1.Below is the table containing the information of Customer price index, discounts, offers where the organization wants to predict the sales based on the cpi, discounts,offers.

|  |  |  |  |
| --- | --- | --- | --- |
| CPI | discounts | offers | Sales |
| 2600 | 3 | 20 | 550000 |
| 3000 | 4 | 15 | 565000 |
| 3200 | 5 | 18 | 610000 |
| 3600 | 3 | 30 | 595000 |
| 4000 | 5 | 8 | 760000 |
| 4100 | 6 | 8 | 810000 |

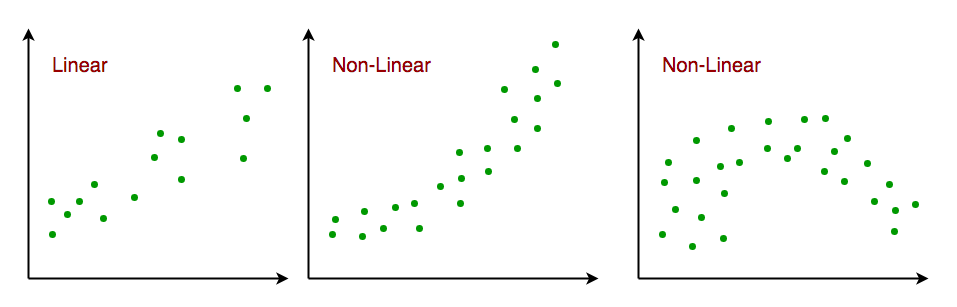
For the above question i am going to use Linear Regression Model

**What is Linear Regression?**

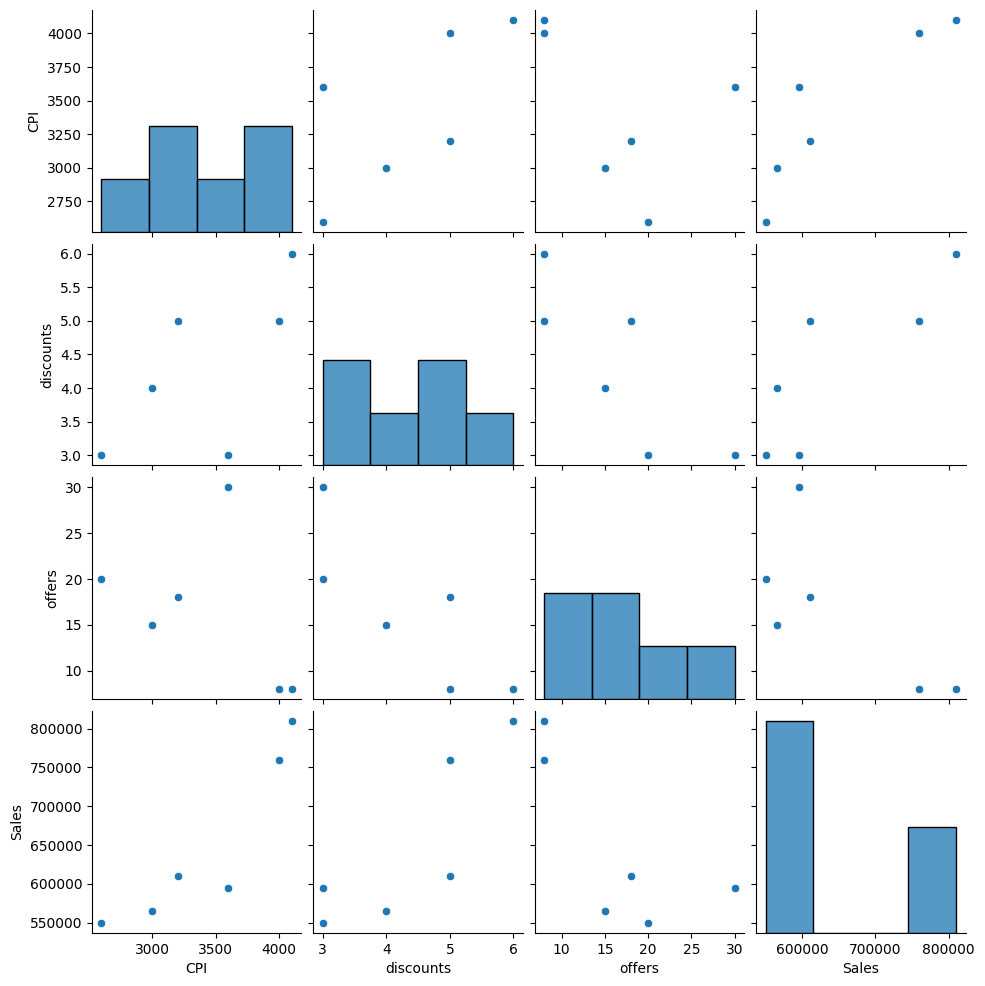
[**Linear regression**](https://www.geeksforgeeks.org/ml-linear-regression/)is a statistical method that is used to predict a continuous dependent variable (target variable) based on one or more independent variables (predictor variables). This technique assumes a linear relationship between the dependent and independent variables, which implies that the dependent variable changes proportionally with changes in the independent variables.

### **Assumptions We Make in a Linear Regression Model:**

Given below are the basic assumptions that a linear regression model makes regarding a dataset on which it is applied:



**Output:**



**LINEAR REGRESSION ASSUMPTIONS:**

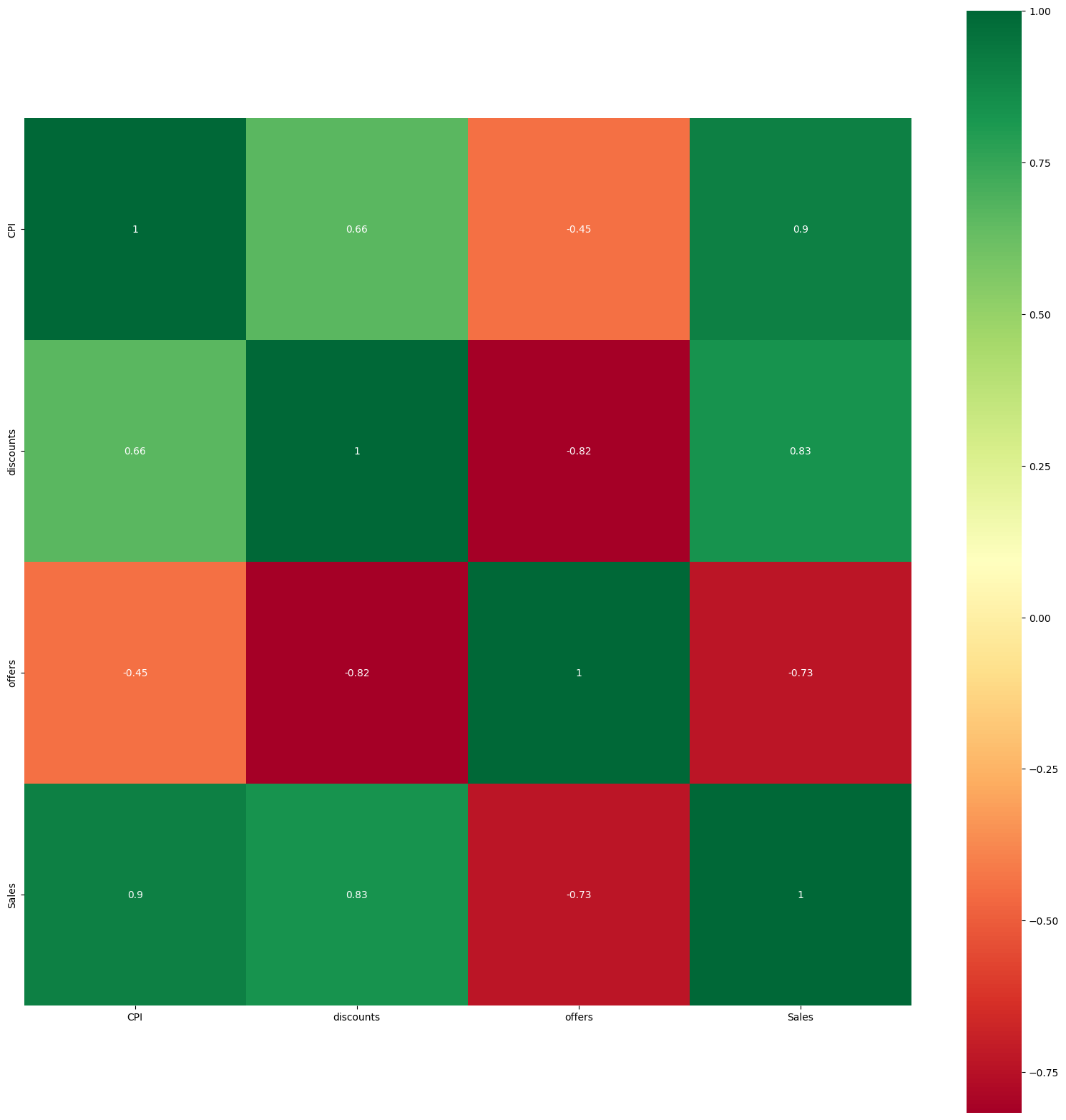
The relationship between response and feature variables should be linear. The linearity assumption can be tested using scatter plots.

Let's use a pair plot to check the relation of independent variables with the Sales variable

R squared: 0.9517053277318955

plt.figure(figsize=(20,20)) # on this line I just set the size of figure to 12 by 10.

p=sns.heatmap(sales\_data.corr(), annot=True,cmap='RdYlGn',square=True) # seaborn has very simple solution for heatmap



OLS Regression Results

==============================================================================

Dep. Variable: Sales R-squared: 0.952

Model: OLS Adj. R-squared: 0.879

Method: Least Squares F-statistic: 13.14

Date: Tue, 30 Jan 2024 Prob (F-statistic): 0.0716

Time: 22:59:58 Log-Likelihood: -68.476

No. Observations: 6 AIC: 145.0

Df Residuals: 2 BIC: 144.1

Df Model: 3

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

Intercept 2.648e+05 1.64e+05 1.613 0.248 -4.41e+05 9.71e+05

CPI 128.4351 39.639 3.240 0.083 -42.120 298.990

offers -4902.5460 3641.815 -1.346 0.311 -2.06e+04 1.08e+04

discounts 5913.5196 2.99e+04 0.198 0.861 -1.23e+05 1.34e+05

==============================================================================

Omnibus: nan Durbin-Watson: 2.185

Prob(Omnibus): nan Jarque-Bera (JB): 0.238

Skew: -0.031 Prob(JB): 0.888

Kurtosis: 2.026 Cond. No. 3.69e+04

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 3.69e+04. This might indicate that there are

strong multicollinearity or other numerical problems.