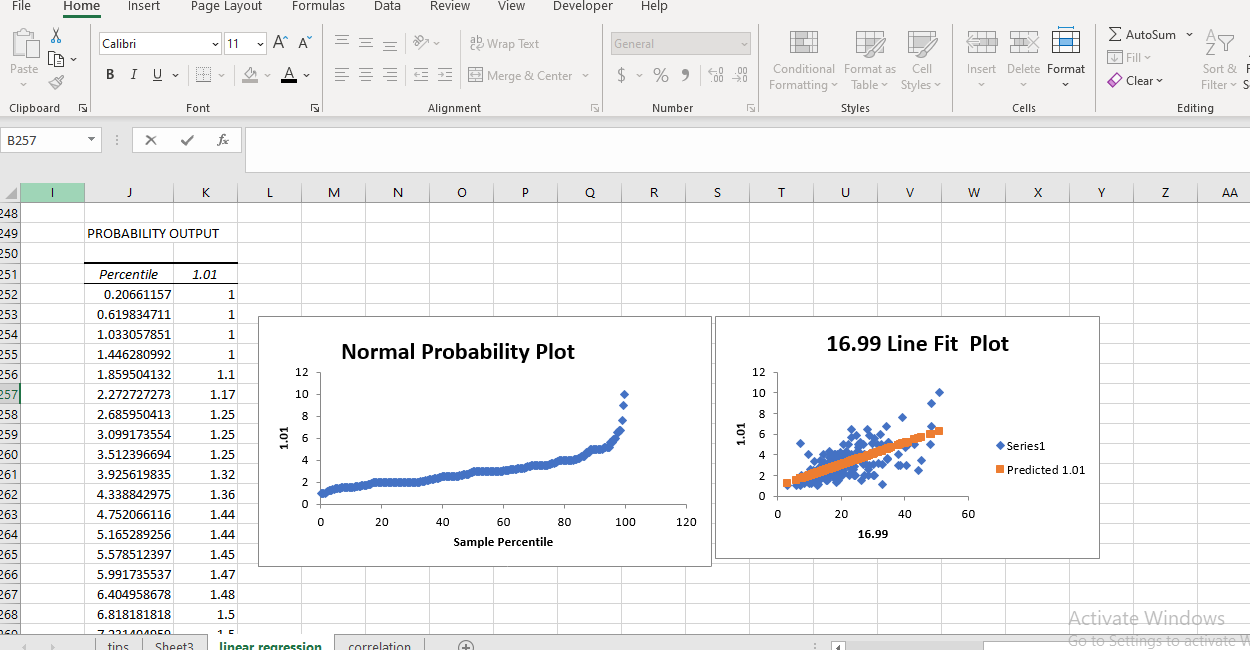
Here linear regression is best fit model for this problem.

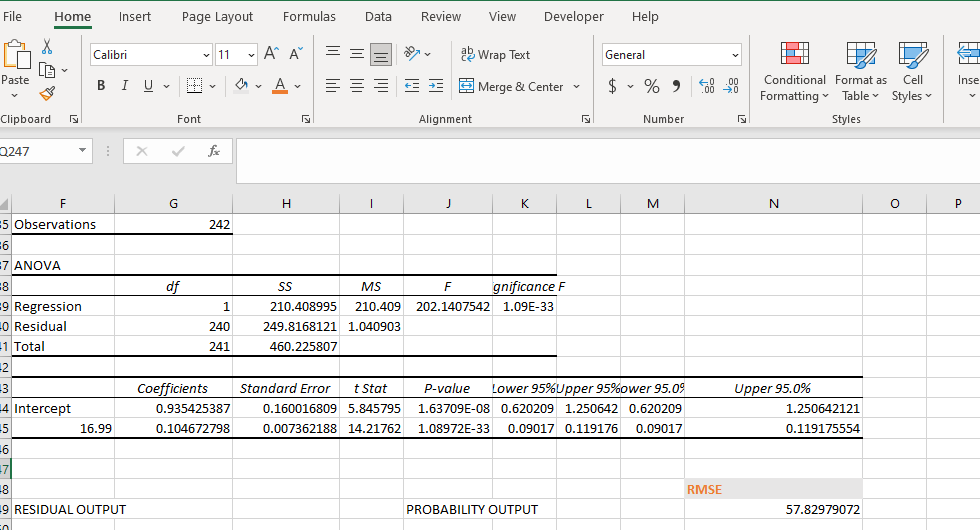


Here we calculated correlation coefficient for all input features with tip and we need high value for feature to be chose as best independent feature to pridict values .here total bill is highly correlated with tip that is 0.67 so we chose that as input feature.

* 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.
* 

Here RMSE value is calculated by

=SQRT(SUMSQ(J252:J493,K252:K493)/COUNTA(J252:J493)) i.e 57.82979072

**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.