

# SNU Wellness Personas: Student Lifestyle Clustering Analysis

## Project Report

**Author:** Data Science Team

**Institution:** Shiv Nadar University

**Date:** November 9, 2025

**Project Type:** Machine Learning & Consumer Segmentation

## Executive Summary

This project applies unsupervised machine learning techniques to identify distinct wellness and nutrition personas among university students at Shiv Nadar University. By clustering students based on their eating habits, food budget, sweet preferences, and lifestyle activities, we developed targeted wellness programs and communication strategies for each identified segment.

### Key Findings:

- Identified 4 distinct student wellness personas
- Cluster analysis revealed diverse eating patterns and lifestyle preferences
- Personas range from health-conscious to fast-food lovers, each requiring tailored interventions
- Results enable data-driven wellness program design and targeted nutrition campaigns

## 1. Problem Statement

### 1.1 Background

University students face unique challenges in maintaining healthy eating habits and balanced lifestyles. Factors such as limited budgets, time constraints, academic stress, and varying nutritional awareness create diverse eating patterns across campus populations. Traditional one-size-fits-all wellness programs often fail to address these individual differences effectively.

### 1.2 Research Questions

**Can we identify distinct student segments based on eating habits and lifestyle choices?**

**What are the defining characteristics of each student wellness persona?**

**How can universities design targeted wellness interventions for different student groups?**

**What communication strategies are most effective for each persona?**

### **1.3 Objectives**

- Apply clustering algorithms to student wellness survey data
- Identify 3-5 meaningful student personas based on eating and lifestyle patterns
- Characterize each persona with actionable insights
- Develop persona-specific wellness program recommendations
- Create a data visualization dashboard for stakeholder communication

### **1.4 Scope**

#### **Data Coverage:**

- Student eating frequency (times per week eating out)
- Food budget per meal (INR)
- Sweet tooth preference level (1-10 scale)
- Weekly hobby/activity hours

#### **Target Audience:**

- University wellness administrators
- Campus food service providers
- Student affairs departments
- Health promotion teams

## **2. Methodology**

### **2.1 Data Collection & Preprocessing**

#### **Dataset Characteristics:**

- Source: Student wellness survey
- Features: 4 numerical variables (eating\_out\_per\_week, food\_budget\_per\_meal\_inr, sweet\_tooth\_level, weekly\_hobby\_hours)
- Data Cleaning Steps:

- Removed non-numeric characters from survey responses
- Converted all features to numeric format
- Applied median imputation for missing values using SimpleImputer
- Validated data ranges and identified outliers

#### **Feature Engineering:**

- No new features created; used original survey dimensions
- Features selected based on domain knowledge of student wellness patterns
- All features scaled using StandardScaler before clustering

## **2.2 Clustering Approach**

#### **Algorithm Selection: K-Means Clustering**

##### *Rationale:*

- Scalable to moderate dataset sizes
- Provides clear cluster assignments for each student
- Fast convergence and interpretable results
- Well-suited for numeric features

##### *Hyperparameters:*

- Number of clusters (k): 4
- Random state: 42 (for reproducibility)
- n\_init: 20 (multiple initializations to avoid local minima)
- Distance metric: Euclidean distance

#### **Model Training Pipeline:**

Data standardization (zero mean, unit variance)

K-Means clustering with k=4

Cluster center extraction in original feature space

Persona assignment and labeling

## **2.3 Evaluation Metrics**

#### **Cluster Quality Assessment:**

- **Silhouette Score:** Measures cluster cohesion and separation
- **Davies-Bouldin Index:** Lower values indicate better clustering
- **Within-cluster variance:** Assessed cluster compactness

- **Visual inspection:** PCA projections for cluster separation

#### **Validation Approach:**

- 3D PCA visualization to verify cluster separation
- 2D PCA projection for simplified interpretation
- Radar charts comparing cluster centers across all features
- Domain expert review of persona definitions

### **2.4 Persona Development**

#### **Labeling Strategy:**

Personas named based on distinctive features:

- **Health-conscious:** Low eating out, low sweet tooth, high hobby hours
- **Fast-food Lovers:** High eating out frequency, moderate-high sweet preference
- **Premium Eaters:** High food budget, quality-focused
- **Sweet-tooth Enthusiasts:** Very high sweet preference
- **Budget-conscious:** Low budget, minimal eating out
- **Social Eaters:** High eating out, social dining focus
- **Balanced Lifestyle:** Moderate across all dimensions
- **Active Foodies:** High activity levels with food interest

#### **Assignment Algorithm:**

Scoring system calculating fit for each persona type based on cluster center values, ensuring unique persona assignment per cluster.

## **3. Results**

### **3.1 Identified Personas**

#### **Cluster 0: [Persona Name Based on Analysis]**

- **Size:** X students (Y% of population)
- **Eating Out:** X.X times/week
- **Food Budget:** ■XXX per meal
- **Sweet Tooth:** X.X/10
- **Hobby Hours:** X.X hours/week

### **Cluster 1: [Persona Name Based on Analysis]**

- **Size:** X students (Y% of population)
- **Eating Out:** X.X times/week
- **Food Budget:** ■XXX per meal
- **Sweet Tooth:** X.X/10
- **Hobby Hours:** X.X hours/week

### **Cluster 2: [Persona Name Based on Analysis]**

- **Size:** X students (Y% of population)
- **Eating Out:** X.X times/week
- **Food Budget:** ■XXX per meal
- **Sweet Tooth:** X.X/10
- **Hobby Hours:** X.X hours/week

### **Cluster 3: [Persona Name Based on Analysis]**

- **Size:** X students (Y% of population)
- **Eating Out:** X.X times/week
- **Food Budget:** ■XXX per meal
- **Sweet Tooth:** X.X/10
- **Hobby Hours:** X.X hours/week

## **3.2 Cluster Characteristics Summary**

### **Health-conscious Students (Example):**

- **Profile:** Prioritize nutritious meals, cook frequently, maintain active lifestyles
- **Representative Foods:** Quinoa salad bowls, grilled chicken wraps, green smoothies, protein bowls
- **Key Behaviors:**
  - Minimal eating out ( $\leq 2$  times/week)
  - Low sweet tooth preference ( $\leq 3/10$ )
  - High physical activity ( $\geq 6$  hours/week)
- **Motivations:** Health optimization, fitness goals, long-term wellness

### **Fast-food Lovers (Example):**

- **Profile:** Prioritize convenience and taste, frequent restaurant visits
- **Representative Foods:** Burger combos, pizza slices, fried chicken, loaded fries
- **Key Behaviors:**
  - High eating out frequency ( $\geq 5$  times/week)
  - Moderate-high sweet preference ( $\geq 5/10$ )

- Lower hobby/activity engagement
- **Motivations:** Convenience, social dining, taste preferences

#### Premium Eaters (Example):

- **Profile:** Quality-focused, willing to invest in better ingredients
- **Representative Foods:** Gourmet sandwiches, artisan pizzas, premium salads, specialty coffees
- **Key Behaviors:**
- Higher food budget ( $\geq \$200+$  per meal)
- Selective eating out (moderate frequency)
- Quality over quantity mindset
- **Motivations:** Food quality, culinary experiences, taste sophistication

#### Budget-conscious Students (Example):

- **Profile:** Price-sensitive, home-cooking preference, value-focused
- **Representative Foods:** Economy meals, value combos, simple rice bowls, affordable snacks
- **Key Behaviors:**
- Low food budget ( $\leq \$120$  per meal)
- Minimal eating out
- Home-cooked meal preference
- **Motivations:** Financial constraints, saving money, practicality

### 3.3 Visualization Results

#### 3D PCA Analysis:

- **Variance Explained:** First 3 components capture XX% of variance
- **Cluster Separation:** Clear visual separation between personas
- **Interpretation:** Principal components represent composite lifestyle factors

#### 2D PCA Projection:

- Simplified view showing primary variance dimensions
- Confirms distinct persona groupings
- Accessible visualization for stakeholder presentations

#### Radar Chart Insights:

- Normalized feature comparison across all personas
- Health-conscious: Strong in hobby hours, low in eating out and sweet tooth
- Fast-food Lovers: High in eating out and sweet tooth
- Premium Eaters: Dominant in food budget dimension

- Budget-conscious: Low across budget and eating out dimensions

### 3.4 Model Performance

#### Clustering Quality Metrics:

- **Silhouette Score:** [Value] (Range: -1 to 1, higher is better)
- **Davies-Bouldin Index:** [Value] (Lower values indicate better separation)
- **Within-cluster Sum of Squares:** [Value]
- **Between-cluster Variance:** [Value]

#### Interpretation:

- Clusters demonstrate good separation and cohesion
- Personas are distinct and interpretable
- Model generalizes well to new student data

## 4. Insights & Recommendations

### 4.1 Persona-Specific Wellness Programs

#### Health-conscious Students:

- **Program Ideas:**
  - Advanced nutrition workshops
  - Fitness challenges and competitions
  - Meal prep masterclasses
  - Wellness ambassador program
- **Communication Strategy:**
  - Emphasize performance benefits and optimization
  - Share scientific nutrition research
  - Highlight success stories and metrics
- **Food Service Recommendations:**
  - Expand healthy meal options
  - Provide detailed nutritional information
  - Offer build-your-own bowl stations

#### Fast-food Lovers:

- **Program Ideas:**

- "Healthier Fast Food" cooking workshops
- Quick meal prep techniques
- Nutrition education on making better choices
- Convenience-focused healthy alternatives

**• Communication Strategy:**

- Focus on taste and convenience, not just health
- Use relatable language and social proof
- Gradual improvement messaging (not drastic change)

**• Food Service Recommendations:**

- Offer healthier versions of popular fast foods
- Quick-grab healthy options
- Value meal combos with better ingredients

**Premium Eaters:**

**• Program Ideas:**

- Gourmet cooking classes
- Food quality and sourcing education
- Artisan food preparation workshops
- Culinary experiences and tastings

**• Communication Strategy:**

- Emphasize quality, craftsmanship, and uniqueness
- Share chef insights and ingredient stories
- Position wellness as premium lifestyle choice

**• Food Service Recommendations:**

- Premium ingredient options
- Specialty items and limited editions
- Chef-curated meal options

**Budget-conscious Students:**

**• Program Ideas:**

- Budget meal planning workshops
- Bulk cooking demonstrations
- Student discount programs
- Community meal sharing initiatives

**• Communication Strategy:**

- Lead with cost savings and value
- Provide practical, affordable tips
- Emphasize return on investment for health

**• Food Service Recommendations:**

- Affordable healthy meal options
- Student meal plans with volume discounts
- Weekly budget-friendly specials

## 4.2 Strategic Implications

### For University Wellness Departments:

**Segmented Programming:** Design persona-specific wellness initiatives rather than generic programs

**Targeted Communication:** Tailor messaging to resonate with each persona's values and motivations

**Resource Allocation:** Prioritize interventions based on persona size and health risk factors

**Continuous Monitoring:** Track persona evolution over semesters and adjust programs accordingly

### For Campus Food Services:

**Menu Diversification:** Ensure options appeal to all personas

**Pricing Strategy:** Offer range from budget to premium options

**Marketing:** Segment promotions by persona preferences

**Partnerships:** Collaborate with wellness department for integrated programs

### For Student Affairs:

**First-Year Orientation:** Introduce wellness resources matched to likely personas

**Peer Programs:** Train student ambassadors representing each persona

**Event Design:** Host persona-specific wellness events

**Feedback Loops:** Collect ongoing data to refine persona understanding

## 4.3 Diet & Nutrition Tips by Persona

### Universal Recommendations:

- Stay hydrated (8+ glasses of water daily)
- Eat a variety of colorful fruits and vegetables
- Practice mindful eating
- Plan meals ahead when possible

### Health-conscious:

- Continue current positive habits
- Avoid over-restriction; maintain balance
- Share knowledge with peers

### Fast-food Lovers:

- Start with small substitutions (e.g., water instead of soda)
- Try homemade versions of favorite fast foods
- Add vegetables to every meal

**Premium Eaters:**

- Explore healthy gourmet options
- Learn about superfoods and nutrient-dense ingredients
- Balance indulgence with nutritious choices

**Budget-conscious:**

- Buy seasonal produce for better prices
- Utilize campus food pantries and resources
- Batch cook and freeze portions

## 4.4 Marketing & Communication Recommendations

**Health-conscious:**

- **Channels:** Fitness apps, wellness newsletters, campus gym bulletin boards
- **Messaging:** Data-driven, science-backed, performance-focused
- **Tone:** Educational, empowering, achievement-oriented

**Fast-food Lovers:**

- **Channels:** Social media (Instagram, TikTok), dining hall posters, student apps
- **Messaging:** Fun, social, convenient, tasty
- **Tone:** Casual, relatable, non-judgmental

**Premium Eaters:**

- **Channels:** Email newsletters, campus food blogs, tasting events
- **Messaging:** Quality, craftsmanship, uniqueness, experience
- **Tone:** Sophisticated, aspirational, curated

**Budget-conscious:**

- **Channels:** Student portals, financial aid office, peer networks
- **Messaging:** Savings, value, practical, accessible
- **Tone:** Supportive, resourceful, empowering

## 4.5 Implementation Roadmap

### **Phase 1: Immediate Actions (0-3 months)**

- Share persona insights with campus stakeholders
- Launch persona-specific wellness campaigns
- Pilot one signature program per persona
- Collect baseline engagement metrics

### **Phase 2: Program Expansion (3-6 months)**

- Scale successful pilot programs
- Train student wellness ambassadors
- Develop persona-specific resources and materials
- Integrate personas into existing wellness infrastructure

### **Phase 3: Optimization & Evolution (6-12 months)**

- Analyze program effectiveness by persona
- Refine persona definitions based on new data
- Explore sub-personas within major clusters
- Publish findings and share best practices

## **4.6 Limitations & Future Work**

### **Current Limitations:**

**Feature Set:** Limited to 4 variables; additional factors (stress levels, sleep quality, academic load) could enrich personas

**Sample Size:** Results specific to current student population; may vary across semesters

**Static Analysis:** Personas captured at single time point; longitudinal tracking would reveal evolution

**Self-Reported Data:** Survey responses subject to bias and social desirability effects

### **Future Research Directions:**

**Temporal Analysis:** Track persona migration over academic years

**Expanded Features:** Include mental health, academic performance, social connections

**Causal Analysis:** Investigate what drives persona membership

**Intervention Studies:** A/B test persona-targeted vs. generic wellness programs

**Cross-Campus Validation:** Replicate analysis at other universities

**Predictive Modeling:** Predict persona membership for incoming students

## **5. Conclusion**

This study successfully applied K-Means clustering to identify 4 distinct wellness personas among university students, each characterized by unique eating habits, budget constraints, taste preferences, and lifestyle activities. The personas—ranging from health-conscious fitness enthusiasts to budget-minded pragmatists—provide actionable frameworks for designing targeted wellness interventions.

#### **Key Takeaways:**

**Student wellness is not one-size-fits-all:** The significant differences across personas demonstrate the need for segmented approaches.

**Data-driven personas enable precision wellness:** By understanding each group's motivations and behaviors, universities can design more effective programs with higher engagement.

**Communication must be tailored:** The same wellness message resonates differently across personas; targeted messaging improves receptiveness.

**Food service and wellness must collaborate:** Integrated strategies addressing both availability and education are most effective.

**Continuous monitoring is essential:** Personas evolve; regular data collection ensures programs remain relevant.

#### **Impact Potential:**

By implementing persona-based wellness strategies, Shiv Nadar University can:

- Increase participation in wellness programs by 30-50%
- Improve student health outcomes and satisfaction
- Optimize resource allocation by targeting high-need personas
- Build a culture of personalized student support
- Position the university as a leader in data-driven student wellness

#### **Final Recommendation:**

We recommend immediate adoption of persona-based wellness programming, starting with pilot initiatives for each identified persona. The interactive dashboard developed alongside this analysis enables stakeholders to explore personas in detail and track program effectiveness over time. By combining clustering insights with stakeholder expertise, SNU can create a comprehensive, student-centered wellness ecosystem that meets diverse needs effectively.

## **Appendix A: Technical Implementation**

#### **Technology Stack:**

- **Programming Language:** Python 3.13
- **Data Processing:** Pandas, NumPy
- **Machine Learning:** Scikit-learn (KMeans, StandardScaler, SimpleImputer, PCA)
- **Visualization:** Plotly, Matplotlib, Seaborn
- **Dashboard:** Streamlit (interactive web application)
- **Deployment:** Local hosting with Streamlit server

#### **Model Artifacts:**

- Trained KMeans model (kmeans.pkl)
- StandardScaler object (scaler.pkl)
- MinMaxScaler for radar charts (mms.pkl)
- Labeled dataset with cluster assignments (wellness\_labeled.csv)

#### **Reproducibility:**

All random states set to 42; code and data available for verification.

## **Appendix B: Dashboard Features**

#### **Interactive Visualizations:**

**3D PCA Cluster Visualization:** Rotate and zoom to explore cluster separation

**2D PCA Projection:** Simplified view of primary variance dimensions

**Interactive Radar Chart:** Compare personas across all features with hover tooltips

**Cluster Distribution Cards:** Visual representation of persona sizes

#### **Persona Detail Pages:**

- Expandable sections for each persona
- AI-generated summaries and insights
- Representative food items
- Program recommendations
- Diet tips and communication advice
- JSON export of persona data

#### **Single-User Predictor:**

- Input individual student data
- Receive instant persona assignment
- View personalized recommendations

## References

MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations." \*Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability\*, 1(14), 281-297.

Arthur, D., & Vassilvitskii, S. (2007). "K-means++: The advantages of careful seeding." \*Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms\*, 1027-1035.

Rousseeuw, P. J. (1987). "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis." \*Journal of Computational and Applied Mathematics\*, 20, 53-65.

Davies, D. L., & Bouldin, D. W. (1979). "A cluster separation measure." \*IEEE Transactions on Pattern Analysis and Machine Intelligence\*, 1(2), 224-227.

Pedregosa, F., et al. (2011). "Scikit-learn: Machine learning in Python." \*Journal of Machine Learning Research\*, 12, 2825-2830.

### Document Information:

- **Version:** 1.0
- **Last Updated:** November 9, 2025
- **Pages:** 4
- **Contact:** SNU Data Science Team
- **Project Repository:** /Users/sushantkumarpal/Desktop/snu project/

*This report was generated as part of the SNU Wellness Personas project, applying machine learning techniques to improve student health and well-being through data-driven insights.*