What is Variance Inflation Factor?

► **Variance inflation factor** (VIF) is a tool to detect multicollinearity, which is when independent variables in a regression model are correlated with each other.

$$VIF_i = \frac{1}{1 - R_i^2} = \frac{1}{\text{Tolerance}}$$

In Variance Inflation Factor...

Employee Id	First Name	Last Name	Department	Age	Experience	Salary
1	Joy	Bass	Sales and Marketing	28	3	32889
2	Sheila	Garza	Sales and Marketing	22	1	15944
3	John	Bryant	Customer Relations	22	1	40343
4	Christian	Farley	Customer Relations	22	1	19018
5	Colorado	Bowen	Accounting	27	0	24795
6	Elijah	Matthews	Accounting	26	5	15387
7	Bo	Mcleod	Customer Relations	24	6	34117

▷ $VIF_i = \dfrac{1}{1 - R_i^2}$

For Example if we take Employee Salary Data, How VIF is calculated means, First EmployeeID(X1) is Taken and compare with all the Input of Columns and find the VIF value and take it as X1.
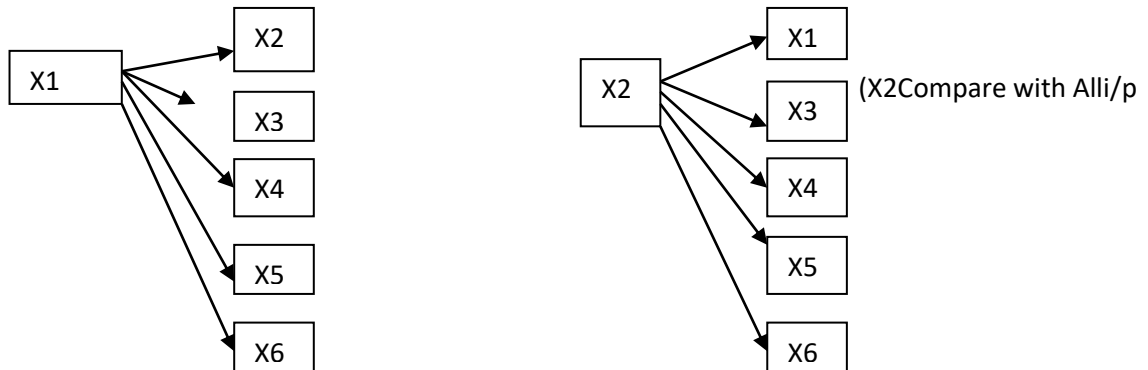
Same like X3,X4,X5 will find variation Inflation Factor.And Result display will be

Result Display...

	Variation Inflation Factor	Features
0	1.344	X1
1	1.2	X2
2	5.5443	X3
3	8	X4
4	1.55	X5

R²	VIF
<0.9	1
0.9	10
0.99	100

- VIF = 1 means no collinearity (orthogonal)
- VIF between 5-10 or higher indicates collinearity
- **Solution: Find VIF, remove the redundant term**

The VIF ranges More Than 5 can remove the redundant term.

```python
from statsmodels.stats.outliers_influence import variance_inflation_factor

def calc_vif(X):

    # Calculating VIF
    vif = pd.DataFrame()
    vif["variables"] = X.columns
    vif["VIF"] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

    return(vif)
```

```python
vif["VIF"] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
```

In this code:

X.values accesses the values of the DataFrame X.

X.shape[1] retrieves the number of columns in the DataFrame X.

The list comprehension **[variance_inflation_factor(X.values, i) for i in range(X.shape[1])]** iterates over each column index in X and calculates the VIF for each predictor variable.

This code efficiently computes the **VIF values** for each predictor variable in the DataFrame X, providing insights into the degree of multicollinearity present in the regression model.

```
[14]: calc_vif(dataset[numeric_columns])
```

[14]:

	variables	VIF
0	ssc_p	78.168671
1	hsc_p	61.882196
2	degree_p	114.820554
3	etest_p	32.720365
4	mba_p	116.034378
5	salary	4.171783

Here mba_p mark is more Collinearity. So we can Remove.