



# Spotify

## **From Free to Premium:** **Understanding Spotify's User** **Segments**

Dream Stream Team – Team 18B

Shivani Vallandas, Christie Shin, Gema Zhu, Vishal Srivastava, Shuai Zhao

BANA 205: Foundations of Marketing

# | The Problem

How can Spotify predict churn risk using cluster-level behavioral profiles, and which clusters should be prioritized for retention and conversion?



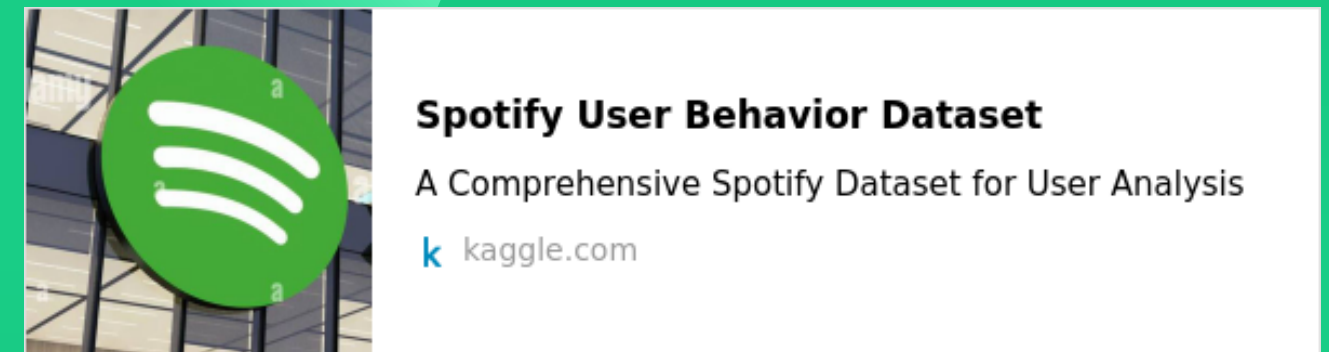
# | Our Data

We used a dataset from Kaggle titled  
**“Spotify User Behavior Dataset”**

A few important questions that the dataset includes is:

- Are you willing to take a premium subscription or willing to continue with premium subscription in future?
- What is your favorite time slot to listen to music?
- When do you listen to music more often?

The dataset includes 520 unique rows and 20 columns





# | Data Pipeline



# | Data Transformation & Segmentation

## Raw Data

User | Sub | Mood  
U1 | Prem | Relax, Sad  
U2 | Free | Party

Explode  
→

## Processed Data

User | Sub | Mood  
U1 | Prem | Relax  
U1 | Prem | Sad  
U2 | Free | Party

## Free Users

C0: The Almost-Premiums  
C1: The Free-Tier Fanatics  
C2: The Casual Dabblers

Goal: **Conversion**

Split by Plan  
↙ ↘

## Premium Users

C0: The Wanderers  
C1: The Nighttime Loyalist  
C2: The Everyday Groovers

Goal: **Retention**

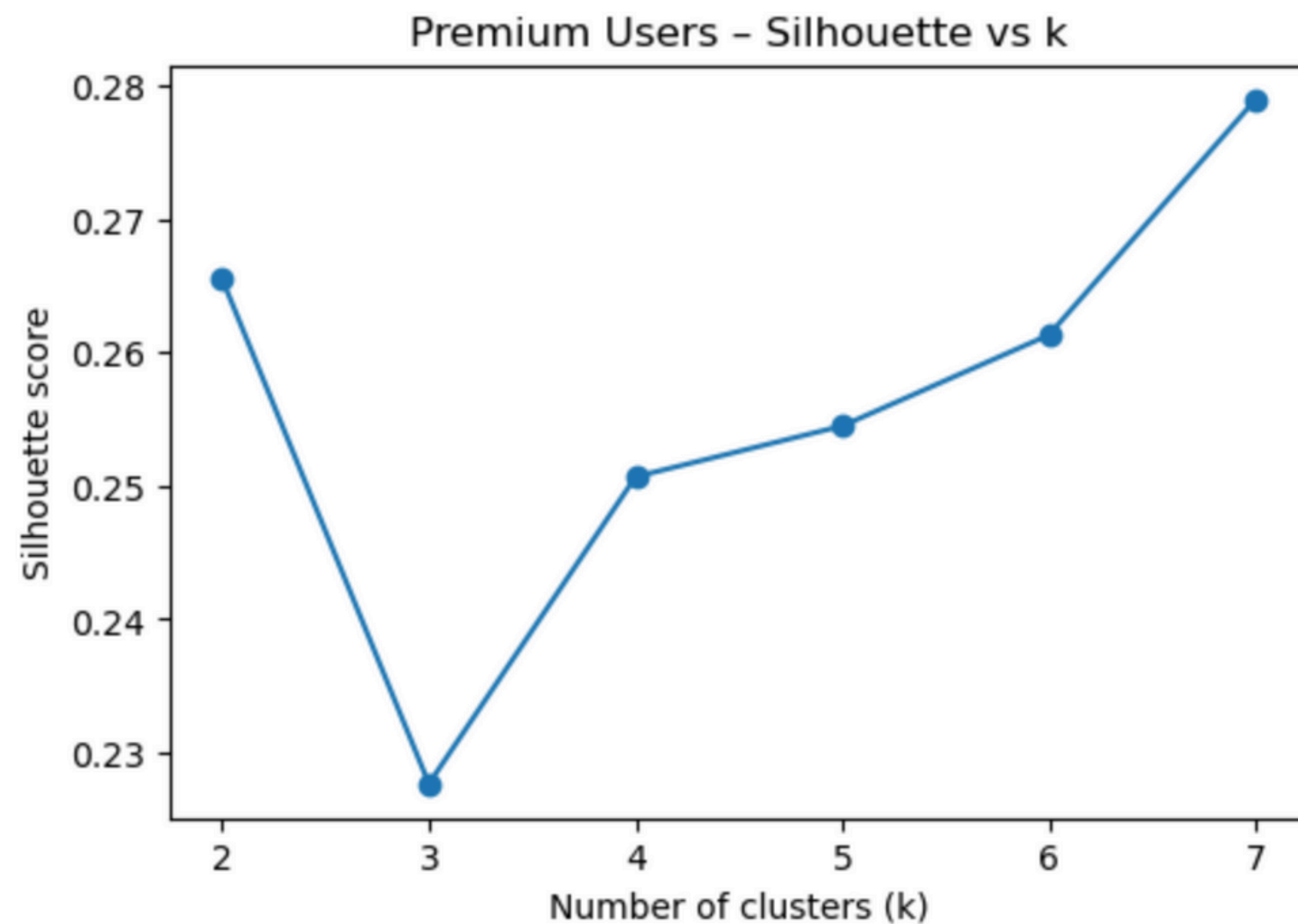
# | K-means Clustering

**We separated the data into 2 groups: Premium Users (357 rows) & Free Users (1359 rows)**

- Each group had 8 variables: 1 target variable and 7 independent variables
- We applied k-means clustering (silhouette score) by dividing each set of variables into subsets and assigning them to numbers
- K=3 produced the strongest balance of cohesion and separation for both user groups
- We calculated cluster-level averages for all variables and visualized them into a heat map to profile each segment
- We also created a scatterplot to visually display the spatial separation and distribution of clusters

# Insights

## Premium users – Silhouette line graph



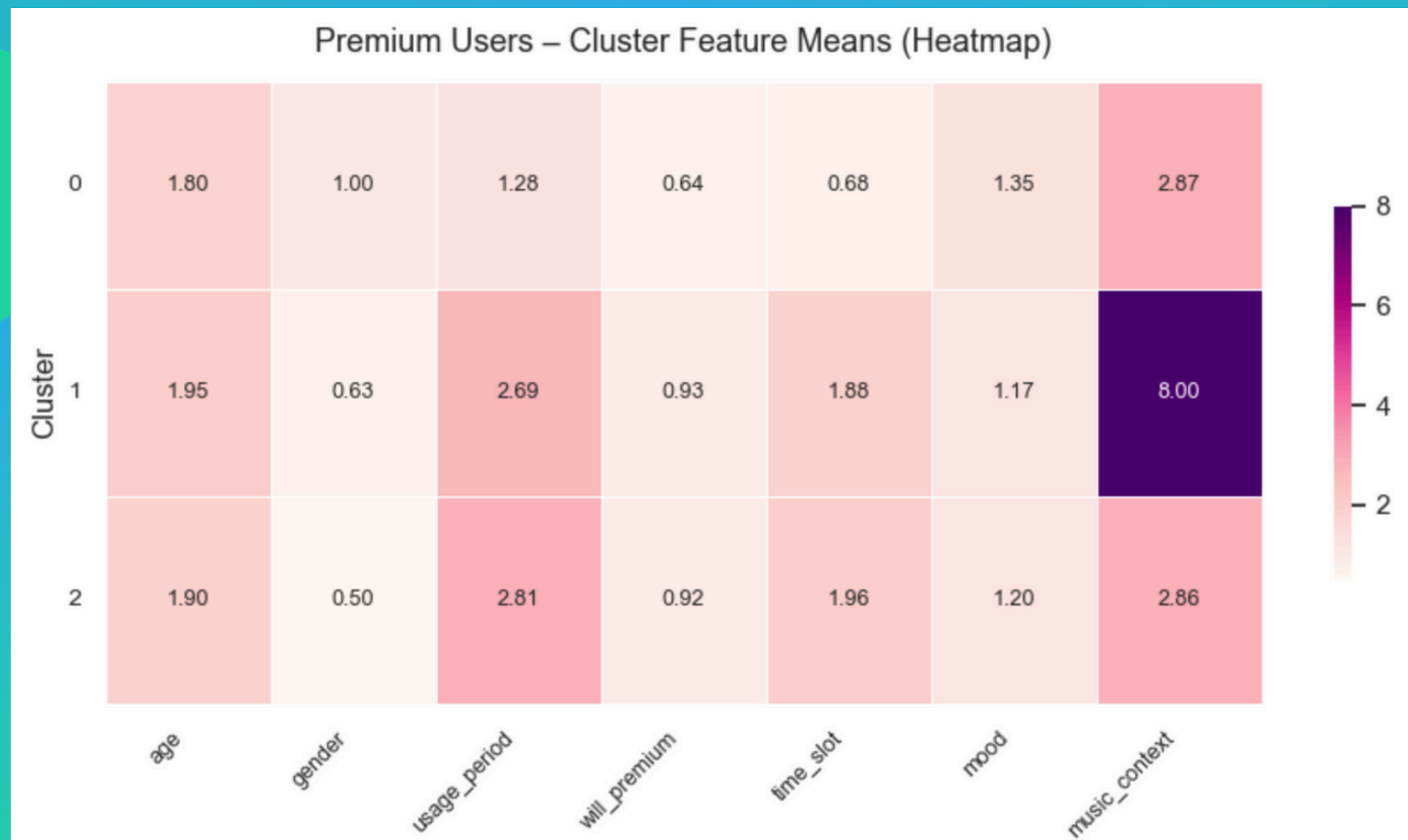
We evaluated multiple k values using the silhouette score to determine optimal cluster quality.

K=3 produced the strongest balance of cohesion and separation for both user groups



# Insights

## Premium users – Heatmap



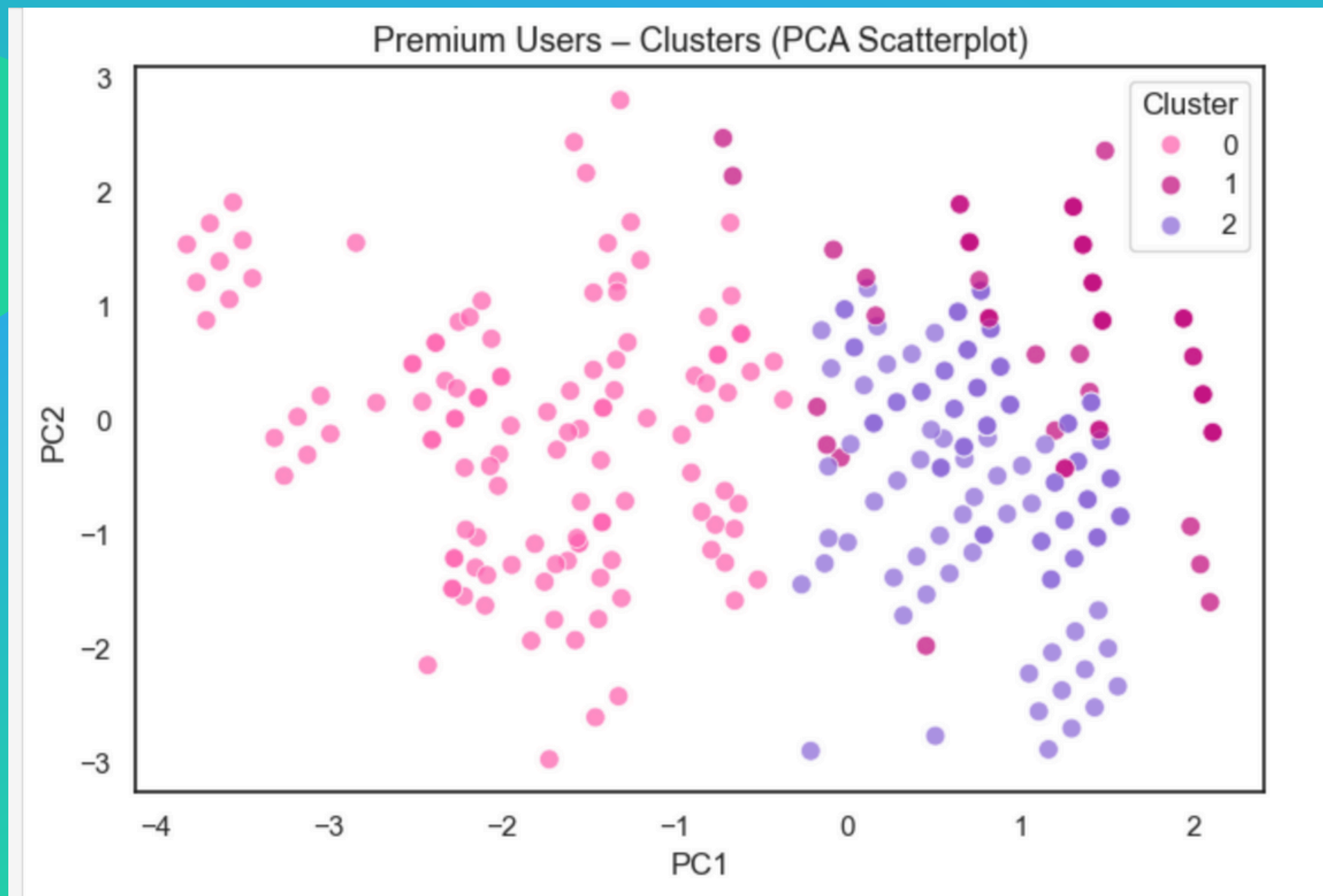
Premium willingness is highest in Cluster 1 (0.93) and Cluster 2 (0.92), compared with 0.64 in Cluster 0.

Cluster 1 also shows an extreme music-context value of 8.00, and Cluster 2 has the longest usage period at 2.81. These numbers highlight the key differences across the Premium clusters.



# Insights

## Premium users – Scatter Plot



**PC1** captures overall engagement level, separating heavy, long-term users from lighter, casual listeners.

**PC2** reflects differences in listening style, such as time-of-day preferences, mood, or usage context.

Three visually distinct user groups:

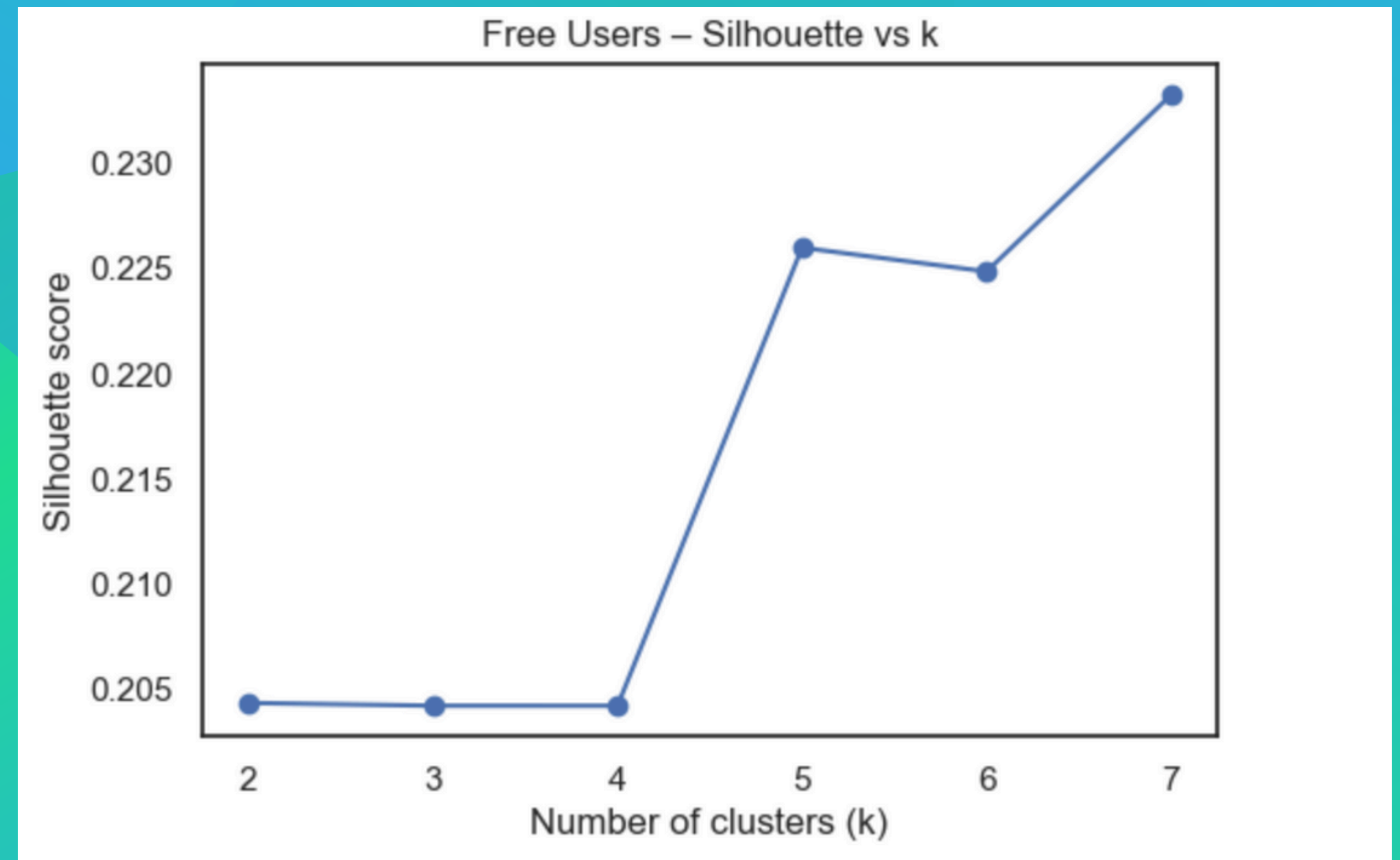
- 1) low-engagement
- 2) context-driven
- 3) routine nightly listener segments

# | Insights

## Free users – Silhouette line graph

We evaluated multiple  $k$  values using the silhouette score to determine optimal cluster quality.

$K=3$  produced the strongest balance of cohesion and separation for both user groups.



# | Insights

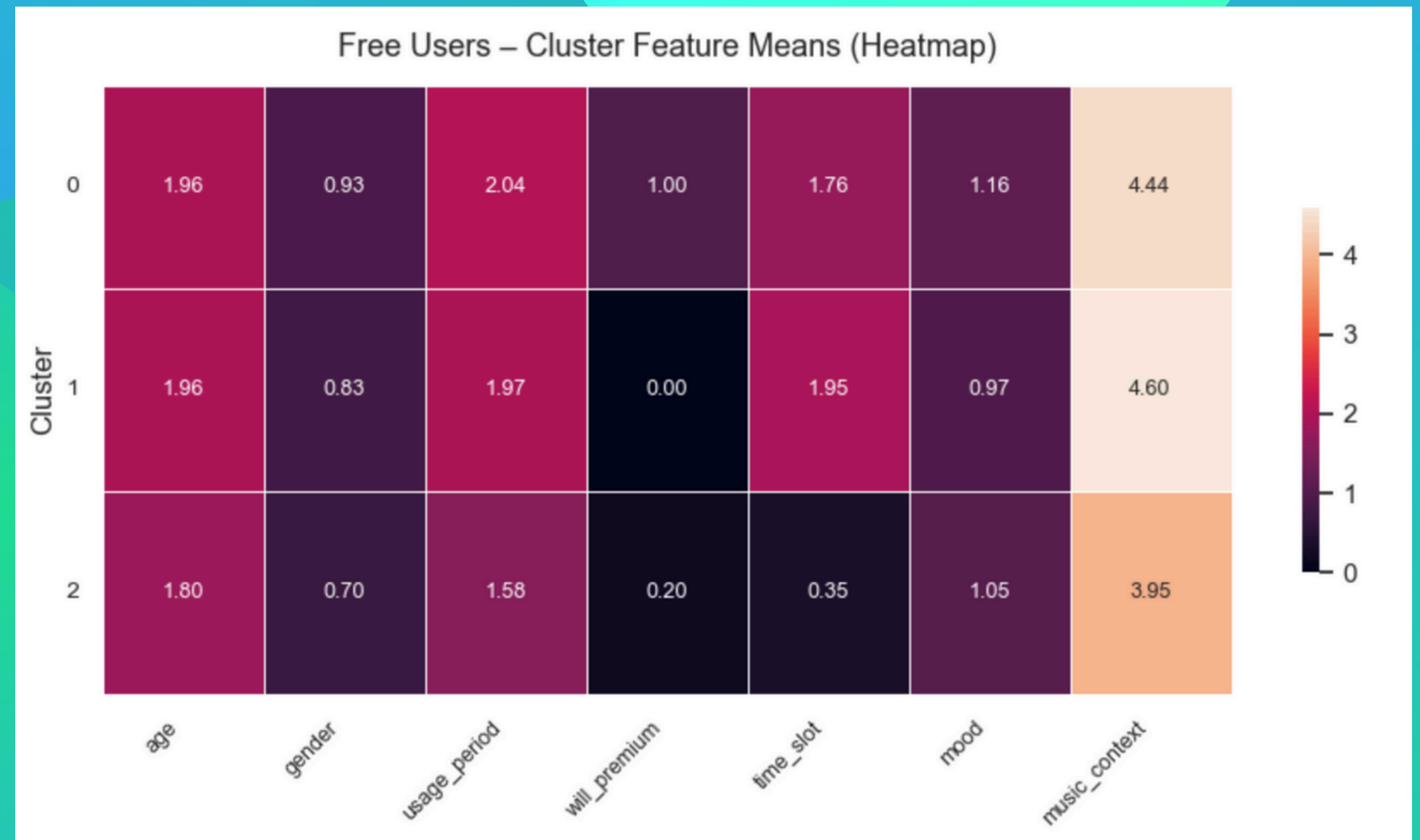
## Free users – Heatmap

Free users show clear differences across clusters.

Cluster 0 has the highest music-context score (4.44), meaning they listen in more varied situations.

Cluster 1 has the lowest premium willingness (0.00), showing no interest in upgrading.

Cluster 2 shows low engagement overall, with lower usage period (1.58) and time-slot activity (0.35).





# | Insights

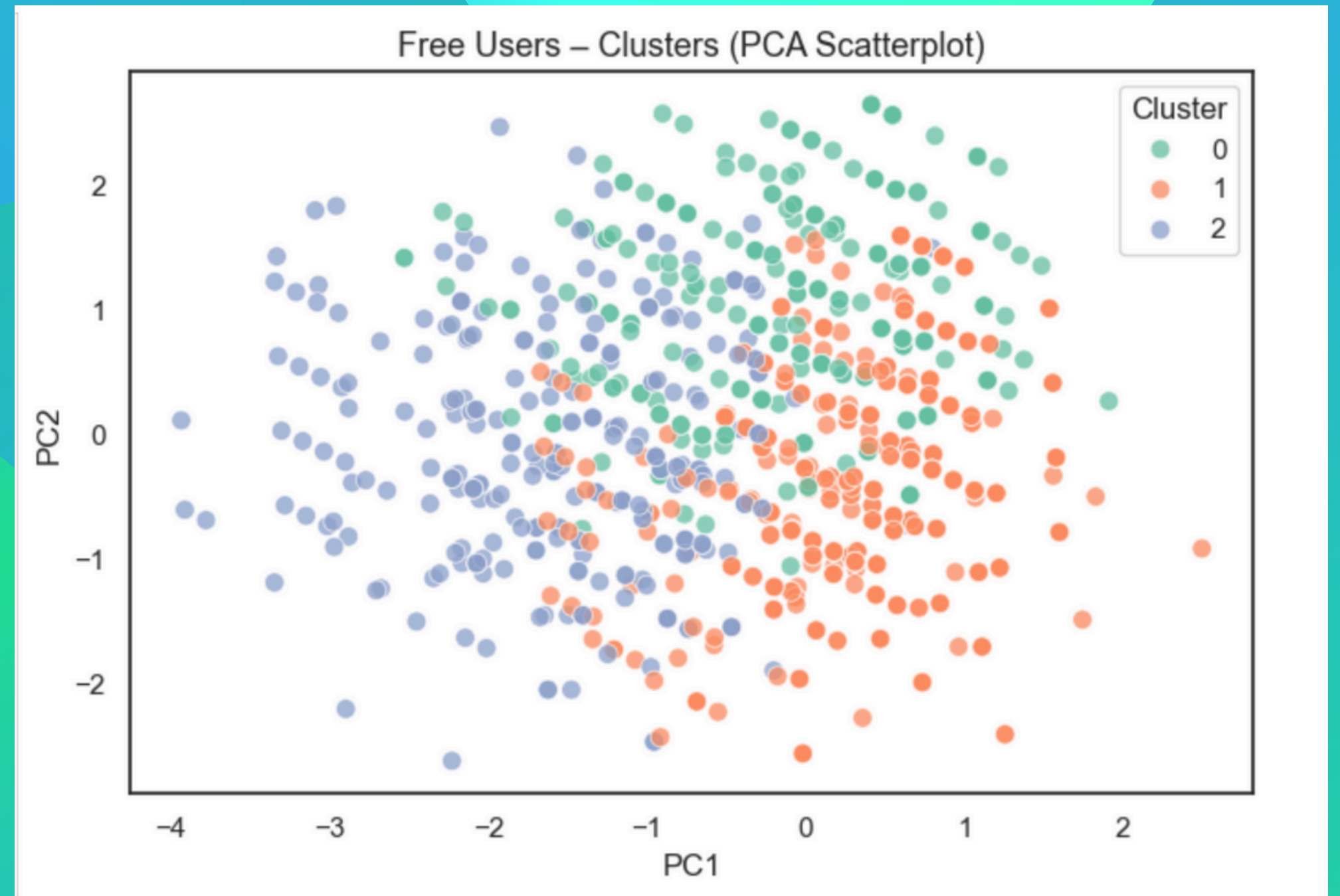
## Free users – Scatter Plot

**PC1** captures overall engagement level, separating frequent, long-term users from lighter, casual listeners.

**PC2** reflects differences in listening style, such as time-of-day preferences, mood, or usage context.

Three free-user groups:

- 1) casual light listeners
- 2) more active routine users
- 3) unique or irregular listening patterns





# | Summary of Results

## *Premium Users*

### Cluster 0

#### **"The Wanderers"**

- Relatively new Premium users with shorter usage periods
- Lower willingness to continue their subscription
- Uses Spotify in lighter, more casual situations

**Low engagement Premium users, high risk of churn**

### Cluster 1

#### **"The Nighttime Loyalists"**

- Strong, loyal Premium users
- Long usage history
- Extremely high willingness to keep Premium
- Listens mostly during night time
- High context dependence

**Most engaged, high retention user group**

### Cluster 2

#### **"The Everyday Groovers"**

- Long-term Premium users
- Highest usage period
- Strong willingness to renew
- Rely less on special contexts

**Steady, habitual users**

# | Summary of Results

## *Free Users*

### Cluster 0

#### **"The Almost-Premiums"**

- Moderate usage history
- Highest willingness to upgrade
- High engagement with Spotify

**High engagement users,  
strong upgrade potential**

### Cluster 1

#### **"The Free-Tier Fanatics"**

- Relatively long usage period
- Zero willingness to upgrade
- High context dependence

**Highly active but  
resistant to Premium**

### Cluster 2

#### **"The Casual Dabblers"**

- Shortest usage periods
- Low willingness to upgrade
- Uses Spotify in habitual, casual situations

**Low engagement users,  
low-conversion to Premium**

# | Managerial Implications

## **Premium user**

*Cluster 0 (Lowest Engagement) – Short Usage History & Lower Renewal Intention*

- Weak perceived benefit from Premium
  - Gap in Spotify's onboarding experience
  - Haven't experienced enough Premium-specific differentiators
- Premium is not integrated into parts of their daily life
  - Lack strong emotional or contextual attachment to Spotify

## **Free user**

*Cluster 0 (Promising Conversion Target) – Moderate Usage History & Highest Premium Interest*

- Uses Spotify in multiple daily situations
  - Formed habits around Spotify, reducing churn risk
- High psychological readiness to convert



# | Recommendations

## **Premium User Retention Strategy**

*Cluster 0 (Lowest Engagement)*

### **1. Targeted campaigns**

- “Getting Started” tutorials
- Playlist recommendations
- Habit-building nudges to increase early engagement

### **2. Context-based push notifications**

### **3. Improve early lifecycle experience**

- Strengthen first 30-day experience with guided playlist creation or AI – personalized mixes

## **Free user Conversion Strategy**

*Cluster 0 (Promising Conversion Target)*

### **1. Premium nudges**

### **2. Showcase better benefits**

- Offline listening for commuting
- High-Quality audio for focused evening sessions
- Unlimited skips for mood dependent playlists

### **3. Short-term Premium trials during high engagement periods**

- Tailored messaging to their behavior: “You’ve been listening a lot – want to try uninterrupted Premium for the next 48 hours?”



# | Limitations

A few problems we encountered were:

- Lack of a significant amount of data
  - This can lead to overfitting
- The data does not reveal if a user later became a premium subscriber or if a premium subscriber cancelled their subscription
- The responses in the dataset are 75% female and 22% male which can lead to biased and female dominant insights



# Thank You

Dream Stream Team – Team 18B

**Shivani Vallamdas, Christie Shin, Gema Zhu, Vishal Srivastava, Shuai Zhao**

BANA 205: Foundations of Marketing