**FINAL PROJECT**

**PROJECT - 3 (INSTAGRAM INFLUENCERS)**

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

**A:** Yes, there are correlated features in the given dataset. Below are the pair of

Features with their related coefficients:

**Correlation Matrix**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Influence Score | Followers | Avg. Likes | Posts | 60-Day Eng Rate | New Post Avg. Likes | Total Likes |
| Influence Score | 1.000000 | 0.365428 | 0.047803 | 0.168343 | -0.072460 | 0.039693 | 0.235259 |
| Followers | 0.365428 | 1.000000 | 0.346644 | 0.069219 | -0.107439 | 0.260220 | 0.667469 |
| Avg. Likes | 0.047803 | 0.346644 | 1.000000 | -0.348120 | 0.789366 | 0.892447 | 0.361560 |
| Posts | 0.168343 | 0.069219 | -0.348120 | 1.000000 | -0.333880 | -0.283638 | 0.232106 |
| 60-Day Eng Rate | -0.072460 | -0.107439 | 0.789366 | -0.333880 | 1.000000 | 0.866303 | -0.001412 |
| New Post Avg. Likes | 0.039693 | 0.260220 | 0.892447 | -0.283638 | 0.866303 | 1.000000 | 0.351079 |
| Total Likes | 0.235259 | 0.667469 | 0.361560 | 0.232106 | -0.001412 | 0.351079 | 1.000000 |

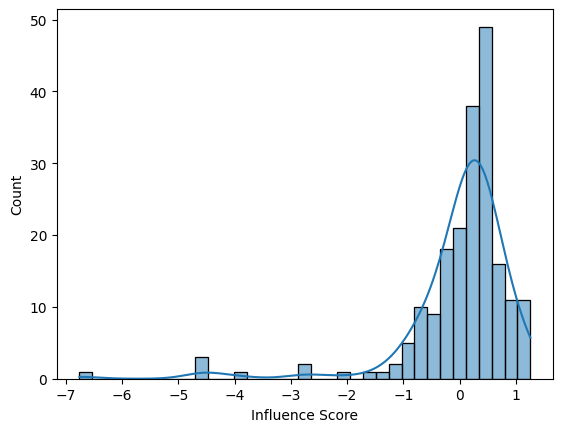
**Highly Correlated Features**

|  |  |  |  |
| --- | --- | --- | --- |
| **FEATURE 1** | **FEATURE 2** | **CORRELATION** | **COEFFICIENT** |
| Avg. Likes | New Post Avg. Likes | Positive | 0.892447 |
| 60-Day Eng Rate | New Post Avg. Likes | Positive | 0.866303 |
| Avg. Likes | 60-Day Eng Rate | Positive | 0.789366 |
| Followers | Total Likes | Positive | 0.667469 |

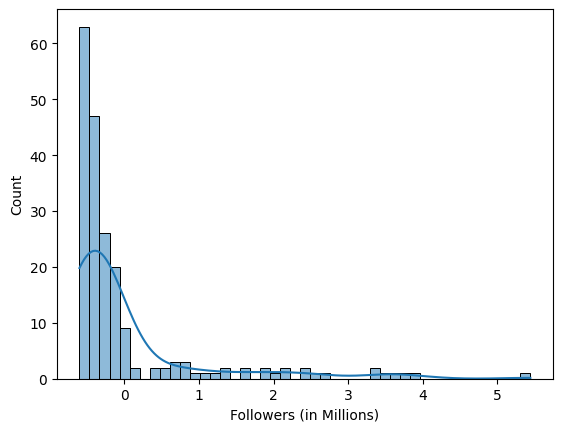
1. **Q:** What is the frequency distribution of the features Influence Score, Followers, Posts?

**A:** Below are the answers:

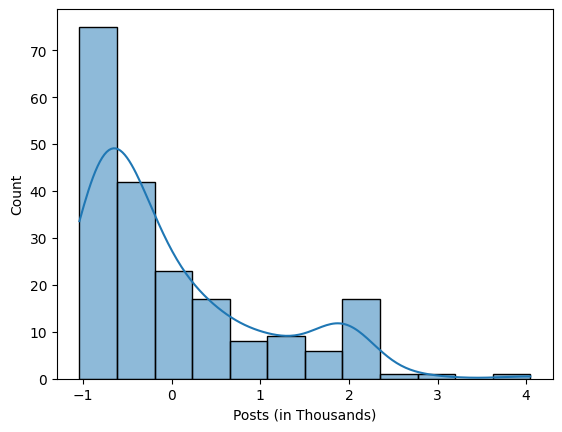
Influence Score - **Right Skewed Normal Distribution**



Followers – **Left Skewed Normal Distribution**

****

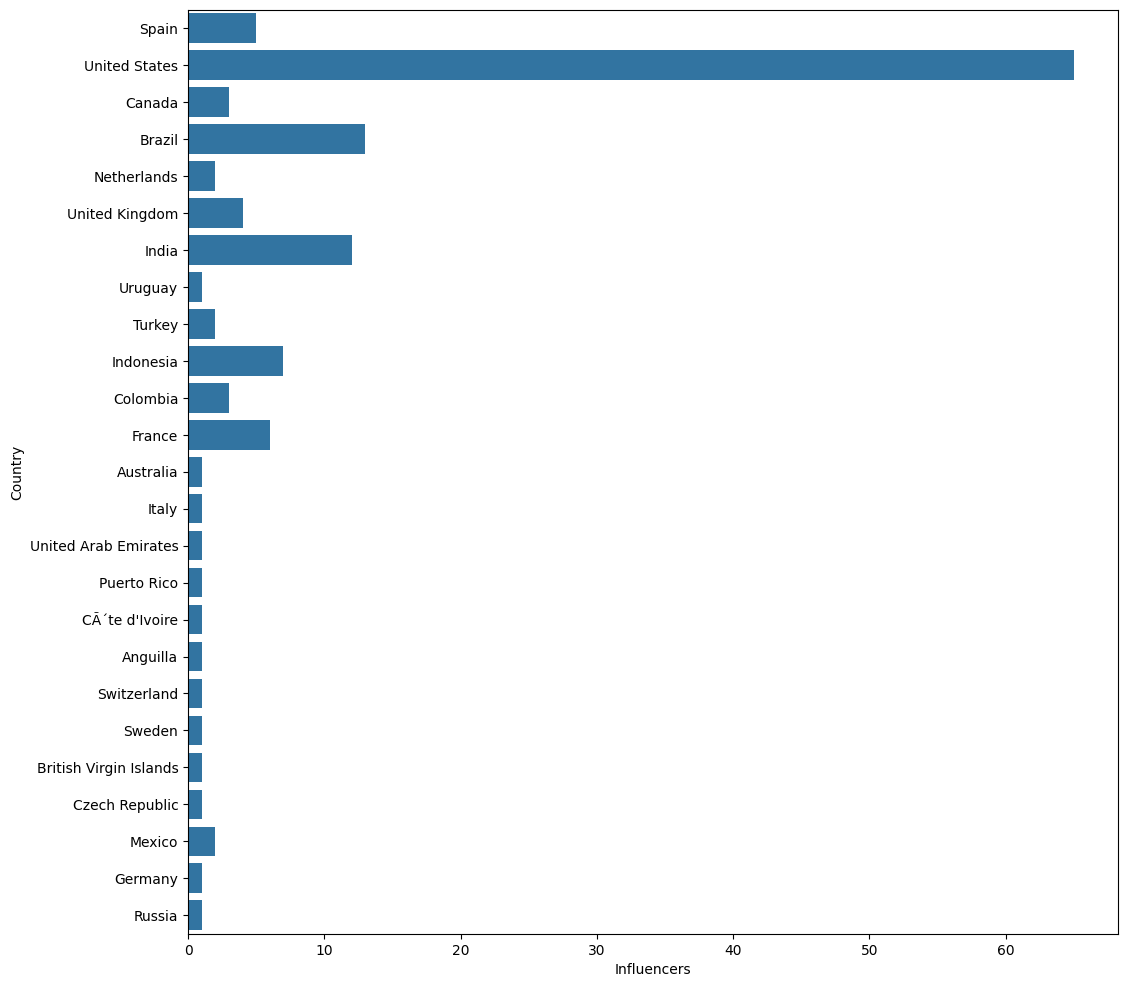
Posts – **Left Skewed Normal Distribution**



1. **Q:** Which country houses the highest number of Instagram Influencers? Please show the

count of Instagram influencers in different countries using bar chart.

**A:** United States with **65 influencers**



1. **Q:** Who are the top 10 influencers in the given dataset based on the Followers, Average

Likes, Total Likes?

**A: Top 10 Influencers based on Followers**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **CHANNEL INFO** | **FOLLOWERS (IN MILLIONS)** |
| 1 | cristiano | 485.2 |
| 2 | kyliejenner | 370.7 |
| 3 | leomessi | 363.9 |
| 4 | selenagomez | 348.8 |
| 5 | therock | 339.4 |
| 6 | arianagrande | 333.0 |
| 7 | kimkardashian | 330.7 |
| 8 | beyonce | 276.1 |
| 9 | khloekardashian | 273.9 |
| 10 | justinbieber | 260.0 |

**Top 10 Influencers based on Average Likes**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **CHANNEL INFO** | **AVERAGE LIKES (IN MILLIONS)** |
| 1 | thv | 15.4 |
| 2 | j.m | 14.4 |
| 3 | rkive | 10.6 |
| 4 | cristiano | 8.7 |
| 5 | billieeilish | 8.4 |
| 6 | kyliejenner | 8.2 |
| 7 | leomessi | 6.7 |
| 8 | selenagomez | 6.1 |
| 9 | lalalalisa\_m | 6.0 |
| 10 | zendaya | 5.9 |

**Top 10 Influencers based on Total Likes**

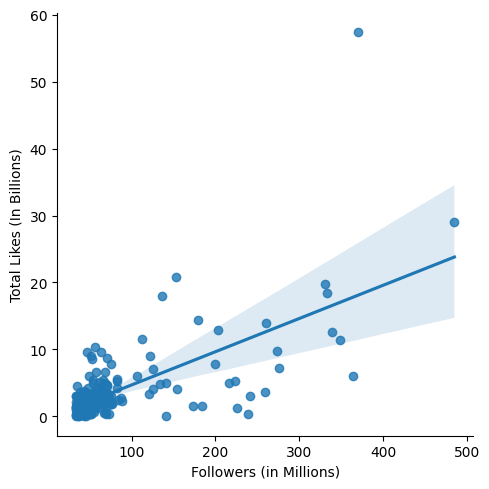
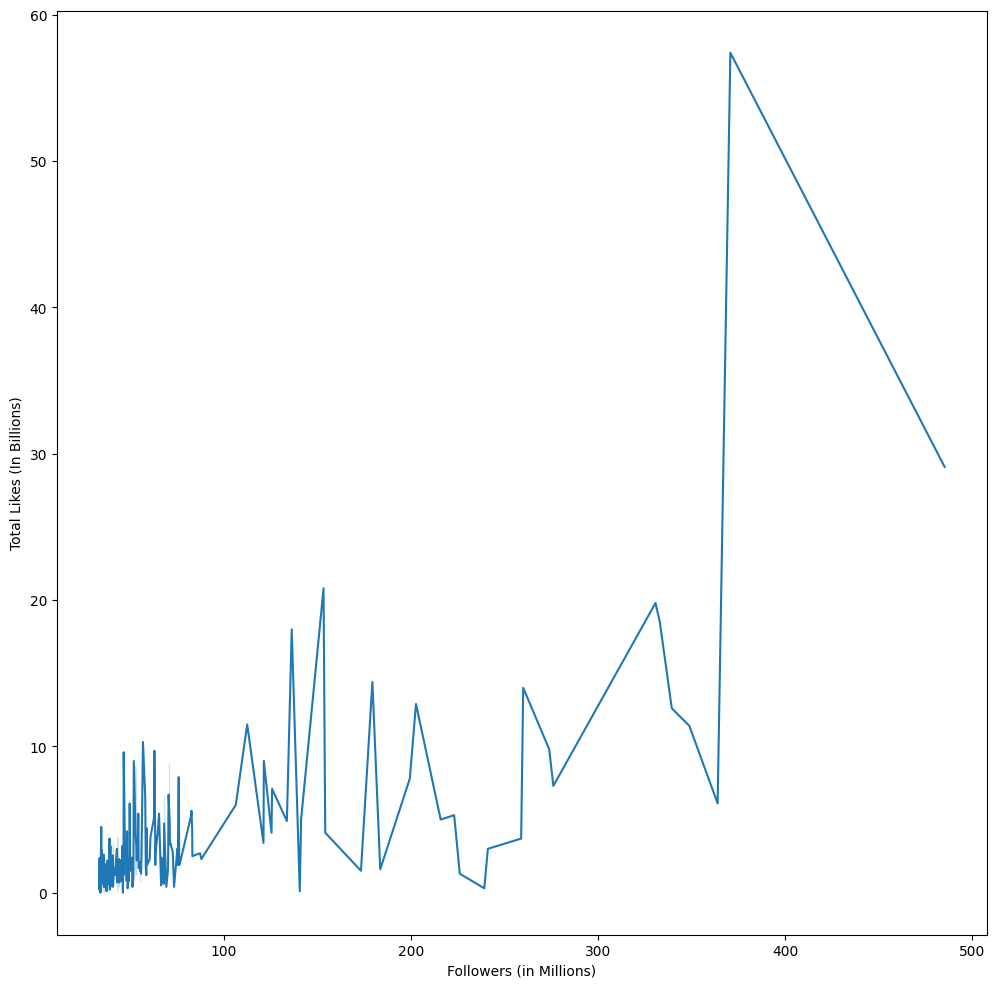
|  |  |  |
| --- | --- | --- |
| **S.NO** | **CHANNEL INFO** | **TOTAL LIKES (IN BILLIONS)** |
| 1 | kyliejenner | 57.4 |
| 2 | cristiano | 29.1 |
| 3 | zendaya | 20.8 |
| 4 | kimkardashian | 19.8 |
| 5 | arianagrande | 18.5 |
| 6 | badgalriri | 18.0 |
| 7 | neymarjr | 14.4 |
| 8 | justinbieber | 14.0 |
| 9 | nickiminaj | 12.9 |
| 10 | therock | 12.6 |

1. **Q:** Describe the relationship between the following pairs of features using a suitable graph:

* Followers & Total Likes
* Followers and Influence Score
* Posts and Average Likes
* Posts and Influence Score

**A:** Below are the answers:

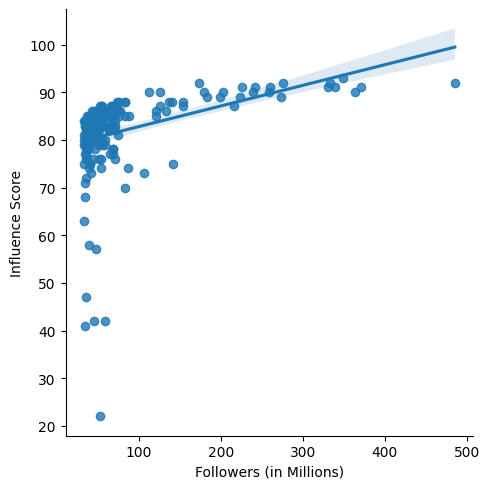
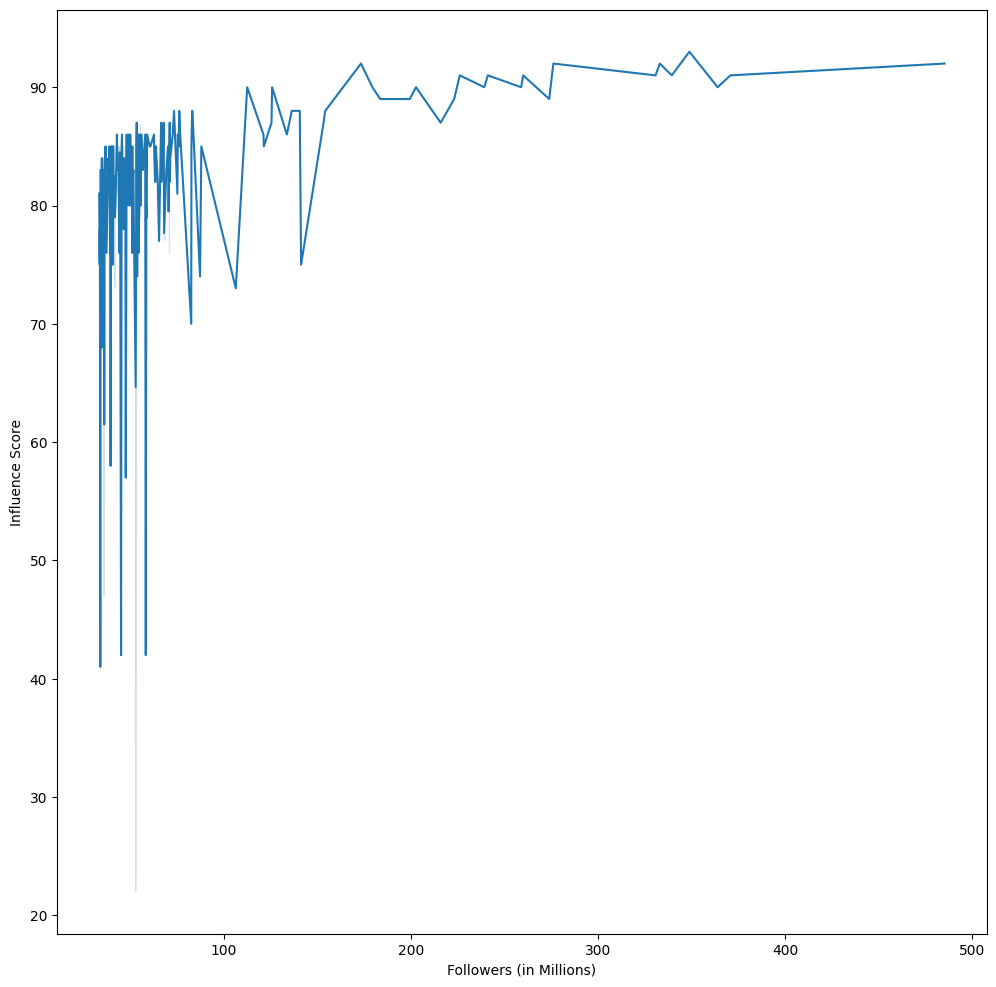
**Followers vs Total Likes**

**Inferences:**

* Followers have a very good correlation with Total Likes with **67%**
* Followers have a positive correlation with Total Likes
* Followers has a non-linear relationship with Total Likes
* Majority of relationship exists with **Followers upto 100 Millions & Total Likes upto 10 Billions**
* Above 100 Million Followers, the relationship starts to fade away between the both
* The maximum number of likes is obtained with followers close to 400 Millions

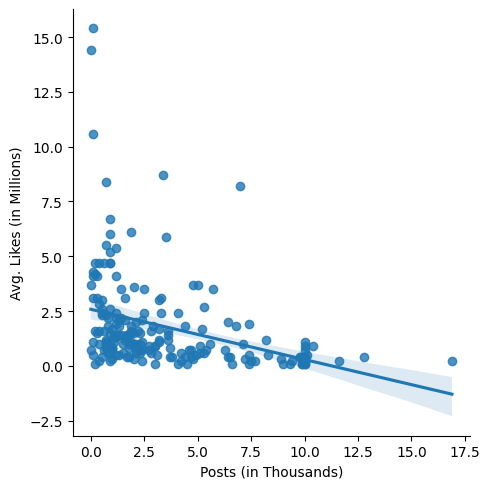
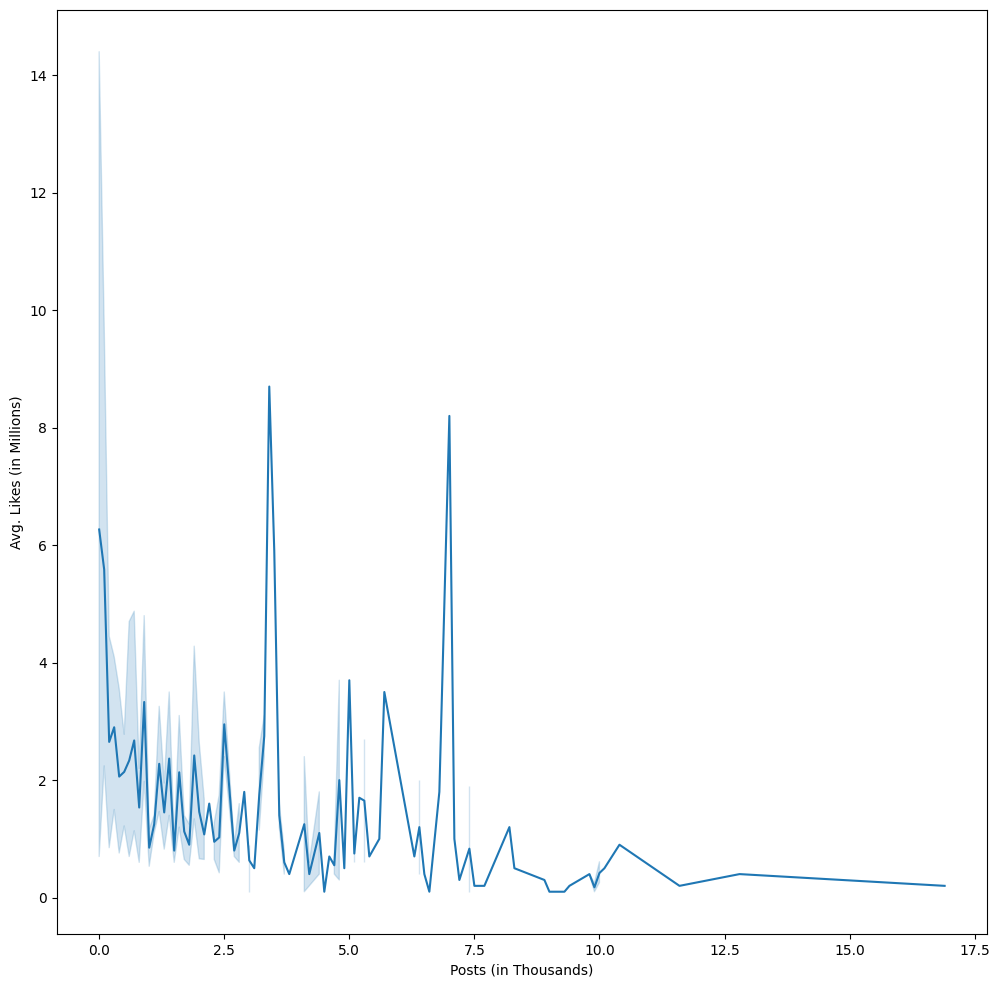
**Followers vs Influence Score**

**Inferences:**

* Followers have a below average correlation with Influence Score with **36.5%**
* Followers have a positive correlation with Influence Score
* Followers has a non-linear relationship with Influence Score
* Majority of relationship exists with **Followers upto 100 Millions & Influence Score between 70 to 90**
* With above 100 Million Followers, the relationship starts to fade away between the both.
* Also, with Followers below 100 Million and Influence Score of less than 70, the relationship starts to fade away between the both.
* The maximum Influence Score is obtained with followers above 100 Million and the minimum Influence Score is obtained for 50 Million followers

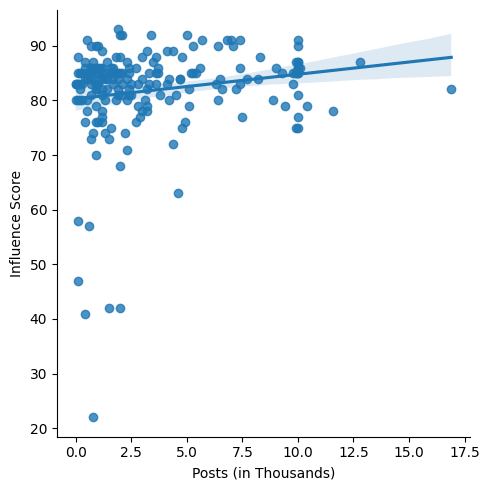
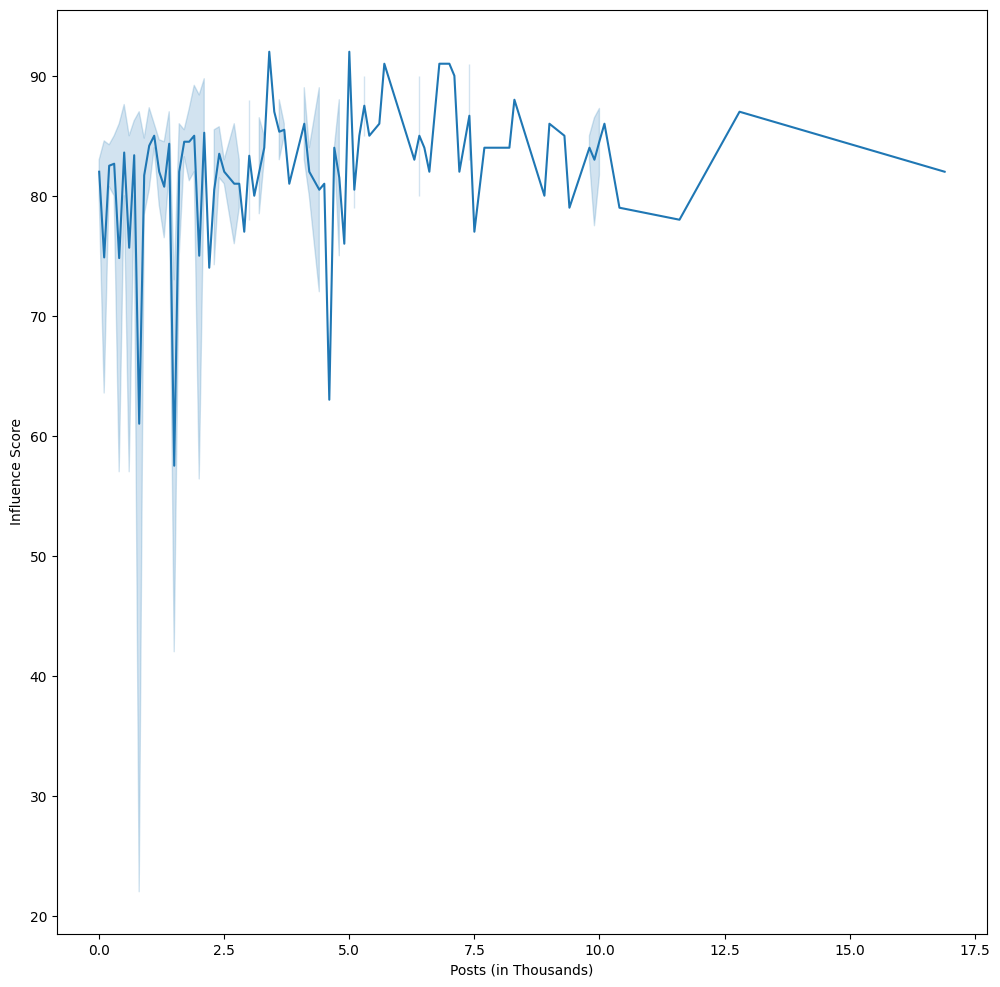
**Posts vs Average Likes**

** **

**Inferences:**

* Posts have a below average correlation with Average Likes with **34.8%**
* Posts have a negative correlation with Average Likes
* Posts has a non-linear relationship with Average Likes
* Majority of relationship exists with **Number of Posts upto 10 Thousands & Average Likes between 0 to 2.5 Millions**
* Above 10 Thousand Posts, the relationship starts to fade away between the both
* Also, with number of posts below 7.5 Thousands and Average Likes above 2.5 Millions, the relationship starts to fade away between the both.
* The maximum number of likes of **15 Millions** is obtained is at minimum number of Posts (almost 0) and the minimum number of likes of closer to **0.25 Millions** is obtained at maximum number of Posts (17.5 Thousands)

**Posts vs Influence Score**

** **

**Inferences:**

* Posts have a very below average correlation with Influence Score with **16.8%**
* Posts have a positive correlation with Influence Score
* Posts has a non-linear relationship with Influence Score
* Majority of relationship exists with **Number of Posts upto 5 Thousands & Influence Score over 70 and above**
* Above 10 Thousand Posts, the relationship starts to fade away between the both
* Also, with number of posts below 5 Thousands and Influence Score of below 70, the relationship starts to fade away between the both.
* The maximum Influence Score is obtained is at minimum number of Posts between 0 to 10 Thousands and the minimum number of likes is obtained at number of Posts around 1 Thousands