

# **Anàlisi de la salut mental dels estudiants mitjançant Machine Learning**

Joan Colillas, Ilias Dahchouri i Bernat Vidal

**UAB**  
Universitat Autònoma  
de Barcelona

# Table of contents

**01**

**Presentació del  
projecte**

**02**

**Dataset**

**03**

**Estudi de la  
importància de les  
variables**

**04**

**Clusterings**

**05**

**Resultats i anàlisis**

**06**

**Conclusions**

**01**

# **Presentació del projecte**

# Objectiu

- Què volem aconseguir?
- Com ho volem aconseguir?
- Quines eines utilitzarem?

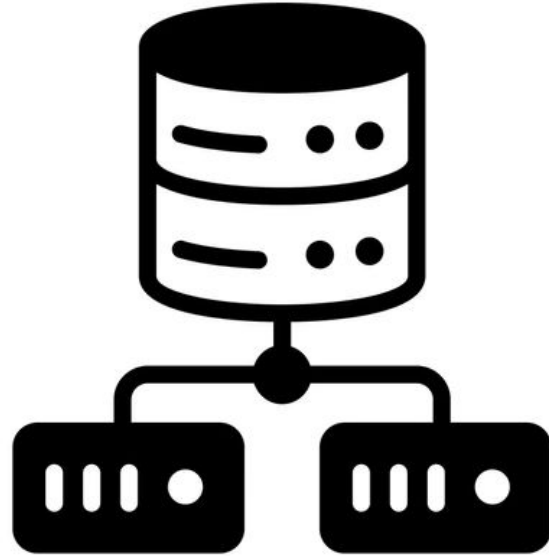


**02**

# **Dataset**

# Características del dataset

- Dimensions
- Preprocessament



# Variables principals i descripcions

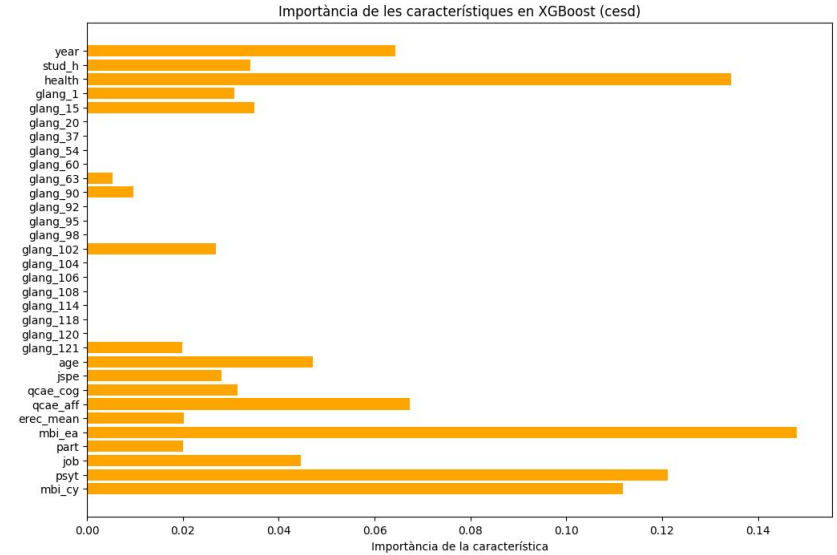
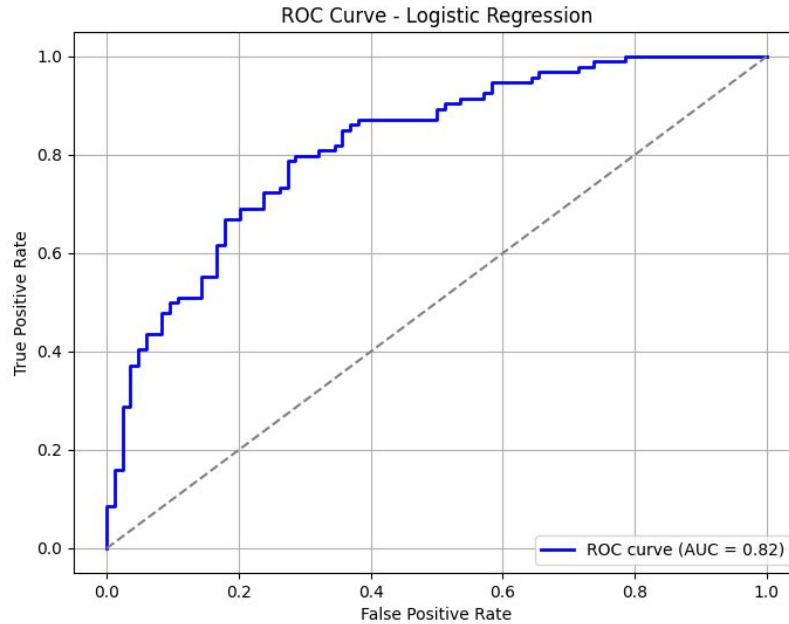
- Variables psicològiques objectiu:
  - cesd (depressió)
  - stai\_t (ansietat)
  - mbi\_ex (cansament emocional)
- Variables numèriques
  - age (edat)
  - stud\_h (hores d'estudi per setmana)
  - mbi\_ea (eficàcia acadèmica)
- Variables ordinals:
  - sex (gènere)
  - part (parella)
  - glang (llengua materna)
- Variables nominals:
  - year (any acadèmic)
  - health (nivell de satisfacció amb la salut)

**03**

# **Estudi de la importància de les variables**

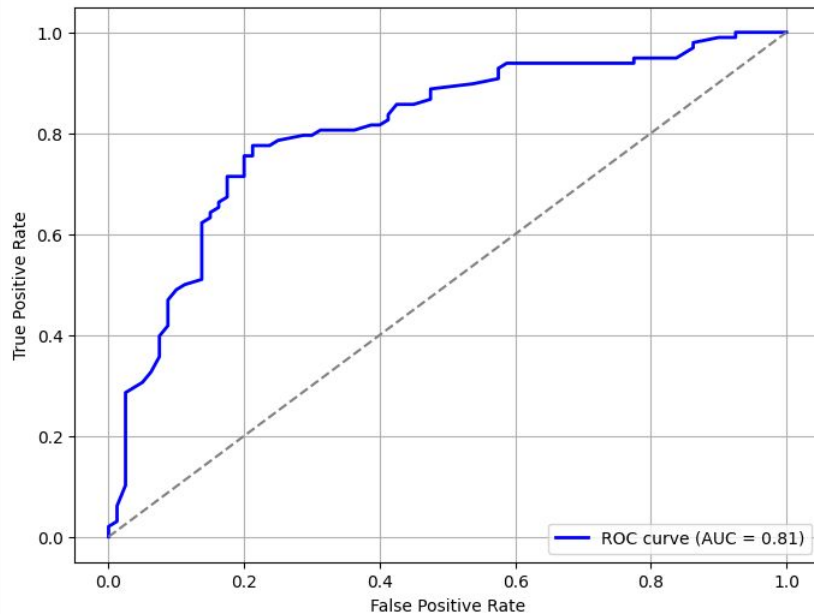


# XGBoost

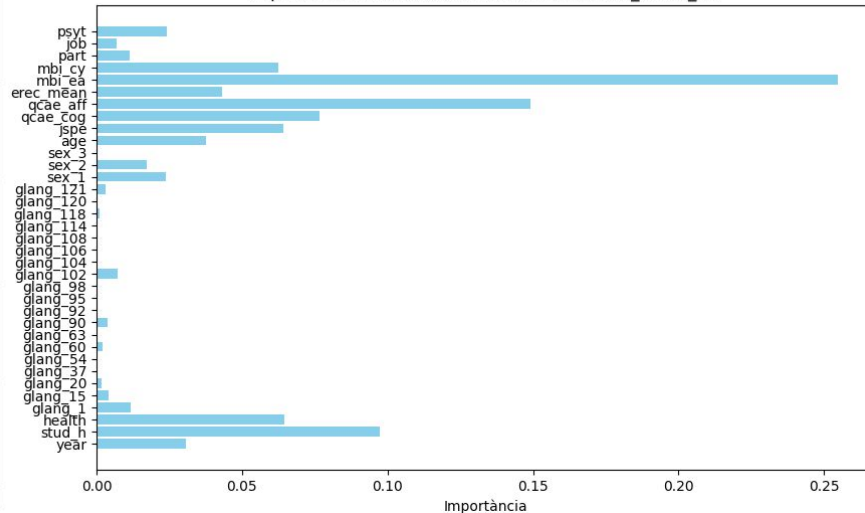


# Random Forest

ROC Curve - Random Forest Classifier



Importància de les variables (sense ccsd, stai\_t, mbi\_ex)

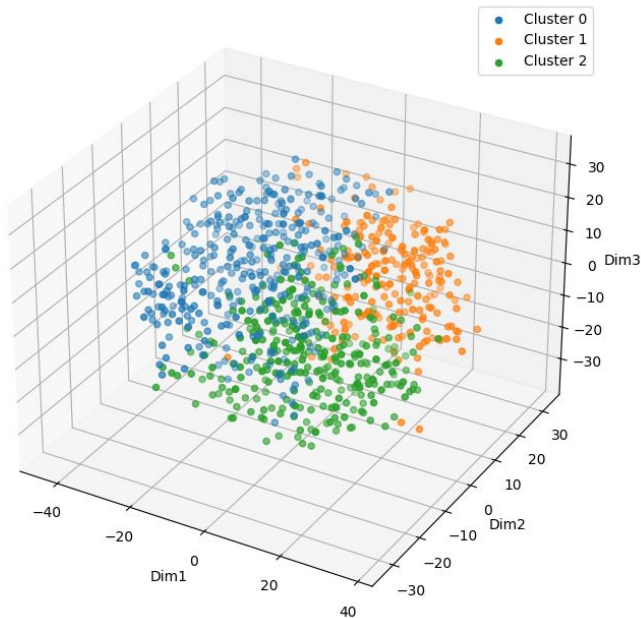


**04**

# **Clusterings**

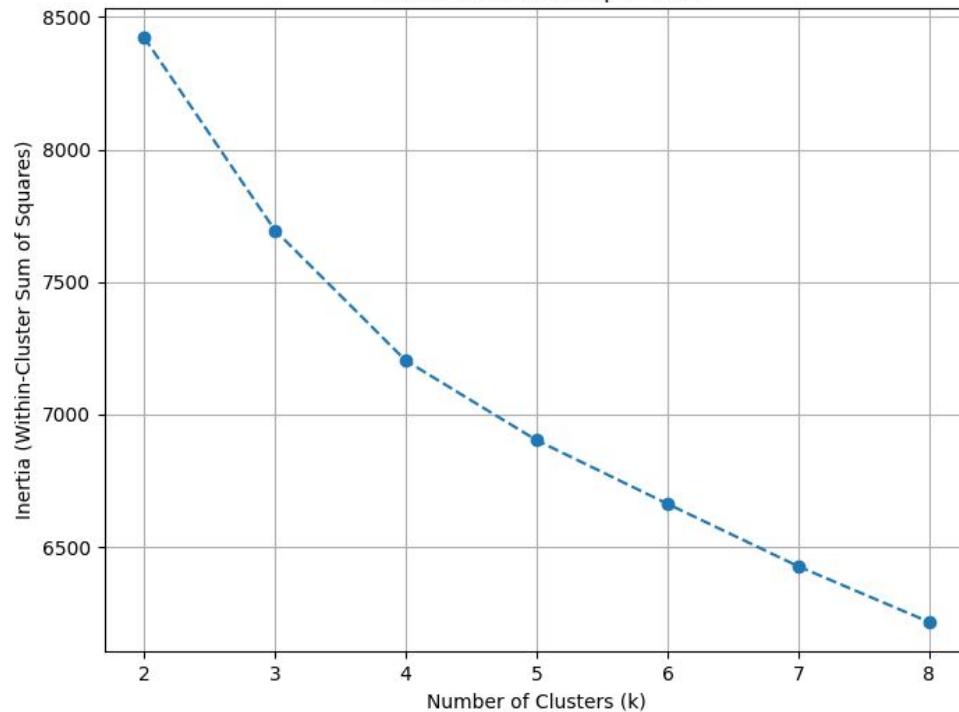
# Kmeans

Clusters Visualitzats amb t-SNE (3D)



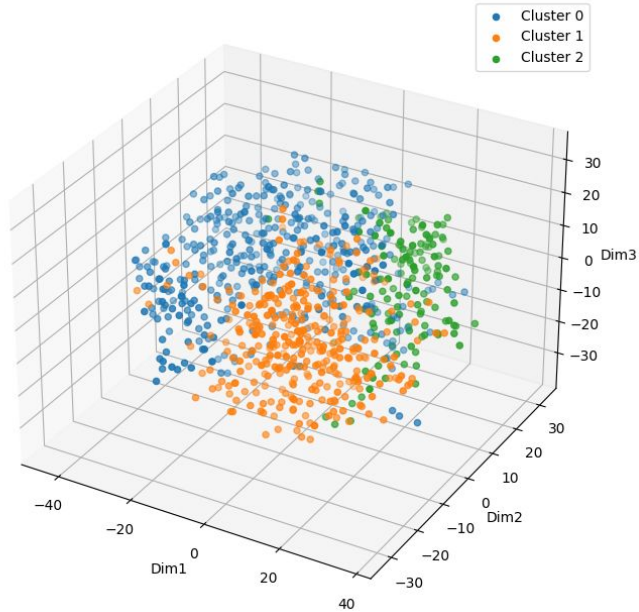
Millor  $k$  silhouette: 3

Elbow Method for Optimal  $k$



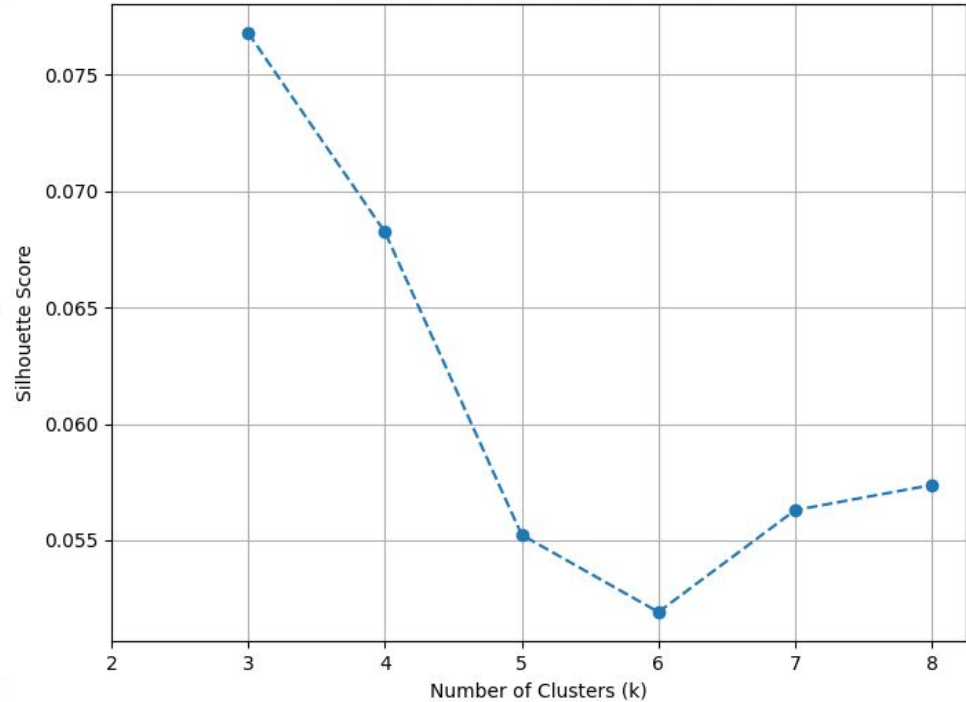
# Agglomerative

Clusters Visualitzats amb t-SNE (3D)



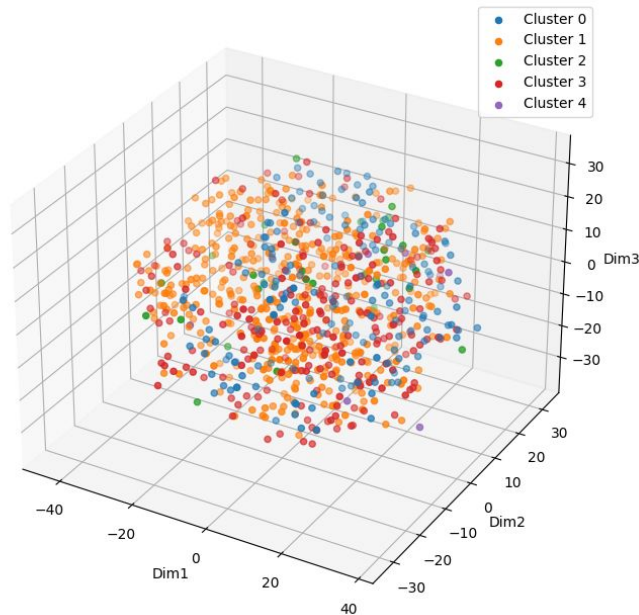
**Millor k silhouette: 3**

Silhouette Score per a diferents k



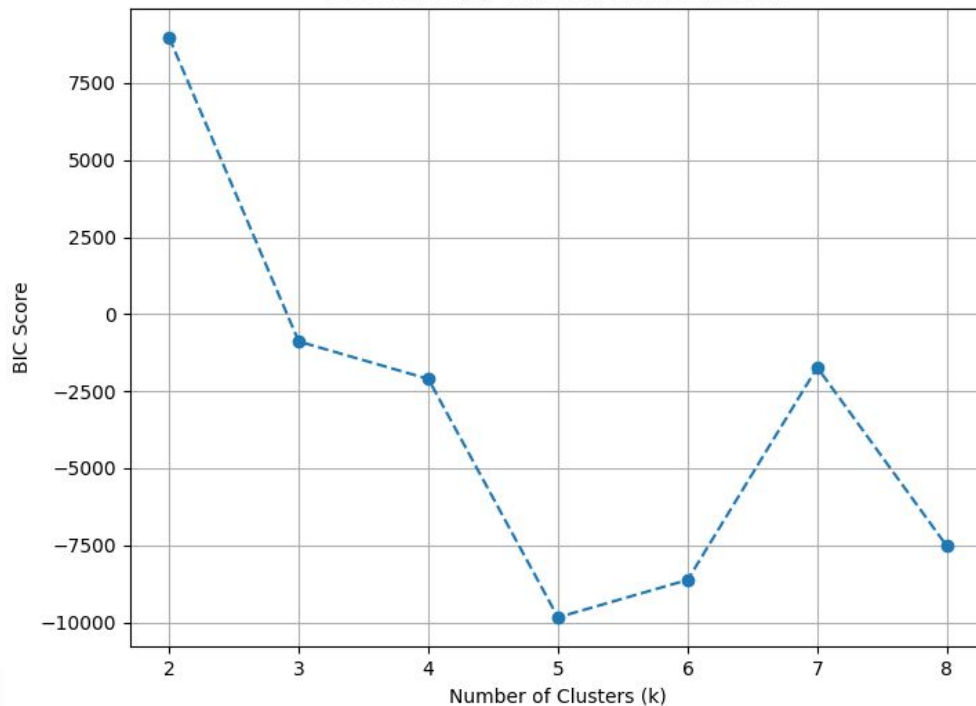
# Gmm (Gaussian Mixture Model)

Clusters Visualitzats amb t-SNE (3D)



**Millor k segons BIC: 5**

BIC Scores for Gaussian Mixture Models

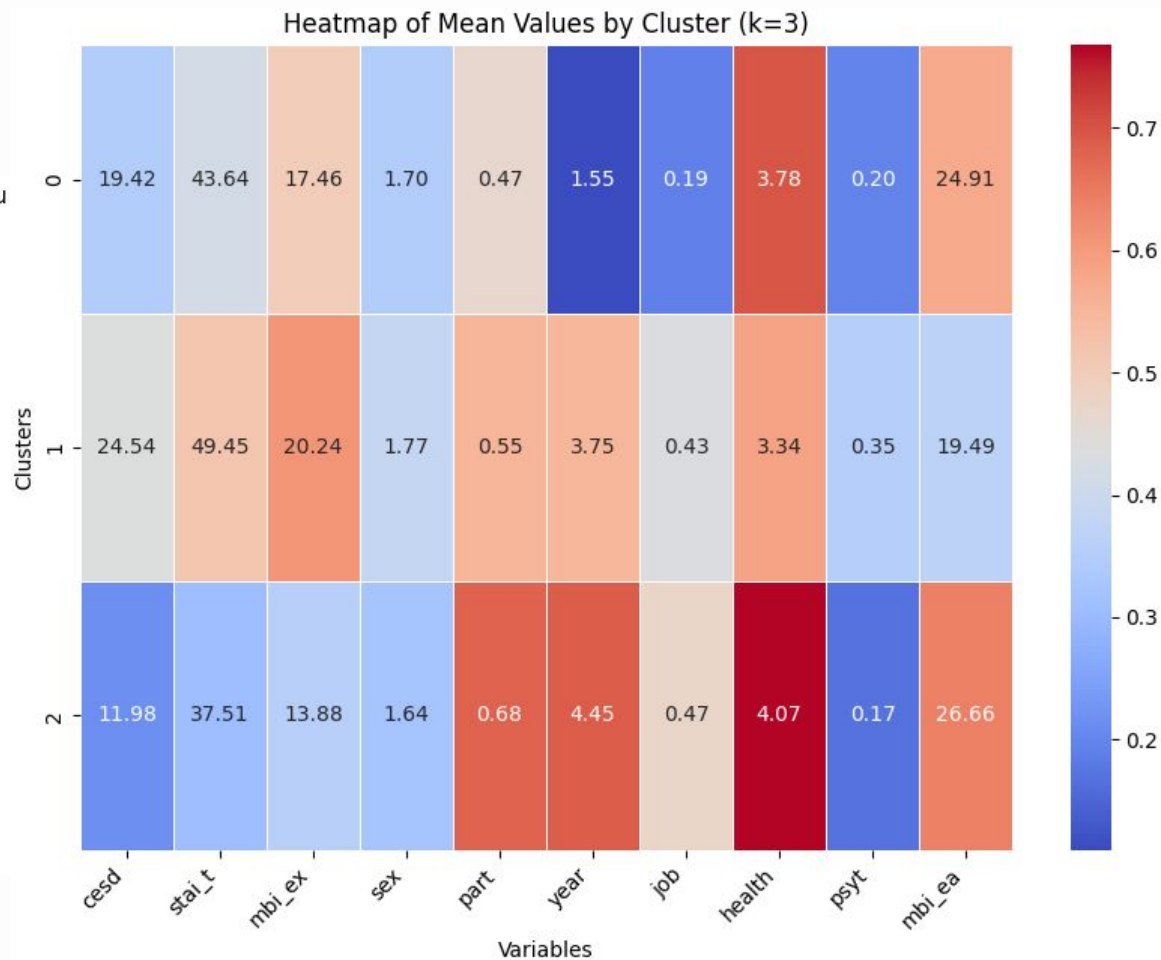
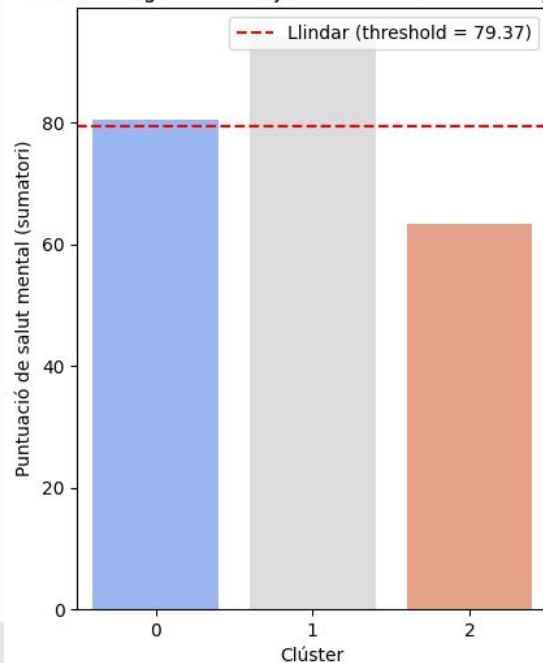


**05**

# **Resultats i anàlisis**

# Kmeans

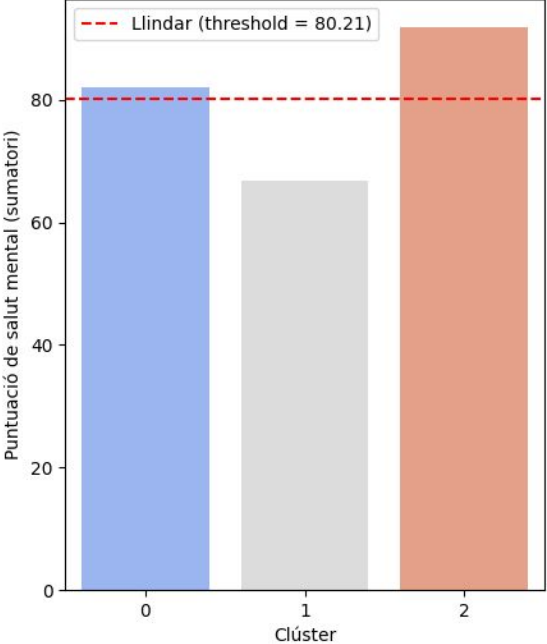
Clusters segons les mitjanes de les variables objectiu



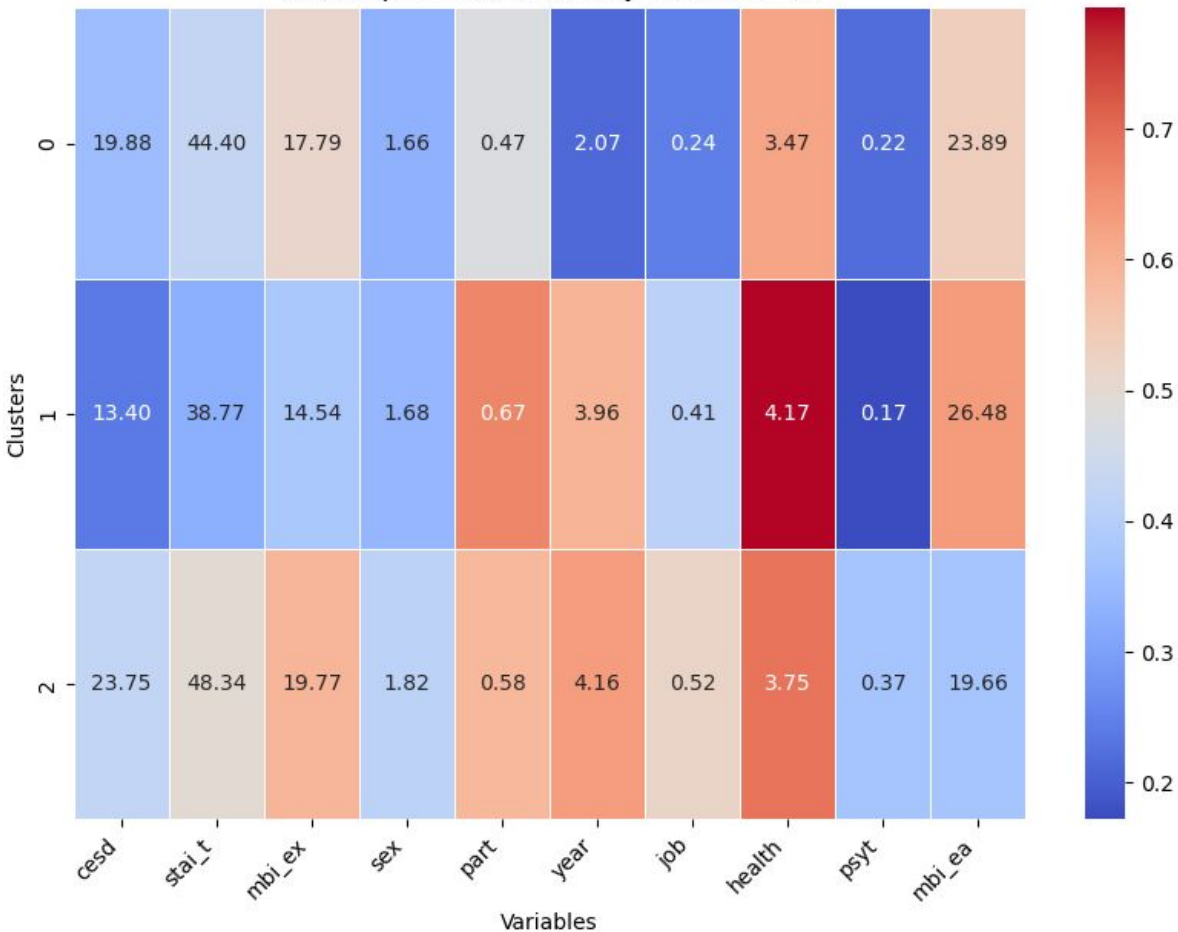


# Agglomerative

Clusters segons les mitjanes de les variables objectiu

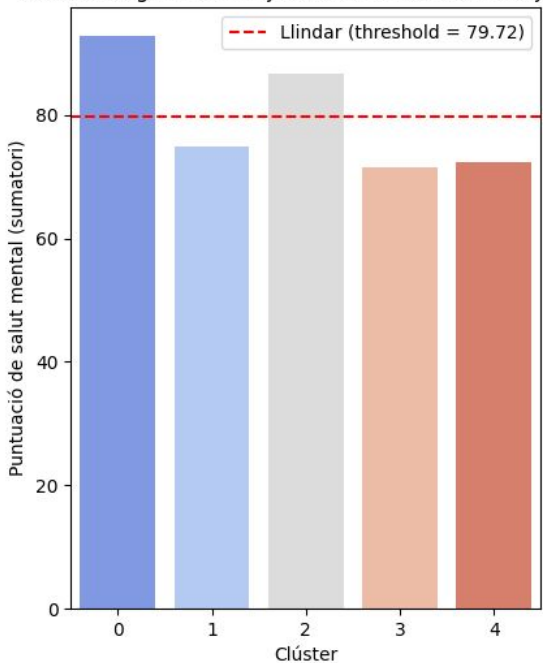


Heatmap of Mean Values by Cluster (k=3)

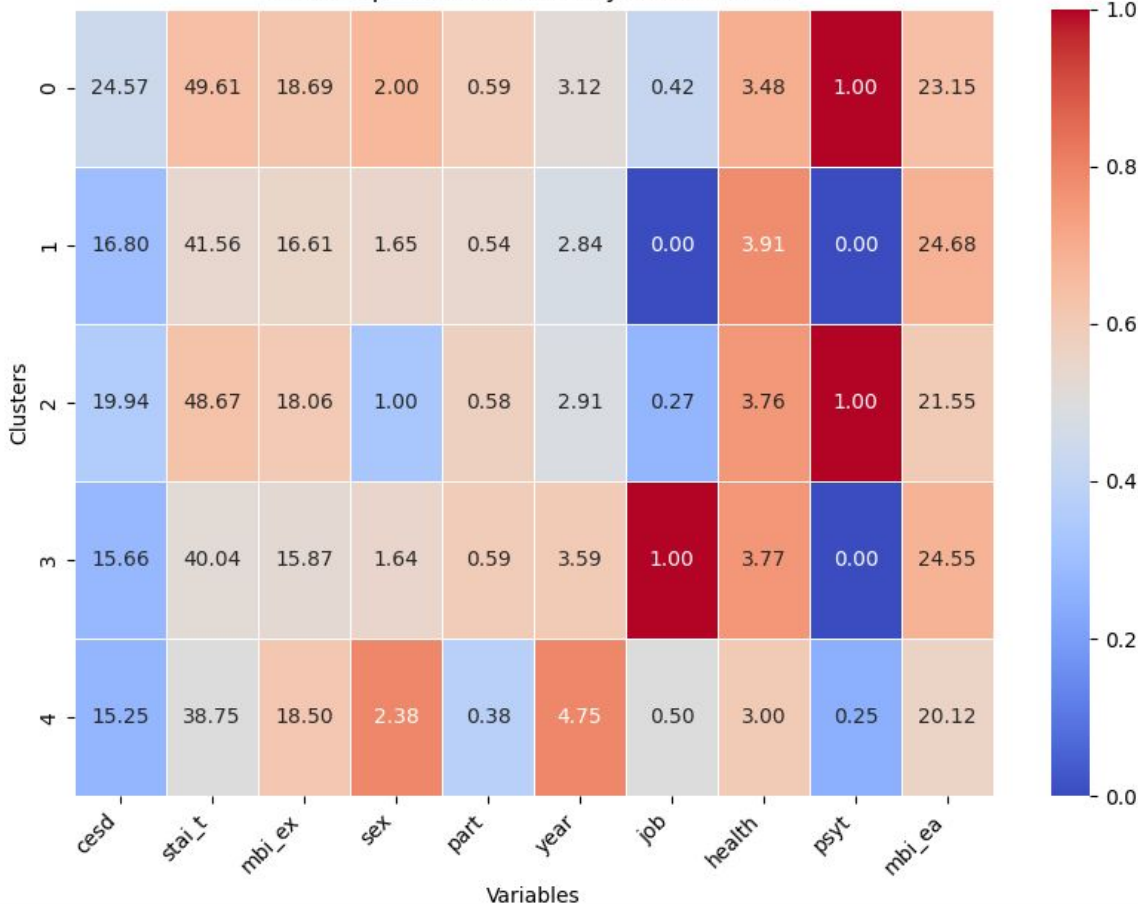


# Gaussian Mixture Models (GMM)

Clusters segons les mitjanes de les variables objectiu



Heatmap of Mean Values by Cluster (k=5)



**06**

# **Conclusions**

# **Moltes gràcies per l'atenció**

Teniu cap pregunta?