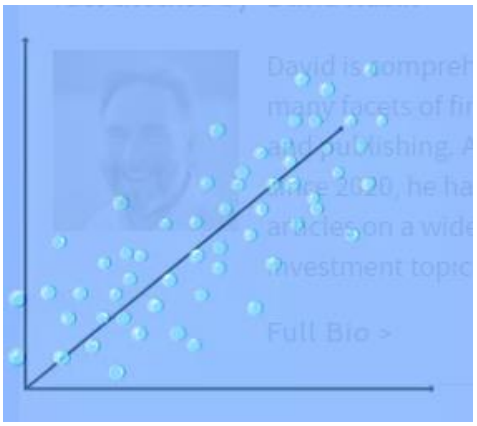
Multi Collinearity

What is Multi collinearity?

- Multicollinearity is the occurrence of high intercorrelations among two or more independent variables in a multiple regression model.



Two ways of remove Multi collinearity :-

- VIF -Variance Inflation factor
- Correlation

VIF -Variance Inflation Factor:-

- A variance inflation factor (VIF) is a measure of the amount of multicollinearity in regression analysis.
- Multi collinearity exists when there is a correlation between multiple independent variables in a multiple regression model.

Formula and Calculation of VIF

The formula for VIF is:

$$\mathrm{VIF}_i = \frac{1}{1 - R_i^2}$$

where:

$R_i^2 = \text{Unadjusted coefficient of determination for regressing the ith independent variable on the remaining ones}$

```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)
```

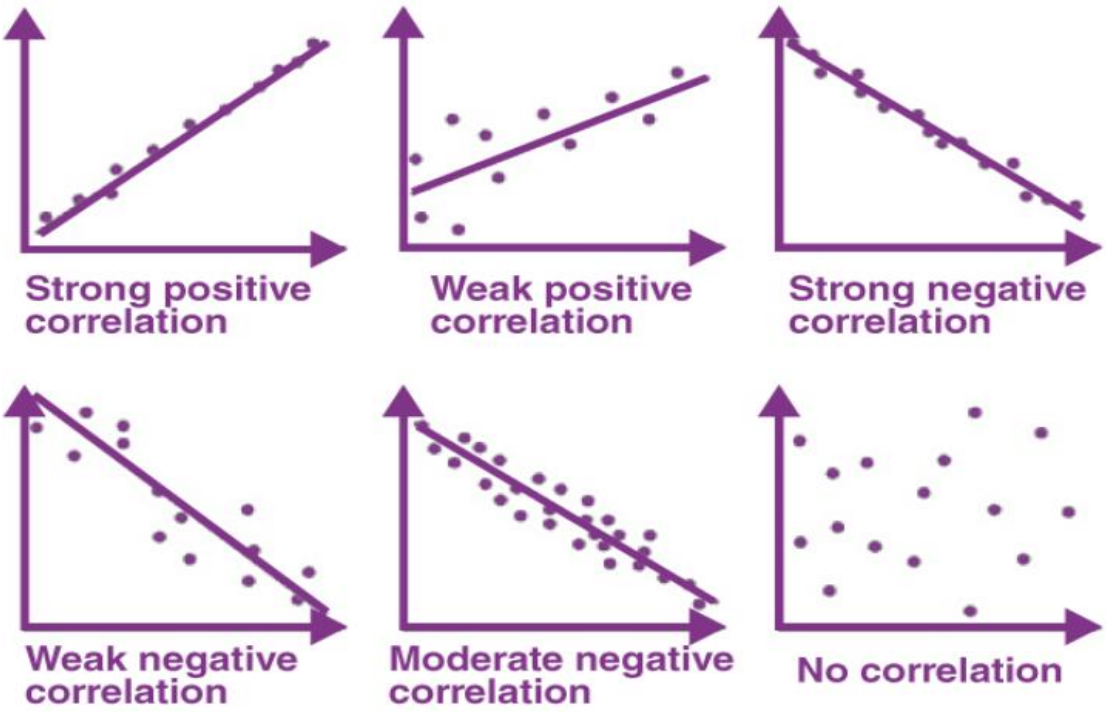
```python
calc_vif(dataset[["etest_p","salary"]])
```

	variables	VIF
0	etest_p	11.595795
1	salary	11.595795

Correlation:-

- Correlation is a statistical measure that expresses the extent to which two variables are linearly related (meaning they change together at a constant rate).
- It's a common tool for describing simple relationships without making a statement about cause and effect.

Types Of Correlation:-



```python
#Calculating correlation in dataset
def correlation(df, threshold):
    correlated_cols = set()
    corr_matrix = df.corr(numeric_only=True)
    for i in range(len(corr_matrix.columns)):
        for j in range(i):
            if abs(corr_matrix.iloc[i,j]) > threshold:
                colname = corr_matrix.columns[i]
                correlated_cols.add(colname)
    return correlated_cols
```

```python
correlation(X_train, 0.4)
```

```
{'degree_p', 'hsc_p', 'mba_p'}
```

- Then we can remove the highly correlate independent variables and create a model