**Predicting restaurant tips using predictive analytics**

**Predicting restaurant tips using predictive analytics** means analyzing data to identify patterns and build a prediction model that forecasts tip amounts based on factors like Total bill, size, time of day, day of the week, etc.

By using statistical tools like **correlation, regression, moving averages**, etc. a restaurant can gain insights into factors that affect tips and make data-driven decisions, potentially improving service and revenue.

This dataset typically includes variables such as total bill, tip amount, customer gender, smoker status, day, time (lunch or dinner), and party size.

Here, **Categorical variables** are encoded numerically with **IF** condition (e.g., 0 for female and 1 for male) to make them compatible. It as follows:

* Sex (0 for female, 1 for male)
* Smoker (0 for non-smokers, 1 for smokers)
* Day (e.g., Thur, Fri, Sat, Sun), as 1
* Time (0 for dinner, 1 for lunch)

**Correlation** measures the relationship between two variables, showing whether they increase or decrease together.Correlation Coefficient (r): Ranges from -1 to +1(+1: Perfect positive correlation and -1: Perfect negative correlation).

In this predictive analytics, correlation helps to identify which factors are most related to the target variable that is the **tip** amount.

It suggests that **total bills** and **size** has a **high** positive correlation with **Tips** which indicate that larger bills and party size tend to result in larger tips.

**Covariance** measures how two variables vary together, indicating the direction of their relationship but not its strength.

**Positive Covariance** Indicates that as one variable increases, the other also tends to increase. For instance, a positive covariance between **Total Bill** and **Tip** means higher bills are generally associated with higher tips.

A **Moving Average** is a technical tool used to make it easier to observe underlying trends over time.

Here, percentages for tips relative to total bills, includes a moving average, likely to smooth variations and observe tipping trends over time. A moving average can help identify tipping trends and patterns across different time periods, such as days, weeks. etc. It also shows a moving average chart trends over time.

**Regression** focuses on building a model to predict tip amounts based on the variables. **Residual** means the difference between the **actual tip(1.01)** and the **predicted tip(2.62)**, representing**(-1.61)** the model’s error for each observation.

**Multiple R** represents the correlation between the actual and predicted values. By applying CORREL function(=correl(array1,array2) will give 0.68.

A higher value suggests a stronger relationship. For example here, **Multiple R** of 0.68 suggests that **68%** of the variability in tips is explained by the model.

**R Square**: This metric, also known as the coefficient of determination, indicates the proportion of variability in the tip amount explained by the model’s predictors. For example, an R Square of 0.46 suggests that **46%** of the variability in tips is explained by the model.

**RMSE (Root Mean Squared Error)** is a common metric in regression analysis that measures the average magnitude of the prediction errors. Here, RMSE tells us how far the predicted tip amounts are from the actual tips.

The RMSE (Root Mean Squared Error) for the regression model is approximately **1.007**. This means that, on average, the model's predictions for tips deviate from the actual tips by about 1.007 units.

The variables size and total bill are statistically significant, suggesting that these two variables have a meaningful impact on predicting tips.