QUESTIONS AND ANSWERS:

Answer the following questions based on the given data set:

1. Are there any correlated features in the given dataset? If yes, state the correlation coefficient of the pair of features which are highly correlated.

Sol:  
  
-Yes, there are correlated features in the given dataset.

Feature 1 Feature 2 Correlation coefficient

| 60day\_eng\_rate | avg\_likes | 0.79 |
| --- | --- | --- |
| new\_post\_avg\_likes | avg\_likes | 0.89 |
| new\_post\_avg\_likes | 60day\_eng\_rate | 0.87 |
| Followers | Rank | -0.7 |

2. What is the frequency distribution of the following features?

○ Influence Score

○ Followers

○ Posts

Sol:

influence\_score

85 30

83 20

86 19

84 18

82 12

80 12

87 9

81 9

90 7

88 7

76 6

79 6

91 6

78 5

77 4

92 4

75 4

89 4

74 3

73 2

42 2

70 1

93 1

22 1

57 1

58 1

47 1

72 1

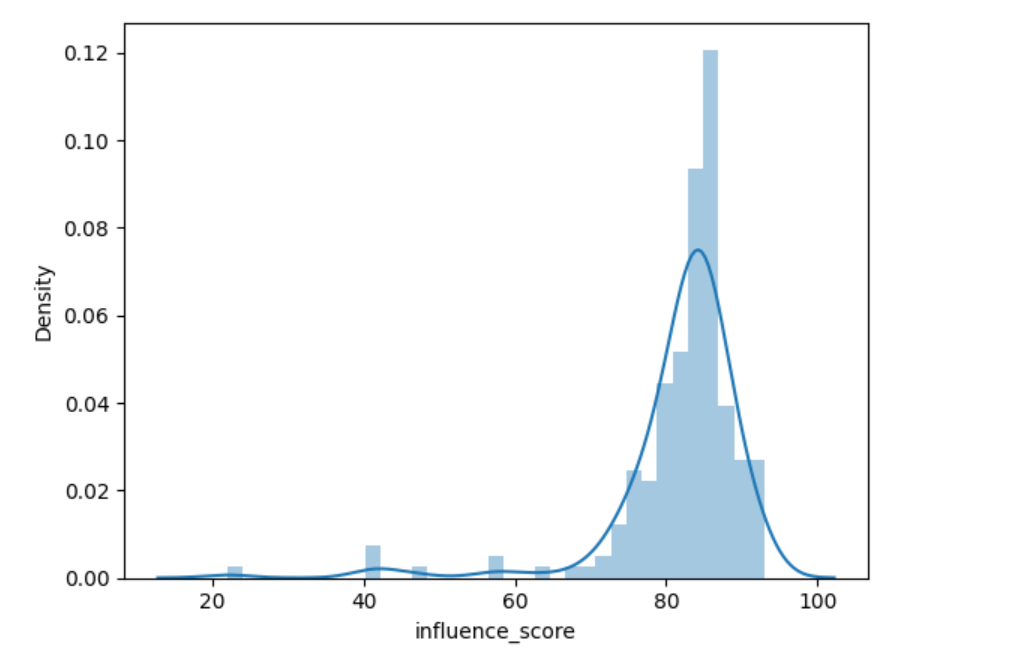
68 1

41 1

71 1

63 1

Name: count, dtype: int64



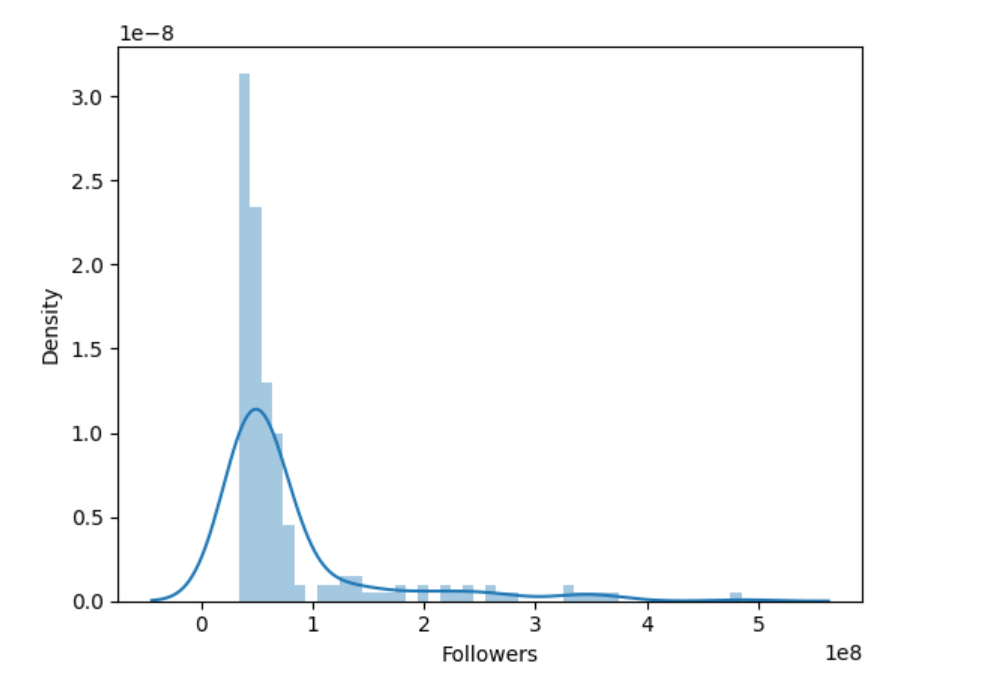
Distribution is= left skewed distribution

Skewness for 'influence\_score' column: -3.4524004652506624

The distribution is negatively skewed.

Followers:

Counter({68100000: 3, 55100000: 3, 52900000: 3, 36000000: 3, 33800000: 3, 33600000: 3, 70500000: 2, 63500000: 2, 53300000: 2, 47300000: 2, 46200000: 2, 44400000: 2, 44200000: 2, 43500000: 2, 43400000: 2, 41800000: 2, 41400000: 2, 40600000: 2, 39200000: 2, 38400000: 2, 36100000: 2, 35400000: 2, 485200000: 1, 370700000: 1, 363900000: 1, 348800000: 1, 339400000: 1, 333000000: 1, 330700000: 1, 276100000: 1, 273900000: 1, 260000000: 1, 258900000: 1, 241100000: 1, 239200000: 1, 226100000: 1, 223100000: 1, 215900000: 1, 202700000: 1, 199400000: 1, 183600000: 1, 179400000: 1, 173300000: 1, 154200000: 1, 153300000: 1, 141300000: 1, 140600000: 1, 136300000: 1, 133700000: 1, 125800000: 1, 125500000: 1, 121400000: 1, 121200000: 1, 112500000: 1, 106400000: 1, 88000000: 1, 87300000: 1, 83200000: 1, 83100000: 1, 82700000: 1, 82600000: 1, 76700000: 1, 76200000: 1, 75900000: 1, 75400000: 1, 75200000: 1, 73400000: 1, 72700000: 1, 71400000: 1, 71200000: 1, 71100000: 1, 70200000: 1, 69300000: 1, 67900000: 1, 67500000: 1, 66600000: 1, 66500000: 1, 65400000: 1, 64700000: 1, 64400000: 1, 63600000: 1, 63400000: 1, 63300000: 1, 62900000: 1, 62700000: 1, 60800000: 1, 60400000: 1, 59000000: 1, 58800000: 1, 58600000: 1, 58300000: 1, 58000000: 1, 56800000: 1, 56000000: 1, 55800000: 1, 55500000: 1, 55200000: 1, 54500000: 1, 54300000: 1, 54100000: 1, 53700000: 1, 53500000: 1, 51900000: 1, 51700000: 1, 51200000: 1, 51100000: 1, 51000000: 1, 50900000: 1, 50300000: 1, 50000000: 1, 49900000: 1, 49700000: 1, 49600000: 1, 49200000: 1, 49100000: 1, 49000000: 1, 48900000: 1, 48600000: 1, 48500000: 1, 48300000: 1, 48200000: 1, 48100000: 1, 47600000: 1, 47400000: 1, 46500000: 1, 46400000: 1, 46100000: 1, 45800000: 1, 45600000: 1, 45400000: 1, 45100000: 1, 44600000: 1, 44100000: 1, 43800000: 1, 43100000: 1, 42900000: 1, 42500000: 1, 42100000: 1, 41100000: 1, 40800000: 1, 40700000: 1, 40500000: 1, 40300000: 1, 39800000: 1, 39700000: 1, 39500000: 1, 39400000: 1, 39100000: 1, 39000000: 1, 38900000: 1, 38500000: 1, 38300000: 1, 37800000: 1, 37300000: 1, 36800000: 1, 36700000: 1, 36600000: 1, 36200000: 1, 35900000: 1, 35800000: 1, 35700000: 1, 35600000: 1, 35100000: 1, 34900000: 1, 34800000: 1, 34700000: 1, 34500000: 1, 34100000: 1, 34000000: 1, 33700000: 1, 33500000: 1, 33400000: 1})



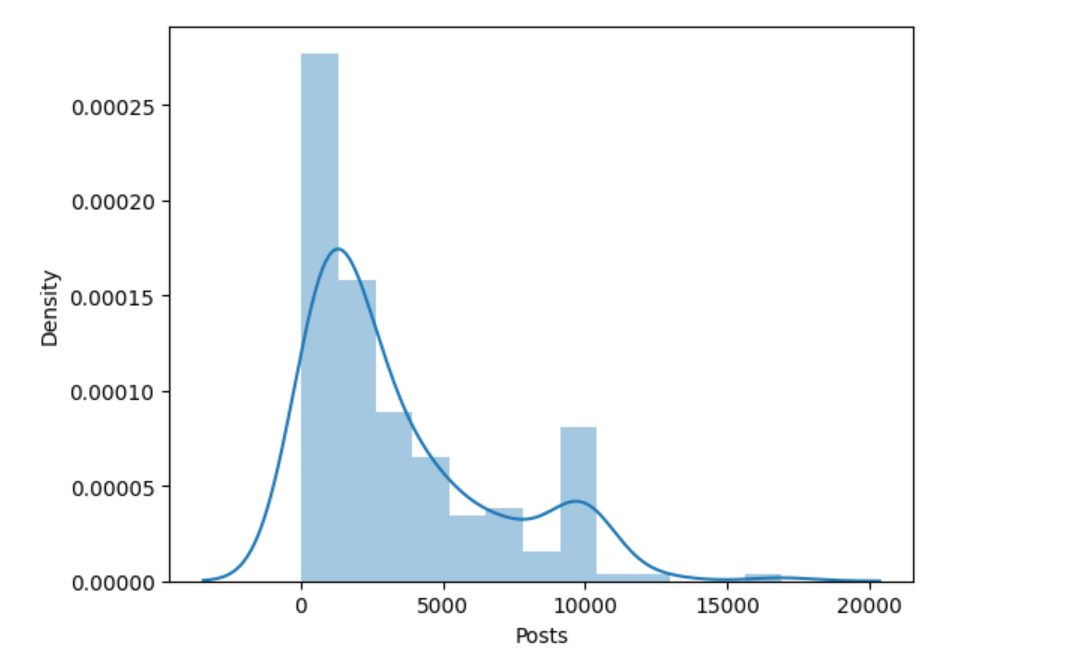
Distribution is= right skewed distribution

Skewness for 'Followers' column: 2.78788081585234

The distribution is positively skewed.

Posts:

Counter({10000: 11, 1200: 9, 1900: 5, 2000: 5, 3200: 4, 2100: 4, 2400: 4, 1700: 4, 1300: 4, 1800: 4, 2300: 4, 9900: 4, 7400: 3, 1400: 3, 1600: 3, 3600: 3, 3000: 3, 1000: 3, 4100: 2, 6400: 2, 4400: 2, 5300: 2, 4800: 2, 3300: 2, 2500: 2, 549: 2, 9800: 2, 5100: 2, 2700: 2, 2800: 2, 4700: 2, 1100: 2, 3700: 2, 4200: 2, 1500: 2, 3400: 1, 7000: 1, 915: 1, 6800: 1, 5000: 1, 5700: 1, 671: 1, 973: 1, 542: 1, 8300: 1, 3500: 1, 105: 1, 7100: 1, 5200: 1, 710: 1, 893: 1, 12800: 1, 887: 1, 285: 1, 379: 1, 7200: 1, 785: 1, 2900: 1, 876: 1, 854: 1, 8200: 1, 10100: 1, 290: 1, 899: 1, 881: 1, 669: 1, 5600: 1, 2200: 1, 439: 1, 849: 1, 10400: 1, 69: 1, 620: 1, 423: 1, 6500: 1, 179: 1, 9000: 1, 6600: 1, 207: 1, 989: 1, 618: 1, 165: 1, 35: 1, 16900: 1, 695: 1, 992: 1, 28: 1, 617: 1, 4900: 1, 735: 1, 9300: 1, 20: 1, 3100: 1, 104: 1, 4500: 1, 328: 1, 107: 1, 11600: 1, 751: 1, 514: 1, 742: 1, 5400: 1, 115: 1, 8900: 1, 397: 1, 890: 1, 691: 1, 59: 1, 120: 1, 946: 1, 682: 1, 7700: 1, 7500: 1, 6300: 1, 180: 1, 377: 1, 9400: 1, 525: 1, 4600: 1, 865: 1, 3800: 1})



Distribution is:

Skewness for 'Posts' column: 1.224704590123004

The distribution is positively skewed.

The distribution is multimodal with 3 modes.

Chi-square Statistic: 91.97359409105114

P-value: 0.9731325772678894

Degrees of Freedom: 120

The data follows a Poisson distribution.

Anderson-Darling Test Statistic: 1.067628923258411

At 10.0% significance level, the data follows an exponential distribution.

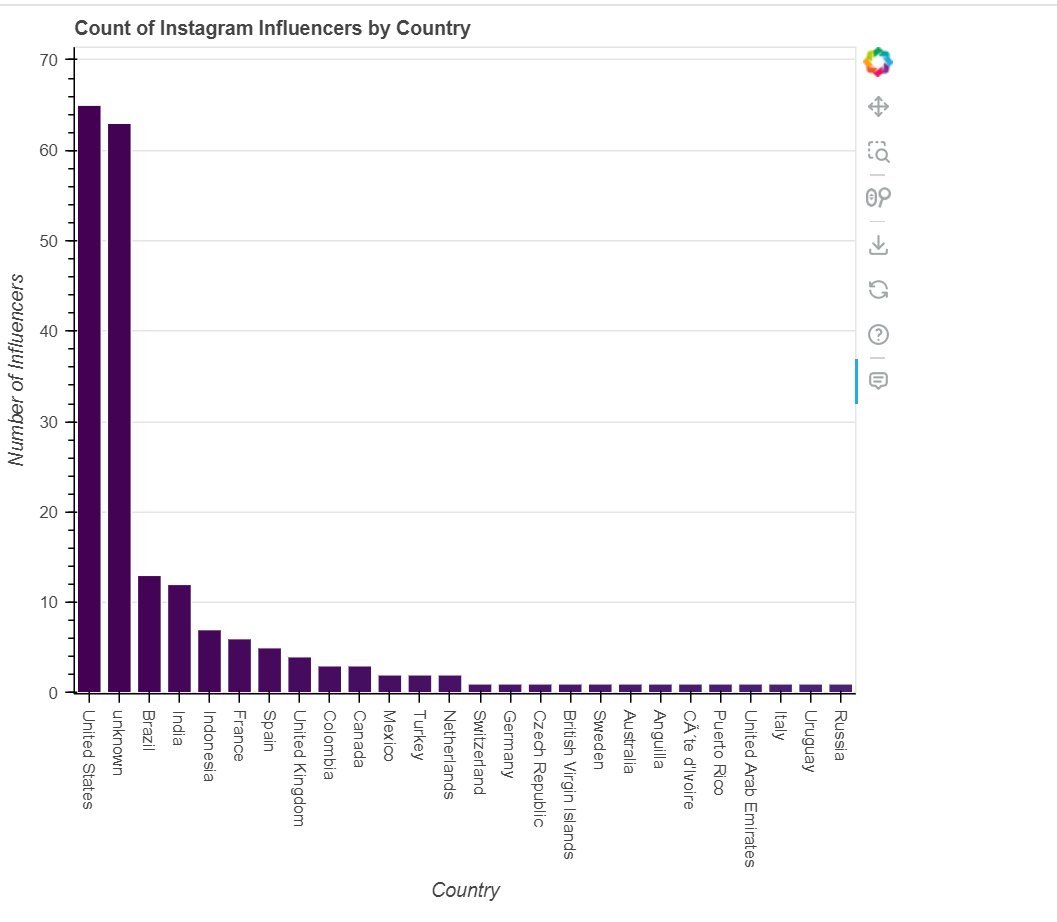
3. Which country houses the highest number of Instagram Influencers? Please show the count of Instagram influencers in different countries using barchart.

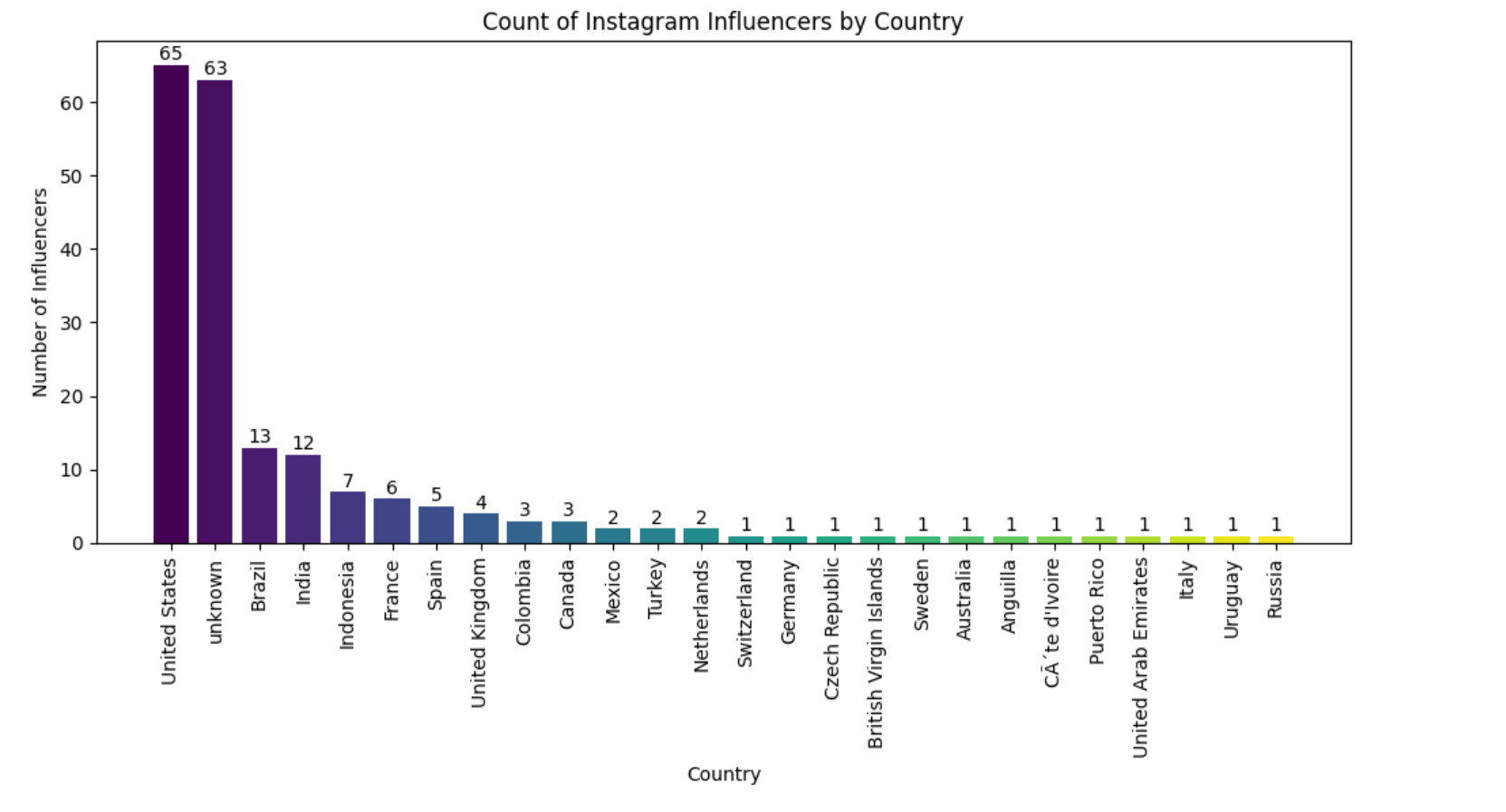
Sol:

| **Rank** | **channel\_info** | **influence\_score** | **Followers** | **avg\_likes** | **Posts** | **60day\_eng\_rate** | **new\_post\_avg\_likes** | **total\_likes** | **country\_or\_region** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | \ncristiano | 92 | 485200000 | 8700000 | 3400 | 0.013 | 6300000 | 29100000000 | Spain |
| 2 | kyliejenner | 91 | 370700000 | 8200000 | 7000 | 0.014 | 5000000 | 57400000000 | United States |
| 3 | \nleomessi | 90 | 363900000 | 6700000 | 915 | 0.010 | 3500000 | 6100000000 | unknown |
| 4 | \nselenagomez | 93 | 348800000 | 6100000 | 1900 | 0.005 | 1700000 | 11400000000 | United States |

Country with the most influencers: United States

Number of influencers: 65





4. Who are the top 10 influencers in the given dataset based on the following features

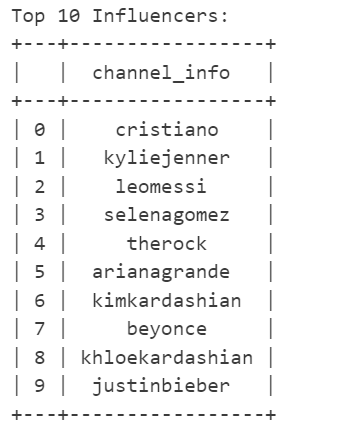
● Followers

● Average likes

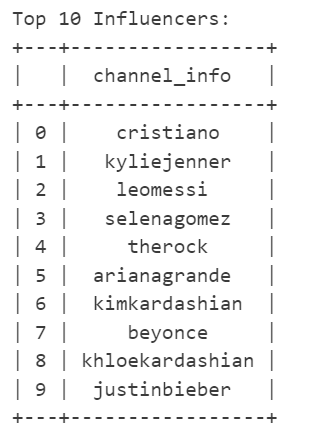
● Total Likes

Sol:

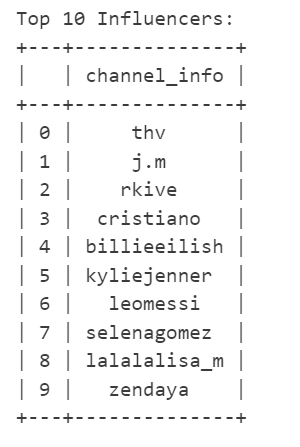
All 3 included together:



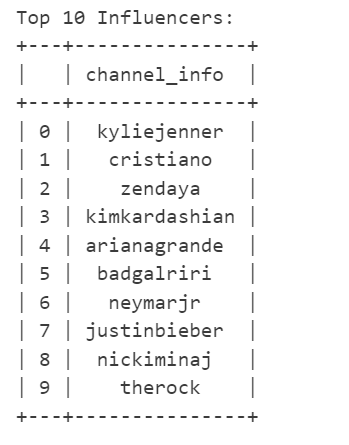
Only Followers:



Only Average likes:



Only Total Likes:



5. Describe the relationship between the following pairs of features using a suitable graph

● Followers and Total Likes

● Followers and Influence Score

● Posts and Average likes

● Posts and Influence Score

Sol:

Correlation between Followers and Total Likes: 0.67

P-value: 0.0000

The correlation is statistically significant.

The correlation is considered moderate.

Correlation between Followers and Influence Score: 0.37

P-value: 0.0000

The correlation is statistically significant.

The correlation is considered moderate.

Correlation between Posts and Average Likes: -0.35

P-value: 0.0000

The correlation is statistically significant.

The correlation is considered moderate.

Correlation between Posts and Influence Score: 0.17

P-value: 0.0172

The correlation is statistically significant.

The correlation is considered weak.

