**Experiment 4**

**Shashwat Tripathi**

**D15A Batch C  
RollNo: 64**

**AIM : Implementation of Statistical Hypothesis Test using Scipy and Sci-kit learn**

**THEORY :**

Correlation Tests are statistical methods used to measure the strength and direction of the relationship between two variables. Here's a brief overview of each:

**1. Pearson’s Correlation Coefficient:**

- Pearson’s correlation coefficient, also known as Pearson's r or simply correlation coefficient, measures the linear relationship between two continuous variables.

- It ranges from -1 to 1, where:

- 1 indicates a perfect positive linear relationship,

- -1 indicates a perfect negative linear relationship, and

- 0 indicates no linear relationship.

- Pearson's correlation coefficient assumes that the variables are normally distributed and have a linear relationship.

**2. Spearman’s Rank Correlation:**

- Spearman’s rank correlation coefficient, denoted by ρ (rho), measures the strength and direction of the monotonic relationship between two variables.

- Unlike Pearson's correlation, Spearman's correlation does not require the variables to be normally distributed or have a linear relationship.

- Instead of using the actual values of the variables, it uses the ranks of the values to compute the correlation coefficient.

- It is suitable for ordinal and non-linear data.

**3. Kendall’s Rank Correlation:**

- Kendall’s rank correlation coefficient, denoted by τ (tau), is similar to Spearman’s rank correlation coefficient.

- It also measures the strength and direction of the association between two variables using the ranks of the data.

- However, Kendall’s tau takes into account the number of concordant and discordant pairs of data points to compute the correlation coefficient.

- Like Spearman's correlation, it is also suitable for ordinal and non-linear data.

**4. Chi-Squared Test:**

- The Chi-Squared Test is a statistical test used to determine whether there is a significant association between two categorical variables.

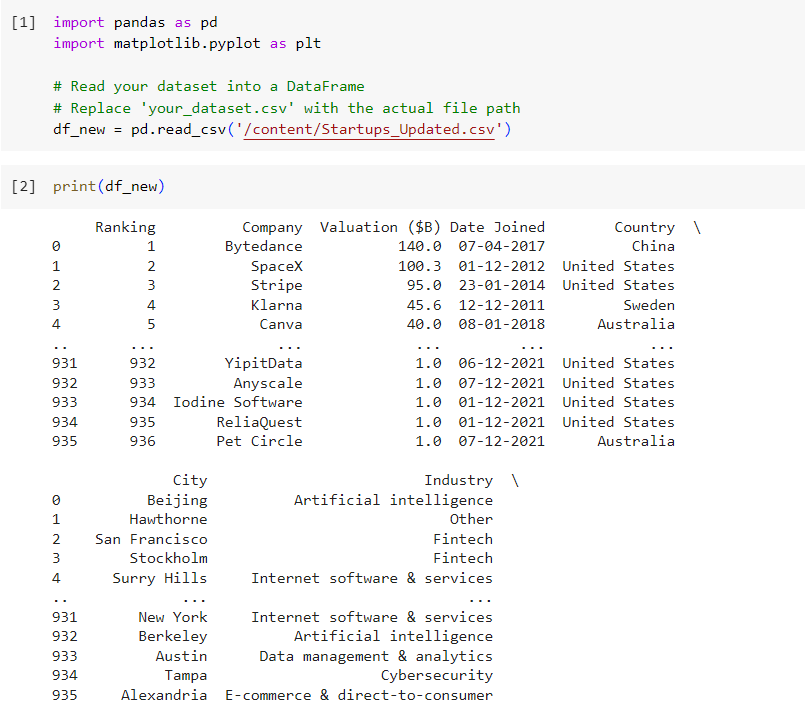
- It compares the observed frequencies of data with the frequencies that would be expected if the variables were independent.

- It is commonly used in contingency tables to assess the relationship between categorical variables.

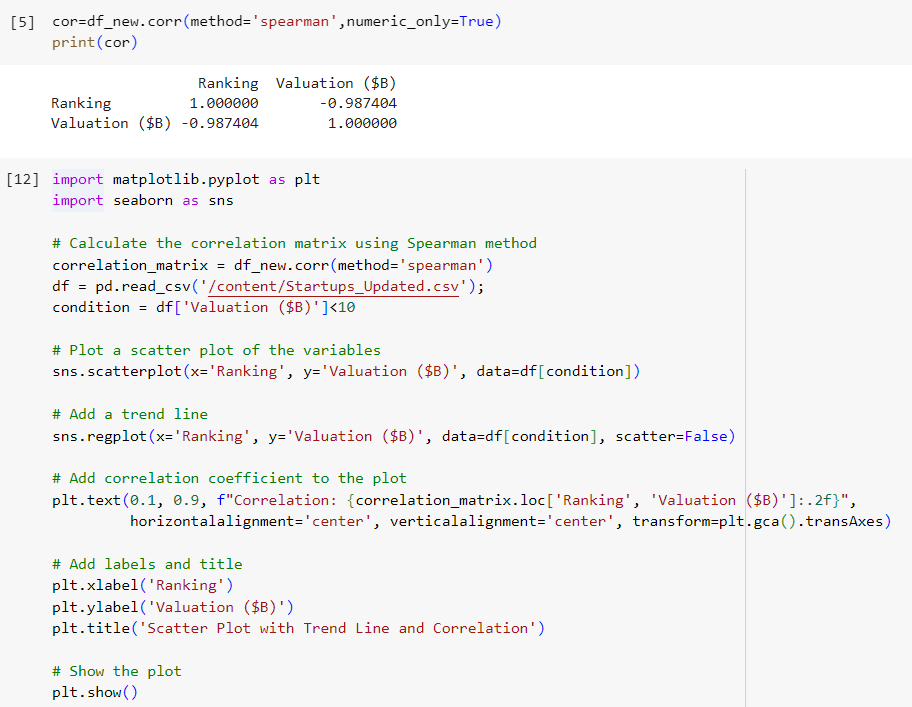
- The test statistic follows a chi-squared distribution, and its significance is determined by comparing it to a critical value from the chi-squared distribution with appropriate degrees of freedom.

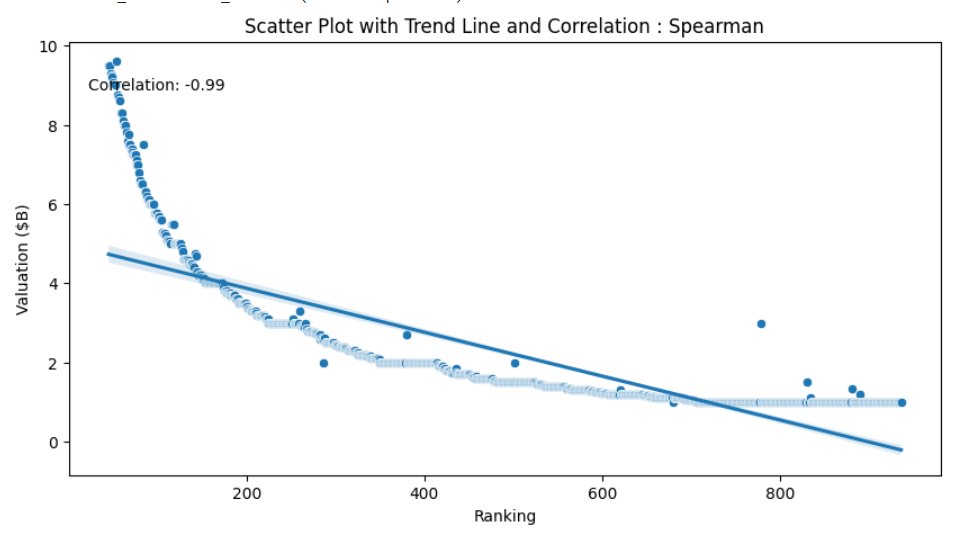
These correlation tests are essential tools in statistical analysis for understanding the relationships between variables and making informed decisions in various fields.

**OUTPUT :**

**Importing the Dataset** [https://www.kaggle.com/datasets/niekvanderzwaag/unicorn-startups](https://www.kaggle.com/datasets/niekvanderzwaag/unicorn-startups-cleaned)****

**1. Spearman Correlation**

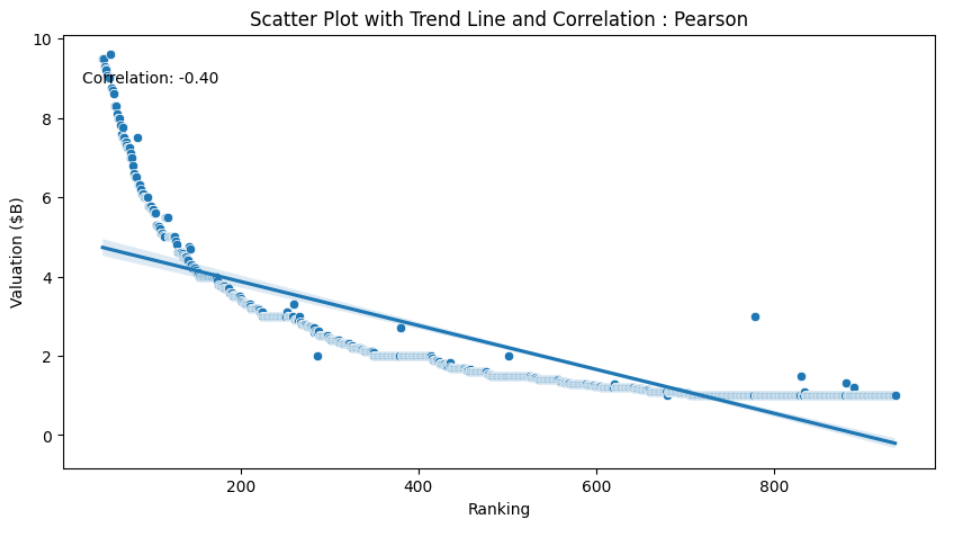


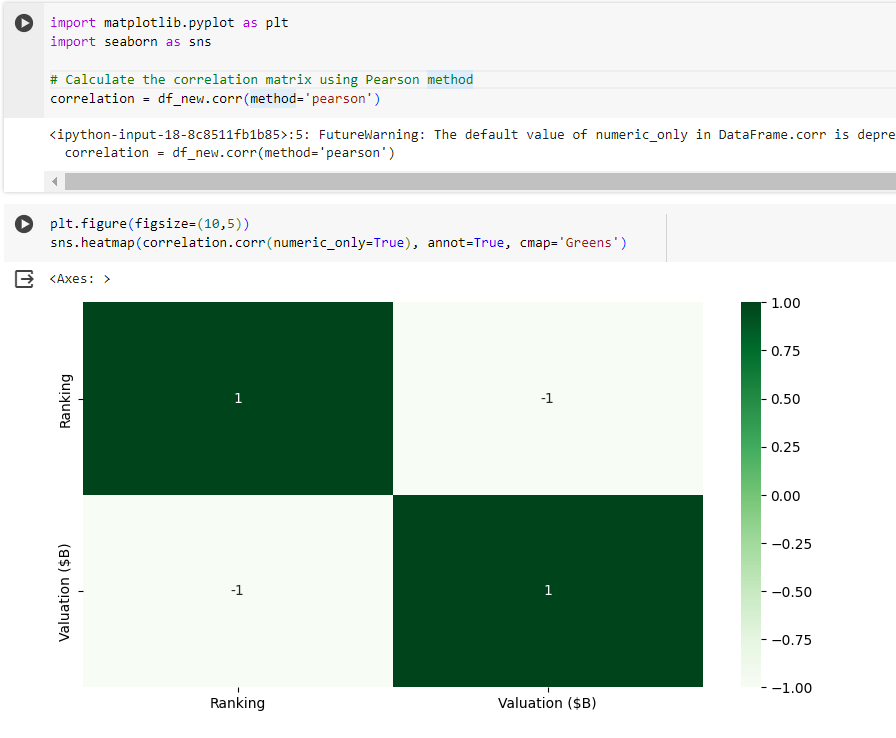


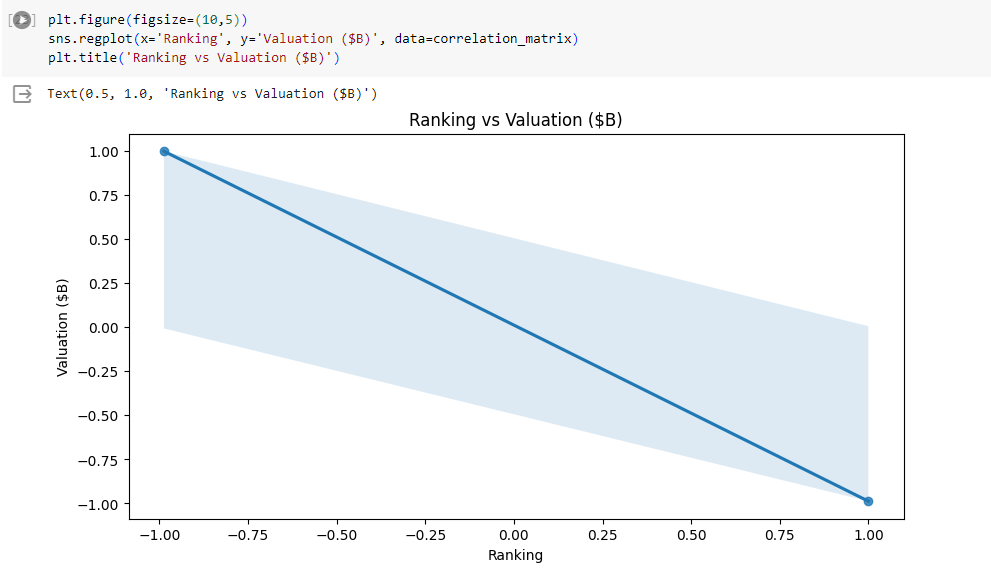


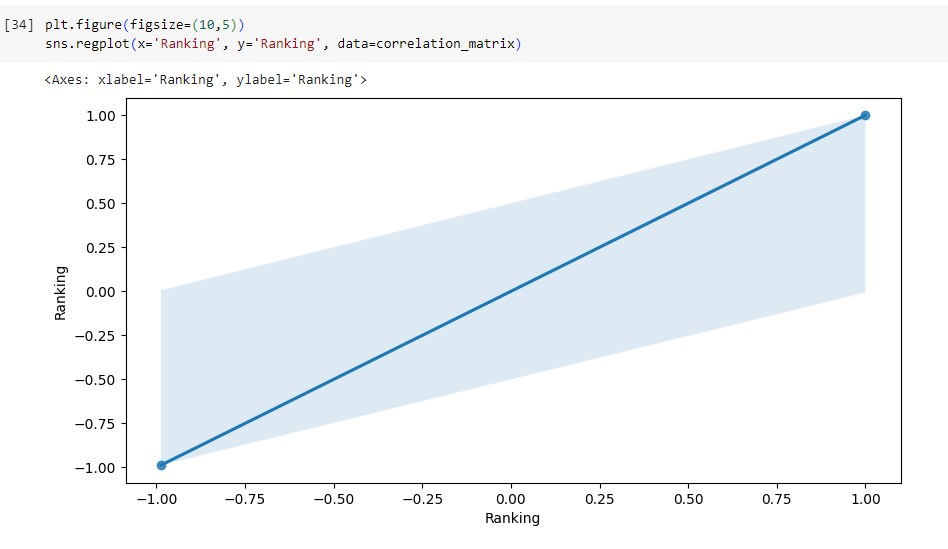
**2. Pearson Correlation**



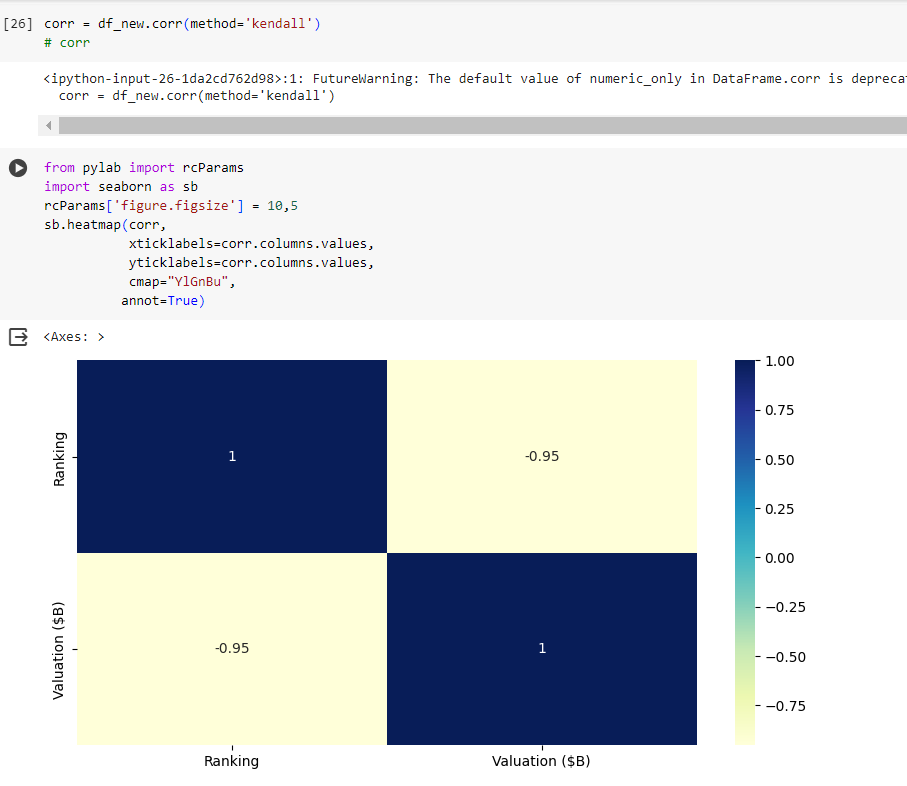


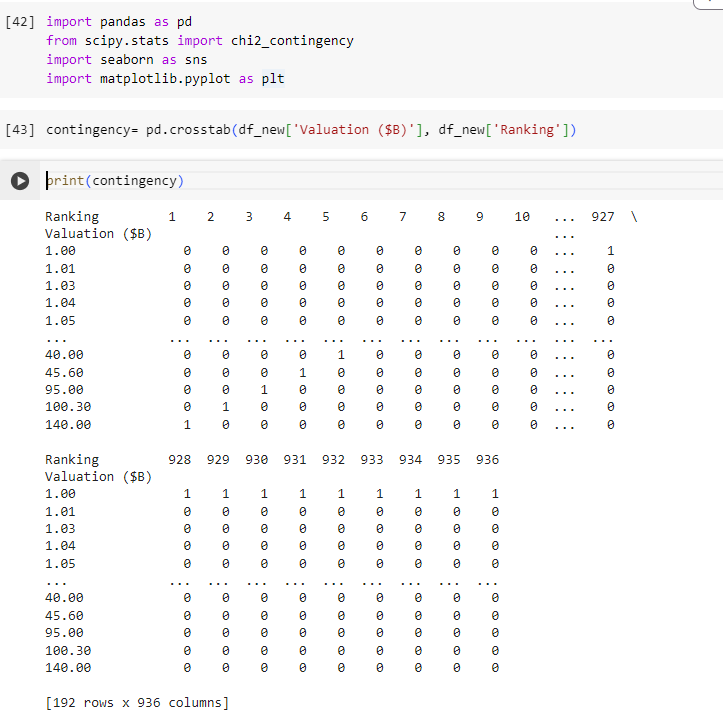


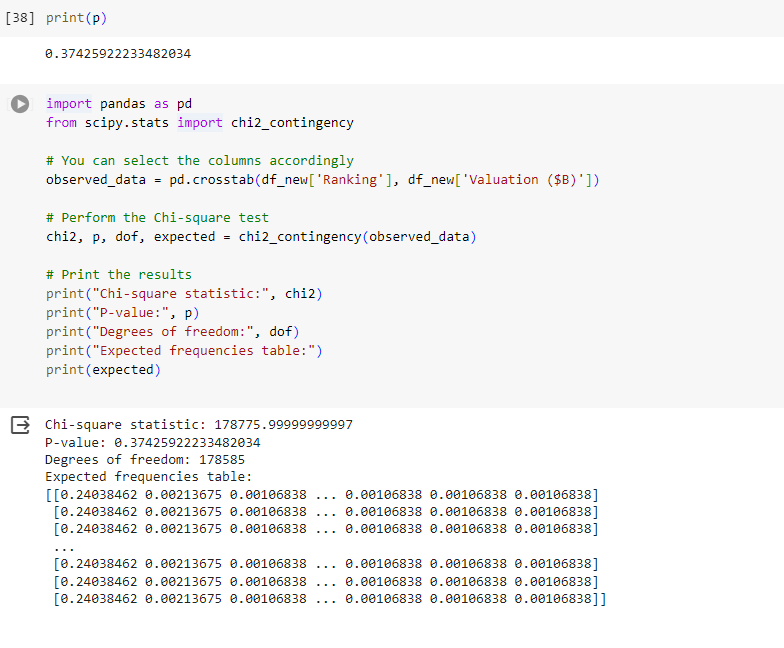




**3. Kendall’s Method**







**CONCLUSION : Thus, we have performed various correlation tests and got the output for my dataset.**