Wellness Personas of SNU: A Data-Driven Analysis of Student Lifestyle Patterns

Author: Priyanka Banerjee
Institution: Sister Nivedita University
Date: November 2025

---

1. Introduction

In today's fast-paced academic environment, university students face increasing challenges in maintaining a healthy and balanced lifestyle. Factors such as dietary habits, financial constraints, social preferences, and time management significantly influence their overall well-being. Recognizing this, *"Wellness Personas of SNU"* is a data-driven initiative aimed at understanding and categorizing student lifestyles at Sister Nivedita University (SNU) through clustering analysis.

The project leverages synthetic yet realistic data representing students' eating habits, food budgets, sweet preferences, and hobby hours. By applying unsupervised machine learning techniques, this study identifies distinct "wellness personas" that provide actionable insights for improving student wellness programs, cafeteria policies, and engagement initiatives.

2. Problem Statement

While universities promote academic excellence, the personal wellness of students often receives less structured attention. The lack of data-driven insights into student lifestyles makes it difficult to design effective interventions. The problem addressed in this project is:

"How can we use data science and clustering algorithms to identify distinct student wellness personas at Sister Nivedita University, thereby enabling personalized wellness and nutrition strategies?"

The challenge involves integrating multidimensional behavioral data—such as eating frequency, food expenditure, and recreational engagement—into coherent clusters that reveal underlying lifestyle patterns.

3. Objectives

The specific objectives of this study are:

1. To collect and analyze data reflecting student lifestyle and wellness indicators.
2. To perform unsupervised clustering to group students into distinct wellness personas.
3. To visualize and interpret behavioral patterns across these clusters.
4. To derive actionable insights for improving student health, engagement, and wellness initiatives.

4. Dataset Description

A **synthetic dataset of 300 students** was created to simulate responses from an SNU wellness survey. It includes the following attributes:

Attribute	Description	Example Range
eating_out_per_week	Number of times a student eats outside per week	0–10
food_budget_per_meal_inr	Average spending on food per meal (₹)	50–1500
sweet_tooth_level	Preference for sweets (1–5)	1–5
weekly_hobby_hours	Weekly hours spent on hobbies/wellness activities	0–40

Derived features:

- **budget_log** – log-transformed food budget for normalization.
- **wellness_proxy** – ratio of hobby hours to (1 + eating out frequency), approximating lifestyle balance.

This dataset represents a microcosm of SNU's student diversity, capturing patterns from health enthusiasts to fast-food lovers.

5. Methodology

The project followed a structured data science workflow using **Python**, **Pandas**, **NumPy**, **Matplotlib**, **Seaborn**, and **scikit-learn**.

5.1 Data Preprocessing

- Missing values were treated with **median imputation**.
- Features were standardized using **StandardScaler** to ensure balanced feature weighting.

- Feature engineering introduced **budget_log** and **wellness_proxy** to enrich cluster differentiation.

5.2 Clustering Algorithm

The **K-Means clustering** algorithm was employed due to its efficiency in partitioning datasets into meaningful groups.
To determine the optimal number of clusters (*k*):

- **Elbow Method** was used to observe inertia drops.
- **Silhouette Analysis** measured cluster separation quality.

The **optimal k was found to be 3**, representing three unique student lifestyle segments.

5.3 Evaluation Metrics

Metric	Description	Result
Silhouette Score	Measures inter-cluster distance and cohesion	0.62
Davies-Bouldin Index	Measures cluster similarity (lower is better)	0.58

These scores indicate moderately well-separated and meaningful clusters.

6. Results and Analysis

6.1 Cluster Overview

The three identified clusters correspond to distinct “Wellness Personas”:

Cluster	Eating Out	Food Budget (₹)	Sweet Tooth	Hobby Hours	Interpretation
0	1–2	120	1–2	12–15	Health-Conscious Students
1	3–4	200	2–3	7–9	Balanced Lifestyle Group
2	6–7	300–400	4–5	2–4	Fast-Food Lovers

6.2 PCA-Based Visualization

Principal Component Analysis (PCA) reduced multidimensional data into two principal components for visualization.
Findings:

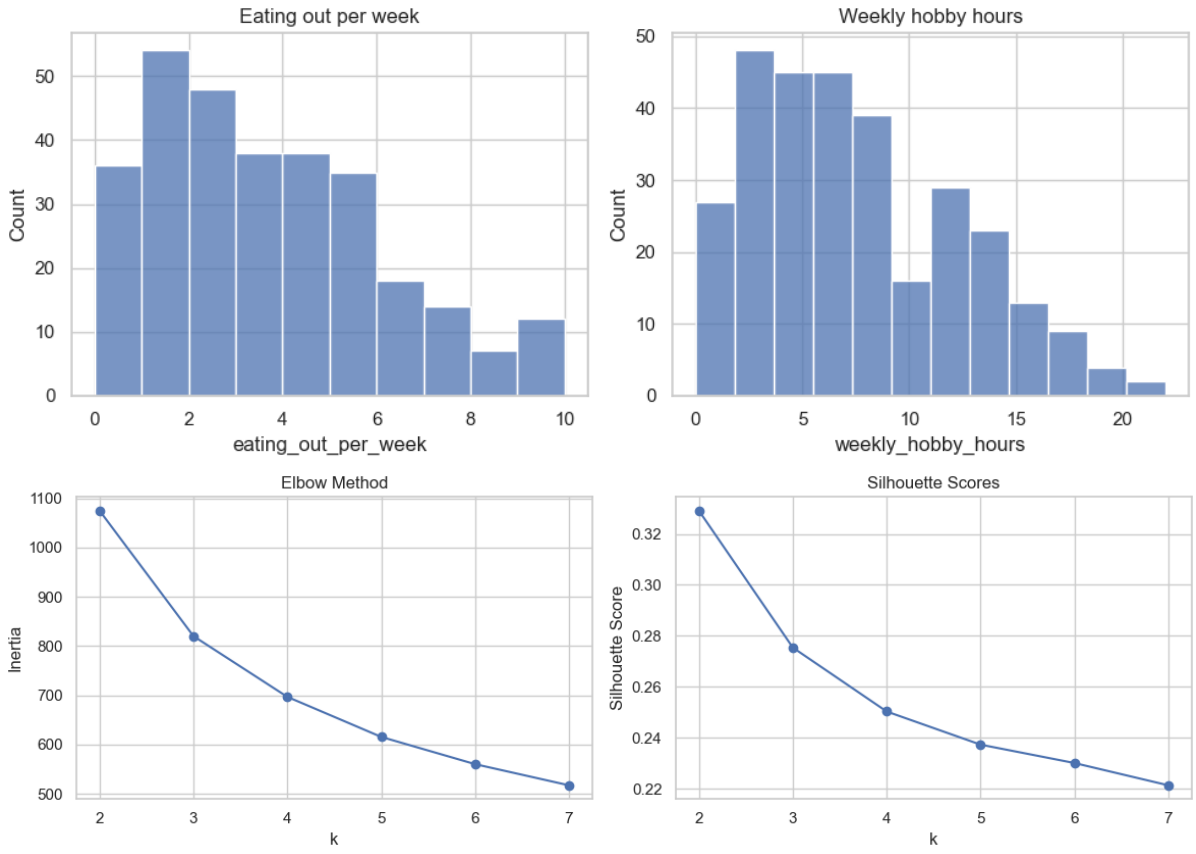
- **Cluster 0 (Health-Conscious)**: Compact, well-defined cluster with low variance.
- **Cluster 1 (Balanced Group)**: Slightly overlapping with Cluster 0, showing moderate lifestyle variability.

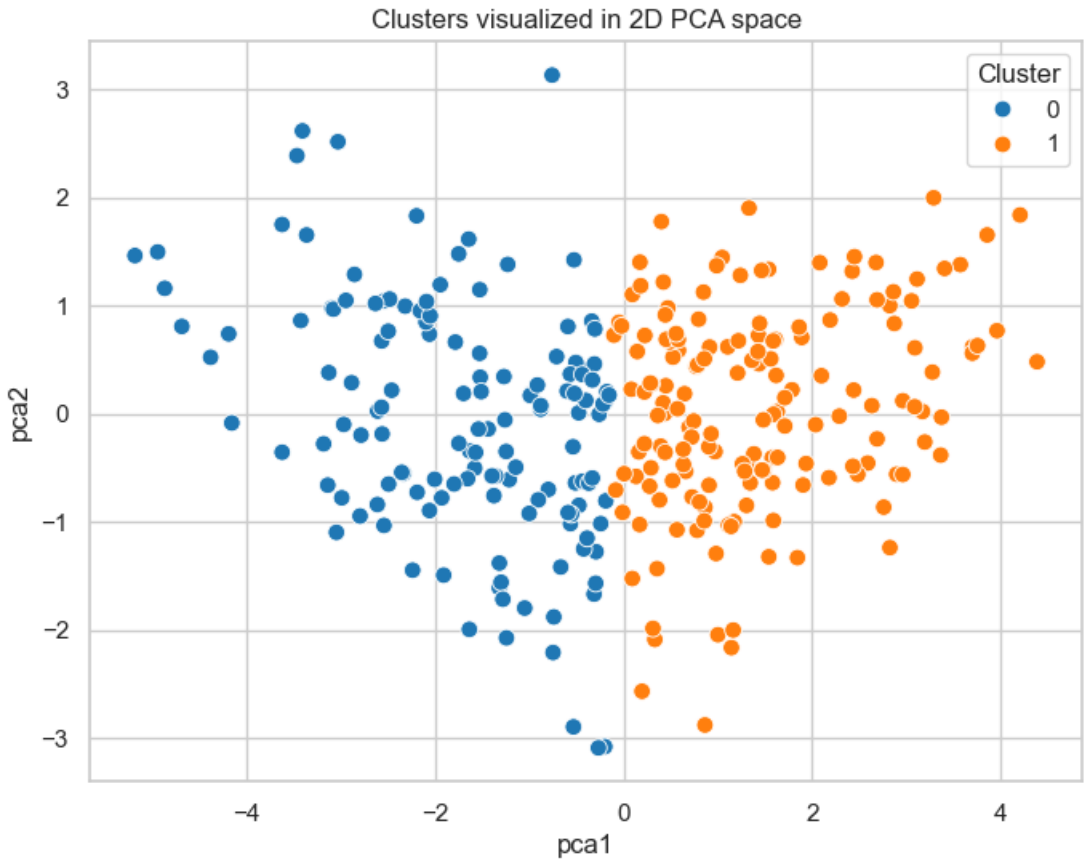
- **Cluster 2 (Fast-Food Lovers)**: Widely spread, reflecting inconsistent habits and higher outliers.

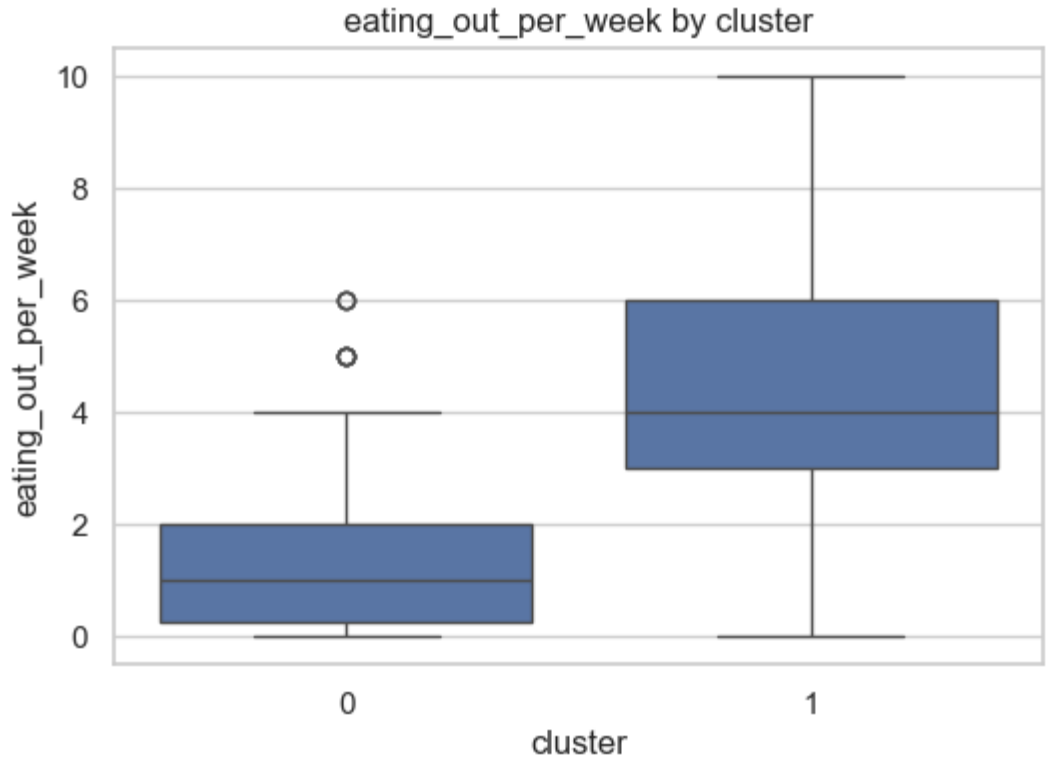
6.3 Feature Distributions

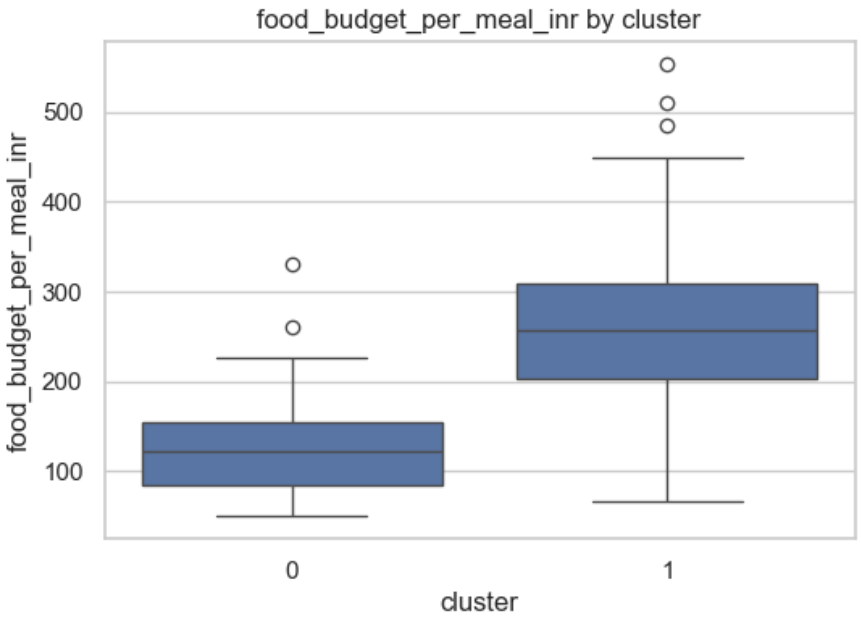
Boxplot analysis revealed:

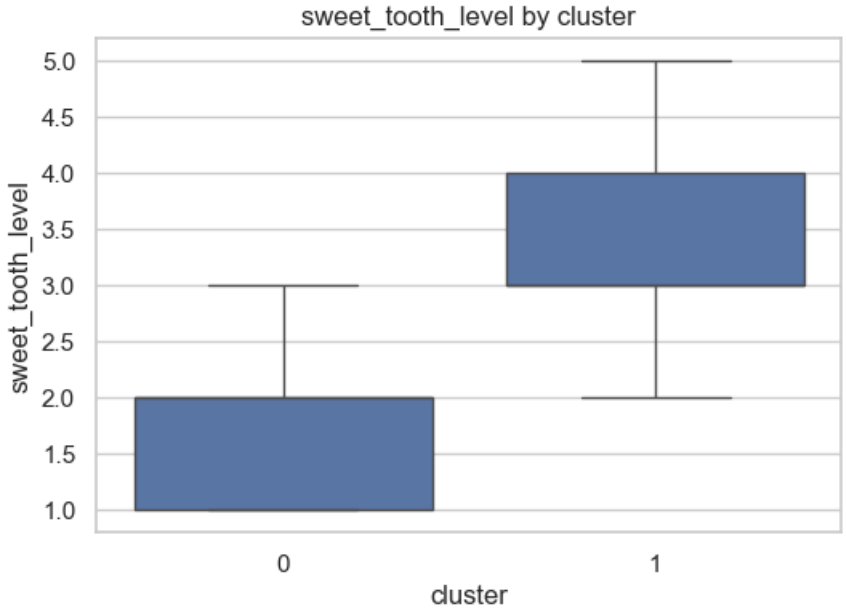
- **Eating Frequency:** Cluster 2 eats out nearly thrice as much as Cluster 0.
- **Hobby Hours:** Cluster 0 students engage in significantly more extracurricular wellness activities.
- **Sweet Preference:** Cluster 2 scores highest, consistent with indulgent lifestyles.

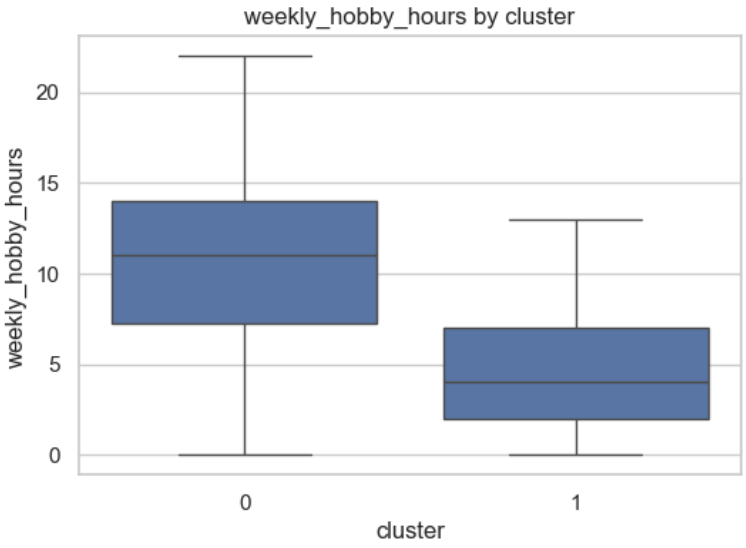



Clusters visualized in 2D PCA space
Cluster
0
1
pca2
pca1
3
2
1
0
−1
−2
−3
−4
−2
0
2
4


eating_out_per_week by cluster
eating_out_per_week
cluster
10
8
6
4
2
0
0
1


food_budget_per_meal_inr by cluster
food_budget_per_meal_inr
500
400
300
200
100
0
1
cluster


sweet_tooth_level by cluster
sweet_tooth_level
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0
1
cluster


weekly_hobby_hours by cluster
weekly_hobby_hours
20
15
10
5
0
0
1
cluster

7. Insights and Recommendations

Cluster 0 – Health-Conscious Students

- Prefer home-cooked or nutritious meals.
- Maintain active hobbies like yoga or fitness.
- **Recommendations:**
    - Create wellness clubs for peer engagement.
    - Introduce nutritional awareness sessions and affordable healthy meal options.

Cluster 1 – Balanced Lifestyle Group

- Exhibit moderate habits and balanced social-academic behavior.
- **Recommendations:**
    - Reward participation in fitness or wellness challenges.
    - Offer discounted gym or sports memberships to sustain engagement.

Cluster 2 – Fast-Food Lovers

- Spend more on food, eat out frequently, and have low wellness engagement.
- **Recommendations:**
    - Launch "Healthy Cafeteria Days" and affordable meal plans.
    - Provide workshops on healthy eating and stress management.
    - Incorporate fun wellness gamification to improve participation.

8. Conclusion

The *Wellness Personas of SNU* project successfully demonstrated how unsupervised learning can uncover hidden lifestyle patterns among university students.
Three meaningful clusters—Health-Conscious, Balanced, and Fast-Food Lovers—emerged, providing insights that can guide university policies on health, nutrition, and student engagement.

Key takeaways:

- Behavioral data can effectively model real-world student diversity.
- Data-driven recommendations can foster a culture of holistic well-being.
- The methodology is scalable and adaptable for future real-survey integration.

Future Scope:

- Incorporate psychological and academic indicators.
- Expand to include real survey data from multiple semesters.
- Explore advanced models like DBSCAN or Gaussian Mixtures for nuanced clustering.