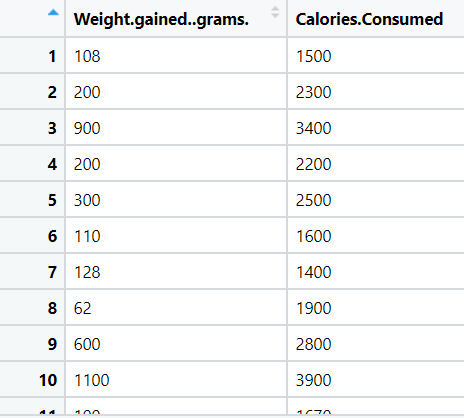
# Topic(s): Simple Linear Regression

**Problem Statement: -**

A certain food based company conducted a survey with the help of a fitness company spread across the country to find relationship between a person’s weight gain and the no of calories consumed by them in order to come up a diet plan for individuals that fall under different weight groups. Approach - A Simple Linear regression model needs to be built with target variable ‘Calories.Consumed’. Apply necessary transformations and record the RMSE values, Correlation coefficient values for different transformation models.



**Business Problem**

**Objective :-** Calories.Consumed’

**Python code details :**

Data Frame name is df. It has 14 rows and 2 columns.

**Work on each feature of the dataset to create a data dictionary as displayed in the below image:**

Then we create a data frame that’s contain details of each columns ,like- description ,data types ,and save the details named as data\_details .all of them are important .

**Data Pre-processing**

**Data Cleaning and Data Mining.**

Now we check info and describe for df .Check for data types ,unique value and variance .

Then we check for unique value in each columns

:-

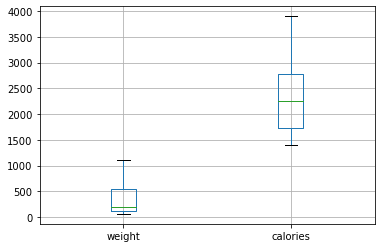
Weight gained (grams) 13

Calories Consumed 13

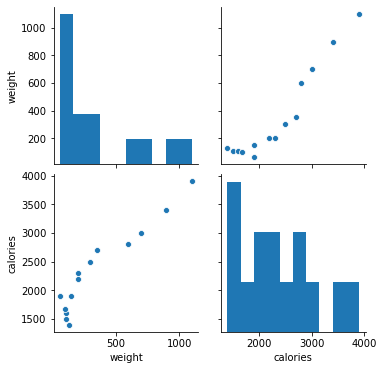
Dataframe has no missing values in columns .

We have done EDA for each columns and save the details as EDA. covariance for data set save as covariance . historgam and scatter plot for each column all data are normally distributed as well as we check for boxplot .there is no outliers present.

Boxplot:-



Histogram and Scatter plot:-



1. **Model Building:**
   1. **Perform Simple Linear Regression on the given datasets**
   2. **Apply different transformations such as exponential, log, polynomial transformations and calculate RMSE values, R-Squared values, Correlation Coefficient for each model**
   3. **Build the models and choose the best fit model**
   4. **Briefly explain the model output in the documentation**

We build ML models that predict calories consumed gained using weight. here y is continuous .so we tried simple linear regression with different transformation.

**Result = MODEL RMSE R-squared Adj. R-squared Correlation coefficient**

**SLR 232.83 0.897 0.888 0.94**

**Log model 253.55 0.878 0.867 0.89**

**Exp model 272.42 0.808 0.792 0.93**

**Poly model 240.82 0.852 0.825 0.89**

**Model details saved Model\_details.**

Applying transformation is decreasing Multiple R Squared Value. So model does not need further transformation, Multiple R-squared: 0.89.

Model reg is best fit model= SLR ,root mean square error 232.83 , R\_squared 0.897 , . . Adj. R-squared 0.888 .

We split the data 50-50 in train and test.

So , we build our final model on SLR build on train data and test on test data .

Now we split our data in X\_train, X\_test, Y\_train, Y\_test 50% data on train and 20% test . Preparing a Simple linear regression on training data set

R-squared: 0.911

Adj. R-squared: 0.893

,then test on test data , Evaluation on Test Data as result root mean square error=233.

Evaluation on Train Data also as result root mean square error=231.

* used library :- pandas, numpy, sk learn, matplotlib, statsmodels