Implementation of Unsupervised Machine Learning to Create Clusters in Main-Game Dataset

By: Azis Muslim
Role: Data Scientist

Background

In Main-Game Datasets consisted by 631 rows and 127 columns and It would be laborious to groups the dataset to some clusters. The easiest way to create clusters in large dataset especially with high dimensional features was by utilizing unsupervised machine learning. Unsupervised that would be used in this research was k-means.

Main-Game Dataset Overview

|	Unnamed: 0	Country	Gender	Game	Total Follower	Broadcast Hours	PaidStarPerWatchedHour	Character_Facet_Cont_Rigidity	Character_Facet_Cont_AchievementStriving
0	0	ID	Male	MLBB	96751.0	104	0.004113	0.663442	0.718679
1	1	ID	Male	8 Ball Pool	4781.0	174	0.002855	0.543255	0.427887
2	2	ID	Female	Free Fire 0 Battlegrounds	68246.0	83	0.003007	0.527279	0.399873
3	3	ID	Male	MLBB	142037.0	127	0.005171	0.677222	0.603534
4	4	ID	Male	Free Fire 0 Battlegrounds	15416.0	86	0.001213	0.238194	0.543285
...	...	...	...	...	...	...	...	...	...
626	942	VN	Male	PUBG	330969.0	182	0.001082	0.534637	0.399738
627	943	VN	Male	MU Online	33450.0	102	817000.000000	0.664025	0.597952
628	944	VN	Male	Age of Empires	29000.0	103	0.000724	0.561336	0.378814
629	945	VN	Female	PUBG	19771.0	143	0.000433	0.600844	0.364150
630	946	VN	Male	League of Legends	515000.0	105	0.002266	0.664025	0.597952

631 rows × 127 columns

Feature Importance Analysis
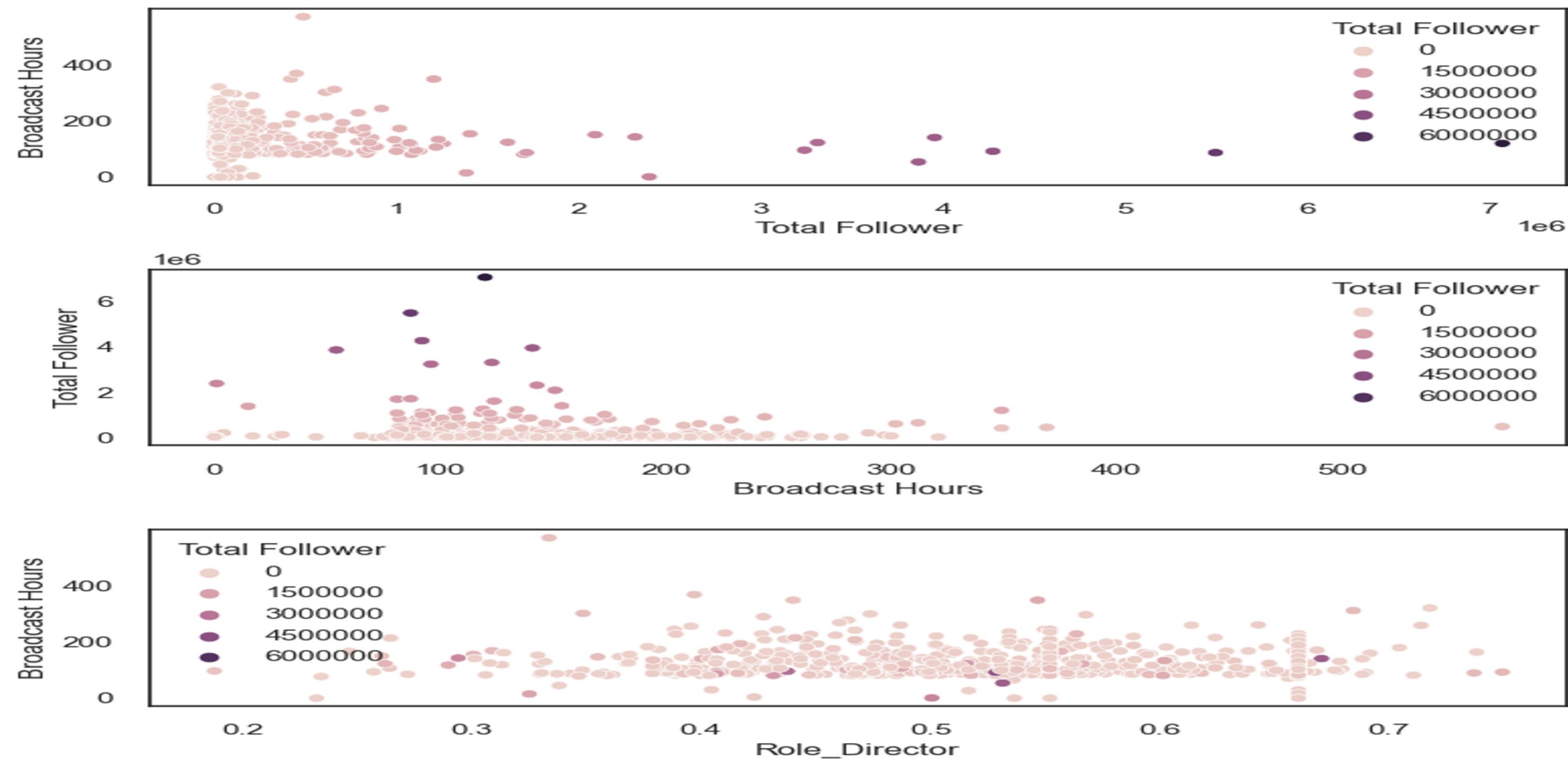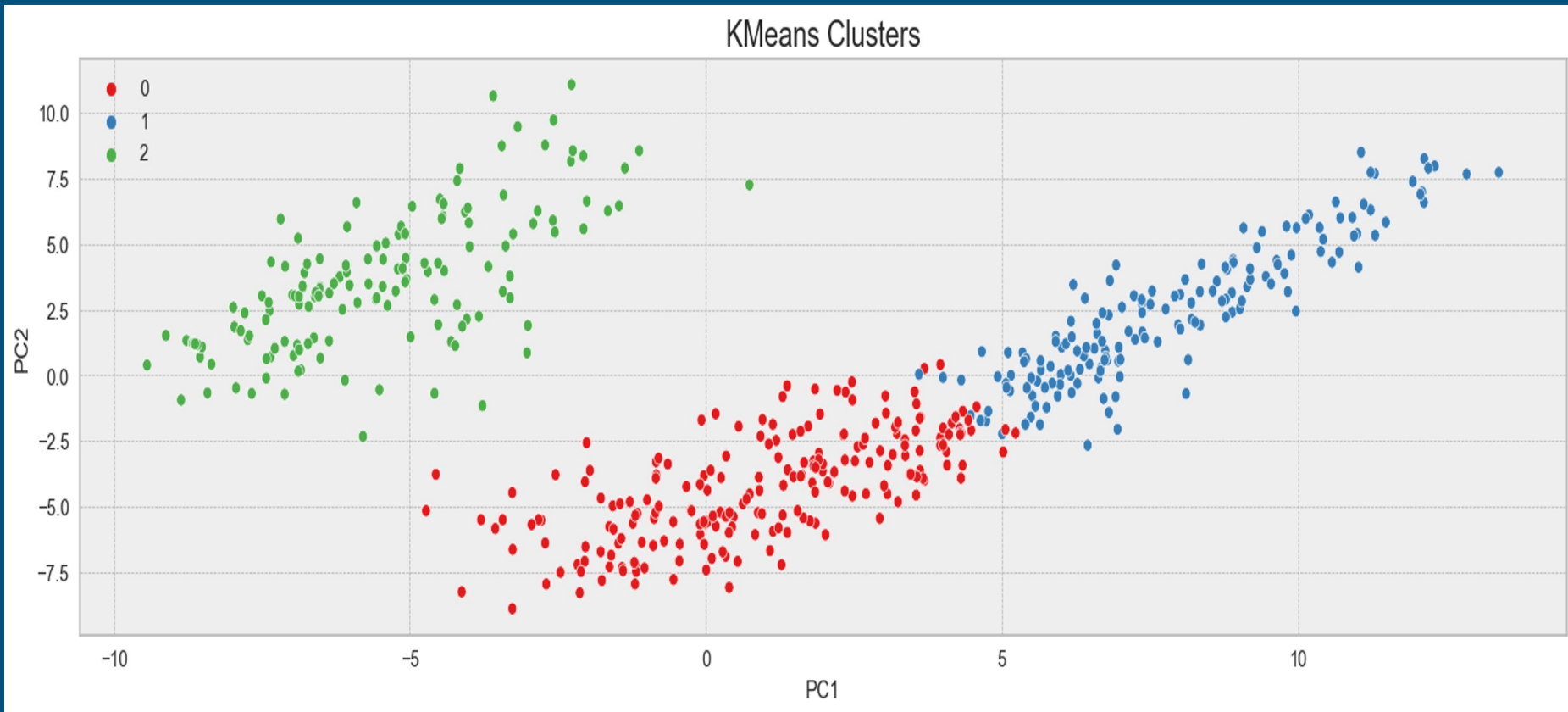
Feature Importance Analysis
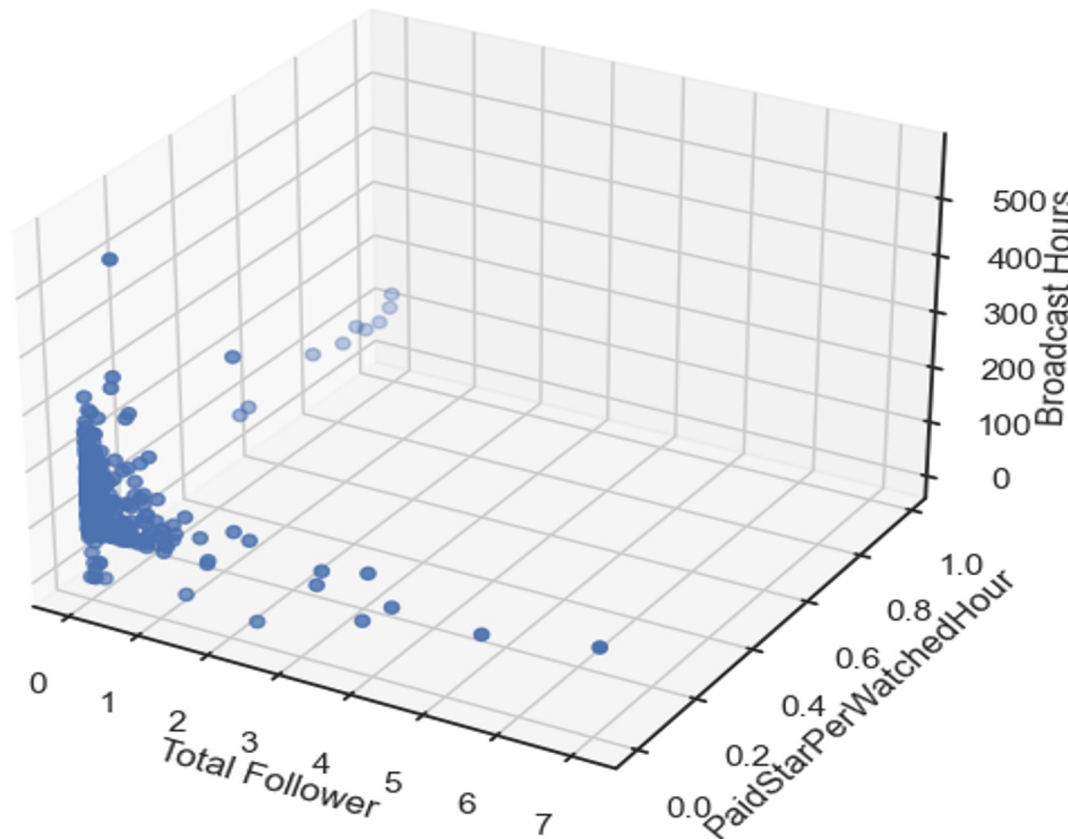
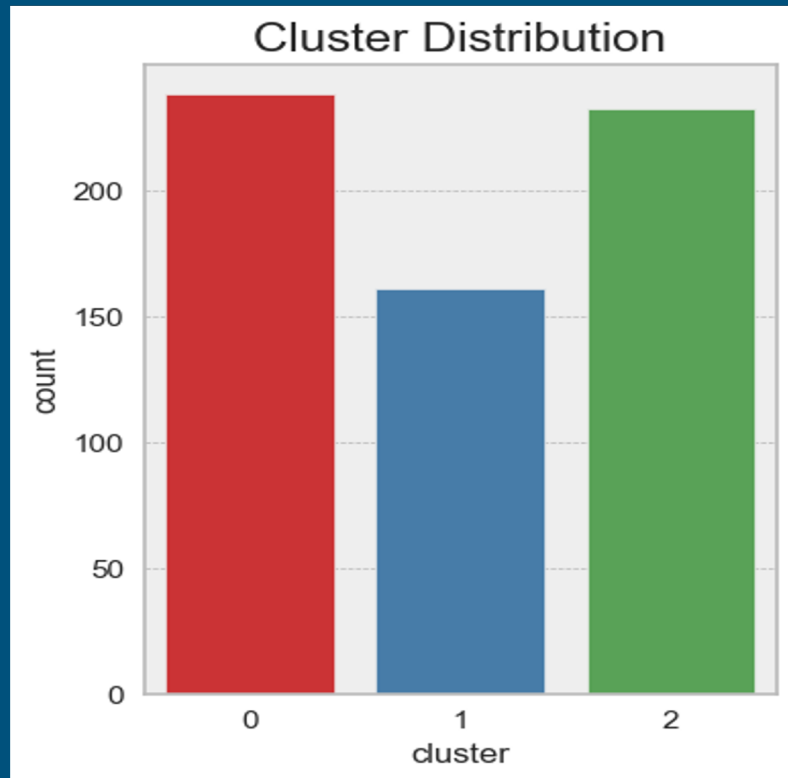# Principal Component Analysis (PCA)

PCA with 2 components
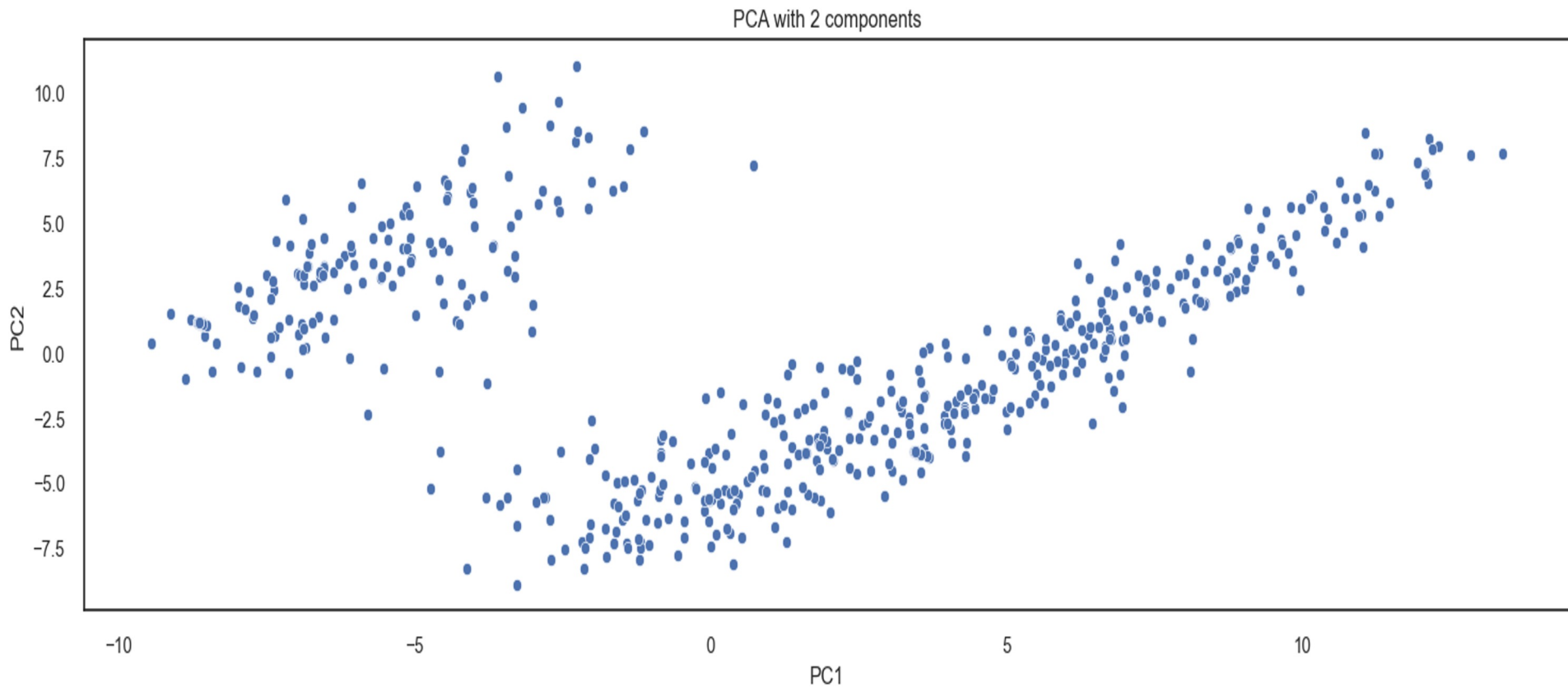
Elbow Method
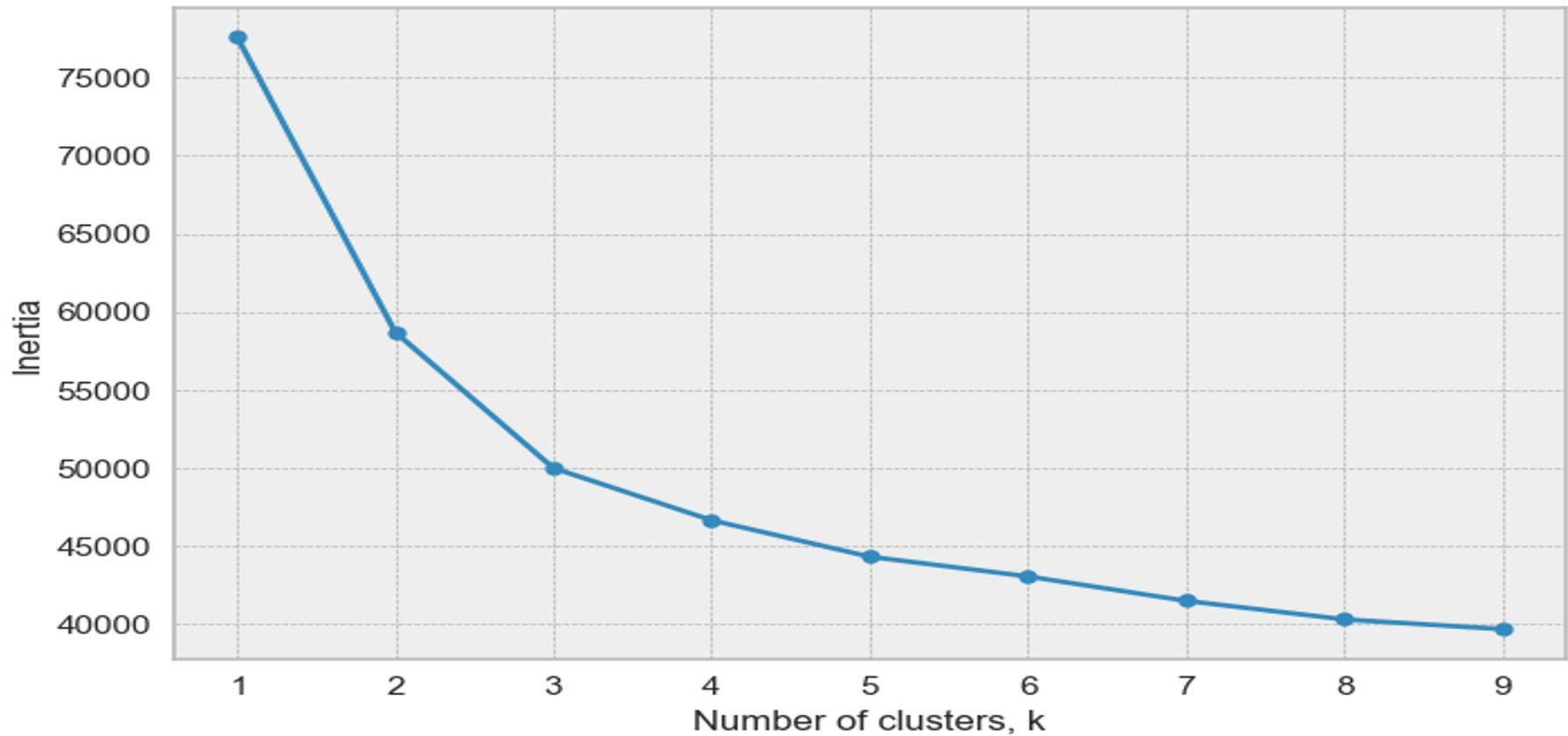
PCA Visualization

Clusters Count for K-means

Final Result

Country	Game	Total Follower	Broadcast Hours	PaidStarPerWatchedHour	Cluster
VN	Free Fire 0 Battlegrounds	311785.0	100	0.000803	1
ID	Free Fire 0 Battlegrounds	57532.0	96	0.010847	1
VN	The Last of Us	196718.0	145	0.005211	1
PH	Grand Theft Auto V	37413.0	168	0.005550	1
ID	MLBB	7987.0	141	0.003610	0
PH	MLBB	51215.0	109	0.012097	1
PH	League of Legends	151274.0	214	0.031874	2
ID	Worms Zone	12886.0	91	0.002302	1
ID	PUBG	25606.0	81	0.000590	0
VN	Arena of Valor	74978.0	121	0.000177	1

From the sample of final result above we can see that games such as Free Fire, The Last of Us, GTA V, MLBB, Worms Zone, and Arena of Valor were in cluster 1

Conclusion

Based on the final result of k-means three clusters were the optimum clusters that could be implemented on this research. That means one group of data can be stated significantly different than the other group of clusters.

The k-means algorithm has successfully separated the data into three clusters. Nevertheless further research to break down the feature importance from each cluster in the dataset after clustering is still needed to get the point of what cluster 0,1,2 really are.