**Student Name: Ashab Shaikh**

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

**Expected Deliverables:**Model to predict restaurant tips given input values with the mathematical equation for predicting the tips value. Project predict tips based on characteristics of dinners and dining.

**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.

1. There are no missing values.
2. Dependent variable: Tip, independent variables: sex, smoker, day, time, size, total\_bill, tip

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *size* |  | *total\_bill* |  | *tip* |  |
|  |  |  |  |  |  |
| Mean | 2.569672 | Mean | 19.78594 | Mean | 2.998279 |
| Standard Error | 0.060888 | Standard Error | 0.569919 | Standard Error | 0.088578 |
| Median | 2 | Median | 17.795 | Median | 2.9 |
| Mode | 2 | Mode | 13.42 | Mode | 2 |
| Standard Deviation | 0.9511 | Standard Deviation | 8.902412 | Standard Deviation | 1.383638 |
| Sample Variance | 0.904591 | Sample Variance | 79.25294 | Sample Variance | 1.914455 |
| Kurtosis | 1.7317 | Kurtosis | 1.218484 | Kurtosis | 3.648376 |
| Skewness | 1.447882 | Skewness | 1.133213 | Skewness | 1.465451 |
| Range | 5 | Range | 47.74 | Range | 9 |
| Minimum | 1 | Minimum | 3.07 | Minimum | 1 |
| Maximum | 6 | Maximum | 50.81 | Maximum | 10 |
| Sum | 627 | Sum | 4827.77 | Sum | 731.58 |
| Count | 244 | Count | 244 | Count | 244 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *sex\_label* |  | *smoker\_label* |  | *day\_label* |  | *time\_label* |  |
|  |  |  |  |  |  |  |  |
| Mean | 0.356557 | Mean | 0.381148 | Mean | 2.72541 | Mean | 0.721311 |
| Standard Error | 0.030727 | Standard Error | 0.031156 | Standard Error | 0.073991 | Standard Error | 0.028762 |
| Median | 0 | Median | 0 | Median | 3 | Median | 1 |
| Mode | 0 | Mode | 0 | Mode | 3 | Mode | 1 |
| Standard Deviation | 0.479967 | Standard Deviation | 0.486667 | Standard Deviation | 1.155774 | Standard Deviation | 0.449276 |
| Sample Variance | 0.230368 | Sample Variance | 0.236845 | Sample Variance | 1.335813 | Sample Variance | 0.201848 |
| Kurtosis | -1.6504 | Kurtosis | -1.77207 | Kurtosis | -1.2507 | Kurtosis | -1.02173 |
| Skewness | 0.602658 | Skewness | 0.492471 | Skewness | -0.44761 | Skewness | -0.99333 |
| Range | 1 | Range | 1 | Range | 3 | Range | 1 |
| Minimum | 0 | Minimum | 0 | Minimum | 1 | Minimum | 0 |
| Maximum | 1 | Maximum | 1 | Maximum | 4 | Maximum | 1 |
| Sum | 87 | Sum | 93 | Sum | 665 | Sum | 176 |
| Count | 244 | Count | 244 | Count | 244 | Count | 244 |

1. Predict tip based on other characteristics such as total bill, sex, smoker, day, time, size.
2. tip= b0+b1\*sex+b2\*smoker+b3\*day+b4\*time+b5\*size+b6\*total\_bill+e

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Regression output** |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |
| Multiple R | 0.684980787 |  |  |  |  |
| R Square | 0.469198679 |  |  |  |  |
| Adjusted R Square | 0.455760671 |  |  |  |  |
| Standard Error | 1.020745565 |  |  |  |  |
| Observations | 244 |  |  |  |  |
|  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 6 | 218.2770796 | 36.37951327 | 34.91579067 | 4.09922E-30 |
| Residual | 237 | 246.9353974 | 1.041921508 |  |  |
| Total | 243 | 465.212477 |  |  |  |
|  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |  |
| Intercept | 0.636618577 | 0.268289652 | 2.372877869 | 0.018449024 |  |
| sex\_label | 0.034644964 | 0.141081963 | 0.245566218 | 0.806230561 |  |
| smoker\_label | -0.075663089 | 0.140198277 | -0.539686293 | 0.589920088 |  |
| day\_label | 0.05273982 | 0.120334639 | 0.4382763 | 0.661585219 |  |
| time\_label | -0.112477769 | 0.307526134 | -0.365750277 | 0.714877667 |  |
| size | 0.174819618 | 0.089187194 | 1.960142606 | 0.051150876 |  |
| total\_bill | 0.094325088 | 0.009538173 | 9.889219168 | 1.57818E-19 |  |

1. Showing some of the Calculated residuals

|  |  |  |  |
| --- | --- | --- | --- |
| *Observation* | *Predicted tip* | *Residuals* | *Observed/actual tips* |
| 1 | 2.72196753 | -1.71196753 | 1.01 |
| 2 | 2.234880351 | -0.574880351 | 1.66 |
| 3 | 3.241329037 | 0.258670963 | 3.50 |
| 4 | 3.318357404 | -0.008357404 | 3.31 |
| 5 | 3.788477433 | -0.178477433 | 3.61 |
| 6 | 3.819860031 | 0.890139969 | 4.71 |
| 7 | 1.911970345 | 0.088029655 | 2.00 |
| 8 | 3.96983692 | -0.84983692 | 3.12 |

1. Calculated RMSE

|  |
| --- |
| **RMSE** |
| 1.005997176 |

Project predict tips based on characteristics of dinners and dining.

1. Equation used to calculate predicted values:

tip=0.636618577+0.034644964\*sex+-0.075663089\*smoker+0.05273982\*day+-0.112477769\*time+0.174819618\*size+0.094325088\*total\_bill+e

1. Equation used to calculate RMSE

Error= Observation-Forcast (actual tip-predicted tip)

RMSE=Square root of Mean of Square of errors

RMSE evaluates the quality of predictions.