

# Ciencia de Datos: Visualización de la Información

UNAM-IIMAS, Semestre 2023-2

## Tarea-04: Reducción de dimensionalidad e interpolación

### Actividad 1

Utilice el conjunto de datos [spotify.csv](#) para análisis. Los datos contienen métricas numéricas generadas por spotify que miden la capacidad de baile (danceability), el estado de ánimo (mood), la vivacidad (liveness), etc. de las canciones. Los datos también contienen el título y el artista de las canciones.

- Primero puede importar todos los módulos necesarios (Python) y generar una tabla a analizar.
- Utilice un *diagrama de dispersión* basado en los valores de valencia (valence) y bailabilidad (danceability) de las canciones. Además, también puede usar la *regresión lineal* para encontrar su correlación.
- Muestre un gráfico que ilustre la *distribución* de las canciones en función de sus niveles de energía (energy), así como otro gráfico que ilustre el número de canciones encontradas en todos los valores de valencia (valence) y bailabilidad (danceability).

A continuación, genere una lista de rasgos "elegidos" y use el **análisis de componentes principales** (PCA) para reducir las dimensiones de esa lista. Cuanto menor sea la distancia entre dos canciones, mayores serán las similitudes en sus rasgos.

1. Use una lista "elegida" más pequeña y PCA para generar un gráfico de dos dimensiones.
2. Por último, genere un gráfico similar usando t-SNE y otra lista "elegida" en su lugar.

### Actividad 2

Utilice el conjunto de datos [cancer.csv](#) para el análisis. Los datos contienen métricas numéricas sobre análisis de cáncer de mama.

- Muestre una gráfica que indique cuántos casos son benignos (B) y malignos (M).
- Puede convertir el diagnóstico en un número (0 o 1) para facilitar el trabajo. Por ejemplo, para benignos: 0, malignos: 1.
- Realice una búsqueda de correlaciones de acuerdo a los datos "mean", "SE", "worst".
- También realice búsqueda de correlaciones de acuerdo a características.

A partir de las correlaciones observadas, intente eliminar alguna de las características para ver si afecta a la clasificación. En caso de no afectar, sería positivo eliminarlas, ya que reduce el número de características a tratar, lo que es beneficioso de cara al coste computacional. Por ejemplo: radio (radio), perímetro (perimeter) y área (area), Concavidad (Concavity), puntos cóncavos (concave points) y compacidad (compactness).

- Muestre un análisis de **PCA** y **t-SNE** que indique los casos benignos (B) y malignos (M).

### Actividad 3

Considere la siguiente imagen, la cual se puede visualizar en el **archivo adjunto**: [unam.png](#)



- Muestre una *ecualización de histograma* de la imagen [unam.png](#) para mejorar el contraste, es decir, modificar las intensidades de pixel.
- Describa brevemente si hay diferencias con respecto a la imagen original tomando en cuenta el *contraste*.
- Muestre tres regiones diferentes (zoom) de la imagen [unam.png](#) con tres métodos de interpolación distintos (por ejemplo, bilineal, bicubico, gaussiano, etc.).
- Para cada region compare la imagen (zoom) *sin interpolación* con respecto a la imágenes que si están interpoladas y explique brevemente sus observaciones.