**VIF REPORT**

1. vif=pd.DataFrame()
2. X=self.dataset[self.quan]
3. from statsmodels.stats.outliers\_influence import variance\_inflation\_factor
4. vif["Variables"]=X.columns
5. vif["VIF"]=[variance\_inflation\_factor(X.values,i) for i in range(X.shape[1])]
6. return(vif)

**Explanation for above code:**

1. A table vif has been created to store VIF values of each column

2. Assigning quantitative columns with the respective values to the variable X

3. variance\_inflation\_factor method has been imported.

4. Assigning column names to vif

5. Calculating vif: X.shape[1] takes the total number of columns and each number assigned to i from starting value on each iteration for eg: X.shape[1]=5 means 0 to 5

So i=0 to 5

If i =0 x.values take the value of zeroth column that is first column and calculate vif for those values.

**VIF for all independent variables:**

Variables VIF

0 ssc\_p 67.087602

1 hsc\_p 59.100713

2 degree\_p 114.085694

3 etest\_p 32.704858

4 mba\_p 99.767641

**Dropping one of the correlated variables:**

One of the simplest methods for handling multicollinearity is to drop one of the correlated variables from the regression model. This method works well when there are only two correlated variables, but it may not be practical when there are multiple correlated variables.

**VIF leaving one independent variables:**

Variables VIF

0 ssc\_p 65.789610

1 hsc\_p 56.447314

2 degree\_p 81.045948

3 etest\_p 30.560290

**VIF leaving two independent variables:**

Variables VIF

0 ssc\_p 64.565289

1 hsc\_p 54.728410

2 degree\_p 71.827077

**VIF leaving three independent variables:**

Variables VIF

0 ssc\_p 41.494645

1 hsc\_p 41.494645

Using principal component analysis (PCA):

PCA is a statistical technique that can be used to reduce the dimensionality of the predictor variables by transforming them into a smaller number of uncorrelated variables called principal components. PCA can help to reduce the effects of multicollinearity by creating new variables that explain most of the variance in the original variables.

0 1

0 -0.119982 -1.220955

1 2.478016 0.313948

2 -0.452895 0.373799

3 -2.489203 0.372589

4 1.741619 1.350622

... ... ...

210 3.415286 0.411804

211 -0.934783 0.346964

212 0.701875 -1.279954

213 -0.457330 0.043262

214 -1.279318 1.818872

The following methods can also be used:

**Ridge regression:**

Ridge regression is a type of regression analysis that adds a penalty term to the regression coefficients to reduce the effects of multicollinearity. The penalty term helps to shrink the regression coefficients towards zero, which can improve the stability and reliability of the estimates.

**Lasso regression:**

Lasso regression is another type of regression analysis that adds a penalty term to the regression coefficients to reduce the effects of multicollinearity. The penalty term in lasso regression is different from ridge regression, as it encourages some of the regression coefficients to be exactly zero. This can help to reduce the effects of multicollinearity by selecting only the most important predictor variables.