Improving Wellness Programs through Clustering Analysis

Abstract
This paper will summarize my approach, findings, and recommendations for enhancing the wellness program based on the dataset of a wellness program of n=200. The goal is to use an unsupervised learning method, K-Means clustering, to uncover hidden patterns or intrinsic structures (Itauma, 2015). It will discuss how K-Means clustering can help healthcare organizations tailor interventions for different patient segments.

Introduction
As technology advances, the importance of machine learning in healthcare is increasingly recognized. Healthcare organizations and physicians utilize clustering results to analyze similarities with patients (Yang et al., 2024). By clustering patients in terms of risk factors, lifestyles, or other key factors, clustering results can help physicians gain insights into patients' needs and provide personalized treatments (Yang et al., 2024). This paper explores how our clustering analysis and results could help the wellness program identify and create programs to better aid and assist their patients.

Methodology
First, performing an EDA analysis and preprocessing the dataset is essential. Through research and EDA, the K-Means clustering method was applied to segment patients based on key health metrics. The features included: healthy meals, sleep hours, stress levels, BMI, and Exercise time in minutes. From personal experience, as a professional athlete, the EDA analysis and researching literature on health, I believe it was important to focus on features BMI, and Exercise time in minutes. The K-Means clustering revealed 6 patient clusters, supported by the WCSS (Figure 1) method and Silhouette score (Figure 2). Lastly, applying the principal component analysis (PCA) to see if dimensionality could be reduced to capture the most important variance and reduce the noise (Kaloyanova, 2020).

Figure 1 - *WCSS*

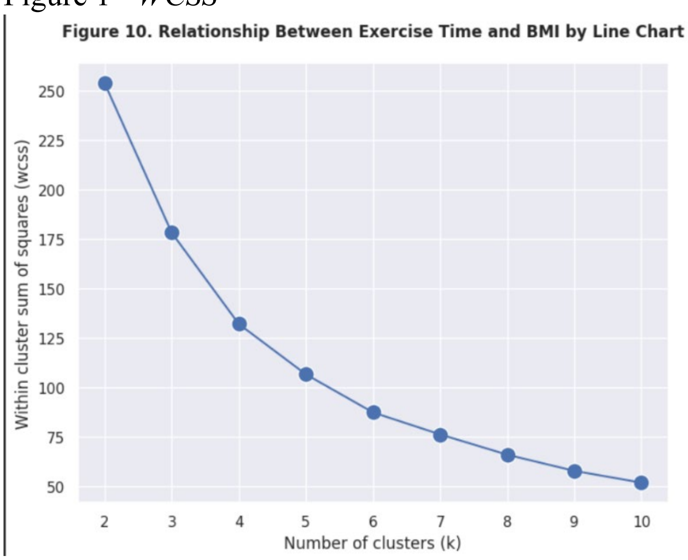
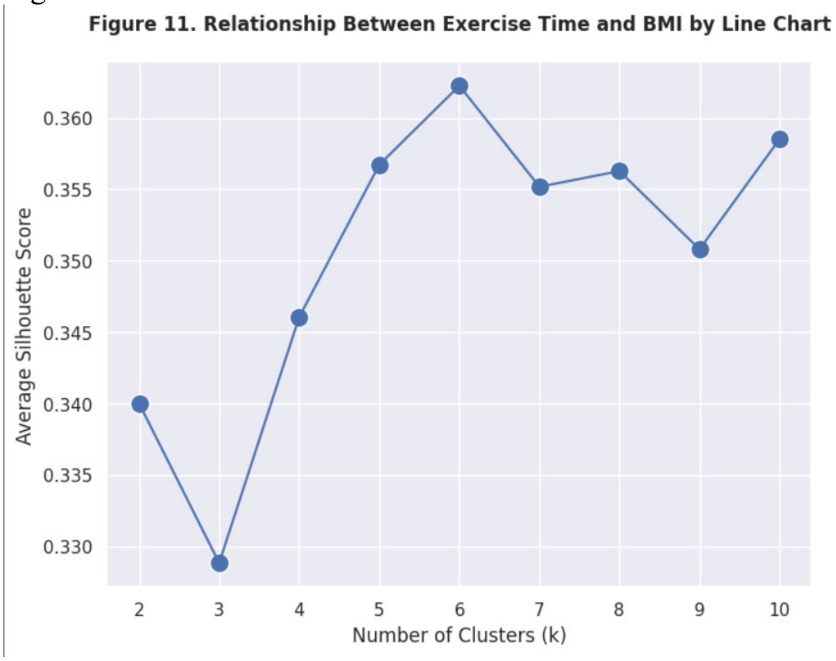


Figure 2 – *Silhouette Score*



Findings

In Figure 3, we observe that each cluster is differentiated yet no clear trend of the assumption that an increase in exercise time results in a lower BMI. However, in Figure 4, we can derive some insight. For instance, Cluster 0 exhibited high exercise minutes and moderate BMI, while Cluster 1 showed lower exercise minutes yet an optimal BMI.

Figure 3 – K-Means clustering (k=6)

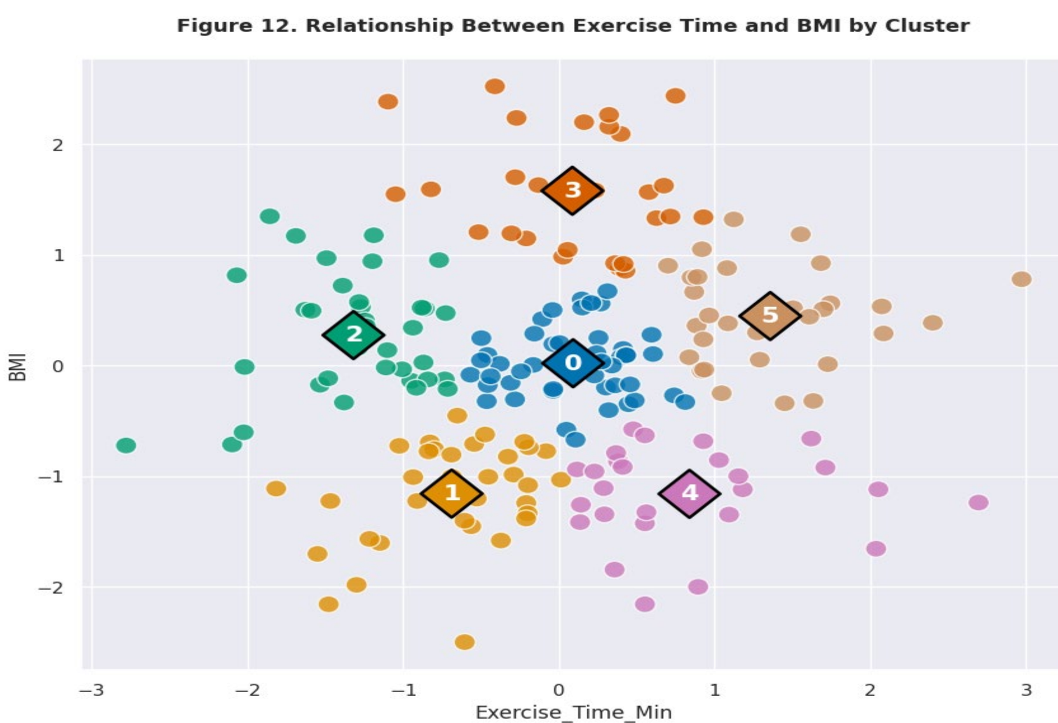


Figure 4 – *Table comparison of clusters and their features*

Cluster	Exercise_Time_Min	Healthy_Meals_Per_Day	Sleep_Hours_Per_Night	Stress_Level	BMI
0	30.46	3.24	6.55	4.65	25.25
1	23.21	2.50	7.05	6.00	19.29
2	17.36	2.74	7.06	4.71	26.53
3	30.41	2.62	7.34	4.62	33.13
4	37.39	3.00	7.09	5.27	19.28
5	42.21	3.03	6.71	4.83	27.40

Recommendations

Clusters 3 and 5 show that exercise alone does not guarantee a low BMI. This highlights the importance of healthy meal intake (cluster 3 has the lowest meal count despite 30.41 min/day exercise). The program should emphasize the importance of having a healthy diet, as exercise alone is not sufficient. Cluster 1's high stress (6.0) coexists with adequate sleep (7.05 hrs), which could suggest that stress may not disrupt sleep duration but could affect the quality of sleep. It would be beneficial for the program not just to track sleep hours but quality as well. Cluster 2 (lowest exercise time in minutes) should start with low-impact activities such as walking and gradually increase to 30 minutes per day.

Discussion/Conclusion

By leveraging data-driven insights and embracing emerging technologies, healthcare systems can navigate challenges effectively and drive meaningful change in healthcare delivery, ultimately leading to better outcomes for patients (Queen-Mary & Aderonke, 2024). In this lab report, we observed 6 patient clusters focusing on two features: BMI and Exercise time; however, it's important to note that by adding more features into our model, a clearer picture of relationships could improve our analysis and interpretation of the clusters. K-Means clustering is a great support decision-making tool and can aid in inferring problems, like the wellness program. I must remind myself that this is not a causational-based approach but a relationship-based one.

References

Itauma, I. (2015). *Machine Learning using Python - 4 Chapter 4: Unsupervised Learning - Clustering and Dimensionality Reduction*. Quarto.pub. https://amightyo.quarto.pub/machine-learning-using-python/Chapter_4.html

Kaloyanova, E. (2020, March 10). *How to Combine PCA and K-means Clustering in Python?* 365 Data Science. https://365datascience.com/tutorials/python-tutorials/pca-k-means/

Queen-Mary Akudo Ebugosi, & Aderonke, J. (2024). Optimizing healthcare resource allocation through data-driven demographic and psychographic analysis. *Computer Science & IT Research Journal*, *5*(6), 1488–1504. https://doi.org/10.51594/csitrj.v5i6.1249

Yang, W.-C., Lai, J.-P., Liu, Y.-H., Lin, Y.-L., Hou, H.-P., & Pai, P.-F. (2024). *Using Medical Data and Clustering Techniques for a Smart Healthcare System. Electronics*, *13*(1), 140. https://doi.org/10.3390/electronics13010140