| FULL LEGAL NAME | **LOCATION (COUNTRY)** | **EMAIL ADDRESS** | **MARK X FOR ANY NON-CONTRIBUTING MEMBER** |
| --- | --- | --- | --- |
| Kamlesh Kumar Bhatt | India | kbhatt2005@gmail.com |  |
| Mohamad Sleiman |  |  |  |
| JOHN OPEYEMI DADA |  |  |  |

|  |  |
| --- | --- |
| **Statement of integrity:** By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an “X” above). | |
| **Team member 1** |  |
| **Team member 2** |  |
| **Team member 3** |  |

|  |
| --- |
| Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed.  **Note:** You may be required to provide proof of your outreach to non-contributing members upon request. |
|  |

**Multicollinearity**

**Definition:** Multicollinearity is a statistical phenomenon in which two or more independent variables in a regression model are highly correlated with each other.

**Description**: For a regression model to be effective, the independent (exogenous) variables must be uncorrelated with each other. Multicollinearity introduces redundancy that makes it difficult to determine the individual impact of each variable on the dependent (endogenous) variable. As a result, multicollinearity inflates the variances of the estimated regression coefficients, leading to less reliable and stable model predictions.

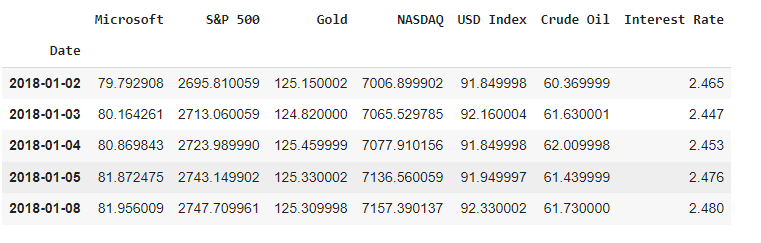
**Demonstration**: We are using real world data to analyze Multicollinearity where:

**Endogenous Variable:** Microsoft (MSFT) stock price

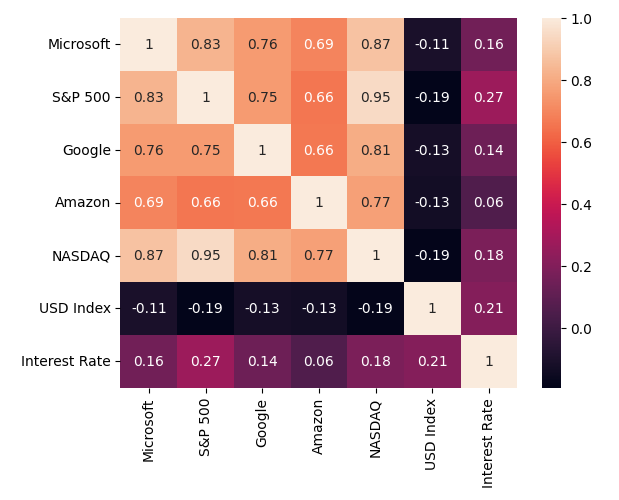
Following **Exogenous variables** used for analysis:

S&P 500 Index, Google stock price, NASDAQ Composite Index, US Dollar Index, Amazon stock price, 10-Year Treasury Yield.

We are using data from 2018-01-02 to 2024-08-16. Sample of data is shown below. Please refer to price\_data.xlsx in the zipped folder for complete data.

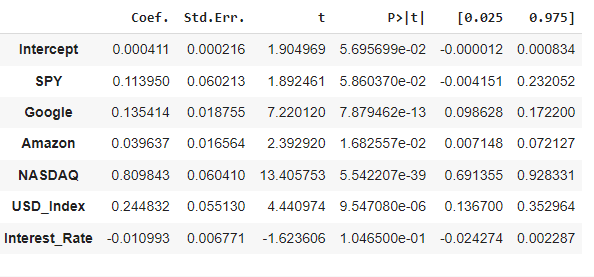


**Diagram:** We are demonstrating a heat map to analyze Multicollinearity.



**Diagnosis:** Above diagrams clearly demonstrate collinearity between independent variables. There is high correlation between S&P 500-Google (0.75), S&P 500 -NASDAQ (0.95), S&P 500-Amazon(0.66), NASDAQ-Google(0.81), NASDAQ-Amazon(0.77). These are clearly multicollinearity problems.

**Damage:** Due to the correlation among the independent variables, the regression model may exhibit large variances, standard errors, and covariances of the coefficients. As a result, the coefficient estimates become less precise, with wider confidence intervals. Additionally, the high standard errors of coefficient make it statistically less significant.



**Directions:** There are various methods to address multicollinearity problems. The simpler approach could be to drop One method could be to drop independent variables having high correlation (>0.8). The other approach is to calculate Variance Inflation factor (VIF) and drop the variables having VIF greater than 5.

For the current example, SPY and NASDAQ VIF values are 12.5 and 18 respectively. So they may be dropped from the regression model.

If the number of dependent variables is large, we may consider sophisticated methods like PCA and factor analysis.