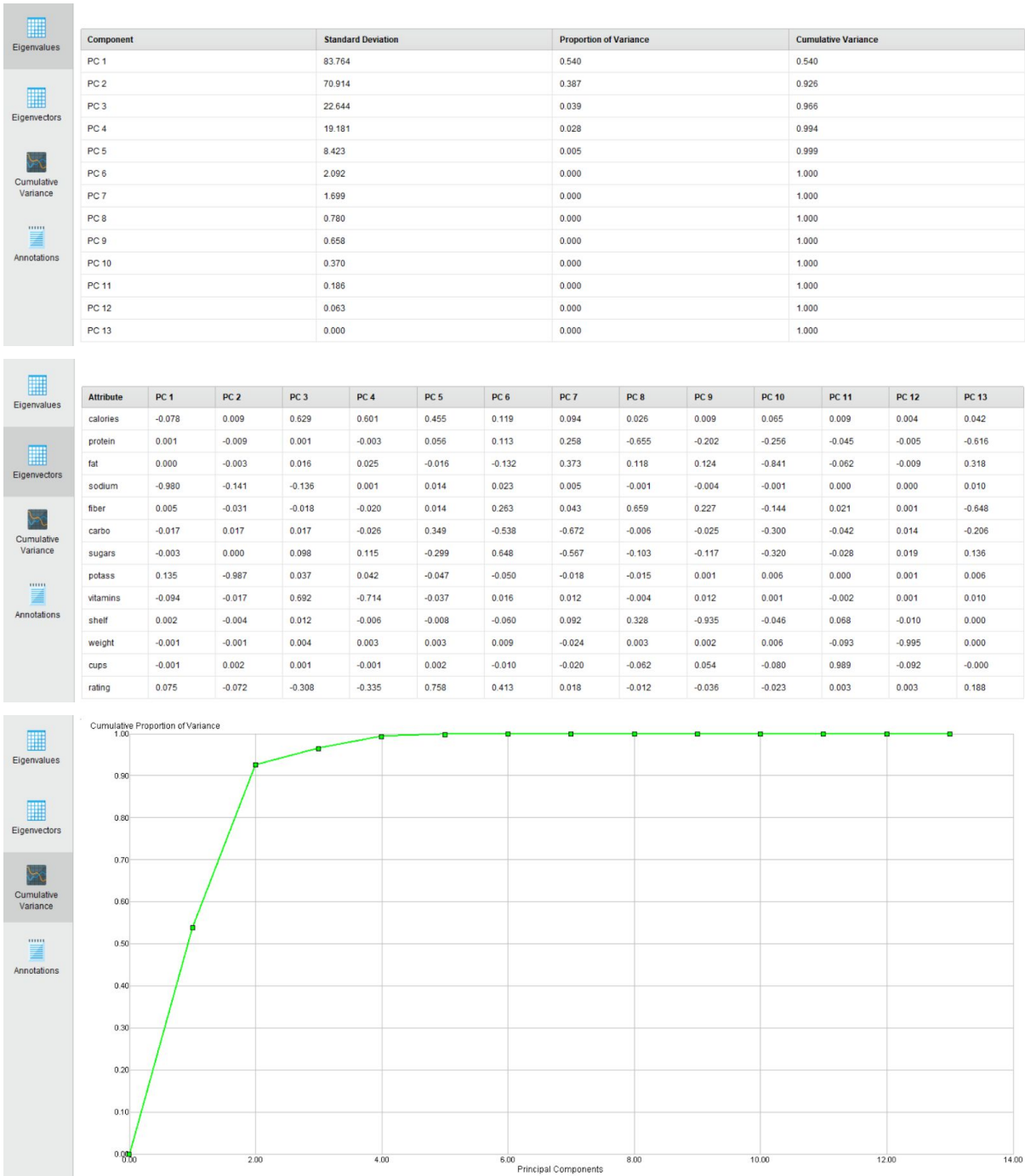


TA3 - Enzo Cozza - Agustín Fernández

Ejercicio 1

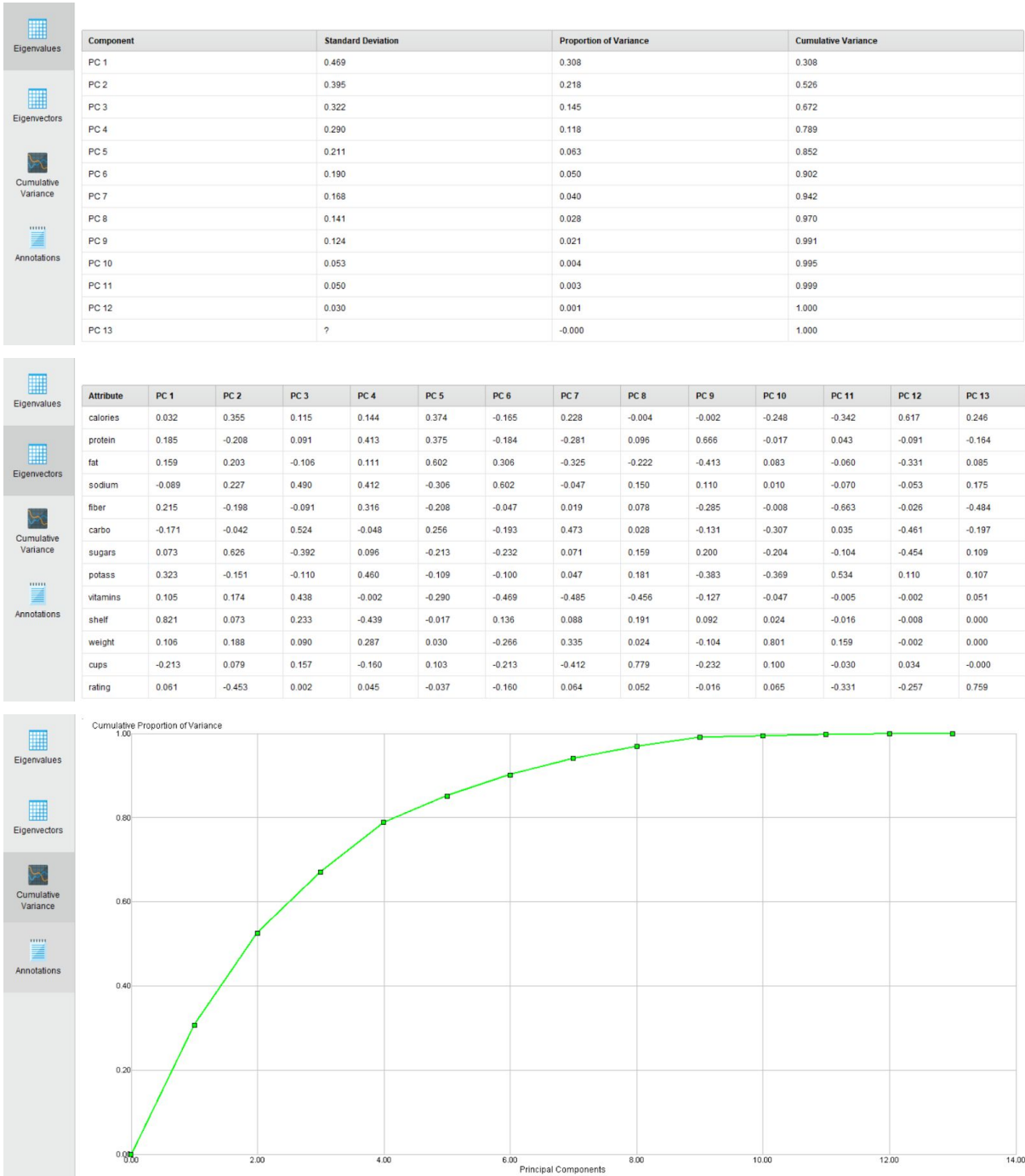
Contexto: los datos con los que se cuenta fueron recaudados por alumnos, que recuperaron información nutricional de las etiquetas de paquetes de cereales de un supermercado.



De los eigenvectores obtenidos para PC1, PC2 y PC3 se puede obtener que los atributos que más impactan son sodio, potasio, calorías, vitaminas y rating.

Estos atributos identificados como más significativos son en verdad los que tienen los rangos más grandes. Se sabe que PCA es sensible a la escala de los datos, por lo que se debería normalizar los datos en la etapa de preparación de los mismos.

Normalizando:



Los componentes principales necesarios para cubrir el 95% de la varianza son los PC 1 al 8. Los atributos que más impactan en estos componentes son: estante, potasio, fibra, azúcares, rating, calorías, sodio, carbo, vitaminas, proteínas, grasas, tazas.

Con estos resultados, todos los atributos menos el peso tienen un gran impacto para obtener la varianza deseada. Se tienen 12 atributos luego de aplicarlo, dejando de percibirse una notoria reducción en la cantidad atributos. En cambio, sin normalizar los datos, se obtuvieron 5 atributos de gran impacto, donde sí se capta una significativa mejora. Sin embargo, esto termina siendo así ya que PCA se ve fuertemente afectado por las escalas de los atributos, y tal es así que los 5 atributos obtenidos resultaron ser los 5 de mayor escala en el dataset, por lo que no se puede concluir nada certero.